

IBM GDPS Family

An Introduction to Concepts and Capabilities



David Clitherow
Sim Schindel
John Thompson
Marie-France Narbey



z Systems



International Technical Support Organization

**IBM GDPS Family: An Introduction to
Concepts and Capabilities**

April 2017

Note: Before using this information and the product it supports, read the information in “Notices” on page xi.

Thirteenth Edition (April 2017)

This edition applies to Version 3, Release 14, Modification 0 of the GDPS family of offerings.

© Copyright International Business Machines Corporation 2005, 2017. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Notices	xi
Trademarks	xii
Preface	xiii
Authors	xiii
Now you can become a published author, too	xiv
Comments welcome	xv
Stay connected to IBM Redbooks	xv
Summary of changes	xvii
March 2017, Thirteenth Edition	xvii
June 2016, Twelfth Edition	xviii
June 2015, Eleventh Edition	xviii
August 2014, Update to Tenth Edition	xix
October 2013, Update to Ninth Edition	xx
March 2013, Ninth Edition	xx
July 2012, Eighth Edition	xxi
June 2011, Seventh Edition	xxiii
August 2010, Sixth Edition	xxv
September 2009, Fifth Edition	xxv
September 2008, Fourth Edition	xxvii
March 2007, Third Edition	xxvii
December 2005, Second Edition	xxviii
Chapter 1. Introduction to business resilience and the role of GDPS	1
1.1 Objective	2
1.2 Layout of this book	2
1.3 IT resilience	2
1.3.1 Disaster recovery	3
1.3.2 The next level	4
1.3.3 Other considerations	6
1.4 Characteristics of an IT resilience solution	7
1.5 GDPS offerings	8
1.6 Automation and disk replication compatibility	11
1.7 Summary	12
Chapter 2. Infrastructure planning for availability and GDPS	13
2.1 Parallel Sysplex overview	14
2.1.1 Maximizing application availability	14
2.1.2 Multisite sysplex considerations	15
2.2 Data consistency	17
2.2.1 Dependent write logic	18
2.3 Synchronous versus asynchronous data transfer	19
2.4 Data replication technologies	22
2.4.1 PPRC (IBM Metro Mirror)	23
2.4.2 XRC (z/OS Global Mirror)	27
2.4.3 Global Mirror	32
2.4.4 Combining disk remote copy technologies for CA and DR	35
2.4.5 IBM software replication products	36

2.5	Tape resident data	38
2.6	FlashCopy	38
2.7	Automation	41
2.7.1	Recovery time objective	41
2.7.2	Operational consistency	41
2.7.3	Skills impact	42
2.7.4	Summary	42
2.8	Flexible server capacity	42
2.8.1	Capacity Backup upgrade	42
2.8.2	On/Off Capacity on Demand	43
2.8.3	GDPS CBU and On/Off CoD handling	43
2.9	Cross-site connectivity considerations	44
2.9.1	Server-to-disk links	44
2.9.2	Data replication links	45
2.9.3	Coupling links	45
2.9.4	Server Time Protocol	46
2.9.5	XCF signaling	47
2.9.6	HMC and consoles	47
2.9.7	Connectivity options	47
2.9.8	Single points of failure	49
2.10	Testing considerations	49
2.11	Summary	51
Chapter 3.	GDPS/PPRC	53
3.1	Introduction to GDPS/PPRC	54
3.1.1	Protecting data integrity and data availability with GDPS/PPRC	54
3.1.2	Protecting tape data	65
3.1.3	Protecting distributed (FB) data	66
3.1.4	Protecting other CKD data	67
3.2	GDPS/PPRC configurations	68
3.2.1	Controlling system	68
3.2.2	Single-site workload configuration	70
3.2.3	Multisite workload configuration	72
3.2.4	Business Recovery Services (BRS) configuration	72
3.2.5	GDPS/PPRC in a 3-site or 4-site configuration	74
3.2.6	GDPS/PPRC in a single site	74
3.2.7	Other considerations	74
3.3	GDPS/PPRC management of distributed systems and data	75
3.3.1	Multiplatform Resiliency for z Systems (also known as xDR)	75
3.3.2	Distributed Cluster Management	75
3.3.3	IBM zEnterprise BladeCenter Extension (zBX) hardware management	75
3.4	Management of z/OS systems outside of the GDPS sysplex	76
3.4.1	z/OS Proxy disk and disk subsystem sharing	78
3.5	Managing the GDPS/PPRC environment	78
3.5.1	NetView interface	78
3.5.2	GDPS scripts	85
3.5.3	System Management actions	91
3.6	GDPS/PPRC monitoring and alerting	91
3.6.1	GDPS/PPRC health checks	92
3.7	Other facilities related to GDPS	93
3.7.1	HyperSwap coexistence	93
3.7.2	Reduced impact initial copy and resynchronization	94
3.7.3	Reserve Storage Pool	95

3.7.4 Query Services	95
3.7.5 Concurrent Copy cleanup	95
3.8 GDPS/PPRC flexible testing and resync protection	96
3.8.1 Use of space-efficient FlashCopy volumes	96
3.9 GDPS tools for GDPS/PPRC	97
3.10 GDPS/PPRC co-operation with GDPS/Active-Active	98
3.11 Services component	98
3.12 GDPS/PPRC prerequisites	99
3.13 Comparison of GDPS/PPRC versus other GDPS offerings	99
3.14 Summary	100
Chapter 4. GDPS/PPRC HyperSwap Manager	103
4.1 Introduction to GDPS/PPRC HM	104
4.1.1 Protecting data integrity and data availability with GDPS/PPRC HM	104
4.1.2 Protecting distributed (FB) data	116
4.1.3 Protecting other CKD data	116
4.2 GDPS/PPRC HM configurations	117
4.2.1 Controlling system	117
4.2.2 GDPS/PPRC HM in a single site	119
4.2.3 GDPS/PPRC HM in a 2-site configuration	120
4.2.4 GDPS/PPRC HM in a 3-site configuration	120
4.2.5 Other important considerations	121
4.3 Managing the GDPS/PPRC HM environment	121
4.3.1 NetView interface	121
4.3.2 NetView commands	125
4.4 GDPS/PPRC HM monitoring and alerting	126
4.4.1 GDPS/PPRC HM health checks	127
4.5 Other facilities related to GDPS	128
4.5.1 HyperSwap coexistence	128
4.5.2 GDPS/PPRC HM reduced impact initial copy and resynchronization	129
4.5.3 Reserve Storage Pool	130
4.5.4 GDPS/PPRC HM Query Services	130
4.5.5 Concurrent Copy cleanup	130
4.6 GDPS/PPRC HM flexible testing and resync protection	131
4.6.1 Use of space-efficient FlashCopy volumes	131
4.7 GDPS tools for GDPS/PPRC HM	132
4.8 Services component	133
4.9 GDPS/PPRC HM prerequisites	133
4.10 Comparison of GDPS/PPRC HM to other GDPS offerings	134
4.11 Summary	135
Chapter 5. GDPS/XRC	137
5.1 Introduction to GDPS/XRC	138
5.1.1 Protecting data integrity	138
5.2 GDPS/XRC configuration	140
5.2.1 GDPS/XRC in a 3-site configuration	141
5.3 GDPS/XRC management of distributed systems and data	141
5.4 Managing the GDPS environment	142
5.4.1 NetView interface	142
5.4.2 GDPS scripts	146
5.4.3 System management actions	150
5.5 GDPS/XRC monitoring and alerting	151
5.5.1 GDPS/XRC health checks	152

5.6 Other facilities related to GDPS	154
5.6.1 FlashCopy disk definition in the GDPS systems	154
5.6.2 GDPS/XRC FlashCopy locking	154
5.6.3 GDPS/XRC Configuration checking	154
5.6.4 Vary-After-Clip automation	155
5.6.5 GDPS use of the XRC offline volume support	155
5.6.6 Query Services	156
5.6.7 Easy Tier Heat Map Transfer	156
5.7 Flexible testing	157
5.8 GDPS tools for GDPS/XRC	158
5.9 Services component	158
5.10 GDPS/XRC prerequisites	159
5.11 Comparison of GDPS/XRC versus other GDPS offerings	159
5.12 Summary	161
Chapter 6. GDPS/Global Mirror	163
6.1 Introduction to GDPS/Global Mirror	164
6.1.1 Protecting data integrity	164
6.2 GDPS/Global Mirror configuration	165
6.2.1 GDPS/GM in a 3-site or 4-site configuration	168
6.2.2 Other considerations	168
6.3 GDPS/GM management for distributed systems and data	169
6.4 Managing the GDPS environment	170
6.4.1 NetView panel interface	170
6.4.2 System Management actions	178
6.5 GDPS/GM monitoring and alerting	179
6.5.1 GDPS/GM health checks	180
6.6 Other facilities related to GDPS	181
6.6.1 GDPS/GM Copy Once facility	181
6.6.2 GDPS/GM Query Services	182
6.6.3 Global Mirror Monitor integration	182
6.6.4 Easy Tier Heat Map Transfer	182
6.7 Flexible testing	183
6.7.1 Use of space-efficient FlashCopy	184
6.7.2 Creating a test copy using GM CGPause and testing on isolated disks	184
6.8 GDPS tools for GDPS/GM	185
6.9 Services component	185
6.10 GDPS/GM prerequisites	186
6.11 Comparison of GDPS/GM versus other GDPS offerings	186
6.12 Summary	188
Chapter 7. GDPS/MTMM	189
7.1 Introduction to GDPS/MTMM	190
7.1.1 Protecting data integrity and data availability with GDPS/MTMM	191
7.1.2 Protecting other CKD data	203
7.2 GDPS/MTMM configurations	204
7.2.1 Controlling system	205
7.2.2 Single-site workload configuration	206
7.2.3 Multisite workload configuration	208
7.2.4 Business Recovery Services (BRS) configuration	208
7.2.5 Combining GDPS/MTMM with GDPS/XRC	209
7.2.6 Combining GDPS/MTMM with GDPS/GM in a 4-site configuration	210
7.2.7 Other considerations	210

7.3	Multiplatform Resiliency for System z (also known as xDR)	210
7.4	Managing the GDPS environment	211
7.4.1	NetView interface	211
7.4.2	GDPS scripts	216
7.4.3	System Management actions	221
7.5	GDPS/MTMM monitoring and alerting	222
7.5.1	GDPS/MTMM health checks	223
7.6	Other facilities related to GDPS	224
7.6.1	HyperSwap and TDMF coexistence	224
7.6.2	Reduced impact initial copy and resynchronization	225
7.6.3	Concurrent Copy cleanup	225
7.6.4	Easy Tier Heat Map Transfer	226
7.7	GDPS/MTMM flexible testing and resync protection	226
7.7.1	Use of space-efficient FlashCopy volumes	227
7.8	GDPS tools for GDPS/MTMM	227
7.9	Services component	228
7.10	GDPS/MTMM prerequisites	228
7.11	Comparison of GDPS/MTMM versus other GDPS offerings	228
7.12	Summary	230
Chapter 8.	GDPS/Active-Active solution	231
8.1	Overview of GDPS/Active-Active	232
8.1.1	Positioning GDPS/Active-Active	232
8.1.2	GDPS/Active-Active sites concept	233
8.2	GDPS/Active-Active solution products	235
8.2.1	The GDPS/Active-Active product	236
8.2.2	Tivoli NetView for z/OS	237
8.2.3	IBM Tivoli Monitoring	237
8.2.4	System Automation for z/OS	238
8.2.5	IBM Multi-site Workload Lifeline for z/OS	238
8.2.6	Middleware	239
8.2.7	Replication software	239
8.2.8	Other optional components	240
8.3	GDPS/Active-Active environment	240
8.3.1	GDPS/Active-Active: A closer look	244
8.3.2	Considerations for other non-Active-Active workloads	247
8.4	GDPS/Active-Active functions and features	248
8.4.1	GDPS/Active-Active web interface	249
8.4.2	GDPS/Active-Active monitoring and alerting	257
8.4.3	GDPS/Active-Active scripts	259
8.4.4	GDPS/Active-Active Query Services	264
8.5	GDPS/Active-Active co-operation with GDPS/PPRC or GDPS/MTMM	264
8.6	GDPS/Active-Active disk replication integration	267
8.7	Zero Data Loss Configuration	268
8.8	Flexible testing with GDPS/Active-Active	271
8.9	GDPS/Active-Active services	271
8.10	GDPS/Active-Active prerequisites	272
8.11	GDPS/Active-Active comparison to other GDPS offerings	273
8.12	Summary	273
Chapter 9.	GDPS Virtual Appliance	275
9.1	Introduction to the GDPS Virtual Appliance	276
9.2	GDPS Virtual Appliance configuration components	276

9.2.1	GDPS Virtual Appliance	277
9.2.2	Multiplatform Resiliency for z Systems	277
9.3	Protecting data integrity and data availability with the GDPS Virtual Appliance	278
9.3.1	GDPS Freeze function for mirroring failures	278
9.3.2	GDPS HyperSwap function	279
9.3.3	GDPS use of DS8000 functions	281
9.3.4	Protecting secondary disks from accidental update	283
9.4	Managing the GDPS environment	283
9.4.1	GDPS graphic user interface	284
9.4.2	GDPS scripts	289
9.4.3	System Management actions	290
9.5	GDPS monitoring and alerting	290
9.6	Services component	291
9.7	GDPS Virtual Appliance prerequisites	292
9.8	GDPS Virtual Appliance compared to other GDPS offerings	292
9.9	Summary	294
Chapter 10.	GDPS extensions for heterogeneous systems and data	295
10.1	Open LUN Management function	296
10.2	GDPS/PPRC Multiplatform Resiliency for z Systems	299
10.2.1	Guest Linux under z/VM	299
10.2.2	Native Linux on z Systems	303
10.2.3	Support for two GDPS Controlling systems	305
10.2.4	Customization Verification Program	306
10.2.5	xDR Extended Monitoring	307
10.3	Distributed Cluster Management	307
10.3.1	Distributed Cluster Management terminology	308
10.3.2	DCM support for VCS	308
10.3.3	DCM support for SA AppMan	319
10.3.4	GDPS/PPRC Support for IBM zEnterprise BladeCenter Extension (zBX)	329
10.3.5	Summary	329
Chapter 11.	Combining local and metro continuous availability with out-of-region disaster recovery	331
11.1	Introduction	332
11.2	Design considerations	333
11.2.1	Three-copy solutions versus 3-site solutions	333
11.2.2	Multi-target and cascading topologies	336
11.2.3	Four-copy solutions	337
11.2.4	Cost considerations	337
11.2.5	Operational considerations	337
11.3	GDPS Metro/Global Mirror 3-site solution	337
11.3.1	GDPS/MGM 3-site overview	338
11.3.2	GDPS/MGM Site1 failures	342
11.3.3	GDPS/MGM Site2 failures	342
11.3.4	GDPS/MGM region switch and return home	343
11.3.5	Scalability in a GDPS/MGM 3-site environment	343
11.3.6	Other considerations in a GDPS/MGM 3-site environment	344
11.3.7	Managing the GDPS/MGM 3-site environment	344
11.3.8	Flexible testing in a GDPS/MGM 3-site environment	344
11.3.9	GDPS Query Services in a GDPS/MGM 3-site environment	345
11.3.10	Prerequisites for GDPS/MGM 3-site	345
11.3.11	GDPS/Active-Active disk replication integration with GDPS/MGM	346
11.4	GDPS Metro/Global Mirror 4-site solution	346

11.4.1	Benefits of a GDPS/MGM 4-site configuration	349
11.5	GDPS Metro z/OS Global Mirror 3-site solution	349
11.5.1	GDPS/MzGM overview	350
11.5.2	GDPS/MzGM Site1 failures.	351
11.5.3	GDPS/MzGM Site2 failures.	352
11.5.4	GDPS/MzGM region switch and return home	352
11.5.5	Management of the GDPS/MzGM environment	353
11.5.6	Flexible testing of the GDPS/MzGM environment.	353
11.5.7	Prerequisites for GDPS/MzGM	354
11.6	GDPS Metro z/OS Global Mirror 4-site solution	354
11.6.1	Benefits of a GDPS/MzGM 4-site configuration	356
Chapter 12. Sample continuous availability and disaster recovery scenarios		357
12.1	Introduction	358
12.2	Continuous availability in a single data center	358
12.3	DR across two data centers at metro distance	361
12.4	DR and CA across two data centers at metro distance	362
12.4.1	Active/active workload	365
12.5	DR and CA across two data centers at metro distance for z/VM and Linux on z Systems only.	366
12.6	Local CA and remote DR across two data centers at long metropolitan distance . .	368
12.7	DR in two data centers, global distance	370
12.8	Other configurations	371
Glossary		373
Related publications		377
IBM Redbooks publications		377
Other publications		377
Online resources		378
Help from IBM		378
Index		379

Notices

This information was developed for products and services offered in the US. This material might be available from IBM in other languages. However, you may be required to own a copy of the product or product version in that language in order to access it.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, MD-NC119, Armonk, NY 10504-1785, US

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this IBM product and use of those websites is at your own risk.

IBM may use or distribute any of the information you provide in any way it believes appropriate without incurring any obligation to you.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to actual people or business enterprises is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at <http://www.ibm.com/legal/copytrade.shtml>

The following terms are trademarks or registered trademarks of International Business Machines Corporation, and might also be trademarks or registered trademarks in other countries.

AIX®	HyperSwap®	System Storage®
CICS®	IBM®	System z®
DB2®	IBM z Systems®	System z10®
Distributed Relational Database Architecture™	IMS™	System z9®
DS8000®	InfoSphere®	Tivoli®
Easy Tier®	MVS™	VTAM®
Enterprise Storage Server®	NetView®	WebSphere®
FICON®	OMEGAMON®	z Systems®
FlashCopy®	Parallel Sysplex®	z/OS®
GDPS®	RACF®	z/VM®
Geographically Dispersed Parallel Sysplex™	Redbooks®	z/VSE®
Global Technology Services®	Redpaper™	z10™
HACMP™	Redbooks (logo)  ®	z9®
	Resource Link®	zEnterprise®
	System i®	

The following terms are trademarks of other companies:

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product, or service names may be trademarks or service marks of others.

Preface

This IBM® Redbooks® publication presents an overview of the IBM Geographically Dispersed Parallel Sysplex™ (IBM GDPS®) offerings and the roles they play in delivering a business IT resilience solution.

The book begins with general concepts of business IT resilience and disaster recovery, along with issues related to high application availability, data integrity, and performance. These topics are considered within the framework of government regulation, increasing application and infrastructure complexity, and the competitive and rapidly changing modern business environment.

Next, it describes the GDPS family of offerings with specific reference to how they can help you achieve your defined goals for disaster recovery and high availability. Also covered are the features that simplify and enhance data replication activities, the prerequisites for implementing each offering, and tips for planning for the future and immediate business requirements. Tables provide easy-to-use summaries and comparisons of the offerings, and the additional planning and implementation services available from IBM are explained.

Then, several practical client scenarios and requirements are described, along with the most suitable GDPS solution for each case.

The introductory chapters of this publication are intended for a broad technical audience, including IT System Architects, Availability Managers, Technical IT Managers, Operations Managers, System Programmers, and Disaster Recovery Planners. The subsequent chapters provide more technical details about the GDPS offerings, and each can be read independently for those readers who are interested in specific topics. Therefore, if you do read all the chapters, be aware that some information is intentionally repeated.

Authors

This book was produced by a team of specialists from around the world working at the International Technical Support Organization, Poughkeepsie Center.

David Clitherow is a Consulting IT Specialist with IBM Global Technology Services® in the UK. He has over 30 years of experience working directly with IBM customers, focused primarily on the mainframe platform and specializing in high availability. Dave is a member of the GDPS development team, the GDPS Customer Design Council and the Active/Active Sites Design Council.

Sim Schindel is an Executive IT Specialist with over 35 years of experience in IT, working with enterprise clients and systems. Her areas of expertise include IBM Parallel Sysplex and GDPS. Sim is a member of the GDPS development team with primary responsibility for GDPS information development. She is a member of the IBM z® Systems eBusiness Leadership Council and the GDPS Customer Design Council.

John Thompson is a Senior Technical Staff Member with more than 25 years of experience in IBM z/OS® software design and development. He is currently involved in architecture and strategy for IBM z Systems® Business Continuity. John is a member of the IBM System z® eBusiness Leadership Council (zBLC) and serves as a co-leader of the zBLC Business Availability Workgroup. He is also a member of the leadership team for the GDPS Customer Design Council.

Marie-France Narbey is a Certified Project Manager having worked in various areas of IBM System z for the past 16 years. Her System z hardware and large account experience gained through EMEA Product Engineering, Manufacturing support (MOP/DUB), Poughkeepsie Labs, Customer Satisfaction Project Office, pre sales activities and field enablement led her to recently join the GDPS team in the test infrastructure group and as information development team lead.

Thanks to the authors of the previous editions of this book:

- ▶ Brian Cooper
- ▶ Noshir Dhondy
- ▶ Mike Hrencecin
- ▶ Frank Kyne
- ▶ Udo Pimiskern
- ▶ Mark Ratte
- ▶ Gene Sale
- ▶ Sim Schindel

Thanks to the following people for their contributions to this project:

George Kozakos
IBM Australia

Thomas Bueche
IBM Germany

Nick Clayton
IBM UK

Stephen Anania
Charlie Burger
Alan McClure
David Petersen
Judy Ruby-Brown
John Sing
IBM USA

Mike Ebbers
Frank Kyne
Bill White
Keith Winnard
IBM ITSO Poughkeepsie, NY, USA

Now you can become a published author, too

Here's an opportunity to spotlight your skills, grow your career, and become a published author, all at the same time. Join an ITSO residency project and help write a book in your area of expertise, while honing your experience using leading-edge technologies. Your efforts will help to increase product acceptance and customer satisfaction, as you expand your network of technical contacts and relationships. Residencies run from two to six weeks, and you can participate either in person or as a remote resident working from your home base.

Find out more about the residency program, browse the residency index, and apply online:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us.

We want our books to be as helpful as possible. Send us your comments about this book or other IBM Redbooks publications in one of the following ways:

- Use the online **Contact us** review form:

ibm.com/redbooks

- Send your comments by email:

redbook@us.ibm.com

- Mail your comments:

IBM Corporation, International Technical Support Organization
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400

Stay connected to IBM Redbooks

- Find us on Facebook:

<http://www.facebook.com/IBMRedbooks>

- Follow us on Twitter:

<http://twitter.com/ibmredbooks>

- Look for us on LinkedIn:

<http://www.linkedin.com/groups?home=&gid=2130806>

- Explore new Redbooks publications, residencies, and workshops with the IBM Redbooks weekly newsletter:

<https://www.redbooks.ibm.com/Redbooks.nsf/subscribe?OpenForm>

- Stay current on recent Redbooks publications with RSS Feeds:

<http://www.redbooks.ibm.com/rss.html>

Summary of changes

This section describes the technical changes made in this edition of the book and in previous editions. This edition also includes minor corrections and editorial changes that are not identified.

Summary of Changes
for IBM Redbooks publication SG24-6374-12
for IBM GDPS Family: An Introduction to Concepts and Capabilities
as created or updated on April 17, 2017.

March 2017, Thirteenth Edition

New and changed information

- ▶ The following sections are updated or added in support of the new GDPS/MGM Multi-Target 3-site solution:
 - 1.5, “GDPS offerings” on page 8.
 - 3.13, “Comparison of GDPS/PPRC versus other GDPS offerings” on page 99.
 - 4.10, “Comparison of GDPS/PPRC HM to other GDPS offerings” on page 134.
 - 5.11, “Comparison of GDPS/XRC versus other GDPS offerings” on page 159.
 - 6.11, “Comparison of GDPS/GM versus other GDPS offerings” on page 186.
 - 7.11, “Comparison of GDPS/MTMM versus other GDPS offerings” on page 228.
 - 7.2.6, “Combining GDPS/MTMM with GDPS/GM in a 4-site configuration” on page 210.
 - Chapter 11, “Combining local and metro continuous availability with out-of-region disaster recovery” on page 331.
- ▶ Updated section 3.1.4, “Protecting other CKD data” on page 67, to reflect that GDPS/PPRC now provides foreign systems support for Linux on z Systems running as a guest under KVM on z Systems.
- ▶ The following sections are updated/added in support of the new Zero Data Loss specialized configuration for GDPS/Active-Active:
 - 8.3, “GDPS/Active-Active environment” on page 240.
 - 8.7, “Zero Data Loss Configuration” on page 268.
- ▶ Updated section 10.2.1, “Guest Linux under z/VM” on page 299, to reflect that GDPS will now automatically switch the dump volume pointers when an IBM HyperSwap® occurs.
- ▶ Updated section 11.3.1, “GDPS/MGM 3-site overview” on page 338, to remove a reference to the MGM Incremental Resynchronization Tool, which is no longer supported.
- ▶ The GDPS/PPRC, GDPS/PPRC HM, and GDPS Virtual Appliance web interfaces are no longer supported. All references have been removed.
- ▶ All occurrences of the term “Fixed Block Architecture” (FBA) have been removed or changed to “Fixed Block” (FB) for clarity.
- ▶ Several small changes have been made for clarity or to correct minor errors.

June 2016, Twelfth Edition

New and changed information

- ▶ Information regarding the GDPS/PPRC and GDPS/PPRC HM web interfaces has been refreshed in the following sections:
 - “Managing the GDPS/PPRC environment” on page 78.
 - “Managing the GDPS/PPRC HM environment” on page 121.

This is due to the introduction of a new, modernized GDPS graphical user interface for those offerings.
- ▶ 5.6.7, “Easy Tier Heat Map Transfer” on page 156 is new.
- ▶ 6.6.4, “Easy Tier Heat Map Transfer” on page 182 is new.
- ▶ 7.6.4, “Easy Tier Heat Map Transfer” on page 226 is new.
- ▶ The following sections are updated/added in support of the new GDPS/MGM Multi-Target 4-site solution:
 - 7.2.6, “Combining GDPS/MTMM with GDPS/GM in a 4-site configuration” on page 210.
 - 11.4, “GDPS Metro/Global Mirror 4-site solution” on page 346 is updated to describe the new GDPS/MGM 4-site with multi-target configuration.
 - 1.5, “GDPS offerings” on page 8.
 - 6.2.1, “GDPS/GM in a 3-site or 4-site configuration” on page 168.
 - 3.13, “Comparison of GDPS/PPRC versus other GDPS offerings” on page 99.
 - 4.10, “Comparison of GDPS/PPRC HM to other GDPS offerings” on page 134.
 - 5.11, “Comparison of GDPS/XRC versus other GDPS offerings” on page 159.
 - 6.11, “Comparison of GDPS/GM versus other GDPS offerings” on page 186.
 - 7.11, “Comparison of GDPS/MTMM versus other GDPS offerings” on page 228.
- ▶ 11.6, “GDPS Metro z/OS Global Mirror 4-site solution” on page 354 is new.
- ▶ Several small changes have been made for clarity or to correct minor errors.

June 2015, Eleventh Edition

New and changed information

- ▶ All references to the IBM TotalStorage 3494-based Virtual Tape Subsystem are removed, because these devices are no longer supported.
- ▶ GDPS/MTMM and the GDPS Virtual Appliance are added to 1.5, “GDPS offerings” on page 8.
- ▶ Chapter 7, “GDPS/MTMM” on page 189 is new.
- ▶ Several sections are updated in Chapter 2, “Infrastructure planning for availability and GDPS” on page 13, with information on MTMM and GDPS/MTMM.
- ▶ Chapter 9, “GDPS Virtual Appliance” on page 275 is new.
- ▶ Scenarios are added in Chapter 12, “Sample continuous availability and disaster recovery scenarios” on page 357.

- ▶ Section 3.4, “Management of z/OS systems outside of the GDPS sysplex” on page 76 is new.
- ▶ The following sections are updated:
 - 3.13, “Comparison of GDPS/PPRC versus other GDPS offerings” on page 99
 - 4.10, “Comparison of GDPS/PPRC HM to other GDPS offerings” on page 134
 - 5.11, “Comparison of GDPS/XRC versus other GDPS offerings” on page 159
 - 6.11, “Comparison of GDPS/GM versus other GDPS offerings” on page 186
- ▶ Two subsections, 3.14, “Summary” on page 100 and 4.11, “Summary” on page 135, are updated with information related to the new GDPS/MTMM and GDPS Virtual Appliance offerings.
- ▶ References to GDPS Sysplex Timer (ETR) related processing are removed because support for the 9037 Sysplex Timer has been discontinued.
- ▶ Several small changes have been made for clarity or to correct minor errors.

August 2014, Update to Tenth Edition

New and changed information

- ▶ New information about the GDPS/MGM 4-site solution is added to several sections:
 - 1.5, “GDPS offerings” on page 8
 - 3.2.5, “GDPS/PPRC in a 3-site or 4-site configuration” on page 74
 - 11.2.3, “Four-copy solutions” on page 337
 - 11.4, “GDPS Metro/Global Mirror 4-site solution” on page 346
- ▶ Where required, instances of GDPS/MGM and MGM are updated to indicate whether a 3-site or 4-site configuration is being referred to.
- ▶ Updated 11.3.8, “Flexible testing in a GDPS/MGM 3-site environment” on page 344 to reference 6.7.2, “Creating a test copy using GM CGPause and testing on isolated disks” on page 184.
- ▶ GDPS/PPRC references to an Active/Active and Active/Standby configuration are changed to single-site and multisite workload respectively and minor changes are made for clarification.
- ▶ Added new GDPS/PPRC section, “GDPS use of DS8000 functions” on page 63. Equivalent section “GDPS use of DS8000 functions” on page 113 is added.
- ▶ Added GDPS use of IBM DS8000® Soft Fence information to “Protecting secondary disks from accidental update” on page 65 and “Protecting secondary disks from accidental update” on page 115.
- ▶ Updated 3.2.5, “GDPS/PPRC in a 3-site or 4-site configuration” on page 74 to include MGM 4-site.
- ▶ New section 3.3.3, “IBM zEnterprise BladeCenter Extension (zBX) hardware management” on page 75 is added.
- ▶ Updated 10.2.1, “Guest Linux under z/VM” on page 299 with information regarding support for alternate subchannel sets to increase scalability.
- ▶ Added information about support for stand-alone servers and high availability options for the SA AppMan server to 10.3, “Distributed Cluster Management” on page 307.
- ▶ New section 10.3.4, “GDPS/PPRC Support for IBM zEnterprise BladeCenter Extension (zBX)” on page 329 is added.

- ▶ New section 6.7.2, “Creating a test copy using GM CGPause and testing on isolated disks” on page 184 is added.
- ▶ Updated 3.13, “Comparison of GDPS/PPRC versus other GDPS offerings” on page 99, 4.10, “Comparison of GDPS/PPRC HM to other GDPS offerings” on page 134, 5.11, “Comparison of GDPS/XRC versus other GDPS offerings” on page 159, and 6.11, “Comparison of GDPS/GM versus other GDPS offerings” on page 186.
- ▶ New section, 5.6.5, “GDPS use of the XRC offline volume support” on page 155 is added.
- ▶ Several other small changes are made for clarity or to correct minor errors.

October 2013, Update to Ninth Edition

New and changed information

- ▶ The following sections were added or updated in support of the GDPS/Active-Active V1R4 enhancements:
 - Section added, “InfoSphere Data Replication for VSAM for z/OS” on page 37
 - Section added, “GDPS/PPRC co-operation with GDPS/Active-Active” on page 98
 - Section updated, “What is a workload” on page 233
 - Section updated, “GDPS/Active-Active solution products” on page 235
 - Section “GDPS/Active-Active environment” on page 240 is updated.
 - Section updated, “Considerations for other non-Active-Active workloads” on page 247
 - Some screen samples refreshed in “GDPS/Active-Active web interface” on page 249 and “GDPS/Active-Active monitoring and alerting” on page 257
 - Section updated, “GDPS/Active-Active scripts” on page 259
 - Section added, “GDPS/Active-Active co-operation with GDPS/PPRC or GDPS/MTMM” on page 264
 - Section added, “GDPS/Active-Active disk replication integration” on page 267
 - Section added “GDPS/Active-Active disk replication integration with GDPS/MGM” on page 346

March 2013, Ninth Edition

New and changed information

- ▶ New section, “Addressing z/OS device limits in a GDPS/GM environment” on page 34
- ▶ New section, “Scalability in a GDPS/MGM 3-site environment” on page 343
- ▶ New section, “Scalability in a GDPS/XRC environment” on page 31
- ▶ “Protecting tape data” on page 65 is updated to include information about the TS7700 “in-doubt” tape support that is added in GDPS 3.10.
- ▶ “Region Switch” on page 148 extended to describe the different possible configurations when performing a GDPS region switch
- ▶ New section, “GDPS/MGM region switch and return home” on page 343
- ▶ New section, “GDPS/MzGM region switch and return home” on page 352
- ▶ New section, “GDPS Query Services in a GDPS/MGM 3-site environment” on page 345

- ▶ New section, “GDPS/XRC integrated XRC performance monitoring” on page 151
- ▶ The following sections are updated with information about the GDPS/PPRC and GDPS/PPRC HM support of consistent IBM FlashCopy® using the IBM FlashCopy Freeze capability of the disk subsystems:
 - “FlashCopy” on page 38
 - “GDPS/PPRC flexible testing and resync protection” on page 96
 - “GDPS/PPRC HM flexible testing and resync protection” on page 131
- ▶ The following sections are updated with the GDPS Health Check management panel information:
 - “GDPS/PPRC health checks” on page 92
 - “GDPS/PPRC HM health checks” on page 127
 - “GDPS/XRC health checks” on page 152
 - “GDPS/GM health checks” on page 180
- ▶ Information about the legacy Freeze option is removed from the following sections:
 - “Protecting data integrity and data availability with GDPS/PPRC” on page 54
 - “Protecting data integrity and data availability with GDPS/PPRC HM” on page 104
- ▶ The following new sections are added to describe add-on GDPS tools available for each of the GDPS products:
 - “GDPS tools for GDPS/PPRC” on page 97
 - “GDPS tools for GDPS/PPRC HM” on page 132
 - “GDPS tools for GDPS/XRC” on page 158
 - “GDPS tools for GDPS/GM” on page 185
- ▶ Some screen captures are refreshed.
- ▶ Several sections are renamed or reorganized for clarity.
- ▶ Several small changes were made for clarity or to correct minor errors.

July 2012, Eighth Edition

New information

- ▶ “Integrated configuration of GDPS/GM and SA AppMan” on page 327 is added.
- ▶ A new section “HyperSwap with less than full channel bandwidth” on page 61 is added to the GDPS/PPRC chapter. An equivalent section is added to the GDPS/HM chapter.
- ▶ 3.6.1, “GDPS/PPRC health checks” on page 92 in Chapter 3, “GDPS/PPRC” on page 53, has been updated to include information that GDPS now ships sample coexistence policy definitions for the GDPS checks that are known to be conflicting with those provided with IBM z/OS.
- ▶ 4.4.1, “GDPS/PPRC HM health checks” on page 127 in Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103, has been updated to include information that GDPS now ships sample coexistence policy definitions for the GDPS checks that are known to be conflicting with those provided with IBM z/OS.
- ▶ A new section “Reserve Storage Pool” has been added in Chapter 3, “GDPS/PPRC” on page 53 and in Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103.
- ▶ A new section “System Management actions” has been added in Chapter 3, “GDPS/PPRC” on page 53 to accommodate customer procedures to perform specific LOAD and RESET actions

- ▶ A new section “SYSRES Management” on page 91 has been added to remove the requirement for clients to manage and maintain their own procedures when IPLing a system on a different alternate SYSRES device.
- ▶ The ability to use commands for taking a FlashCopy is added to 4.3.2, “NetView commands” on page 125.
- ▶ A new section “Query Services” has been added to Chapter 5, “GDPS/XRC” on page 137. It describes new capabilities in GDPS/XRC that improve system management and availability through autonomic capabilities.
- ▶ A new section “Region Switch” has been added in 5.4, “Managing the GDPS environment” on page 142 to reflect a process for performing a planned Site Switch between the two sites that act as the application and recovery sites. New capabilities of the GDPS/XRC product assist with and simplify various procedural aspects of a Site Switch or Return Home operation.
- ▶ A new section 5.6.4, “Vary-After-Clip automation” on page 155 has been added to Chapter 5, “GDPS/XRC” on page 137. With XRC primary devices online in the SDM systems, Vary After Clip processing detects an XRC primary volume has been relabelled and automates varying the subject device online with the new volume label in SDM systems, and schedules a VOLSER REFRESH-type GDPS Configuration update.
- ▶ A new section 6.6.1, “GDPS/GM Copy Once facility” on page 181 has been added in Chapter 6, “GDPS/Global Mirror” on page 163 to provide information about the Copy Once facility to copy volumes that have data sets on them that are required for recovery but the content is not critical, so they do not need to be copied all the time.
- ▶ A new section 10.2, “GDPS/PPRC Multiplatform Resiliency for z Systems” on page 299 has been added in 10.2, “GDPS/PPRC Multiplatform Resiliency for z Systems” on page 299 to reflect the GDPS system management capabilities for IBM z/VM® and Linux systems. GDPS controlled shutdown of z/VM (also referred to as a *graceful shutdown*) has also been included.

Changed information

- ▶ 2.8, “Flexible server capacity” on page 42 in Chapter 2, “Infrastructure planning for availability and GDPS” on page 13 has been updated to reflect the enhanced GDPS CBU and On/Off Capacity on Demand (OOCOD) handling.
- ▶ Section “Addressing z/OS device limits in GDPS/PPRC and GDPS/MTMM environments” in 2.4.1, “PPRC (IBM Metro Mirror)” on page 23 is updated to reflect the new support, the PPRC secondary of IPL, IODF and Stand-Alone Dump devices for z/OS systems in the GDPS sysplex can be defined in the alternate subchannel set (MSS1).
- ▶ Section “PPRC and MTMM-based solutions” in 2.9, “Cross-site connectivity considerations” on page 44 is updated to include considerations about possible use of IBM HyperSwap with less than full bandwidth cross-site connectivity.
- ▶ Section “Sysplex resource management” in 3.5, “Managing the GDPS/PPRC environment” on page 78 has been updated to include GDPS Coupling Facility Management with a single CFRM policy.
- ▶ Section “GDPS HyperSwap function” in Chapter 3, “GDPS/PPRC” on page 53 and in Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103 has been updated to include information about the new, proactive unplanned HyperSwap trigger in conjunction with a new function available in z/OS 1.13 and the new Storage Controller Health Message capability in the IBM DS8000 disk subsystems.
- ▶ 5.6.1, “FlashCopy disk definition in the GDPS systems” on page 154 in Chapter 5, “GDPS/XRC” on page 137 has been updated to include the new FlashCopy protection

support, a new GDPS logic to ensure that when a FlashCopy is taken, it is only taken if the FlashCopy source devices represent a valid recovery point.

- ▶ 10.2.1, “Guest Linux under z/VM” on page 299 in 10.2, “GDPS/PPRC Multiplatform Resiliency for z Systems” on page 299 has been updated to reflect various enhancements in GDPS/PPRC xDR such as definition of two Proxy nodes for each z/VM host, shutdown of an xDR-managed z/VM system in multiple phases, and z/VM system shutdown using the GDPS Stop Standard Action (or equivalent script statement) with all xDR-managed guests stopped in parallel.
- ▶ 10.2.1, “Guest Linux under z/VM” on page 299 in 10.2, “GDPS/PPRC Multiplatform Resiliency for z Systems” on page 299 reflects that GDPS 3.9 xDR for the z/VM guest environment no longer supports REXEC as the communication protocol between the xDR Proxy and the GDPS controlling systems. SOCKET protocol remains the only supported protocol.
- ▶ 10.2, “GDPS/PPRC Multiplatform Resiliency for z Systems” on page 299 has been updated to reflect IBM z/VSE® guests of xDR-managed z/VM systems can be enabled for special GDPS xDR monitoring and management.
- ▶ 10.2, “GDPS/PPRC Multiplatform Resiliency for z Systems” on page 299 has been updated to provide added protection for an xDR-managed z/VM or native Linux system that must be reset during HyperSwap, to ensure that the subject system does not continue to run and update the former primary disks.
- ▶ 2.5, “Tape resident data” on page 38 and 3.1.2, “Protecting tape data” on page 65 has been updated to reflect the GDPS/PPRC support for configuration management of the IBM Virtualization Engine TS7700.
- ▶ 6.3, “GDPS/GM management for distributed systems and data” on page 169 has been updated to reflect the DCM capability in GDPS/GM for managing local cluster sets using IBM Tivoli® System Automation Application Manager.
- ▶ The RCMF offerings are generally replaced by more full featured GDPS offering peers, so the RCMF appendixes have been removed.

June 2011, Seventh Edition

New information

- ▶ This document was updated to reflect the changes and new capabilities of GDPS V3.8 and the GDPS/Active-Active solution
- ▶ A new section “IBM software replication products” has been added to Chapter 2, “Infrastructure planning for availability and GDPS” on page 13 to provide an introduction to the supported IBM software-based replication products within the GDPS/Active-Active solution.
- ▶ A new section “Freeze policy (PPRCFAILURE policy) options” has been updated to include information about the new PPRCFailure and PRIMARYFailure freeze/swap policy options and enhanced Freeze and Stop Conditional processing.
- ▶ A new section Protecting secondary disks from accidental update has been added to Chapter 3, “GDPS/PPRC” on page 53 and to Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103.
- ▶ A new section “Automated response to STP sync WTORs” has been added to Chapter 3, “GDPS/PPRC” on page 53 and to Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103.

- ▶ New sections “STP CTN role reassignments: Planned operations” on page 87, “STP CTN role reassignments: Unplanned failure” on page 90, and “STP WTOR IEA394A response: Unplanned failure” on page 90 have been added to Chapter 3, “GDPS/PPRC” on page 53. They describe the new scripting capability to reassign roles in an STP-only Coordinated Timing Network (CTN), and the capability to respond to WTORs posted by IBM z/OS.
- ▶ A new section “Concurrent Copy cleanup” has been added to Chapter 3, “GDPS/PPRC” on page 53 and to Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103.
- ▶ A new section 4.3.2, “NetView commands” on page 125 has been added to Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103 to describe the HYPERSW command and the new GDPSTIME command to reassign roles in an STP-only CTN.
- ▶ New sections “FlashCopy disk definition in the GDPS systems” and “GDPS/XRC Configuration checking” have been added to Chapter 5, “GDPS/XRC” on page 137. They describe miscellaneous enhancements that improve scalability, management, and data consistency for GDPS/XRC.
- ▶ New sections “GDPS/GM Query Services”, and “Global Mirror Monitor integration” have been added to Chapter 6, “GDPS/Global Mirror” on page 163. They describe new capabilities in GDPS/GM that improve system management and availability through autonomic capabilities.
- ▶ A new chapter, Chapter 8, “GDPS/Active-Active solution” on page 231 is added to describe the GDPS/Active-Active solution.
- ▶ A new section 10.2.4, “Customization Verification Program” on page 306 was added to describe new capabilities to verify that installation and customization activities have been carried out correctly for both xDR native and guest Linux on z Systems environments.
- ▶ A new section 10.2.5, “xDR Extended Monitoring” on page 307 was added to describe the extended monitoring function now available for xDR systems that previously was available only for z/OS systems.
- ▶ A new section “GDPS/MGM Procedure Handler” has been added to Chapter 9, “Combining local and metro continuous availability with out-of-region disaster recovery” on page 331 to describe its function and to mention that this tool is a fully supported tool in GDPS/MGM V3.8.

Changed information

- ▶ The introductory chapters, Chapter 1, “Introduction to business resilience and the role of GDPS” on page 1 and Chapter 2, “Infrastructure planning for availability and GDPS” on page 13 are updated to include new information about the considerations when using software replication for the GDPS/Active-Active solution.
- ▶ Section “Freeze policy (PPRCFAILURE policy) options” in has been updated.
- ▶ Section 3.1.3, “Protecting distributed (FB) data” on page 66 has been updated to add support for SCSI attached FB disk used by native Linux on z Systems systems under GDPS xDR control.
- ▶ Section “Improved controlling system availability: Enhanced STP support” has been updated in Chapter 3, “GDPS/PPRC” on page 53 and Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103 to indicate under what conditions WTORs IEA015A and IEA394A are posted.
- ▶ Section “Flexible testing” has been updated in Chapter 5, “GDPS/XRC” on page 137 to indicate changes in IBM Zero Suspend FlashCopy support.
- ▶ Section 10.1, “Open LUN Management function” on page 296 was updated to include support for SCSI attached FB disk used by native Linux z Systems under GDPS xDR control.

- ▶ Section 10.2, “GDPS/PPRC Multiplatform Resiliency for z Systems” on page 299 was reformatted to include separate sections 10.2.1, “Guest Linux under z/VM” on page 299 and 10.2.2, “Native Linux on z Systems” on page 303 and updated to include support for SCSI-attached FB disk used by native Linux on z Systems under GDPS xDR control.

August 2010, Sixth Edition

New information

- ▶ This document has been updated to reflect changes and new capabilities in GDPS V3.7 including support for PPRC secondary devices defined in an alternate subchannel set and xDR improvements with starting and stopping Linux systems on z Systems.
- ▶ Added references to Microsoft Windows clusters (particularly in “Integrated configuration of GDPS/GM and VCS clusters” on page 313) as part of the GDPS DCM for VCS function.

Changed information

- ▶ Chapter 1, “Introduction to business resilience and the role of GDPS” on page 1 has been rewritten to remove references to SHARE workgroup material and the focus on disaster recovery. The overview of the GDPS family of offerings was also updated to help this chapter act as a stand-alone high level overview.
- ▶ Chapter 2, “Infrastructure planning for availability and GDPS” on page 13 has been modified, moving some of the more technical details to subsequent chapters, but retaining the broad overview of numerous areas of technology infrastructure that are touched or leveraged by a GDPS solution.
- ▶ Section 3.1.1, “Protecting data integrity and data availability with GDPS/PPRC” on page 54 was re-written to remove the discussion about detailed “CRIT” settings not recommended for use with GDPS. References to other documentation with details about these settings are still included. Similar changes were made to Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103.
- ▶ References to *Metro Mirror* were replaced by *PPRC* when discussing the IBM synchronous mirroring architecture. The brand name of *IBM Metro Mirror* continues to be used for the implementation of the PPRC architecture included on the IBM Enterprise Storage Server® and DS8000 family of storage products. A similar change was made for *XRC* and the IBM brand name of *z/OS Global Mirror*.
- ▶ There was a minor reordering of the chapters following the overview of the four primary GDPS offerings.
- ▶ Removed Peer-to-Peer tape from “DR in two data centers, global distance” on page 370 because the configuration of this legacy hardware (with immediate mode) is not appropriate as a recommendation for global distances.

September 2009, Fifth Edition

New information

- ▶ This document has been updated to reflect changes and new capabilities in GDPS V3.6.
- ▶ A new section “Combining disk remote copy technologies for CA and DR” on page 35 has been added to Chapter 2, “Infrastructure planning for availability and GDPS” on page 13.

- ▶ A new section “Improved controlling system availability: Enhanced STP support” on page 69 has been added to Chapter 3, “GDPS/PPRC” on page 53, and to Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103.
- ▶ A new section “GDPS/PPRC in a 3-site or 4-site configuration” has been added to Chapter 3, “GDPS/PPRC” on page 53 and to Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103.
- ▶ A new section “GDPS/PPRC management of distributed systems and data” has been added to Chapter 3, “GDPS/PPRC” on page 53.
- ▶ A new section “GDPS/PPRC monitoring and alerting” has been added to Chapter 3, “GDPS/PPRC” on page 53, to Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103, to Chapter 5, “GDPS/XRC” on page 137, and to Chapter 6, “GDPS/Global Mirror” on page 163.
- ▶ A new section “Other facilities related to GDPS” was created as a repository for miscellaneous topics such as HyperSwap coexistence available in previous versions of GDPS, and new topics available with GDPS V3.6 such as Reduced impact initial copy and resynchronization and Query Services.
- ▶ A new section, “Disk and LSS sharing” on page 302 has been added to Chapter 10, “GDPS extensions for heterogeneous systems and data” on page 295.
- ▶ A new section, “Integrated configuration of GDPS/GM and VCS clusters” on page 313 has been added to Chapter 10, “GDPS extensions for heterogeneous systems and data” on page 295.
- ▶ A new section, “DCM support for SA AppMan” on page 319 has been added to Chapter 10, “GDPS extensions for heterogeneous systems and data” on page 295.

Changed information

- ▶ 2.1.2, “Multisite sysplex considerations” on page 15 was updated to change the maximum fiber distance from 100 km to 200 km (with RPQ).
- ▶ 2.4.3, “Global Mirror” on page 32 was rewritten.
- ▶ 2.8.1, “Capacity Backup upgrade” on page 42 and 2.8.2, “On/Off Capacity on Demand” on page 43 have been updated to indicate the general availability of new functions for GDPS V3.5 and higher.
- ▶ Multiple changes were made in 2.9, “Cross-site connectivity considerations” on page 44 to reflect the recently available Parallel Sysplex InfiniBand technology for coupling and STP and HMC connectivity requirements for STP.
- ▶ 2.9.7, “Connectivity options” on page 47 was updated.
- ▶ 11.3, “GDPS Metro/Global Mirror 3-site solution” on page 338 and 11.5, “GDPS Metro z/OS Global Mirror 3-site solution” on page 349 have been updated to include the Incremental Resynchronization function.
- ▶ Chapter 3, “GDPS/PPRC” on page 53 and Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103 were restructured to introduce concepts in 3.1.1, “Protecting data integrity and data availability with GDPS/PPRC” on page 54 prior to the discussion of configurations in 3.2, “GDPS/PPRC configurations” on page 68 and 4.2, “GDPS/PPRC HM configurations” on page 117.
- ▶ “HyperSwap policy (Primary Failure policy) options” on page 61 was added.
- ▶ Tables provided in Chapter 3, “GDPS/PPRC” on page 53, Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103, Chapter 5, “GDPS/XRC” on page 137, and Chapter 6, “GDPS/Global Mirror” on page 163 that compare functions offered by each GDPS offering have been updated to include a comprehensive list of GDPS functions available to date.

September 2008, Fourth Edition

New information

This document has been updated to reflect changes and new capabilities in GDPS V3.5.

March 2007, Third Edition

New information

- ▶ This document has been updated to reflect changes and new capabilities in GDPS V3.3 and GDPS V3.4.
- ▶ A new section, “Synchronous versus asynchronous data transfer” on page 19, was added to explain the business impact of using Synchronous and Asynchronous remote copy technologies.
- ▶ A new chapter, Chapter 11, “Combining local and metro continuous availability with out-of-region disaster recovery” on page 331, discusses the GDPS/MGM and GDPS/MzGM offerings.
- ▶ Multiplatform for z Systems now supports native zLinux LPARs.
- ▶ GDPS/PPRC and GDPS HyperSwap Manager have been enhanced to provide coexistence support for HyperSwap and IBM TDMF.
- ▶ IBM IMS™ XRF coexistence support added to GDPS/PPRC and GDPS HyperSwap Manager.
- ▶ GDPS/PPRC enables use of the PPRC failover and failback support, if available, in all disk subsystems.
- ▶ GDPS/Global Mirror has been enhanced to provide “No UCB FlashCopy” support.
- ▶ Zero Suspend FlashCopy support added to GDPS/XRC and GDPS/MzGM.
- ▶ Availability of a GDPS Qualification Program for vendor storage subsystems.
- ▶ New WEB GUI interface support added for GDPS/PPRC.
- ▶ GDPS/PPRC and GDPS HyperSwap Manager have been enhanced so that a HyperSwap can now be triggered by a non-responsive primary device, in addition to the existing error conditions that can cause a HyperSwap.
- ▶ GDPS/PPRC has been enhanced to support the new GDPS Enhanced Recovery Services in IBM z/OS 1.8.
- ▶ The ability has been added to GDPS/PPRC to do a planned freeze covering both CKD and FBA devices.
- ▶ FlashCopy support for Open LUN devices has been added to GDPS/PPRC and GDPS HyperSwap Manager.
- ▶ GDPS/XRC has been enhanced to support the new asynchronous write support for system logger staging data sets added in z/OS 1.7.

Changed information

- ▶ The GDPS/PPRC BRS configuration has moved to Chapter 3, “GDPS/PPRC” on page 53.
- ▶ GDPS/XRC scalability enhancements allow up to 20 SDMs in a single LPAR, of which 13 can be coupled together into a cluster. Up to 14 clusters can be coupled together increasing the architectural limit to 182 SDMs.

December 2005, Second Edition

New information

- ▶ Information about the GDPS/GM offering has been added.



Introduction to business resilience and the role of GDPS

In this chapter, we discuss the objective of this book and briefly introduce the contents and layout. We discuss the topic of business IT resilience from a technical perspective (we refer to it as *IT resilience*).

The chapter includes a general description that is not specific to mainframe platforms, although the topics are covered from an enterprise systems and mainframe perspective. Finally, we introduce the members of the IBM Geographically Dispersed Parallel Sysplex (GDPS) family of offerings and provide a brief description of the aspects of an IT resilience solution that each offering addresses.

1.1 Objective

Business IT resilience is a high profile topic across many industries and businesses. Apart from the business drivers requiring near-continuous application availability, government regulations in various industries now take the decision about whether to have an IT resilience capability out of your hands.

This book was developed to provide an introduction to the topic of business resilience from an IT perspective, and to share how GDPS can help you address your IT resilience requirements.

1.2 Layout of this book

This chapter starts by presenting an overview of IT resilience and disaster recovery. These practices have existed for many years. However, recently they have become more complex because of a steady increase in the complexity of applications, the increasingly advanced capabilities of available technology, competitive business environments, and government regulations.

In Chapter 2, “Infrastructure planning for availability and GDPS” on page 13, we briefly describe the available technologies typically used in a GDPS solution to achieve IT resilience goals. To understand the positioning and capabilities of the various offerings (which encompass hardware, software, and services), it is also useful to have at least a basic understanding of the underlying technology.

Following these two introductory chapters and starting with Chapter 3, “GDPS/PPRC” on page 53, we describe the capabilities and prerequisites of each offering in the GDPS family of offerings. Because each offering addresses fundamentally different requirements, each member of the GDPS family of offerings is described in a chapter of its own.

Most enterprises today have a heterogeneous IT environment including various hardware and software platforms. After covering the GDPS family of offerings, Chapter 10, “GDPS extensions for heterogeneous systems and data” on page 295 describes the GDPS facilities that can provide a single point of control to manage data across all the server platforms within an enterprise IT infrastructure.

Finally, we include a section with examples illustrating how the various GDPS offerings can satisfy your requirements for IT resilience and disaster recovery.

1.3 IT resilience

IBM defines *IT resilience* as the ability to rapidly adapt and respond to any internal or external disruption, demand, or threat, and continue business operations without significant impact.

IT resilience is related to, but broader in scope, than *disaster recovery*. Disaster recovery concentrates solely on recovering from an *unplanned* event.

When you investigate IT resilience options, these two terms must be at the forefront of your thinking:

- Recovery time objective (RTO)

This term refers to *how long* your business can afford to wait for IT services to be resumed following a disaster.

If this number is not clearly stated now, think back to the last time that you had a significant service outage. How long was that outage, and how much difficulty did your company suffer as a result? This can help you get a sense of whether to measure your RTO in days, hours, or minutes.

► **Recovery point objective (RPO)**

This term refers to *how much data* your company is willing to re-create following a disaster. In other words, what is the acceptable time difference between the data in your production system and the data at the recovery site?

As an example, if your disaster recovery solution depends on daily full volume tape dumps, your RPO is 24 - 48 hours depending on when the tapes are taken offsite. If your business requires an RPO of less than 24 hours, you will almost certainly be forced to do some form of offsite real-time data replication instead of relying on these tapes alone.

The terms *RTO* and *RPO* are used repeatedly in this book because they are core concepts in the methodology that you can use to meet your IT resilience needs.

1.3.1 Disaster recovery

As mentioned, the practice of preparing for disaster recovery (DR) is something that has been a focus of IT planning for many years. In turn, there is a wide range of offerings and approaches available to accomplish DR. Several options rely on offsite or even outsourced locations that are contracted to provide data protection or even servers if there is a true IT disaster. Other options rely on in-house IT infrastructures and technologies that can be managed by your own teams.

There is no one correct answer for which approach is better for every business. However, the first step in deciding what makes the most sense for *you* is to have a good view of your IT resiliency objectives, specifically your RPO and RTO.

Although Table 1-1 does not cover every possible DR offering and approach, it does provide a view of what RPO and RTO might typically be achieved with some common options.

Table 1-1 Typical achievable RPO and RTO for some common DR options

Description	Typically achievable recovery point objective (RPO)	Typically achievable recovery time objective (RTO)
No disaster recovery plan	Not applicable: all data is lost	Not applicable
Tape vaulting	Measured in days since last stored backup	Days
Electronic vaulting	Hours	Hours (hot remote location) to days
Active replication to remote site (without recovery automation)	Seconds to minutes	Hours to days (dependent on availability of recovery hardware)
Active storage replication to remote "in-house" site	Zero to minutes (dependent on replication technology and automation policy)	One or more hours (dependent on automation)
Active software replication to remote "active" site	Seconds to minutes	Seconds to minutes (dependent on automation)

Generally a form of real-time software or hardware replication is required to achieve an RPO of minutes or less, but the only technologies that can provide an RPO of zero (0) are synchronous replication technologies (see 2.3, “Synchronous versus asynchronous data transfer” on page 19) coupled with automation to ensure that no data is written to one location and not the other.

The recovery time is largely dependent on the availability of hardware to support the recovery and control over that hardware. You might have real-time software or hardware-based replication in place, but without server capacity at the recovery site you will have hours to days before you can recover this previously current data.

Furthermore, even with all the spare capacity and current data, you might find that you are relying on people to perform the recovery actions. In this case, you will undoubtedly find that these same people are not necessarily available in the case of a true disaster or, even more likely, they find that processes and procedures for the recovery are not practiced or accurate. This is where automation comes in to mitigate the risk introduced by the human element and to ensure that you actually meet the RTO required of the business.

Also, you might decide that one DR option is not appropriate for all aspects of the business. Various applications might tolerate a much greater loss of data and might not have an RPO as low as others. At the same time, some applications might not require recovery within hours whereas others most certainly do.

Although there is obvious flexibility in choosing different DR solutions for each application, the added complexity this can bring needs to be balanced carefully against the business benefit. The preferred approach, supported by GDPS, is to provide a single optimized solution for the enterprise. This generally leads to a simpler solution and, because less infrastructure and software might need to be duplicated, often a more cost-effective solution, too. Consider a different DR solution only for your most critical applications, where their requirements cannot be catered for with a single solution.

1.3.2 The next level

In addition to the ability to recover from a disaster, many businesses now look for a greater level of availability covering a wider range of events and scenarios. This larger requirement is called *IT resilience*. In this book, we concentrate on two aspects of IT resilience: *Disaster recovery*, as discussed previously, and *continuous availability (CA)*, which encompasses recovering from disasters and keeping your applications up and running throughout the far more common planned and unplanned outages that do not constitute an actual disaster.

For some organizations, a proven disaster recovery capability that meets their RTO and RPO can be sufficient. Other organizations might need to go a step further and provide near-continuous application availability.

There are several market factors that make IT resilience imperative:

- ▶ High and constantly increasing client and market requirements for continuous availability of IT processes
- ▶ Financial loss because of lost revenue, punitive penalties or fines, or legal actions that are a direct result of disruption to critical business services and functions
- ▶ An increasing number of security-related incidents, causing severe business impact
- ▶ Increasing regulatory requirements
- ▶ Major potential business impact in areas such as market reputation and brand image from security or outage incidents

For a business today, few events affect a company as much as having an IT outage, even for a matter of minutes, and then finding a report of the incident splashed across the newspapers and the evening news. Today, your clients, employees, and suppliers expect to be able to do business with you around the clock and from around the globe.

To help keep business operations running 24x7, you need a comprehensive business continuity plan that goes beyond disaster recovery. Maintaining high availability and continuous operations in normal day-to-day operations are also fundamental for success. Businesses need resiliency to help ensure two essentials:

- ▶ Key business applications and data are protected and available
- ▶ If a disaster occurs, business operations continue with a minimal impact

Regulations

In some countries, government regulations specify how organizations must handle data and business processes. An example is the Health Insurance Portability and Accountability Act (HIPAA) in the United States. This law defines how an entire industry, the US healthcare industry, must handle and account for patient-related data.

Other well-known examples include the US government-released *Interagency Paper on Sound Practices to Strengthen the Resilience of the US Financial System*,¹ which loosely drove changes in the interpretation of IT resilience within the US financial industry, and the Basel II rules for the European banking sector, which stipulate that banks must have a resilient back-office infrastructure.

This is also an area that accelerates as financial systems around the world become more interconnected. Although a set of recommendations published in Singapore (such as the *S 540-2008 Standard on Business Continuity Management*)² might be directly addressing only businesses in a relatively small area, it is common for companies to do business in many countries around the world, where these might be requirements for ongoing business operations of any kind.

Business requirements

An important concept to understand is that the cost and complexity of a solution can increase as you get closer to true continuous availability, and that the value of a potential loss must be borne in mind when deciding which solution you *need*, and which one you can *afford*. You do not want to spend more money on a continuous availability solution than the financial loss you can incur as a result of an outage.

A solution must be identified that balances the costs of the solution with the financial impact of an outage. Several studies have been done to identify the cost of an outage; however, most of them are several years old and do not accurately reflect the degree of dependence most modern businesses have on their IT systems.

Therefore, your company must calculate the impact in your specific case. If you have not already conducted such an exercise, you might be surprised at how difficult it is to arrive at an accurate number. For example, if you are a retailer and you suffer an outage in the middle of the night after all the batch work has completed, the financial impact is far less than if you had an outage of equal duration in the middle of your busiest shopping day. Nevertheless, to understand the value of the solution, you must go through this exercise, using assumptions that are fair and reasonable.

¹ <http://www.sec.gov/news/studies/34-47638.htm>

² <http://www.ss540.org/>

1.3.3 Other considerations

In addition to the increasingly stringent availability requirements for traditional mainframe applications, there are other considerations, including those described in this section.

Increasing application complexity

The mixture of disparate platforms, operating systems, and communication protocols found within most organizations intensifies the already complex task of preserving and recovering business operations. Reliable processes are required for recovering the mainframe data and also, perhaps, data accessed by multiple types of UNIX, Microsoft Windows, or even a proliferation of virtualized distributed servers.

It is becoming increasingly common to have business transactions that span and update data on multiple platforms and operating systems. If a disaster occurs, your processes must be designed to recover this data in a consistent manner.

Just as you would not consider recovering half an application's IBM DB2® data to 8:00 a.m. and the other half to 5:00 p.m., the data touched by these distributed applications must be managed to ensure that *all* of this data is recovered with consistency to a single point in time. The exponential growth in the amount of data generated by today's business processes and IT servers compounds this challenge.

Increasing infrastructure complexity

Have you looked in your computer room recently? If you have, you probably found that your mainframe systems are only a small part of the equipment in that room. How confident are you that all those other platforms can be recovered? And if they can be recovered, will it be to the same point in time as your mainframe systems? And how long will that recovery take?

Figure 1-1 shows a typical IT infrastructure. If you have a disaster and recover the mainframe systems, will you be able to recover your service without all the other components that sit between the user and those systems? It is important to remember why you want your applications to be available, so that users can access them.

Therefore, part of your IT resilience solution must include more than addressing the non-mainframe parts of your infrastructure. It must also ensuring that recovery is integrated with the mainframe plan.

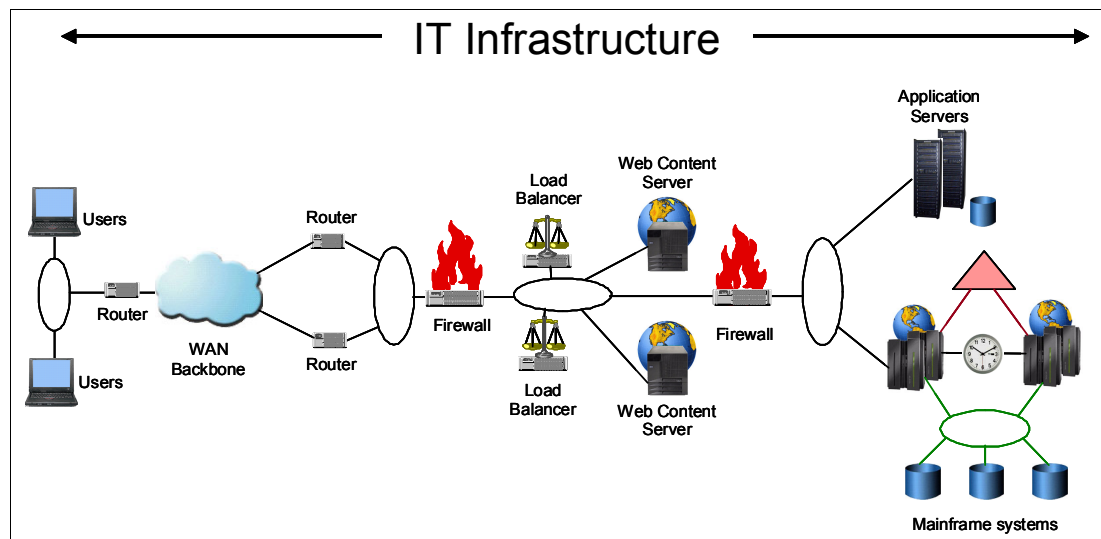


Figure 1-1 Typical IT infrastructure

Outage types

In the early days of computer data processing, planned outages were relatively easy to schedule. Most of the users of your systems were within your company, so the impact to system availability was able to be communicated to all users in advance of the outage. Examples of planned outages are software or hardware upgrades that require the system to be brought down. These outages can take minutes or even hours.

Most outages are planned, and even among unplanned outages, most are not disasters. However, in the current business world of 24x7 Internet presence and web-based services shared across and also between enterprises, even planned outages can be a serious disruption to your business.

Unplanned outages are unexpected events. Examples of unplanned outages are software or hardware failures. Although various of these outages might be quickly recovered from, others might be considered a disaster.

You will undoubtedly have both planned and unplanned outages while running your organization, and your business resiliency processes must cater to both types. You will likely find, however, that coordinated efforts to reduce the numbers of and impacts of unplanned outages often are complementary to doing the same for planned outages.

Later in this book we discuss the technologies available to you to make your organization more resilient to outages, and perhaps avoid them altogether.

1.4 Characteristics of an IT resilience solution

As the previous sections demonstrate, IT resilience encompasses much more than the ability to get your applications up and running after a disaster with “some” amount of data loss, and after “some” amount of time.

When investigating an IT resilience solution, keep in mind the following points:

- ▶ Support for planned system outages
 - Does the proposed solution provide the ability to stop a system in an orderly manner?
 - Does it provide the ability to move a system from the production site to the backup site in a planned manner? Does it support server clustering, data sharing, and workload balancing, so the planned outage can be masked from users?
- ▶ Support for planned site outages
 - Does the proposed solution provide the ability to move the entire production environment (systems, software subsystems, applications, and data) from the production site to the recovery site? Does it provide the ability to move production systems back and forth between production and recovery sites with minimal or no manual intervention?
- ▶ Support for data that spans more than one platform
 - Does the solution support data from more systems than just z/OS? Does it provide data consistency across all supported platforms, or only within the data from each platform?
- ▶ Support for managing the data replication environment
 - Does the solution provide an easy-to-use interface for monitoring and managing the data replication environment? Will it automatically react to connectivity or other failures in the overall configuration?

- ▶ Support for data consistency

Does the solution provide consistency across all replicated data? Does it provide support for protecting the consistency of the second copy if it is necessary to resynchronize the primary and secondary copy?
- ▶ Support for continuous application availability

Does the solution support continuous application availability? From the failure of any component? From the failure of a complete site?
- ▶ Support for hardware failures

Does the solution support recovery from a hardware failure? Is the recovery disruptive (reboot or IPL again) or transparent (HyperSwap, for example)?
- ▶ Support for monitoring the production environment

Does the solution provide monitoring of the production environment? Is the operator notified in a failure? Can recovery be automated?
- ▶ Dynamic provisioning of resources

Does the solution have the ability to dynamically allocate resources and manage workloads? Will critical workloads continue to meet their service objectives, based on business priorities, if there is a failure?
- ▶ Support for recovery across database managers

Does the solution provide recovery with consistency independent of the database manager? Does it provide data consistency across multiple database managers?
- ▶ End-to-end recovery support

Does the solution cover all aspects of recovery, from protecting the data through backups or remote copy, through to automatically bringing up the systems following a disaster?
- ▶ Cloned applications

Do your critical applications support data sharing and workload balancing, enabling them to run concurrently in more than one site? If so, does the solution support and use this capability?
- ▶ Support for recovery from regional disasters

What distances are supported by the solution? What is the impact on response times? Does the distance required for protection from regional disasters permit a continuous application availability capability?

You then need to compare your company's requirements in each of these categories against your existing or proposed solution for providing IT resilience.

1.5 GDPS offerings

GDPS is a collection of several offerings, each addressing a different set of IT resiliency goals that can be tailored to meet the RPO and RTO for your business. Each offering uses a combination of server and storage hardware or software-based replication and automation and clustering software technologies, many of which are described in more detail in Chapter 2, "Infrastructure planning for availability and GDPS" on page 13.

In addition to the infrastructure that makes up a given GDPS solution, IBM also includes services, particularly for the first installation of GDPS and optionally for subsequent installations to ensure that the solution meets and fulfills your business objectives.

The following list briefly describes each offering, with a view of which IT resiliency objectives it is intended to address. Extra details are included in separate chapters of this book:

► GDPS/PPRC

A near-CA or DR solution across two sites separated by metropolitan distances. The solution is based on the IBM PPRC synchronous disk mirroring technology.

► GDPS/PPRC HyperSwap Manager

A near-CA solution for a single site or entry-level DR solution across two sites separated by metropolitan distances. The solution is based on the same technology as GDPS/PPRC, but does not include much of the system automation capability that makes GDPS/PPRC a more complete DR solution.

► GDPS/MTMM

A near-CA *and* DR solution across two sites separated by metropolitan distances. The solution is based on the IBM Multi-Target Metro Mirror (MTMM) synchronous disk mirroring technology. It is very similar to the GDPS/PPRC offering except that, rather than a single synchronous copy, two synchronous copies are managed. This provides additional protection.

► IBM GDPS Virtual Appliance

A near-CA or DR solution across two sites separated by metropolitan distances. The solution is based on the IBM PPRC synchronous disk mirroring technology. The solution provides Near-CA or DR protection for IBM z/VM and Linux on z Systems in environments that do not have IBM z/OS operating systems.

► GDPS/XRC

A DR solution across two sites separated by virtually unlimited distance between sites. The solution is based on the IBM Extended Remote Copy (XRC) asynchronous disk mirroring technology (also called IBM z/OS Global Mirror).

► GDPS/Global Mirror

A DR solution across two sites separated by virtually unlimited distance between sites. The solution is based on the IBM System Storage® Global Mirror technology, which is a disk subsystem-based asynchronous form of remote copy.

► GDPS/Metro-Global Mirror (GDPS/MGM)

Either a 3-site or a symmetrical 4-site configuration is supported:

– GDPS/MGM 3-site

A 3-site solution that provides CA across two sites within metropolitan distances and DR to a third site, in a different region, at virtually unlimited distances. It is based on either a cascading mirroring technology that combines PPRC and Global Mirror or a multi-target mirroring technology that combines MTMM and Global Mirror.

– GDPS/MGM 4-site

A symmetrical 4-site solution that is similar to the 3-site solution in that it provides CA within region and DR cross region. In addition, in the 4-site solution, the two regions are configured symmetrical so that the same levels of CA and DR protection is provided, no matter which region production runs in.

- ▶ GDPS Metro-z/OS Global Mirror (GDPS/MzGM)
 - GDPS/MzGM 3-site

A 3-site solution that provides CA across two sites within metropolitan distances and DR to a third site at virtually unlimited distances. It is based on a multi-target mirroring technology that combines PPRC and XRC (also known as z/OS Global Mirror on IBM storage subsystems).
 - GDPS/MzGM 4-site

A symmetrical 4-site solution that is similar to the 3-site solution in that it provides CA within region and DR cross region. In addition, in the 4-site solution, the two regions are configured symmetrically so that the same levels of CA and DR protection is provided, no matter which region that production runs in.
- ▶ GDPS/Active-Active

A multisite CA/DR solution at virtually unlimited distances. This solution is based on software-based asynchronous mirroring between two active production sysplexes running the same applications with the ability to process workloads in either site.

As mentioned briefly at the beginning of this section, each of these offerings provides the following benefits:

- ▶ GDPS automation code

This code has been developed and enhanced over several years to use new hardware and software capabilities to reflect preferred practices, based on IBM experience with GDPS clients since the inception of GDPS, in 1998, and to address the constantly changing requirements of our clients.
- ▶ Can use underlying hardware and software capabilities

IBM software and hardware products have support to surface problems that can affect the availability of those components, and to facilitate repair actions.
- ▶ Services

There is perhaps only one factor in common across all GDPS implementations, namely that each has a unique requirement or attribute that makes it different from every other implementation. The services aspect of each offering provides you with invaluable access to experienced GDPS practitioners.

The amount of service included depends on the scope of the offering. For example, more function-rich offerings such as GDPS/PPRC include a larger services component than GDPS/PPRC HyperSwap Manager.

Note: Detailed information about each of the offerings is provided in the following chapters. It is not necessary to read all chapters if you are interested only in a specific offering. If you do read all of the chapters, you might notice that some information is repeated in multiple chapters.

1.6 Automation and disk replication compatibility

The GDPS automation code relies on the runtime capabilities of IBM Tivoli NetView® and IBM Tivoli System Automation (SA). Although these products provide tremendous first-level automation capabilities in and of themselves, there are alternative solutions you might already have from other vendors.

GDPS continues to deliver features and functions that take advantage of properties unique to the IBM Tivoli products (such as support for alert management through Tivoli System Automation for Integrated Operations Management), but Tivoli NetView and Tivoli SA also work well alongside other first-level automation solutions. Therefore, although there are definitely benefits to having a comprehensive solution from IBM, you do not have to replace your current automation investments before moving forward with a GDPS solution.

Most of the GDPS solutions rely on the IBM developed disk replication technologies³ of PPRC for GDPS/PPRC, Multi-Target Metro Mirror for GDPS/MTMM, XRC for GDPS/XRC, and Global Mirror for GDPS/GM. These architectures are implemented on several IBM enterprise storage products. Specifically, PPRC has been implemented and branded as IBM System Storage Metro Mirror for the IBM Enterprise Storage Server and the IBM DS8000 family of products. Similarly, the XRC technology has been implemented on the same storage servers under the brand name of IBM System Storage z/OS Global Mirror.

The external interfaces for all of these disk replication technologies (PPRC, MTMM, XRC, GM, and FlashCopy) have also been licensed by many major enterprise storage vendors. This gives clients the flexibility to select the disk subsystems that best match their requirements and to mix and match disk subsystems from different storage vendors within the context of a single GDPS solution. Although most GDPS installations do rely on IBM storage products, there are several production installations of GDPS around the world that rely on storage products from other vendors.

IBM has a GDPS Qualification Program⁴ for other enterprise storage vendors to validate that their implementation of the advanced copy services architecture meets the GDPS requirements.

The GDPS Qualification Program offers the following arrangement to vendors:

- ▶ IBM provides the system environment.
- ▶ Vendors install their disk in this environment.
- ▶ Testing is conducted jointly.
- ▶ A qualification report is produced jointly, describing details of what was tested and results.

Recognize that this qualification program does not imply that IBM provides defect or troubleshooting support for a qualified vendor's products. It does, however, indicate at least a point-in-time validation that the products are functionally compatible and demonstrates that they work in a GDPS solution.

Check directly with non-IBM storage vendors if you are considering using their products with a GDPS solution because they can share their own approaches and capability to support the specific GDPS offering you are interested in.

³ Disk replication technology is independent of the GDPS/Active-Active solution, which uses software replication.

⁴ <http://www.ibm.com/systems/z/gdps/qualification.html>

1.7 Summary

At this point we have discussed why it is important to have an IT resilience solution, and have provided information about key objectives to consider when developing your own solution. We have also introduced the GDPS family of offerings with a brief description of which objectives of IT resiliency each offering is intended to address.

In Chapter 2, “Infrastructure planning for availability and GDPS” on page 13 we introduce key infrastructure technologies related to IT resilience focused on the mainframe platform. After that, we describe how the various GDPS offerings use those technologies. And finally, we position the various GDPS offerings against typical business scenarios and requirements.

We intend to update this book as new GDPS capabilities are delivered.



Infrastructure planning for availability and GDPS

In this chapter, we discuss several technologies that are available to help you achieve your goals related to IT resilience, recovery time, and recovery point objectives. To understand how the IBM GDPS offerings described in this book can help you, it is important to have at least conceptual knowledge of the functions, capabilities, and limitations of these underlying technologies.

2.1 Parallel Sysplex overview

As discussed in Chapter 1, “Introduction to business resilience and the role of GDPS” on page 1, *IT resilience* covers more than just recovery from a disaster. It also encompasses ensuring high availability on a day-to-day basis, protecting your applications from normal planned, and unplanned outages. You cannot expect to be able to provide continuous or near-continuous application availability across a disaster if you are unable to provide that in normal operations.

Parallel Sysplex is the primary mechanism used by IBM to provide the highest levels of application availability on the z Systems¹ platform. The logical first step in a business resiliency project is to do all you can to deliver the highest levels of service from your existing configuration. Implementing Parallel Sysplex with data sharing and dynamic workload routing provides higher levels of availability now. It also provides a foundation to achieve greater resiliency if you implement GDPS.

In the following sections we briefly discuss Parallel Sysplex, the benefits you can derive by using the technology, and the points to consider if you decide to implement GDPS/PPRC, GDPS/MTMM, or GDPS/Active-Active. Because GDPS/XRC and GDPS/Global Mirror do not have a continuous availability (CA) aspect, there are no Parallel Sysplex considerations specifically relating to GDPS/XRC and GDPS/Global Mirror. There are also no Parallel Sysplex considerations for the IBM GDPS Virtual Appliance because the GDPS Virtual Appliance protects only IBM z/VM and Linux on z Systems platforms.

2.1.1 Maximizing application availability

There is only one way to protect applications from the loss of a single component (such as an IBM CICS® region or a z/OS system), and that is to run multiple, failure-isolated copies. This infers an ability to share data at the record level, with integrity, and to *dynamically* route incoming work requests across the available servers. Parallel Sysplex uses hardware and software components to link individual systems together in a cluster. Because all systems in the sysplex are able to share the same resources and data, they appear as a single image to applications and users, while providing the ability to eliminate single points of failure.

Having more than one instance of an application within the sysplex can shield your users from both planned and unplanned outages. With Parallel Sysplex, parts of the cluster can be brought down for maintenance, upgrades, or any other type of outage, while the applications continue to be available on other members of the sysplex.

GDPS/Active-Active further extends this concept with the ability to switch the workload between two sysplexes separated by virtually unlimited distance for both planned and unplanned outage situations.

Although it is not necessary to have a Parallel Sysplex before implementing most GDPS solutions, it is important to understand the role that Parallel Sysplex plays in supporting the continuous availability aspect of IT resilience. Technical information about implementing and using Parallel Sysplex is available in other IBM documentation, so it is not covered in this book.

¹ In this book, we use the term *z Systems* to refer to the IBM z Systems, System z, and zSeries ranges of processors. If something applies only to System z or zSeries processors, we point that out at the time.

2.1.2 Multisite sysplex considerations

The considerations for a multisite sysplex depend on whether you plan to run production systems in both sites at the same time or if all the production systems will be in a single site at any one time. Configurations where production systems can run in both sites at the same time are referred to as multisite workload configurations. Configurations where the production systems run together in one site or the other (but not split across multiple sites) are referred to as single-site workload configurations or sometimes as Active/Standby configurations. Other variations on this, where production systems are predominantly running at one site but where partially active systems or systems enabled only for queries are running at the other site, are still considered multisite workloads.

Terminology: This section is focused on a *multisite sysplex*, which is a single sysplex spread across multiple (typically two) sites, and how the workload is configured to run in those sites to provide near-continuous availability and metro distance DR.

Do not confuse it with the GDPS/Active-Active solution that uses some of the same terminology, but is related to multiple sysplexes (limited to two, currently) and how the workload is configured between the two sysplexes, not within any single sysplex.

In a GDPS/Active-Active environment, it is anticipated that each of the participating sysplexes will themselves be in an Active/Active configuration, providing local and continuous availability with GDPS/PPRC and GDPS/Active-Active providing a solution for unlimited distance CA/DR. For more details about the GDPS/Active-Active solution, see Chapter 8, “GDPS/Active-Active solution” on page 231.

Several phrases are often used to describe variations of multisite workload. Brief definitions are included here for the more commonly implemented variations.

Active/Active	This refers to a multisite workload configuration where z/OS systems are actively running in the same sysplex with active subsystems in more than one site at the same time. Typically this term also implies that applications take advantage of data sharing and dynamic workload routing in such a way that applications can freely move from one site to another. Finally, critical Parallel Sysplex resources are duplexed or replicated in such a way that if one site fails, the remaining site can recover workload within minutes after contending locks and communications timeouts clear. When combined with HyperSwap, an Active/Active configuration has the potential to provide near-continuous availability for applications even in the case of a site outage.
Active/Warm	This refers to a multisite workload configuration that is similar to the Active/Active configuration, with production systems running at more than one site. The difference is that workload generally runs in one site at a time, with the systems in the other site simply IPLed without subsystems or other resources active. This configuration is intended to save IPL time when moving workload between sites. It can be most effective for supporting the planned movement of workload because in many unplanned scenarios, the “warm” systems might also not survive.

Active/Query	This refers to a multisite workload configuration that is quite close to the Active/Active configuration, but where workload at the second site is partitioned or restricted (possibly to queries only) in such a way as to limit impacts because of serialization, thereby protecting shared resources when delays because of distance between the sites is a concern. Again, depending on the configuration of the coupling facility structures (that is, whether they are duplexed across sites or basically in one site at a time), this configuration might provide value only for planned scenarios because in many unplanned scenarios the “query” or “hot standby” subsystems might not survive.

You can devise potentially many more configuration variations, but from a Parallel Sysplex and GDPS² perspective, all of these fall into either the single-site or the multisite workload category.

Single-site or multisite workload configuration

When first introduced, Parallel Sysplexes were typically contained within a single site. Extending the distance between the operating system images and the coupling facility has an impact on the response time of requests using that coupling facility (CF). Also, even if the systems sharing the data are spread across more than one site, all of the primary disk subsystems are normally contained in the same site, so a failure affecting the primary disks affects the systems in both sites. As a result, a multisite workload configuration does not, in itself, provide significantly greater availability than a single-site workload configuration during unplanned outages. To achieve the optimal benefit from a multisite workload configuration for planned outages, HyperSwap should be used; this enables you to move applications and their data from one site to the other nondisruptively.

More specifically, be careful when planning a multisite workload configuration if the underlying Parallel Sysplex cannot be configured to spread the important coupling facility structures across the sites and still achieve the required performance. As discussed later in this chapter and illustrated in Table 2-1 on page 46, the Coupling Link technology can support links upwards of 100 km with qualified dense wavelength division multiplexing (DWDM). However, this does not mean that your workload will tolerate even 1 km of distance between the z/OS images and the CF. Individual coupling operations will be delayed by 10 microseconds per kilometer. Although this time can be calculated, there is no safe way to predict the increased queuing effects caused by the increased response times and the degree of sharing that is unique to each environment. In other words, you will need to run your workload with connections at distance to evaluate the tolerance and impacts of distance.

The benefits of a multisite workload come with more complexity. This must be taken into account when weighing the benefits of such configurations.

CF structure duplexing

Two mechanisms exist for duplexing CF structures.

- User-Managed Structure Duplexing is supported for use only with DB2 group buffer pool (GBP) structures. Duplexing the GBP structures can significantly reduce the time to recover the structures following a CF or CF connectivity failure. The performance impact of duplexing the GBP structures is small. Therefore, it is best to duplex the GBP structures used by a production DB2 data sharing group.

² Not including the GDPS/Active-Active solution which relates to a multiple sysplex configuration that can be either single-site or multisite workloads.

- **System-Managed Coupling Facility Structure Duplexing** (referred to as *SM duplexing*) provides a general purpose, hardware-assisted and easy-to-use mechanism for duplexing CF structures. This feature is primarily intended to allow installations to do data sharing without having to have a failure-isolated CF. However, the design of SM duplexing means that having the CFs a significant distance (kilometers) apart can have a dramatic impact on CF response times for the duplexed structures, and thus your applications, and needs careful planning and testing.

In addition to the response time question, there is another consideration relating to the use of cross-site SM Duplexing. Because communication between the CFs is independent of the communication between mirrored disk subsystems, a failure that results in remote copy being suspended would not necessarily result in duplexing being suspended at the same instant. In case of a potential disaster, you want the data in the “remote” CF to be frozen in time at the same instant the “remote” disks are frozen, so you can restart your applications from the moment of failure.

If you are using duplexed structures, it might seem that you are guaranteed to be able to use the duplexed instance of your structures if you must recover and restart your workload with the frozen secondary copy of your disks. However, this is not always the case. There can be rolling disaster scenarios where before, after, or during the freeze event, an interruption occurs (perhaps failure of CF duplexing links) that forces CFRM to drop out of duplexing. There is no guarantee that the structure instance in the surviving site is the one that will be kept. It is possible that CFRM keeps the instance in the site that is about to totally fail. In this case, there will not be an instance of the structure in the site that survives the failure.

Furthermore, during a rolling disaster event, if you freeze secondary disks at a certain point but continue to update the primary disks and the CF structures, then the CF structures, whether duplexed or not, will not be usable if it is necessary to recover on the frozen secondary disks. This depends on some of your installation’s policies.

To summarize, if there is a surviving, accessible instance of application-related structures, this might or might be consistent with the frozen secondary disks and therefore might or might not be usable. Furthermore, depending on the circumstances of the failure, even with structures duplexed across two sites, you are not 100% guaranteed to have a surviving, accessible instance of the application structures. Therefore, you must have procedures in place to restart your workloads without the structure contents.

For more information, see the white paper titled *System-Managed CF Structure Duplexing*, GM13-0103:

<http://ibm.co/1I6k60k>

2.2 Data consistency

In an unplanned outage or disaster situation the ability to perform a database restart, rather than a database recovery, is essential to meet the recovery time objective (RTO) of many businesses, which typically are less than an hour. Database restart allows starting a database application (as you would follow a database manager abend or system abend) without having to restore it from backups. Database recovery is normally a process measured in many hours (especially if you have hundreds or thousands of databases to recover), and it involves restoring the last set of image copies and applying log changes to bring the databases up to the point of failure.

But, there is more to consider than simply the data for one data manager. What if you have an application that updates data in IMS, DB2, and VSAM? If you need to do a recover for these, will your recovery tools allow you to recover them to the same point in time and to the level of granularity that ensures that either all or none of the updates made by one transaction are recovered? Being able to do a restart rather than a recover avoids these issues.

Data consistency across all copies of replicated data, spread across any number of storage subsystems, and in some cases across multiple sites, is essential to providing data integrity and the ability to perform a normal database restart if there is a disaster.

2.2.1 Dependent write logic

Database applications commonly ensure the consistency of their data by using *dependent write logic* regardless of whether data replication techniques are being used. Dependent write logic states that if I/O B must logically follow I/O A, so B will not be started until A completes successfully. This logic would normally be included in all software to manage data consistency. There are numerous instances within the software subsystem, such as databases, catalog/VTOC, and VSAM file updates, where dependent writes are issued.

As an example, in Figure 2-1, LOG-P is the disk subsystem containing the database management system (DBMS) logs, and DB-P is the disk subsystem containing the DBMS data segments. When the DBMS updates a database, it will do these tasks:

1. Write an entry to the log about the intent of the update.
2. Update the database.
3. Write another entry to the log indicating that the database was updated.

If you will be doing a remote copy of these volumes, be sure that *all* of the updates are mirrored to the secondary disks.

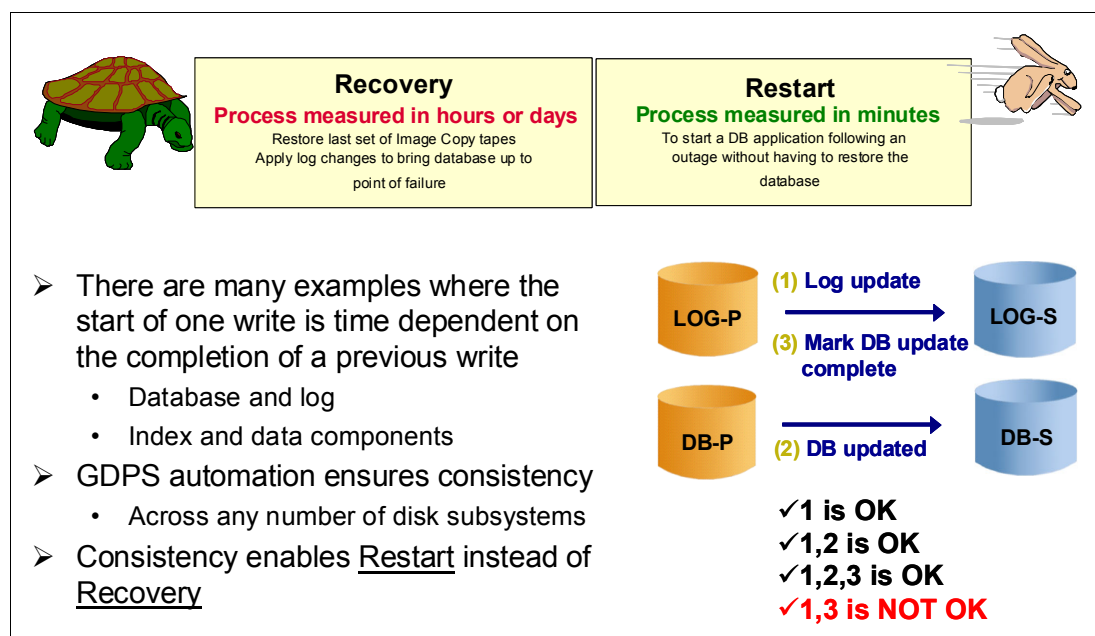


Figure 2-1 The need for data consistency

It is unlikely that all the components in a data center will fail at the same instant, even in the rare case of a full data center outage. The networks might fail first, or possibly one disk subsystem, or any other component in unpredictable combinations. No matter what happens, the remote image of the data must be managed so that cross-volume and subsystem data consistency is preserved during intermittent and staged failures that might occur over many seconds, even minutes. Such a staged failure is generally referred to as a *rolling disaster*.

Data consistency during a rolling disaster is difficult to achieve for synchronous forms of remote copy because synchronous remote copy is entirely implemented within disk subsystem pairs.

For example, in Figure 2-1 on page 18 the synchronously mirrored data sets are spread across multiple disk subsystems for optimal performance. The volume containing the DBMS log on the LOG-P disk subsystem in Site1 is mirrored to the secondary volume in the LOG-S disk subsystem in Site2, and the volume containing the data segments in the DB-P disk subsystem in Site1 is mirrored to the secondary volume in the DB-S disk subsystem in Site2.

Assume that a disaster is in progress in Site1, causing the link between DB-P and DB-S to be lost before the link between LOG-P and LOG-S is lost. With the link between DB-P and DB-S lost, a write sequence of (1), (2), and (3) might be completed on the primary devices (depending on how the remote copy pair was defined) and the LOG writes (1) and (3) would be mirrored to the LOG-S device, but the DB write (2) would not have been mirrored to DB-S. A subsequent DBMS restart using the secondary copy of data in Site2 would clean up in-flight transactions and resolve in-doubt transactions, but the missing DB write (2) would not be detected. In this example of the missing DB write the DBMS integrity was compromised.³

We discuss data consistency for synchronous remote copy in more detail in “PPRC data consistency” on page 24.

For the two IBM asynchronous remote copy offerings, the consistency of the volumes in the recovery site is ensured because of the way these offerings work. This is described further in 2.4.3, “Global Mirror” on page 32 and “XRC data consistency” on page 28.

For GDPS/Active-Active, which relies on asynchronous software replication as opposed to the use of PPRC, XRC, or Global Mirror, consistency is managed within the replication software products discussed further in 2.4.5, “IBM software replication products” on page 36.

2.3 Synchronous versus asynchronous data transfer

Synchronous data transfer and *asynchronous* data transfer are two methods used to replicate data. Before selecting a data replication technology, you must understand the differences between the methods used and the business impact.

³ The way the disk subsystem reacts to a synchronous PPRC remote copy failure depends on the options you specify when setting up the remote copy session. The behavior described here is the default if no overrides are specified.

Terminology: In this book, we continue to use the term Peer-to-Peer Remote Copy (PPRC) when referring to the synchronous disk replication architecture. The rebranded name of the IBM implementation of this architecture is IBM Metro Mirror, which is used when specifically referring to the IBM implementation on the IBM Enterprise Storage Server and the IBM DS8000 family of products.

Similarly, we continue to use the term Extended Remote Copy (XRC) when referring to the asynchronous disk copy technology that uses the z/OS System Data Mover (SDM). The rebranded name of the IBM disk storage implementation is z/OS Global Mirror, which is used specifically when referring to the IBM implementation on the IBM Enterprise Storage Server and the IBM DS8000 family of products.

When using synchronous data transfer, illustrated in Figure 2-2 using PPRC, the application writes are first written to the primary disk subsystem (1) and then forwarded on to the secondary disk subsystem (2). When the data has been committed to *both* the primary and secondary disks (3), an acknowledgment that the write is complete (4) is sent to the application. Because the application must wait until it receives the acknowledgment before executing its next task, there will be a slight performance impact. Furthermore, as the distance between the primary and secondary disk subsystems increases, the write I/O response time increases because of signal latency⁴.

The goals of synchronous replication are zero or near-zero loss of data, and quick recovery times from failures that occur at the primary site. Synchronous replication can be costly because it requires high-bandwidth connectivity.

One other characteristic of synchronous replication is that it is an enabler for nondisruptive switching between the two copies of the data that is known to be identical.

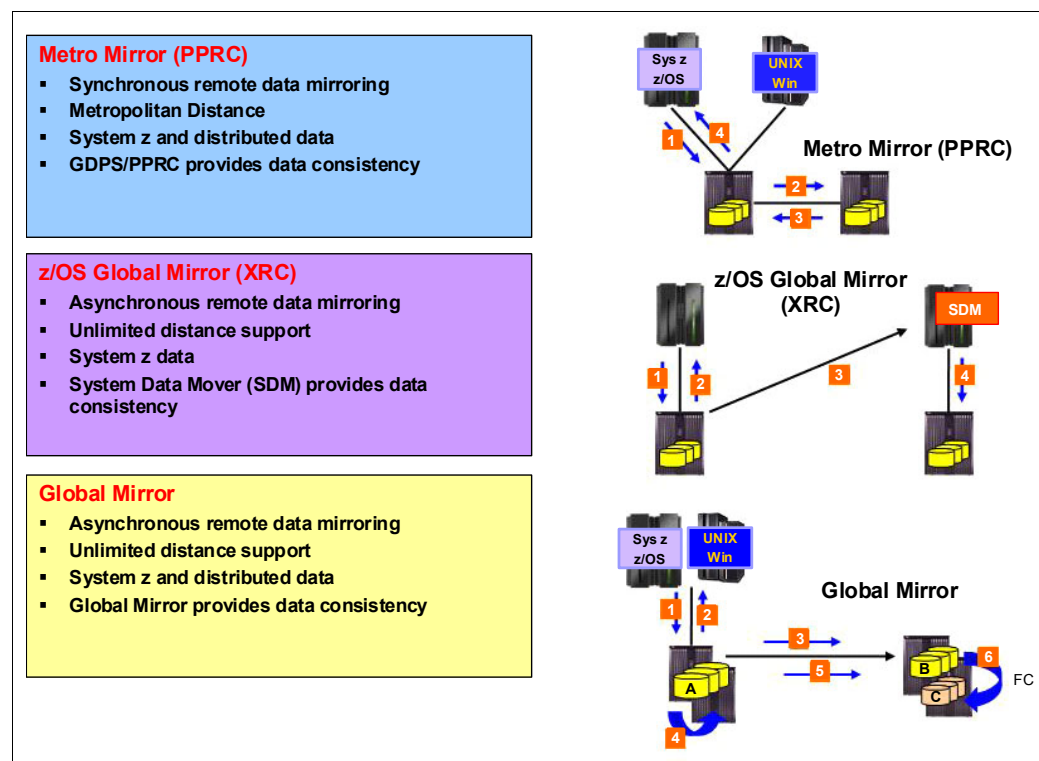


Figure 2-2 Synchronous versus asynchronous storage replication

⁴ Signal latency is related to the speed of light over fiber and is 10 microseconds per km, round trip.

With asynchronous replication, illustrated in Figure 2-2 on page 20, with either z/OS Global Mirror (XRC) or with Global Mirror, the application writes to the primary disk subsystem (1) and receives an acknowledgment that the I/O is complete as soon as the write is committed on the primary disk (2). The write to the secondary disk subsystem is completed in the background. Because applications do not have to wait for the completion of the I/O to the secondary device, asynchronous solutions can be used at virtually unlimited distances with negligible impact to application performance. In addition, asynchronous solutions do not require as much bandwidth as the synchronous solutions.

With software-based asynchronous replication, as used in a GDPS/Active-Active environment, the process is similar to that described for XRC. Data is captured from the database subsystem logs at the source copy when a transaction commits data to the database. That captured data is then sent asynchronously to a second location where it is applied to the target copy of the database in near real time.

When selecting a data replication solution, perform a business impact analysis to determine which solution meets the businesses requirements while ensuring your service delivery objectives continue to be met; see Figure 2-3. The maximum amount of transaction loss that is acceptable to the business (RPO) is one measurement used to determine which remote copy technology should be deployed. If the business is able to tolerate loss of committed transactions, then an asynchronous solution will likely provide the most cost-effective solution. When no loss of committed transactions is the objective, then synchronous remote copy must be deployed. In this case, the distance between the primary and secondary remote copy disk subsystems, and the application's ability to tolerate the increased response times, must be factored into the decision process.

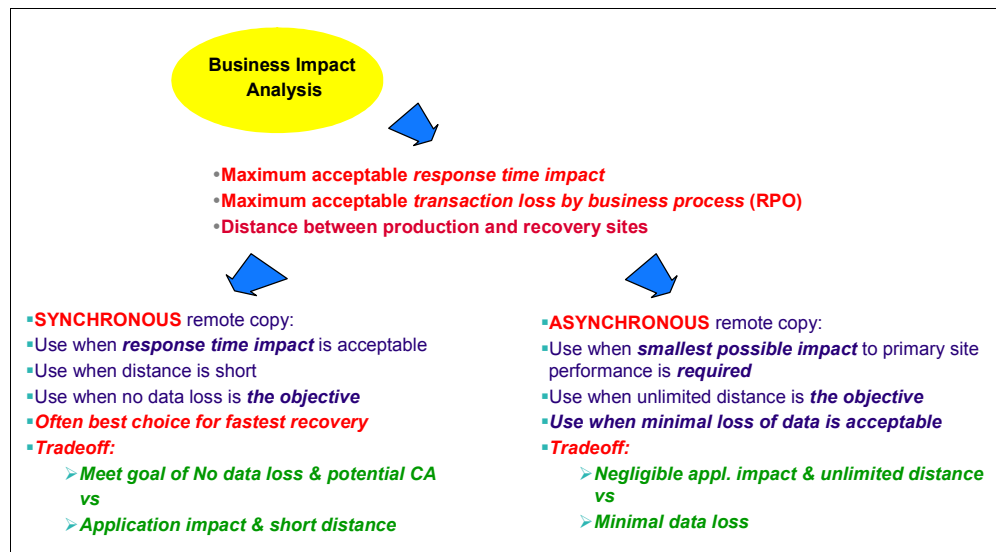


Figure 2-3 Business impact analysis

Many enterprises have both business and regulatory requirements to provide near-continuous data availability, without loss of transactional data, while protecting critical business data if there is a wide-scale disruption. This can be achieved by implementing three-copy (sometimes referred to as *3-site*) mirroring solutions that use both synchronous and asynchronous replication technologies. Synchronous solutions are used to protect against the day-to-day disruptions with no loss of transactional data. Asynchronous replication is used to provide out-of-region data protection, with some loss of committed data, for wide-spread disruptions. The key is to ensure cross-disk subsystem data integrity and data consistency is maintained through any type of disruption.

For more information about three-copy replication solutions see Chapter 11, “Combining local and metro continuous availability with out-of-region disaster recovery” on page 331.

2.4 Data replication technologies

The two primary ways to make your data available following a disaster are as follows:

- ▶ By using a form of tape-based backup
- ▶ By using data replication to a recovery site (also known as remote copy)

This can be hardware-based or software-based replication.

For companies with an RTO of a small number of hours or less, a tape-based solution is unlikely to be acceptable, because it is simply not possible to restore all your volumes and apply all database logs in the time available. Therefore, we are assuming that if you are reading this book you already have, or are planning to implement, some form of data replication technology.

Remotely copying your data eliminates the time that would be required to restore the data from tape and addresses the problem of having to recover data that is generated between the last backup of an application system and the time when the application system fails. Depending on the technology used, remote copy implementations provide a real-time (or near real-time) continuing copy of data between a source and a target.

IBM offers three basic technologies to provide this type of mirroring for disk storage:

- ▶ PPRC: Updates to the primary volumes are synchronously mirrored to the remote volumes and all interactions related to this activity are done between the disk subsystems. MTMM is based on PPRC and allows multiple secondary copies from the same primary.
- ▶ XRC: The task of retrieving the updates from the primary disk subsystem and applying those changes to the secondary volumes is done by a z/OS component named the System Data Mover (SDM).
- ▶ Global Mirror: This offering mirrors the data asynchronously. However, unlike XRC, all interactions are done between the disk subsystems rather than by an SDM.

These technologies are described more fully in the following sections. For an even more detailed explanation of the remote copy technologies described in the following sections, see *IBM System Storage DS8000: Copy Services for IBM System z*, SG24-6787.

IBM also offers several software-based replication products. Unlike the technologies listed for mirroring disk storage (which are application independent), most software replication products are specific to the database source and target in use. The following products are currently supported in a GDPS/Active-Active environment:

- ▶ IBM InfoSphere® Data Replication for IMS for z/OS
- ▶ IBM InfoSphere Data Replication for VSAM for z/OS
- ▶ IBM InfoSphere Data Replication for DB2 for z/OS

These products are introduced in the following sections; more information can be found at the following location:

<http://publib.boulder.ibm.com/infocenter/dzichelp/v2r2/index.jsp>

2.4.1 PPRC (IBM Metro Mirror)

PPRC ensures that after the volume pair has been established and remains synchronized that the secondary volume will always contain exactly the same data as the primary. The IBM implementation of PPRC, known as IBM Metro Mirror, provides synchronous data mirroring at distances up to 300 km (and potentially even greater distances, after technical review and approval).

Important: Always use caution when considering long distances. When we say that something is “supported up to xx km,” it means that the technology will work at that distance if you have qualified cross-site connectivity technology that supports that protocol. See 2.9, “Cross-site connectivity considerations” on page 44 for more details.

You must also consider the impact the increased response time will have on your applications. Some applications can tolerate the response time increase associated with cross-site distances of 100 km, but the same distance in another installation might make it impossible for the applications to deliver acceptable levels of performance.

So, carefully evaluate the projected response time impact, and apply that increase to your environment to see if the result is acceptable. Your vendor storage specialist can help you determine the disk response time impact of the proposed configuration.

Recovery point objective with PPRC

If you have a recovery point objective of zero (0), meaning zero data loss, PPRC is the only IBM remote copy option that can achieve that objective.

That is not to say that you will always have zero data loss if using PPRC. Zero data loss means that there will never be updates made to the primary disks that are not mirrored to the secondaries. The only way to ensure that zero data loss is to immediately stop all update activity to the primary disks if the remote copy relationship ceases to exist (if you lose connectivity between the primary and secondary devices, for example).

Thus, choosing to have zero data loss really means that you must have automation in place that will stop all update activity in the appropriate circumstances. It also means that you accept the possibility that the systems can be stopped for a reason other than a real disaster; for example, if the failure was caused by a broken remote copy link rather than a fire in the computer room. Completely avoiding single points of failure in your remote copy configuration, however, can reduce the likelihood of such events to an acceptably low level.

Supported platforms with PPRC

PPRC replication is supported for any IBM or non-IBM disk subsystem that supports the PPRC architecture, specifically the Freeze/Run capability. PPRC can mirror both fixed-block (FB) devices that are typically used by z Systems and platforms other than z Systems and by CKD devices that are traditionally used by mainframe operating systems, such as IBM z/OS, IBM z/VM, and IBM z/VSE. Not all operating systems necessarily support an interface to control the remote copy function. However, the PPRC function for FB devices can be controlled from a connected z/OS system if there are sufficient CKD formatted volumes defined in the storage subsystem (as described more fully in 3.1.3, “Protecting distributed (FB) data” on page 66 for GDPS/PPRC and 4.1.2, “Protecting distributed (FB) data” on page 116 for GDPS/PPRC HyperSwap Manager).

With current implementations of PPRC, the primary and secondary disk subsystems must be from the same vendor, although vendors (including IBM) often support PPRC mirroring between different disk subsystem models of their own product lines. This can help with migrations and technology upgrades.

Distance with PPRC

The maximum distance supported for IBM Metro Mirror is 300 km (without an RPQ). Typical GDPS/PPRC, GDPS/HyperSwap manager, and GDPS/MTMM configurations are limited to distances less than this because of Coupling Link configurations. See 2.9.3, “Coupling links” on page 45 for more details about the supported distances for these Parallel Sysplex connections. You will also need to contact other storage vendors to understand the maximum distances supported by their PPRC compatible mirroring implementations.

Performance with PPRC

As the distance between your primary and secondary disk subsystems increases, the time it takes for your data to travel between the subsystems also increases. This might have a performance impact on your applications because they cannot proceed until the write to the secondary device completes.

Be aware that as response times increase, link use also increases. Depending on the type and number of PPRC links you configured, additional links and the use of Parallel Access Volumes (PAVs) might help to provide improved response times at longer distances.

Disk Magic, a tool available to your IBM storage specialist, can be used to predict the impact of various distances, link types, and link numbers for IBM disk implementation. We consider access to the information provided by such a tool essential to a GDPS project using PPRC.

PPRC connectivity

Connectivity between the primary and secondary disk subsystems can be provided by direct connections between the primary and secondary disk subsystems, by IBM FICON® switches, by DWDMs, and by channel extenders.

The type of intersite connection (dark fiber or telecommunications link) available determines the type of connectivity you use: telecommunication links can be used by channel extenders, and the other types of connectivity require dark fiber.

You can find information about connectivity options and considerations for z Systems in the most recent version of *IBM System z Connectivity Handbook*, SG24-5444.

PPRC data consistency

When using PPRC, the following sequence of actions occurs when an update I/O is issued to a primary volume:

1. Write to the primary volume (disk subsystem cache and non-volatile store (NVS)).

Your production system writes data to a primary volume and a cache hit occurs.

2. Write to the secondary (disk subsystems cache and NVS).

The primary disk subsystem's microcode then sends the update to the secondary disk subsystem's cache and NVS.

3. Signal write is complete on the secondary.

The secondary disk subsystem signals write complete to the primary disk subsystem when the updated data is in its cache and NVS.

4. Post I/O is complete.

When the primary disk subsystem receives the write complete from the secondary disk subsystem, it returns Device End (DE) status to your application program. Now, the application program can continue its processing and move on to any dependent writes that might have been waiting for this one to complete.

However, PPRC on its own provides this consistency only for a single write. Guaranteeing consistency across multiple logical subsystems and even across multiple disk subsystems requires automation on top of the PPRC function itself. This is where GDPS comes in with freeze automation, described more fully in these sections:

- ▶ 3.1.1, “Protecting data integrity and data availability with GDPS/PPRC” on page 54 for GDPS/PPRC
- ▶ 4.1.1, “Protecting data integrity and data availability with GDPS/PPRC HM” on page 104 for GDPS/PPRC HyperSwap Manager
- ▶ 7.1.1, “Protecting data integrity and data availability with GDPS/MTMM” on page 191 for GDPS/MTMM.

PPRC transparent disk swap

Because under normal conditions, the primary and secondary disks are known to be identical, with PPRC it is possible to swap to using the secondary copy of the disks in a manner that is transparent to applications that are using those disks. This task is not simple, and it requires tight control and coordination across many devices shared by multiple systems in a timely manner. GDPS/PPRC, GDPS/PPRC HyperSwap Manager, and GDPS/MTMM automation, with support provided in z/OS, z/VM, and specific distributions of Linux on z Systems, provide such a transparent swap capability and it is known as HyperSwap.

HyperSwap is a key availability-enabling technology. For more information about GDPS HyperSwap, see the following sections:

- ▶ “GDPS HyperSwap function” on page 58 for GDPS/PPRC
- ▶ “GDPS HyperSwap function” on page 108 for GDPS/PPRC HyperSwap Manager
- ▶ “GDPS HyperSwap function” on page 196 for GDPS/MTMM
- ▶ “GDPS HyperSwap function” on page 279 for the GDPS Virtual Appliance.

Addressing z/OS device limits in GDPS/PPRC and GDPS/MTMM environments

As clients implement IT resiliency solutions that rely on multiple copies of data, more are finding that the z/OS limit of 64K (65,536) devices is limiting their ability to grow or even to take advantage of technologies like HyperSwap. Clients can consolidate data sets to fewer larger volumes, but even with that, there are times when this might not make operational sense for all types of data.

As a result, z/OS introduced the concept of an “alternate subchannel set,” which can include the definition for certain types of disk devices. An alternate subchannel set provides another set of 64K devices for the following device types:

- ▶ Parallel Access Volume (PAV) alias devices
- ▶ PPRC secondary devices (defined as 3390D)
- ▶ FlashCopy target devices

Including PAV alias devices in an alternate subchannel set is transparent to GDPS and is common practice for current GDPS/PPRC, GDPS/PPRC HyperSwap Manager, and GDPS/MTMM environments.

Support is included in GDPS/PPRC, GDPS/PPRC HyperSwap Manager, and GDPS/MTMM to allow definition of PPRC secondary devices in an alternate subchannel set. With this feature, GDPS can support PPRC configurations with nearly 64K device pairs. GDPS/PPRC and GDPS/PPRC HyperSwap Manager allow the secondary devices for z/OS systems in the GDPS sysplex, as well as for managed z/VM systems (and guests) to be defined in an alternate subchannel set. GDPS/MTMM only supports alternate subchannel sets for z/OS systems in the sysplex.

There are limitations to keep in mind when considering the use of this feature, specifically:

Enhanced support is provided in IBM zEnterprise® 196 or 114 servers that allow the PPRC secondary copy of the IPL, IODF, and stand-alone dump devices for z/OS systems in the GDPS sysplex to also be defined in the alternate subsystem set (MSS1). With this support, a client can define all z/OS PPRCed devices belonging to the GDPS sysplex uniformly with their secondary in the alternate subchannel set. This removes the necessity to define IPL, IODF, and stand-alone dump devices differently in MSS0.

The following limitations still apply:

- ▶ Fixed Block (FB) open-LUN devices managed by GDPS (these devices are not defined in any z/OS systems and do not consume any unit control blocks (UCBs), and therefore do not contribute to UCB constraints).
- ▶ FlashCopy target devices managed by GDPS.

Multi-Target Metro Mirror (MTMM)

Multi-target PPRC, also known as MT-PPRC, is based on the PPRC technology. The MT-PPRC architecture allows multiple secondary, synchronous, or asynchronous PPRC targets from a single primary device.

Multi-Target Metro Mirror (MTMM) is a specific topology based on the MT-PPRC technology which allows maintaining two synchronous PPRC secondary targets (two PPRC legs) from a single primary device. Each leg is tracked and managed independently.

- ▶ Data is transferred to both targets in parallel.
- ▶ Pairs operate independent of each other.
- ▶ Pairs may be established, suspended or removed separately.
- ▶ A replication problem on one leg does not affect the other leg.
- ▶ HyperSwap is possible on either leg.

MTMM provides all the benefits of PPRC synchronous mirroring plus has the additional protection of a second synchronous leg.

Summary

PPRC synchronous mirroring gives you the ability to remote copy your data in real time, with the potential for no data loss at the recovery site. PPRC is your only choice if your RPO is zero. PPRC is the underlying remote copy capability that the GDPS/PPRC, GDPS/PPRC HyperSwap Manager, and the GDPS Virtual Appliance offerings are built on. The multi-target version of PPRC, called MTMM, is the underlying remote copy architecture that the GDPS/MTMM offering is built on.

2.4.2 XRC (z/OS Global Mirror)

The Extended Remote Copy (XRC) solution consists of a combination of software and hardware functions. XRC maintains a copy of the data asynchronously at a remote location. It involves a System Data Mover (SDM) that is a component of the z/OS operating system working with supporting microcode in the primary disk subsystems. One or more SDMs running in the remote location are channel-attached to the primary disk subsystems. They periodically pull the updates from the primary disks, sort them in time stamp order, and apply the updates to the secondary disks. This provides point-in-time consistency for the secondary disks. The IBM implementation of XRC is branded as z/OS Global Mirror. This name is used interchangeably with XRC in many places, including in this book.

Recovery point objective

Because XRC collects the updates from the primary disk subsystem some time after the I/O has completed, there will always be an amount of data that has not been collected when a disaster hits. As a result, XRC can be used only when your recovery point objective is greater than zero (0). The amount of time that the secondary volumes lag behind the primary depends mainly on the following items:

- ▶ The performance of the SDM

The SDM is responsible for collecting, sorting, and applying all updates. If insufficient capacity (MIPS, storage, and I/O resources) is available to the SDM, longer delays collecting the updates from the primary disk subsystems will occur, causing the secondaries to drift further behind during peak times.

- ▶ The amount of bandwidth

If there is insufficient bandwidth to transmit the updates in a timely manner, contention on the remote copy links can cause the secondary volumes to drift further behind at peak times.

- ▶ The use of device blocking

Enabling blocking for devices results in I/O write activity to be paused for devices with high update rates. This allows the SDM to offload the write I/Os from cache, resulting in a smaller RPO.

- ▶ The use of write pacing

Enabling write pacing for devices with high write rates results in delays being inserted into the application's I/O response to prevent the secondary disk from falling behind. This option slows the I/O activity, resulting in a smaller RPO; it is less disruptive than device blocking. Write pacing, if wanted, can be used in conjunction with the z/OS Workload Manager (WLM).

Because XRC is able to pace the production writes, it is possible to provide an average RPO of 1 to 5 seconds and maintain a guaranteed maximum RPO, if sufficient bandwidth and resources are available. However, it is possible that the mirror will suspend, or that production workloads will be impacted, if the capability of the replication environment is exceeded because of either of the following reasons:

- ▶ Unexpected peaks in the workload
- ▶ An underconfigured environment

To minimize the lag between the primary and secondary devices, you must have sufficient connectivity and a well-configured SDM environment. For more information about planning for the performance aspects of your XRC configuration, see the chapter about capacity planning in *DFSMS Extended Remote Copy Installation Planning Guide*, GC35-0481.

Supported platforms

There are two aspects to “support” for XRC. The first aspect is the ability to append a time stamp to all write I/Os so the update can subsequently be remotely copied by an SDM. This capability is provided in the following operating systems:

- ▶ Any supported release of z/OS
- ▶ Linux on z Systems when using CKD format disks
- ▶ z/VM with STP and appropriate updates (contact IBM support for the most current details)

Note: XRC does *not* support FB devices.

It is also possible to use XRC to remote copy volumes being used by z Systems operating systems that do *not* time stamp their I/Os. However, in this case, it is not possible to provide consistency across multiple LSSs. The devices must all be in the same LSS to provide consistency. For more information, see the section about understanding the importance of timestamped writes in the most recent revision of *z/OS DFSMS Advanced Copy Services* manual.

The other aspect is which systems can run the System Data Mover function. In this case, the only system that supports this is any supported release of z/OS.

Distance and performance

Because XRC is an asynchronous remote copy capability, the amount of time it takes to mirror the update to the remote disks does not affect the response times to the primary volumes. As a result, virtually unlimited distances between the primary and secondary disk subsystems are supported, with minimal impact to the response time of the primary devices.

Connectivity

If the recovery site is within the distance supported by a direct FICON connection, switches/directors, or DWDM, then you can use one of these methods to connect the SDM system to the primary disk subsystem. Otherwise, you must use channel extenders and telecommunication lines.

XRC data consistency

XRC uses time stamps and consistency groups to ensure that your data is consistent across the copy operation. When an XRC pair is established, the primary disk subsystem notifies all systems with a logical path group for that device, and the host system DFSMSdfp software starts to time stamp all write I/Os to the primary volumes. This is necessary to provide data consistency.

XRC is implemented in a cooperative way between the disk subsystems in the primary site and the SDMs, which typically are in the recovery site. A brief outline of the data flow follows (see Figure 2-4):

1. The primary system writes to the primary volumes.
2. Primary disk subsystem posts I/O complete.

Your application I/O is signaled completed when the data is written to the primary disk subsystem's cache and NVS. Channel End (CE) and Device End (DE) are returned to the writing application. These signal that the updates have completed successfully. A time stamped copy of the update is kept in the primary disk subsystems cache. Dependent writes can proceed now.

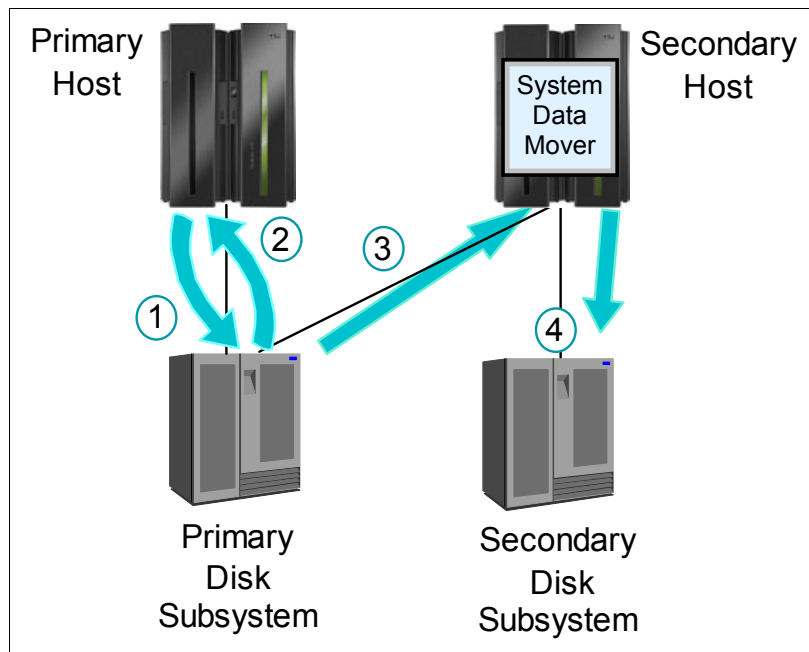


Figure 2-4 Data flow when using z/OS Global Mirror

3. Offload data from primary disk subsystem to SDM.

Every so often (several times a second), the SDM requests each of the primary disk subsystems to send any updates that have been received. The updates are grouped into record sets, which are asynchronously offloaded from the cache to the SDM system.

Within the SDM, the record sets, perhaps from multiple primary disk subsystems, are processed into consistency groups (CGs) by the SDM. The CG contains records that have their order of update preserved across multiple disk subsystems participating in the same XRC session. This preservation of order is vital for dependent write I/Os such as databases and logs. The creation of CGs guarantees that XRC applies the updates to the secondary volumes with update sequence integrity for any type of data.

4. Write to secondary.

When a CG is formed, it is written from the SDM's buffers to the SDM's journal data sets. Immediately after the CG has been hardened on the journal data sets, the records are written to their corresponding secondary volumes. Those records are also written from the SDM's buffers.

5. The XRC control data set is updated to reflect that the records in the CG have been written to the secondary volumes.

Coupled Extended Remote Copy

XRC is an effective solution for mirroring many thousands of volumes. However, a single SDM instance can manage replication only for a finite number of devices. You can use the Coupled XRC (CXRC) support to extend the number of devices for added scalability.

CXRC provides the ability to couple multiple SDMs running in the same or different LPARs together into a *master* session. CXRC coordinates the consistency of data across coupled sessions in a master session, allowing recovery of data for all the volumes in the coupled sessions to a consistent time.

If the sessions are not coupled, recoverable consistency is provided only within each individual SDM, not across SDMs. All logically related data (for example, all the data used by a single sysplex) should be copied by one SDM, or a single group of coupled SDMs.

Multiple Extended Remote Copy

In addition to the additional capacity enabled by Coupled XRC, there is also an option called Multiple XRC (MXRC). MXRC allows you to have up to 20 SDMs in a single LPAR, of which 13 can be coupled together into a cluster. These can then be coupled to SDMs or clusters running in other LPARs through CXRC. Up to 14 SDM clusters can then be coupled together, allowing for an architectural limit of coupled consistency across 182 SDMs.

Multiple Reader

XRC Multiple Reader (also known as Extended Reader) allows automatic load balancing over multiple readers in an XRC environment. A *reader* is a task that is responsible for reading updates from a primary LSS. Depending on the update rate for the disks in an LSS, a reader task might not be able to keep up with pulling these updates and XRC could fall behind. The function can provide increased parallelism through multiple SDM readers and improved throughput for XRC remote mirroring configurations.

It can allow XRC to do these tasks:

- ▶ Better sustain peak workloads for a given bandwidth
- ▶ Increase data currency over long distances
- ▶ Replicate more capacity while maintaining the same recovery point objective
- ▶ Help avoid potential slowdowns or suspends caused by I/Os that are not being processed fast enough

Before the introduction of Multiple Readers, you needed to plan carefully to balance the primary volume update rate versus the rate at which the SDM could “drain” the data. If the drain rate was unable to keep up with the update rate, there was a potential to affect application I/O performance.

GDPS/XRC can use this multireader function, and thus provide these benefits.

Extended Distance FICON

Extended Distance FICON is an improvement focused on providing XRC clients a choice of selecting less complex channel extenders built on frame forwarding technology rather than channel extenders that need to emulate XRC read commands to optimize the channel transfer through the channel extender to get the best performance.

Extended distance FICON enables mirroring over longer distances without substantial reduction of effective data rate. It can significantly reduce the cost of remote mirroring over FICON for XRC.

Extended Distance FICON is supported only on the IBM system z10™ and later servers, and the IBM System Storage DS8000 disk subsystems.

SDM offload to zIIP

The System Data Mover (SDM) is allowed to run on one of the specialty engines that are referred to as a z Systems Integrated Information Processor (zIIP), which are offered on IBM System z9® and later processors. By offloading some of the SDM workload to a zIIP, better price performance and improved use of resources at the mirrored site can be achieved.

One benefit is that DFSMS SDM processing is redirected to a zIIP processor, which can lower server use at the mirrored site. Another benefit is that with an investment of a zIIP specialty processor at the mirrored site, you might now be able to cost-justify the investment in and implementation of a disaster recovery solution that can lower server use at the mirrored site, while at the same time reduce software and hardware fees.

Scalability in a GDPS/XRC environment

As clients implement IT resiliency solutions that rely on multiple copies of data, more are finding that the z/OS limit of 64K (65,536) devices is limiting their ability to grow. Clients can consolidate data sets to fewer larger volumes, but even with that, there are times when this might not make operational sense for all types of data.

In an XRC replication environment, the SDM system or systems are responsible for performing replication. An SDM system will need to address a small number of XRC infrastructure volumes plus the primary and secondary XRC devices that it is responsible for and possibly the FlashCopy target devices. This means, assuming target FlashCopy devices are also defined to the SDM system, that each SDM system can manage XRC replication for up to roughly 21K primary devices. However, as described in “Multiple Extended Remote Copy” on page 30 and “Coupled Extended Remote Copy” on page 30, it is possible to run multiple clustered and coupled SDMs across multiple z/OS images. As you can see, you have more than ample scalability.

Furthermore, in a GDPS/XRC environment it is possible to use “no UCB” FlashCopy, in which case you do not need to define the FlashCopy target devices to the SDM systems. This further increases the number of devices each SDM system can handle.

Hardware prerequisites

XRC requires, on IBM disk subsystems, that primary IBM disk subsystems have the IBM z/OS Global Mirror feature code installed. It is not necessary for the primary and secondary disks to be the same device type, although they must both have the same geometry and the secondary device must be at least as large as the primary device.

XRC is also supported on disk subsystems from other vendors that have licensed and implemented the interfaces from IBM, and it is possible to run with a heterogeneous environment with multiple vendors’ disks. Target XRC volumes can also be from any vendor, even if the target subsystem does not support XRC, thus enabling investment protection.

Note: Keep in mind that at some point, you might have to remote copy from the recovery site back to the production site. GDPS/XRC defines procedures and provides specific facilities for switching your production workload between the two regions.

To reverse the XRC direction, the IBM z/OS Global Mirror feature code must also be installed in the secondary disk subsystems that will become primary when you reverse the replication direction. To reverse the replication direction, the primary and secondary devices must be the same size.

In summary, it makes sense to maintain a symmetrical configuration across both primary and secondary devices.

An extra requirement is that all the systems writing to the primary volumes must be connected to the same STP network. It is not necessary for them all to be in the same *sysplex*, simply that they all share the same *time source*.

Summary

XRC offers a proven disk mirroring foundation for an enterprise disaster recovery solution that provides large scalability and good performance.

XRC is a preferred solution if your site has these requirements:

- ▶ Extended distances between primary and recovery site
- ▶ Consistent data, at all times, in the recovery site
- ▶ Ability to maintain the highest levels of performance on the primary system
- ▶ Can accept a small time gap between writes on the primary system and the subsequent mirroring of those updates on the recovery system
- ▶ Scale with performance to replicate a large number of devices with consistency
- ▶ Run with a heterogeneous environment with multiple vendors' disks

2.4.3 Global Mirror

Global Mirror is an asynchronous remote copy technology that enables a 2-site disaster recovery and backup solution for the z Systems and distributed systems environments. Using asynchronous technology, Global Mirror operates over Fibre Channel Protocol (FCP) communication links and maintains a consistent and restartable copy of data at a remote site that can be located at virtually unlimited distances from the local site.

Global Mirror works by using three sets of disks, as shown in Figure 2-5 on page 33. Global Copy (PPRC Extended Distance, or PPRC-XD), which is an asynchronous form of PPRC, is used to continually transmit data from the primary (A) to secondary (B) volumes, using the out-of-sync bitmap to determine what needs to be transmitted. Global Copy does not guarantee that the arriving writes at the local site are applied to the remote site in the same sequence. Therefore, Global Copy by itself does not provide data consistency.

If there are multiple physical primary disk subsystems, one of them is designated as the *Master* and is responsible for coordinating the creation of consistency groups. The other disk subsystems are subordinates to this Master.

Each primary device maintains two bitmaps. One bitmap tracks incoming changes. The other bitmap tracks which data tracks must be sent to the secondary before a consistency group can be formed in the secondary.

Periodically, depending on how frequently you want to create consistency groups, the Master disk subsystem will signal the subordinates to pause application writes and swap the change recording bitmaps. This identifies the bitmap for the next consistency group. While the I/Os are paused in all LSSs in the Global Mirror session, any dependent writes will not be issued because the CE/DE has not been returned. This maintains consistency across disk subsystems. The design point to form consistency groups is 2 - 3 ms.

After the change recording bitmaps are swapped, write I/Os are resumed and the updates that remain on the Global Mirror primary for the current consistency group will be drained to the secondaries. After all of the primary devices have been drained, a FlashCopy command is sent to the Global Mirror secondaries (B), which are also the FlashCopy source volumes, to perform a FlashCopy to the associated FlashCopy target volumes (C). The tertiary or C copy is a consistent copy of the data. Remember the B volumes are secondaries to Global Copy and are not guaranteed to be consistent. The C copy provides a “golden copy” which can be used to make the B volumes consistent in case recovery is required. Immediately after the FlashCopy process is logically complete, the primary disk subsystems are notified to continue with the Global Copy process. For more information about FlashCopy, see 2.6, “FlashCopy” on page 38.

After Global Copy is resumed, the secondary or B volumes are inconsistent. However, if there is a need for recovery, the FlashCopy target volumes provide the consistent data for recovery.

All this processing is done under the control of microcode in the disk subsystems. You can have up to 16 mirrored pairs in a pool, one of which is the Master primary and secondary pair.

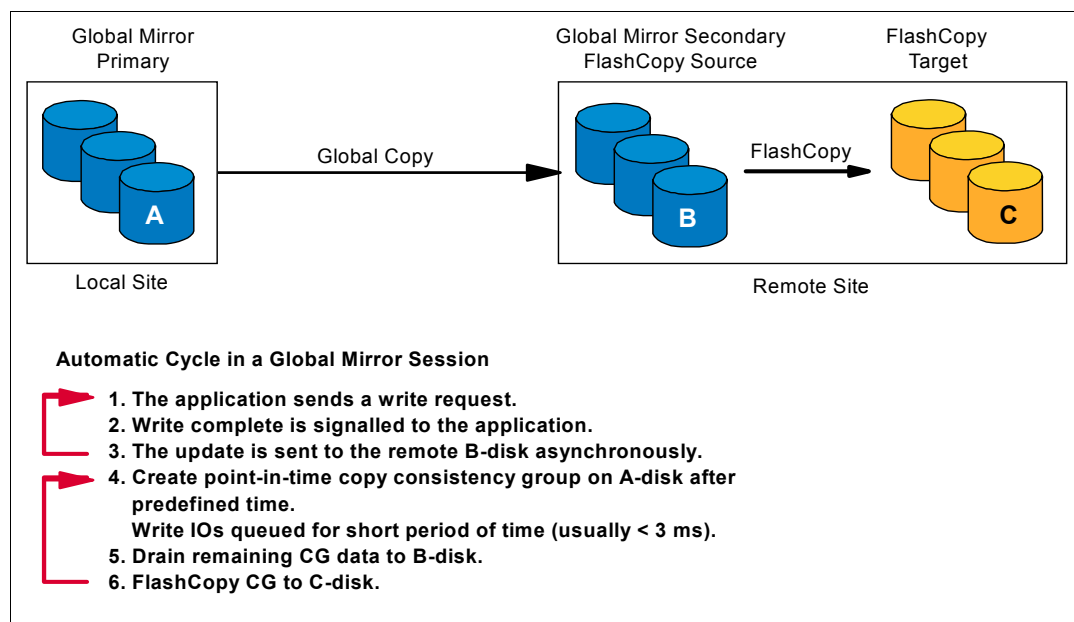


Figure 2-5 Global Mirror: How it works

Recovery point objective

Because Global Mirror is an asynchronous remote copy solution, there will always be an amount of data that must be re-created following a disaster. As a result, Global Mirror can be used only when your recovery point objective (RPO) requirement is greater than zero (0). The amount of time that the FlashCopy target volumes lag behind the primary depends mainly on the following items:

- How often consistency groups are built

This is controlled by the installation and can be specified in terms of seconds.

- The amount of bandwidth

If there is insufficient bandwidth to transmit the updates in a timely manner, contention on the remote copy links can cause the secondary volumes to drift further behind at peak times. The more frequently you create consistency groups, the more bandwidth you will require.

Although it is not unusual to have an average RPO of 5 - 10 seconds with Global Mirror, it is possible that the RPO will increase significantly if production write rates exceed the available resources. However, unlike z/OS Global Mirror, the mirroring session will not be suspended and the production workload will not be impacted if the capacity of the replication environment is exceeded because of unexpected peaks in the workload or an underconfigured environment.

To maintain a consistent lag between the primary and secondary disk subsystems, you must have sufficient connectivity. For more information about planning for the performance aspects of your Global Mirror configuration, see *IBM DS8870 Copy Services for IBM z Systems*, SG24-6787.

Supported platforms

The IBM Enterprise Storage Server and DS8000 families of disk subsystems support Global Mirror. For other enterprise disk vendors, contact your vendor to determine whether they support Global Mirror and if so, on which models.

Distance and connectivity

Because Global Mirror is an asynchronous remote copy capability, the amount of time it takes to mirror the update to the remote disks does not affect the response times to the primary volumes. As a result, virtually unlimited distances between the primary and secondary disk subsystems are supported.

Global Mirror requires FCP links on the disk subsystems. If the recovery site is within the distance supported by FCP direct connect, switches, or DWDM, you can use one of those methods to connect the primary and secondary disk subsystems. Otherwise, you must use network extension technology that supports FCP links.

Addressing z/OS device limits in a GDPS/GM environment

As clients implement IT resiliency solutions that rely on multiple copies of data, more are finding that the z/OS limit of 64K (65,536) devices is limiting their ability to grow. Clients can consolidate data sets to fewer larger volumes, but even with that, there are times when this might not make operational sense for all types of data.

To this end, z/OS introduced the concept of an *alternate subchannel set*, which can include the definition for certain types of disk devices. An alternate subchannel set provides another set of 64K devices for the following device types:

- Parallel Access Volume (PAV) alias devices
- PPRC secondary devices (defined as 3390D)
- FlashCopy target devices

Including PAV alias devices in an alternate subchannel set is transparent to GDPS and is common practice for many client configurations.

The application site controlling system requires access to the GM primary devices and can address up to nearly 64K devices. The recovery site controlling system requires access to both the GM secondary and the GM FlashCopy devices. GDPS supports defining the GM FlashCopy devices in an alternate subchannel set (MSS1). This allows up to nearly 64K devices to be replicated in a GDPS/GM environment.

Summary

Global Mirror provides an asynchronous remote copy offering that supports virtually unlimited distance, without the requirement of an SDM system to move the data from primary to secondary volumes. Global Mirror also supports a wider variety of platforms because it supports FB devices and removes the requirement for timestamped updates that is imposed by XRC.

Conversely, Global Mirror is currently not as scalable as XRC because it supports only a maximum of 17 storage subsystems. In addition, Global Mirror does not have the multiple vendor flexibility provided by XRC.

2.4.4 Combining disk remote copy technologies for CA and DR

In this section we briefly describe Metro/Global Mirror and Metro/z/OS Global Mirror. For more detailed information, see Chapter 11, “Combining local and metro continuous availability with out-of-region disaster recovery” on page 331. Combining the technologies of Metro Mirror and HyperSwap with either Global Mirror or XRC (also referred to as *z/OS Global Mirror* in this section) allows clients to meet requirements for continuous availability (CA) with zero data loss locally within metropolitan distances for most failures, along with providing a disaster recovery (DR) solution in the case of a region-wide disaster. This combination might also allow clients to meet increasing regulatory requirements.

Metro Global Mirror

Metro Global Mirror (MGM) is a cascading data replication solution that combines the capabilities of Metro Mirror and Global Mirror.

Synchronous replication between a primary and secondary disk subsystem located either within a single data center, or between two data centers located within metropolitan distances, is implemented using Metro Mirror. Global Mirror is used to asynchronously replicate data from the secondary disks to a third disk subsystem in a recovery site typically out of the local metropolitan region. As described in 2.4.3, “Global Mirror” on page 32, a fourth set of disks, also in the recovery site, are the FlashCopy targets used to provide the consistent data for disaster recovery. Because both Metro Mirror and Global Mirror are hardware-based remote copy technologies, CKD and FB devices can be mirrored to the recovery site protecting both z Systems and open system data.

For enterprises that require consistency across both distributed systems and z Systems data, MGM provides a comprehensive three-copy data replication strategy to protect against day-to-day disruptions, while protecting critical business data and functions if there is a wide-scale disruption.

Metro z/OS Global Mirror

GDPS Metro/z/OS Global Mirror (MzGM) is a multi-target data replication solution that combines the capabilities of Metro Mirror and XRC (z/OS Global Mirror).

Synchronous replication between a primary and secondary disk subsystem located either within a single data center, or between two data centers located within metropolitan distances, is implemented using Metro Mirror. XRC is used to asynchronously replicate data from the primary disks to a third disk system in a recovery site, typically out of the local metropolitan region.

Because XRC supports only CKD devices, only z Systems data can be mirrored to the recovery site. However, because both PPRC and XRC are supported by multiple storage vendors, this solution provides flexibility that MGM cannot.

For enterprises looking to protect z Systems data, MzGM delivers a three-copy replication strategy to provide continuous availability for day-to-day disruptions, while protecting critical business data and functions if there is a wide-scale disruption.

2.4.5 IBM software replication products

This section does not aim to provide a comprehensive list of all IBM software-based replication products. Instead, it provides an introduction to the following supported products within the GDPS/Active-Active solution:

- ▶ InfoSphere Data Replication for IMS for z/OS
- ▶ InfoSphere Data Replication for VSAM for z/OS
- ▶ InfoSphere Data Replication for DB2 for z/OS

These products provide the capability to asynchronously copy changes to data held in IMS or DB2 databases or VSAM files from a source to target copy. Fine-grained controls allow you to precisely define what data is critical to your workload and needs to be copied in real time between the source and target. Unlike disk replication solutions that are application, or data-agnostic and work at the z/OS volume level, software replication does not provide a mechanism for copying all possible data types in your environment. As such, it is suited to provide only a CA/DR solution for specific workloads that can tolerate only the IMS, DB2 or VSAM database-resident information to be copied between locations. This is also discussed in Chapter 8, “GDPS/Active-Active solution” on page 231.

InfoSphere Data Replication for IMS for z/OS

IMS Replication provides the mechanisms for producing copies of your IMS databases and maintaining the currency of the data in near real time, typically between two systems separated by geographic distances. There is essentially no limit to the distance between source and target systems because the copy technique is asynchronous and uses TCP/IP as the protocol to transport the data over your wide area network (WAN).

IMS replication employs Classic data servers in the source and target systems to provide the replication services.

Classic source server

The Classic source server reads the IMS log data and packages changes to the specified databases into messages that are then sent through TCP/IP to the target location.

Classic target server

The Classic target server, running in the target location, receives messages from the source server and applies the changes to a replica of the source IMS database in near real time. IMS replication provides mechanisms to ensure that updates to a given record in the source database are applied in the same sequence in the target replica. Furthermore, IMS replication maintains a *bookmark* to know where it has reached in processing the IMS log data so that if any planned or unplanned outage occurs, it can later catch up knowing where it was at the time of the outage.

For details, see “InfoSphere IMS Replication for z/OS V10.1” in IBM Knowledge Center:

<http://ibm.co/1FSsSPc>

InfoSphere Data Replication for VSAM for z/OS

VSAM replication is similar in structure to IMS replication. For CICS/VSAM workloads, the transaction data for selected VSAM data sets is captured using the CICS log streams as the source. For non-CICS workloads, CICS VSAM Recovery (CICS VR) logs are used as the source for capturing VSAM update information. The updates are transmitted to the target using TCP/IP, where they are applied to the target data sets upon receipt.

InfoSphere Data Replication for DB2 for z/OS

InfoSphere Replication Server for z/OS, as used in the GDPS/Active-Active solutions, is also known as *Q replication*. It provides a high capacity and low latency replication solution that uses IBM WebSphere® MQ message queues to transmit data updates between source and target tables of a DB2 database.

Q replication is split into two distinct pieces:

- ▶ Q capture program or engine
- ▶ Q apply program or engine

Q capture

The Q capture program reads the DB2 logs or changes to the source table or tables that you want to replicate. These changes are then put into WebSphere MQ messages and sent across the WebSphere MQ infrastructure to the system where the target table resides. There, they are read and applied to the target table by the Q apply program.

The Q capture program is flexible in terms of what can be included or excluded from the data sent to the target and even the rate at which data is sent can be modified if required.

By the nature of the method of Q replication, the replication of data is an asynchronous process. Even so, an RPO of a few seconds is possible even in high update environments.

Q apply

The Q apply program takes WebSphere MQ messages from a receive queue, or queues and then applies the changes held within the message to the target tables. The Q apply program is designed in such a way to use parallelism to keep up with updates to multiple targets while maintaining any referential integrity constraints between related target tables.

Both the Q capture and Q apply programs have mechanisms to track what has been read from the logs and sent to the target site, and what has been read from the receive queues and applied to the target tables, including any dependencies between updates. This in turn provides data consistency and allows for restart of both the capture and apply programs, if this is required or in case of failures.

For more information about Q replication, see “Q Replication and Event Publishing” in IBM Knowledge Center:

<http://ibm.co/1QnW6sc>

2.5 Tape resident data

Operational data, that is, data that is used directly by applications supporting users, is normally found on disk. However, there is another category of data (called *support data*) that supports the operational data; this often resides in tape subsystems. Support data typically covers migrated data, point-in-time backups, archive data, and so on. For sustained operation in the failover site, the support data is indispensable. Furthermore, some enterprises have mission-critical data that resides only on tape. You need a solution to ensure that tape data is readily accessible at your recovery site.

Just as you mirror your disk-resident data to protect it, similarly you can mirror your tape-resident data. GDPS provides support for management of the IBM Virtualization Engine TS7700⁵. See 3.1.2, “Protecting tape data” on page 65 for details about GDPS TS7700 support. The IBM Virtualization Engine TS7700 provides comprehensive support for replication of tape data. See *IBM Virtualization Engine TS7700 with R 2.0*, SG24-7975 for more information about the TS7700 technology that complements GDPS for tape data.

2.6 FlashCopy

FlashCopy provides a point-in-time (PIT) copy of a volume, with almost instant availability for the user of both the source and target volumes. There is also a data set-level FlashCopy supported for z/OS volumes. Only a minimal interruption is required for the FlashCopy relationship to be established. The copy is then created by the disk subsystem, with minimal impact on other disk subsystem activities. The volumes created when you use FlashCopy to copy your secondary volumes are called *tertiary volumes*.

FlashCopy and disaster recovery

FlashCopy has specific benefits in relation to disaster recovery. For example, consider what happens if you temporarily lose connectivity between primary and secondary PPRC volumes. At the point of failure, the secondary volumes will be consistent. However, during the period when you are resynchronizing the primary and secondary volumes, the secondary volumes are inconsistent (because the updates are not applied in the same time sequence that they were written to the primaries). So, what happens if you have a disaster during this period? If it is a real disaster, your primary disk subsystem will be a smoldering lump of metal on the computer room floor. And your secondary volumes are inconsistent, so those volumes are of no use to you either.

So, how do you protect yourself from such a scenario? One way (our suggested way) is to take a FlashCopy of the secondary volumes just before you start the resynchronization process. This at least ensures that you have a consistent set of volumes in the recovery site. The data might be several hours behind the primary volumes, but even data a few hours old that is consistent is better than current, but unusable, data.

⁵ The TS7700 management support is available only in GDPS/PPRC at this time.

An additional benefit of FlashCopy is that it provides the ability to perform disaster recovery tests while still retaining disaster recovery readiness. The FlashCopy volumes you created when doing the resynchronization (or subsequently) can be used to enable frequent testing (thereby ensuring that your recovery procedures continue to be effective) without having to use the secondary volumes for that testing.

FlashCopy can operate in several modes. GDPS uses one of the following modes of FlashCopy, depending on the GDPS offering:

COPY	When the volumes are logically copied, the FlashCopy session continues as a background operation, physically copying all the data from the source volume to the target. When the volumes have been physically copied, the FlashCopy session ends. In this mode, the FlashCopy target physical volume will be a mirror image of the source volume at the time of the FlashCopy.
NOCOPY	When the volumes are logically copied, a FlashCopy session continues as a background operation, physically copying only those tracks subsequently updated by write operations to the source volume. In this mode, the FlashCopy target physical volume contains only data that was changed on the source volume after the FlashCopy.
NOCOPY2COPY	Change existing FlashCopy relationship from NOCOPY to COPY. This can be done dynamically. When one or more NOCOPY relationships exist for a source volume, NOCOPY2COPY will initiate a background copy for all target relationships with intersecting source extents from the point in time the NOCOPY was issued. Upon completion of the background copy, the converted relationship or relationships will be terminated.
INCREMENTAL	This allows repetitive FlashCopies to be taken, but only the tracks that have changed since the last FlashCopy will be copied to the target volume. This provides the ability to refresh a FlashCopy relationship and bring the target up to the source's newly established point-in-time. Incremental FlashCopy helps reduce the background copy completion time when only a subset of data on either the source or target has changed, thus giving you the option to perform a FlashCopy on a more frequent basis.
CONSISTENT	This option is applicable to GDPS/PPRC and GDPS/HM environments. It creates a consistent set of tertiary disks without suspending the PPRC mirror. It uses the FlashCopy Freeze capability which, similar to PPRC Freeze, puts all source disks in Extended Long Busy to ensure that the FlashCopy source disks are consistent before the point-in-time copy is made. After the source disks are consistent, the FlashCopy is taken (quite fast) and the Freeze is thawed. Without this support, to produce a consistent point-in-time copy of the secondary disks, you would need to suspend the PPRC mirror (planned freeze) and then resynchronize PPRC. HyperSwap would remain disabled from the time you suspended PPRC until the mirror is full-duplex again; however, this can take a long time depending on how much data was updated while PPRC remained suspended. In comparison, with Consistent FlashCopy, HyperSwap is only disabled during the FlashCopy Freeze, which should be simply a few seconds. GDPS gives you the capability to restrict the FlashCopy Freeze duration and to abort the FlashCopy operation if the FlashCopy Freeze time exceeds your threshold.

To create a consistent point-in-time copy of the primary disks without Consistent FlashCopy, you would need to somehow make sure that there is no I/O on the primary disks (effectively, you would need to stop the production systems). With Consistent FlashCopy, production systems continue to run and I/O is prevented during the few seconds until the FlashCopy Freeze completes. After the FlashCopy Freeze completes, the primary disks are in a consistent state, the FlashCopy operation itself is quite fast, and then the freeze is thawed and production systems resume I/O. Consistent FlashCopy can be used in conjunction with COPY, NOCOPY, or INCREMENTAL FlashCopy.

Zero Suspend

This option is applicable to GDPS/XRC environments. It creates a recoverable set of tertiary disks for recovery testing with no suspension of the XRC operation. This allows DR testing to be performed without ever losing the DR capability. Before this support, to produce a consistent tertiary copy you needed to suspend XRC for all volumes, FlashCopy secondary volumes, and then resynchronize XRC sessions.

If you plan to use FlashCopy, remember that the source and target volumes must be within the same physical disk subsystem. This is a capacity planning consideration when configuring and planning for the growth of your disk subsystems.

Also remember that if you performed a site switch to run in the recovery site, at some point you will want to return to the production site. To provide equivalent protection and testing capability no matter which site you are running in, consider providing FlashCopy capacity in both sites.

Furthermore, GDPS does not perform FlashCopy for simply selected volumes. The GDPS use of FlashCopy is for the purposes of protection during resynchronization and for testing. Both of these tasks require that a point-in-time copy for the entire configuration is made. GDPS FlashCopy support assumes that you will provide FlashCopy target devices for the entire configuration and that every time GDPS performs a FlashCopy, it will be for all secondary devices (GDPS/PPRC also supports FlashCopy for primary devices).

User-initiated FlashCopy

User-initiated FlashCopy supports FlashCopy of all defined FlashCopy volumes using panel commands, GDPS scripts, or GDPS Tivoli NetView for z/OS commands, depending on which GDPS product is used.

Space-efficient FlashCopy (FlashCopy SE)

FlashCopy SE is functionally not much different from the standard FlashCopy. The concept of *space-efficient* with FlashCopy SE relates to the attributes or properties of a DS8000 volume. As such, a space-efficient volume can be used like any other DS8000 volume. However, the intended and only *preferred* use is as a target volume in a FlashCopy relationship.

When a normal volume is created, it occupies the defined capacity on the physical drives. A space-efficient volume does not occupy physical capacity when it is initially created. Space gets allocated when data is actually written to the volume. This allows the FlashCopy target volume capacity to be thinly provisioned (that is, smaller than the full capacity of the source volume). In essence this means that when planning for FlashCopy, you may provision less disk capacity when using FlashCopy SE than when using standard FlashCopy, which can help lower the amount of physical storage needed by many installations.

All GDPS products support FlashCopy SE. Details of how FlashCopy SE is used by each offering is described in the chapter related to that offering.

2.7 Automation

If you have challenging recovery time and recovery point objectives, implementing disk remote copy, software-based replication, tape remote copy, FlashCopy, and so on are necessary prerequisites for you to be able to recover from a disaster and meet your objectives. However, be sure you realize that they are only enabling technologies. To achieve the stringent objectives placed on many IT departments today, it is necessary to tie those technologies together with automation and sound systems management practices. In this section we discuss your need for automation to recover from an outage.

2.7.1 Recovery time objective

If you have reached this far in the document, we presume that your recovery time objective (RTO) is a “challenge” to you. If you have performed tape-based disaster recovery tests, you know that ensuring that all your data is backed up is only the start of your concerns. In fact, even getting all those tapes restored does not result in a mirror image of your production environment. You also need to get all your databases up to date, get all systems up and running, and then finally start all your applications.

Trying to drive all this manually will, without question, prolong the whole process. Operators must react to events as they happen, while consulting recovery documentation. However, automation responds at machine speeds, meaning your recovery procedures will be executed without delay, resulting in a shorter recovery time.

2.7.2 Operational consistency

Think about an average computer room scene immediately following a system failure. All the phones are ringing. Every manager within reach moves in to determine when everything will be recovered. The operators are frantically scrambling for procedures that are more than likely outdated. And the systems programmers are all vying with the operators for control of the consoles; in short, chaos.

Imagine, instead, a scenario where the only manual intervention is to confirm how to proceed. From that point on, the system will recover itself using well-tested procedures. How many people watch it does not matter because it will not make mistakes. And you can yell at it all you like, but it will still behave in exactly the manner it was in which it was programmed to behave. You do not need to worry about outdated procedures being used. The operators can concentrate on handling calls and queries from the assembled managers. And the systems programmers can concentrate on pinpointing the cause of the outage, rather than trying to get everything up and running again.

And all of this is just for a system outage. Can you imagine the difference that well-designed, coded, and tested automation can make in recovering from a real disaster? Apart from speed, perhaps the biggest benefit that automation brings is consistency. If your automation is thoroughly tested, you can be assured that it will behave in the same way, time after time. When recovering from as rare an event as a real disaster, this consistency can be a lifesaver.

2.7.3 Skills impact

Recovering a computing center involves many complex activities. Training staff takes time. People come and go. You cannot be assured that the staff that took part in the last disaster recovery test will be on hand to drive recovery from this real disaster. In fact, depending on the nature of the disaster, your skilled staff might not even be available to drive the recovery.

The use of automation removes these concerns as potential pitfalls to your successful recovery.

2.7.4 Summary

The technologies you will use to recover your systems all have various control interfaces. Automation is required to tie them all together so they can be controlled from a single point and your recovery processes can be executed quickly and consistently.

Automation is one of the central tenets of the GDPS offerings. By using the automation provided by GDPS, you save all the effort to design and develop this code yourself, and also benefit from the IBM experience with hundreds of clients across your industry and other industries.

2.8 Flexible server capacity

In this section we discuss options for increasing your server capacity concurrently, for either planned upgrades or unplanned upgrades, to quickly provide the additional capacity you will require on a temporary basis. These capabilities can be used for server or site failures, or they can be used to help meet the temporary peak workload requirements of clients.

The only capabilities described in this section are the ones used by GDPS. Other capabilities exist to upgrade server capacity, either on a temporary or permanent basis, but they are not covered in this section.

2.8.1 Capacity Backup upgrade

Capacity Backup (CBU) upgrade for z Systems processors provides reserved emergency backup server capacity that can be activated in lieu of capacity that is lost as a result of an unplanned event elsewhere. CBU helps you to recover by adding reserved capacity on a designated z Systems system. A CBU system normally operates with a base server configuration and with a preconfigured number of additional processors reserved for activation in case of an emergency.

CBU can be used to install (and pay for) less capacity in the recovery site than you have in your production site, while retaining the ability to quickly provision the additional capacity that would be required in a real disaster.

CBU can be activated manually, using the HMC. It can also be activated automatically by GDPS, either as part of a disaster recovery test, or in reaction to a real disaster. Activating the additional processors is nondisruptive. That is, you do not need to power-on reset (POR) the server or even IPL the LPARs that can benefit from the additional capacity (assuming that an appropriate number of reserved CPs were defined in the LPAR Image profiles).

CBU is available for all processor types on IBM z Systems.

The CBU contract allows for an agreed-upon number of tests over the period of the contract. GDPS supports activating CBU for test purposes.

For more information about CBU, see *System z Capacity on Demand User's Guide*, SC28-6846.

2.8.2 On/Off Capacity on Demand

On/Off Capacity on Demand (On/Off CoD) is a function that enables concurrent and temporary capacity growth of the server. The difference between CBU and On/Off CoD is that On/Off CoD is for planned capacity increases, and CBU is intended to replace capacity lost as a result of an unplanned event elsewhere. On/Off CoD can be used for client peak workload requirements, for any length of time, and it has a daily hardware and software charge.

On/Off CoD helps clients, with business conditions that do not justify a permanent upgrade in capacity, to contain workload spikes that might exceed permanent capacity so that Service Level Agreements cannot be met. On/Off CoD can concurrently add processors (CPs, IFLs, ICFs, zAAPs, and zIIPs) up to the limit of the installed books of an existing server. It is restricted to double the currently installed capacity.

2.8.3 GDPS CBU and On/Off CoD handling

The GDPS temporary capacity management capabilities are related to the capabilities provided by the particular server system being provisioned. Processors before the IBM System z10® required that the full capacity for a Capacity Backup (CBU) upgrade or On/Off Capacity on Demand (OOCOD) be activated, even though the full capacity might not be required for the particular situation at hand.

GDPS, in conjunction with System z10 and later generation systems, provides support for activating temporary capacity, such as CBU or OOCOD, based on a preinstalled capacity-on-demand record. In addition to the capability to activate the full record, GDPS also provides the ability to define profiles that determine what will be activated. The profiles are used in conjunction with a GDPS script statement and provide the flexibility to activate the full record or a partial record.

When temporary capacity upgrades are performed using GDPS facilities, GDPS tracks activated CBU and OOCOD resources at a CEC level.

GDPS provides keywords in GDPS scripts to support activation and deactivation of the CBU and On/Off CoD function.

GDPS allows definition of capacity profiles to add capacity to already running systems. Applicable types of reserved engines (CPs, zIIPs, zAAPs, IFLs, and ICFs) can be configured online to GDPS z/OS systems, to xDR-managed z/VM systems, and to coupling facilities that are managed by GDPS.

When a GDPS z/OS system is IPLed, GDPS automatically configures online any applicable reserved engines (CPs, zIIPs, and zAAPs) based on the LPAR profile. The online configuring of reserved engines is done only if temporary capacity was added to the CEC where the system is IPLed using GDPS facilities.

2.9 Cross-site connectivity considerations

When setting up a recovery site, there might be a sizeable capital investment to get started, but you might find that one of the largest components of your ongoing costs is related to providing connectivity between the sites. Also, the type of connectivity available to you can affect the recovery capability you can provide. Conversely, the type of recovery capability you want to provide will affect the types of connectivity you can use.

In this section, we list the connections that must be provided, from a simple disk remote copy configuration through to an Active/Active workload configuration. We briefly review the types of cross-site connections that you must provide for the different GDPS solutions and the technology that must be used to provide that connectivity. All of these descriptions relate solely to cross-site connectivity. We assume that you already have whatever intrasite connectivity is required.

2.9.1 Server-to-disk links

If you want to be able to use disks installed remotely from a system in the production site, you must provide channel connections to those disk control units.

PPRC and MTMM-based solutions

For PPRC and MTMM with GDPS, all of the secondary disks (both sets for MTMM) must be defined to, and be channel-accessible to, the production systems for GDPS to be able to manage those devices.

If you foresee a situation where systems in the production site will be running off the secondary disks (for example, if you will use HyperSwap), you need to provide connectivity equivalent to that provided to the corresponding primary volumes in the production site. The HyperSwap function provides the ability to nondisruptively swap from the primary volume of a mirrored pair to what had been the secondary volume. For more information about HyperSwap, see the following sections in this book:

- ▶ “GDPS HyperSwap function” on page 58 for GDPS/PPRC
- ▶ “GDPS HyperSwap function” on page 108 for GDPS/PPRC HyperSwap Manager
- ▶ “GDPS HyperSwap function” on page 196 for GDPS/MTMM
- ▶ “GDPS HyperSwap function” on page 279 for GDPS Virtual Appliance

If you will not have any cross-site disk accessing, then minimal channel bandwidth (two FICON channel paths from each system to each disk subsystem) is sufficient.

Depending on your director and switch configuration, you might be able to share the director-to-director links between channel and PPRC connections. For more information, see *IBM System z Connectivity Handbook*, SG24-5444.

HyperSwap across sites with less than full channel bandwidth

You might consider enabling unplanned HyperSwap to the secondary disks in the remote site even if you do not have sufficient cross-site channel bandwidth to sustain your production workload for normal operations. Assuming that a disk failure is likely to cause an outage and you will need to switch to using a disk in the other site, the unplanned HyperSwap might at least give you the opportunity to perform an orderly shutdown of your systems first. Shutting down your systems cleanly avoids the complications and longer restart time that is associated with crash-restart of application subsystems.

For GDPS/MTMM environments, the same consideration applies to enabling HyperSwap to the remote secondary copy: Channel bandwidth to the local secondary copy should not be an issue.

XRC-based and Global Mirror-based solutions

For any of the asynchronous remote copy implementations (XRC or Global Mirror), the production systems would normally not have channel access to the secondary volumes.

Software replication-based solutions

As with other asynchronous replication technologies, given that effectively unlimited distances are supported, there is no requirement for the source systems to have host channel connectivity to the data in the target site.

2.9.2 Data replication links

You will need connectivity for your data replication activity:

- ▶ Between storage subsystems (for PPRC or Global Mirror)
- ▶ From the SDM system to the primary disks (for XRC)
- ▶ Across the wide area network for software-based replication

PPRC-based and Global Mirror-based solutions

The IBM Metro Mirror (including MTMM) and Global Mirror implementations use Fibre Channel Protocol (FCP) links between the primary and secondary disk subsystems. The FCP connection can be direct, through a switch, or through other supported distance solutions (for example, Dense Wave Division Multiplexer, DWDM, or channel extenders). Even though some of the older technology disk subsystems support IBM ESCON connections for PPRC, we strongly suggest using FCP links for best performance over distance.

XRC-based solutions

If you are using XRC, the System Data Movers (SDMs) are typically in the recovery site. The SDMs must have connectivity to both the primary volumes and the secondary volumes. The cross-site connectivity to the primary volumes is a FICON connection, and depending on the distance between sites, either a supported DWDM can be used (distances less than 300 km) or a channel extender can be used for longer distances. As discussed in “Extended Distance FICON” on page 30, an enhancement to the industry standard FICON architecture (FC-SB-3) helps avoid degradation of performance at extended distances, and this might also benefit XRC applications within 300 km where channel extension technology had previously been required to obtain adequate performance.

Software-based solutions

Both IMS replication and DB2 replication use your wide area network (WAN) connectivity between the data source and the data target. Typically, for both, either natively or through IBM WebSphere MQ, TCP/IP is the transport protocol used, although other protocols, such as LU6.2, are supported. It is beyond the scope of this book to go into detail about WAN design, but ensure that any such connectivity between the source and target have redundant routes through the network to ensure resilience from failures. There are effectively no distance limitations on the separation between source and target. However, the greater the distance between them will affect the latency and, therefore, the RPO that can be achieved,

2.9.3 Coupling links

Coupling links are required in a Parallel Sysplex configuration to provide connectivity from the z/OS images to the coupling facility. Coupling links are also used to transmit timekeeping messages when Server Time Protocol (STP) is enabled. If you have a multisite Parallel Sysplex, you will need to provide coupling link connectivity between sites.

For distances greater than 10 km, either ISC3 or Parallel Sysplex InfiniBand (PSIFB) Long Reach links must be used to provide this connectivity. The maximum supported distance depends on several things, including the particular DWDMs that are being used and the quality of the links.

Table 2-1 summarizes the distances supported by the various link types.

Table 2-1 Supported CF link distances

Link type	Link data rate	Maximum unrepeated distance	Maximum repeated distance
ISC-3	2 Gbps ^a 1 Gbps ^b	10 km 20 km ^c	200 km
PSIFB Long Reach 1X	5.0 Gbps 2.5 Gbps ^d	10 km	175 km
PSIFB 12X, for use within a data center	6 GBytes/sec 3 GBytes/sec ^e	150 meters	Not applicable

a. Gbps (gigabits per second)

b. RPQ 8P2197 provides an ISC-3 Daughter Card that clocks at 1 Gbps.

c. Requires RPQ 8P2197 and 8P2263 (z Systems Extended Distance).

d. The PSIFB Long Reach feature will negotiate to 1x IB-SDR link data rate of 2.5 Gbps if connected to qualified DWDM infrastructure that cannot support the 5 Gbps (1x IB-DDR) rate.

e. The PSIFB links negotiate to 12x IB-SDR link data rate of 3 GBytes/sec when connected to System z9 servers.

2.9.4 Server Time Protocol

Server Time Protocol (STP) is a server-wide facility that is implemented in the Licensed Internal Code (LIC) of the IBM z Systems servers. It provides the capability for multiple servers to maintain time synchronization with each other. STP is the successor to the 9037 Sysplex Timer.

STP is designed for servers that have been configured to be in a Parallel Sysplex or a basic sysplex (without a coupling facility), and servers that are not in a sysplex, but need to be time-synchronized. STP is a message-based protocol in which timekeeping information is passed over data links between servers. The timekeeping information is transmitted over externally defined coupling links. Coupling links are used to transport STP messages.

If you are configuring a sysplex across two or more sites, you need to synchronize servers in multiple sites. For more information about Server Time Protocol, see *Server Time Protocol Planning Guide*, SG24-7280, and *Server Time Protocol Implementation Guide*, SG24-7281.

2.9.5 XCF signaling

One of the requirements for being a member of a sysplex is the ability to maintain XCF communications with the other members of the sysplex. XCF uses two mechanisms to communicate between systems: XCF signaling structures in a CF and channel-to-channel adapters. Therefore, if you are going to have systems in both sites that are members of the same sysplex, you must provide CF connectivity, CTC connectivity, or preferably both, between the sites.

If you provide both CF structures and CTCs for XCF use, XCF will dynamically determine which of the available paths provides the best performance and use that path. For this reason, and for backup in case of a failure, we suggest providing *both* XCF signaling structures and CTCs for XCF cross-site communication.

2.9.6 HMC and consoles

To be able to control the processors in the remote center, you need to have access to the LAN containing the SEs and HMCs for the processors in that location. Such connectivity is typically achieved using bridges or routers.

If you are running systems at the remote site, you will also want to be able to have consoles for those systems. Two options are 2074 control units and OSA-ICC cards. Alternatively, you can use SNA consoles, but be aware that they cannot be used until IBM VTAM® is started, so they cannot be used for initial system loading.

2.9.7 Connectivity options

Note: WAN connectivity options are not covered in this book. Table 2-2, with the exception of HMC connectivity, is predominantly related to disk replication solutions.

Now that we have explained what you need to connect across the two sites, we briefly review the most common options for providing that connectivity. There are several ways to provide all this connectivity, from direct channel connection through to DWDMs. Table 2-2 shows the different options. The distance supported varies by device type and connectivity method.

Table 2-2 Cross-site connectivity options

Connection type	Direct (unrepeated)	Switch and director or cascaded directors	DWDM	Channel extender
Server to disk	Yes	Yes	Yes	Yes
Disk Remote copy	Yes	Yes	Yes	Yes
Coupling links	Yes	No	Yes	No
Sysplex Timer	Yes	No	Yes	No
STP (coupling links)	Yes	No	Yes	No
XCF signaling	Yes	Yes (CTC) No (coupling links)	Yes	Yes (CTC only) No (coupling links)
HMC/consoles	Yes	Yes	Yes	Yes

For more information about options and distances that are possible, see *IBM System z Connectivity Handbook*, SG24-5444.

FICON switches/directors

For information about z Systems qualified FICON and Fibre Channel Protocol (FCP) products and products that support mixing FICON and FCP within the same physical FC switch or FICON director, see the I/O Connectivity web page:

<http://www.ibm.com/systems/z/connectivity/products/fc.html>

The maximum unrepeatable distance for FICON is typically 10 km. However, FICON switches can be used to extend the distance from the server to the control unit further with the use of a cascaded configuration. The maximum supported distance for the interswitch links (ISL) in this configuration is technology- and vendor-specific.

No matter what the case might be, if the property between the two sites is not owned by your organization, you will need a vendor to provide dark fiber between the two sites because FICON switches/directors cannot be directly connected to telecommunication lines.

For more information about this topic, see *IBM System z Connectivity Handbook*, SG24-5444.

Wavelength Division Multiplexing

A Wavelength Division Multiplexor (WDM) is a high-speed, high-capacity, scalable fiber optic data transport system that uses Dense Wavelength Division Multiplexing (DWDM) or Course Wavelength Division Multiplexing (CWDM) technology to multiplex several independent bit streams over a single fiber link, thereby making optimal use of the available bandwidth.

WDM solutions that support the protocols described in this book generally support metropolitan distances in the range of tens to a few hundred kilometers. The infrastructure requirements and the supported distances vary by vendor, model, and even by features on a given model.

More specifically, several qualified WDM solutions support the following key protocols used in a GDPS solution:

- ▶ Enterprise Systems Connection (ESCON)
- ▶ Fiber Connection (FICON)
- ▶ InterSystem Channel (ISC-3)
- ▶ Parallel Sysplex InfiniBand (PSIFB) Long Reach links
- ▶ Server Time Protocol (STP) over ISC-3 Peer Mode or PSIFB Long Reach
- ▶ Potentially, protocols that are not z Systems protocols

Given the criticality of these links for transport of data and timing information, it is important to use only qualified WDM vendor solutions when extending Parallel Sysplexes to more than one site (as is often done as part of a GDPS configuration).

The latest list of qualified WDM vendor products, along with links to corresponding IBM Redpaper™ publications for each product, is available at the IBM Resource Link® web page:

<https://www.ibm.com/servers/resourcelink/>

See “Hardware products for servers” on the Library page.

Channel extenders

Channel extenders are special devices that are connected in the path between a server and a control unit, or between two control units. Channel extenders provide the ability to extend connections over much greater distances than that provided by DWDM. Distances supported with channel extenders are virtually unlimited⁶.

Unlike DWDMs, channel extenders support connection to telecom lines, removing the need for dark fiber. This can make channel extenders more flexible because access to high-speed telecoms is often easier to obtain than access to dark fiber.

However, channel extenders typically do not support the same range of protocols as DWDMs. In a z Systems context, channel extenders support IP connections (for example, connections to OSA adapters), FCP and FICON channels, but not coupling links or time synchronization-related links.

For much more detailed information about the options and distances that are possible, see *IBM System z Connectivity Handbook*, SG24-5444.

More information about channel extenders that have been qualified to work with IBM storage is available to download from the DS8000 Series Copy Services Fibre Channel Extension Support Matrix web page:

<http://www.ibm.com/support/docview.wss?uid=ssg1S7003277&rs=1329>

2.9.8 Single points of failure

When planning to connect systems across sites, it is vital to do as much as you possibly can to avoid all single points of failure. Eliminating all single points of failure makes it significantly easier to distinguish between a connectivity failure and a failure of the remote site. The recovery actions you take are quite different, depending on whether the failure you just detected is a connectivity failure or a real site failure.

If you have only a single path, you do not know if it was the path or the remote site that went down. If you have no single points of failure and everything disappears, there is an extremely good chance that it was the site that went down. Any other mechanism to distinguish between a connectivity failure and a site failure (most likely human intervention) cannot react with the speed required to drive effective recovery actions.

2.10 Testing considerations

Testing your DR solution is a required and essential step in maintaining DR readiness. Many enterprises have business or regulatory requirements to conduct periodic tests to ensure the business is able to recover from a wide-scale disruption and recovery processes meet RTO and RPO requirements. The only way to determine the effectiveness of the solution and your enterprise's ability to recover from a disaster is through comprehensive testing.

One of the most important test considerations in developing a DR test plan is to make sure that the testing you conduct truly represents the way you would recover your data and enterprise. This way, when you actually need to recover following a disaster, you can recover the way you have been testing, thus improving the probability that you will be able to meet the RTO and RPO objectives established by your business.

⁶ For information about the impact of distance on response times when using channel extenders, contact your IBM representative to obtain the white paper titled *The effect of IU pacing on XRC FICON performance at distance*, which is available at this website:

<http://w3.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP100440>

Testing disk mirroring-based solutions

When conducting DR drills to test your recovery procedures, without additional disk capacity to support FlashCopy, the mirroring environment will be suspended so the secondary disks can be used to test your recovery and restart processes. When testing is completed, the mirror must be brought back to a duplex state again. During this window, until the mirror is back to a duplex state, the enterprises ability to recover from a disastrous event is compromised.

If this is not acceptable or your enterprise has a requirement to perform periodic disaster recovery tests while maintaining a disaster readiness posture, you will need to provide additional disk capacity to support FlashCopy. The additional FlashCopy device can be used for testing your recovery and restart procedures while the replication environment is running. This ensures that a current and consistent copy of the data is available, and that disaster readiness is maintained throughout the testing process.

The additional FlashCopy disk can also be used to create a copy of the secondary devices to ensure a consistent copy of the data is available if a disaster-type event occurs during primary and secondary volume resynchronization.

From a business perspective, installing the additional disk capacity to support FlashCopy will mean incurring additional expense. Not having it, however, can result in compromising the enterprise's ability to recover from a disastrous event, or in extended recovery times and exposure to additional data loss.

Testing software replication solutions

Similar in some instances to the situation described for testing disk-based mirroring solutions, if you test on your target copy of your database or databases, you will have to pause the replication process. Potentially, you might have to also re-create the target copy from scratch by using the source copy as input when the test is complete.

It would be normal to test the recovery procedures and operational characteristics of a software replication solution in a pre-production environment that as close as possible reflects the production environment.

However, because of the nature of software replication solutions, there is limited recovery required in the target site. Updates will either have been sent (and applied) from the source site, or they will not; the apply process is based on completed units of work, so there should be no issue with incomplete updates arriving from the source site. The testing is more likely to be related to the process for handling the potential data loss and any possible handling of collisions caused by the later capture/apply of stranded transactions with other completed units of work that might have occurred following an outage or disaster.

Testing methodology

How you approach your DR testing is also an important consideration. Most enterprises aim to do the majority of disruptive testing in a test or "sandbox" environment. This ideally will closely resemble the production environment so that the testing scenarios done in the sandbox are representative of what is applicable also in your production environment.

Other enterprises might decide to simulate a disaster in the production environment to really prove the processes and technology deliver what is required. Remember, however, that a disaster can surface to the technology in different ways (for example, different components failing in different sequences), so the scenarios you devise and test should consider these possible variations.

A typical approach to DR testing in production is to perform some form of a planned site switch. In such a test, the production service is closed down in a controlled manner where it normally runs, and then restarted in the DR site. This type of test will demonstrate that the infrastructure in the DR site is capable of running the services within the scope of the test, but given the brief duration of such tests (often over a weekend only) not all possible workload scenarios can be tested.

For this reason, consider the ability to move the production services to the DR site for an extended period (weeks or months), to give an even higher degree of confidence. This ability to “toggle” production and DR locations can provide other operational benefits, such as performing a preemptive switch because of an impending event, along with increased confidence being able to run following a DR invocation.

With this approach it is important to continue to test the actual DR process in your test environment, because a real disaster is unlikely to happen in a way where a controlled shutdown is possible. Those processes must then be carefully mapped across to the production environment to ensure success in a DR invocation.

In some industries, regulation might dictate or at least suggest guidelines about what constitutes a valid DR test, and this also needs to be considered.

2.11 Summary

In this chapter we covered the major building blocks of an IT resilience solution. We discussed providing continuous availability for normal operations, the options for keeping a consistent offsite copy of your disk and tape-based data, the need for automation to manage the recovery process, and the areas you need to consider when connecting across sites.

In the next few chapters, we discuss the functions provided by the various offerings in the GDPS family.



GDPS/PPRC

In this chapter, we discuss the capabilities and prerequisites of the GDPS/PPRC offering. GDPS/PPRC supports both planned and unplanned situations, helping to maximize application availability and provide business continuity. In particular, a GDPS/PPRC solution can deliver the following capabilities:

- ▶ Near-continuous availability solution
- ▶ Disaster recovery (DR) solution across metropolitan distances
- ▶ Recovery time objective (RTO) less than an hour
- ▶ Recovery point objective (RPO) of zero

The functions provided by GDPS/PPRC fall into two categories: protecting your data and controlling the resources managed by GDPS. Many of those provided functions are listed here:

- ▶ Protecting your data:
 - Ensuring the consistency of the secondary data if there is a disaster or suspected disaster, including the option to also ensure zero data loss
 - Transparent switching to the secondary disk using HyperSwap
 - Management of the remote copy configuration for z Systems and data that is not z Systems platform data
- ▶ Controlling the resources managed by GDPS during normal operations, planned changes, and following a disaster:
 - Monitoring and managing the state of the production z/OS systems and LPARs (shutdown, activating, deactivating, IPL, and automated recovery)
 - Monitoring and managing z/VM guests and native Linux z Systems LPARs (shutdown, activating, deactivating, IPL, and automated recovery)
 - Monitoring and managing distributed cluster resources (starting, stopping, and automated recovery supporting the movement of resources to another site)
 - Managing the couple data sets and coupling facility recovery
 - Support for switching your disk, or systems, or both, to another site
 - User-customizable scripts that control how GDPS/PPRC reacts to specified error situations, which can also be used for planned events

3.1 Introduction to GDPS/PPRC

GDPS/PPRC is a continuous availability and disaster recovery solution that handles many types of planned and unplanned outages. As mentioned in Chapter 1, “Introduction to business resilience and the role of GDPS” on page 1, most outages are planned, and even among unplanned outages, most are not disasters. GDPS/PPRC provides capabilities to help provide the required levels of availability across these outages and in a disaster scenario. These capabilities are described in this chapter.

3.1.1 Protecting data integrity and data availability with GDPS/PPRC

In 2.2, “Data consistency” on page 17, we point out that data integrity across primary and secondary volumes of data is essential to perform a database restart and accomplish an RTO of less than hour. This section provides details about how GDPS automation in GDPS/PPRC provides both data consistency if there are mirroring problems and data availability if there are disk problems.

Two types of disk problems trigger a GDPS automated reaction:

- ▶ PPRC Mirroring problems (Freeze triggers)

No problem exists writing to the primary disk subsystem, but a problem exists mirroring the data to the secondary disk subsystem. This is described in “GDPS Freeze function for mirroring failures” on page 54.

- ▶ Primary disk problems (HyperSwap triggers)

There is a problem writing to the primary disk: either a hard failure, or the disk subsystem is not accessible or not responsive. This is described in “GDPS HyperSwap function” on page 58.

GDPS Freeze function for mirroring failures

GDPS uses automation that is triggered by events or messages to stop all mirroring when a remote copy failure occurs. Specifically, the GDPS automation uses the IBM PPRC Freeze and Run architecture, which is implemented as part of Metro Mirror on IBM disk subsystems and also by other enterprise disk vendors. In this way, if the disk hardware supports the Freeze and Run architecture, GDPS can ensure consistency across all data in the sysplex (consistency group) regardless of disk hardware type. This preferred approach differs from proprietary hardware approaches that work only for one type of disk hardware. For a related introduction to data consistency with synchronous disk mirroring, see “PPRC data consistency” on page 24.

When a mirroring failure occurs, this problem is classified as a Freeze trigger and GDPS stops activity across *all* disk subsystems at the time the initial failure is detected, thus ensuring that the dependant write consistency of the remote disks is maintained. This is what happens when a GDPS performs a Freeze:

- ▶ Remote copy is suspended for all device pairs in the configuration.
- ▶ While the suspend command is being processed for each LSS, each device goes into a long busy state. When the suspend completes for each device, z/OS marks the device unit control block (UCB) in all connected operating systems to indicate an Extended Long Busy (ELB) state.

- ▶ No I/Os can be issued to the affected devices until the ELB is thawed with PPRC Run or until it times out (the consistency group timer setting commonly defaults to 120 seconds, although for most configurations a longer ELB is preferred).
- ▶ All paths between the PPRCed disks are removed, preventing further I/O to the secondary disks if PPRC is accidentally restarted.

Because no I/Os are processed for a remote-copied volume during the ELB, dependent write logic ensures the consistency of the remote disks. GDPS performs a Freeze for all LSS pairs that contain GDPS managed mirrored devices.

Important: Because of the dependent write logic, it is not necessary for all LSSs to be frozen at the same instant. In a large configuration with many thousands of remote copy pairs, it is not unusual to see short gaps between the times when the Freeze command is issued to each disk subsystem. Because of the ELB, however, such gaps are not a problem.

After GDPS performs the Freeze and the consistency of the remote disks is protected, what GDPS does depends on the client's PPRCFAILURE policy (also known as Freeze policy). The policy as described in "Freeze policy (PPRCFAILURE policy) options" on page 56 will tell GDPS to take one of these three possible actions:

- ▶ Perform a Run action against all LSSs. This will remove the ELB and allow production systems to continue using these devices. The devices will be in remote copy-suspended mode, meaning that any further writes to these devices are no longer being mirrored. However, changes are being tracked by the hardware so that, later, only the changed data will be resynchronized to the secondary disks. See "Freeze and Go" on page 57 for more detail on this policy option.
- ▶ System-reset all production systems. This ensures that no more updates can occur to the primary disks because such updates would not be mirrored, meaning that it would not be possible to achieve an RPO of zero (zero data loss) if a failure occurs (or if the original trigger was an indication of a catastrophic failure). See "Freeze and Stop" on page 56 for more detail about this option.
- ▶ Try to determine if the cause of the PPRC suspension event was a permanent or temporary problem with any of the secondary disk subsystems in the GDPS configuration. If GDPS can determine that the PPRC failure was caused by the secondary disk subsystem, this would not be a potential indicator of a disaster in the primary site. In this case, GDPS would perform a Run action and allow production to continue using the suspended primary devices. If, however, the cause cannot be determined to be a secondary disk problem, GDPS would reset all systems, guaranteeing zero data loss. See "Freeze and Stop conditionally" on page 57 for further details.

GDPS/PPRC uses a combination of storage subsystem and sysplex triggers to automatically secure, at the first indication of a potential disaster, a data-consistent secondary site copy of your data using the Freeze function. In this way, the secondary copy of the data is preserved in a consistent state, perhaps even before production applications are aware of any issues. Ensuring the data consistency of the secondary copy ensures that a normal system restart can be performed instead of having to perform DBMS forward recovery actions. This is an essential design element of GDPS to minimize the time to recover the critical workloads if there is a disaster in the primary site.

You can appreciate why such a process must be automated. When a device suspends, there is not enough time to launch a manual investigation process. The entire mirror must be frozen by stopping further I/O to it, and then the policy indicates whether production will continue to run with mirroring temporarily suspended, or whether all systems should be stopped to guarantee zero data loss.

In summary, freeze is triggered as a result of a PPRC suspension event for any primary disk in the GDPS configuration, that is, at the first sign that a duplex mirror that is going out of the duplex state. When a device suspends, all attached systems are sent a “State Change Interrupt” (SCI). A message is issued in all of those systems and then each system must issue multiple I/Os to investigate the reason for the suspension event.

When GDPS performs a freeze, all primary devices in the PPRC configuration suspend. This can result in significant SCI traffic and many messages in all of the systems. GDPS, in z/OS and microcode on the DS8000 disk subsystems, supports reporting suspensions in a summary message per LSS instead of at the individual device level. When compared to reporting suspensions on a per devices basis, the Summary Event Notification for PPRC Suspend (PPRCSUM) dramatically reduces the message traffic and extraneous processing associated with PPRC suspension events and freeze processing.

For more information about the implementation of PPRC and IBM Metro Mirror, see *IBM DS8870 Copy Services for IBM z Systems*, SG24-6787.

Freeze policy (PPRCFAILURE policy) options

As we have described, when a mirroring failure is detected, GDPS automatically and unconditionally performs a Freeze to secure a consistent set of secondary volumes in case the mirroring failure could be the first indication of a site failure. Because the primary disks are in the Extended Long Busy state as a result of the freeze and the production systems are locked out, GDPS must take some action. Here, there is no time to interact with the operator on an event-by-event basis. The action must be taken immediately. The action to be taken is determined by a customer policy setting, that is, the PPRCFailure policy option (also known as the Freeze policy option). GDPS will use this same policy setting after every Freeze event to determine what its next action should be. The options are as follows:

- ▶ PPRCFailure=Stop (Freeze and Stop)
GDPS resets production systems while I/O is suspended.
- ▶ PPRCFailure=Go (Freeze and Go)
GDPS allows production systems to continue operation after mirroring is suspended.
- ▶ PPRCFailure=COND (Freeze and Stop, conditionally)
GDPS tries to determine if a secondary disk caused the mirroring failure. If so, GDPS performs a Go. If not, GDPS performs a Stop.

Freeze and Stop

If your RPO is zero (that is, you cannot tolerate any data loss), you must select the Freeze and Stop policy to reset all production systems. With this setting, you can be assured that no updates are made to the primary volumes after the Freeze because all systems that can update the primary volumes are reset. You can choose to restart them when you want. For example, if this was a false freeze (that is, a false alarm), then you can quickly resynchronize the mirror and restart the systems only after the mirror is duplex.

If you are using duplexed coupling facility (CF) structures along with a Freeze and Stop policy, it might seem that you are guaranteed to use the duplexed instance of your structures if you must recover and restart your workload with the frozen secondary copy of your disks. However, this is not always the case. There can be rolling disaster scenarios where before,

after, or during the freeze event, there is an interruption (perhaps failure of CF duplexing links) that forces CFRM to drop out of duplexing. There is no guarantee that it is the structure instance in the surviving site that is kept.

It is possible that CFRM keeps the instance in the site that is about to totally fail. In this case, there will not be an instance of the structure in the site that survives the failure.

To summarize, with a Freeze and Stop policy, if there is a surviving, accessible instance of application-related CF structures, this instance will be consistent with the frozen secondary disks. However, depending on the circumstances of the failure, even with structures duplexed across two sites you are not 100% guaranteed to have a surviving, accessible instance of the application structures and therefore you must have the procedures in place to restart your workloads without the structures.

A Stop policy ensures no data loss. However, if this was a false freeze event, that is, it was a transient failure that did not necessitate recovering using the frozen disks, then it results in unnecessarily stopping the systems.

Freeze and Go

If you can accept an RPO that is not *necessarily* zero, you might decide that the production systems can *continue* operation after the secondary volumes have been protected by the Freeze. In this case, you would use a Freeze and Go policy. With this policy you avoid an unnecessary outage for a false freeze event, that is, if the trigger is simply a transient event.

However, if the trigger turns out to be the first sign of an actual disaster, you might continue operating for an amount of time before all systems fail. Any updates made to the primary volumes during this time are not replicated to the secondary disk, and therefore are lost. In addition, because the CF structures were updated after the secondary disks were frozen, the CF structure content is not consistent with the secondary disks. Therefore, the CF structures in either site cannot be used to restart workloads and log-based restart must be used when restarting applications.

Note this is not full forward recovery. It is forward recovery of any data such as DB2 Group buffer pools that might have existed in a CF, but might not have been written to disk yet. This results in prolonged recovery times. The duration of this elongation depends on how much such data existed in the CFs at that time. With a Freeze and Go policy, you might consider tuning to applications such as DB2, which can harden such data on disk more frequently than otherwise.

Freeze and Go is a high availability option that avoids production outage for false freeze events. However, it carries a potential for data loss.

Freeze and Stop conditionally

Field experience has shown that most of the Freeze triggers are not necessarily the start of a rolling disaster, but are “False Freeze” events that do not necessitate recovery on the secondary disk. Examples of such events include connectivity problems to the secondary disks and secondary disk subsystem failure conditions.

With a COND specification, the action that GDPS takes after it performs the Freeze is conditional. GDPS tries to determine if the mirroring problem was as a result of a permanent or temporary secondary disk subsystem problem:

- If GDPS can determine that the freeze was triggered as a result of a secondary disk subsystem problem, then GDPS performs a Go. That is, it allows production systems to continue to run using the primary disks. However, updates will not be mirrored until the secondary disk can be fixed and PPRC can be resynchronized.

- If GDPS cannot ascertain that the cause of the freeze was a secondary disk subsystem, then GDPS deduces that this could still be the beginning of a rolling disaster in the primary site and it performs a Stop, resetting all the production systems to guarantee zero data loss. GDPS cannot always detect that a particular freeze trigger was caused by a secondary disk, and that some freeze events that are in fact caused by a secondary disk could still result in a Stop.

For GDPS to determine whether a freeze trigger might have been caused by the secondary disk subsystem, the IBM DS8000 disk subsystems provide a special query capability known as the *Query Storage Controller Status* microcode function. If all disk subsystems in the GDPS managed configuration support this feature, GDPS uses this special function to query the secondary disk subsystems in the configuration to understand the state of the secondaries and if one of these secondaries might have caused the freeze. If you use the COND policy setting but all disks your configuration do not support this function, then GDPS cannot query the secondary disk subsystems and the resulting action is a Stop.

This option can provide a good compromise where you can minimize the chance that systems would be stopped for a false freeze event and increase the chance of achieving zero data loss for a real disaster event.

PPRCFAILURE policy selection considerations

As explained, the PPRCFailure policy option specification directly relates to Recovery Time and recovery point objectives, which are business objectives. Therefore, the policy option selection is really a business decision, rather than an IT decision. If data associated with your transactions is high value, it might be more important to ensure that no data associated with your transactions is ever lost, so you might decide on a Freeze and Stop policy. If you have huge volumes of relatively low value transactions, you might be willing to risk some lost data in return for avoiding unneeded outages with a Freeze and Go policy. The Freeze and Stop CONDitional policy attempts to minimize the chance of unnecessary outages and the chance of data loss; however, there is still a risk of either, however small.

Most installations start out with a Freeze and Go policy. Companies that have an RPO of zero will typically then move on and implement a Freeze and Stop Conditional or Freeze and Stop policy after the implementation is proven stable.

GDPS HyperSwap function

If there is a problem writing or accessing the *primary* disk because of a failing, failed, or non-responsive primary disk, then there is a need to swap from the primary disks to the secondary disks.

GDPS/PPRC delivers a powerful function known as HyperSwap. HyperSwap provides the ability to swap from using the primary devices in a mirrored configuration to using what had been the secondary devices, not apparent to the production systems and applications using these devices. Before the availability of HyperSwap, a transparent disk swap was not possible. All systems using the primary disk would have been shut down (or might have failed, depending on the nature and scope of the failure) and would have been re-IPLed using the secondary disks. Disk failures were often a single point of failure for the entire sysplex.

With HyperSwap, such a switch can be accomplished without IPL and with just a brief hold on application I/O. The HyperSwap function is completely controlled by automation, thus allowing all aspects of the disk configuration switch to be controlled through GDPS.

HyperSwap can be invoked in two ways:

► **Planned HyperSwap**

A planned HyperSwap is invoked by operator action using GDPS facilities. One example of a planned HyperSwap is where a HyperSwap is initiated in advance of planned disruptive maintenance to a disk subsystem.

► **Unplanned HyperSwap**

An unplanned HyperSwap is invoked automatically by GDPS, triggered by events that indicate the primary disk problem.

Primary disk problems can be detected as a direct result of an I/O operation to a specific device that fails because of a reason that indicates a primary disk problem such as:

- No paths available to the device
- Permanent error
- I/O timeout

In addition to a disk problem being detected as a result of an I/O operation, it is also possible for a primary disk subsystem to proactively report that it is experiencing an acute problem. The IBM DS8000 models have a special microcode function known as the *Storage Controller Health Message Alert* capability. Problems of different severity are reported by disk subsystems that support this capability. Those problems classified as acute are also treated as HyperSwap triggers. After systems are swapped to use the secondary disks, the disk subsystem and operating system can try to perform recovery actions on the former primary without impacting the applications using those disks.

Planned and unplanned HyperSwap have requirements in terms of the physical configuration, such as having it be symmetrically configured, and so on. While a client's environment meets these requirements, there is no special enablement required to perform planned swaps. Unplanned swaps are not enabled by default and must be enabled explicitly as a policy option. This is described in further detail in "HyperSwap policy (Primary Failure policy) options" on page 61.

When a swap is initiated, GDPS always validates various conditions to ensure that it is safe to swap. For example, if the mirror is not fully duplex, that is, not all volume pairs are in a duplex state, a swap cannot be performed. The way that GDPS reacts to such conditions changes depending on the condition detected and whether the swap is a planned or unplanned swap.

Assuming that there are no show-stoppers and the swap proceeds, for both planned and unplanned HyperSwap, the systems that are using the primary volumes will experience a temporary pause in I/O processing. GDPS blocks I/O both at the channel subsystem level by performing a Freeze which results in all disks going into Extended Long Busy, and also in all systems, where I/O is quiesced at the operating system (UCB) level. This is to ensure that no systems use the disks until the switch is complete. During the time when I/O is paused, the following steps happen:

1. The PPRC configuration is *physically switched*. This includes physically changing the secondary disk status to primary. Secondary disks are protected and cannot be used by applications. Changing their status to primary allows them to come online to systems and be used.
2. The disks will be *logically switched* in each of the systems in the GDPS configuration. This involves switching the internal pointers in the operating system control blocks (UCBs). After the switch, the operating system will point to the former secondary devices instead of the current primary devices.
3. For planned swaps, optionally, the mirroring direction can be reversed.

4. The systems resume operation using the new, swapped-to primary devices. The applications are not aware of the fact that different devices are now being used.

This brief pause during which systems are locked out of performing I/O is known as the *User Impact Time*. In benchmark measurements at IBM using currently supported releases of GDPS and IBM DS8000 disk subsystems, the User Impact Time to swap 10,000 pairs across 16 systems during an unplanned HyperSwap was less than 10 seconds. Most implementations are actually much smaller than this and typical impact times using the most current storage and server hardware are measured in seconds. Although results will depend on your configuration, these numbers give you a high-level idea of what to expect.

GDPS/PPRC HyperSwaps all devices in the managed configuration. Just as the Freeze function applies to the entire consistency group, similarly HyperSwap is for the entire consistency group. For example, if a single mirrored volume fails and HyperSwap is invoked, processing is swapped to the secondary copy of *all* mirrored volumes in the configuration, including those in other, unaffected, subsystems. This is because, to maintain disaster readiness, all primary volumes *must* be in the same site. If HyperSwap were to swap the only failed LSS, you would then have several primaries in one site, and the remainder in the other site. This would also make for a significantly complex environment to operate and administer I/O configurations.

Why is this necessary? Consider the configuration shown in Figure 3-1 on page 61. This is what might happen if only the volumes of a single LSS or subsystem were hyper-swapped without swapping the whole consistency group. What happens if there is a remote copy failure at 15:00? The secondary disks in both sites are frozen at 15:00 and the primary disks (in the case of a Freeze and Go policy) continue to receive updates.

Now assume that either site is hit by another failure at 15:10. What do you have? Half the disks are now at 15:00 and the other half are at 15:10 and *neither* site has consistent data. In other words, the volumes are of virtually no value to you.

If you had *all* the secondaries in Site2, all volumes in that site would be consistent. If you had the disaster at 15:10, you would lose 10 minutes of data with the Go policy, but at least all the data in Site2 would be usable. Using a Freeze and Stop policy is no better for this partial swap scenario because with a mix of primary disks in either site, you have to maintain I/O configurations that can match *every* possible combination simply to IPL any systems.

More likely, you first have to restore mirroring across the entire consistency group before recovering systems and this is not really practical. Therefore, for disaster recovery readiness, it is necessary that *all* the primary volumes are in one site, and *all* the secondaries in the other site.

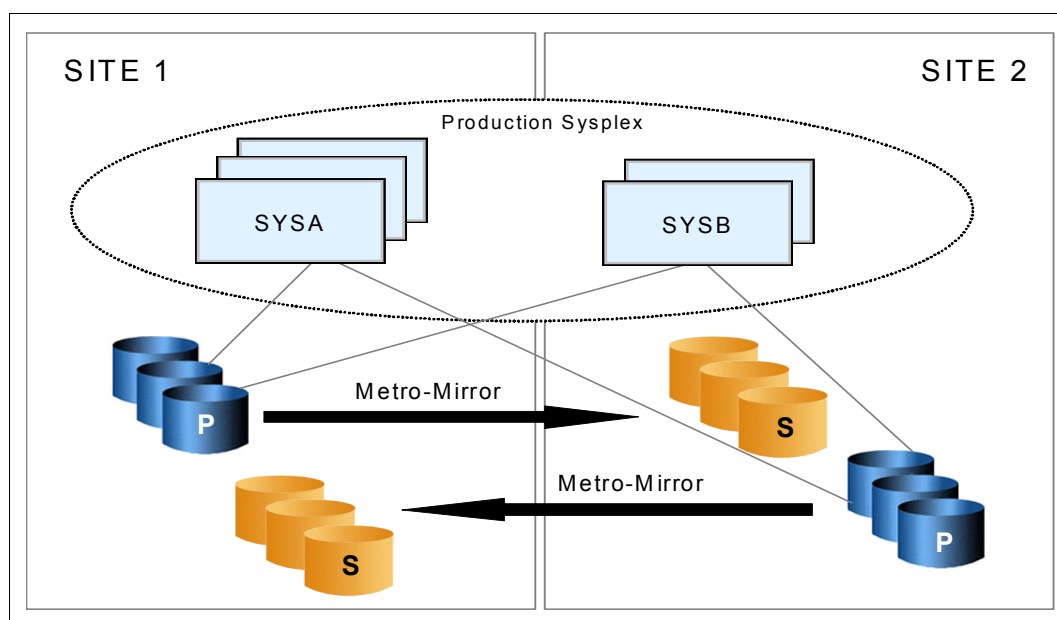


Figure 3-1 Unworkable Metro Mirror disk configuration

HyperSwap with less than full channel bandwidth

You may consider enabling unplanned HyperSwap even if you do not have sufficient cross-site channel bandwidth to sustain the full production workload for normal operations. Assuming that a disk failure is likely to cause an outage and you will need to switch to using disk in the other site, the unplanned HyperSwap might at least present you with the opportunity to perform an orderly shutdown of your systems first. Shutting down your systems cleanly avoids the complications and restart time elongation associated with a crash-restart of application subsystems.

HyperSwap policy (Primary Failure policy) options

Clients might prefer not to immediately enable their environment for unplanned HyperSwap when they first implement GDPS. For this reason, HyperSwap is not enabled by default. However, we strongly suggest that all GDPS/PPRC clients enable their environment for unplanned HyperSwap.

An unplanned swap is the action that makes most sense when a primary disk problem is encountered. However, other policy specifications that will not result in a swap are available. When GDPS detects a primary disk problem trigger, the first thing it will do will be a Freeze (the same as is performed when a mirroring problem trigger is encountered). GDPS then uses the selected Primary Failure policy option to determine what action it will take next:

► PRIMARYFAILURE=GO

No swap is performed. The action GDPS takes is the same as for a freeze event with policy option PPRCFAILURE=GO. A Run action is performed, which allows systems to continue using the original primary disks. PPRC is suspended and therefore updates are not being replicated to the secondary.

However, depending on the scope of the primary disk problem, it might be that some or all production workloads simply cannot run or cannot sustain required service levels. Such a situation might necessitate restarting the systems on the secondary disks. Because of the freeze, the secondary disks are in a consistent state and can be used for restart. However, any transactions that ran after the Go action will be lost.

► PRIMARYFAILURE=Stop

No swap is performed. The action GDPS takes is the same as for a freeze event with policy option PPRCFAILURE=STOP. GDPS system-resets all the production systems. This ensures that no further I/O occurs. After performing situation analysis, if it is determined that this was not a transient issue and that the secondaries should be used to IPL the systems again, no data will be lost.

► PRIMARYFAILURE=SWAP,swap_disabled_action

The first parameter, SWAP, indicates that after performing the Freeze, GDPS will proceed with performing an unplanned HyperSwap. When the swap is complete, the systems will be running on the new, swapped-to primary disks (former secondaries). PPRC will be in a suspended state; because the primary disks are known to be in a problematic state, there is no attempt to reverse mirroring. After the problem with the primary disks is fixed, you can instruct GDPS to resynchronize PPRC from the current primaries to the former ones (which are now considered to be secondaries).

The second part of this policy, swap_disabled_action, indicates what GDPS should do if HyperSwap had been temporarily disabled by operator action at the time the trigger was encountered. Effectively, an operator action has instructed GDPS not to perform a HyperSwap, even if there is a swap trigger. GDPS has already performed a freeze. The second part of the policy control what action GDPS will take next.

The following options (which are in effect only if HyperSwap is disabled by the operator) are available for the second parameter (remember that the disk is already frozen):

GO This is the same action as GDPS would have performed if the policy option had been specified as PRIMARYFAILURE=GO.

STOP This is the same action as GDPS would have performed if the policy option had been specified as PRIMARYFAILURE=STOP.

Primary Failure policy specification considerations

As indicated previously, the action that best serves RTO/RPO objectives when there is a primary disk problem is to perform an unplanned HyperSwap. Therefore, the SWAP policy option is the recommended policy option.

For the Stop or Go choice, either as the second part of the SWAP specification or if you will not be using SWAP, similar considerations apply as discussed for the PPRCFAILURE policy options to Stop or Go. Go carries the risk of data loss if it should be necessary to abandon the primary disk and restart systems on the secondary. Stop carries the risk of taking an unnecessary outage if the problem was transient. The key difference is that with a mirroring failure, the primary disks are not broken. When you allow the systems to continue to run on the primary disk with the Go option, other than a disaster which is low probability, the systems are likely to run with no problems. With a primary disk problem, with the Go option, you are allowing the systems to continue running on what are known to be disks that experienced a problem just seconds ago. If this was a serious problem with widespread impact such as an entire disk subsystem failure, the applications are going to experience severe problems. Some transactions might continue to commit data to those disks that are not broken. Other transactions might be failing or experiencing serious service time issues. Finally, if there is a decision to restart systems on the secondary because the primary disks are simply not able to support the workloads, there will be data loss. The probability that a primary disk problem is a real problem that will necessitate restart on the secondary disks is much higher when

compared to a mirroring problem. A Go specification in the Primary Failure policy increases your overall risk for data loss.

If the primary failure was of a transient nature, a Stop specification results in an unnecessary outage. However, with primary disk problems the probability that the problem could necessitate restart on the secondary disks is high, so a Stop specification in the Primary Failure policy avoids data loss and facilitates faster restart.

The considerations relating to CF structures with a PRIMARYFAILURE event are similar to a PPRCFAILURE event. If there is an actual swap, the systems continue to run and continue to use the same structures as they did before the swap; the swap is transparent. With a Go action, because you continue to update the CF structures along with the primary disks after the Go action, if you need to abandon the primary disks and restart on the secondary, the structures are inconsistent with the secondary disks and are not usable for restart purposes. This will prolong the restart, and therefore your recovery time. With Stop, if you decide to restart the systems using the secondary disks, there is no consistency issue with the CF structures because no further updates occurred on either set of disks after the trigger was captured.

GDPS use of DS8000 functions

GDPS strives to use (when it makes sense) enhancements to the IBM DS8000 disk technologies. In this section we provide information about the key DS8000 technologies that GDPS supports and uses.

PPRC Failover/Failback support

When a primary disk failure occurs and the disks are switched to the secondary devices, PPRC Failover/Failback (FO/FB) support eliminates the need to do a full copy when reestablishing replication in the opposite direction. Because the primary and secondary volumes are often in the same state when the freeze occurred, the only differences between the volumes are the updates that occur to the secondary devices after the switch. Failover processing sets the secondary devices to primary suspended status and starts change recording for any subsequent changes made. When the mirror is reestablished with failback processing, the original primary devices become secondary devices and a resynchronization of changed tracks takes place.

GDPS/PPRC requires PPRC FO/FB capability to be available on all disk subsystems in the managed configuration.

PPRC eXtended Distance (PPRC-XD)

PPRC-XD (also known as Global Copy) is an asynchronous form of the PPRC copy technology. GDPS uses PPRC-XD rather than synchronous PPRC to reduce the performance impact of certain remote copy operations that potentially involve a large amount of data. See 3.7.2, “Reduced impact initial copy and resynchronization” on page 94 for details.

Storage Controller Health Message Alert

This facilitates triggering an unplanned HyperSwap proactively when the disk subsystem reports an acute problem that requires extended recovery time. See “GDPS HyperSwap function” on page 58 for information about unplanned HyperSwap triggers.

PPRCS Summary Event Messages

GDPS supports the DS8000 PPRC Summary Event Messages (PPRCSUM) function which is aimed at reducing the message traffic and the processing of these messages for Freeze events. This is described in “GDPS Freeze function for mirroring failures” on page 54.

Soft Fence

Soft Fence provides the capability to block access to selected devices. As discussed in “Protecting secondary disks from accidental update” on page 65, GDPS uses Soft Fence to avoid write activity on disks that are exposed to accidental update in certain scenarios.

On-demand dump (also known as non-disruptive statesave)

When problems occur with disk subsystems such as those which result in an unplanned HyperSwap, a mirroring suspension or performance issues, a lack of diagnostic data from the time the event occurs can result in difficulties in identifying the root cause of the problem. Taking a full statesave can lead to temporary disruption to host I/O and is often frowned upon by clients for this reason. The on-demand dump (ODD) capability of the disk subsystem facilitates taking a non-disruptive statesave (NDSS) at the time such an event occurs. The microcode does this automatically for certain events, such as taking a dump of the primary disk subsystem that triggers a PPRC freeze event, and also allows an NDSS to be requested. This enables first failure data capture (FFDC) and thus ensures that diagnostic data is available to aid problem determination. Be aware that not all information that is contained in a full statesave is contained in an NDSS and therefore there may still be failure situations where a full statesave is requested by the support organization.

GDPS provides support for taking an NDSS by using the remote copy panels (or GDPS GUI). In addition to this support, GDPS autonomically takes an NDSS if there is an unplanned Freeze or HyperSwap event.

Query Host Access

When a PPRC disk pair is being established, the device that is the target (secondary) must not be used by any system. The same is true when establishing a FlashCopy relationship to a target device. If the target is in use, the establishment of the PPRC or FlashCopy relationship fails. When such failures occur, it can be a tedious task to identify which system is holding up the operation. The Query Host Access disk function provides the means to query and identify what system is using a selected device. GDPS uses this capability and adds usability in several ways:

- ▶ Query Host Access identifies the LPAR that is using the selected device through the CPC serial number and LPAR number. It is still a tedious job for operations staff to translate this information to a system or CPC and LPAR name. GDPS does this translation and presents the operator with more readily usable information, avoiding this additional translation effort.
- ▶ Whenever GDPS is requested to perform a PPRC or FlashCopy establish operation, GDPS first performs Query Host Access to see if the operation is expected to succeed or fail as a result of one or more target devices being in use. GDPS alerts the operator if the operation is expected to fail, and identifies the target devices in use and the LPARs holding them.
- ▶ GDPS continually monitors the target devices defined in the GDPS configuration and alerts operations to the fact that target devices are in use when they should not be. This allows operations to fix the reported problems in a timely manner.
- ▶ GDPS provides the ability for the operator to perform ad hoc Query Host Access to any selected device using the GDPS panels (or GUI).

Easy Tier Heat Map Transfer

IBM DS8000 Easy Tier® optimizes data placement (placement of logical volumes) across the various physical tiers of storage within a disk subsystem in order to optimize application performance. The placement decisions are based on learning the data access patterns and can be changed dynamically and transparently to the applications using this data.

PPRC mirrors the data from the primary to the secondary disk subsystem however, the Easy Tier learning information is not included in PPRC scope. The secondary disk subsystems are optimized according to the workload on these subsystems which is different than the activity on the primary (there is only write workload on the secondary whereas there is read/write activity on the primary). As a result of this difference, during a disk switch or disk recovery, the secondary disks that you switch to are likely to display different performance characteristics compared to the former primary.

Easy Tier Heat Map Transfer is the DS8000 capability to transfer the Easy Tier learning from a PPRC primary to the secondary disk subsystem so that the secondary disk subsystem can also be optimized, based on this learning, and will have similar performance characteristics if it is promoted to become the primary.

GDPS integrates support for Heat Map Transfer. The appropriate Heat Map Transfer actions (such as start/stop of the processing and reversing transfer direction) are incorporated into the GDPS managed processes. For example, if PPRC is temporarily suspended by GDPS for a planned or unplanned secondary disk outage, Heat Map Transfer is also suspended, or if PPRC direction is reversed as a result of a HyperSwap, Heat Map Transfer direction is also reversed.

Protecting secondary disks from accidental update

A system cannot be IPLed using a disk that is physically a PPRC secondary disk because PPRC secondary disks cannot be brought online to any systems. However, a disk can be secondary from a GDPS (and application use) perspective but physically have simplex or primary status from a PPRC perspective.

For both planned and unplanned HyperSwap, and a disk recovery, GDPS changes former secondary disks to primary or simplex state. However, these actions do not modify the state of the former primary devices, which remain in the primary state. Therefore, the former primary devices remain accessible and usable even though they are considered to be the secondary disks from a GDPS perspective. This makes it is possible to accidentally update or IPL from the wrong set of disks. Accidentally using the wrong set of disks can result in a potential data integrity or data loss problem.

GDPS/PPRC provides protection against using the wrong set of disks in different ways:

- ▶ If you attempt to load a system through GDPS (either script or panel or GUI) using the wrong set of disks, GDPS rejects the load operation.
- ▶ If you used the HMC rather than GDPS facilities for the load, then early in the IPL process, during initialization of GDPS, if GDPS detects that the system coming up has just been IPLed using the wrong set of disks, GDPS will quiesce that system, preventing any data integrity problems that could be experienced had the applications been started.
- ▶ GDPS uses a DS8000 disk subsystem capability, which is called Soft Fence for configurations where the disks support this function. Soft Fence provides the means to *fence*, which means block access to a selected device. GDPS uses Soft Fence when appropriate to fence devices that would otherwise be exposed to accidental update.

3.1.2 Protecting tape data

Although most of your critical data will be resident on disk, it is possible that other data you require following a disaster resides on tape. Just as you mirror your disk-resident data to protect it, equally you can mirror your tape-resident data. GDPS/PPRC provides support for management of the IBM Virtualization Engine TS7700. GDPS provides TS7700 configuration management and displays the status of the managed TS7700s on GDPS panels.

TS7700s that are managed by GDPS are monitored and alerts are generated for non-normal conditions. The capability to control TS7700 replication from GDPS scripts and panels using TAPE ENABLE and TAPE DISABLE by library, grid, or site is provided for managing TS7700 during planned and unplanned outage scenarios.

Another important aspect with replicated tape is identification of “in-doubt” tapes. Tape replication is not exactly like disk replication in that the replication is not done every time a record is written to the tape. The replication is typically performed at tape unload rewind time or perhaps even later. This means that if there is an unplanned event or interruption to the replication, some volumes could be back-level in one or more libraries in the grid. If you have to perform a recovery operation in one site because the other site has failed, it is important to identify if any of the tapes in the library in the site where you are recovering are back-level. Depending on the situation with any in-doubt tapes in the library or libraries you will use in the recovery site, you might need to perform special recovery actions. For example, you might need to rerun one or more batch jobs before resuming batch operations.

GDPS provides support for identifying in-doubt tapes in a TS7700 library. The TS7700 provides a capability called Bulk Volume Information Retrieval (BVIR). Using this BVIR capability, GDPS, if there is an unplanned interruption to tape replication, automatically collects information about all volumes in all libraries in the grid where the replication problem occurred. GDPS can then use this information to report on in-doubt volumes in any given library in that grid if the user requests a report. In addition to this automatic collection of in-doubt tape information, it is possible to request GDPS to perform BVIR processing for a selected library using the GDPS panel interface at any time.

The IBM Virtualization Engine TS7700 provides comprehensive support for replication of tape data. See *IBM Virtualization Engine TS7700 with R 2.0*, SG24-7975 for more information about the TS7700 technology that complements GDPS for tape data.

3.1.3 Protecting distributed (FB) data

Terminology: The introduction of Open LUN support in GDPS has caused several changes in the terminology we use when referring to disks in this book, as explained here.

- z Systems or CKD disks

GDPS can manage disks that are used by z Systems, although disks could be z/VM, VSE, or Linux on z Systems disks. All these disks are formatted as Count-Key-Data (CKD) disks, the traditional mainframe format.

In most places, we refer to the disks used by a system running on the mainframe as “z Systems disks,” although there are a small number of cases where the term “CKD disks” is also used; both terms are used interchangeably.

- Open LUN or FB disks

Disks that are used by systems other than those running on z Systems are traditionally formatted as Fixed Block (FB). In this book, we generally use the term “Open LUN disks” or “FB disks” interchangeably to refer to such devices.

GDPS/PPRC provides support for SCSI-attached FB disk used by native Linux for z Systems under GDPS xDR^a control. There is a need to differentiate between FB disk used by other distributed systems (including Linux on z Systems not under xDR control) and the FB disk used by native Linux on z Systems xDR.

a. For more information, see 3.3.1, “Multiplatform Resiliency for z Systems (also known as xDR)” on page 75.

GDPS/PPRC can manage the mirroring of FB devices used by non-mainframe operating systems; this also includes SCSI disks written by Linux on z Systems. The FB devices can be part of the same consistency group as the mainframe CKD devices, or they can be managed separately in their own consistency group. CKD and xDR-managed FB disks are always in the same consistency group: they are always frozen and swapped together.

For more information about Open LUN management, see 10.1, “Open LUN Management function” on page 296.

3.1.4 Protecting other CKD data

Systems that are fully managed by GDPS are known as *GDPS managed systems* or *GDPS systems*. These are as follows:

- ▶ z/OS systems in the GDPS sysplex
- ▶ z/VM systems managed by GDPS/PPRC MultiPlatform Resiliency for z Systems (xDR)
- ▶ Linux on z Systems running natively in an LPAR managed by GDPS/PPRC MultiPlatform Resiliency for z Systems (xDR)
- ▶ z/OS systems outside of the GDPS sysplex that are managed by the GDPS/PPRC z/OS Proxy (the z/OS Proxy)

We describe the z/OS Proxy in 3.4, “Management of z/OS systems outside of the GDPS sysplex” on page 76. MultiPlatform Resiliency for z Systems is described in 10.2, “GDPS/PPRC Multiplatform Resiliency for z Systems” on page 299.

In this section, we describe GDPS/PPRC management for the disk mirroring of CKD disks used by systems outside of the sysplex that are *not using* the z/OS Proxy or MultiPlatform Resiliency for z Systems. These systems are other z/OS systems, Linux on z Systems (running in a native LPAR or as a guest under z/VM or as a guest under KVM on z Systems), z/VM systems, and VSE systems that are not running any GDPS/PPRC function. These systems are known as “foreign systems.”

Because GDPS manages PPRC for the disks used by these systems, these disks will be attached to the GDPS controlling systems. With this setup, GDPS is able to capture mirroring problems and will perform a freeze. All GDPS managed disks belonging to the GDPS systems and these foreign systems are frozen together, regardless of whether the mirroring problem is encountered on the GDPS systems’ disks or the foreign systems’ disks.

GDPS/PPRC is not able to directly communicate with these foreign systems. For this reason, GDPS automation will not be aware of certain other conditions such as a primary disk problem that is detected by these systems. Because GDPS will not be aware of such conditions that would have otherwise driven autonomic actions such as HyperSwap, GDPS will not react to these events.

If an unplanned HyperSwap occurs (because it was triggered on a GDPS managed system), the foreign systems cannot and will not swap to using the secondaries. A setup is prescribed to set a long Extended Long Busy timeout (the maximum is 18 hours) for these systems so that when the GDPS managed systems swap, these systems hang. The ELB prevents these systems from continuing to use the former primary devices. You can then use GDPS automation facilities to reset these systems and re-IPL them using the swapped-to primary disks.

3.2 GDPS/PPRC configurations

At its most basic, a GDPS/PPRC configuration consists of at least one production system, at least one controlling system in a sysplex, primary disks, and secondary disks. The actual configuration depends on your business and availability requirements. The following three configurations are most common:

- ▶ Single-site workload configuration

In this configuration, all the production systems normally run in the same site, referred to as Site1, and the GDPS controlling system runs in Site2. In effect, Site1 is the active site for all production systems; the controlling system in Site2 is running and resources are available to move production to Site2, if necessary, for a planned or unplanned outage of Site1. Although you might also hear this referred to as an *Active/Standby GDPS/PPRC configuration*, we avoid the *Active/Standby* term to avoid confusion with the same term used in conjunction with the GDPS/Active-Active product.

- ▶ Multisite workload configuration

In this configuration, the production systems run in both sites, Site1 and Site2. This configuration typically uses the full benefits of data sharing available with a Parallel Sysplex. Having two GDPS controlling systems, one in each site, is preferable. Although you might also hear this referred to as an *Active/Active GDPS/PPRC configuration*, we avoid the *Active/Active* term to avoid confusion with the same term used in conjunction with the GDPS/Active-Active product.

- ▶ Business Recovery Services (BRS) configuration

In this configuration, the production systems and the controlling system are all in the same site, referred to as Site1. Site2 can be a client site or can be owned by a third-party recovery services provider (thus the name BRS). You might hear this referred to as an *Active/Cold* configuration.

These configuration options are described in further detail in the following sections.

3.2.1 Controlling system

Why does a GDPS/PPRC configuration need a controlling system? At first, you might think this is an additional infrastructure overhead. However, when you have an unplanned outage that affects production systems or the disk subsystems, it is crucial to have a system such as the controlling system that can survive failures that might have impacted other portions of your infrastructure. The controlling system allows you to perform situation analysis after the unplanned event to determine the status of the production systems or the disks, and then to drive automated recovery actions. The controlling system plays a vital role in a GDPS/PPRC configuration.

The controlling system must be in the same sysplex as the production system (or systems) so it can see all the messages from those systems and communicate with those systems. However, it shares an absolute minimum number of resources with the production systems (typically just the sysplex couple data sets). By being configured to be as self-contained as possible, the controlling system is unaffected by errors that can stop the production systems (for example, an Extended Long Busy event on a primary volume).

The controlling system must have connectivity to all the Site1 and Site2 primary and secondary devices that it will manage. If available, it is preferable to isolate the controlling system infrastructure on a disk subsystem that is not housing mirrored disks that are managed by GDPS.

The controlling system is responsible for carrying out all recovery actions following a disaster or potential disaster; for managing the disk mirroring configuration; for initiating a HyperSwap; for initiating a freeze and implementing the freeze/swap policy actions; reassigning STP roles; re-IPLing failed systems, and so on.

The availability of the dedicated GDPS controlling system (or systems) in *all* configurations is a fundamental requirement of GDPS. It is not possible to merge the function of the controlling system with any other system that accesses or uses the primary volumes or other production resources.

Configuring GDPS/PPRC with two controlling systems, one in each site is highly recommended. This is because a controlling system is designed to survive a failure in the opposite site of where the primary disks are. Primary disks are normally in Site1 and the controlling system in Site2 is designed to survive if Site1 or the disks in Site1 fail. However, if you reverse the configuration so that primary disks are now in Site2, the controlling system is in the same site as the primary disks. It will certainly not survive a failure in Site2 and might not survive a failure of the disks in Site2 depending on the configuration. Configuring a controlling system in both sites ensures the same level of protection, no matter which site is the primary disk site. When two controlling systems are available, GDPS manages assigning a Master role to the controlling system that is in the same site as the secondary disks and switching the Master role if there is a disk switch.

Improved controlling system availability: Enhanced STP support

Normally, a loss of synchronization with the sysplex timing source will generate a disabled console WTOR that suspends all processing on the LPAR, until a response is made to the WTOR. The WTOR message is IEA394A.

In a GDPS environment, z/OS is aware that a given system is a GDPS controlling system and will allow a GDPS controlling system to continue processing even when the server it is running on loses its time source and becomes unsynchronized. The controlling system is therefore able to complete any freeze or HyperSwap processing it might have started and is available for situation analysis and other recovery actions, instead of being in a disabled WTOR state.

In addition, because the controlling system is operational, it can be used to help in problem determination and situation analysis during the outage, thus reducing further the recovery time needed to restart applications.

The controlling system is required to perform GDPS automation if a failure occurs. Actions might include these tasks:

- ▶ Reassigning STP roles
- ▶ Performing the freeze processing to guarantee secondary data consistency
- ▶ Coordinating HyperSwap processing
- ▶ Executing a takeover script
- ▶ Aiding with situation analysis

Because the controlling system needs to run with only a degree of time synchronization that allows it to correctly participate in heartbeat processing with respect to the other systems in the sysplex, this system should be able to run unsynchronized for a period of time (80 minutes) using the local time-of-day (TOD) clock of the server (referred to as *local timing mode*), rather than generating a WTOR.

Automated response to STP sync WTORs

GDPS on the controlling systems, using the BCP Internal Interface, provides automation to reply to WTOR IEA394A when the controlling systems are running in local timing mode. See “Improved controlling system availability: Enhanced STP support” on page 69. A server in an STP network might have recovered from an unsynchronized to a synchronized timing state without client intervention. By automating the response to the WTORs, potential time outs of subsystems and applications in the client’s enterprise might be averted, thus potentially preventing a production outage.

If either WTOR IEA015A or IEA394A is posted for production systems, GDPS uses the BCP Internal Interface to automatically reply RETRY to the WTOR. If z/OS determines that the CPC is in a synchronized state, either because STP recovered or the CTN was reconfigured, it will no longer spin and continue processing. If the CPC is still in an unsynchronized state when GDPS automation responded with RETRY to the WTOR, however, the WTOR will be reposted.

The automated reply for any given system is retried for 60 minutes. After 60 minutes, you will need to manually respond to the WTOR.

3.2.2 Single-site workload configuration

A GDPS/PPRC single-site workload environment typically consists of a multisite sysplex, with all production systems running in a single site, normally Site1, and the GDPS controlling system in Site2. The controlling system (or systems, because you may have two in some configurations) will normally run in the site containing the secondary disk volumes.

The multisite sysplex can be a base sysplex or a Parallel Sysplex; a coupling facility is not strictly required. The multisite sysplex must be configured with redundant hardware (for example, a coupling facility and a Sysplex Timer in each site), and the cross-site connections must also be redundant. Instead of using Sysplex Timers to synchronize the servers, you can also use Server Time Protocol (STP) to synchronize the servers.

Figure 3-2 shows a typical GDPS/PPRC single-site workload configuration. The LPARs in blue (P1, P2, P3, and K1) are in the production sysplex, as are the coupling facilities CF1 and CF2. The primary disks are all in Site1, with the secondaries in Site2. All the production systems are running in Site1, with only the GDPS controlling system (K1) running in Site2. You will notice that system K1's disks (those marked K) are also in Site2. The unlabeled boxes represent work that can be displaced, such as development or test systems.

The GDPS/PPRC code itself runs under NetView and System Automation, and runs in every system in the GDPS sysplex.

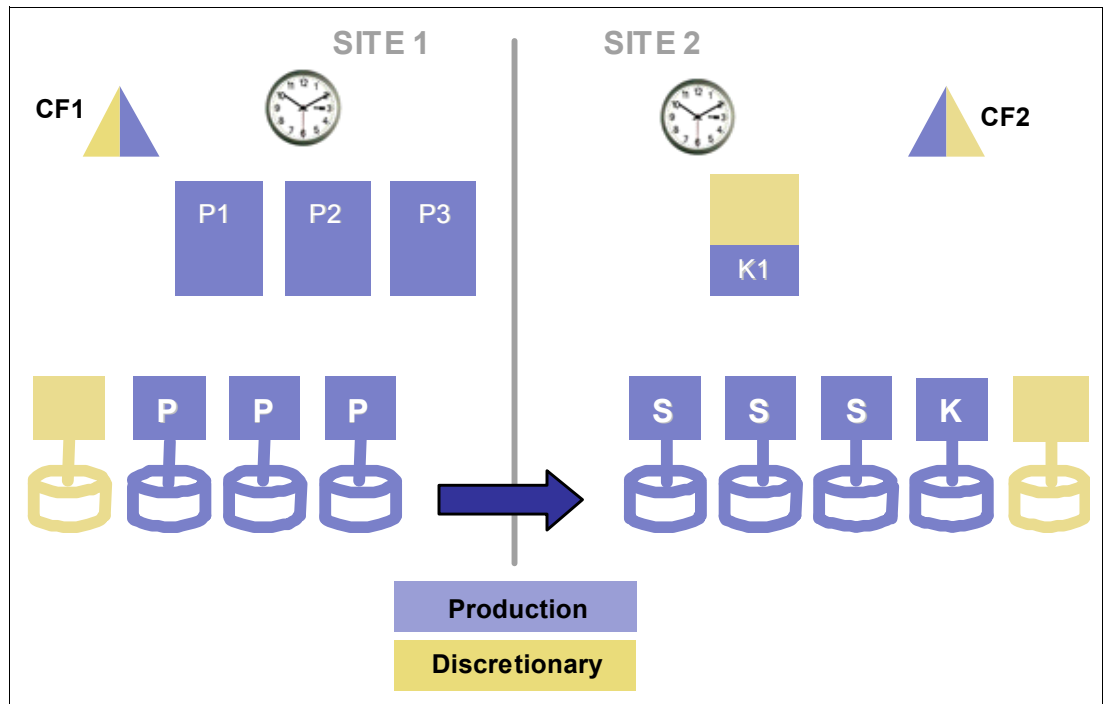


Figure 3-2 GDPS/PPRC Active/Standby workload configuration

3.2.3 Multisite workload configuration

A multisite workload configuration, shown in Figure 3-3, differs from a single-site workload in that production systems are running in *both* sites. Although running a multisite workload as a base sysplex is possible, seeing this configuration as a base sysplex (that is, without coupling facilities) is unusual. This is because a multisite workload is usually a result of higher availability requirements, and Parallel Sysplex and data sharing are core components of such an environment.

Because in this example we have production systems in both sites, we need to provide the capability to recover from a failure in either site. So, in this case, there is also a GDPS controlling system with its own local (not mirrored) disk running in Site1, namely System K2. Therefore, if there is a disaster that disables Site2, there will still be a GDPS controlling system available to decide how to react to that failure and what recovery actions are to be taken.

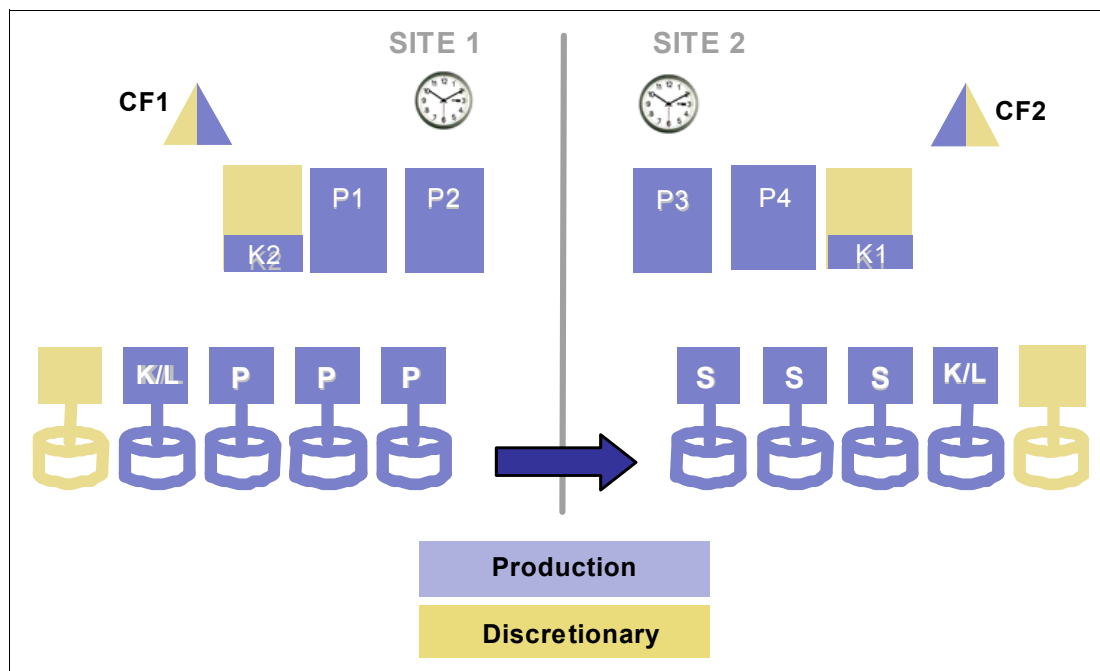


Figure 3-3 GDPS/PPRC Active/Active workload configuration

3.2.4 Business Recovery Services (BRS) configuration

A third configuration is known as the *BRS configuration*, and is illustrated in Figure 3-4 on page 73. In this configuration, all the systems in the GDPS configuration, including the controlling system, are in a sysplex in the same site, namely Site1. The sysplex does not span the two sites. The second site, Site2, might be a client site or might be owned by a third-party recovery services provider; thus the name BRS.

Site2 will contain the secondary disks and the alternate couple data sets (CDS), and might also contain processors that will be available in case of a disaster, but are not part of the configuration. This configuration can also be used when the distance between the two sites exceeds the distance supported for a multisite sysplex, but is within the maximum distance supported by FICON and Metro Mirror.

Even though there is no need for a multisite sysplex with this configuration, you must have channel connectivity from the GDPS systems to the secondary disk subsystems. Also, as explained in the next paragraph, the controlling system in Site1 will need channel connectivity to its disk devices in Site2. Therefore, FICON link connectivity from Site1 to Site2 will be required. See 2.9.7, “Connectivity options” on page 47, and *IBM z Systems Connectivity Handbook*, SG24-5444, for options available to extend the distance of FICON links between sites.

In the BRS configuration one of the two controlling systems must have its disk devices in Site2. This permits that system to be restarted manually in Site2 after a disaster is declared. After it restarts in Site2, then system runs a GDPS script to recover the secondary disk subsystems, reconfigure the recovery site and restart the production systems from the disk subsystems in Site2.

If you have only a single controlling system and you have a total cross-site fiber connectivity failure, the controlling system running on Site2 disks might not be able to complete the Freeze operation because it will lose access to its disk in Site2. Having a second controlling system running on Site1 local disks in will guarantee that the freeze operation completes successfully if the controlling system running on Site2 disks is down or is unable to function because of a cross-site fiber loss. GDPS will attempt to maintain the current Master system in the controlling system by using the secondary disks.

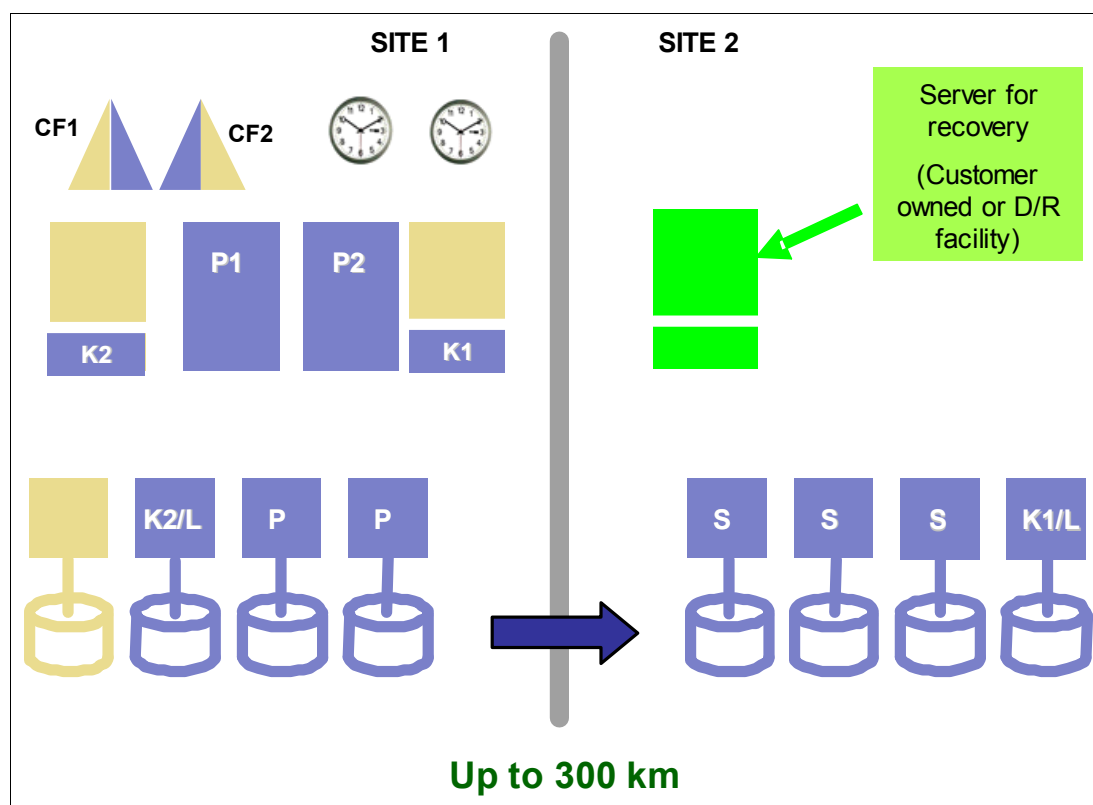


Figure 3-4 GDPS/PPRC BRS configuration

3.2.5 GDPS/PPRC in a 3-site or 4-site configuration

GDPS/PPRC can be combined with GDPS/XRC in a 3-site configuration or GDPS/GM in a three or 4-site configuration. In such configuration, GDPS/PPRC (when combined with Parallel Sysplex use and HyperSwap) in one region provides continuous availability across a metropolitan area or within the same local site, and GDPS/XRC or GDPS/GM provides disaster recovery capability using a remote site in a different region.

For 3-site configurations, the second region is predominantly for disaster recovery purposes because there is no HyperSwap protection in the recovery region.

The 4-site configuration is configured in a symmetric manner so that GDPS/PPRC is available in both regions to provide continuous availability (CA) with GDPS/GM to provide cross-region DR, no matter in which region production is running at any time.

We call these combinations GDPS/Metro Global Mirror (GDPS/MGM) or GDPS/Metro z/OS Global Mirror (GDPS/MzGM). In these configurations, GDPS/PPRC, GDPS/XRC, and GDPS/GM provide additional automation capabilities.

See Chapter 11, “Combining local and metro continuous availability with out-of-region disaster recovery” on page 331 for more information about GDPS/MGM and GDPS/MzGM.

3.2.6 GDPS/PPRC in a single site

The final configuration is where you want to benefit from the capabilities of GDPS/PPRC to extend the continuous availability attributes of a Parallel Sysplex to planned and unplanned disk reconfigurations, but you do not have the facilities to mirror disk across two sites. In this case, you can implement GDPS/PPRC HyperSwap Manager (GDPS/PPRC HM or GDPS/HM).

GDPS/PPRC HM is similar to the full function GDPS/PPRC offering, except that it does not include the scripts for management of the LPARs and workloads. GDPS/PPRC HM is upgradeable to a full GDPS/PPRC implementation. GDPS/PPRC HM is described in Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 103.

Because configuring GDPS/PPRC (or GDPS/HM) within a single site does not provide protection against site failure events, such a configuration is likely to be used within the context of a 3-site or 4-site solution rather than a stand-alone solution.

Another possibility is that this is for a client environment that has aggressive recovery time objectives for failures other than a disaster event and some mechanism such as tape vaulting is used for disaster protection. This means that long recovery times and a fair amount of data loss can be tolerated during a disaster.

3.2.7 Other considerations

The availability of the dedicated GDPS controlling system (or systems) in *all* scenarios is a fundamental requirement in GDPS. Merging the function of the controlling system with any other system that accesses or uses the primary volumes is not possible.

Equally important is that certain functions (stopping and restarting systems and changing the couple data set configuration) are done through the scripts and panel interface provided by GDPS. Because events such as systems going down or changes to the couple data set configuration are indicators of a potential disaster, such changes must be initiated using GDPS functions so that GDPS understands that these are planned events.

3.3 GDPS/PPRC management of distributed systems and data

As mentioned in 3.1.3, “Protecting distributed (FB) data” on page 66, it is possible for GDPS/PPRC to manage *FB disks* on behalf of distributed systems either in the same consistency group as the z Systems CKD disks or in a separate group.

GDPS/PPRC also provides capabilities to extend management of *distributed systems* in the following ways:

- ▶ GDPS/PPRC Multiplatform Resiliency for z Systems (also known as xDR)
- ▶ GDPS/PPRC Distributed Cluster management (DCM)

3.3.1 Multiplatform Resiliency for z Systems (also known as xDR)

To reduce IT costs and complexity, many enterprises are consolidating open servers into Linux on z Systems servers. Linux on z Systems can be implemented either as guests running under z/VM or native Linux on z Systems. Several examples exist of an application server running on Linux on z Systems and a database server running on z/OS. Two examples are as follows:

- ▶ WebSphere Application Server running on Linux and CICS, DB2 running under z/OS
- ▶ SAP application servers running on Linux and database servers running on z/OS

With a multitiered architecture, there is a need to provide a coordinated near-continuous availability and disaster recovery solution for both z/OS and Linux on z Systems. The GDPS/PPRC function that provides this capability is called Multiplatform Resiliency for z Systems, and it can be implemented if the disks being used by z/VM and Linux are CKD disks. For more details about this function, see 10.2, “GDPS/PPRC Multiplatform Resiliency for z Systems” on page 299.

3.3.2 Distributed Cluster Management

GDPS Distributed Cluster Management (DCM) is a capability that allows the management and coordination of disaster recovery across clustered distributed servers and the z Systems workload (or workloads) that GDPS is responsible for.

The DCM support is provided in GDPS/PPRC for both Symantec Veritas Cluster Server (VCS) clusters and IBM Tivoli System Automation Application Manager (SA AppMan). GDPS/PPRC can support both VCS and SA AppMan concurrently. DCM provides advisory and coordination functions between GDPS and one or more VCS or SA AppMan managed clusters.

For more information about the DCM function, see 10.3, “Distributed Cluster Management” on page 307.

3.3.3 IBM zEnterprise BladeCenter Extension (zBX) hardware management

GDPS provides support for activating and deactivating blades and virtual servers on zBX hardware. This support can be combined with the GDPS Distributed Cluster Management for managing workloads controlled by either SA AppMan or VCS, further extending GDPS’ end-to-end reach.

Additional information is provided in 10.3.4, “GDPS/PPRC Support for IBM zEnterprise BladeCenter Extension (zBX)” on page 329.

3.4 Management of z/OS systems outside of the GDPS sysplex

In 3.1.4, “Protecting other CKD data” on page 67, we describe a method which allows GDPS to monitor and manage PPRC on behalf of systems that are not running in the GDPS sysplex. We refer to such non-GDPS systems outside of the sysplex as *foreign systems* and we refer to the disk of these systems as *foreign disk*. Managing foreign systems and the foreign disk using the method described in the above referenced section has a key limitation in that this method does not support HyperSwap for the foreign systems. Although the foreign disks are included in the swap scope, the foreign systems are required to be stopped before a planned swap and will hang on Extended Long Busy as a result of an unplanned swap after which they must be reset and reloaded.

GDPS/PPRC does however, provide a feature known as the *z/OS Proxy* that extends the near continuous availability protection of HyperSwap to z/OS systems that are running outside of the GDPS sysplex. This includes stand-alone z/OS systems (MONOPLEX or XCFLOCAL) and systems running in a multi-system sysplex other than the GDPS sysplex.

In a z/OS Proxy environment, a GDPS/PPRC agent runs in each of the z/OS Proxy-managed systems outside of the GDPS sysplex. This agent, known as the z/OS Proxy, communicates with the Master GDPS controlling system, which facilitates coordinated planned and unplanned HyperSwap, as well as coordinated freeze processing across the systems in the GDPS sysplex as well as all z/OS systems that are managed by the z/OS Proxy.

In addition to PPRC, Freeze and HyperSwap management, basic hardware management (for example, automated system resets and IPLs) of the z/OS Proxy-managed systems is also provided. However, some GDPS/PPRC functions such as the Stop action, as well as CDS and CF management functions for z/OS Proxy-managed systems running in foreign sysplexes, are not available.

Figure 3-5 represents a basic configuration to help explain the support that GDPS provides in monitoring and managing the z/OS Proxy-managed systems and the PPRCed disks that are used by these systems.

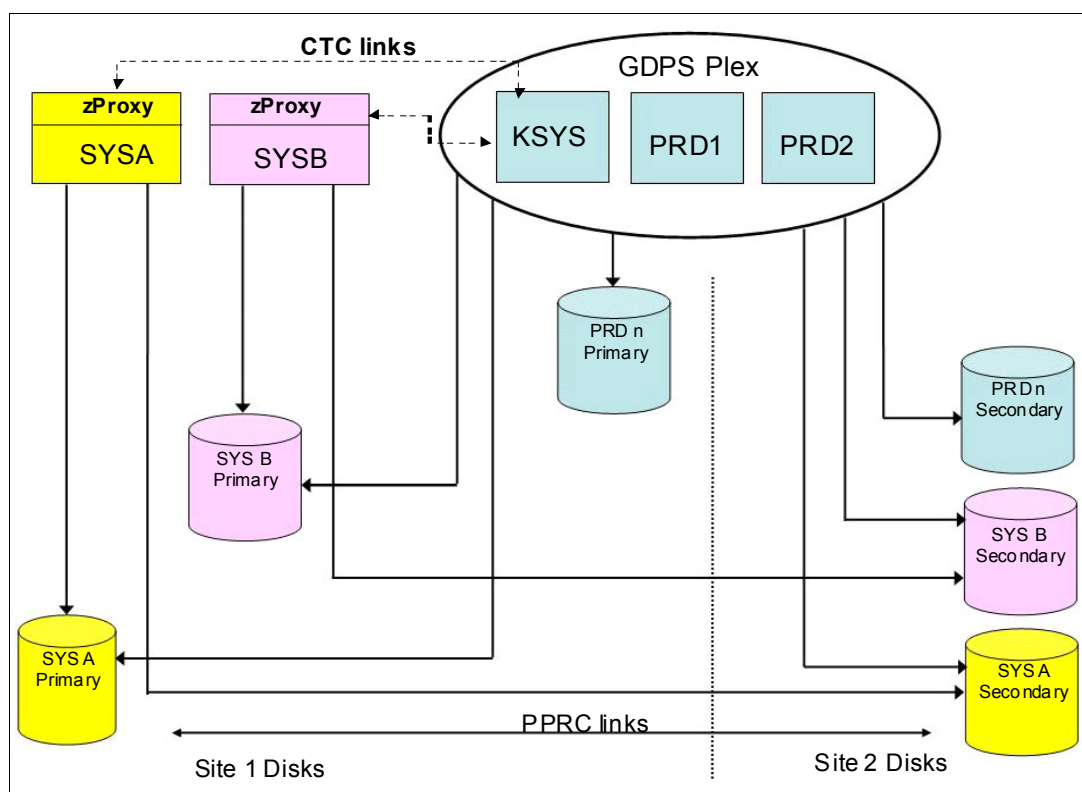


Figure 3-5 Sample z/OS Proxy Environment

In Figure 3-5, the traditional GDPS sysplex environment consists of production systems PRD1 and PRD2 and the controlling system KSYS. The primary disks for these GDPS production systems in Site1 are PPRC mirrored to Site2. This environment represents a standard GDPS/PPRC installation.

The systems SYSA and SYSB are z/OS Proxy-managed systems. They are outside of the GDPS sysplex and do not run GDPS NetView or System Automation code. Instead, they run the z/OS Proxy agent which communicates and coordinates actions with the Master GDPS controlling system.

The z/OS Proxy-managed systems are connected to the controlling systems via FICON Channel-to-Channel connections.

The z/OS Proxy-managed systems do not need host attachment to the disks belonging to the systems in the GDPS sysplex and do not need to define those disks. However, the systems in the GDPS sysplex *do* need to have UCBs for and have host channel attachment to, all PPRCed disk, their own as well as all disks belonging to the z/OS Proxy-managed disks.

Alternate subchannel sets (MSS1) *cannot be used* for secondary devices for z/OS Proxy-managed systems. All disks belonging to z/OS Proxy-managed systems, both Site1 and Site2, must be defined to MSS0 in all systems having these disks defined.

3.4.1 z/OS Proxy disk and disk subsystem sharing

The PPRCed disk attached to the z/OS Proxy-managed systems can reside either on separate physical disk subsystems or in the same physical disk subsystems as the disk belonging to the systems in the GDPS sysplex. PPRCed disks for the systems in the GDPS sysplex and the z/OS Proxy-managed systems can also be co-located in the same LSS.

As hardware reserves are not allowed in a GDPS HyperSwap environment, GDPS systems and z/OS Proxy-managed systems cannot share the GDPS-managed PPRCed disks. Systems in the GDPS sysplex can share disks among themselves provided that reserves in the GDPS sysplex are converted to global enqueues. Similarly, systems in any given foreign sysplex can share disks with each other if reserves in the foreign sysplex are converted to global enqueues. No other sharing is possible.

3.5 Managing the GDPS/PPRC environment

We have seen how GDPS/PPRC can protect just about any type of data that can reside in a disk subsystem. Further, it can provide data consistency across all platforms. However, as discussed in Chapter 1, “Introduction to business resilience and the role of GDPS” on page 1, the overwhelming majority of z Systems outages are not disasters. Most are planned outages, with a small percentage of unplanned ones.

In this section, we describe the other aspect of GDPS/PPRC, that is, its ability to monitor and manage the resources in its environment. GDPS provides two mechanisms to help you manage the GDPS sysplex and resources within that sysplex. One mechanism is the NetView interface and the other is support for scripts. We review both of these mechanisms here.

3.5.1 NetView interface

Two primary user interface options are available for GDPS/PPRC: The NetView 3270 panels and a browser-based graphical user interface (also referred to as the *GDPS GUI* in this book).

An example of the main GDPS/PPRC 3270-based panel is shown in Figure 3-6.

```

VPCPPNLI          GDPS - Disaster/Recovery System          GDPS V3.R10.M0

System             = G0C2      A6PS3      Primary Dasd = OK   SITE1  MOP1
Current Master     = G0C2      A6PS3
Parallel mode      = YES
HyperSwap F0/FB    = ENABLED YES      Dasd Config = 2013-02-08 15:33:25
Debug              = OFF              FREEZE Date =
                                   Time =

      1            Dasd Remote Copy

      3            Standard Actions

      6            Planned Actions
      7            Sysplex Resource Management
      8            Debug ON/OFF
      9            View/Alter Definitions
      H            Health Checks
      C            Config Management
      M            Run Monitor1/Monitor3

Selection ==> _
                Licensed Materials - Property of IBM
                6942-35B © Copyright IBM Corp. 1998, 2012 All Rights Reserved.

```

Figure 3-6 Main GDPS/PPRC 3270-based panel

This panel has a summary of configuration status at the top, and a menu of selectable choices. As an example, to view the disk mirroring (Dasd Remote Copy) panels enter 1 at the Selection prompt, and then press Enter.

GDPS graphical user interface

The GDPS GUI is a browser-based interface designed to improve operator productivity. The GDPS GUI provides the same functional capability as the 3270-based panel, such as providing management capabilities for Remote Copy Management, Standard Actions, Sysplex Resource Management, SDF Monitoring, and browsing the CANZLOG using simple point-and-click procedures. Advanced sorting and filtering is available in most of the views provided by the GDPS GUI. In addition, users can open multiple windows or tabs to allow for continuous status monitoring, while performing other GDPS/PPRC management functions.

The GDPS GUI display has four main sections:

1. The application header at the top of the page that provides an Actions button for carrying out a number of GDPS tasks, along with the help function and the ability to logoff or switch between target systems.
2. The application menu is down the left side of the screen. This menu gives access to various features and functions available through the GDPS GUI.
3. The active screen that shows context-based content, depending on the selected function. This tabbed area is where the user can switch context by clicking a different tab.
4. A status summary area is shown at the bottom of the display.

The initial status panel of the GDPS/PPRC GUI is shown in Figure 3-7. This panel provides an instant view of the status and direction of replication, HyperSwap status, and systems and systems availability. Hovering over the various icons provides more information by using pop-up windows.

Note: For the remainder of this section, only the GDPS GUI is shown to illustrate the various GDPS management functions. The equivalent traditional 3270 panels are not shown here.

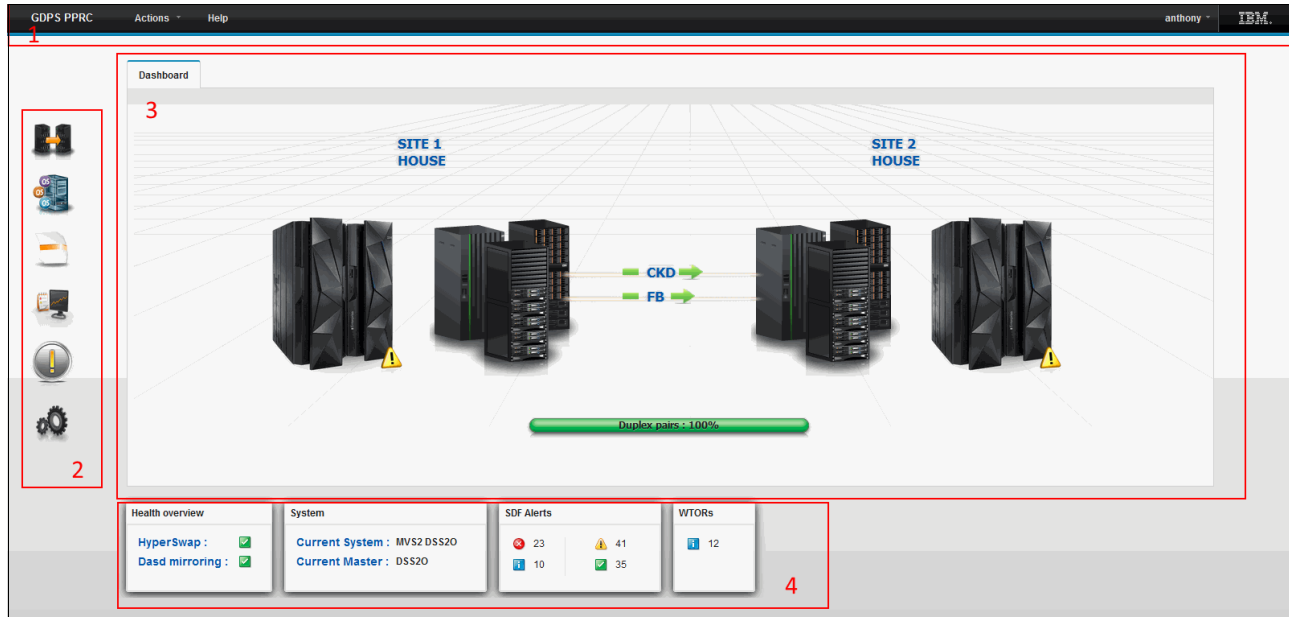


Figure 3-7 Full view of GDPS GUI main panel

Monitoring function: Status Display Facility

GDPS also provides many monitors to check the status of disks, sysplex resources, and so on. Any time there is a configuration change, or something in GDPS that requires manual intervention, GDPS will raise an alert. GDPS uses the Status Display Facility (SDF) provided by System Automation as the primary status feedback mechanism for GDPS.

GDPS provides a dynamically updated panel, as shown in Figure 3-8. There is a summary of all current alerts at the bottom of each panel. The initial view presented is for the SDF trace entries so you can follow, for example, script execution. Simply click one the other alert categories to view the different alerts associated with automation or remote copy in either site, or select **All** to see all alerts. You can sort and filter the alerts based on a number of the fields presented, such as severity.

The GDPS GUI refreshes the alerts automatically every 10 seconds by default. As with the 3270 panel, if there is a configuration change or a condition that requires special attention, the color of the fields will change based on the severity of the alert. By pointing to and clicking any of the highlighted fields, you can obtain detailed information regarding the alert.

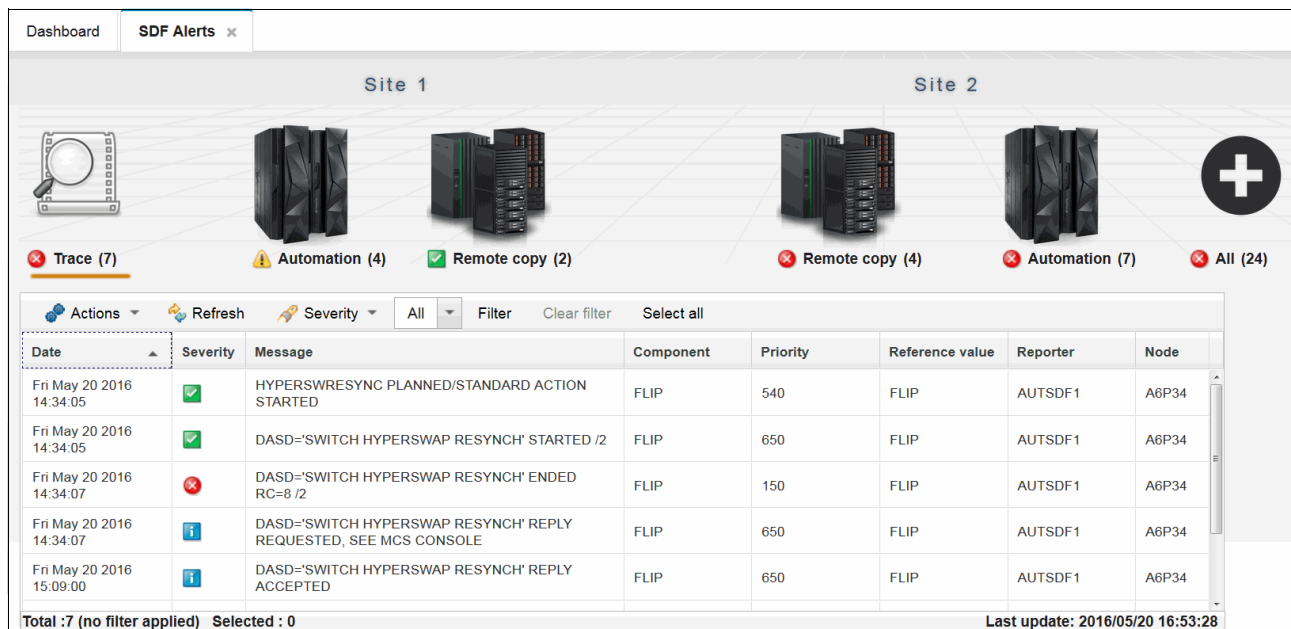


Figure 3-8 GDPS GUI SDF panel

Remote copy panels

The z/OS Advanced Copy Services capabilities are powerful, but the native command-line interface (CLI), z/OS TSO, and ICKDSF interfaces are not as user-friendly as the DASD remote copy panels are. To more easily check and manage the remote copy environment, use the DASD remote copy panels provided by GDPS.

For GDPS to manage the remote copy environment, you must first define the configuration (primary and secondary LSSs, primary and secondary devices, and PPRC links) to GDPS in a file called the GEOPARM file. This GEOPARM file can be edited and introduced to GDPS directly from the GDPS GUI.

After the configuration is known to GDPS, you can use the panels to check that the current configuration matches the one you want. You can start, stop, suspend, and resynchronize mirroring at the volume or LSS level. These actions can be done at the device or LSS level, or both, as appropriate. Figure 3-9 shows the mirroring panel for CKD devices at the LSS level.

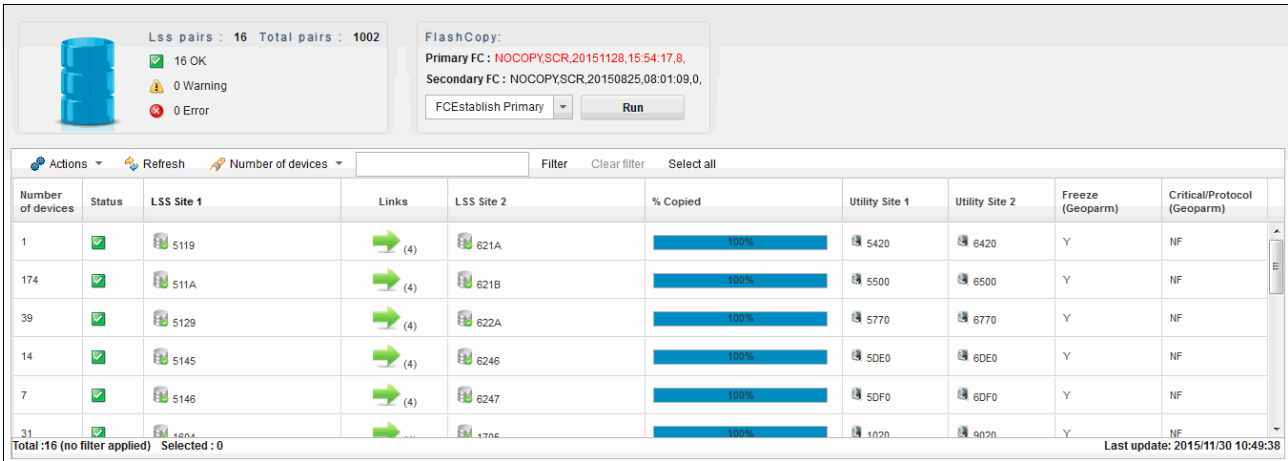


Figure 3-9 GDPS GUI Dasd Remote Copy SSID panel

The Dasd Remote Copy panel is organized into three sections:

- ▶ Upper left provides a summary of the device pairs in the configuration and their status.
- ▶ Upper right provides the ability to invoke GDPS-managed FlashCopy operations.
- ▶ A table with one row for each LSS pair in your GEOPARM. In addition to the rows for each LSS, there is a header row with an **Action** menu to enable you to carry out the various DASD management tasks, and the ability to filter the information presented.

To perform an action on a single SSID-pair, double click a row in the table. A panel is then displayed, where you can perform the same actions as those available as line commands on the top section of the 3270 panel.

After an individual SSID-pair is selected, the frame shown in Figure 3-10 is displayed. The table in this frame shows each of the mirrored device pairs within a single SSID-pair, along with the current status of each pair. In this example, all the pairs are fully synchronized and in duplex status, as summarized in the upper left area. Additional details can be viewed for each pair by double clicking the row, or by selecting the row with a single click and then selecting **Query** from the **Actions** menu.

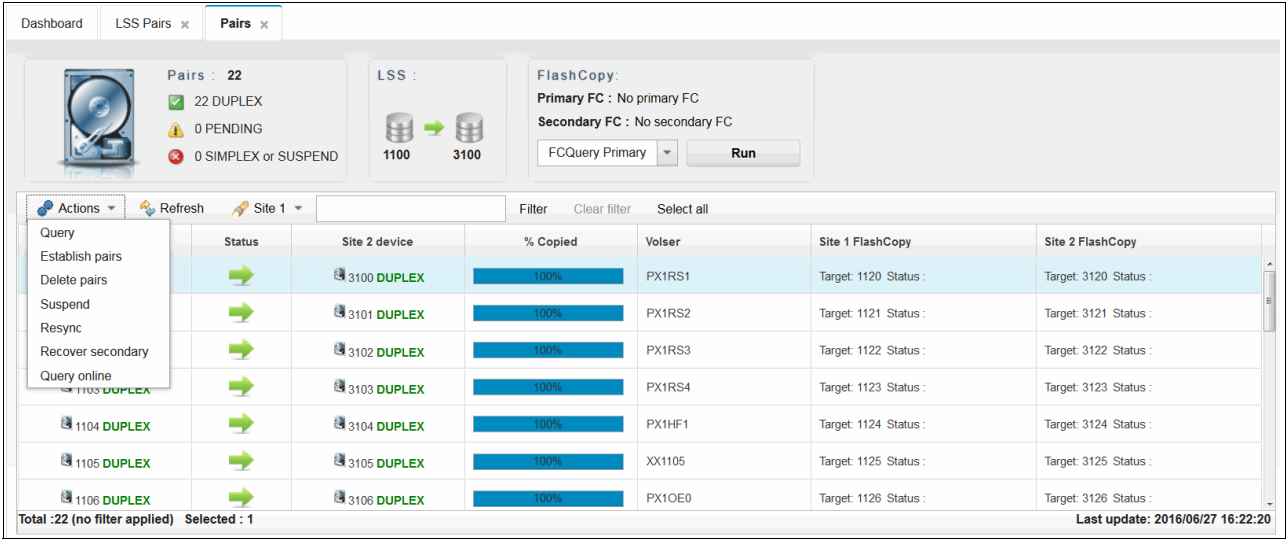


Figure 3-10 GDPS GUI Dasd Remote Copy: View Devices detail panel

If you are familiar with using the TSO or ICKDSF interfaces, you might appreciate the ease of use of the DASD remote copy panels.

Remember that these panels provided by GDPS are *not* intended to be a remote copy monitoring tool. Because of the overhead involved in gathering the information for every device to populate the panels, GDPS gathers this data only on a timed basis, or on demand following an operator instruction. The normal interface for finding out about remote copy status or problems is the Status Display Facility (SDF).

Similar panels are provided for controlling the FB devices.

Standard Actions

GDPS provides facilities to help manage many common system-related planned actions. There are two reasons to use the GDPS facilities to perform these *Standard Actions*:

- ▶ They are well tested and based on IBM preferred procedures.
- ▶ Using the GDPS interface lets GDPS know that the changes that it is seeing (couple data sets (CDS) being deallocated or systems going out of the sysplex, for example) are planned changes, and therefore GDPS is not to react to these events.

There are two types of resource-altering actions you can initiate from the panels. Those that GDPS calls *Standard Actions* are actually single steps, or are intended to impact only one resource. Examples are starting a system IPL, maintaining the various IPL address and load parameters that can be used to IPL a system, selecting the IPL address and load parameters to be used the next time a system IPL is started, or activating an LPAR. So if you want to stop a system, change its IPL address, then perform an IPL, you initiate three separate Standard Actions.

The GDPS/PPRC Standard Actions 3270 panel is shown in Figure 3-11. It displays all the systems being managed by GDPS/PPRC, and for each one it shows the current status and various IPL information. To perform actions on each system, you simply use a line command letter (L to load, X to reset and so on) next to the selected system.

VPCPSTD1		Standard Actions							G7C2
Actions:		S Stop	R ReIPL	M odify	Q QryxDR				
		L Load	X Reset	A Activate	D Deactivate	U Dump			
Sysname	CA	Status	IPLtype	LPAR	IPLmode	Auto	L-addr	Loadparm	
— SITE1		MOP1							
— G7C1	CA	ACTIVE	NORMAL	S691	SITE1	YN	A200	A207G7	
— G7P1	A	ACTIVE	NORMAL	S692	SITE1	YN	A300	A308G7	
— CF1		MANUAL	NORMAL	S695	NORMAL	NN			
— SITE2		MOP2							
— G7C2	CA	MASTER	NORMAL	S291	SITE2	YN	A220	A227G7	
— G7P2	A	ACTIVE	NORMAL	S292	SITE1	YN	A300	A308G7	
— CF2		MANUAL	NORMAL	S295	NORMAL	NN			
1 CPC Ops									
Selection ==>									
F1=Help		F3=Return		F6=Roll		F11=Right			

Figure 3-11 GDPS/PPRC Standard Actions panel

GDPS supports taking a stand-alone dump using the GDPS Standard Actions panel. The stand-alone dump can be performed for any z Systems operating system defined to GDPS, either a GDPS system or a foreign system, running native in an LPAR. Clients using GDPS facilities to perform HMC actions no longer need to use the HMC for taking stand-alone dumps.

Sysplex resource management

There are certain resources that are vital to the health and availability of the sysplex. In a multisite sysplex, it can be quite complex trying to manage these resources to provide the required availability while ensuring that any changes do not introduce a single point of failure.

The GDPS/PPRC Sysplex Resource Management GUI panel, as shown in Figure 3-12 on page 85, provides you with the ability to manage the resources, with knowledge about where the resources exist. Simply click the resource type (couple data sets or coupling facilities) to open a panel to manage each resource type. For example, normally you have your Primary CDS in Site1, and your alternates in Site2. However, if you will be shutting down Site1, you still want to have a Primary and Secondary set of CDS, but both must be in Site2. The GDPS Sysplex Resource Management panels provide this capability, without you having to know specifically where each CDS is located.

GDPS provides facilities to manage coupling facilities (CFs) in your sysplex. These facilities allow for isolating all of your structures in the CF or CFs in a single site and returning to your normal configuration with structures spread across (and possibly duplexed across) the CFs in the two sites.

Isolating structures into CFs in one site, or returning to normal use with structures spread across CFs in both sites, can be accomplished through the GDPS Sysplex Resource Management panel interface or GDPS scripts. This provides an automated means for managing CFs for planned and unplanned site or disk subsystem outages.

The maintenance mode switch allows you to start or stop maintenance mode on a single CF (or multiple CFs, if all selected CFs are in the same site). DRAIN, ENABLE, and POPULATE function is still available for single CFs.

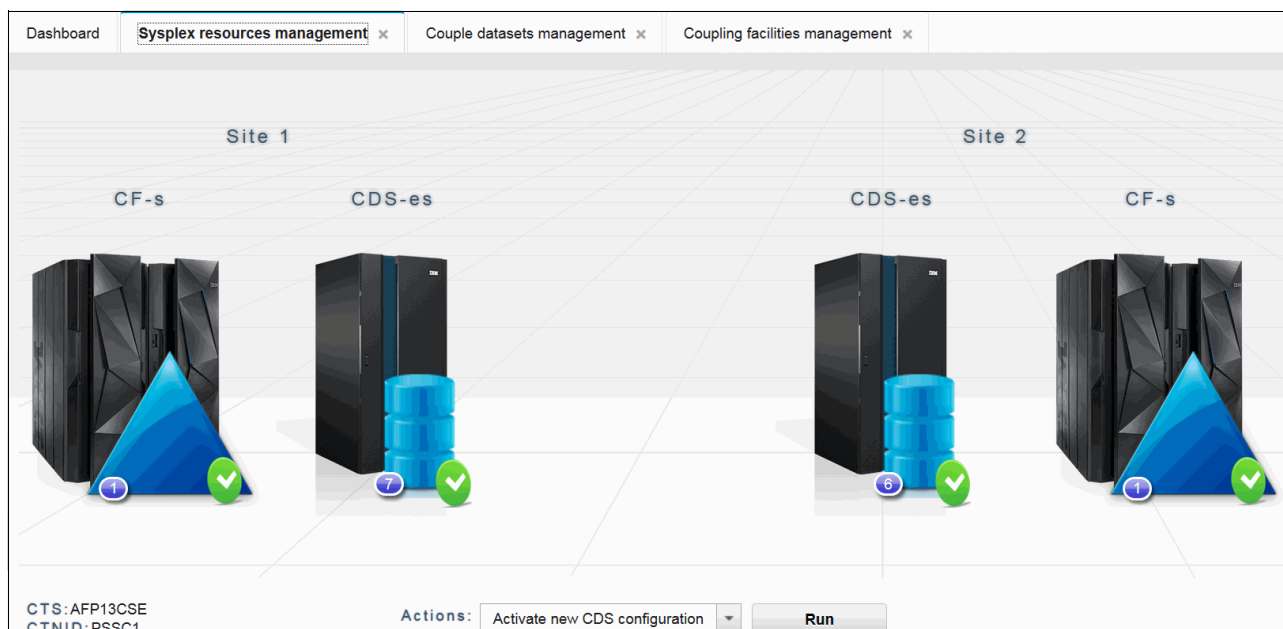


Figure 3-12 GDPS/PPRC Sysplex Resource Management GUI panel

3.5.2 GDPS scripts

At this point we have shown how GDPS panels provide powerful functions to help you manage GDPS resources. However, using GDPS panels is only one way of accessing this capability. Especially when you need to initiate what might be a complex, compound, multistep procedure, it is much simpler to use a script which in effect is a workflow.

Nearly all of the main functions that can be initiated through the GDPS panels are also available using GDPS scripts. Scripts also provide additional capabilities that are not available using the panels.

A “script” is simply a procedure recognized by GDPS that pulls together one or more GDPS functions. Scripts can be initiated manually for a planned activity through the GDPS panels (using the Planned Actions interface), automatically by GDPS in response to an event (Unplanned Actions), or through a batch interface. GDPS performs the first statement in the list, checks the result, and only if it is successful, proceeds to the next statement. If you perform the same steps manually, you would have to check results, which can be time-consuming, and initiate the next action. With scripts, the process is automated.

Scripts can easily be customized to automate the handling of various situations, both to handle planned changes and unplanned situations. This is an extremely important aspect of GDPS. Scripts are powerful because they can access the full capability of GDPS. The ability to invoke all the GDPS functions through a script provides the following benefits:

- **Speed**

The script will execute the requested actions and check results at machine speeds. Unlike a human, it does not need to search for the latest procedures or the commands manual.

- **Consistency**

If you were to look into most computer rooms immediately following a system outage, what would you see? Mayhem, with operators frantically scrambling for the latest system programmer instructions. All the phones ringing. Every manager within reach asking when the service will be restored. And every systems programmer with access vying for control of the keyboards. All this results in errors because humans naturally make mistakes when under pressure. But with automation, your well-tested procedures will execute in exactly the same way, time after time, regardless of how much you shout at them.

- **Thoroughly tested procedures**

Because they behave in a consistent manner, you can test your procedures over and over until you are sure they do everything that you want, in exactly the manner that you want. Also, because you need to code everything and cannot assume a level of knowledge (as you might with instructions intended for a human), you are forced to thoroughly think out every aspect of the action the script is intended to undertake. And because of the repeatability and ease of use of the scripts, they lend themselves more easily to frequent testing than manual procedures.

Planned Actions

As mentioned earlier, GDPS scripts are simply procedures that pull together into a list one or more GDPS functions. For the scripted procedures that you might use for a planned change, these scripts can be initiated from the panels called *Planned Actions* (option 6 on the main GDPS panel as shown in Figure 3-6 on page 79).

As one example, you can have a short script that stops an LPAR and then re-IPLs it in an alternate LPAR location, as shown in Example 3-1.

Example 3-1 Sample script to re-IPL a system

```
COMM='Example script to re-IPL system SYS1 on alternate ABNORMAL LPAR location'  
SYSPLEX='STOP SYS1'  
IPLTYPE='SYS1 ABNORMAL'  
SYSPLEX='LOAD SYS1'
```

A more complex example of a Planned Action is shown in Figure 3-13 on page 87. In this example, a single action in GDPS executing a planned script of only a few lines results in a complete planned site switch. Specifically, the following actions are done by GDPS:

1. The systems in Site1, P1 and P3, are stopped (P2 and P4 remain active in this example).
2. The sysplex resources (CDS and CF) are switched to use only those in Site2.
3. A HyperSwap is executed to use the disk in Site2. The IPL parameters (IPL address and load parameters) are updated automatically by GDPS to reflect the new configuration.
4. The IPL location for the P1 and P3 systems are changed to the backup LPAR location in Site2.

5. The backup LPARs for P1 and P3 are activated.
6. P1 and P3 are IPLed in Site2 using the disk in Site2.

Using GDPS removes the reliance on out-of-date documentation, provides a single repository for information about IPL addresses and load parameters, and ensures that the process is done the same way every time with no vital steps accidentally overlooked.

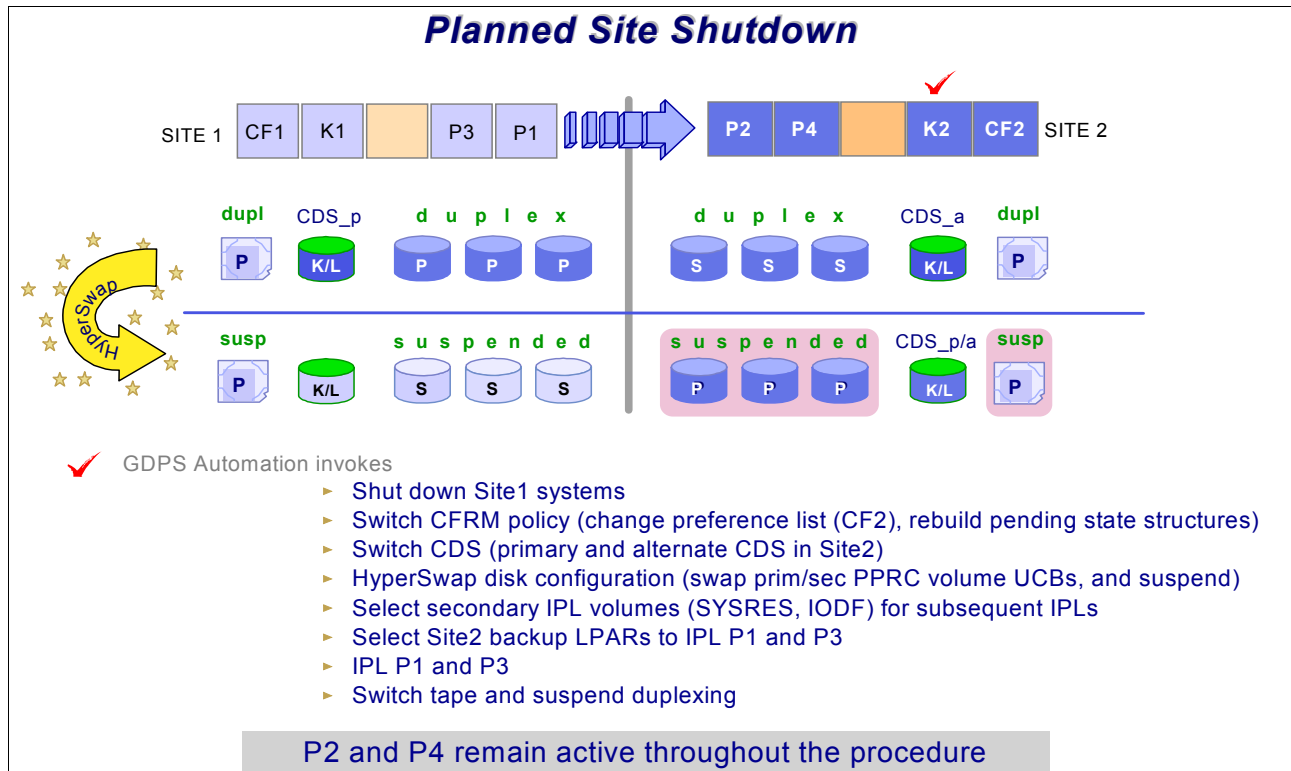


Figure 3-13 GDPS/PPRC Planned Action

STP CTN role reassignments: Planned operations

GDPS provides a script statement that allows you to reconfigure an STP-only CTN by reassigning the STP-only CTN server roles. In an STP CTN servers (CPCs) are assigned special roles to identify which CPC is preferred to be the clock source (Preferred Time Server, or PTS), which CPC is able to take over as the clock source for planned and unplanned events (Backup Time Server, or BTS), which CPC is the active clock source (Current Time Server, or CTS), and which CPC assists in STP recovery (Arbiter).

It is strongly recommended that the server roles be reassigned before performing planned disruptive actions on any of these special role servers. Examples of planned disruptive actions are power-on reset (POR) and Activate/Deactivate. The script statement can be integrated as part of your existing control scripts to perform these planned disruptive actions.

For example, if you are planning to deactivate the CPC that is the PTS/CTS, you can now execute a script to perform the following tasks:

- ▶ Reassign the PTS/CTS role to a different CPC in the CTN
- ▶ Optionally also reassign the BTS and Arbiter roles if required
- ▶ Execute script statements you might already have in place today to deactivate the PTS/CTS CPC

After the disruptive action is completed you can execute a second script to restore the STP roles to their normal operational state, as listed here:

- ▶ Script statement to activate the CPC
- ▶ Reassign the STP server roles to their normal operational state
- ▶ Statements you might already have in existing scripts to perform IPLs and so on

Recovery scripts

There are scripts also known as Takeover scripts that are designed to be invoked in case of a disaster or potential disaster. In the case of a Freeze-inducing event, GDPS/PPRC will immediately issue a freeze for all applicable primary devices. This is done automatically to protect the integrity of the secondary data.

After the freeze and the action indicated in the freeze policy (STOP or G0) has completed, GDPS will present the operator with a prompt listing the Takeover scripts that can be executed at this time. At this time, some situation analysis must be performed before accepting to run one of the proposed scripts. The freeze could be caused by a failing secondary disk or by a disaster in the primary site. The scripts to run for these two conditions will be different. After the analysis is complete, the operator selects the appropriate action and GDPS does the rest. You could view such actions as “preplanned unplanned actions.”

An example is shown in Figure 3-14. In this example, the operator has selected to abandon Site1 and move everything to Site2. GDPS will isolate CDSs and CF structures in Site2. It will also update all the IPL information to point to what were the secondary volumes in Site2; stop expendable systems (if there are any); invoke capacity backup (CBU) to provide the required capacity on the CPC in Site2; and re-IPL all the production systems in LPARs on that CPC. All these tasks are done with a single operator instruction.

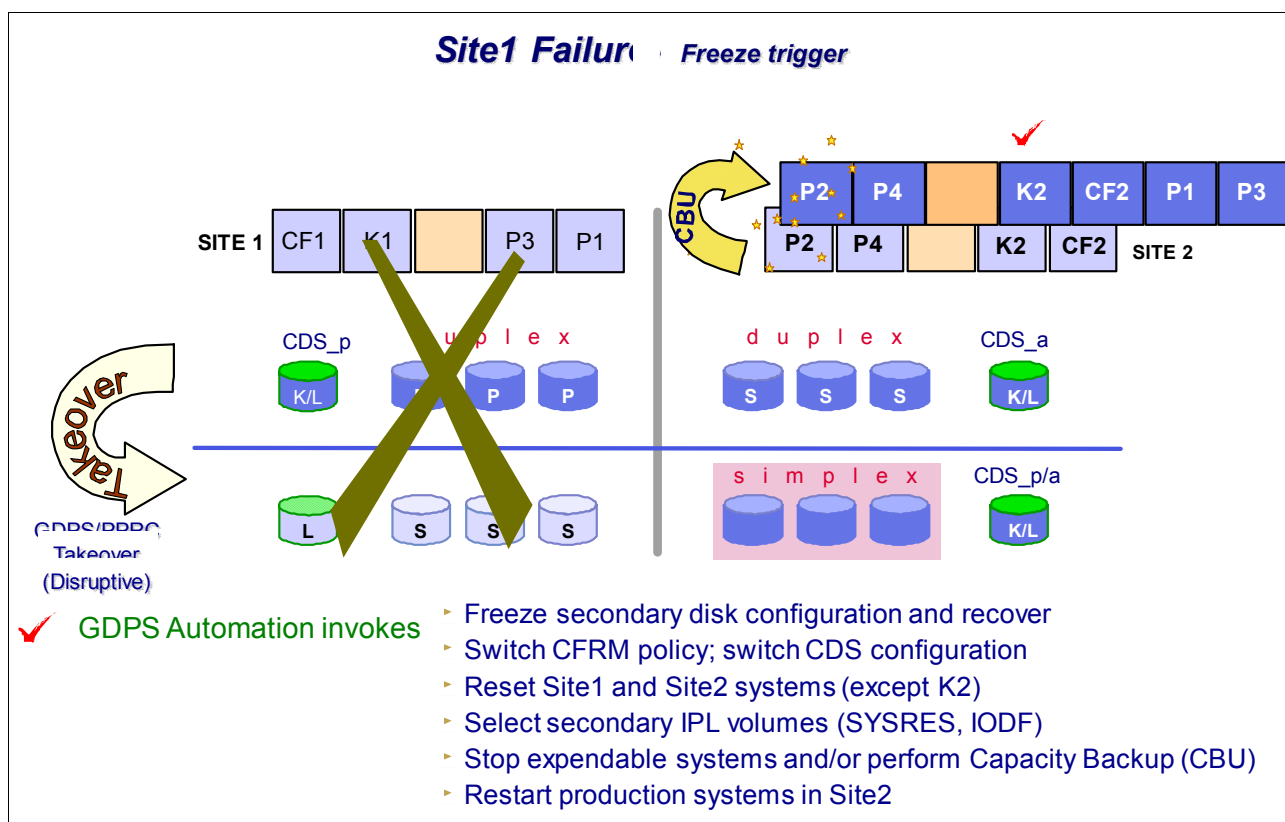


Figure 3-14 GDPS managed recovery from site failure

Another important aspect to disaster recovery is returning to the normal configuration after the unplanned outage. GDPS can help with this also, again using a GDPS script - a planned action script for this case. The actions to return to “normal” are similar, with one important difference. When you moved from Site1 to Site2, the data on the primary and secondary disks was identical (synchronized) at the time of the move. But when you move back, the disks in Site1 do not contain the updates made when production was running using the Site2 disks and must first be resynchronized before the move back to normal. During the period when they are being resynchronized, the secondary volumes have no consistency; remember that the missing updates are not applied in chronological order. What would happen if you had a disaster in Site2 during this window?

If the disaster were a fire, your current primary volumes would be a pool of molten metal on the computer room floor. The secondary disks’ data would be inconsistent. It would be as though you had a random collection of data.

To ensure that you at least have a set of consistent disks in Site1, even if they are not completely current, GDPS can be instructed to first take a FlashCopy of those volumes while they are still consistent before it starts the resynchronization. Thus, if you are unfortunate enough to lose Site2 during this resynchronization period, you at least have a consistent set of disks in Site1 that you can fall back on.

Because it might take a long time for the data to come back to sync, you would need two scripts for your return-to-normal operation: one script to take a FlashCopy and then resynchronize the disk, and a second script which you would initiate to bring systems, disk, and sysplex resources back to normal.

GDPS monitors data-related events, and scripts can be provided to guide recovery from a system failure. As seen in Figure 3-15, when GDPS detects that a z/OS system is no longer active, it verifies whether the policy definition indicates that Auto IPL is enabled, that the threshold of the number of IPLs in the predefined time window has not been exceeded, and that no planned action is active. If these conditions are met, GDPS can automatically re-IPL the system in place, bring it back into the Parallel Sysplex, and restart the application workload.

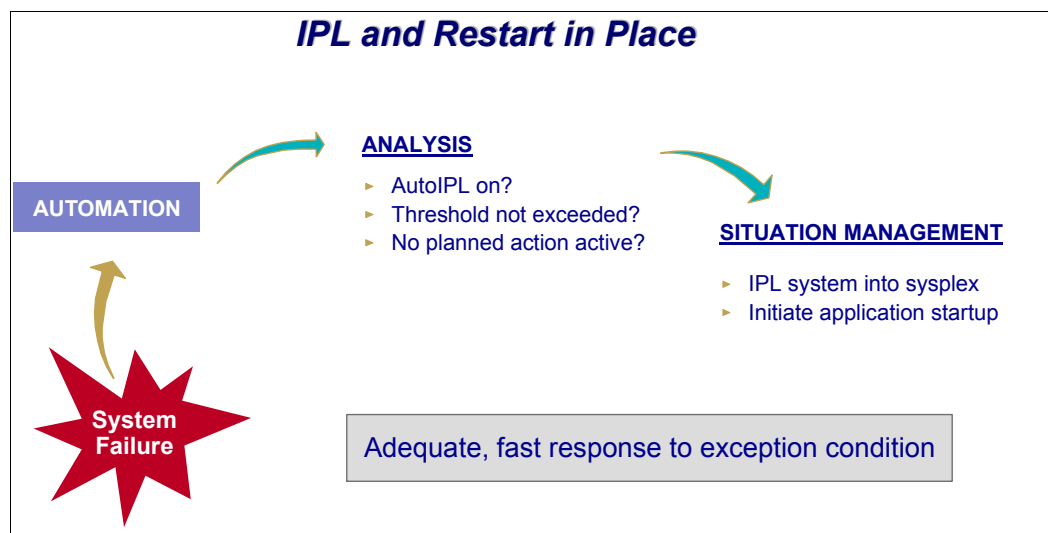


Figure 3-15 Recovering a failed image

Similarly, if a complete processor fails, you can have a script prepared to provide recovery support for all the LPARs on that processor that are managed by GDPS. This script would activate backup partitions for the failed systems, activate CBU if appropriate, and IPL these systems. You could have one such script prepared in advance for every server in your configuration.

STP CTN role reassignments: Unplanned failure

After a failure condition that has resulted in the PTS, BTS, or Arbiter being assigned on a CPC that is no longer operational or synchronized in the CTN, it is best to reassign any affected roles to operational CPCs after any STP recovery actions have completed. The reassignment reduces the potential for a sysplex outage if a second failure or planned action affects one of the remaining special role CPCs.

The script statement capability described in “STP CTN role reassignments: Planned operations” on page 87 can be used to integrate the STP role reassignment as part of an existing script such as a site takeover script, and eliminate the requirement for the operator to perform the STP reconfiguration task manually at the HMC.

STP WTOR IEA394A response: Unplanned failure

As described in “Improved controlling system availability: Enhanced STP support” on page 69, a loss of synchronization with the sysplex timing source will generate a disabled console WTOR. This suspends all processing on the LPAR until a response to the WTOR is provided. The WTOR message is IEA394A if the CPC is in STP timing mode (either in an STP mixed CTN or STP-only CTN).

GDPS, using scripts, can reply (either ABORT or RETRY) to the IEA394A sync WTOR for STP on systems that are spinning because of a loss of synchronization with their Current Time Source. As described in “Automated response to STP sync WTORS” on page 70, autonomic function exists to reply RETRY automatically for 60 minutes on any GDPS systems that have posted this WTOR.

The script statement complements and extends this function, as described:

- ▶ It provides a way to reply to the message after the 60-minute automatic reply window expires.
- ▶ It can reply to the WTOR on systems that are not GDPS systems (foreign systems) that are defined to GDPS. The autonomic function replies only on GDPS systems.
- ▶ It provides the ability to reply **ABORT** on any systems that you do not want to restart for a specified failure scenario before reconfiguration and synchronization of STP.

Batch scripts

GDPS also provides a flexible batch interface to invoke planned action scripts. These scripts can be invoked:

- ▶ As a REXX program from a user terminal
- ▶ By using the IBM MVS™ MODIFY command to the NetView task
- ▶ From timers in NetView
- ▶ Triggered through the SA automation tables

This capability, along with the query services interface described in 3.7.4, “Query Services” on page 95, provides a rich framework for user-customizable systems management procedures.

3.5.3 System Management actions

Most of the GDPS Standard Actions require actions to be done on the HMC. The interface between GDPS and the HMC is through the BCP internal interface (BCPii). This allows GDPS to communicate directly with the hardware for automation of HMC actions, such as Load, Stop (graceful shutdown), RESET, Activate LPAR, and Deactivate LPAR. GDPS can also perform ACTIVATE (power-on reset), CBU ACTIVATE/UNDO, OOCOD ACTIVATE/UNDO, and STP role reassignment actions against an HMC object that represents a CPC.

The GDPS LOAD and RESET Standard Actions (available through the Standard Actions panel or the SYSPLEX script statement) allow specification of a *CLEAR* or *NOCLEAR* operand. This provides operational flexibility to accommodate client procedures, thus eliminating the requirement to use the HMC to perform specific LOAD and RESET actions.

Furthermore, when you LOAD a system using GDPS (panels or scripts), GDPS can listen for operator prompts from the system being IPLed and reply to such prompts. GDPS provides support for optionally replying to such IPL-time prompts automatically, removing reliance on operator skills and eliminating operator error for any messages that require replies.

SYSRES Management

Today many clients maintain multiple alternate z/OS SYSRES devices (also known as *IPLSETs*) as part of their maintenance methodology. GDPS provides special support to allow clients to identify IPLSETs. This removes the requirement for clients to manage and maintain their own procedures when IPLing a system on a different alternate SYSRES device.

GDPS can automatically update the IPL pointers after any disk switch or disk recovery action that changes the GDPS primary site indicator for PPRC disks. This removes the requirement for clients to perform additional script actions to switch IPL pointers after disk switches, and greatly simplifies operations for managing alternate SYSRES “sets.”

3.6 GDPS/PPRC monitoring and alerting

The GDPS SDF panel, discussed in “Monitoring function: Status Display Facility” on page 80, is where GDPS dynamically displays color-coded alerts.

Alerts can be posted as a result of an unsolicited error situation that GDPS listens for. For example, if one of the multiple PPRC links that provide the path over which PPRC operations take place is broken, there is an unsolicited error message issued. GDPS listens for this condition and will raise an alert on the SDF panel, notifying the operator of the fact that a PPRC link is not operational. Clients run with multiple PPRC links and if one is broken, PPRC continues over any remaining links.

However, it is important for operations to be aware that a link is broken and fix this situation because a reduced number of links results in reduced PPRC bandwidth and reduced redundancy. If this problem is not fixed in a timely manner and more links have a failure, it can result in production impact because of insufficient mirroring bandwidth or total loss of PPRC connectivity (which results in a freeze).

Alerts can also be posted as a result of GDPS periodically monitoring key resources and indicators that relate to the GDPS/PPRC environment. If any of these monitoring items are found to be in a state deemed to be not normal by GDPS, an alert is posted on SDF.

Various GDPS monitoring functions are executed on the GDPS controlling systems and on the production systems. This is because, from a software perspective, it is possible that different production systems have different views of some of the resources in the environment, and although status can be normal in one production system, it can be not normal in another. All GDPS alerts generated on one system in the GDPS sysplex are propagated to all other systems in the GDPS. This propagation of alerts provides for a single focal point of control. It is sufficient for the operator to monitor SDF on the master controlling system to be aware of all alerts generated in the entire GDPS complex.

When an alert is posted, the operator will have to investigate (or escalate, as appropriate) and corrective action will need to be taken for the reported problem as soon as possible. After the problem is corrected, this is detected during the next monitoring cycle and the alert is cleared by GDPS automatically.

GDPS/PPRC monitoring and alerting capability is intended to ensure that operations are notified and can take corrective action for any problems in their environment that can affect the ability of GDPS/PPRC to do recovery operations. This maximizes the chance of achieving your availability and RPO and RTO commitments.

3.6.1 GDPS/PPRC health checks

In addition to the GDPS/PPRC monitoring described, GDPS provides health checks. These health checks are provided as a plug-in to the z/OS Health Checker infrastructure to check that certain settings related to GDPS adhere to preferred practices.

The z/OS Health Checker infrastructure is intended to check a variety of settings to determine whether these settings adhere to z/OS optimum values. For settings found to be not in line with preferred practices, exceptions are raised in the Spool Display and Search Facility (SDSF). If these settings do not adhere to recommendations, this can hamper the ability of GDPS to perform critical functions in a timely manner.

Often, if there are changes in the client environment, this might necessitate adjustment of various parameter settings associated with z/OS, GDPS, and other products. It is possible that you can miss making these adjustments, which can affect GDPS. The GDPS health checks are intended to detect such situations and avoid incidents where GDPS is unable to perform its job because of a setting that is perhaps less than ideal.

For example, GDPS/PPRC provides facilities for management of the couple data sets (CDS) for the GDPS sysplex. One of the health checks provided by GDPS/PPRC checks that the couple data sets are allocated and defined to GDPS in line with the GDPS preferred practices recommendations.

Similar to z/OS and other products that provide health checks, GDPS health checks are optional. Several optimum values that are checked and the frequency of the checks can be customized to cater to unique client environments and requirements.

There are a few z/OS preferred practices that conflict with GDPS preferred practices. The related z/OS and GDPS health checks result in conflicting exceptions being raised. For such health check items, to avoid conflicting exceptions, z/OS provides the ability to define a coexistence policy where you can indicate which practice is to take precedence; GDPS or z/OS. GDPS provides sample coexistence policy definitions for the GDPS checks that are known to be conflicting with z/OS.

GDPS also provides a convenient interface for managing the health checks using the GDPS panels. You can use it to perform actions such as activate/deactivate or run any selected health check, view the customer overrides in effect for any optimum values, and so on. Figure 3-16 shows a sample of the GDPS Health Checks Information Management panel. In this example, you see that all the health checks are enabled. The status of the last run is also shown, indicating that some were successful and some resulted in raising a medium exception. The exceptions can also be viewed using other options on the panel.

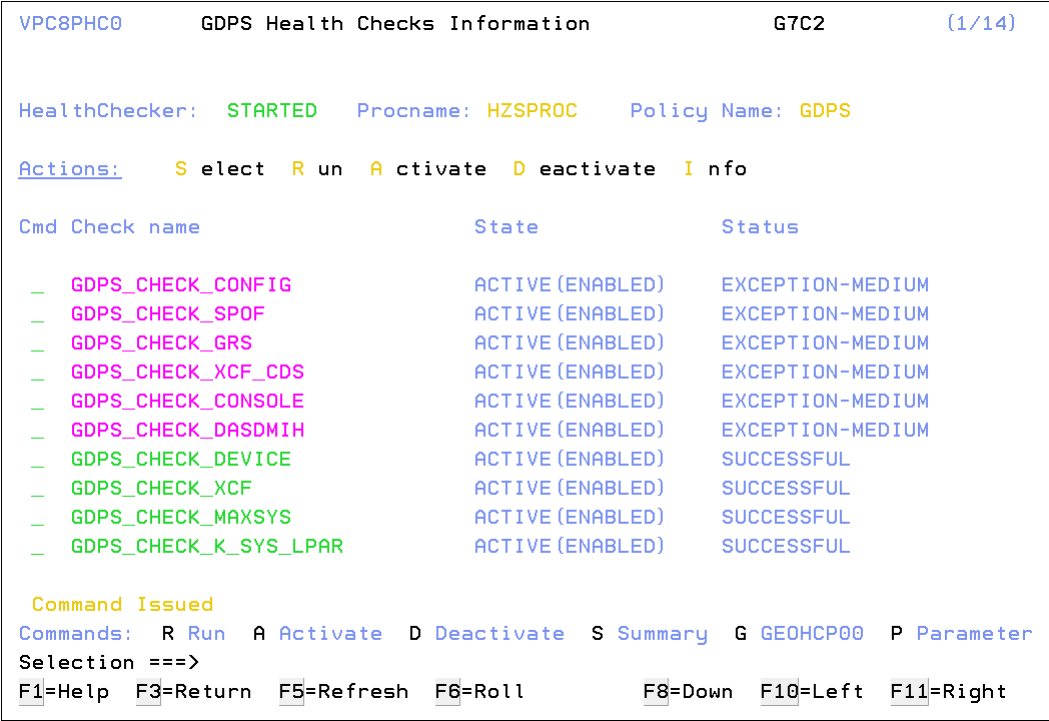


Figure 3-16 GDPS/PPRC Health Checks Information Management panel

3.7 Other facilities related to GDPS

Miscellaneous facilities that GDPS/PPRC provides can assist in various ways, such as reducing the window during which disaster recovery capability is not available.

3.7.1 HyperSwap coexistence

In the following sections we discuss the GDPS enhancements that remove some of the restrictions that had existed regarding HyperSwap coexistence with products such as Softek Transparent Data Migration Facility (TDMF) and IMS Extended Recovery Facility (XRF).

HyperSwap and TDMF coexistence

To minimize disruption to production workloads and service levels, many enterprises use TDMF for storage subsystem migrations and other disk relocation activities. The migration process is transparent to the application, and the data is continuously available for read and write activities throughout the migration process.

However, the HyperSwap function is mutually exclusive with software that moves volumes around by switching UCB pointers. The currently supported versions of TDMF and GDPS allow operational coexistence. With this support, TDMF automatically temporarily disables HyperSwap as part of the disk migration process only during the brief time when it switches UCB pointers.

Manual operator interaction is not required. Without this support, through operator intervention, HyperSwap is disabled for the entire disk migration, including the lengthy data copy phase.

HyperSwap and IMS XRF coexistence

HyperSwap also has a technical requirement that RESERVEs cannot be allowed in the hardware because the status cannot be reliably propagated by z/OS during the HyperSwap to the new primary volumes. For HyperSwap, all RESERVEs must be converted to GRS global enqueue through the GRS RNL lists.

IMS/XRF is a facility by which IMS can provide one active subsystem for transaction processing, and a backup subsystem that is ready to take over the workload. IMS/XRF issues hardware RESERVE commands during takeover processing and these cannot be converted to global enqueues through GRS RNL processing. This coexistence problem has also been resolved so that GDPS is informed before IMS issuing the hardware RESERVE, allowing it to automatically disable HyperSwap. After IMS has finished processing and releases the hardware RESERVE, GDPS is again informed and re-enables HyperSwap.

3.7.2 Reduced impact initial copy and resynchronization

Performing PPRC copy of a large amount of data across a large number of devices while the same devices are used in production by application workloads can potentially affect production I/O service times if such copy operations are performed synchronously. Your disk subsystems and PPRC link capacity are typically sized for steady state update activity, but not for bulk, synchronous replication. Initial copying of disks and resynchronization of disks are examples of bulk copy operations that can affect production if performed synchronously.

There is no need to perform initial copy or resynchronizations using synchronous copy, because the secondary disks cannot be made consistent until all disks in the configuration have reached duplex state.

GDPS supports initial copy and resynchronization using asynchronous PPRC-XD (also known as Global Copy). When GDPS initiates copy operations in asynchronous copy mode, GDPS monitors progress of the copy operation and when the volumes are near full duplex state, GDPS converts the replication from the asynchronous copy mode to synchronous PPRC. Initial copy or resynchronization using PPRC-XD eliminates the performance impact of synchronous mirroring on production workloads.

Without asynchronous copy it might be necessary to defer these operations or reduce the number of volumes being copied at any given time. This would delay the mirror from reaching a duplex state, thus impacting a client's ability to recovery. Use of the XD-mode asynchronous copy allows clients to establish or resynchronize mirroring during periods of high production workload, and can potentially reduce the time during which the configuration is exposed.

This function requires that all disk subsystems in the GDPS configuration support PPRC-XD.

3.7.3 Reserve Storage Pool

Reserve Storage Pool (RSP) is a type of resource introduced with the z/OS Management Facility (z/OSMF) that can simplify the management of defined but unused volumes. GDPS provides support for including RSP volumes in the PPRC configuration that is managed by GDPS. PPRC primary volumes are expected to be online in controlling systems, and GDPS monitoring on the GDPS controlling systems results in an alert being raised for any PPRC primary device that is found to be offline. However, because z/OS does not allow RSP volumes to be brought online to any system, GDPS monitoring recognizes that an offline primary device is an RSP volume and suppresses alerting for these volumes.

3.7.4 Query Services

GDPS maintains configuration information and status information in NetView variables for the various elements of the configuration that it manages. Query Services is a capability that allows client-written NetView REXX programs to query the value for numerous GDPS internal variables. The variables that can be queried pertain to the PPRC configuration, the system and sysplex resources managed by GDPS, and other GDPS facilities such as HyperSwap and GDPS Monitors.

Query Services allows clients to complement GDPS automation with their own automation code. In addition to the Query Services function that is part of the base GDPS product, GDPS provides several samples in the GDPS SAMPLIB library to demonstrate how Query Services can be used in client-written code.

GDPS also makes available to clients a sample tool called the Preserve Mirror Tool (PMT), which facilitates adding new disks to the GDPS PPRC configuration and bringing these disks to duplex. The PMT tool, which is provided in source format, makes extensive use of GDPS Query Services and thereby provides clients with an excellent example of how to write programs to benefit from Query Services.

3.7.5 Concurrent Copy cleanup

The DFSMS Concurrent Copy (CC) function uses a “sidefile” that is kept in the disk subsystem cache to maintain a copy of changed tracks that have not yet been copied. For a PPRCed disk, this sidefile is not mirrored to the secondary subsystem. If a HyperSwap is executed while a Concurrent Copy operation is in progress, the application using Concurrent Copy will fail after the completion of the HyperSwap. GDPS will not allow a planned swap when a Concurrent Copy session exists against your primary PPRC devices. However, unplanned swaps will still be allowed. Therefore, if you plan to use HyperSwap for primary disk subsystem failures (unplanned HyperSwap), try to eliminate any use of Concurrent Copy, because you cannot plan when a failure will occur.

Checking for CC is performed by GDPS immediately before performing a planned HyperSwap. SDF trace entries are generated if one or more CC sessions exist, and the swap command will end with no PPRC device pairs being swapped. You must identify and terminate any CC against the PPRC primary devices before the swap.

When attempting to resynchronize your disks, checking is performed to ensure that the secondary devices do not retain CC status from the time when they were primary devices. These are not supported as PPRC secondary devices. Therefore, GDPS will not attempt to establish a duplex pair with secondary devices if it detects a CC session.

GDPS provides a function to discover and terminate Concurrent Copy sessions that would otherwise cause errors during a resync operation. The function is controlled by a keyword that provides options to disable, to conditionally enable, or to unconditionally enable the cleanup of Concurrent Copy sessions on the target disks. This capability eliminates the manual task of identifying and cleaning up orphaned Concurrent Copy sessions before resynchronizing a suspended PPRC mirror.

3.8 GDPS/PPRC flexible testing and resync protection

Configuring point-in-time copy (FlashCopy) capacity in your PPRC environment provides two significant benefits:

- ▶ You can conduct regular DR drills or other tests using a copy of production data while production continues to run.
- ▶ You can save a consistent, “golden” copy of the PPRC secondary data, which can be used if the primary disk or site is lost during a PPRC resynchronization operation.

FlashCopy and the various options related to FlashCopy are discussed in 2.6, “FlashCopy” on page 38. GDPS/PPRC supports taking a FlashCopy of either the current primary or the current secondary disks. The COPY, NOCOPY, NOCOPY2COPY, and INCREMENTAL options are supported. CONSISTENT FlashCopy is supported in conjunction with COPY, NOCOPY, and INCREMENTAL FlashCopy.

FlashCopy can also be used, for example, to back up data without the need for extended outages to production systems; to provide data for data mining applications; for batch reporting, and so on.

3.8.1 Use of space-efficient FlashCopy volumes

As discussed in “Space-efficient FlashCopy (FlashCopy SE)” on page 40, by using *space-efficient (SE) volumes*, you might be able to lower the amount of physical storage needed and thereby reduce the cost associated with providing a tertiary copy of the data. GDPS provides support allowing space-efficient FlashCopy volumes to be used as FlashCopy target disk volumes. Whether a target device is space-efficient or not is transparent to GDPS; if any of the FlashCopy target devices defined to GDPS are space-efficient volumes, GDPS will simply use them. All GDPS FlashCopy operations with the NOCOPY option, whether through GDPS scripts, panels, or FlashCopies automatically taken by GDPS, can use space-efficient targets.

Space-efficient volumes are ideally suited for FlashCopy targets when used for resync protection. The FlashCopy is taken before the resync and can be withdrawn as soon as the resync operation is complete. As changed tracks are sent to the secondary for resync, the time zero (T0) copy of this data is moved from the secondary to the FlashCopy target device. This means that the total space requirement for the targets is equal to the number of tracks that were out of sync, which typically will be significantly less than a full set of fully provisioned disks.

Another potential use of space-efficient volumes is if you want to use the data for limited disaster recovery testing.

Understanding the characteristics of FlashCopy SE is important to determine whether this method of creating a point-in-time copy will satisfy your business requirements. For example, will it be acceptable to your business if, because of an unexpected workload condition, the repository on the disk subsystem for the space-efficient devices becomes full and your FlashCopy is invalidated so that you are unable to use it? If your business requirements dictate that the copy must always be guaranteed to be usable, space-efficient might not be the best option and you can consider using standard FlashCopy instead.

3.9 GDPS tools for GDPS/PPRC

GDPS/PPRC includes tools that provide function that is complementary to GDPS function. The tools represent the kind of function that all or many clients are likely to develop themselves to complement GDPS. Using these tools eliminates the need for you to develop similar function yourself. The tools are provided in source code format which means that if the tool does not exactly meet your requirements, you can modify the code to suit your needs.

The following tools are available with GDPS/PPRC:

- ▶ **Preserve Mirror Tool (PMT)**

This tool is intended to simplify and automate, to a great extent, the process of bringing new devices to PPRC duplex state. It also adds these devices to your running GDPS environment while keeping the time during which the GDPS managed PPRC mirror is not full duplex (and therefore, not protected by Freeze and HyperSwap) to a minimum.

PMT also provides facilities to aid with migration procedures when using Global Copy (PPRC-XD) and PPRC to migrate data to new disk subsystems.

- ▶ **Configuration Checker Tool (GEOCHECK)**

This checks whether all devices that are online to a GDPS production system are PPRCed under GDPS control, and raises alerts if violations are encountered. It provides identification and facilitates correction for any production devices that are inadvertently left out of the GDPS managed PPRC configuration. Not replicating some devices can prevent HyperSwap and also recovery from catastrophic disk or site failures.

- ▶ **GDPS Console Interface Tool (GCI)**

This facilitates using the MVS system console as an interface for submitting GDPS scripts for execution or for executing individual script commands. This provides operators who do not have access to the NetView interfaces but do have access to the console with some GDPS operational capability.

- ▶ **GDPS Distributed Systems Hardware Management Toolkit**

This provides an interface for GDPS to monitor and control distributed systems' hardware and virtual machines (VMs) by using script procedures that can be integrated into GDPS scripts. This tool provides REXX script templates that show examples of how to monitor/control IBM AIX® HMC; VMware ESX server; IBM BladeCenters; and stand-alone x86 servers with Remote Supervisor Adapter II (RSA) cards. This tool is complementary to the heterogeneous, distributed management capabilities provided by GDPS, such as the Distributed Cluster Management (DCM) and Open LUN management functions.

- ▶ **GDPS Configuration Assistant (GeoAssistant) Tool**

This can help you manage the GDPS/PPRC configuration definition file (GEOPARM file). It allows you to create a graphical view of your GEOPARM that can be easily shared and displayed on a variety of devices (such as workstation, tablet, smartphone, and so on). It can analyze and extract various statistics about your configuration. GeoAssistant can also provide step-by-step guidance for coding the GEOPARM statements when adding new devices to an existing configuration.

3.10 GDPS/PPRC co-operation with GDPS/Active-Active

GDPS/PPRC provides facilities for co-operation with GDPS/Active-Active if GDPS/Active-Active is used to provide workload level protection for selected workloads that are running on the systems that are in the GDPS/PPRC sysplex. See 8.5, “GDPS/Active-Active co-operation with GDPS/PPRC or GDPS/MTMM” on page 264 for details.

3.11 Services component

As you have learned, GDPS touches on much more than simply remote copy. It also includes sysplex, automation, database management and recovery, testing processes, disaster recovery processes, and other areas.

Most installations do not have skills in all these areas readily available. And it is extremely rare to find a team that has this range of skills across many implementations. However, the GDPS/PPRC offering includes exactly that: access to a global team of specialists in all the disciplines you need to ensure a successful GDPS/PPRC implementation.

Specifically, the Services component includes several or all of the following services:

- ▶ Planning to determine availability requirements, configuration recommendations, and implementation and testing plans
- ▶ Installation and necessary customization of NetView and System Automation
- ▶ Remote copy implementation
- ▶ IBM Virtualization Engine TS7700 implementation.
- ▶ GDPS/PPRC automation code installation and policy customization
- ▶ Assistance in defining Recovery Point and Recovery Time objectives
- ▶ Education and training on GDPS/PPRC setup and operations
- ▶ Onsite implementation assistance
- ▶ Project management and support throughout the engagement

The sizing of the Services component of each project is tailored for that project, based on many factors including what automation is already in place, whether remote copy is already in place, whether the two centers are already in place with a multisite sysplex, and so on. This means that the skills provided are tailored to the specific needs of each particular implementation.

3.12 GDPS/PPRC prerequisites

For more information about GDPS/PPRC prerequisites, see the following GDPS web page:

<http://www.ibm.com/systems/z/advantages/gdps/getstarted/gdpspprc.html>

3.13 Comparison of GDPS/PPRC versus other GDPS offerings

So many features and functions are available in the various members of the GDPS family that recalling them all and remembering which offerings support them is sometimes difficult. To position the offerings, Table 3-1 lists the key features and functions and indicates which ones are delivered by the various GDPS offerings.

Table 3-1 Supported features matrix

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/MTMM	GDPS Virtual Appliance	GDPS/XRC	GDPS/GM
Continuous availability	Yes	Yes	Yes	Yes	No	No
Disaster recovery	Yes	Yes	Yes	Yes	Yes	Yes
CA/DR protection against multiple failures	No	No	Yes	No	No	No
Continuous availability for foreign z/OS systems	Yes with z/OS proxy	No	No	No	No	No
Supported distance	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	Virtually unlimited	Virtually unlimited
Zero Suspend FlashCopy support	Yes, using Consistent	Yes, using Consistent for secondary only	Yes, using Consistent	No	Yes, using Zero Suspend FlashCopy	Yes, using CGPause
Reduced impact initial copy/resync	Yes	Yes	Yes	Yes	Not applicable	Not applicable
Tape replication support	Yes	No	No	No	No	No
Production sysplex automation	Yes	No	Yes	Not applicable	No	No

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/MTMM	GDPS Virtual Appliance	GDPS/XRC	GDPS/GM
Span of control	Both sites	Both sites (disk only)	Both sites	Both sites	Recovery site	Disk at both sites; recovery site (CBU or LPARs)
GDPS scripting	Yes	No	Yes	Yes	Yes	Yes
Monitoring, alerting and health checks	Yes	Yes	Yes	Yes (except health checks)	Yes	Yes
Query Services	Yes	Yes	No	No	Yes	Yes
MSS support for added scalability	Yes (secondary in MSS1)	Yes (secondary in MSS1)	Yes (H2 in MSS1, H3 in MSS2)	No	No	Yes (GM FC and primary for MGM in MSS1)
MGM 3-site and 4-site	Yes (all configurations)	Yes (3-site only and non-IR only)	Yes (all configurations)	No	Not applicable	Yes (all configurations)
MzGM	Yes	Yes	Yes (non-IR only)	No	Yes	Not applicable
Open LUN	Yes	Yes	No	No	No	Yes
z/OS equivalent function for Linux for IBM z Systems	Yes	No	Yes (Linux for IBM z Systems running as a z/VM guest only)	Yes (Linux for IBM z Systems running as a z/VM guest only)	Yes	Yes
Heterogeneous support through DCM	Yes (VCS and SA AppMan)	No	No	No	Yes (VCS only)	Yes (VCS and SA AppMan)
z/BX hardware management	Yes	No	No	No	No	No
GDPS GUI	Yes	Yes	No	Yes	No	Yes

3.14 Summary

GDPS/PPRC is a powerful offering that provides disaster recovery, continuous availability, and system/sysplex resource management capabilities. HyperSwap, available with GDPS/PPRC, provides the ability to dynamically swap disks between two sites. The power of automation allows you to test and perfect the actions to be taken, either for planned or unplanned changes, thus minimizing or eliminating the risk of human error.

This offering is one of the offerings in the GDPS family, along with GDPS/MTMM, GDPS/HM, and GDPS Virtual Appliance, that offers the potential of zero data loss and that can achieve the shortest recovery time objective, typically less than one hour following a complete site failure.

It is also one of the only members of the GDPS family, along with GDPS/MTMM and GDPS Virtual Appliance, that is based on hardware replication and that provides the capability to manage the production LPARs. Although GDPS/XRC and GDPS/GM offer LPAR management, their scope for system management includes only the systems in the recovery site, not the production systems running in Site1.

GDPS/PPRC is the only offering in the GDPS family that provides continuous availability for z/OS foreign systems.

In addition to the disaster recovery and planned reconfiguration capabilities, GDPS/PPRC provides a user-friendly interface for monitoring and managing the various elements of the GDPS configuration.



GDPS/PPRC HyperSwap Manager

In this chapter, we discuss the capabilities and prerequisites of the GDPS/PPRC HyperSwap Manager (GDPS/PPRC HM) offering.

GDPS/PPRC HM extends the availability attributes of a Parallel Sysplex to disk subsystems, whether the Parallel Sysplex and disk subsystems are in a single site, or whether the Parallel Sysplex and the primary/secondary disk subsystems span across two sites.

It provides the ability to transparently switch primary disk subsystems with the secondary disk subsystems for either a planned or unplanned disk reconfiguration. It also supports disaster recovery capability across two sites by enabling the creation of a consistent set of secondary disks in case of a disaster or potential disaster.

However, unlike the full GDPS/PPRC offering, GDPS/PPRC HM does not provide any resource management or recovery management capabilities.

The following functions are for protecting data provided by GDPS/PPRC HM:

- ▶ Ensuring the consistency of the secondary data in case there is a disaster or suspected disaster, including the option to also ensure zero data loss
- ▶ Switching to the secondary disk by using HyperSwap
- ▶ Managing the remote copy configuration for z Systems and other platform data

Because GDPS/PPRC HM is a subset of the GDPS/PPRC offering, you might want to review the comparison presented in Table 4-1 on page 134 if you have already read Chapter 3, “GDPS/PPRC” on page 53.

4.1 Introduction to GDPS/PPRC HM

GDPS/PPRC HM provides a subset of GDPS/PPRC capability with the emphasis being more on the remote copy and disk management aspects. At its most basic, GDPS/PPRC HM extends Parallel Sysplex availability to disk subsystems by delivering the HyperSwap capability to mask disk outages caused by planned disk maintenance or unplanned disk failures. It also provides monitoring and management of the data replication environment, including the freeze capability.

In the multisite environment, GDPS/PPRC HM provides an entry-level disaster recovery offering. Because GDPS/PPRC HM does not include the systems management and automation capabilities of GDPS/PPRC, it cannot provide in and of itself the short RTO that is achievable with GDPS/PPRC. However, GDPS/PPRC HM does provide a cost-effective route into full GDPS/PPRC at a later time if your recovery time objectives change.

4.1.1 Protecting data integrity and data availability with GDPS/PPRC HM

In 2.2, “Data consistency” on page 17, we point out that data integrity across primary and secondary volumes of data is essential to perform a database restart and accomplish an RTO of less than an hour. This section provides details about how GDPS automation in GDPS/PPRC HM provides both data consistency if there are mirroring problems and data availability if there are disk problems.

Two types of disk problems trigger a GDPS automated reaction:

- ▶ PPRC Mirroring problems (Freeze triggers)

There is no problem with writing to the primary disk subsystem, but there is a problem mirroring the data to the secondary disk subsystem. This is described in “GDPS Freeze function for mirroring failures” on page 104.”

- ▶ Primary disk problems (HyperSwap triggers)

There is a problem writing to the primary disk: Either a hard failure, or the disk subsystem is not accessible or not responsive. This is described in “GDPS HyperSwap function” on page 108.

GDPS Freeze function for mirroring failures

GDPS uses automation, keyed off events or messages, to stop all mirroring when a remote copy failure occurs. In particular, the GDPS automation uses the IBM PPRC Freeze and Run architecture, which has been implemented as part of Metro Mirror on IBM disk subsystems and also by other enterprise disk vendors. In this way, if the disk hardware supports the Freeze/Run architecture, GDPS can ensure consistency across all data in the sysplex (consistency group) regardless of disk hardware type. This preferred approach differs from proprietary hardware approaches that work only for one type of disk hardware. For an introduction to data consistency with synchronous disk mirroring, see “PPRC data consistency” on page 24.

When a mirroring failure occurs, this problem is classified as a Freeze trigger and GDPS stops activity across *all* disk subsystems at the time the initial failure is detected, thus ensuring that the dependant write consistency of the remote disks is maintained. This is what happens when a GDPS performs a Freeze:

- ▶ Remote copy is suspended for all device pairs in the configuration.
- ▶ While the suspend command is being processed for each LSS, each device goes into a long busy state. When the suspend completes for each device, z/OS marks the device unit control block (UCB) in all connected operating systems to indicate an Extended Long Busy (ELB) state.
- ▶ No I/Os can be issued to the affected devices until the ELB is thawed with the PPRC action or until it times out (the consistency group timer setting commonly defaults to 120 seconds, although for most configurations a longer ELB is recommended).
- ▶ All paths between the PPRCed disks are removed, preventing further I/O to the secondary disks if PPRC is accidentally restarted.

Because no I/Os are processed for a remote-copied volume during the ELB, dependent write logic ensures the consistency of the remote disks. GDPS performs Freeze for all LSS pairs that contain GDPS managed mirrored devices.

Important: Because of the dependent write logic, it is not necessary for all LSSs to be frozen at the same instant. In a large configuration with many thousands of remote copy pairs, it is not unusual to see short gaps between the times when the Freeze command is issued to each disk subsystem. However, because of the ELB, such gaps are not a problem.

After GDPS performs the Freeze and the consistency of the remote disks is protected, what GDPS does depends on the client's PPRC Failure policy (also known as Freeze policy). The policy, as described in "Freeze policy (PPRC Failure policy) options" on page 106, tells GDPS to take one of these three possible actions:

- ▶ Perform a **Run** action against all LSSs. This will remove the ELB and allow production systems to continue using these devices. The devices will be in remote copy-suspended mode, meaning that any further writes to these devices are no longer being mirrored. However, the changes are being tracked by the hardware so that, later, only the changed data will be resynchronized to the secondary disks. See "Freeze and Go" on page 107 for more detail about this policy option.
- ▶ System-reset all production systems. This ensures that no more updates can occur to the primary disks, because such updates would not be mirrored and it would not be possible to achieve RPO zero (zero data loss) if a failure occurs (or if the original trigger was an indication of a catastrophic failure). See "Freeze and Stop" on page 106 for more detail about this option.
- ▶ Try to determine if the cause of the PPRC suspension event was a permanent or temporary problem with any of the secondary disk subsystems in the GDPS configuration. If GDPS can determine that the PPRC failure was caused by the secondary disk subsystem, this would not be a potential indicator of a disaster in the primary site. In this case, GDPS would perform a Run action and allow production to continue using the suspended primary devices. If, however, the cause cannot be determined to be a secondary disk problem, GDPS would reset all systems, guaranteeing zero data loss. See "Freeze and Stop conditionally" on page 107 for further details.

GDPS/PPRC HM uses a combination of storage subsystem and sysplex triggers to automatically secure, at the first indication of a potential disaster, a data-consistent secondary site copy of your data using the Freeze function. In this way, the secondary copy of the data is preserved in a consistent state, perhaps even before production applications are aware of any issues. Ensuring the data consistency of the secondary copy ensures that a normal system restart can be performed, instead of having to perform DBMS forward recovery actions. This is an essential design element of GDPS to minimize the time to recover the critical workloads if there is a disaster in the primary site.

You will appreciate why such a process must be automated. When a device suspends, there is simply not enough time to launch a manual investigation process. The entire mirror must be frozen by stopping further I/O to it, and then letting production run with mirroring temporarily suspended, or stopping all systems to guarantee zero data loss based on the policy.

In summary, freeze is triggered as a result of a PPRC suspension event for any primary disk in the GDPS configuration, that is, at the first sign of a duplex mirror that is going out of duplex state. When a device suspends, all attached systems are sent a State Change Interrupt (SCI). A message is issued in all of those systems and then each system must issue multiple I/Os to investigate the reason for the suspension event.

When GDPS performs a freeze, all primary devices in the PPRC configuration suspend. This will result in significant SCI traffic and many messages in all of the systems. GDPS, in conjunction with z/OS and microcode on the DS8000 disk subsystems, supports reporting suspensions in a summary message per LSS instead of at the individual device level. When compared to reporting suspensions on a per device basis, the Summary Event Notification for PPRC Suspends (PPRCSUM) dramatically reduces the message traffic and extraneous processing associated with PPRC suspension events and freeze processing.

For more information about the implementation of PPRC and IBM Metro Mirror, see *IBM DS8870 Copy Services for IBM z Systems*, SG24-6787.

Freeze policy (PPRC Failure policy) options

As described, when a mirroring failure is detected, GDPS automatically and unconditionally performs a Freeze to secure a consistent set of secondary volumes in case the mirroring failure could be the first indication of a site failure. Because the primary disks are in an Extended Long Busy state as a result of the freeze and the production systems are locked out, GDPS must take some action. There is no time to interact with the operator on an event-by-event basis. The action must be taken immediately and is determined by a customer policy setting, namely the PPRC Failure policy option (also known as the Freeze policy option). GDPS will use this same policy setting after every Freeze event to determine what its next action should be. The options are listed here:

- ▶ PPRCFAILURE=STOP (Freeze and STOP)
GDPS resets production systems while I/O is suspended.
- ▶ PPRCFAILURE=GO (Freeze and Go)
GDPS allows production systems to continue operation after mirroring is suspended.
- ▶ PPRCFAILURE=COND (Freeze and Stop, conditionally)
GDPS tries to determine if a secondary disk caused the mirroring failure. If so, GDPS performs a Go. If not, GDPS performs a Stop.

Freeze and Stop

If your RPO is zero (that is, you cannot tolerate any data loss), you must select the Freeze and Stop policy to reset all production systems. With this setting, you can be assured that no updates are made to the primary volumes after the Freeze because all systems that can

update the primary volumes are reset. You can choose to restart them when you see fit. For example, if this was a false freeze (that is, a false alarm), then you can quickly resynchronize the mirror and restart the systems only after the mirror is duplex.

If you are using duplexed coupling facility (CF) structures along with a Freeze and Stop policy, it might seem that you are guaranteed to be able to use the duplexed instance of your structures if you have to recover and restart your workload with the frozen secondary copy of your disks. However, this is not always the case. There can be rolling disaster scenarios where before, following, or during the freeze event, there is an interruption (perhaps a failure of CF duplexing links) that forces CFRM to drop out of duplexing.

There is no guarantee that the structure instance in the surviving site is the one that is kept. It is possible that CFRM keeps the instance in the site that is about to totally fail. In this case, there will not be an instance of the structure in the site that survives the failure.

To summarize, with a Freeze and Stop policy, if there is a surviving, accessible instance of application-related CF structures, that instance will be consistent with the frozen secondary disks. However, depending on the circumstances of the failure, even with structures duplexed across two sites you are not 100% guaranteed to have a surviving, accessible instance of the application structures and therefore you must have the procedures in place to restart your workloads without the structures.

A Stop policy ensures no data loss. However, if this was a false Freeze event, that is, it was a transient failure that did not necessitate recovering using the frozen disks, it will stop the systems unnecessarily.

Freeze and Go

If you can accept an RPO that is not *necessarily* zero, you might decide to let the production systems *continue* operation after the secondary volumes have been protected by the Freeze. In this case you would use a Freeze and Go policy. With this policy you avoid an unnecessary outage for a false freeze event, that is, if the trigger is simply a transient event.

However, if the trigger turns out to be the first sign of an actual disaster, you might continue operating for an amount of time before all systems actually fail. Any updates made to the primary volumes during this time will not have been replicated to the secondary disk, and therefore are lost. In addition, because the CF structures were updated after the secondary disks were frozen, the CF structure content is not consistent with the secondary disks. Therefore, the CF structures in either site cannot be used to restart workloads and log-based restart must be used when restarting applications.

This is not full forward recovery. It is forward recovery of any data such as DB2 group buffer pools that might have existed in a CF but might not have been written to disk yet. This results in prolonged recovery times. The extent of this elongation will depend on how much such data existed in the CFs at that time. With a Freeze and Go policy you may consider tuning applications such as DB2 to harden such data on disk more frequently than otherwise.

Freeze and Go is a high-availability option that avoids production outage for false Freeze events. However, it carries a potential for data loss.

Freeze and Stop conditionally

Field experience has shown that most occurrences of freeze triggers are not necessarily the start of a rolling disaster, but are “false freeze” events, which do not necessitate recovery on the secondary disk. Examples of such events include connectivity problems to the secondary disks and secondary disk subsystem failure conditions.

With a COND (conditional) specification, the action that GDPS takes after it performs the Freeze is conditional. GDPS tries to determine if the mirroring problem was as a result of a permanent or temporary secondary disk subsystem problem:

- ▶ If GDPS can determine that the freeze was triggered as a result of a secondary disk subsystem problem, then GDPS performs a Go. That is, it allows production systems to continue to run using the primary disks. However, updates will not be mirrored until the secondary disk can be fixed and PPRC can be resynchronized.
- ▶ If GDPS cannot ascertain that the cause of the freeze was a secondary disk subsystem, then GDPS deduces that this could be the beginning of a rolling disaster in the primary site. Therefore, it performs a Stop, resetting all of the production systems to ensure zero data loss. GDPS cannot always detect that a particular freeze trigger was caused by a secondary disk, and some freeze events that are truly caused by a secondary disk could still result in a Stop.

For GDPS to determine whether a freeze trigger could have been caused by the secondary disk subsystem, the IBM DS8000 disk subsystems provide a special query capability known as the *Query Storage Controller Status* microcode function. If all disk subsystems in the GDPS managed configuration support this feature, GDPS uses this special function to query the secondary disk subsystems in the configuration to understand the state of the secondaries and whether one of those secondaries could have caused the freeze. If you use the COND policy setting but all disks your configuration do not support this function, then GDPS cannot query the secondary disk subsystems, and the resulting action will be a Stop.

This option could provide a useful compromise wherein you can minimize the chance that systems would be stopped for a false freeze event, and increase the chance of achieving zero data loss for a real disaster event.

PPRC Failure policy selection considerations

As described, the PPRC Failure policy option specification directly relates to Recovery Time and recovery point objectives, which are business objectives. Therefore, the policy option selection is really a business decision, rather than an IT decision. If data associated with your transactions is of high value, it might be more important to ensure that no data associated with your transactions is ever lost, so you might decide on a Freeze and Stop policy. If you have huge volumes of relatively low value transactions, you might be willing to risk some lost data in return for avoiding unneeded outages with a Freeze and Go policy. The Freeze and Stop Conditional policy attempts to minimize the chance of unnecessary outages and the chance of data loss; however, there is still a risk of either, however small.

Most installations start with a Freeze and Go policy. Companies that have an RPO of zero typically then move on and implement a Freeze and Stop Conditional or Freeze and Stop policy after the implementation is proven to be stable.

GDPS HyperSwap function

If there is a problem writing or accessing the *primary* disk because of a failing, failed, or non-responsive primary disk, there is a need to swap from the primary disks to the secondary disks.

GDPS/PPRC HM delivers a powerful function known as *HyperSwap*, which provides the ability to swap from using the primary devices in a mirrored configuration to using what had been the secondary devices, apparent to the production systems and applications that are using these devices. Before the availability of HyperSwap, a transparent disk swap was not possible. All systems using the primary disk would have been shut down (or could have failed, depending on the nature and scope of the failure) and would have been re-IPLed using the secondary disks. Disk failures were often a single point of failure for the entire sysplex.

With HyperSwap, such a switch can be accomplished without IPL and with simply a brief hold on application I/O. The HyperSwap function is completely controlled by automation, thus allowing all aspects of the disk configuration switch to be controlled through GDPS.

HyperSwap can be invoked in two ways:

- ▶ Planned HyperSwap

A planned HyperSwap is invoked by operator action using GDPS facilities. One example of a planned HyperSwap is where a HyperSwap is initiated in advance of planned disruptive maintenance to a disk subsystem.

- ▶ Unplanned HyperSwap

An unplanned HyperSwap is invoked automatically by GDPS, triggered by events that indicate a primary disk problem.

Primary disk problems can be detected as a direct result of an I/O operation to a specific device that fails because of a reason that indicates a primary disk problem such as these:

- No paths available to the device
- Permanent error
- I/O timeout

In addition to a disk problem being detected as a result of an I/O operation, it is also possible for a primary disk subsystem to proactively report that it is experiencing an acute problem. The IBM DS8000 models have a special microcode function known as the *Storage Controller Health Message Alert* capability. Problems of different severity are reported by disk subsystems that support this capability. Those problems classified as acute are also treated as HyperSwap triggers. After systems are swapped to use the secondary disks, the disk subsystem and operating system can try to perform recovery actions on the former primary without impacting the applications using those disks.

Planned and unplanned HyperSwap have requirements in terms of the physical configuration, such as having a symmetrically configured configuration. If a client's environment meets these requirements, there is no special enablement required to perform planned swaps. Unplanned swaps are not enabled by default and must be enabled explicitly as a policy option. This is described in detail in "HyperSwap (Primary Failure) policy options" on page 111.

When a swap is initiated, GDPS will always validate various conditions to ensure that it is safe to swap. For example, if the mirror is not fully duplex, that is, not all volume pairs are in a duplex state, then a swap cannot be performed. The way that GDPS reacts to such conditions will change depending on the condition detected and whether the swap is a planned or unplanned swap.

Assuming that there are no show-stoppers and the swap proceeds, for both planned and unplanned HyperSwap, the systems that are using the primary volumes will experience a temporary pause in I/O processing. GDPS will block I/O both at the channel subsystem level by performing a Freeze, which will result in all disks going into Extended Long Busy, and also in all systems I/O being quiesced at the operating system (UCB) level. This is to ensure that no systems will use the disks until the switch is complete. During this time when I/O is paused:

- ▶ The PPRC configuration is *physically switched*. This involves physically changing the secondary disk status to primary. Secondary disks are protected and cannot be used by applications. Changing their status to primary allows them to come online to systems and be used.

- ▶ The disks will be *logically switched* in each of the systems in the GDPS configuration. This involves switching the internal pointers in the operating system control blocks (UCBs). The operating system will point to the former secondary devices instead of the current primary devices.
- ▶ For planned swaps, the mirroring direction can be reversed (optional).
- ▶ Finally, the systems resume operation using the new, swapped-to primary devices even though applications are not aware of the fact that different devices are now being used.

This brief pause during which systems are locked out of performing I/O is known as the User Impact Time. In benchmark measurements at IBM using currently supported releases of GDPS and IBM DS8000 disk subsystems, the User Impact Time to swap 10,000 pairs across 16 systems during an unplanned HyperSwap was less than 10 seconds. Most implementations are actually much smaller than this and typical impact times using the most current storage and server hardware are measured in seconds. Although results will depend on your configuration, these numbers give you a high-level idea of what to expect.

GDPS/PPRC HM HyperSwaps all devices in the managed configuration. Just as the Freeze function applies to the entire consistency group, HyperSwap is similarly for the entire consistency group. For example, if a single mirrored volume fails and HyperSwap is invoked, processing is swapped to the secondary copy of *all* mirrored volumes in the configuration, including those in other, unaffected, subsystems. This is because, to maintain disaster readiness, all primary volumes *must* be in the same site. If HyperSwap were to swap only the failed LSS, you would then have several primaries in one site, and the remainder in the other site. This would also make for a complex environment to operate and administer I/O configurations.

Why is this necessary? Consider the configuration shown in Figure 4-1 on page 111. This is what might happen if only the volumes of a single LSS or subsystem were hyperswapped without swapping the whole consistency group. What happens if a remote copy failure occurs at 15:00? The secondary disks in both sites are frozen at 15:00 and the primary disks (in the case of a Freeze and Go policy) continue to receive updates.

Now assume that either site is hit by another failure at 15:10. What do you have? Half the disks are now at 15:00 and the other half are at 15:10 and *neither* site has consistent data. In other words, the volumes are of virtually no value to you.

If you had *all* of the secondaries in Site2, all volumes in that site would be consistent. If you had the disaster at 15:10, you would lose 10 minutes of data with the Go policy, but at least all of the data in Site2 would be usable. Using a Freeze and Stop policy is no better for this partial swap scenario because, with a mix of primary disks in either site, you must maintain I/O configurations that can match *every* possible combination simply to IPL any systems.

More likely, you must first restore mirroring across the entire consistency group before recovering systems, and this is not really practical. Therefore, for disaster recovery readiness, it is necessary that *all* of the primary volumes are in one site and *all* of the secondaries in the other site.

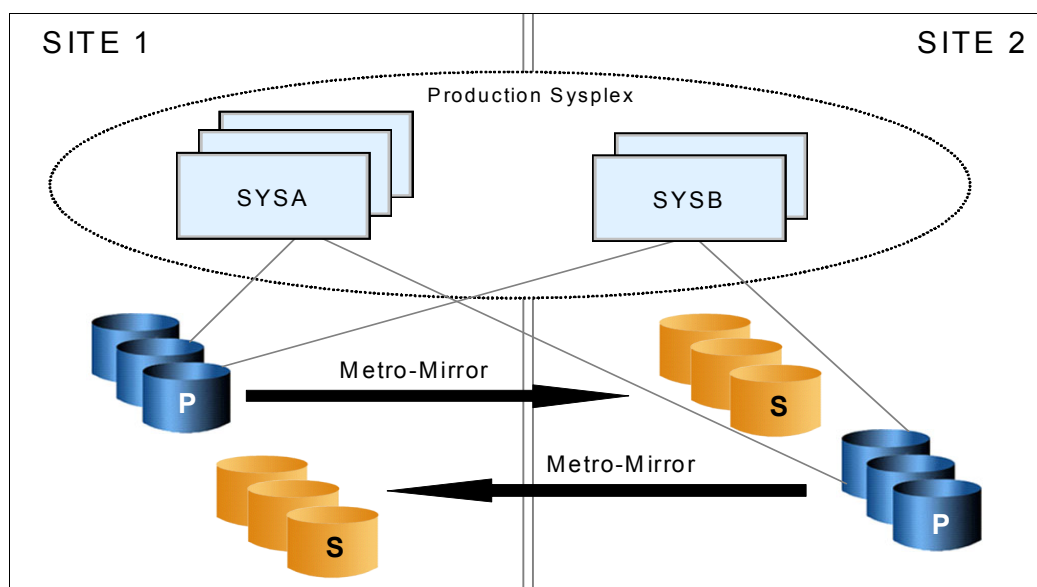


Figure 4-1 Unworkable Metro Mirror disk configuration

HyperSwap with less than full channel bandwidth

You may consider enabling unplanned HyperSwap even if you do not have sufficient cross-site channel bandwidth to sustain the full production workload for normal operations. Assuming that a disk failure is likely to cause an outage and you will need to switch to using disk in the other site, the unplanned HyperSwap might at least present you with the opportunity to perform an orderly shutdown of your systems first. Shutting down your systems cleanly avoids the complications and restart time elongation associated with a crash-restart of application subsystems.

HyperSwap (Primary Failure) policy options

Clients might prefer not to immediately enable their environment for unplanned HyperSwap when they first implement GDPS. For this reason, HyperSwap is not enabled by default. However, we strongly suggest that all GDPS/PPRC HM clients enable their environment for unplanned HyperSwap.

An unplanned swap is the action that makes most sense when a primary disk problem is encountered. However, other policy specifications that will not result in a swap are available. When GDPS detects a primary disk problem trigger, the first thing it will do will be a Freeze (the same as it performs when a mirroring problem trigger is detected).

GDPS then uses the selected Primary Failure policy option to determine what action it will take next:

► PRIMARYFAILURE=GO

No swap is performed. The action GDPS takes is the same as for a freeze event with policy option PPRCFAILURE=GO. A Run action is performed, which allows systems to continue using the original primary disks. PPRC is suspended and therefore updates are not replicated to the secondary. However, depending on the scope of the primary disk problem, it might be that all or some production workloads simply cannot run or cannot sustain required service levels. Such a situation might necessitate restarting the systems on the secondary disks. Because of the freeze, the secondary disks are in a consistent state and can be used for restart. However, any transactions that ran after the Go action will be lost.

► PRIMARYFAILURE=STOP

No swap is performed. The action GDPS takes is the same as for a freeze event with policy option PPRCFAILURE=STOP. GDPS resets all the production systems. This ensures that no further I/O occurs. After performing situation analysis, if it is determined that this was not a transient issue and that the secondaries should be used to re-IPL the systems, no data will be lost.

► PRIMARYFAILURE=SWAP,*swap_disabled_action*

The first parameter, SWAP, indicates that after performing the Freeze, GDPS will proceed with an unplanned HyperSwap. When the swap is complete, the systems will be running on the new, swapped-to primary disks (the former secondaries). PPRC will be in a suspended state; because the primary disks are known to be in a problematic state, there is no attempt to reverse mirroring. After the problem with the primary disks is fixed, you can instruct GDPS to resynchronize PPRC from the current primaries to the former ones (which are now considered to be secondaries).

The second part of this policy, *swap_disabled_action*, indicates what GDPS should do if HyperSwap had been temporarily disabled by operator action at the time the trigger was encountered. Effectively, an operator action has instructed GDPS not to perform a HyperSwap, even if there is a swap trigger. GDPS has already done a freeze. So the second part of the policy says what GDPS should do next.

The following options are available for the second parameter, which comes into play only if HyperSwap was disabled by the operator (remember, disk is already frozen):

GO This is the same action as GDPS would have performed if the policy option had been specified as PRIMARYFAILURE=GO.

STOP This is the same action as GDPS would have performed if the policy option had been specified as PRIMARYFAILURE=STOP.

Primary Failure policy specification considerations

As indicated previously, the action that best serves RTO/RPO objectives when there is a primary disk problem is to perform an unplanned HyperSwap. Therefore, the SWAP policy option is the recommended policy option.

For the Stop or Go choice, either as the second part of the SWAP specification or if you will not be using SWAP, similar considerations apply as discussed for the PPRC Failure policy options to Stop or Go. Go carries the risk of data loss if it becomes necessary to abandon the primary disk and restart systems on the secondary. Stop carries the risk of taking an unnecessary outage if the problem was transient.

The key difference is that with a mirroring failure, the primary disks are not broken. When you allow the systems to continue to run on the primary disk with the Go option, other than a disaster which is low probability, the systems are likely to run with no problems. With a primary disk problem, with the Go option, you are allowing the systems to continue running on what are known to be disks that experienced a problem just seconds ago. If this was a serious problem with widespread impact, such as an entire disk subsystem failure, the applications are going to experience severe problems. Some transactions might continue to commit data to those disks that are not broken. Other transactions might be failing or experiencing serious service time issues.

Finally, if there is a decision to restart systems on the secondaries because the primary disks are simply not able to support the workloads, there will be data loss. The probability that a primary disk problem is a real problem that will necessitate restart on the secondary disks is much higher when compared to a mirroring problem. A Go specification in the Primary Failure policy increases your overall risk for data loss.

If the primary failure was of a transient nature, a Stop specification results in an unnecessary outage. However, with primary disk problems it is likely that the problem could necessitate restart on the secondary disks. Therefore, a Stop specification in the Primary Failure policy avoids data loss and facilitates faster restart.

The considerations relating to CF structures with a PRIMARYFAILURE event are similar to a PPRCFAILURE event. If there is an actual swap, the systems continue to run and continue to use the same structures as they did before the swap; the swap is transparent. With a Go action, you continue to update the CF structures along with the primary disks after the Go. If you need to abandon the primary disks and restart on the secondary, the structures are inconsistent with the secondary disks and are not usable for restart purposes. This will prolong the restart and, therefore, your recovery time. With Stop, if you decide to restart the systems using the secondary disks, there is no consistency issue with the CF structures because no further updates occurred on either set of disks after the trigger was detected.

GDPS use of DS8000 functions

GDPS strives to use (when it makes sense) enhancements to the IBM DS8000 disk technologies. In this section we provide information about the key DS8000 technologies that GDPS supports and uses.

Failover/Failback support

When a primary disk failure occurs and the disks are switched to the secondary devices, PPRC Failover/Failback (FO/FB) support eliminates the need to do a full copy when reestablishing replication in the opposite direction. Because the primary and secondary volumes are often in the same state when the freeze occurred, the only differences between the volumes are the updates that occur to the secondary devices after the switch. Failover processing sets the secondary devices to primary suspended status and starts change recording for any subsequent changes made. When the mirror is reestablished with failback processing, the original primary devices become secondary devices and a resynchronization of changed tracks takes place.

GDPS/PPRC HM requires PPRC FO/FB capability to be available on all disk subsystems in the managed configuration.

PPRC Extended Distance (PPRC-XD)

PPRC-XD (also known as Global Copy) is an asynchronous form of the PPRC copy technology. GDPS uses PPRC-XD rather than synchronous PPRC in order to reduce the performance impact of certain remote copy operations that potentially involve a large amount of data. See 4.5.2, “GDPS/PPRC HM reduced impact initial copy and resynchronization” on page 129 for details.

Storage Controller Health Message Alert

This facilitates triggering an unplanned HyperSwap proactively when the disk subsystem reports an acute problem that requires extended recovery time. See “GDPS HyperSwap function” on page 108 for more information about unplanned HyperSwap triggers.

PPRCS Summary Event Messages

GDPS supports the DS8000 PPRC Summary Event Messages (PPRCSUM) function, which is aimed at reducing the message traffic and the processing of these messages for Freeze events. This is described in “GDPS Freeze function for mirroring failures” on page 104.

Soft Fence

Soft Fence provides the capability to block access to selected devices. As discussed in “Protecting secondary disks from accidental update” on page 115, GDPS uses Soft Fence to avoid write activity on disks that are exposed to accidental update in certain scenarios.

On-demand dump (also known as non-disruptive statesave)

When problems occur with disk subsystems, such as those that result in an unplanned HyperSwap, mirroring suspension or performance issues can happen. The lack of diagnostic information can be associated with any of the things that can happen. This function is designed to reduce the likelihood of missing diagnostic information.

Taking a full statesave can lead to temporary disruption to the host I/O and is often disliked by clients for this reason. The on-demand dump (ODD) capability of the disk subsystem facilitates taking a non-disruptive statesave (NDSS) at the time that such an event occurs. The microcode does this automatically for certain events, such as taking a dump of the primary disk subsystem that triggers a PPRC freeze event, and also allows an NDSS to be requested. This enables first failure data capture (FFDC) and thus ensures that diagnostic data is available to aid problem determination. Be aware that not all information that is contained in a full statesave is contained in an NDSS. Therefore, there might still be failure situations where a full statesave is requested by the support organization.

GDPS provides support for taking an NDSS using the remote copy panels (or GDPS GUI). In addition to this support, GDPS autonomically takes an NDSS if there is an unplanned freeze or HyperSwap event.

Query Host Access

When a PPRC disk pair is being established, the device that is the target (secondary) must not be used by any system. The same is true when establishing a FlashCopy relationship to a target device. If the target is in use, the establishment of the PPRC or FlashCopy relationship fails. When such failures occur, it can be a tedious task to identify which system is holding up the operation.

The Query Host Access disk function provides the means to query and identify what system is using a selected device. GDPS uses this capability and adds usability in several ways:

- ▶ Query Host Access identifies the LPAR that is using the selected device through the CPC serial number and LPAR number. It is still a tedious job for operations staff to translate this information to a system or CPC and LPAR name. GDPS does this translation and presents the operator with more readily usable information, avoiding this additional translation effort.
- ▶ Whenever GDPS is requested to perform a PPRC or FlashCopy establish operation, GDPS first performs Query Host Access to see if the operation is expected to succeed or fail because of one or more target devices being in use. It alerts operations if the operation is expected to fail, and identifies the target devices in use and the LPARs holding them.
- ▶ GDPS continually monitors the target devices defined in the GDPS configuration and alerts operations that target devices are in use when they should not be. This allows operations to fix the reported problems in a timely manner.
- ▶ GDPS provides the ability for the operator to perform ad hoc Query Host Access to any selected device using the GDPS panels (or GDPS GUI).

Easy Tier Heat Map Transfer

IBM DS8000 Easy Tier optimizes data placement (placement of logical volumes) across the various physical tiers of storage within a disk subsystem in order to optimize application performance. The placement decisions are based on learning the data access patterns and can be changed dynamically and transparently to the applications using this data.

PPRC mirrors the data from the primary to the secondary disk subsystem however, the Easy Tier learning information is not included in PPRC scope. The secondary disk subsystems are optimized according to the workload on these subsystems, which differs from the activity on the primary (there is only write workload on the secondary; there is read/write activity on the primary). As a result of this difference, during a disk switch or disk recovery, the secondary disks that you switch to are likely to display different performance characteristics compared to the former primary.

Easy Tier Heat Map Transfer is the DS8000 capability to transfer the Easy Tier learning from a PPRC primary to the secondary disk subsystem so that the secondary disk subsystem can also be optimized based on this learning and will have similar performance characteristics if it is promoted to become the primary.

GDPS integrates support for Heat Map Transfer. The appropriate Heat Map Transfer actions (such as start or stop of the processing and reversing transfer direction) are incorporated into the GDPS managed processes. For example, if PPRC is temporarily suspended by GDPS for a planned or unplanned secondary disk outage, Heat Map Transfer is also suspended; or if PPRC direction is reversed as a result of a HyperSwap, Heat Map Transfer direction is also reversed.

Protecting secondary disks from accidental update

A system cannot be IPLed using a disk that is physically a PPRC secondary disk because PPRC secondary disks cannot be brought online to any systems. However, a disk can be secondary from a GDPS (and application use) perspective but physically have a simplex or primary status from a PPRC perspective.

For planned and unplanned HyperSwap, and a disk recovery, GDPS changes former secondary disks to primary or simplex state. However, these actions do not modify the state of the former primary devices, which remain in the primary state. Therefore, the former primary devices remain accessible and usable even though they are considered to be the secondary disks from a GDPS perspective. This makes it possible to accidentally update or IPL from the wrong set of disks. Accidentally using the wrong set of disks can result in a potential data integrity or data loss problem.

GDPS/PPRC HM provides IPL protection early in the IPL process. During initialization of GDPS, if GDPS detects that the system coming up has just been IPLed using the wrong set of disks, GDPS will quiesce that system, preventing any data integrity problems that could be experienced had the applications been started.

GDPS also uses an IBM DS8000 disk subsystem capability, which is called Soft Fence, for configurations where the disks support this function. Soft Fence provides the means to fence, that is to block access to a selected device. GDPS uses Soft Fence when appropriate to fence devices that might otherwise be exposed to accidental update.

4.1.2 Protecting distributed (FB) data

Terminology: The introduction of Open LUN support in GDPS has caused several changes in the terminology we use when referring to disks in this book, as explained here.

- **z Systems or CKD disks**

GDPS can manage disks used by z Systems, including z/VM, VSE, or Linux on z Systems disks. All these disks are formatted as Count-Key-Data (CKD) disks, the traditional mainframe format.

In most places, we refer to the disks used by a system running on the mainframe as *z Systems disks* or *CKD disks*. Both terms are used interchangeably.

- **Open LUN or FB disks**

Disks that are used by systems other than those running on z Systems are traditionally formatted as Fixed Block (FB). In this book, we generally use the term *Open LUN disks* or *FB disks*. These terms are used interchangeably.

GDPS/PPRC HM can manage the mirroring of FB devices used by non-mainframe operating systems; which includes SCSI disks written by Linux on z Systems. The FB devices can be part of the same consistency group as the mainframe CKD devices, or they can be managed separately in their own consistency group.

For more information about Open LUN management, see 10.1, “Open LUN Management function” on page 296.

4.1.3 Protecting other CKD data

Systems that are fully managed by GDPS are known as *GDPS managed systems* or *GDPS systems*. These are the z/OS systems in the GDPS sysplex.

GDPS/PPRC HM can also manage the disk mirroring of CKD disks used by systems outside the sysplex: Other z/OS systems, Linux on z Systems, VM, and VSE systems that are not running any GDPS/PPRC or xDR automation. These are known as “foreign systems.”

Because GDPS manages PPRC for the disks used by these systems, these disks must be attached to the GDPS controlling systems. With this setup, GDPS is able to capture mirroring problems and will perform a freeze. All GDPS managed disks belonging to the GDPS systems and these foreign systems are frozen together, regardless of whether the mirroring problem is encountered on the GDPS systems' disks or the foreign systems' disks.

GDPS/PPRC HM is not able to directly communicate with these foreign systems. For this reason, GDPS automation will not be aware of certain other conditions such as a primary disk problem that is detected by these systems. Because GDPS will not be aware of such conditions that would have otherwise driven autonomic actions such as HyperSwap, GDPS cannot react to these events.

If an unplanned HyperSwap occurs (because it triggered on a GDPS managed system), the foreign systems cannot and will not swap to using the secondaries. A setup is prescribed to set a long Extended Long Busy timeout (the maximum is 18 hours) for these systems so that when the GDPS managed systems swap, these systems hang. The ELB prevents these systems from continuing to use the former primary devices. You can then use GDPS automation facilities to reset these systems and re-IPL them using the swapped-to primary disks.

4.2 GDPS/PPRC HM configurations

A basic GDPS/PPRC HM configuration consists of at least one production system, at least one controlling system, primary disks, and secondary disks. The entire configuration can be in either a single site to provide protection from disk outages with HyperSwap, or it can be spread across two data centers within metropolitan distances as the foundation for a disaster recovery solution. The actual configuration depends on your business and availability requirements.

4.2.1 Controlling system

Why does a GDPS/PPRC HM configuration need a controlling system? At first, you might think this is an additional infrastructure overhead. However, when you have an unplanned outage that affects production systems or the disk subsystems, it is crucial to have a system such as the controlling system that can survive failures that might have impacted other portions of your infrastructure. The controlling system allows you to perform situation analysis after the unplanned event to determine the status of the production systems or the disks. The controlling system plays a vital role in a GDPS/PPRC HM configuration.

The controlling system must be in the same sysplex as the production system (or systems) so it can see all the messages from those systems and communicate with those systems. However, it shares an absolute minimum number of resources with the production systems (typically only the sysplex couple data sets). By being configured to be as self-contained as possible, the controlling system will be unaffected by errors that might stop the production systems (for example, an ELB event on a primary volume).

The controlling system must have connectivity to all the Site1 and Site2 primary and secondary devices that it will manage. If available, it is preferable to isolate the controlling system infrastructure on a disk subsystem that is not housing mirrored disks that are managed by GDPS.

The controlling system is responsible for carrying out all PPRC and STP-related recovery actions following a disaster or potential disaster, for managing the disk mirroring configuration, for initiating a HyperSwap, for initiating a freeze and implementing the freeze policy actions following a freeze event, for reassigning STP roles, and so on.

The availability of the dedicated GDPS controlling system (or systems) in *all* configurations is a fundamental requirement of GDPS. It is not possible to merge the function of the controlling system with any other system that accesses or uses the primary volumes or other production resources.

Especially in 2-site configurations, configuring GDPS/HM with two controlling systems, one in each site is highly recommended. This is because a controlling system is designed to survive a failure in the opposite site of where the primary disks are. Primary disks are normally in Site1 and the controlling system in Site2 is designed to survive if Site1 or the disks in Site1 fail. However, if you reverse the configuration so that primary disks are in Site2, the controlling system is in the same site as the primary disks. It will certainly not survive a failure in Site2 and might or might not survive a failure of the disks in Site2, depending on the configuration. Configuring a controlling system in both sites ensures the same level of protection no matter which site is the primary disk site. When two controlling systems are available, GDPS manages by assigning a Master role to the controlling system that is in the same site as the secondary disks and switching the Master role if there is a disk switch.

Improved controlling system availability: Enhanced timer support

Normally, a loss of synchronization with the sysplex timing source will generate a disabled console WTOR that suspends all processing on the LPAR, until a response is made to the WTOR. The WTOR message is IEA394A in STP timing mode.

In a GDPS environment, z/OS is aware that a given system is a GDPS controlling system and will allow a GDPS controlling system to continue processing even when the server it is running on loses its time source and becomes unsynchronized. The controlling system is therefore able to complete any freeze or HyperSwap processing it might have started and is available for situation analysis and other recovery actions, instead of being in a disabled WTOR state.

In addition, because the controlling system is operational, it can be used to help in problem determination and situation analysis during the outage, thus reducing further the recovery time needed to restart applications.

The controlling system is required to perform GDPS automation if there is a failure. That might include these actions:

- ▶ Performing the freeze processing to guarantee secondary data consistency
- ▶ Coordinating HyperSwap processing
- ▶ Aiding with situation analysis

Because the controlling system needs to run only with a degree of time synchronization that allows it to correctly participate in heartbeat processing with respect to the other systems in the sysplex, this system should be able to run unsynchronized for a period of time (80 minutes) using the local time-of-day (TOD) clock of the server (referred to as *local timing mode*), rather than generating a WTOR.

Automated response to STP sync WTORs

GDPS on the controlling systems, using the BCP Internal Interface, provides automation to reply to WTOR IEA394A when the controlling systems are running in local timing mode. See “Improved controlling system availability: Enhanced timer support” on page 118. A server in an STP network might have recovered from an unsynchronized to a synchronized timing state without client intervention. By automating the response to the WTORs, potential time outs of subsystems and applications in the client’s enterprise might be averted, thus potentially preventing a production outage.

If either WTOR IEA394A is posted for production systems, GDPS uses the BCP Internal Interface to automatically reply RETRY to the WTOR. If z/OS determines that the CPC is in a synchronized state, either because STP recovered or the CTN was reconfigured, it will no longer spin and continue processing. If the CPC is still in an unsynchronized state when GDPS automation responded with RETRY to the WTOR, however, the WTOR will be reposted.

The automated reply for any given system is retried for 60 minutes. After 60 minutes, you will need to manually respond to the WTOR.

4.2.2 GDPS/PPRC HM in a single site

In the single-site configuration, the controlling systems, primary disks, and secondary disks are all in the same site, as shown in Figure 4-2 on page 119. This configuration allows you to benefit from the capabilities of GDPS/PPRC HM to manage the mirroring environment, and HyperSwap across planned and unplanned disk reconfigurations. A single site configuration does *not* provide disaster recovery capabilities, because all the resources are in the same site, and if that site suffers a disaster, then the systems and disk are all gone.

Note: We continue to refer to Site1 and Site2 in this section, although this terminology here refers to the two copies of the production data in the same site.

Even though having a single controlling system might be acceptable, we suggest having two controlling systems to provide the best availability and protection. The K1 controlling system can use Site2 disks, and K2 can use the Site1 disks. In this manner, a single failure will not affect availability of at least one of the controlling systems, and it will be available to perform GDPS processing.

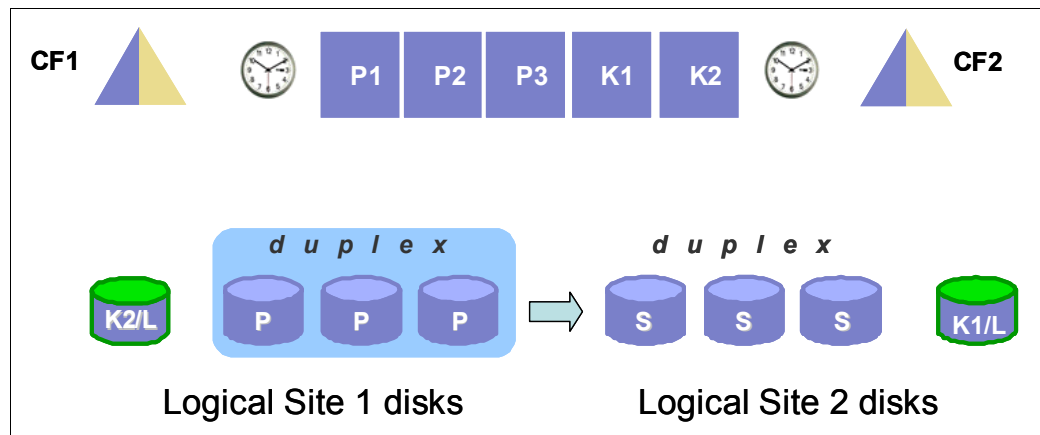


Figure 4-2 GDPS/PPRC HM single-site configuration

4.2.3 GDPS/PPRC HM in a 2-site configuration

Another option is to use GDPS/PPRC HM with the primary disk in one site, and the secondaries in a second site, as shown in Figure 4-3. This configuration *does* provide the foundation for disaster recovery because the secondary copy of disk is in a separate site protected from a disaster in Site1. GDPS/PPRC HM also delivers the freeze capability, which ensures a consistent set of secondary disk in case of a disaster.

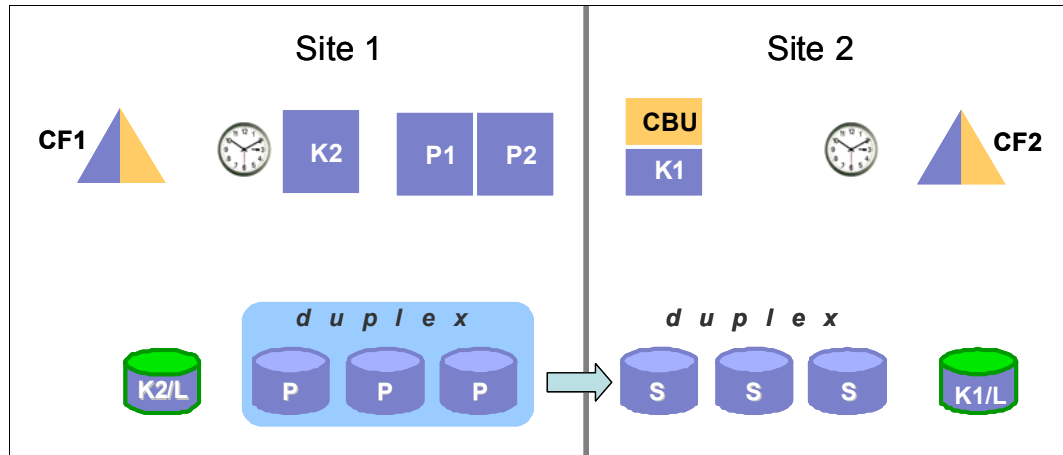


Figure 4-3 GDPS/PPRC HM 2-site configuration

If you have a 2-site configuration, and chose to implement only one controlling system, it is highly recommended that you place the controlling system in the recovery site. The advantage of this is that the controlling system will continue to be available even if a disaster takes down the whole production site. Placing the controlling system in the second site creates a multisite sysplex, meaning that you must have the appropriate connectivity between the sites. To avoid cross-site sysplex connections you might also consider the BRS configuration described in more detail in 3.2.4, “Business Recovery Services (BRS) configuration” on page 72.

To get the full benefit of HyperSwap and the second site, ensure that there is sufficient bandwidth for the cross-site connectivity from the primary site servers to the secondary site disk. Otherwise, although you might be able to successfully perform the HyperSwap to the second site, the I/O performance following the swap might not be acceptable.

4.2.4 GDPS/PPRC HM in a 3-site configuration

GDPS/PPRC HM can be combined with GDPS/XRC or GDPS/GM in a 3-site configuration. In this configuration, GDPS/PPRC HM provides protection from disk outages across a metropolitan area or within the same local site, and GDPS/XRC or GDPS/GM provides disaster recovery capability in a remote site.

We call these combinations GDPS/Metro Global Mirror (GDPS/MGM) or GDPS/Metro z/OS Global Mirror (GDPS/MzGM). In these configurations, GDPS/PPRC, GDPS/XRC, and GDPS/GM provide additional automation capabilities.

See Chapter 11, “Combining local and metro continuous availability with out-of-region disaster recovery” on page 331 for more information about the capabilities and limitations of using GDPS/PPRC HM in a GDPS/MGM and GDPS/MzGM solution.

4.2.5 Other important considerations

The availability of the dedicated GDPS controlling system (or systems) in *all* scenarios is a fundamental requirement in GDPS. It is not possible to merge the function of the controlling system with any other system that accesses or uses the primary volumes.

4.3 Managing the GDPS/PPRC HM environment

The bulk of the functions delivered with GDPS/PPRC HM relate to maintaining the integrity of the secondary disks and being able to nondisruptively switch to the secondary volume of the Metro Mirror pair.

However, there is an additional aspect of remote copy management that is available with GDPS/PPRC HM, namely the ability to query and manage the remote copy environment using the GDPS panels.

In this section, we describe this other aspect of GDPS/PPRC HM. Specifically, GDPS/PPRC HM provides facilities to let you:

- ▶ Be alerted to any changes in the remote copy environment
- ▶ Display the remote copy configuration
- ▶ Stop, start, and change the direction of remote copy
- ▶ Stop and start FlashCopy

Note: GDPS/PPRC HM does not provide script support. For scripting support with added capabilities, the full function GDPS/PPRC product is required.

4.3.1 NetView interface

There are two primary user interface options available for GDPS/PPRC: The NetView 3270 panels and a browser-based graphical user interface (also referred to as the GDPS GUI in this book).

An example of the main GDPS/PPRC 3270-based panel is shown in Figure 4-4.

VPCPPNLP	GDPS HyperSwap Manager				GDPS V3.R10.M0	
System	=	G0C2	A6PS3	Primary Dasd =	OK	SITE1 M0P1
Current Master	=	G0C2	A6PS3			
Parallel mode	=	YES		Primary FB =	OK	SITE1
HyperSwap F0/FB	=	ENABLED	YES			
Debug	=	OFF				
1	Dasd Remote Copy					
3	Standard Actions					
6	Planned Actions					
7	Sysplex Resource Management					
8	Debug ON/OFF					
9	View/Alter Definitions					
H	Health Checks					
C	Config Management					
M	Run Monitor1/Monitor3					
Selection ==> -						
Licensed Materials - Property of IBM						
6942-35B © Copyright IBM Corp. 1998, 2013 All Rights Reserved.						

Figure 4-4 GDPS/PPRC HyperSwap Manager main GDPS panel

Notice that several option choices are dimmed to the color *blue* instead of *black*; these blue options are supported by the GDPS/PPRC offering, but are not part of GDPS/PPRC HM.

This panel has a summarized configuration status at the top and a menu of choices. For example, to view the disk mirroring (Dasd Remote Copy) panels, enter 1 at the Selection prompt, and then press Enter.

GDPS graphical user interface

The GDPS GUI is a browser-based interface designed to improve operator productivity. The GDPS GUI provides the same functional capability as the 3270-based panels, such as providing management capabilities for Remote Copy Management, Configuration Management, SDF Monitoring, and browsing the CANZLOG using simple point-and-click procedures. Advanced sorting and filtering is available in most of the views provided by the GDPS GUI. In addition, users can open multiple windows or tabs to allow for continuous status monitoring, while performing other GDPS/PPRC HM management functions.

The GDPS GUI display has four main sections:

1. The application header at the top of the page that provides an Actions button for carrying out a number of GDPS tasks, along with the help function and the ability to logoff or switch between target systems.
2. The application menu is down the left hand side of the window. This menu gives access to various features and functions available through the GDPS GUI.
3. The active window that shows context-based content depending on the selected function. This tabbed area is where the user can switch context by clicking a different tab.
4. A status summary area is shown at the bottom of the display.

The initial status panel of the GDPS/PPRC HM GDPS GUI is shown in Figure 4-5 on page 123. This panel provides an instant view of the status and direction of replication, and HyperSwap status. Hovering over the various icons provides more information by using pop-up windows.

Note: For the remainder of this section, only the GDPS GUI is shown to illustrate the various GDPS management functions. The equivalent traditional 3270 panels are not shown here.



Figure 4-5 Full view of GDPS GUI main panel

Monitoring function: Status Display Facility

GDPS also provides many monitors to check the status of disks, sysplex resources, and so on. Any time there is a configuration change, or something in GDPS that requires manual intervention, GDPS will raise an alert. GDPS uses the Status Display Facility (SDF) provided by System Automation as the primary status feedback mechanism for GDPS.

GDPS provides a dynamically updated window, as shown in Figure 4-6. There is a summary of all current alerts at the bottom of each window. The initial view presented is for the SDF trace entries so you can follow, for example, script execution. Simply click one the other alert categories to view the different alerts associated with automation or remote copy in either site, or select **All** to see all alerts. You can sort and filter the alerts based on a number of the fields presented, such as severity.

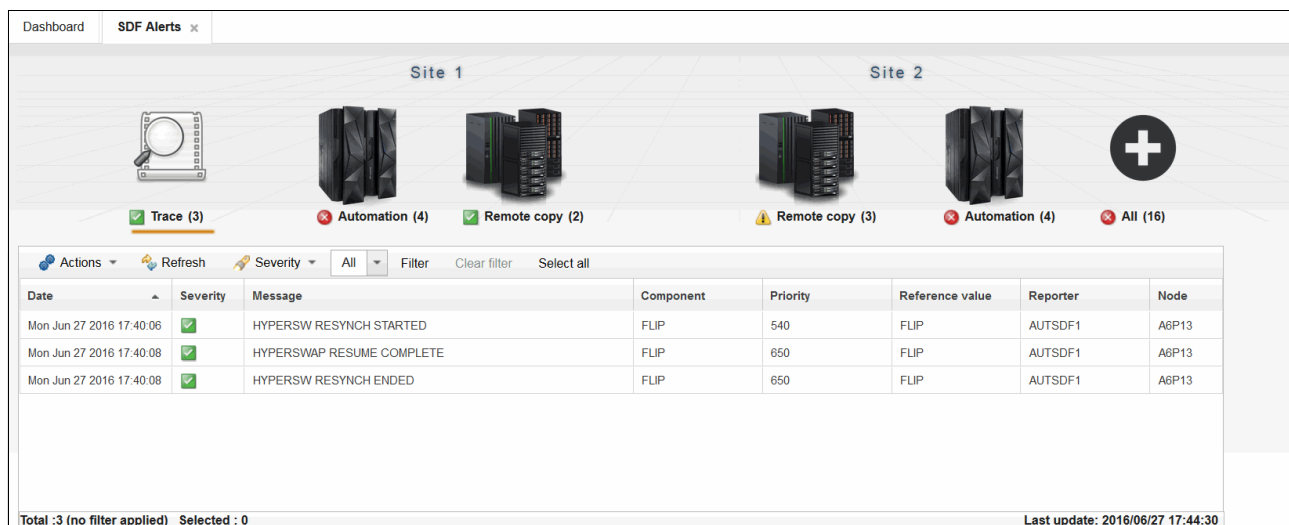


Figure 4-6 GDPS GUI SDF panel

By default, the GDPS GUI refreshes the alerts automatically every 10 seconds. As with the 3270 window, if there is a configuration change or a condition that requires special attention, the color of the icons will change based on the severity of the alert. By pointing to and clicking any of the highlighted fields, you can obtain detailed information regarding the alert.

Remote copy panels

The z/OS Advanced Copy Services capabilities are powerful, but the native command-line interface (CLI), z/OS TSO, and ICKDSF interfaces are not as user-friendly as the DASD remote copy panels are. To more easily check and manage the remote copy environment, use the DASD remote copy panels provided by GDPS.

For GDPS to manage the remote copy environment, you must first define the configuration (primary and secondary LSSs, primary and secondary devices, and PPRC links) to GDPS in a file called the GEOPARM file. This GEOPARM file can be edited and introduced to GDPS directly from the GDPS GUI.

After the configuration is known to GDPS, you can use the panels to check that the current configuration matches the one you want. You can start, stop, suspend, and resynchronize mirroring at the volume or LSS level. These actions can be done at the device or LSS level, or both, as appropriate. Figure 4-7 shows the mirroring panel for CKD devices at the LSS level.

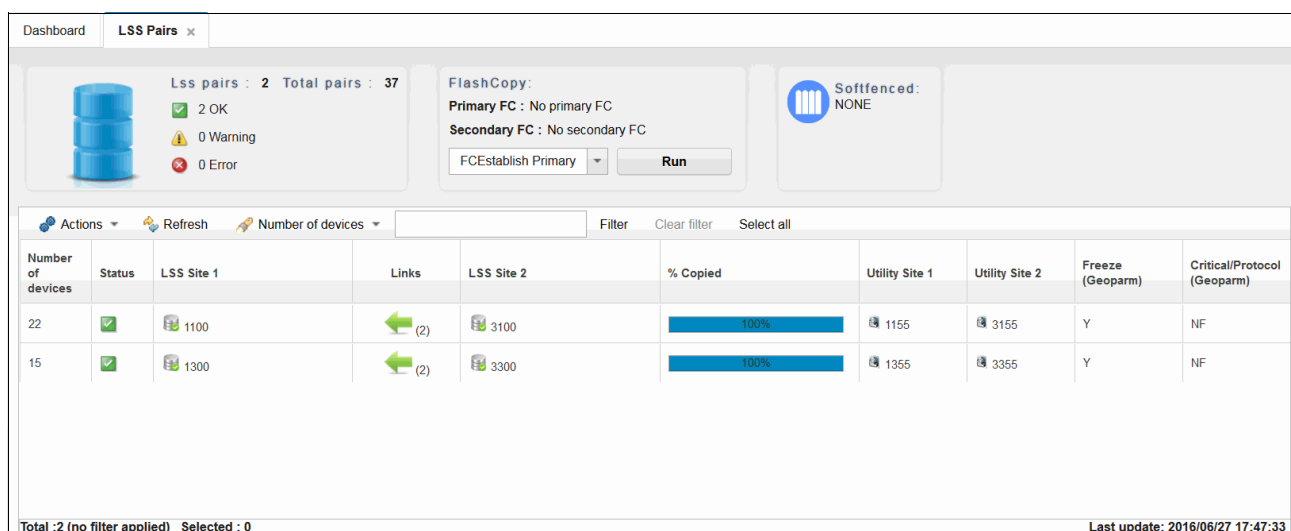


Figure 4-7 GDPS GUI Dasd Remote Copy SSID panel

The Dasd Remote Copy panel is organized into three sections:

- ▶ Upper left provides a summary of the device pairs in the configuration and their status.
- ▶ Upper right provides the ability to invoke GDPS-managed FlashCopy operations.
- ▶ A table with one row for each LSS pair in your GEOPARM. In addition to the rows for each LSS, there is a header row with an **Action** menu to enable you to carry out the various DASD management tasks, and the ability to filter the information presented.

To perform an action on a single SSID-pair, double click a row in the table. A panel is then displayed, where you can perform the same actions as those available as line commands on the top section of the 3270 panel.

After an individual SSID-pair is selected, the frame shown in Figure 4-8 is displayed. The table in this frame shows each of the mirrored device pairs within a single SSID-pair, along with the current status of each pair. In this example, all the pairs are fully synchronized and in duplex status, as summarized in the upper left area. Additional details can be viewed for each pair by double-clicking the row, or selecting the row with a single click and then selecting **Query** from the **Actions** menu.

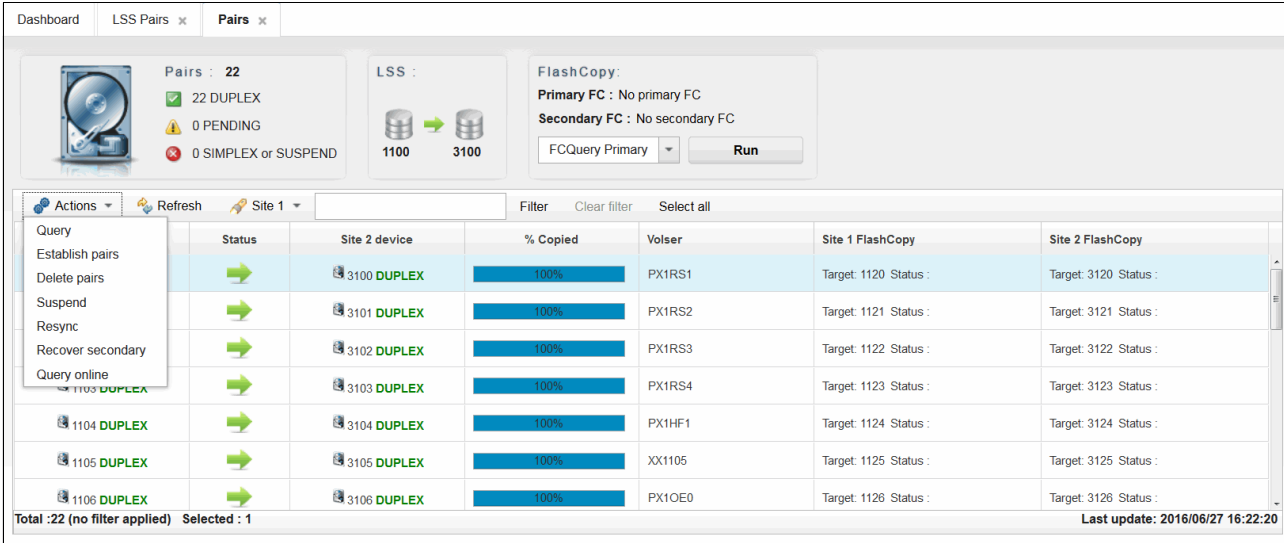


Figure 4-8 GDPS GUI Dasd Remote Copy: View Devices detail panel

If you are familiar with using the TSO or ICKDSF interfaces, you might appreciate the ease of use of the DASD remote copy panels.

Remember that these panels provided by GDPS are *not* intended to be a remote copy monitoring tool. Because of the overhead involved in gathering the information for every device to populate the panels, GDPS gathers this data only on a timed basis, or on demand following an operator instruction. The normal interface for finding out about remote copy status or problems is the Status Display Facility (SDF).

Similar panels are provided for controlling the Open LUN devices.

4.3.2 NetView commands

Even though GDPS/PPRC HM does not support using scripts as GDPS/PPRC does, certain GDPS operations are initiated through the use of NetView commands that perform similar actions to the equivalent script command. These commands are entered at a NetView command prompt.

There are commands to perform the following types of actions (not an all inclusive list). One command would accomplish all of the following actions:

- ▶ Temporarily disable HyperSwap and subsequently re-enable HyperSwap.
- ▶ List systems in the GDPS and identify which are controlling systems.
- ▶ Perform a planned HyperSwap disk switch.
- ▶ Perform a planned freeze of the disk mirror.
- ▶ Make the secondary disks usable through a PPRC failover or recover action.
- ▶ Restore PPRC mirroring to a duplex state.

- ▶ Take a point-in-time copy of the current set of secondary CKD devices. FlashCopy using COPY or NOCOPY options and the NOCOPY2COPY option to convert an existing FlashCopy taken with NOCOPY to COPY are supported. The CONSISTENT option (FlashCopy Freeze) is supported in conjunction with the COPY and NOCOPY options.
- ▶ Reconfigure an STP-only CTN by reassigning the Preferred Time Server (PTS) and Current Time Server (CTS) roles and the Backup Time Server (BTS) and Arbiter (ARB) roles to one or more CPCs.
- ▶ Unfence disks that were blocked by Soft Fence.

4.4 GDPS/PPRC HM monitoring and alerting

The GDPS SDF panel, discussed in “Monitoring function: Status Display Facility” on page 123, is where GDPS dynamically displays color-coded alerts.

Alerts can be posted as a result of an unsolicited error situation that GDPS listens for. For example, if one of the multiple PPRC links that provide the path over which PPRC operations take place is broken, there is an unsolicited error message issued. GDPS listens for this condition and will raise an alert on the SDF panel notifying the operator of the fact that a PPRC link is not operational.

Clients run with multiple PPRC links and if one is broken, PPRC still continues over any remaining links. However, it is important for the operations staff to be aware of the fact that a link is broken and fix this situation, because a reduced number of links results in reduced PPRC bandwidth and reduced redundancy. If this problem is not fixed in a timely manner, and more links have a failure, it can result in production impact because of insufficient mirroring bandwidth or total loss of PPRC connectivity (which results in a freeze).

Alerts can also be posted as a result of GDPS periodically monitoring key resources and indicators that relate to the GDPS/PPRC HM environment. If any of these monitoring items are found to be in a state deemed to be not normal by GDPS, an alert is posted on SDF.

Various GDPS monitoring functions are executed on the GDPS controlling systems and on the production systems. This is because, from a software perspective, it is possible that different production systems have a different view of some of the resources in the environment and although status can be normal in one production system, it might be not normal in another. All GDPS alerts generated on one system in the GDPS sysplex are propagated to all other systems in the GDPS. This propagation of alerts provides a single focal point of control. It is sufficient for operators to monitor SDF on the master controlling system to be aware of all alerts that are generated in the entire GDPS complex.

When an alert is posted, the operator will have to investigate (or escalate, as appropriate) and corrective action will need to be taken for the reported problem as soon as possible. After the problem is corrected, this is detected during the next monitoring cycle and the alert is cleared by GDPS automatically.

GDPS/PPRC HM monitoring and alerting capability is intended to ensure that operations are notified of and can take corrective action for any problems in their environment that can affect the ability of GDPS/PPRC HM to do recovery operations. This will maximize the chance of achieving your IT resilience commitments.

4.4.1 GDPS/PPRC HM health checks

In addition to the GDPS/PPRC HM monitoring described, GDPS provides health checks. These health checks are provided as a plug-in to the z/OS Health Checker infrastructure to check that certain settings related to GDPS adhere to preferred practices.

The z/OS Health Checker infrastructure is intended to check a variety of settings to see whether these settings adhere to z/OS optimum values. For settings found to be not in line with preferred practices, exceptions are raised in the Spool Display and Search Facility (SDSF). If these settings do not adhere to recommendations, this can hamper the ability of GDPS to perform critical functions in a timely manner.

Often, if there are changes in the client environment, this might necessitate adjustment of various parameter settings associated with z/OS, GDPS, and other products. It is possible that you can miss making these adjustments, which might affect GDPS. The GDPS health checks are intended to detect such situations and avoid incidents where GDPS is unable to perform its job because of a setting that is perhaps less than ideal.

For example, GDPS/PPRC HM requires that the controlling systems' data sets are allocated on non-mirrored disks in the same site where the controlling system runs. The Site1 controlling systems' data sets must be on a non-mirrored disk in Site1 and the Site2 controlling systems' data sets must be on a non-mirrored disk in Site2. One of the health checks provided by GDPS/PPRC HM checks that each controlling system's data sets are allocated in line with the GDPS preferred practices recommendations.

Similar to z/OS and other products that provide health checks, GDPS health checks are optional. Several optimum values that are checked and the frequency of the checks can be customized to cater to unique client environments and requirements.

Several z/OS preferred practices conflict with GDPS preferred practices. The z/OS and GDPS health checks for these result in conflicting exceptions being raised. For such health check items, to avoid conflicting exceptions, z/OS provides the capability to define a coexistence policy where you can indicate which preferred practice is to take precedence; GDPS or z/OS. GDPS includes sample coexistence policy definitions for the GDPS checks that are known to be conflicting with those for z/OS.

GDPS also provides a useful interface for managing the health checks using the GDPS panels. You can perform actions such as activate/deactivate or run any selected health check, view the customer overrides in effect for any preferred practices values, and so on.

Figure 4-9 shows a sample of the GDPS Health Check management panel. In this example you see that all the health checks are enabled. The status of the last run is also shown indicating that some were successful and some resulted in a medium exception. The exceptions can also be viewed using other options on the panel.

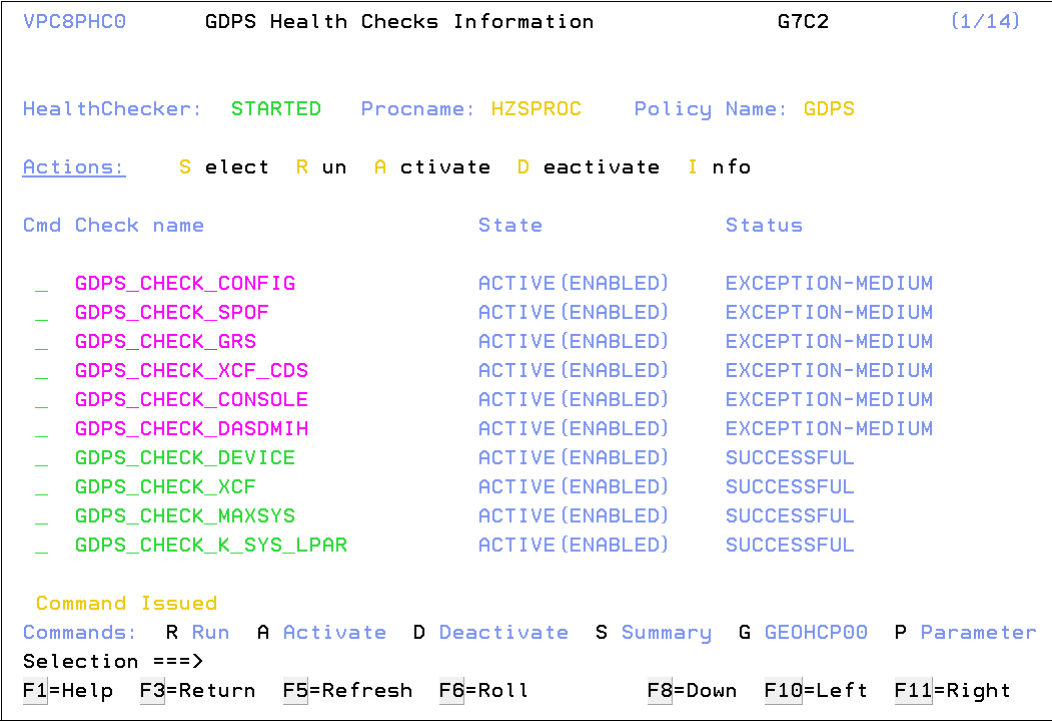


Figure 4-9 GDPS/PPRC HM Health Check management panel

4.5 Other facilities related to GDPS

In this section we describe miscellaneous facilities provided by GDPS/PPRC that can assist in various ways, such as reducing the window during which disaster recovery capability is not available.

4.5.1 HyperSwap coexistence

In the following sections we discuss the GDPS enhancements that remove various restrictions that had existed regarding HyperSwap coexistence with products such as Softek Transparent Data Migration Facility (TDMF) and IMS Extended Recovery Facility (XRF).

HyperSwap and TDMF coexistence

To minimize disruption to production workloads and service levels, many enterprises use TDMF for storage subsystem migrations and other disk relocation activities. The migration process is transparent to the application, and the data is continuously available for read and write activities throughout the migration process.

However, the HyperSwap function is mutually exclusive with software that moves volumes around by switching UCB pointers. The good news is that currently supported versions of TDMF and GDPS allow operational coexistence. With this support, TDMF automatically temporarily disables HyperSwap as part of the disk migration process only during the short time where it switches UCB pointers. Manual operator interaction is not required. Without this support, through operator intervention, HyperSwap is disabled for the entire disk migration, including the lengthy data copy phase.

HyperSwap and IMS XRF coexistence

HyperSwap also has a technical requirement that RESERVEs cannot be allowed in the hardware because the status cannot be reliably propagated by z/OS during the HyperSwap to the new primary volumes. For HyperSwap, all RESERVEs need to be converted to GRS global enqueue through the GRS RNL lists.

IMS/XRF is a facility by which IMS can provide one active subsystem for transaction processing, and a backup subsystem that is ready to take over the workload. IMS/XRF issues hardware RESERVE commands during takeover processing and these cannot be converting to global enqueues through GRS RNL processing. This coexistence problem has also been resolved so that GDPS is informed before IMS issuing the hardware RESERVE, allowing it to automatically disable HyperSwap. After IMS has finished processing and releases the hardware RESERVE, GDPS is again informed and reenables HyperSwap.

4.5.2 GDPS/PPRC HM reduced impact initial copy and resynchronization

Performing PPRC copy of a large amount of data across a large number of devices while the same devices are used in production by application workloads can potentially affect production I/O service times when those copy operations are performed synchronously. Your disk subsystems and PPRC link capacity are typically sized for steady state update activity but not for bulk, synchronous replication. Initial copy of disks and resynchronization of disks are examples of bulk copy operations that can affect production if performed synchronously.

There is no need to perform initial copy or resynchronizations using synchronous copy, because the secondary disks cannot be made consistent until all disks in the configuration have reached duplex state.

GDPS supports initial copy and resynchronization using asynchronous PPRC-XD (also known as Global Copy). When GDPS initiates copy operations in asynchronous copy mode, GDPS monitors progress of the copy operation. When the volumes are near full duplex state, GDPS converts the replication from the asynchronous copy mode to synchronous PPRC. Performing the initial copy or resynchronization using PPRC-XD eliminates the performance impact of synchronous mirroring on production workloads.

Without asynchronous copy it might be necessary to defer these operations or reduce the number of volumes copied at any given time. This would delay the mirror from reaching a duplex state, impacting a client's ability to recovery. Use of the XD-mode asynchronous copy allows clients to establish or resynchronize mirroring during periods of high production workload, and can potentially reduce the time during which the configuration is exposed.

This function requires that all disk subsystems in the GDPS configuration support PPRC-XD.

4.5.3 Reserve Storage Pool

Reserve Storage Pool (RSP) is a type of resource introduced with the z/OS Management Facility (z/OSMF) that can simplify the management of defined but unused volumes. GDPS provides support for including RSP volumes in the PPRC configuration that is managed by GDPS. PPRC primary volumes are expected to be online in controlling systems, and GDPS monitoring on the GDPS controlling systems results in an alert being raised for any PPRC primary device that is found to be offline. However, because z/OS does not allow RSP volumes to be brought online to any system, GDPS monitoring recognizes that an offline primary device is an RSP volume and suppresses alerting for these volumes.

4.5.4 GDPS/PPRC HM Query Services

GDPS maintains configuration information and status information in NetView variables for the various elements of the configuration that it manages. Query Services is a capability that allows client-written NetView REXX programs to query the value of numerous GDPS internal variables. The variables that can be queried pertain to the PPRC configuration, the system and sysplex resources managed by GDPS, and other GDPS facilities such as HyperSwap and GDPS Monitors.

Query Services allows clients to complement GDPS automation with their own automation code. In addition to the Query Services function that is part of the base GDPS product, GDPS provides several samples in the GDPS SAMPLIB library to demonstrate how Query Services can be used in client-written code.

GDPS also makes available to clients a sample tool called the Preserve Mirror Tool (PMT), which facilitates adding new disks to the GDPS PPRC HM configuration and bringing these disks to duplex. The PMT tool, which is provided in source format, makes extensive use of GDPS Query Services and thereby provides clients with an excellent example of how to write programs to benefit from Query Services.

4.5.5 Concurrent Copy cleanup

The DFSMS Concurrent Copy (CC) function uses a “sidefile” that is kept in the disk subsystem cache to maintain a copy of changed tracks that have not yet been copied. For a PPRCed disk, this sidefile is not mirrored to the secondary subsystem. If you perform a HyperSwap while a Concurrent Copy operation is in progress, the application using Concurrent Copy will fail following the completion of the HyperSwap. GDPS will not allow a planned swap when a Concurrent Copy session exists against your primary PPRC devices. However, unplanned swaps will still be allowed. Therefore, if you plan to use HyperSwap for primary disk subsystem failures (unplanned HyperSwap), try to eliminate any use of Concurrent Copy because you cannot plan when a failure will occur.

Checking for CC is performed by GDPS immediately before performing a planned HyperSwap. SDF trace entries are generated if one or more CC sessions exist, and the swap command will end with no PPRC device pairs being swapped. You must identify and terminate any CC and XRC sessions against the PPRC primary devices before the swap.

When attempting to resynchronize your disks, checking is performed to ensure that the secondary devices do not retain CC status from the time when they were primary devices. These are not supported as PPRC secondary devices. Therefore, GDPS will not attempt to establish a duplex pair with secondary devices if it detects a CC session.

GDPS can discover and terminate Concurrent Copy sessions that could otherwise cause errors. The function is controlled by a keyword that provides options to disable, to conditionally enable, or to unconditionally enable the cleanup of Concurrent Copy sessions on the target disks. This capability eliminates the manual task of identifying and cleaning up orphaned Concurrent Copy sessions.

4.6 GDPS/PPRC HM flexible testing and resync protection

Configuring point-in-time copy (FlashCopy) capacity in your PPRC environment provides two main benefits:

- ▶ You can conduct regular DR drills or other tests using a copy of production data while production continues to run.
- ▶ You can save a consistent, “golden” copy of the PPRC secondary data, which can be used if the primary disk or site is lost during a PPRC resynchronization operation.

FlashCopy and the various options related to FlashCopy are discussed in 2.6, “FlashCopy” on page 38. GDPS/PPRC HM supports taking a FlashCopy of the current secondary CKD disks. The COPY, NOCOPY and NOCOPY2COPY options are supported. CONSISTENT FlashCopy is supported in conjunction with COPY and NOCOPY FlashCopy.

In addition, FlashCopy can be used to provide a consistent point-in-time copy of production data to be used for nondisruptive testing of your system and application recovery procedures. FlashCopy can also be used, for example, to back up data without the need for extended outages to production systems; to provide data for data mining applications; and for batch reporting and other uses.

4.6.1 Use of space-efficient FlashCopy volumes

As discussed in “Space-efficient FlashCopy (FlashCopy SE)” on page 40, by using space-efficient (SE) volumes, you might be able to lower the amount of physical storage needed and thereby reduce the cost associated with providing a tertiary copy of the data. GDPS provides support for space-efficient FlashCopy volumes to be used as FlashCopy target disk volumes. Whether a target device is space-efficient or not is transparent to GDPS; if any of the FlashCopy target devices defined to GDPS are space-efficient volumes, GDPS will simply use them. All GDPS FlashCopy operations with the NOCOPY option, through panels or using the FLSHCOPY command or FlashCopies automatically taken by GDPS can use space-efficient targets.

Space-efficient volumes are ideally suited for FlashCopy targets when used for resync protection. The FlashCopy is taken before the resync and can be withdrawn as soon as the resync operation is complete. As changed tracks are sent to the secondary for resync, the time zero (T0) copy of this data is moved from the secondary to the FlashCopy target device. This means that the total space requirement for the targets is equal to the number of tracks that were out of sync, which is typically going to be significantly less than a full set of fully provisioned disks.

Another potential use of space-efficient volumes is if you want to use the data for limited disaster recovery testing.

You must understand the characteristics of space-efficient FlashCopy to determine whether this method of creating a point-in-time copy will satisfy your business requirements. For example, will it be acceptable to your business if, because of some unexpected workload condition, the repository on the disk subsystem for the space-efficient devices gets full and your FlashCopy is invalidated so that you are unable to use it? If your business requirements dictate that the copy must always be guaranteed to be usable, space-efficient might not be the best option and you can consider using standard FlashCopy instead.

4.7 GDPS tools for GDPS/PPRC HM

GDPS/PPRC also includes tools that provide function that is complementary to GDPS function. The tools represent the kind of function that all or many clients are likely to develop themselves to complement GDPS. Using tools eliminates the necessity for you to develop similar function yourself. The tools are provided in source code format, which means that if the tool does not meet your requirements completely, you can modify the code to tailor it to your needs.

The following tools are available with GDPS/PPRC HM:

- **Preserve Mirror Tool (PMT)**

This is intended to simplify and automate to a great extent the process of bringing new devices to PPRC duplex state. It also adds these devices to your running GDPS environment, while keeping the time during which the GDPS managed PPRC mirror is not full duplex (and therefore not protected by Freeze and HyperSwap) to a minimum.

PMT also provides facilities to aid with migration procedures when using Global Copy (PPRC-XD) and PPRC to migrate data to new disk subsystems.

- **Configuration Checker Tool (GEOCHECK)**

This checks whether all devices that are online to a GDPS production system are PPRCed under GDPS control, and raises alerts if violations are encountered. It provides identification and facilitates correction for any production devices that are inadvertently left out of the GDPS managed PPRC configuration. Not replicating some devices can prevent HyperSwap and also recovery from catastrophic disk or site failures.

- **GDPS Configuration Assistant (GeoAssistant) Tool**

This can help you to manage the GDPS/PPRC configuration definition file (GEOPARM file). It allows you to create a graphical view of your GEOPARM that can be easily shared and displayed on a variety of devices (such as a workstation, tablet, smartphone and so on). It can analyze and extract various statistics about your configuration. GeoAssistant can also provide step-by-step guidance for coding the GEOPARM statements when adding new devices to an existing configuration.

4.8 Services component

As explained, GDPS touches on much more than simply remote copy. It also includes automation, testing processes, disaster recovery processes, and others areas.

Most installations do not have skills in all these areas readily available. And it is extremely rare to find a team that has this range of skills across many implementations. However, the GDPS/PPRC HM offering includes exactly that: access to a global team of specialists in all the disciplines you need to ensure a successful GDPS/PPRC HM implementation.

Specifically, the Services component includes some or all of the following services:

- ▶ Planning to determine availability requirements, configuration recommendations, implementation and testing plans
- ▶ Assistance in defining recovery point objectives
- ▶ Installation and necessary customization of the special GDPS/PPRC HM versions of NetView and System Automation
- ▶ Remote copy implementation
- ▶ GDPS/PPRC HM automation code installation and policy customization
- ▶ Education and training on GDPS/PPRC HM setup and operations
- ▶ Onsite implementation assistance
- ▶ Project management and support throughout the engagement

GDPS/PPRC HM projects are typically much smaller than those for the other GDPS offerings. Nevertheless, the sizing of the services component of each project can be tailored for that project based on many factors including what automation is already in place, whether remote copy is already in place, whether the two centers are already in place with a multisite sysplex if required, and so on. This means that the skills provided are tailored to the specific needs of each particular implementation.

4.9 GDPS/PPRC HM prerequisites

For more information about the latest GDPS/PPRC HM prerequisites, see the following GDPS web page:

http://www.ibm.com/systems/z/advantages/gdps/getstarted/gdpspprc_hsm.html

4.10 Comparison of GDPS/PPRC HM to other GDPS offerings

So many features and functions are available in the various members of the GDPS family that recalling them all and remembering which offerings support them is sometimes difficult. To position the offerings.

Table 4-1 lists the key features and functions and indicates which ones are delivered by the various GDPS offerings.

Table 4-1 Supported features matrix

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/MTMM	GDPS Virtual Appliance	GDPS/XRC	GDPS/GM
Continuous availability	Yes	Yes	Yes	Yes	No	No
Disaster recovery	Yes	Yes	Yes	Yes	Yes	Yes
CA/DR protection against multiple failures	No	No	Yes	No	No	No
Continuous Availability for foreign z/OS systems	Yes with z/OS proxy	No	No	No	No	No
Supported distance	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	Virtually unlimited	Virtually unlimited
Zero Suspend FlashCopy support	Yes, using CONSISTENT	Yes, using CONSISTENT for secondary only	Yes, using CONSISTENT	No	Yes, using Zero Suspend FlashCopy	Yes, using CGPause
Reduced impact initial copy/resync	Yes	Yes	Yes	Yes	Not applicable	Not applicable
Tape replication support	Yes	No	No	No	No	No
Production sysplex automation	Yes	No	Yes	Not applicable	No	No
Span of control	Both sites	Both sites (disk only)	Both sites	Both sites	Recovery site	Disk at both sites; recovery site (CBU or LPARs)
GDPS scripting	Yes	No	Yes	Yes	Yes	Yes
Monitoring, alerting and health checks	Yes	Yes	Yes	Yes (except health checks)	Yes	Yes
Query Services	Yes	Yes	No	No	Yes	Yes
MSS support for added scalability	Yes (secondary in MSS1)	Yes (secondary in MSS1)	Yes (H2 in MSS1, H3 in MSS2)	No	No	Yes (GM FC and Primary for MGM in MSS1)
MGM 3-site and 4-site	Yes (all configurations)	Yes (3-site only and non-IR only)	Yes (all configurations)	No	Not applicable	Yes (all configurations)

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/MTMM	GDPS Virtual Appliance	GDPS/XRC	GDPS/GM
MzGM	Yes	Yes	Yes (non-IR only)	No	Yes	Not applicable
Open LUN	Yes	Yes	No	No	No	Yes
z/OS equivalent function for Linux for IBM z Systems	Yes	No	Yes (Linux for IBM z Systems running as a z/VM guest only)	Yes (Linux for IBM z Systems running as a z/VM guest only)	Yes	Yes
Heterogeneous support through DCM	Yes (VCS and SA AppMan)	No	No	No	Yes (VCS only)	Yes (VCS and SA AppMan)
z/BX hardware management	Yes	No	No	No	No	No
GDPS GUI	Yes	Yes	No	Yes	No	Yes

4.11 Summary

GDPS/PPRC HM is a powerful offering that can extend Parallel Sysplex availability to disk subsystems by delivering the HyperSwap capability to mask planned and unplanned disk outages. It also provides monitoring and management of the data replication environment, including the freeze capability. It can provide these capabilities either in a single site, or when the systems and disks are spread across two data centers within metropolitan distances.

In a multisite configuration, GDPS/PPRC HM can also be an entry-level offering, capable of providing zero data loss. The RTO is typically longer than what can be obtained with a full GDPS/PPRC offering. As time goes by, if your business needs to migrate from GDPS/PPRC HM to the full GDPS/PPRC offering, this can be achieved also.

In addition to disaster recovery and continuous availability capabilities, GDPS/PPRC HM provides a user-friendly interface for monitoring and managing the remote copy configuration.



GDPS/XRC

In this chapter, we discuss the capabilities and the prerequisites of the GDPS/XRC offering.

The GDPS/XRC offering extends the benefits of GDPS to installations that have a requirement for extended distance remote copy support. However, it is important to understand that GDPS/XRC is not simply GDPS/PPRC with a longer distance between the sites; there are additional differences which are discussed in this chapter.

This chapter describes the following capabilities of GDPS/XRC:

- ▶ Protecting your data
 - Protecting the integrity of the secondary data in the event of a disaster or suspected disaster
 - Management of the remote copy environment both through scripts and through a NetView panel interface
 - Support for remote copy management and consistency of the secondary volumes for data that is not z/OS data, coordinated with management of the z/OS data
- ▶ Controlling the resources managed by GDPS during normal operations, planned changes, and following a disaster
 - Management of the System Data Mover (SDM) LPARs (shutdown, IPL, and automated recovery)
 - Support for switching your production data and systems to the recovery site
 - User-customizable scripts that control how GDPS/XRC reacts to specified error situations and that can also be used for planned events

5.1 Introduction to GDPS/XRC

Extended Remote Copy (XRC), rebranded to *IBM System Storage z/OS Global Mirror*, is a combined hardware and software asynchronous remote copy solution. Consistency of the data is maintained through the Consistency Group function within the z/OS System Data Mover (SDM).

Because of the asynchronous nature of XRC, it is possible to have the secondary disk at greater distances than acceptable for PPRC. Channel extender technology can be used to place the secondary disks up to thousands of kilometers away. Because XRC is asynchronous, the impact it has on response times is minimal, and is independent of the distance between the primary and secondary volumes.

GDPS/XRC combines the benefits of GDPS with the extended distance capabilities of XRC. It includes automation to manage replication and automates the process of recovering the production environment with limited manual intervention, including invocation of CBU¹, thus providing significant value in reducing the duration of the recovery window and requiring less operator interaction.

Whereas GDPS/PPRC is a high availability and disaster recovery solution for a single multisite sysplex, GDPS/XRC is specifically an automated disaster recovery solution. GDPS/XRC controls the remote mirroring and automates the recovery of production data and workloads in the recovery site. The systems running GDPS/XRC are typically in the recovery site, remote from the production systems, and are not members of the sysplex at the primary site. Also, unlike GDPS/PPRC, GDPS/XRC has *no* knowledge of what is happening in the production systems. The only resources GDPS/XRC is aware of are the replication resources and the hardware resources in the recovery site. Following a disaster, the production systems are restored by GDPS/XRC at the recovery site.

Because XRC is an asynchronous remote copy technology, it is not possible to have zero data loss when using XRC. Therefore, the recovery point objective when using XRC must be more than zero, meaning that some minimal data loss is acceptable. In a typical XRC configuration, an RPO of one minute should be achievable. With sufficient bandwidth, clients with large configurations are able to maintain an RPO of from 1 to 5 seconds.

The recovery time objective for GDPS/XRC is not dissimilar to that achievable with GDPS/PPRC, typically between one and two hours. This is because GDPS/XRC automates the entire process of recovering the XRC mirror, activating temporary backup capacity, and restarting the production systems.

5.1.1 Protecting data integrity

With PPRC, you need to apply some automation (for example, the GDPS/PPRC Freeze function) on top of the standard PPRC functions to guarantee the integrity of the secondary disk across multiple subsystems. However, in GDPS/XRC, the design of XRC guarantees the integrity of the secondary disk data. From a remote copy perspective, the role of GDPS is to manage the remote copy configuration and to drive the recovery process.

¹ Where available.

The following systems support time stamping of I/Os when the target volume is defined as a primary XRC volume:

- ▶ Any supported release of z/OS.
- ▶ Linux on z Systems drivers support timestamping of writes and also contain changes to support device blocking.
- ▶ z/VM and its guests.

CKD volumes used by any number of these systems, which we refer to as production systems, can be managed by GDPS/XRC. Any number of sessions or Master sessions can be managed by GDPS/XRC. The volumes managed by an SDM or multiple SDMs that are coupled under the same Master session can be managed to a single point of consistency. For more information, see “XRC data consistency” on page 28.

If you have two z/OS sysplexes running your production workload, you can choose to XRC the entire data for those sysplexes under a single Master session (that is, as a single consistency group). In this case, however, if there is an incident that forces you to recover one of these sysplexes, you will need to recover both; you cannot recover one in the recovery site and leave the other running in the application site. If you need to recover them individually, then you would use two separate Master sessions, one for the data of each sysplex. A single instance of GDPS/XRC can manage these two different sessions.

It is also possible to use XRC to remote copy volumes being used by z Systems operating systems that do *not* time stamp their I/Os (for example, z/VSE). However, in this case it is not possible to provide consistency across multiple LSSs. For more information, see “Understanding the Importance of Timestamped Writes” in the latest revision of the *z/OS DFSMS Advanced Copy Services* manual.

z/OS is the only operating system that supports running the System Data Mover function that performs the XRC replication. Therefore, in a GDPS/XRC configuration, you need a minimum of two z/OS systems: one to provide the SDM function, and one dedicated GDPS controlling system. More than one SDM system might be required, depending on the amount of data to be replicated. SDM systems and the GDPS controlling system must be clustered into a Base or Parallel Sysplex to facilitate GDPS communication among the systems.

5.2 GDPS/XRC configuration

A GDPS/XRC configuration consists of one or more production systems and sysplexes updating the primary volumes in the production site, one or more SDM systems in the recovery site, and one GDPS controlling system (K-sys), also in the recovery site. The SDM systems and the controlling system must be in the same sysplex. There is no requirement for the production system to be in a sysplex; however, all of the systems updating the primary volumes must be connected to the same Sysplex Timers or the same Server Time Protocol (STP) network. Figure 5-1 shows a simplified illustration of the physical topology of a GDPS/XRC implementation.

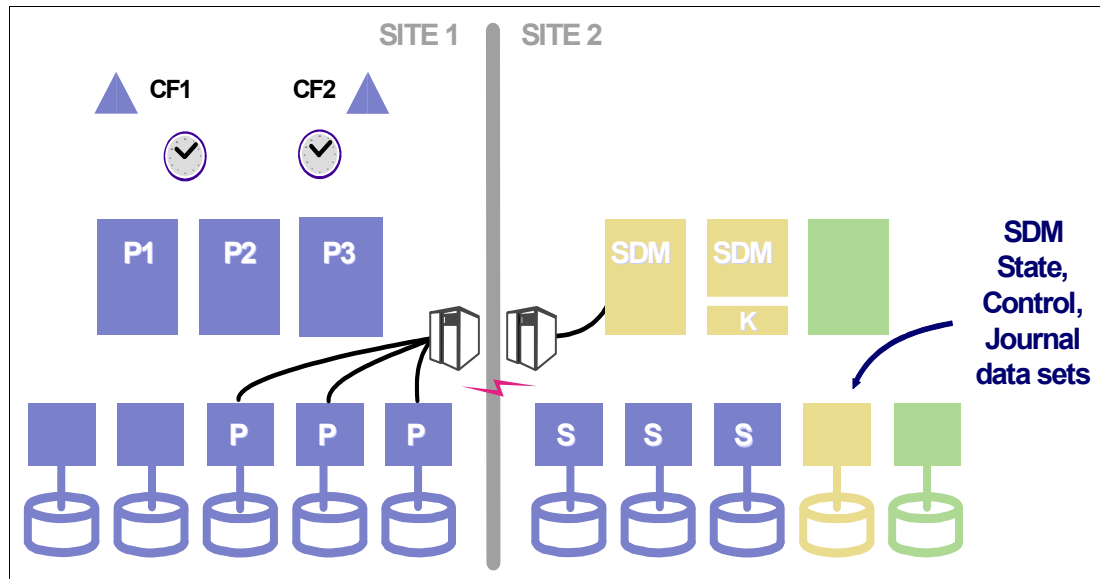


Figure 5-1 GDPS/XRC topology

As with all GDPS products, the GDPS/XRC controlling system is responsible for all remote copy management functions and for managing recovery following a disaster, so its availability is critical. Unlike a GDPS/PPRC configuration, however, there is no requirement to isolate the controlling system disks from the other systems in the GDPS sysplex (the SDM systems). The SDM systems and production systems can share infrastructure disks such as system residency volumes, the master catalog, the IBM RACF® database, and so on.

All critical data resides on storage subsystems in Site1 (the primary copy of data) and is mirrored to the storage subsystems in Site2 (the secondary copy of data) through XRC asynchronous remote copy. The systems in Site2 must have channel connectivity to the primary disk. Most clients use channel extension technology to provide this connectivity; there is no requirement for dark fiber between the sites.

In a more complex configuration, where you have more primary volumes, you might use the Coupled SDM and Multi-SDM support, both of which allow you to have a single point of consistency across multiple SDMs. GDPS/XRC supports both Coupled SDM and Multi-SDM.

In an even more complex configuration, GDPS/XRC can manage multiple master sessions, so you potentially can have two separate production sites, both using XRC to remote copy to a single recovery site, and have a single GDPS/XRC manage that recovery site and all associated XRC sessions.

5.2.1 GDPS/XRC in a 3-site configuration

GDPS/XRC can be combined with GDPS/PPRC (or GDPS/PPRC HM) in a 3-site configuration, where GDPS/PPRC (or GDPS/PPRC HM) is used across two sites within metropolitan distances (or even within a single site) to provide continuous availability through Parallel Sysplex use and GDPS HyperSwap, and GDPS/XRC is used to provide disaster recovery in a remote site.

We call this combination GDPS/Metro z/OS Global Mirror (GDPS/MzGM). In this configuration GDPS/PPRC and GDPS/XRC provide some additional automation capabilities.

After you understand the base capabilities described in 2.4.4, “Combining disk remote copy technologies for CA and DR” on page 35, see Chapter 11, “Combining local and metro continuous availability with out-of-region disaster recovery” on page 331 for more information about GDPS/MzGM.

5.3 GDPS/XRC management of distributed systems and data

GDPS/XRC provides the Distributed Cluster Management (DCM) capability for managing global clusters using Veritas Cluster Server (VCS) with the Global Cluster Option (GCO). When the DCM capability is used, GDPS/XRC does not manage remote copy or consistency for the distributed system data (this is managed by VCS). Therefore, it is not possible to have a common consistency point between the z Systems CKD data and the distributed data. However, for environments where a common consistency point is not a requirement, DCM together with VCS does provide key availability and recovery capabilities which might be of interest.

For more information about DCM, see 10.3.2, “DCM support for VCS” on page 308.

5.4 Managing the GDPS environment

GDPS/XRC monitors only the systems that comprise the GDPS sysplex: The SDM systems and the controlling system. This is shown in Figure 5-2. If all systems in the production site were to go down, GDPS/XRC has no automatic knowledge of this event. However, GDPS/XRC is able to monitor the recovery site server hardware, and it provides capabilities to manage these resources to automate restart of production in the recovery site.

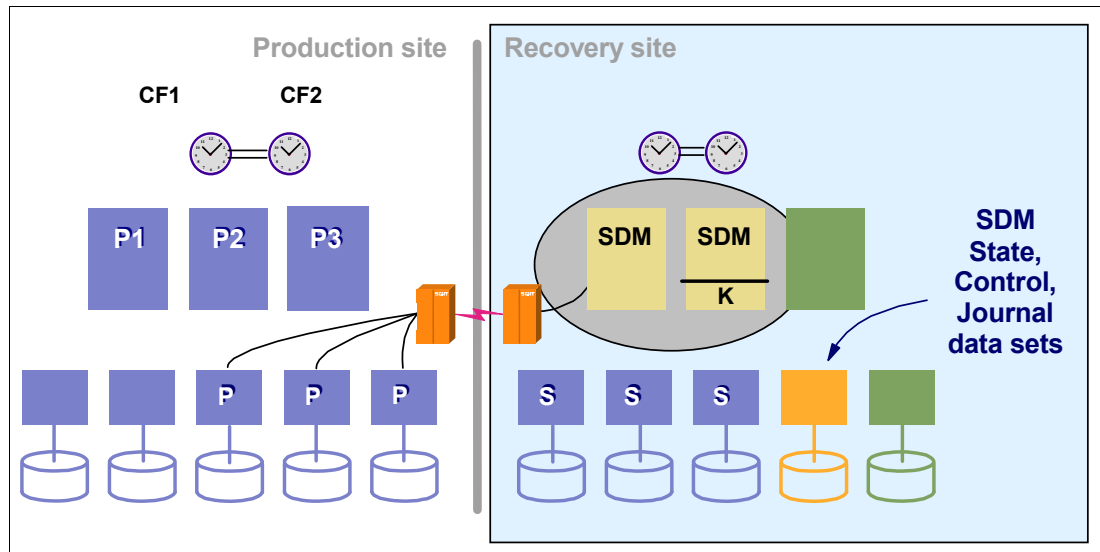


Figure 5-2 GDPS/XRC span of control

5.4.1 NetView interface

The NetView interface for GDPS actually consists of two parts. The first, and potentially the most important, is NetView's Status Display Facility (SDF). Any time there is a configuration change, or something in GDPS that requires manual intervention, GDPS will send an alert to SDF. SDF provides a dynamically-updated color-coded panel that provides the status of the systems and highlights any problems in the remote copy configuration. At all times, the operators should have an SDF panel within view so they will immediately become aware of anything requiring intervention or action.

The other aspect of the NetView interface consists of the panels provided by GDPS to help you manage and inspect the environment. The main GDPS panel is shown in Figure 5-3.

```

VPCPPNLX  GDPS - Disaster/Recovery System                                GDPS V3.R10.M0

System      = G0C1      A6PS1      Dasd Config = 2013-03-25 17:11:09
Current Master = G0C1      A6PS1      Region    = SITE1
Mirroring Status = OK      IDLE

          1      Dasd Remote Copy
          2      Tape Remote Copy
          3      Standard Actions

          6      Planned Actions
          7      Sysplex Resource Management
          8      Debug ON/OFF
          9      View Definitions
          H      Health Checks
          C      Config Management
          M      Run Monitor1/Monitor3

Selection ==>  -
      F1=Help      F3=Return      F6=Roll
  
```

Figure 5-3 Main GDPS/XRC panel

From this panel, you can do the following tasks:

- ▶ Query and control the remote copy configuration
- ▶ Initiate standard actions provided by GDPS against LPARs managed by GDPS (such as IPL, LPAR Deactivate, and so on)
- ▶ Initiate GDPS scripts that you create
- ▶ Manage coupling facilities and couple data sets relating to the SDM sysplex
- ▶ Manage the GDPS Health Checks
- ▶ Change or refresh the remote copy configuration definitions
- ▶ Run GDPS monitors

Remote copy panels

Although z/OS Global Mirror (XRC) provides a powerful replication capability, the operator interface is not as user-friendly as the DASD Remote Copy panels. To more easily check and manage the remote copy environment, use the DASD remote copy panels provided by GDPS.

For GDPS to manage the remote copy environment, you must first define the configuration (primary and secondary device numbers, FlashCopy devices, and information about the sessions and SDMs) to GDPS in a file called the GEOXPARM file.

After the configuration is known to GDPS, you can use the panels to check that the current configuration matches the one you want. You can start, stop, suspend, and resynch mirroring at the volume or LSS level, you can initiate a FlashCopy of the secondary volumes, perform coupled SDM operations, move SDMs to different LPARs, and so on. These actions can be done at the device or LSS level, or both, as appropriate. Figure 5-4 shows the mirroring panel for GDPS/XRC. In this example you see that GDPS is managing four SDM sessions. One of these, SDM04, is a stand-alone session. The remainder are coupled under a single Master named MSDM.

```

VPCPX00L           Mirroring Status = OK           X2C1
Actions           : ST Start      SU Suspend  V Vol      AS uspended  N Normloc
                  E End          R Recover   SE SEt      Q Query     A Altloc
Coupled-SDM:      C Couple       UN UNcouple RE Release  PU PURge     AD ADvance
FlashCopy         : ES EstabSec   WS WithdSec SM StopMonFC
Sessions          : 7
  -Session        Domain  CI  N-Dup/Dup/Pol/Utl  Consist.time
    - SDM04       A6P23    0/2/3/1              09:20:47.35873 IDLE 00:00:06.4
    - MSDM
    - CLUST01     A6P23
    - SDM01       A6P23    YI 0/27/28/1          09:20:47.20133 DELAY 00:00:00.10
    - SDM03       A6P23    YI 0/2/4/2            09:20:47.27640 DELAY 00:00:00.00
    - CLUST02     A6P25
    - SDM05       A6P25    YI 0/4/6/2            09:20:46.75741 DELAY 00:00:01.85

1 Start all sessions      2 End all sessions      4 Errormsg      More info =>
Selection ==>      -
F1=Help            F3=Return      F5=Refresh      F6=Roll
F7=Up              F8=Down              F11=Right

```

Figure 5-4 GDPS/XRC DASD Mirroring Session Status panel: Main view

If you are familiar with the TSO interface to XRC, you will appreciate how user-friendly this panel is.

Remember that these panels provided by GDPS are *not* intended to be a remote copy monitoring tool. Because of the overhead involved in gathering the information from all devices across all SDMs to populate the NetView panels, GDPS gathers this information only on a timed basis, or on demand following an operator instruction. The normal interface for finding out about remote copy problems is through SDF.

Standard Actions

We previously mentioned that the overwhelming majority of z Systems outages are planned outages. Even though GDPS/XRC manages only the SDM systems in the recovery site, it is still important that those systems are available and are correctly managed. GDPS provides facilities to help manage any outages affecting these systems. There are two reasons to use the GDPS facilities:

- ▶ They are well-tested and based on IBM preferred procedures.
- ▶ Using the GDPS interface lets GDPS know that the changes it is seeing (CDSs being deallocated or systems going out of the sysplex, for example) are planned changes, and therefore it is not to react to these events.

There are two types of resource-altering actions you can initiate from the panels: Standard Actions and Planned Actions.

Standard Actions are really single steps and are intended to perform an action on just one resource. Examples include performing a graceful shutdown of one of the systems in the GDPS sysplex, IPLing a system, maintaining the IPL address and the Loadparms that can be used for each system, selecting the IPL address and Loadparm to be used the next time a system is IPLed, and activating or deactivating or resetting an LPAR. For example, if you want to stop a system, change its IPL address and then IPL it again, those are three separate Standard Actions that you will initiate.

The GDPS/XRC Standard Actions panel is shown in Figure 5-5. It displays all the LPARs being managed by GDPS/XRC, and for each one, it shows the current status and various IPL information. It also shows (across the top) the actions that can be done on each system, including Stop, re-IPL (stop followed by IPL), Activate, and Deactivate. You will also see that there are some systems with status MANUAL. These are not systems in the GDPS sysplex. They are the “recovery systems,” which are the systems that GDPS can restart in the recovery site using recovered XRC disks or FlashCopy disks. Therefore, it is possible to perform hardware actions (activate/deactivate the partition, load, reset, and so on) against such foreign systems also.

VPCPSTD1		Standard Actions							X2C1
Actions:		S Stop	R ReIPL	M odify	Q QryxDR				
		L Load	X Reset	A Activate	D Deactivate	U Dump			
Sysname	CA	Status	IPLtype	LPAR	IPLmode	Auto	L-addr	Loadparm	
— X2C1	CA	MASTER	NORMAL	S65D	NORMAL	YN	7400	7407XBM	
— X2C2	A	ACTIVE	NORMAL	S23C	NORMAL	YN	7420	7427XBM	
— X2C3	A	ACTIVE	NORMAL	S23D	NORMAL	YN	7420	7427XBM	
— X2C4	A	????	NORMAL	S23E	NORMAL	YN	7420	7427XBM	
— CFXRC		MANUAL	NORMAL	S25D	NORMAL	NN			
— G2P1		MANUAL	NORMAL	S23D	TERTIARY	NN	7D00	7D07X2M	
— G2P2		MANUAL	NORMAL	S23E	TERTIARY	NN	7D00	7D07XBM	
— CF1		MANUAL	NORMAL	S67A	NORMAL	NN			

1 CPC Ops
Selection ==>
F1=Help F3=Return F6=Roll

Figure 5-5 Standard Actions panel for GDPS/XRC

GDPS provides support for taking a stand-alone dump using the GDPS Standard Actions panel. The stand-alone dump can be used against any z Systems operating system defined to GDPS, either a GDPS system (SDM and controlling systems) or foreign system (production recovery system), running native in an LPAR. Clients using GDPS facilities to perform HMC actions no longer need to use the HMC for taking stand-alone dumps.

5.4.2 GDPS scripts

Nearly all of the functions that can be initiated through the panels are also accessible from GDPS scripts. Additional facilities, not available on the panels, are also available using scripts. A script is a “program” or a workflow consisting of one or more GDPS functions. Scripts can be initiated manually through the GDPS panels (planned actions), and certain scripts can be initiated automatically by GDPS in response to an event (referred to as an *unplanned* action), or through a batch interface.

Scripts are written by you to automate the handling of certain situations, both planned changes and also error situations. This is an extremely important aspect of GDPS.

Scripts are powerful because they can access the full capability of GDPS. The ability to invoke all the GDPS functions through a script provides the following benefits:

- Speed

The script will execute the requested actions as quickly as possible. Unlike a human, it does not need to search for the latest procedures or the commands manual. Results of each command in the script are also analyzed and interpreted quickly. Result checking for such compound/complex actions by a human would require more in-depth skills in a variety of disciplines.

- Consistency

If you were to look into most computer rooms immediately following a system outage, what might you see? Mayhem! Operators frantically scrambling for the latest system programmer instructions. All the phones ringing. Every manager within reach asking when the service will be restored. And every systems programmer with access vying for control of the keyboards. All this results in errors, because humans naturally make mistakes when under pressure. But with automation, your well-tested procedures can execute in exactly the same way, time after time.

- Thoroughly tested procedures

Because scripts behave in a consistent manner, you can test your procedures over and over until you are sure they do everything that you want, in exactly the manner that you want. Also, because you need to code everything and cannot assume a level of knowledge (as you might with instructions intended for a human), you are forced to thoroughly think out every aspect of the action the script is intended to undertake. And because of the repeatability and ease of use of the scripts, they lend themselves more easily to frequent testing than manual procedures.

Planned Actions

Planned Actions are GDPS scripts that are initiated from the GDPS panels (option 6 on the main GDPS panel, as shown in Figure 5-3 on page 143). A Planned Action script might consist of several tasks. For example, you can have a script that will stop an LPAR, change its IPL address to the alternate SYSRES, and then restart it, all from a single script.

A more complex example of a Planned Action is shown in Figure 5-6. In this example, a single action in GDPS results in a tertiary copy of the secondary disks being taken, followed by IPLing the “production” systems in LPARs in the recovery site. This allows you to test your recovery procedure in the recovery site while live production continues to run in the application site and live production data continues to be protected by XRC to maintain up-to-date disaster readiness.

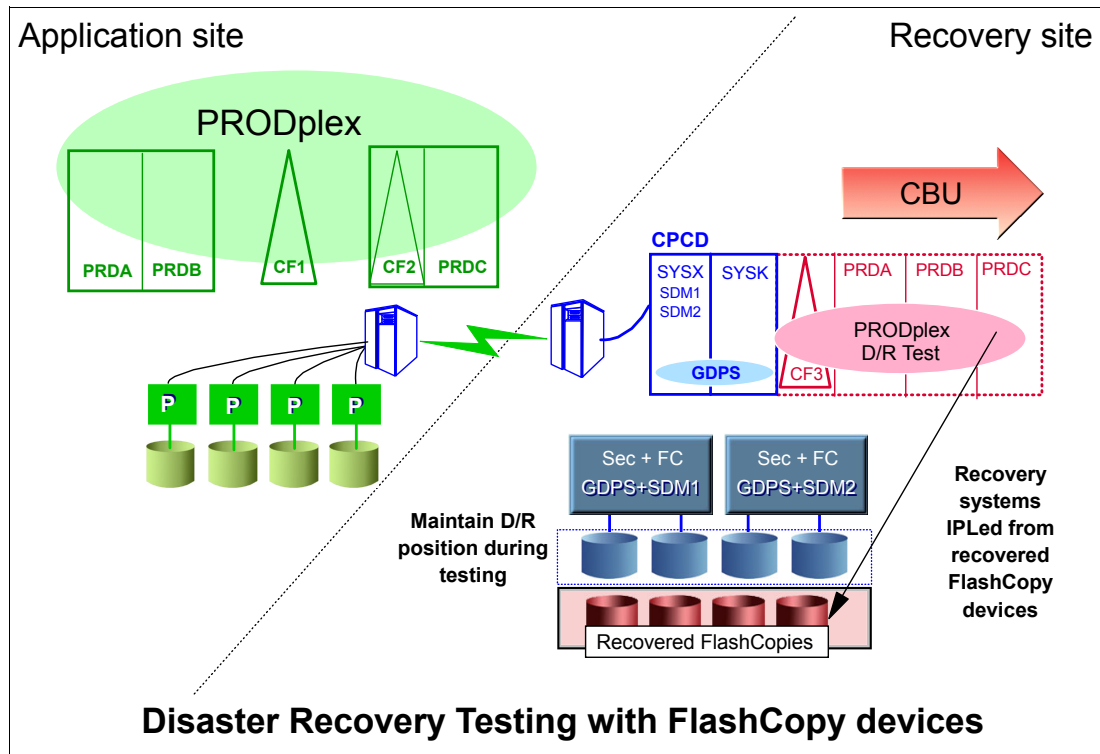


Figure 5-6 GDPS/XRC Planned Action

Specifically, the following actions are done by GDPS in this script:

- ▶ Zero Suspend FlashCopy is initiated:
 - This prevents the SDMs from writing new consistency groups to the secondary disks for a few seconds.
 - A FlashCopy is taken of all XRC secondary devices and the XRC infrastructure devices (the devices housing the XRC state, control, and journal data sets).
 - Zero Suspend FlashCopy completes and SDM processing resumes writing new consistency groups to the secondary disks.
- ▶ An XRC recover on the tertiary devices is performed.
- ▶ Temporary CBU capacity on CPCD is activated.
- ▶ Any test systems whose LPARs will be used for a recovery system in case of a disaster are deactivated.

- ▶ The CF LPARs and the LPARs that will house the recovered production systems are activated.
- ▶ The production recovery systems are started.

As a result of a single action that you performed (initiating the Planned Action), you have stopped discretionary work in the recovery site, created a copy of your production data and systems, and increased capacity, all while live production continued to run and maintain disaster readiness.

The use of a scripting capability removes the reliance on paper procedures which are invariably apt to go out of date, and ensures that the process is done the same way every time, with no vital steps accidentally overlooked.

Region Switch

GDPS defines a process for performing a planned Site Switch (also referred to as a *Region Switch*) between the two sites that act as the application and recovery sites. This process can be used for a planned Region Switch, and to return home to the original application region after an unplanned recovery (failover) to the recovery region.

The GDPS/XRC product provides capabilities that assist with and simplify various procedural aspects of a Region Switch or Return Home operation.

It is most likely that you will perform regular, planned region switches if your two regions are symmetrically configured, although this is not strictly mandatory. A symmetrically configured environment provides the same capabilities and allows you to use nearly identical procedures, no matter which region hosts the production systems and which region is the recovery site (hosting the GDPS/XRC environment). A symmetric configuration where a tertiary FlashCopy capacity is available in both regions is referred to as a *2+2 configuration*. A *1+1 configuration* is also symmetrical but does not provide the benefits associated with tertiary FlashCopy capacity no matter which region is hosting production and which is the recovery region.

Typically you run production in Region A, and Region B is the recovery site where you are likely to also have tertiary disks (FlashCopy capacity). If you do not have FlashCopy capacity in Region A but do in Region B, this is what we call a *1+2 configuration* which is not symmetrical. If you switch production to run in Region B, your recovery site in Region A is not equipped with tertiary disk and does not provide equivalent protection and ability to test, compared to running production in Region A and using Region B for recovery. Some of your operational procedures associated with GDPS will be different when running production in Region B versus when running in Region A.

The procedural steps for switching regions for a 1+1, 1+2, and for a 2+2 configuration will have similarities, but there will also be differences because of the differences in these configurations. The key difference is that the 2+2 configuration Region Switch will benefit from having the FlashCopy capacity in both sites, which will facilitate a faster switch with the least possible downtime to production systems when performing the switch.

At a high level, the sequence for moving production services from one region to the other includes these steps:

1. Assume that your production is running in Region-A and GDPS (controlling system and SDM systems) is running in Region-B.
2. Quiesce the production systems in Region-A and wait for the last updates to drain to Region-B.
3. Start the GDPS environment in Region-A.

4. Reverse replication from Region-B to Region-A and stop the SDM systems in Region-B.
5. Reversing replication does not require any data to copy because the source and target disks have identical content.
6. Start production systems in Region-B using GDPS facilities.

This procedure results in having production running in Region-B, GDPS running in Region-A, and achieving continuous DR protection.

Return Home after an unplanned failover to the recovery region

You might have to recover your production operations in the recovery region as a result of a catastrophic failure in the original application region. After running production in the recovery region for some time, if you want to return operations back to the original application region when it is restored, you can use a modified version of the region switch procedure. The key difference is that return home requires all data to be copied back to the original application region. After all data is copied back, then the operation to return is effectively a region switch as described in “Region Switch” on page 148.

Unplanned Actions

Unplanned Actions are GDPS scripts (also known as Takeover scripts), just like Planned Actions. However, they are used in a different way. Planned Actions are initiated from the GDPS panels. Unplanned Actions are initiated by GDPS in response to a failure event.

Remember that in a GDPS/XRC environment, GDPS has knowledge only about what is happening in the GDPS sysplex in the recovery site. GDPS does not monitor and therefore cannot detect failures in the application site. The script to recover XRC and restart production in the recovery site would be defined as a Planned Action. You could view this as a pre-planned, unplanned action.

In GDPS/XRC, Unplanned Actions are used only to react to failure of an SDM system or the GDPS controlling system (remember that the GDPS code runs in every system, so if the controlling system fails, GDPS in one of the SDM systems will detect that and react with an Unplanned Action script). The intent of such a script would be to re-IPL the failed system. Such scripts are not run automatically. GDPS will detect the failure and propose running the appropriate script. The operator would then have the choice of accepting to run the script in which case GDPS would initiate it, or to do nothing.

Batch scripts

Because the full range of GDPS functions is available to you, you can have scripts that will do normal operational processes for you. This is especially suited to processes that are run regularly, and have some interaction with the GDPS environment.

One of the challenges faced by any medium to large client with high availability requirements is creating a set of consistent tape backups. Backing up tens of terabytes to tape involves stopping the applications for many minutes, which is time that is not available in most installations. However, using a combination of GDPS batch scripts and FlashCopy, you *can* achieve this.

Just as you can have a Planned Action to create a set of tertiary volumes for a DR test, you can have a similar script that creates the tertiary volumes, and then take tape backups of those volumes. The net effect is basically the same as though you had stopped all work in your primary site for the duration of the backup, but without the impact to your applications. A script like this can be initiated from a batch job; such scripts are referred to as *batch scripts*.

Sysplex resource management

There are certain resources that are vital to the health and availability of the sysplex. Even though, in a GDPS/XRC environment, GDPS does not manage your production systems or their sysplex resources, it *does* manage your SDM sysplex. And to ensure the timeliness and consistency of your remote copies, it is important that the SDM systems have similarly high levels of availability.

The GDPS/XRC Sysplex Resource Management panel, as shown in Figure 5-7, provides you with the ability to manage the SDM sysplex resources. For example, if you switch to a new Primary sysplex CDS using the SETXCF PSWITCH command, you end up with a new Primary CDS but no alternate, thereby introducing a single point of failure.

However, if you use the GDPS Sysplex Resource Management functions, part of the function includes adding a new alternate after the switch of the primary had completed successfully, thereby ensuring that you do not have a single point of failure in the CDS configuration.

VPCSPN1		Sysplex Resource Management		X2C1	
Actions: V <u>iew</u> M <u>odify</u>					
TYPE	Primary DSN	Alternate DSN			
— SYSPLEX	SYS1.X2.XCF.CDSA1	SYS1.X2.XCF.CDSA3			
— ARM	SYS1.X2.ARM.CDSA0	SYS1.X2.ARM.CDSA1			
— BPXMCD	SYS1.X2.OMVS.CDSA0	SYS1.X2.OMVS.CDSA1			
— CFRM	SYS1.X2.CFRM.CDSA0	SYS1.X2.CFRM.CDSA1			
— LOGR	SYS1.X2.LOGR.CDSA0	SYS1.X2.LOGR.CDSA1			
— SFM	SYS1.X2.SFM.CDSA0	SYS1.X2.SFM.CDSA1			
— WLM	SYS1.X2.WLM.CDSA0	SYS1.X2.WLM.CDSA1			
		--	CF	--	
POLICY					
— POLXRC	CFX22	CFXRC	CFX23	-	
		-	-		
1 Use Pri/Spare1 CDS-es 2 Use Alt/Spare2 CDS-es 3 USE Normal CDS-es					
Selection ==> _					
F1=Help		F3=Return		F6=Roll	

Figure 5-7 GDPS/XRC Sysplex Resource Management panel

Although it might not receive as much attention as recovering from a disaster, the capability of GDPS to perform Planned Actions is used far more frequently, and it provides tremendous value in terms of faster turnaround and mistake avoidance.

5.4.3 System management actions

Nearly all of the GDPS Standard Actions and several script commands require actions to be done on the HMC. The interface between GDPS and the HMC is through the BCP Internal Interface (BCPii). This allows GDPS to communicate directly with the hardware for automation of HMC actions such as LOAD, DUMP, RESET, and ACTIVATE/DEACTIVATE an LPAR, or ACTIVATE/UNDO CBU or OOCOD.

The GDPS LOAD and RESET Standard Actions (available through the panel or scripts) allow specification of a *CLEAR* or *NOCLEAR* operand. This provides operational flexibility to accommodate your procedures.

Furthermore, when you LOAD a system using GDPS (panels or scripts), GDPS can listen for operator prompts from the system being loaded and reply to such prompts. GDPS provides support for optionally replying to IPL time prompts automatically, removing reliance on operator skills and eliminating operator error for any messages that require replies.

5.5 GDPS/XRC monitoring and alerting

The GDPS SDF panel, which is described in 5.4.1, “NetView interface” on page 142, is where GDPS dynamically displays alerts that are color-coded, based on severity, if and when a non-normal status or situation is detected.

Alerts can be posted as a result of an unsolicited error situation that GDPS listens for. For example, if there is a problem with any of the XRC sessions and the session suspends outside of GDPS control, GDPS will be aware of this because the SDM responsible for the given session will post an error. GDPS listens for this error and will, in turn, raise an alert on the SDF panel notifying the operator of the suspension event. It is important for the operator to initiate action to investigate and fix the reported problem as soon as possible because a suspended session directly translates to eroding RPO.

Alerts can also be posted as a result of GDPS periodically monitoring key resources and indicators that relate to the GDPS/XRC environment. If any of these monitored resources are found to be in a state deemed to be not normal by GDPS, an alert is posted on SDF. For example, GDPS uses the BCP Internal Interface to perform hardware actions to reconfigure the recovery site, either for disaster testing or in a real recovery scenario. To ensure that a recovery operation will not be affected, GDPS monitors the BCP internal interface connection to all CPCs in the recovery site on which the GDPS can perform hardware operations, such as CBU or LPAR activation.

Monitoring takes place on all systems in the GDPS sysplex (that is, the SDM systems and the GDPS controlling system). Alerts generated on any of these systems are propagated to all of the other systems. This allows a single system (normally the GDPS controlling system) to be used as a single focal management and monitoring point.

If an alert is posted, the operator needs to investigate it (or escalate, as appropriate) and a corrective action must be taken for the reported problem as soon as possible. After the problem is corrected, this is detected during the next monitoring cycle and the alert is cleared by GDPS automatically.

The GDPS/XRC monitoring and alerting capability is intended to ensure that operations are notified and can take corrective action for any problems in the environment that can affect the ability of GDPS/XRC to do recovery operations. This maximizes the installation's chance of achieving RPO and RTO commitments.

GDPS/XRC integrated XRC performance monitoring

Traditionally, clients have used the XRC Performance Monitor (XPM) product to monitor XRC performance. You can capture some of the messages issued by XPM to drive some automation. For example, customers typically capture messages issued by an XPM function known as the Batch Exception Monitor to suspend an XRC session that is experiencing excessive delays to suspend the session that appears to be in trouble. This proactive suspending of an XRC session is done to eliminate any risk of this problematic XRC session affecting production workloads and is referred to as the “Big Red Switch.” Such add-on automation is not integrated with GDPS automation even though it is often not desirable to affect GDPS managed resources outside of GDPS control.

In addition to the capabilities offered by XPM, a *GDPS/XRC Performance Monitoring Toolkit* is supplied with GDPS. The toolkit provides functions that are complementary to the capabilities provided by XPM.

In an effort to reduce the various products and tools required for XRC performance monitoring, eliminate the requirement for add-on automation, and to provide tighter integration with GDPS automation, GDPS has started to integrate and provide performance monitoring capability as part of GDPS.

In GDPS/XRC 3.10, the first installment of *GDPS/XRC integrated performance monitoring* is delivered. The objective of this first delivery is to make GDPS/XRC aware of System Data Mover performance data and to start using it to drive alerts and actions. The intent of this first installment is to provide autonomic “self-protection” capabilities that equal or exceed the XPM Batch Exception Monitor function.

The integrated performance monitoring allows you to create a policy to define certain thresholds that you consider are indicative of an XRC session being in trouble. For example, the exposure time, the percentage of cache used by an XRC session, or an increase in the amount of residual data in the primary storage controller’s side file can be indications of an XRC session in trouble. You can define the thresholds, and when these thresholds are exceeded, GDPS will raise SDF alerts for you to review the situation and take corrective action if required. Also, you can choose whether or not GDPS should automatically suspend a session that exceeds its exposure time threshold (that is, whether GDPS should throw the Big Red Switch on the session).

5.5.1 GDPS/XRC health checks

In addition to the GDPS/XRC monitoring, GDPS provides health checks. These health checks are provided as a plug-in to the z/OS Health Checker infrastructure to check that certain settings related to GDPS adhere to GDPS preferred practices recommendations.

The z/OS Health Checker infrastructure is intended to check a variety of settings to see whether these settings adhere to z/OS preferred practices values. For settings that are found to be not in line with preferred practices, exceptions are raised in Spool Display and Search Facility (SDSF). Many products, including GDPS, provide health checks as a plug-in to the z/OS Health Checker. There are various parameter settings related to GDPS, such as z/OS PARMLIB settings or NetView settings, and we document the recommendations and preferred practices for these settings in the GDPS publications. If these settings do not adhere to recommendations, this can hamper the ability of GDPS to perform critical functions in a timely manner.

Although GDPS monitoring will detect that GDPS was not able to perform a particular task and raise an alert, the monitor alert might be too late, at least for that particular instance of an incident. Often, if there are changes in the client environment, this can necessitate adjustment of various parameter settings associated with z/OS, GDPS, and other products. It is possible that you can miss making these adjustments, which might result in affecting GDPS. The GDPS health checks are intended to detect such situations and avoid incidents where GDPS is unable to perform its job because of a setting that is perhaps less than ideal.

For example, several address spaces are associated with GDPS/XRC, and preferred practices recommendations are documented for these. GDPS code itself runs in the NetView address space and there are DFSMS System Data Mover (SDM) address spaces that GDPS interfaces with to perform XRC copy services operations. GDPS recommends that these address spaces are assigned specific WLM service classes to ensure that they are dispatched in a timely manner and do not lock each other out. One of the GDPS/XRC health checks, for example, checks that these address spaces are set up and running with the GDPS recommended characteristics.

Similar to z/OS and other products that provide health checks, GDPS health checks are optional. Several preferred practices values that are checked and the frequency of the checks can be customized to cater to unique client environments and requirements.

GDPS also provides a useful interface for managing the health checks using the GDPS panels. You can perform actions such as activate, deactivate, or run any selected health check, view the customer overrides in effect for any preferred practices values, and so on. Figure 5-8 shows a sample of the GDPS Health Check management panel. In this example you see that all the health checks are enabled. The status of the last run is also shown, indicating whether the last run was successful or whether it resulted in an exception. Any exceptions can also be viewed using other options on the panel.

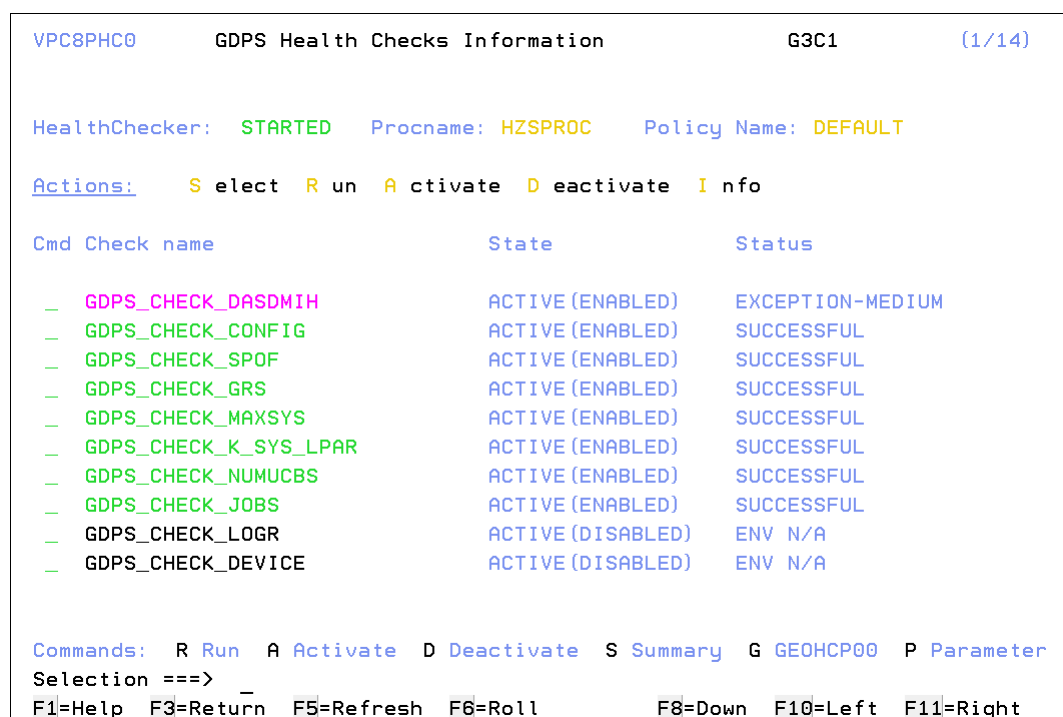


Figure 5-8 GDPS/XRC Health Check management panel

5.6 Other facilities related to GDPS

In this section, we describe miscellaneous facilities provided by GDPS/XRC that might assist in various different ways, such as reducing the window of when DR capability is not available.

5.6.1 FlashCopy disk definition in the GDPS systems

In a GDPS/XRC environment, many disks such as the primary, secondary, and FlashCopy disks need to be defined to the SDM systems. If all of the devices needed to be uniquely identified, this would restrict the number of devices that could be managed.

GDPS provides an option that allows alternatives to defining FlashCopy devices in the systems in the GDPS sysplex. No-UCB FlashCopy support accommodates performing FlashCopy in configurations where the FlashCopy target devices are not defined to some or all of the systems in the GDPS/XRC sysplex. This removes the requirement to define the FlashCopy devices in SDM systems and in any systems in the GDPS sysplex.

Removing the requirement to define FlashCopy devices to all systems in the GDPS/XRC sysplex provides device connectivity (“UCB”) constraint relief to clients with large configurations, allowing a larger number of volume pairs to be managed by GDPS/XRC.

5.6.2 GDPS/XRC FlashCopy locking

GDPS FlashCopy support provides critical protection for the FlashCopy target devices. GDPS logic ensures that when a FlashCopy is taken, it is taken only if the FlashCopy source devices represent a valid recovery point. This eliminates exposures that can result from accidentally overwriting a valid consistent FlashCopy with an invalid one.

There is also support to allow users to “lock out” FlashCopy target devices, effectively not allowing GDPS to take a FlashCopy, even when the FlashCopy source devices do represent a valid recovery point. This facility is useful for clients that are using the FlashCopy target devices for a specific activity (such as testing or dumping to tape), and do not want them to be overwritten until this activity has completed. The lock can then be released after the specific activity is complete.

5.6.3 GDPS/XRC Configuration checking

The SDMs, the LPARs where the SDMs can run, the devices that each SDM will manage, primary and secondary devices, and FlashCopy target devices are all defined in the GDPS GEOXPARM file.

When you introduce the configuration to GDPS and subsequently make changes, GDPS performs thorough checking of the specifications in the GEOXPARM file. In large configurations with multiple SDMs, with each SDM managing many devices, it is possible to make errors. One of the more common errors is specifying the same physical device to be used for multiple purposes. The same physical device could have been specified in the configuration as a secondary device for one SDM and as a FlashCopy target for another SDM. If such a configuration error went undetected it could cause issues with recovery, and the error might go undetected until it is too late to fix.

GDPS performs several checks when it is processing the GEOXPARM configuration file, including a check to ensure that each Primary, Secondary, XRC Infrastructure, and FlashCopy target device is a unique physical device.

5.6.4 Vary-After-Clip automation

GDPS simplifies definition of the XRC configuration, allowing device ranges to be used. This allows defining up to 255 contiguous devices to be mirrored with a single statement in the GEOXPARM configuration definition file. If each device had to be defined individually with its unique volume serial numbers, the configuration management and maintenance task would be virtually impossible. However, the XRC copy technology is actually based on volume serial numbers rather than device numbers. Therefore, when the GEOXPARM information is introduced to GDPS, GDPS queries the devices to determine the volume serial numbers and is then able to perform management actions that rely on volume serials.

When an XRC primary device is relabeled on a production system, this causes the volume serial information in the SDM system control blocks and the GDPS internal information to be incorrect. SDM and GDPS still have the old volume serial information. This can lead to problems with certain operations and can be tedious to fix.

GDPS provides a function known as *Vary After Clip (VAC)* automation. When a primary device is relabeled, the SDM captures this event and issues a message. GDPS captures this message to drive automation that performs the necessary actions to refresh both the SDM and the GDPS volume serial information for the relabeled device.

5.6.5 GDPS use of the XRC offline volume support

The XRC copy technology has required the primary volumes to be online in the SDM system managing XRC for several XRC operations.

Clients prefer to keep the application volumes which are the XRC primary volumes offline to the SDM systems for several reasons. For example, keeping the application volumes online in SDM systems increases the risk of accidental access and update to these volumes from the SDM systems.

Clients that have preferred to run with their primary application volumes offline in the SDMs have had to vary the volumes online when performing operations where XRC requires the volumes to be online. This varying online can take a long time, especially in channel-extended environments with a large distance between the application site where the primary volumes reside and the recovery site where the SDMs run.

In z/OS 2.1, XRC was enhanced to remove the requirement to have the primary volumes online in the SDMs for several XRC operations. This is what we call the *XRC offline volume support*. Some operations, however, continue to require a subset of the primary volumes online in the SDMs. For example, when adding new volumes to a running XRC session, the new volumes must be online.

GDPS supports this capability of keeping primary volumes offline in the SDM systems for applicable XRC operations. The client specifies a parameter indicating that they want to use this XRC offline volumes support and have a preference for keeping primary volumes offline.

With this preference specified, whenever an operation is performed that requires a subset of the volumes to be online, the client does not need to deal with bringing these devices online and then varying them offline again. GDPS brings the relevant devices online in the associated SDM system, performs the requested operation, and then varies the devices offline once again.

GDPS also regularly monitors the status of the primary volumes in the SDM systems. If the preference to use XRC offline volume support is specified, then GDPS will alert the operator if any primary volumes are found to be online in the SDM systems.

GDPS use of the XRC offline volume support simplifies XRC operations for those customers who have a preference for keeping application volumes offline in SDM systems.

5.6.6 Query Services

GDPS maintains configuration information and status information in NetView variables for the various elements of the configuration that it manages. GDPS Query Services is a capability that allows client-written NetView REXX programs to query and obtain the value for numerous GDPS internal variables.

The Query Services feature enables clients to extend and complement GDPS automation with their own automation REXX code. This can be used for various purposes such as reporting, monitoring, or problem determination, and for developing GDPS Tools.

In addition to the Query Services function that is part of the base GDPS product, GDPS provides several samples in the GDPS SAMPLIB library to demonstrate how Query Services can be used in client-written code.

5.6.7 Easy Tier Heat Map Transfer

IBM DS8000 Easy Tier optimizes data placement (placement of logical volumes) across the various physical tiers of storage within a disk subsystem to optimize application performance. The placement decisions are based on learning the data access patterns, and can be changed dynamically and transparently using this data.

XRC mirrors the data from the primary to the secondary disk subsystem. However, the Easy Tier learning information is not included in the XRC scope. The secondary disk subsystems are optimized according to the workload on these subsystems, which is different than the activity on the primary (there is only write workload on the secondary whereas there is read/write activity on the primary). And there is little activity on the tertiary disk (FlashCopy target disk), so it will be optimized differently than the primary disk or the secondary disk. As a result of these differences, during a recovery, the disks that you recover on (secondary or tertiary) are likely to display different performance characteristics compared to the former primary.

Easy Tier Heat Map Transfer is the DS8000 capability to transfer the Easy Tier learning from a XRC primary disk to a target set of disks. With GDPS/XRC, the Easy Tier learning can be transferred to either the secondary disk or the tertiary disk so that the disk that you recover on can also be optimized based on this learning, and will have similar performance characteristics to the former primary.

GDPS integrates support for Heat Map Transfer. The appropriate Heat Map Transfer actions (such as start/stop of the processing and reversing transfer direction) are incorporated into the GDPS managed processes. For example, if XRC is temporarily suspended for a planned or unplanned secondary disk outage, Heat Map Transfer is also suspended.

5.7 Flexible testing

Configuring point-in-time copy (FlashCopy) capacity in your XRC environment provides two main benefits:

- ▶ You can conduct regular DR drills or other tests using a copy of production data while production continues to run.
- ▶ You can save a consistent, “golden” copy of the XRC data which can be used if the primary disk or site is lost during an XRC resynchronization operation.

FlashCopy and the various options related to FlashCopy are discussed in 2.6, “FlashCopy” on page 38. GDPS/XRC supports taking a FlashCopy of either the current primary or the current secondary disks. The COPY, NOCOPY, NOCOPY2COPY and INCREMENTAL options are supported. Zero Suspend FlashCopy is supported in conjunction with COPY, NOCOPY, and INCREMENTAL FlashCopy.

FlashCopy can also be used, for example, to back up data without the need for extended outages to production systems, to provide data for data mining applications, and for batch reporting and so on.

Use of space-efficient FlashCopy

As discussed in “Space-efficient FlashCopy (FlashCopy SE)” on page 40, by using space-efficient (SE) FlashCopy volumes, you might be able to lower the amount of physical storage needed, and thereby reduce the cost associated with providing a tertiary copy of the data. GDPS provides support allowing space-efficient FlashCopy volumes to be used as FlashCopy target disk volumes. Whether a target device is space-efficient or not is transparent to GDPS; if any of the FlashCopy target devices defined to GDPS are space-efficient volumes, GDPS will simply use them. All GDPS FlashCopy operations with the NOCOPY option, whether through GDPS scripts, panels, or FlashCopies automatically taken by GDPS, can use space-efficient targets.

Understand the characteristics of space-efficient FlashCopy to determine whether this method of creating a point-in-time copy will satisfy your business requirements. For example, will it be acceptable to your business if, because of some unexpected workload condition, the repository on the disk subsystem for the space-efficient devices gets full and your FlashCopy is invalidated so that you are unable to use it? If your business requirements dictate that the copy must always be guaranteed to be usable, space-efficient might not be the best option and you can consider using standard FlashCopy instead.

5.8 GDPS tools for GDPS/XRC

GDPS provides tools that offer function that is complementary to GDPS function. The tools represent the kind of function that many clients are likely to develop themselves to complement GDPS. Using the tools provided by GDPS might eliminate the necessity for you to develop similar function yourself. The tools are provided in source code format, which means that if the tool does not completely meet your requirements, you can modify the code to tailor it to your needs.

The GDPS/XRC Performance Toolkit is included with GDPS/XRC. This suite of programs complements the XRC Performance Monitor product (XPM). The tools help with implementation, monitoring, and maintenance of z/OS Global Mirror (XRC) systems. These programs are intended for use by GDPS administrators, storage administrators, and capacity planning staff.

5.9 Services component

As you have seen, GDPS touches on much more than simply remote copy. It also includes sysplex, automation, database management and recovery, testing processes, and disaster recovery processes, to name just some of the areas it touches on.

Most installations do not have all these skills readily available. And it is extremely rare to find a team that had this range of skills across many implementations. However, the GDPS/XRC offering includes just that: access to a global team of specialists in all the disciplines you need to ensure a successful GDPS/XRC implementation.

Specifically, the Services component includes some or all of the following items:

- ▶ Planning to determine availability requirements, configuration recommendations, implementation and testing plans. Planning session topics include hardware and software requirements and prerequisites, configuration and implementation considerations, cross-site connectivity planning and potentially bandwidth sizing, and operation and control.
- ▶ Assistance in defining Recovery Point and Recovery Time objectives.
- ▶ Installation and necessary customization of NetView and System Automation.
- ▶ Remote copy implementation.
- ▶ IBM Virtualization Engine TS7700 implementation.
- ▶ GDPS/XRC automation code installation and policy customization.
- ▶ Education and training on GDPS/XRC setup and operations.
- ▶ Onsite implementation assistance.
- ▶ Project management and support throughout the engagement.

The sizing of the Services component of each project is tailored for that project, based on many things including what automation is already in place, whether remote copy is already in place, and so on. This means that the skills provided are tailored to the specific needs of each specific implementation.

5.10 GDPS/XRC prerequisites

Important: For more information about the latest GDPS/XRC prerequisites, see the following GDPS website:

<http://www.ibm.com/systems/z/advantages/gdps/getstarted/gdpsxrc.html>

5.11 Comparison of GDPS/XRC versus other GDPS offerings

So many features and functions are available in the various members of the GDPS family that recalling them all and remembering which offerings support them is sometimes difficult. To position the offerings, Table 5-1 lists the key features and functions and indicates which ones are delivered by the various GDPS offerings.

Table 5-1 Supported features matrix

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/MTMM	GDPS Virtual Appliance	GDPS/XRC	GDPS/GM
Continuous availability	Yes	Yes	Yes	Yes	No	No
Disaster recovery	Yes	Yes	Yes	Yes	Yes	Yes
CA/DR protection against multiple failures	No	No	Yes	No	No	No
Continuous Availability for foreign z/OS systems	Yes with z/OS Proxy	No	No	No	No	No
Supported distance	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	Virtually unlimited	Virtually unlimited
Zero Suspend FlashCopy support	Yes, using Consistent	Yes, using Consistent for secondary only	Yes, using Consistent	No	Yes (using Zero Suspend FlashCopy)	Yes, using CGPause
Reduced impact initial copy/resync	Yes	Yes	Yes	Yes	Not applicable	Not applicable
Tape replication support	Yes	No	No	No	No	No
Production sysplex automation	Yes	No	Yes	Not applicable	No	No

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/MTMM	GDPS Virtual Appliance	GDPS/XRC	GDPS/GM
Span of control	Both sites	Both sites (disk only)	Both sites	Both sites	Recovery site	Disk at both sites; recovery site (CBU or LPARs)
GDPS scripting	Yes	No	Yes	Yes	Yes	Yes
Monitoring, alerting and health checks	Yes	Yes	Yes	Yes (except health checks)	Yes	Yes
Query Services	Yes	Yes	No	No	Yes	Yes
MSS support for added scalability	Yes (secondary in MSS1)	Yes (secondary in MSS1)	Yes (H2 in MSS1, H3 in MSS2)	No	No	Yes (GM FC and Primary for MGM in MSS1)
MGM 3-site and 4-site	Yes (all configurations)	Yes (3-site only and non-IR only)	Yes (all configurations)	No	Not applicable	Yes (all configurations)
MzGM	Yes	Yes	Yes (non-IR only)	No	Yes	Not applicable
Open LUN	Yes	Yes	No	No	No	Yes
z/OS equivalent function for Linux for IBM z Systems	Yes	No	Yes (Linux for IBM z Systems running as a z/VM guest only)	Yes (Linux for IBM z Systems running as a z/VM guest only)	Yes	Yes
Heterogeneous support through DCM	Yes (VCS and SA AppMan)	No	No	No	Yes (VCS only)	Yes (VCS and SA AppMan)
z/BX hardware management	Yes	No	No	No	No	No
Web graphical interface	Yes	Yes	No	Yes	No	Yes

5.12 Summary

GDPS/XRC is a powerful offering that provides an industry leading, long distance, disaster recovery capability. It is based on the XRC technology, which is highly scalable (there are clients with close to 20,000 volumes being remote copied by XRC). XRC is industry-proven, having been available for well over a decade. XRC also has interoperability advantages: it is possible to have different disk subsystem types, and even different vendors, for the primary and secondary devices.

Building on the base of XRC, GDPS adds the powerful script capability that allows you to perfect the actions to be taken, either for planned or unplanned changes, eliminating the risk of human error. Combining its support of FlashCopy with the scripting capabilities significantly reduces the time and complexity to set up a disaster recovery test. And anyone who has been involved in DR planning will confirm that one of the most important factors in a successful disaster recovery process is frequent and realistic testing that is tied into your change management system. Having the ability to test your DR capability any time a significant change is implemented ensures that *all* aspects of application management are addressed.

In addition to its disaster recovery capability, GDPS/XRC also provides a much more user-friendly interface for monitoring and managing the remote copy configuration. This includes the initialization and monitoring of the XRC volume pairs based upon policy and performing routine operations on installed storage subsystems.



GDPS/Global Mirror

In this chapter, we discuss the capabilities and prerequisites of the GDPS/Global Mirror (GM) offering.

The GDPS/GM offering provides a disaster recovery capability for businesses that have an RTO of as little as two hours or less, and an RPO as low as five seconds. It will typically be deployed in configurations where the application and recovery sites are more than 200 km apart and want to have integrated remote copy processing for mainframe and non-mainframe data.

The functions provided by GDPS/GM fall into two categories:

- ▶ Protecting your data:
 - Protecting the integrity of the data on the secondary data in the event of a disaster or suspected disaster.
 - Managing the remote copy environment through GDPS scripts and NetView panels or the web interface.
 - Optionally supporting remote copy management and consistency of the secondary volumes for Fixed Block (FB) data. Depending on your application requirements, the consistency of the FB data can be coordinated with the CKD data.
- ▶ Controlling the disk resources managed by GDPS during normal operations, planned changes, and following a disaster:
 - Support for recovering the production environment following a disaster.
 - Support for switching your data and systems to the recovery site.
 - Support for testing recovery and restart using a practice FlashCopy point-in-time copy of the secondary data while live production continues to run in the application site and continues to be protected with the secondary copy.

6.1 Introduction to GDPS/Global Mirror

GDPS/GM is a disaster recovery solution. It is similar in various respects to GDPS/XRC in that it supports virtually unlimited distances. However, the underlying IBM Global Mirror (GM), remote copy technology also supports both z Systems CKD data and distributed data, and GDPS/GM also includes support for both.

GDPS/GM can be viewed as a mixture of GDPS/PPRC and GDPS/XRC. Just as PPRC (IBM Metro Mirror) is a disk subsystem-based remote copy technology, GM is also disk-based, which means that it supports the same mix of CKD and FB data that is supported by GDPS/PPRC. Also, being disk-based, there is no requirement for a System Data Mover (SDM) system to drive the remote copy process. Also, like PPRC, Global Mirror requires that the primary and secondary disk subsystems are from the same vendor.

Conversely, GDPS/GM resembles GDPS/XRC in that it is asynchronous and supports virtually unlimited distances between the application and recovery sites. Also, similar to GDPS/XRC, GDPS/GM does not provide any automation or management of the production systems. Instead, its focus is on managing the Global Mirror remote copy environment and automating and managing recovery of data and systems in case of a disaster. Like GDPS/XRC, GDPS/GM supports the ability to remote copy data from multiple systems and sysplexes. In contrast, each GDPS/PPRC installation supports remote copy for only a single sysplex.

The capabilities and features of GDPS/GM are described in this chapter.

6.1.1 Protecting data integrity

Because the role of GDPS/GM is to provide disaster recovery support, its highest priority is protecting the integrity of the data, CKD and FB, in the recovery site. This section discusses the support provided by GDPS for these various data types.

Traditional z Systems (CKD) data

As described in 2.4.3, “Global Mirror” on page 32, Global Mirror protects the integrity of the remote-copied data by creating consistency groups, either continuously or at intervals specified by the installation. The whole process is managed by the Master disk subsystem, based on the GDPS/GM configuration. There are no restrictions relating to which operating systems’ data can be supported; any system that writes to CKD devices (z/OS, z/VM, z/VSE, and Linux for System z) is supported. Regardless of which systems are writing to the devices, all management control is from the z/OS system that is running the GDPS/GM local controlling system, also known as the *K-sys*.

How frequently a consistency group can be created depends on the bandwidth provided between the application and recovery site disks; IBM can perform a bandwidth analysis for you to help you identify the required capacity.

GDPS/Global Mirror uses devices in the primary and secondary disk subsystems to execute the commands to manage the environment. Some of these commands directly address a primary device, whereas others are directed to the LSS. To execute these LSS-level commands, you must designate at least one volume in each primary LSS as a GDPS *utility device*, which is the device that serves as the “go between” between GDPS and the LSS. These utility devices do not need to be dedicated devices; that is, they can be one of the devices that are being mirrored as part of your Global Mirror session. In fact, the utility devices also need to be mirrored.

Global Mirror supports both CKD and FB devices. If the CKD and FB devices are in the same Global Mirror session, they will be in the same consistency group. This means that they must be recovered together, which also means that the systems that use these disks must also be recovered together.

Distributed (FB) data

GDPS/GM supports remote copy of FB devices written by distributed systems (including SCSI-attached FB disks written by Linux on z Systems). If the FB devices are in the same Global Mirror session as the CKD devices that are being global mirrored, they will have the same consistency point. If they are in a different Global Mirror session than the CKD disks, they will have a different consistency point. That is, if CKD and FB disks are in different sessions, the data for each session will be consistent within itself, but the data for the two sessions will not be consistent with each other.

There are certain disk subsystem microcode requirements needed to enable GDPS/GM management of FB disks. See “Open LUN (FB disk) management prerequisites” on page 297 for details.

The Distributed Cluster Management (DCM) capability of GDPS can be used to manage the nodes in distributed system clusters that use the replicated data. See 10.3, “Distributed Cluster Management” on page 307 for more information about DCM.

6.2 GDPS/Global Mirror configuration

At its most basic, a GDPS/GM configuration consists of one or more production systems, an application site controlling system (K-sys), a recovery site controlling system (R-sys), primary disks, and two sets of disks in the recovery site.

The GM copy technology uses three sets of disks. 2.4.3, “Global Mirror” on page 32 includes an overview of how GM works and how the disks are used to provide data integrity.

The K-sys is responsible for controlling all remote copy operations and for sending configuration information to the R-sys. In normal operations, most operator and system programmer interaction with GDPS/GM would be through the K-sys. The K-sys role is simply related to remote copy; it does not provide any monitoring, automation, or management of systems in the application site, nor any FlashCopy support for application site disks. There is no requirement for the K-sys to be in the same sysplex as the system or systems it is managing data for. In fact, the K-sys should be placed in a monoplex on its own.

You can also include the K-sys disks in the GDPS managed GM configuration and replicate them. The K-sys does not have the isolation requirements of the controlling system in a GDPS/PPRC configuration.

The R-sys is primarily responsible for validating the configuration, monitoring the GDPS managed resources such as the disks in the recovery site, and carrying out all recovery actions either for test purposes or in the event of a real disaster. See 6.7, “Flexible testing” on page 183 for more information about testing using FlashCopy.

The K-sys and R-sys communicate information to each other using a NetView-to-NetView network communication mechanism over the wide area network (WAN). K-sys and R-sys are dedicated to their roles as GDPS controlling systems.

GDPS/GM can control multiple Global Mirror sessions. Each session can consist of a maximum of 17 disk subsystems (combination of primary and secondary). All the members of the same session will have the same consistency point. Typically the data for all systems that must be recovered together will be managed through one session. For example, a z/OS sysplex is an entity where the data for all systems in the sysplex need to be in the same consistency group. If you have two production sysplexes under GDPS/GM control, the data for each can be managed through a separate GM session, in which case they can be recovered individually. You can also manage the entire data for both sysplexes in a single GM session, in which case if one sysplex fails and you have to invoke recovery, you will need to also recover the other sysplex.

Information about which disks are to be mirrored as part of each session and the intervals at which a consistency point is to be created for each session is defined in the GDPS remote copy configuration definition file (GEOMPARM). GDPS/GM uses this information to control the remote copy configuration. Like the other GDPS offerings, the NetView panel interface (or the web interface) is used as the operator interface to GDPS.

Although the panel interface or web interface support management of GM, they are primarily intended for viewing the configuration and performing some operations against single disks. GDPS scripts are intended to be used for actions against the entire configuration because this is much simpler (with multiple panel actions combined into a single script command) and less error-prone.

The actual configuration depends on your business and availability requirements, the amount of data you will be remote copying, the types of data you will be remote copying (only CKD or both CKD and FB), and your RPO.

Figure 6-1 shows a typical GDPS/GM configuration.

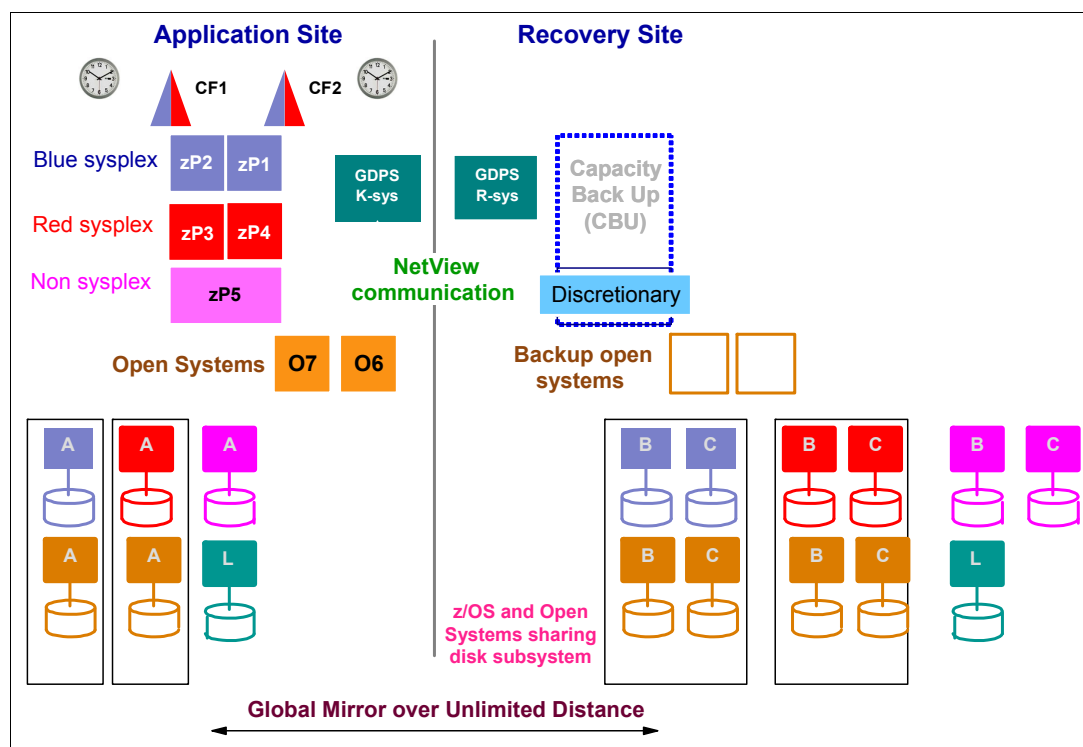


Figure 6-1 GDPS/GM configuration

The application site, as shown in the figure, contains these items:

- ▶ z/OS systems spread across several sysplexes
- ▶ A non-sysplexed z/OS system
- ▶ Two distributed systems
- ▶ The K-sys
- ▶ The primary disks (identified by A)
- ▶ The K-sys' own disks (marked by L)

The recovery site contains these items:

- ▶ The R-sys
- ▶ A CPC with the CBU feature that also contains expendable workloads that can be displaced
- ▶ Two backup distributed servers
- ▶ The Global Mirror secondary disks (marked by B)
- ▶ The Global Mirror FlashCopy targets (marked by C)
- ▶ The R-sys disks (marked by L)

Although there is great flexibility in terms of the number and types of systems in the application site, several items are fixed:

- ▶ All the GM primary disks and the K-sys *must* be in the application site¹.
- ▶ All the GM secondary disks, the FlashCopy targets used by GM, and the GDPS R-sys *must* be in the recovery site².

The following aspects in GDPS/GM differ from the other GDPS offerings:

- ▶ Although the K-sys should be dedicated to its role as a controlling system, it is not necessary to provide the same level of isolation for the K-sys as that required in a GDPS/PPRC or GDPS/HM configuration.
- ▶ GDPS/XRC, because of XRC time stamping, requires that all the systems writing to the primary disks have a common time source (Sysplex Timer or STP). GDPS/GM does not have this requirement.
- ▶ With GDPS/XRC, if there is insufficient bandwidth for XRC operations, writes to the primary disk subsystem will be paced. This means that the RPO will be maintained, but at the potential expense of performance of the primary devices.

With GDPS/GM, if there is insufficient bandwidth, the consistency points will fall behind. This means that the RPO might not be achieved, but performance of the primary devices will be protected.

In both cases, if you want to protect both response times and RPO, you must provide sufficient bandwidth to handle the peak write load.

The GDPS/GM code itself runs under NetView and System Automation, and is run *only* in the K-sys and R-sys.

¹ The application site is where production applications whose data is to be mirrored normally run, and it is the site where the Global Mirror primary disks are located. You might also see this site referred to as the *local site* or the *A-site*.

² The recovery site is where the mirrored copies of the production disks are located, and it is the site to which production systems are failed over in the event of a disaster. You might also see this site referred to as the *remote site* or the *R-site*.

GDPS/GM multiple R-sys collocation

GDPS/GM can support multiple sessions. Therefore, the same instance of GDPS/GM can be used to manage GM replication and recovery for several diverse sysplexes and systems. However, there are certain cases where different instances of GDPS/GM are required to manage different sessions. One example is the GDPS/GM leg of a GDPS/MGM configuration: in such a configuration, GDPS/GM is restricted to managing only one single session. Clients might have other requirements based on workloads or organizational structure for isolating sessions to be managed by different instances of GDPS/GM.

When you have multiple instances of GDPS/GM, each instance will need its own K-sys. However, it is possible to combine the R-sys “functions” of each instance to run in the same z/OS image. Each R-sys function would run in a dedicated NetView address space in the same z/OS. Actions, such as running scripts, can be done simultaneously in these NetView instances. This reduces the overall cost of managing the remote recovery operations for customers that require multiple GDPS/GM instances.

6.2.1 GDPS/GM in a 3-site or 4-site configuration

GDPS/GM can be combined with GDPS/PPRC (or GDPS/HM) in a 3-site or 4-site configuration, where GDPS/PPRC (or GDPS/PPRC HM) is used across two sites within metropolitan distances (or even within a single site) to provide continuous availability through Parallel Sysplex use and GDPS HyperSwap, and GDPS/GM provides disaster recovery in a remote region.

We call this combination the GDPS/Metro Global Mirror (GDPS/MGM) configuration. In such a configuration, both GDPS/PPRC and GDPS/GM provide some additional automation capabilities.

GDPS/GM can also be combined with GDPS/MTMM in a GDPS/MGM Multi-target configuration in a similar manner. When combined with GDPS/MTMM, the environment can benefit from the Multi-Target PPRC technology which provides some advantages when compared to the combination with GDPS/PPRC.

After you understand the base capabilities described in 2.4.4, “Combining disk remote copy technologies for CA and DR” on page 35, see Chapter 11, “Combining local and metro continuous availability with out-of-region disaster recovery” on page 331 for more information about GDPS/MGM.

6.2.2 Other considerations

The availability of the GDPS K-sys in *all* scenarios is a fundamental requirement in GDPS. The K-sys monitors the remote copy process, implements changes to the remote copy configuration, and sends GDPS configuration changes to the R-sys.

Although the main role of the R-sys is to manage recovery following a disaster or to enable DR testing, it is important that the R-sys also be available at all times. This is because the K-sys sends changes to GDPS scripts and changes to the remote copy or remote site configuration to the R-sys at the time the change is introduced on the K-sys. If the R-sys is not available when such configuration changes are made, it is possible that it might not have the latest configuration information in the event of a subsequent disaster, resulting in an impact to the recovery operation.

Also, the R-sys plays a role in validating configuration changes. Therefore, it is possible that a change containing errors that would have been rejected by the R-sys (if it had been running) will not be caught. This, again, impacts the remote copy or recovery operation.

Because GDPS/GM is in essence a disaster recovery offering rather than a continuous availability offering, it does not support the concept of site switches that GDPS/PPRC provides³. It is expected that a switch to the recovery site will be performed only in case of a real disaster.

If you want to move operations back to the application site, you must either set up GDPS/GM in the opposite direction (which means that you will also need two sets of disks in the application site), or use an alternate mechanism, like Global Copy, outside the control of GDPS. If you intend to switch to run production in the recovery site for an extended period of time, then providing two sets of disks and running GDPS/GM in the reverse direction would be the preferable option to provide disaster recovery capability.

6.3 GDPS/GM management for distributed systems and data

As previously mentioned, it is possible for GDPS/GM to manage FB disks on behalf of distributed systems that use these disks, either in the same session as z Systems CKD disks or in a separate session. However, for these distributed systems, although GDPS/GM manages the remote copy and recovery of the disks, it is not able to perform any system recovery actions for the distributed systems in the recovery site.

As an alternative configuration, GDPS/GM also provides the Distributed Cluster Management (DCM) capability for managing global clusters using Veritas Cluster Server (VCS) through the Global Cluster Option (GCO). When the DCM capability is used, GDPS/GM does not manage remote copy or consistency for the distributed system disks (this is managed by VCS). Therefore, it is not possible to have a common consistency point between the z Systems CKD data and the distributed data.

However, for environments where a common consistency point is not a requirement, DCM with VCS does provide various key availability and recovery capabilities that might be of interest. DCM is described further in 10.3.2, “DCM support for VCS” on page 308.

GDPS/GM also provides the DCM capability for managing distributed clusters under IBM Tivoli System Automation Application Manager (SA AppMan) control. DCM provides advisory and coordinated functions between GDPS and SA AppMan-managed clusters. Data for the SA AppMan-managed clusters can be replicated using Global Mirror under GDPS control.

Thus, z/OS and distributed cluster data can be controlled from one point. Distributed data and z/OS data can be managed in the same consistency group (Global Mirror session) if cross-platform data consistency is required. Equally, z/OS and distributed data can be in different sessions and the environments can be recovered independently under GDPS control. For more information, see “Integrated configuration of GDPS/GM and SA AppMan” on page 327.

³ Region switches are supported by GDPS/MGM in an Incremental Resynch configuration.

6.4 Managing the GDPS environment

As previously mentioned, GDPS/GM automation code runs only in one system in the application site, the K-sys, and it does not provide for any monitoring or management of the production systems in this site. The K-sys has the following responsibilities:

- ▶ It is the primary point of GDPS/GM control for operators and system programmers in normal operations.
- ▶ It manages the remote copy environment. Changes to the remote copy configuration (adding new devices into a running GM session or removing devices from a running session) are driven from the K-sys.
- ▶ Changes to the configuration definitions or scripts (including configuration definitions for recovery site resources and scripts destined to be executed on the R-sys) are defined in the K-sys and automatically propagated to the R-sys.

In the recovery site, GDPS/GM runs only in one system: the R-sys. However, the role and capabilities of the R-sys are different from those of the K-sys. Even though both are GDPS controlling systems, there are fundamental differences between them. The R-sys has the following responsibilities:

- ▶ Validate the remote copy configuration in the remote site. This is a key role. GM is a hardware replication technology. Just because the GM primary disks can communicate to the GM secondary disks over remote copy links does not mean that in a recovery situation, systems can use these disks. The disks must be defined in that site's I/O configuration. If you are missing some disks, this can cause recovery to fail because you will not be able to properly restart systems that need those disks.
- ▶ Monitor the GDPS managed resources in the recovery site and raise alerts for not-normal conditions. For example, GDPS will use the BCP Internal Interface (BCPii) to perform hardware actions such as adding temporary CBU capacity to CPCs, deactivating LPARs for discretionary workloads, activating LPARs for recovery systems and so on. The R-sys monitors that it has BCPii connectivity to all CPCs that it will need to perform actions against.
- ▶ Communicate status and alerts to the K-sys that is the focal management point during normal operations.
- ▶ Automate reconfiguration of the recovery site (recovering the Global Mirror, taking a FlashCopy, activating CBU, activating backup partitions and so on) for recovery testing or in the event of a true disaster.

The R-sys has no relation whatsoever to any application site resources. The only connection it has to the application site is the network connection to the K-sys for exchanging configuration and status information.

6.4.1 NetView panel interface

The operator interface for GDPS/GM is provided through NetView 3270 or the GDPS web interface (described in “Web graphical user interface” on page 172), which is also based on NetView facilities. In normal operations the operators interact mainly with the K-sys, but there is also a similar set of interfaces for the R-sys.

The NetView interface for GDPS consists of two parts. The first and potentially the most important part is the Status Display Facility (SDF). Any time that there is a change of status to something that GDPS does not consider “normal” and that can affect the ability to recover, so it is something that requires investigation and manual intervention, GDPS sends an alert to SDF.

SDF provides a dynamically updated color-coded panel that provides the status of the systems and highlights any problems in the remote copy configuration. If something changes in the environment that requires attention, the color of the associated field on the panel will change. K-sys sends alerts the R-sys and R-sys sends alerts to K-sys so that both controlling systems are aware of any problems at all times.

During normal operations, the operators should always have a K-sys SDF panel within view so they will immediately become aware of anything requiring intervention or action. When R-sys is being used for managing testing or recovery operations, then operators should have access to the R-sys SDF panel also.

The other part of the NetView interface consists of the panels provided by GDPS to help you manage and inspect the environment. The main GDPS panel is shown in Figure 6-2. Notice that some of the options are not enabled (options 2 and 7, which are colored in blue). This is because those functions are not part of GDPS/GM.

From this panel, you can perform the following actions:

- ▶ Query and control the disk remote copy configuration.
- ▶ Initiate GDPS standard actions (the ability to control and initiate actions against LPARs):
 - On the K-sys, the only standard action supported is the ability to update IPL information for the recovery site LPARs.
 - On the R-sys, all standard actions are available.
- ▶ Initiate GDPS scripts (Planned Actions).
- ▶ Manage GDPS Health Checks.
- ▶ View and refresh the definitions of the remote copy configuration.
- ▶ Run GDPS monitors.

VPCPPNLM		GDPS Global Mirror		Local		GDPS V3.R10.M0	
System	=	G9C1	A6P91	Mirroring	=	OK	
Current Master	=	G9C1	A6P91	Dasd Config	=	2013-02-07	16:00:01
Debug	=	ON					
1	Dasd Remote Copy						
2	Tape Remote Copy						
3	Standard Actions						
6	Planned Actions						
7	Sysplex Resource Management						
8	Debug ON/OFF						
9	View Definitions						
H	Health Checks						
C	Config Management						
M	Run Monitor1/Monitor3						
Selection ==> _							
F1=Help		F3=Return		F6=Roll			

Figure 6-2 GDPS Main panel (K-sys)

Web graphical user interface

The web interface is a browser-based interface designed to improve operator productivity. The web interface provides the same functional capability as the 3270-based panel, such as providing management capabilities for Remote Copy Management, Standard Actions, Sysplex Resource Management, and SDF Monitoring using simple point-and-click procedures. In addition, users can open multiple windows to allow for continuous status monitoring while performing other GDPS/GM management functions.

The web interface display has three sections:

- ▶ A menu bar on the left with links to the main GDPS options
- ▶ A window list on top allowing switching between multiple open frames
- ▶ An active task frame where the relevant information is displayed and activities are performed for a selected option

The main status panel of the GDPS/GM web interface is shown in Figure 6-3. Various selectable options are listed below the GDPS Mirror Links section in the left frame. These options can be displayed at all times, or you can optionally collapse the frame.

The screenshot displays the Tivoli NetView for z/OS web interface. The top header bar shows the Tivoli logo and the text 'NetView for z/OS'. Below the header, there is a navigation bar with 'About' and 'Status Menu' tabs. The main content area is divided into two sections. On the left, there is a 'My Tasks' section with a list of links including 'Browse Events', 'Browse Sysplex Topology', 'Browse Logs', 'Open Command Console', 'Open NetView Help', 'Launch Procedures', 'View DVIPA Status', 'View DVIPA Distributors', 'Manage TCP/IP Connections', 'Manage IP Packet Trace Data', 'Load/Unload SNMP MIBs', 'Launch MIB Browser', 'Launch Real Time Poller', 'Issue SNMP Commands', 'Open 3270 Console', and a 'GDPS Global Mirror Links' section with links like 'Status Menu', 'Dasd Remote Copy', 'Standard Actions', 'Planned Actions', 'Debug ON/OFF', 'View Definitions', 'Config Management', 'Monitors', 'View GEOMPARM', 'Set User Preferences', and 'About'. The right section is titled 'Status Menu' and contains a table with the following data:

GDPS Global Mirror - Status Menu		LOCAL	WTORs	SDE	GDPS GM V3.R7.M0
LEXA	5 Feb 2010	17:20:14	A6P11	G1C1	GDPS Page: VPCWGMEN

Below the table, there are buttons for 'Refresh', 'AutoRefresh ON', 'Logoff', and 'Help'. Underneath, there is a 'Status Information' section with a table showing the following data:

System/Domain	G1C1 / A6P11	Mirroring	OK
Current Master	G1C1 / A6P11	Dasd Config	2010-02-05 08:49:19
DEBUG	ON	WebDEBUG	NOT ACTIVE

The 'AutoRefresh is: OFF' status is displayed at the bottom right of the 'Status Information' section.

Figure 6-3 Full view of the GDPS main panel with taskbar and status information

Main Status panel

The GDPS web interface status frame shown in Figure 6-4 is the equivalent to the main GDPS panel. The information shown on this panel is what is found on the top portion of the 3270 GDPS Main panel.

GDPS Global Mirror - Status Menu		LOCAL	WTORs	SDF	GDPS GM V3.R7.M0
LEXA	5 Feb 2010	17:20:14	A6P11	G1C1	GDPS Page: VPCWGMEN

Status Information

AutoRefresh is: OFF

System/Domain	G1C1 / A6P11	Mirroring	OK
Current Master	G1C1 / A6P11	Dasd Config	2010-02-05 08:49:19
DEBUG	ON	WebDEBUG	NOT ACTIVE

Figure 6-4 GDPS web interface: Main status panel

Remote copy panels

Although Global Mirror is a powerful copy technology, the z/OS operator interface to it is not particularly intuitive. To make it easier for operators to check and manage the remote copy environment, use the Disk Remote Copy panels provided by GDPS.

For GDPS to manage the remote copy environment, you first define the configuration to GDPS in the GEOMPARM file on the K-sys. The R-sys always gets the configuration information from the K-sys and validates the remote site disk configuration.

After the configuration is known to GDPS, you can use the panels to check that the current configuration matches the one you want. You can start, stop, pause, and resynch mirroring. These actions can be done at the device, LSS, or session level, as appropriate. However, we suggest that GDPS control scripts are used for actions at the session level.

Figure 6-5 shows the mirroring status panel for GDPS/GM as viewed on the K-sys. The panel for the R-sys is similar, except that the R-sys can perform only a limited number of actions (typically only those necessary to take corrective action) against the devices in the recovery site. Control of the GM session can be done only from the K-sys; the R-sys can control only the devices in the recovery site.

GDPS - Global Mirror		LOCAL	WTORs	SDE	GDPS GM V3.R7.M0
LEXA	9 Feb 2010	11:29:22	A6P11	G1C1	GDPS Page: VPCWGDRZ

Defined sessions in this configuration. Last Monitor2 run: 01:00:05

Monitor 2

Session Num Name	GM	Dasd	IR	Last Consistency Time
01 ZPROD01	RUNNING	OK		2010-02-09 10:32:34

Figure 6-5 Disk Mirroring panel for GDPS/GM K-sys

Remember that these panels provided by GDPS are *not* intended to be a remote copy monitoring tool. Because of the overhead involved in gathering information about each and every device in the configuration to populate the NetView panels, GDPS gathers this information only on a timed basis, or on demand following an operator instruction. The normal interface for finding out about remote copy problems is the Status Display Facility, which is dynamically updated if or when a problem is detected.

Standard Actions

As previously explained, the K-sys does not provide any management functions for any systems, either in the application site or in the recovery site. The R-sys manages recovery in the recovery site. As a result, the Standard Actions that are available vary, depending which type of controlling system you are using.

Because the R-sys manages the recovery in the event of a disaster (or IPL for testing purposes) of the production systems in the recovery site, it has a wider range of functions available, as seen in Figure 6-7. Functions are provided to activate and deactivate LPARs, to IPL and reset systems, and to update the IPL information for each system.

Standard Actions

GDPS - Standard Actions G1P1

REMOTE

WTQRs

SDE

GDPS GM V3.R7.M0

LEXA

9 Feb 2010

11:43:08

A6P13

G1C2

GDPS Page: VPCWPFYX

Return

Refresh

Print Screen

Logoff

Help

Load

Reset

Activate LPAR

Deactivate LPAR

Selected IPLtype/mode

System	IPLtype	IPLmode	LoadAdr	LoadParm	Status
G1P1	NORMAL	BDISK	3100	3108M1M	MANUAL

Modify IPLtype/IPLmode

Select

Modify

New

Delete

Optional IPLtypes/modes for system: G1P1

IPLtype	IPLmode	LoadAdr	LoadParm
NORMAL	BDISK	3100	3108M1M
FC1	FC1DISK	3160	3168M1M
NOTRANSL	NOTR	5432	5438M1M

Figure 6-7 Example GDPS/GM R-sys Standard Actions panel for a selected system

There are actually two types of resource-altering actions you can initiate from the panels: *Standard Actions* and *Planned Actions*. Standard Actions are really single steps and will affect one resource, such as deactivating an LPAR after a DR test.

For example, if you want to (1) reset an expendable test system running in the recovery site; (2) deactivate the LPAR of the expendable system; (3) activate the recovery LPAR for a production system; and then (4) IPL the recovery system into the LPAR you just activated, this task consists of four separate Standard Actions that you initiate sequentially from the panel.

GDPS scripts

Nearly all the functions that can be initiated through the panels (and more) are also available from GDPS scripts. A script is a program consisting of one or more GDPS functions to provide a workflow. In addition to the low-level functions available through the panels, scripts can invoke functions with a single command that might require multiple separate steps if performed through the panels. For example, if you have a new disk subsystem and will be adding several LSS, populated with a large number of devices to your Global Mirror configuration, this can require a significantly large number of panel actions.

In comparison, it can be accomplished by a single script command. It is simply faster and more efficient to perform compound/complex operations using scripts.

Scripts can be initiated manually through the GDPS panels or through a batch job. In GDPS/GM, the only way to initiate the recovery of the secondary disks is through a GDPS script on the R-sys; invoking a recovery directly from the mirroring panels is not supported.

Scripts are written by you to automate the handling of certain situations, both planned changes and also error situations. This is an extremely important aspect of GDPS.

Scripts are powerful because they can access the full capability of GDPS. The ability to invoke all the GDPS functions through a script provides the following benefits:

- ▶ **Speed**

The script will execute the requested actions as quickly as possible. Unlike a human, it does not need to search for the latest procedures or the commands manual.

- ▶ **Consistency**

If you were to look into most computer rooms immediately following a system outage, what would you see? Mayhem! Operators will be frantically scrambling for the latest system programmer instructions. All the phones will be ringing. Every manager within reach will be asking when the service will be restored. And every system programmer with access will be vying for control of the keyboards. All this results in errors, because humans often make mistakes when under pressure. But with automation, your well-tested procedures will execute in exactly the same way, time after time, regardless of how much you shout at them.

- ▶ **Automatic checking of results from commands**

Because the results of many GDPS commands can be complex, manual checking of results can be time-consuming and presents the risk of missing something. In contrast, scripts automatically check that the preceding command (remember, that one command can have been six GM commands, each of the six executed against thousands of devices) completed successfully before proceeding with the next command in the script.

- ▶ **Thoroughly tested procedures**

Because scripts behave in a consistent manner, you can test your procedures over and over until you are sure they do everything that you want, in exactly the manner that you want. Also, because you need to code everything and cannot assume a level of knowledge (as you might with instructions intended for a human), you are forced to thoroughly think out every aspect of the action the script is intended to undertake. Finally, because of the repeatability and ease of use of the scripts, they lend themselves more easily to frequent testing than manual procedures.

Planned Actions

GDPS scripts can be initiated from the Planned Actions option on the GDPS main panel. In a GDPS/GM environment, all actions affecting the recovery site are considered planned actions. You can think of this as pre-planned unplanned actions. An example of a planned action in GDPS/GM is a script that prepares the secondary disks and LPARs for a disaster recovery test.

Such a script will do the following actions:

- ▶ Recover the disks in the disaster site; this makes the B disks consistent with the C disks. The B disks will be used for the test; the C disks contain a consistent copy that ages during the test.
- ▶ Activate CBU capacity in the recovery site CPCs.

- ▶ Activate backup partitions that have been predefined for the recovery systems (that is, the production systems running in the recovery site).
- ▶ Activate any backup coupling facility partitions in the recovery site.
- ▶ Load the systems into the partitions in the recovery site using the B disks.

When the test is complete, you run another script in the R-sys to perform the following tasks:

- ▶ Reset the recovery systems that were used for the test
- ▶ Deactivate the LPARs that were activated for the test.
- ▶ Undo CBU on the recovery site CPCs.
- ▶ Issue a message to the operators to manually shut down any open systems servers in the recovery site that were used for the test.
- ▶ Bring the B disks back into sync with the C disks (which are consistent with the primary disks at the time of the start of the test).
- ▶ Finally, you run a script on the K-sys to resynchronize the recovery site disks with the production disks.

Batch scripts

In addition to the ability to initiate GDPS scripts from the GDPS panel interfaces, it is also possible to initiate a script from a batch interface. This is especially suited to processes that are run regularly, and have some interaction with the GDPS environment.

6.4.2 System Management actions

In a GDPS/GM environment, the remote controlling system can use the hardware and system management actions to reconfigure the recovery site by adding temporary capacity, activating backup partitions, and IPLing production systems. This can be either for test purposes or for a real recovery. GDPS does not manage the systems or the hardware in the application site.

Most GDPS Standard Actions and many GDPS script commands require actions to be done on the HMC. The interface between GDPS and the HMC is through the BCP Internal Interface (BCPii), and it allows GDPS to communicate directly with the hardware for automation of HMC actions such as LOAD, RESET, and Activate or Deactivate an LPAR, and Activate or Undo CBU or OOCOD.

The GDPS LOAD and RESET Standard Actions (available through the Standard Actions panel or the SYSPLEX script statement) allow specification of a CLEAR or NOCLEAR operand. This provides the operational flexibility to accommodate client procedures.

Extensive facilities for adding temporary processing capacity to the CPCs in the recovery site are provided by the GDPS scripting capability.

6.5 GDPS/GM monitoring and alerting

We discuss the GDPS SDF panel in 6.4.1, “NetView panel interface” on page 170. This is the panel on which GDPS dynamically displays alerts, which are color-coded based on severity, if and when a non-normal status or situation is detected.

Alerts can be posted as a result of an unsolicited error situation that GDPS listens for. For example, if there is a problem with the GM session and the session suspends outside of GDPS control, GDPS will be aware of this because the disk subsystem that is the Master for the GM session will post an SNMP alert. GDPS listens for these SNMP alerts and, in turn, posts an alert on the SDF panel that notifies the operator of the suspension event.

Alerts can also be posted as a result of GDPS periodically monitoring key resources and indicators that relate to the GDPS/GM environment. If any of these monitoring items are found to be in a state deemed to be not normal by GDPS, an alert is posted on SDF. Because the K-sys and R-sys have different roles and affect different resources, they each monitor a different set of indicators and resources.

For example, the K-sys has TCP/IP connectivity to the A disk through which the GM Master disk subsystem posts SNMP alerts about GM problems. For this reason, it is important that the TCP/IP connectivity between the K-sys and the production disk is functioning properly. The K-sys, among other things, monitors this connection to ensure that it is functional so that if there is a GM problem, the SNMP alert will reach the K-sys.

Likewise, it is the R-sys that uses the BCP Internal Interface to perform hardware actions to reconfigure the recovery site, either for disaster testing or in the event of a real recovery scenario. One of the resources monitored by the R-sys is the BCP Internal Interface connection to all CPCs in the recovery site on which the R-sys can perform hardware operations such as CBU or LPAR activation.

Both the K-sys and the R-sys, in addition to posting alerts on their own SDF panel, will also forward any alerts to the other system for posting. Because the operator will be notified of R-sys alerts on the K-sys SDF panel, it is sufficient for the operator to monitor the K-sys SDF panel during normal operations if the K-sys is up and running.

If an alert is posted, the operator has to investigate (or escalate, as appropriate) and corrective action needs to be taken for the reported problem as soon as possible. After the problem is corrected, this is detected during the next monitoring cycle and the alert is cleared by GDPS automatically.

GDPS/GM monitoring and alerting capability is intended to ensure that operations are notified and can take corrective action for any problems in their environment that can affect the ability of GDPS/GM to do recovery operations. This maximizes the installation's chance of achieving RPO and RTO commitments.

6.5.1 GDPS/GM health checks

In addition to the GDPS/GM monitoring described, GDPS provides health checks. These health checks are provided as a plug-in to the z/OS Health Checker infrastructure to check that certain settings related to GDPS adhere to GDPS preferred practices.

The z/OS Health Checker infrastructure is intended to check a variety of settings to determine whether these settings adhere to z/OS preferred practices values. For settings that are found to be not in line with preferred practices, exceptions are raised in the Spool Display and Search Facility (SDSF). Many products including GDPS provide health checks as a plug-in to the z/OS Health Checker. There are various parameter settings related to GDPS, such as z/OS PARMLIB settings or NetView settings, and the recommendations and preferred practices for these settings have been documented in the GDPS publications. If these settings do not adhere to recommendations, this can hamper the ability of GDPS to perform critical functions in a timely manner.

Although GDPS monitoring will detect that GDPS was not able to perform a particular task and raise an alert, the monitor alert might be too late, at least for that particular instance of an incident. Often, if there are changes in the client environment, this might necessitate adjustment of some parameter settings associated with z/OS, GDPS, and other products. It is possible that you can miss making these adjustments, which might result in affecting GDPS. The GDPS health checks are intended to detect such situations and avoid such incidents where GDPS is unable to perform its job, because of a setting that is perhaps less than ideal.

For example, there are several address spaces associated with GDPS/GM and preferred practices recommendations are documented for these. GDPS code itself runs in the NetView address space and there are DFSMS address spaces that GDPS interfaces with to perform GM copy services operations. GDPS recommends that these address spaces are assigned specific Workload Manager (WLM) service classes to ensure that they are dispatched in a timely manner and do not lock each other out. One of the GDPS/GM health checks, for example, checks that these address spaces are set up and running with the characteristics recommended by GDPS.

Similar to z/OS and other products that provide health checks, GDPS health checks are optional. Several preferred practices values that are checked and the frequency of the checks can be customized to cater to unique client environments and requirements.

GDPS also provides a useful interface for managing the health checks using the GDPS panels. You can perform actions such as activate or deactivate or run any selected health check, view the customer overrides in effect for any preferred practices values, and so on. Figure 6-8 shows a sample of the GDPS Health Check management panel. In this example you see that all the health checks are enabled. The status of the last run is also shown, indicating whether the last run was successful or whether it resulted in an exception. Any exceptions can also be viewed using other options on the panel.

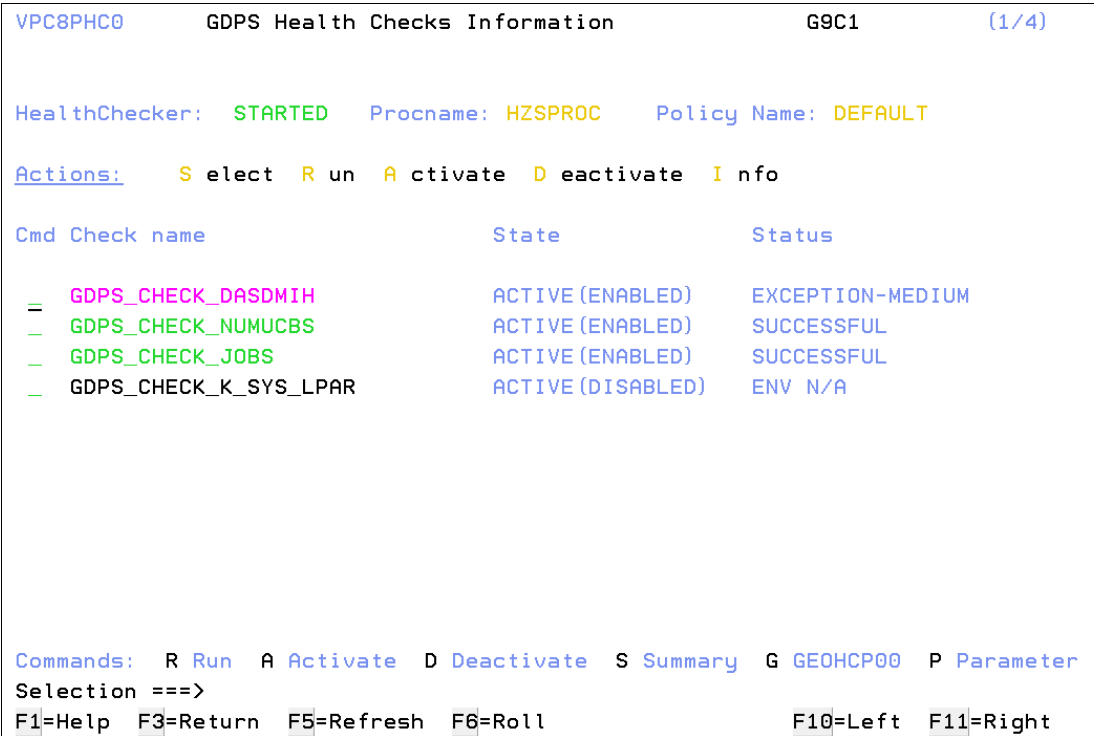


Figure 6-8 GDPS/GM Health Check management panel

6.6 Other facilities related to GDPS

In this section, we describe miscellaneous facilities provided by GDPS/Global Mirror that can assist in various ways.

6.6.1 GDPS/GM Copy Once facility

GDPS provides a *Copy Once* facility to copy volumes that have data sets on them that are required for recovery but the content is not critical, so they do not need to be copied all the time. Page data sets and work volumes that contain only truly temporary data such as sort work volumes are primary examples. The Copy Once facility can be invoked whenever required to refresh the information about these volumes.

To restart your workload in the recovery site, you need to have these devices or data sets available (the content is not required to be up to date). If you do not remote copy all of your production volumes, you will need to either manually ensure that the required volumes and data sets are preallocated and kept up to date at the recovery site or use the GDPS Copy Once function to manage these devices.

For example, if you are not replicating your paging volumes, then you must create the volumes with the proper volume serial with required data sets in the recovery site. Then, each time you change your paging configuration in the application site, you must reflect the changes in your recovery site. The GDPS Copy Once function provides a method of creating an initial copy of such volumes plus the ability to re-create the copy if the need arises as the result of any changes in the application site.

If you plan to use the Copy Once facility, you need to ensure that no data that needs to be continuously replicated is placed on the volumes you define to GDPS as Copy Once because these volumes will not be continuously replicated. The purpose of Copy Once is to ensure that you have a volume with the correct VOLSER, and with the data sets required for recovery allocated, available in the recovery site. The data in the data sets is not time-consistent with the data on the volumes that are continuously mirrored.

6.6.2 GDPS/GM Query Services

GDPS maintains configuration information and status information in NetView variables for the various elements of the configuration that it manages. GDPS Query Services is a facility that allows user-written REXX programs running under NetView to query and obtain the value of various GDPS variables. This allows you to augment GDPS automation with your own automation REXX code for various purposes such as monitoring or problem determination.

Query Services allows clients to complement GDPS automation with their own automation code. In addition to the Query Services function, which is part of the base GDPS product, GDPS provides several samples in the GDPS SAMPLIB library to demonstrate how Query Services can be used in client-written code.

6.6.3 Global Mirror Monitor integration

GDPS provides a Global Mirror Monitor (also referred to as *GM Monitor*) that is fully integrated into GDPS. This function provides a monitoring and historical reporting capability for Global Mirror performance and behavior, and some autonomic capability based on performance. The GM Monitor provides the following capabilities:

- ▶ Ability to view recent performance data for a Global Mirror session, for example to understand if an ongoing incident might be related to Global Mirror.
- ▶ Generation of alerts and messages for Global Mirror behavior based on exceeding thresholds in a defined policy.
- ▶ Ability to perform automatic actions such as pausing a GM session or resuming a previously paused session based on a defined policy.
- ▶ Creation of SMF records with detailed historical Global Mirror performance and behavioral data for problem diagnosis, performance reporting, and capacity planning.

The GM Monitor function runs in the K-sys and supports both CKD and FB environments. An independent monitor can be started for each GM session in your GDPS configuration. GDPS stores the performance data collected by each active monitor. Recent data is viewable using the GDPS 3270 panels.

6.6.4 Easy Tier Heat Map Transfer

IBM DS8000 Easy Tier optimizes data placement (placement of logical volumes) across the various physical tiers of storage within a disk subsystem to optimize application performance. The placement decisions are based on learning the data access patterns, and can be changed dynamically and transparently using this data.

Global Mirror copies the data from the primary to the secondary disk subsystem. However, the Easy Tier learning information is not included in the Global Mirror scope. The secondary disk subsystems are optimized according to the workload on these subsystems, which is different than the activity on the primary (there is only write workload on the secondary whereas there is read/write activity on the primary). And there is very little activity on the tertiary disk (FlashCopy target disk, or FC1 disk), so it will be optimized differently than the primary disk or the secondary disk. As a result of these differences, during a recovery, the disks that you recover on (secondary or tertiary) are likely to display different performance characteristics compared to the former primary.

Easy Tier Heat Map Transfer is the DS8000 capability to transfer the Easy Tier learning from a Global Mirror primary disk to a target set of disk. With GDPS/GM, the Easy Tier learning can be transferred to the secondary disk and the tertiary disk (FC1 disk) so that whatever disk you recover on can also be optimized based on this learning, and will have similar performance characteristics as the former primary.

GDPS integrates support for Heat Map Transfer. The appropriate Heat Map Transfer actions (such as start/stop of the processing and reversing transfer direction) are incorporated into the GDPS managed processes. For example, if Global Mirror is temporarily suspended for a planned or unplanned secondary disk outage, Heat Map Transfer is also suspended.

6.7 Flexible testing

If you want to conduct a disaster recovery test, you can use GDPS/GM to prepare the B disks to be used for the test. However, during the test, remote copying must be suspended. This is because the B disks are being used for the test, and the C disks contain a consistent copy of the production disks at the start of the test. If you were to have a real disaster during the test, the C disks will be used to give you a consistent restart point. All updates made to the production disks after the start of the test will need to be re-created, however. At the completion of the test, GDPS/GM uses the Failover/Failback capability to resynchronize the A and B disks without having to do a complete copy.

GDPS/GM supports an additional FlashCopy disk device, referred to as *F disks* or *FC1 disks*. F disks are additional “practice” FlashCopy target devices that may optionally be created in the recovery site. These devices may be used to facilitate stand-alone testing of your disaster recovery procedures. Disaster testing can be conducted by IPLing recovery systems on the F disk while live production continues to run in the application site and continues to be protected by the B and C disks. In addition, the F disk can be used to create a “gold” or insurance copy of the data in the event of a disaster situation. If you have this additional practice FlashCopy, you will be able to schedule disaster tests on demand much more frequently because such tests will have little or no impact on your RPO and DR capability.

For added scalability, GDPS allows the GM FC disks (C) to be defined in alternate subchannel set MSS1. It is also possible for GDPS/GM to support the FC1 disk without adding a requirement for the FC1 disks to be defined to the R-sys. See “Addressing z/OS device limits in a GDPS/GM environment” on page 34 for more information.

By combining Global Mirror with FlashCopy, you can create a usable copy of your production data to provide for on-demand testing capabilities and other nondisruptive activities. If there is a requirement to perform disaster recovery testing while maintaining the currency of the production mirror or for taking regular additional copies, perhaps once or twice a day, for other purposes, then consider installing the additional disk capacity to support F disks in your Global Mirror environment.

6.7.1 Use of space-efficient FlashCopy

As discussed in “Space-efficient FlashCopy (FlashCopy SE)” on page 40, by using space-efficient (SE) FlashCopy volumes, you might be able to lower the amount of physical storage needed, and thereby reduce the cost associated with providing a tertiary copy of the data. GDPS has support to allow FlashCopy SE volumes to be used as FlashCopy target disk volumes.

This support is transparent to GDPS; if the FlashCopy target devices defined to GDPS are space-efficient volumes, GDPS will simply use them. All GDPS FlashCopy operations with the NOCOPY option, whether through GDPS scripts or panels, can use space-efficient targets.

Because the IBM FlashCopy SE repository is of fixed size, it is possible for this space to be exhausted, thus preventing further FlashCopy activity. Consequently, we suggest using space-efficient volumes for temporary purposes, so that space can be reclaimed regularly.

GDPS/GM may use SE volumes as FlashCopy targets for either the C-disk or the F-disk. In the GM context, where the C-disk has been allocated to space-efficient volumes, each new Consistency Group reclaims used repository space since the previous Consistency Group, as the new flash is established with the C-disk. Therefore, a short Consistency Group Interval in effect ensures the temporary purpose recommendation for FlashCopy data. However, if the Consistency Group Interval grows long because of constrained bandwidth or write bursts, it is possible to exhaust available repository space. This will cause a suspension of GM, because any subsequent FlashCopy will not be possible.

Using space-efficient volumes for F disks depends on how you intend to use the F disks. These can be used for short-term, less-expensive testing, but are suitable for actual recovery because of their non-temporary nature.

6.7.2 Creating a test copy using GM CGPause and testing on isolated disks

The most basic GM configuration requires the GM secondary disk and the GM FlashCopy on the secondary disk subsystems. If you use an additional set of practice FlashCopy disks on the same disk subsystems, while you are performing recovery testing, you will have the I/O activity for GM mirroring and also the I/O activity generated by recovery testing on the same set of secondary disk subsystems. This I/O activity from the testing can potentially affect the GM mirroring.

GDPS/GM supports creating a *test copy* on disk subsystems isolated from the secondary disk subsystems. We call these the X-disks. The GM secondary disks are connected to the X-disks using the Global Copy (PPRC-XD) asynchronous copy technology. The GM secondary disks are the primary disks for the relationship to the X-disks.

To create a consistent test copy on the X-disks, GDPS/GM uses the Consistency Group Pause (CGPause) capability of the DS8000 disk subsystem to make the GM secondary disks consistent. After the GM secondary disks are consistent, GDPS waits until all data on these disks have been replicated to the X-disks and isolates the X-disks. GDPS then resumes the GM session.

The entire process of isolating the test copy on X-disks takes place in a short amount of time, which means minimal impact to GM operations has occurred during the creation of the test copy. Now, with the test copy isolated on disk subsystems other than the secondary disk subsystems, any testing performed does not interfere with or affect GM replication, which continues while you test on the X-disk copy.

GDPS also supports the same technique using CGPause to create practice FlashCopy. For environments that do not support CGPause, the GM secondary disks must first be recovered to make them consistent in order to take the practice FlashCopy. This is a much longer disruption to the GM session when compared to creating the FlashCopy test copy using CGPause.

In summary, CGPause minimizes the interruption to the GM session when creating a test copy. Isolating the test copy on a separate set of disk subsystems (X-disk) eliminates any impact the testing operation may have on the resumed GM session.

6.8 GDPS tools for GDPS/GM

GDPS includes tools that provide function that is complementary to GDPS function. The tools represent the kind of function that many clients are likely to develop themselves to complement GDPS. Using the GDPS tools eliminates the necessity for you to develop similar function yourself. The tools are provided in source code format, which means that if the tool does not completely meet your requirements, you can modify the code to tailor it to your needs.

The GDPS Distributed Systems Hardware Management Toolkit is available for GDPS/GM. It provides an interface for GDPS to monitor and control distributed systems' hardware and virtual machines (VMs) by using script procedures that can be integrated into GDPS scripts. This tool provides REXX script templates that show examples of how to monitor/control; IBM AIX HMC, VMware ESX server, IBM BladeCenter, and stand-alone x86 servers with Remote Supervisor Adapter II (RSA) cards. This tool is complementary to the heterogeneous, distributed management capabilities provided by GDPS, such as the Distributed Cluster Management (DCM) and Open LUN management functions.

6.9 Services component

As demonstrated, GDPS touches on much more than simply remote copy. It also includes automation, disk and system recovery, testing processes, and disaster recovery processes.

Most installations do not have all these skills readily available. Also, it is extremely rare to find a team that possesses this range of skills across many implementations. However, the GDPS/GM offering provides access to a global team of specialists in all the disciplines you need to ensure a successful GDPS/GM implementation.

Specifically, the Services component includes some or all of the following services:

- ▶ Planning to determine availability requirements, configuration recommendations, implementation and testing plans. Planning session topics include hardware and software requirements and prerequisites, configuration and implementation considerations, cross-site connectivity planning and potentially bandwidth sizing, and operation and control.
- ▶ Assistance in defining Recovery Point and Recovery Time objectives.
- ▶ Installation and necessary customization of NetView and System Automation.
- ▶ Remote copy implementation.
- ▶ GDPS/GM automation code installation and policy customization.
- ▶ Education and training on GDPS/GM setup and operations.

- Onsite implementation assistance.
- Project management and support throughout the engagement.

The sizing of the Services component of each project is tailored for that project based on many factors, including what automation is already in place, whether remote copy is already in place, and so on. This means that the skills provided are tailored to the specific needs of each implementation.

6.10 GDPS/GM prerequisites

Important: For the latest GDPS/GM prerequisite information, see the GDPS web page:

<http://www.ibm.com/systems/z/advantages/gdps/getstarted/index.html>

6.11 Comparison of GDPS/GM versus other GDPS offerings

So many features and functions are available in the various members of the GDPS family that recalling them all and remembering which offerings support them is sometimes difficult. To position the offerings Table 6-1 lists the key features and functions and indicates which ones are delivered by the various GDPS offerings.

Table 6-1 Supported features matrix

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/MTMM	GDPS Virtual Appliance	GDPS/XRC	GDPS/GM
Continuous availability	Yes	Yes	Yes	Yes	No	No
Disaster recovery	Yes	Yes	Yes	Yes	Yes	Yes
CA/DR protection against multiple failures	No	No	Yes	No	No	No
Continuous Availability for foreign z/OS systems	Yes with z/OS Proxy	No	No	No	No	No
Supported distance	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	Virtually unlimited	Virtually unlimited
Zero Suspend FlashCopy support	Yes, using Consistent	Yes, using Consistent for secondary only	Yes, using Consistent	No	Yes, using Zero Suspend FlashCopy	Yes, using CGPause
Reduced impact initial copy/resync	Yes	Yes	Yes	Yes	Not applicable	Not applicable

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/MTMM	GDPS Virtual Appliance	GDPS/XRC	GDPS/GM
Tape replication support	Yes	No	No	No	No	No
Production sysplex automation	Yes	No	Yes	Not applicable	No	No
Span of control	Both sites	Both sites (disk only)	Both sites	Both sites	Recovery site	Disk at both sites; recovery site (CBU or LPARs)
GDPS scripting	Yes	No	Yes	Yes	Yes	Yes
Monitoring, alerting and health checks	Yes	Yes	Yes	Yes (except health checks)	Yes	Yes
Query Services	Yes	Yes	No	No	Yes	Yes
MSS support for added scalability	Yes (secondary in MSS1)	Yes (secondary in MSS1)	Yes (H2 in MSS1, H3 in MSS2)	No	No	Yes (GM FC and Primary for MGM in MSS1)
MGM 3-site and 4-site	Yes (all configurations)	Yes (3-site only and non-IR only)	Yes (all configurations)	No	Not applicable	Yes (all configurations)
MzGM	Yes	Yes	Yes (non-IR only)	No	Yes	Not applicable
Open LUN	Yes	Yes	No	No	No	Yes
z/OS equivalent function for Linux for IBM z Systems	Yes	No	Yes (Linux for IBM z Systems running as a z/VM guest only)	Yes (Linux for IBM z Systems running as a z/VM guest only)	Yes	Yes
Heterogeneous support through DCM	Yes (VCS and SA AppMan)	No	No	No	Yes (VCS only)	Yes (VCS and SA AppMan)
z/BX hardware management	Yes	No	No	No	No	No
Web graphical interface	Yes	Yes	No	Yes	No	Yes

6.12 Summary

GDPS/GM provides automated disaster recovery capability over virtually unlimited distances for both CKD and FB devices. It does not have a requirement for a z/OS System Data Mover system as XRC does, but it does require an additional set of recovery disks when compared to GDPS/XRC. It also does not provide the vendor independence that GDPS/XRC provides.

The two controlling systems in a GDPS/GM configuration provide different functions:

- ▶ The K-sys, in the application site, is used to set up and control all remote copy operations.
- ▶ The R-sys, in the recovery site, is used primarily to drive recovery in case of a disaster.

You define a set of scripts that can reconfigure the servers in the recovery site, recover the disks, and start the production systems. The powerful scripting capability allows you to perfect the actions to be taken, either for planned or unplanned changes, thus eliminating the risk of human error. Both the K-sys and R-sys monitor key indicators and resources in their span of control and alert the operator of any non-normal status so that corrective action can be taken in a timely manner to eliminate or minimize RPO and RTO impact.

The B disks in the recovery site can be used for disaster recovery testing. The C disks contain a consistent (although, aging) copy of the production volumes. Optionally, a practice FlashCopy (F disks) can be integrated to eliminate the risk of RPO impact associated with testing on the B disks.

In addition to its DR capabilities, GDPS/GM also provides a user-friendly interface for monitoring and managing the remote copy configuration.



GDPS/MTMM

In this chapter, we discuss the capabilities and prerequisites of the GDPS/MTMM offering. GDPS/MTMM supports both planned and unplanned situations, helping to maximize application availability and provide business continuity. A GDPS/MTMM solution delivers the following benefits:

- ▶ Near-continuous availability
- ▶ Disaster recovery (DR) across metropolitan distances
- ▶ Recovery time objective (RTO) less than an hour
- ▶ Recovery point objective (RPO) of zero

Another key benefit of GDPS/MTMM is that it provides protection against multiple failures. GDPS/MTMM maintains three copies of your data so that even if one copy becomes unavailable, GDPS/MTMM can continue to provide near-continuous availability and DR by using the remaining two copies.

The functions provided by GDPS/MTMM fall into two categories: Protecting your data and controlling the resources managed by GDPS. The following functions are among those that are included:

- ▶ Protecting your data:
 - Ensuring the consistency of the secondary copies of your data in the event of a disaster or suspected disaster, including the option to also ensure zero data loss
 - Transparent switching to the secondary disk using HyperSwap
- ▶ Controlling the resources managed by GDPS during normal operations, planned changes, and following a disaster:
 - Monitoring and managing the state of the production z/OS systems and LPARs (shutdown, activating, deactivating, IPL, and automated recovery)
 - Monitoring and managing z/VM guests (shutdown, activating, deactivating, IPL, and automated recovery)
 - Managing the couple data sets and coupling facility recovery
 - Support for switching your disk, or systems, or both, to another site
 - User-customizable scripts that control how GDPS/MTMM reacts to specified error situations, which can also be used for planned events

7.1 Introduction to GDPS/MTMM

GDPS/MTMM is a continuous availability and disaster recovery solution that handles many types of planned and unplanned outages. As mentioned in Chapter 1, “Introduction to business resilience and the role of GDPS” on page 1, most outages are planned, and even among unplanned outages, most are not disasters. GDPS/MTMM provides capabilities to help provide the required levels of availability across these outages and in a disaster scenario. These capabilities are described in this chapter.

GDPS/MTMM leverages the IBM MTMM disk mirroring technology to maintain two synchronous secondary copies of your data. The primary copy and each of the two secondary copies are also called *disk locations*. The three disk locations, or copies, are H1, H2, and H3. H1 and H2 are assumed to be “local” and are fixed in Site 1. H3 is fixed in Site 2.

At any specific point in time, the production systems run on H1, H2 or H3 disk. Whichever copy the production systems are running on is known as the primary disk, and the other two copies are known as the secondary disks. Although the primary disk role can be with any of three disk locations, in a typical configuration:

- ▶ The primary disk is in Site1, that is, either H1 or H2.
- ▶ The other disk copy in Site1 provides high availability or HA protection.
- ▶ The copy in Site 2 (H3) provides disaster recovery or DR protection.

Each of the replication connections between the H1, H2, and H3 locations is called a *replication leg* or simply a *leg*. The replication legs in an MTMM configuration have fixed names that are based on the two disk locations that they connect:

- ▶ The H1-H2 (or H2-H1) leg is RL1.
- ▶ The H1-H3 (or H3-H1) leg is RL2.
- ▶ The H2-H3 (or H3-H2) leg is RL3.

The name of a given replication leg never changes, even if the replication direction is reversed for that leg. However, the *role* of a leg can change, depending on primary disk location. The two legs from the current primary to each of the two secondaries serve as the *active replication* legs whereas the leg between the two secondary locations serves as the *incremental resync* or *MTIR* leg.

To illustrate this concept, consider the sample GDPS/MTMM configuration that is shown in Figure 7-1.

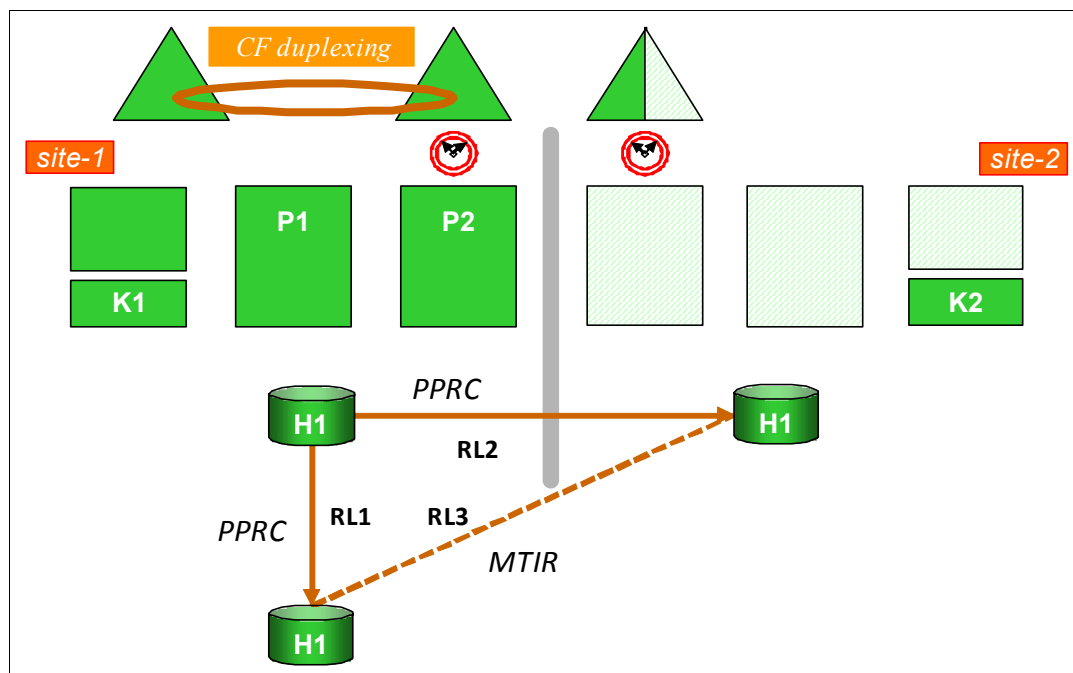


Figure 7-1 Sample GDPS/MTMM Configuration

In this sample configuration, H1 is the primary disk location, RL1 and RL2 are the active replication legs, and RL3 is the MTIR leg.

If there is a disk switch and H2 becomes the new primary disk, RL1 and RL3 become the active replication legs and RL2 becomes the MTIR leg.

7.1.1 Protecting data integrity and data availability with GDPS/MTMM

In 2.2, “Data consistency” on page 17, we point out that data integrity across primary and secondary volumes of data is essential to perform a database restart and accomplish an RTO of less than hour. This section includes details about how GDPS/MTMM automation provides both data consistency if there are mirroring problems and data availability if there are primary disk problems.

Two types of disk problems trigger a GDPS automated reaction:

- ▶ PPRC Mirroring problems (Freeze triggers). No problem exists writing to the primary disk subsystem, but a problem exists mirroring the data to one or both of the secondary disk subsystems. For more information, see “GDPS Freeze function for mirroring failures” on page 192.”
- ▶ Primary disk problems (HyperSwap triggers). There is a problem writing to the primary disk: either a hard failure, or the disk subsystem is not accessible or not responsive. For more information, see “GDPS HyperSwap function” on page 196.

GDPS Freeze function for mirroring failures

GDPS uses automation, keyed off events or messages, to stop all mirroring for a given replication leg when a remote copy failure occurs between one or more of the primary/secondary disk subsystem pairs on that replication leg. In particular, the GDPS automation uses the IBM PPRC Freeze and Run architecture, which has been implemented as part of Metro Mirror on IBM disk subsystems and also by other enterprise disk vendors. In this way, if the disk hardware supports the Freeze and Run architecture, GDPS can ensure consistency across all data in the sysplex (consistency group), regardless of disk hardware type.

This preferred approach differs from proprietary hardware approaches that work only for one type of disk hardware. For more information about data consistency with synchronous disk mirroring, see “PPRC data consistency” on page 24.

When a mirroring failure occurs, this problem is classified as a Freeze trigger and GDPS stops activity across *all* disk subsystems for the affected replication leg at the time the initial failure is detected, thus ensuring that the dependant write consistency of the secondary disks for that replication leg is maintained. Note that mirroring activity for the other replication leg is not affected by the freeze.

This is what happens when a GDPS performs a Freeze:

- ▶ Remote copy is suspended for all device pairs on the affected replication leg.
- ▶ While the suspend command is being processed for each LSS, each device goes into a long busy state. When the suspend completes for each device, z/OS marks the device unit control block (UCB) in all connected operating systems to indicate an Extended Long Busy (ELB) state.
- ▶ No I/Os can be issued to the affected devices until the ELB is thawed with the PPRC Run action or until it times out. (The consistency group timer setting commonly defaults to 120 seconds, although for most configurations a longer ELB is preferable.)
- ▶ All paths between the PPRCed disks on the affected replication leg are removed, preventing further I/O to the associated secondary disks if PPRC is accidentally restarted.

Because no I/Os are processed for a remote-copied volume during the ELB, dependent write logic ensures the consistency of the affected secondary disks. GDPS performs a Freeze for all LSS pairs that contain GDPS managed mirrored devices.

Important: Because of the dependent write logic, it is not necessary for all LSSs to be frozen at the same instant. In a large configuration with many thousands of remote copy pairs, it is not unusual to see short gaps between the times when the Freeze command is issued to each disk subsystem. Because of the ELB, however, such gaps are not a problem.

After GDPS performs the Freeze and the consistency of the secondary disks on the affected leg is protected and the action GDPS takes next depends on the client's PPRCFAILURE policy (also known as Freeze policy). See “Freeze policy (PPRCFAILURE policy) options” on page 193 for details regarding the actions GDPS will take, based on this policy.

GDPS/MTMM uses a combination of storage subsystem and sysplex triggers to automatically secure, at the first indication of a potential disaster, a data-consistent secondary copy of your data using the Freeze function. In this way, the secondary copy of the data is preserved in a consistent state, perhaps even before production applications are aware of any issues. Ensuring the data consistency of the secondary copy ensures that a normal system restart can be performed instead of having to perform DBMS forward recovery actions. This is an

essential design element of GDPS to minimize the time to recover the critical workloads in the event of a disaster in the primary site.

You can appreciate why such a process must be automated. When a device suspends, there is not enough time to launch a manual investigation process. The entire mirror for the affected leg must be frozen by stopping further I/O to it, and then the policy indicates whether production will continue to run with mirroring temporarily suspended, or whether all systems should be stopped to guarantee zero data loss.

In summary, freeze is triggered as a result of a PPRC suspension event for any primary disk in the GDPS configuration, that is, at the first sign that a duplex mirror that is going out of the duplex state. When a device suspends, all attached systems are sent a “State Change Interrupt” (SCI). A message is issued in all of those systems and then each system must issue multiple I/Os to investigate the reason for the suspension event.

When GDPS performs a freeze, all primary devices in the PPRC configuration suspend for the affected replication leg. This can result in significant SCI traffic and many messages in all of the systems. GDPS, in conjunction with z/OS and microcode on the DS8000 disk subsystems, supports reporting suspensions in a summary message per LSS instead of at the individual device level. When compared to reporting suspensions on a per devices basis, the Summary Event Notification for PPRC Suspends (PPRCSUM) dramatically reduces the message traffic and extraneous processing associated with PPRC suspension events and freeze processing.

Freeze policy (PPRCFAILURE policy) options

As we have described, when a mirroring failure is detected on a replication leg, GDPS automatically and unconditionally performs a Freeze of that leg to secure a consistent set of secondary volumes in case the mirroring failure could be the first indication of a site failure. Because the primary disks are in the Extended Long Busy state as a result of the freeze and the production systems are locked out, GDPS must take some action. Here, there is no time to interact with the operator on an event-by-event basis. The action must be taken immediately. The action to be taken is determined by a customer policy setting, that is, the PPRCFailure policy option (also known as the Freeze policy option). GDPS will use this same policy setting after every Freeze event to determine what its next action should be. The policy can be specified at a leg level allowing a different policy specification for the replication legs. The options are as follows:

- ▶ PPRCFailure=GO (Freeze and Go)
GDPS allows production systems to continue operation after mirroring is suspended.
- ▶ PPRCFailure=STOP (Freeze and Stop)
GDPS resets production systems while I/O is suspended.
- ▶ PPRCFailure=STOPLAST
GDPS checks the mirroring status of the other replication leg. If the status of the other leg is OK, GDPS performs a Go. If not, and this is the last viable leg that GDPS has just frozen, GDPS performs a Stop.
- ▶ PPRCFailure=COND (Freeze and Stop conditionally)
GDPS tries to determine if a secondary disk caused the mirroring failure. If so, GDPS performs a Go. If not, GDPS performs a Stop.
- ▶ PPRCFailure=CONDLAST
GDPS checks the mirroring status of the other replication leg. If the status of the other leg is OK, GDPS performs a Go. If not (the freeze was performed on the last viable leg),

GDPS tries to determine if a secondary disk caused the mirroring failure. If so, GDPS performs a Go. If not, GDPS performs a Stop.

Freeze and Go

With this policy, after performing the Freeze, GDPS performs a Run action against all primary LSSs, which is also known as performing a *Go*. Performing a Go removes the ELB and allows production systems to continue using these devices. The devices will be in remote copy-suspended mode in relation to the secondary devices on the affected leg, so any further writes to these devices are no longer being mirrored to the secondary devices on that leg. However, changes are being tracked by the hardware so that, later, only the changed data will be resynchronized to the affected secondary disks.

With this policy you avoid an unnecessary outage for a false freeze event, that is, if the trigger is simply a transient event. However, if the trigger turns out to be the first sign of an actual disaster, you might continue operating for an amount of time before all systems fail. Any updates made to the primary volumes during this time are not replicated to the secondary disk, and therefore are lost if you end up having to recover on the affected secondary disk. In addition, because the CF structures were updated after the secondary disks were frozen, the CF structure content is not consistent with the secondary disks. Therefore, the CF structures in either site cannot be used to restart workloads and log-based restart must be used when restarting applications.

This is not full forward recovery. It is forward recovery of any data, such as DB2 group buffer pools, that might have existed in a CF but might not have been written to disk yet. This results in prolonged recovery times. The duration depends on how much such data existed in the CFs at that time. With a Freeze and Go policy, you might consider tuning applications such as DB2, which can harden such data on disk more frequently than otherwise.

Freeze and Go is a high availability option that avoids production outage for false freeze events. However, it carries a potential for data loss.

Freeze and Stop

With this policy, you can be assured that no updates are made to the primary volumes after the Freeze because all systems that can update the primary volumes are reset. This ensures that no more updates can occur to the primary disks because such updates would not be mirrored to the affected secondary disk, meaning that it would not be possible to achieve zero data loss if a failure occurs (or if the original trigger was an indication of a catastrophic failure) and recovery on the affected secondary disk is required.

You can choose to restart the systems when you want. For example, if this was a false freeze (that is, a false alarm), then you can quickly resynchronize the mirror and restart the systems only after the mirror is duplex.

If you are using duplexed coupling facility (CF) structures along with a Freeze and Stop policy, it might seem that you are guaranteed to use the duplexed instance of your structures if you must recover and restart your workload with the frozen secondary copy of your disks. However, this is not always the case. There can be rolling disaster scenarios where before, after, or during the freeze event, there is an interruption (perhaps failure of CF duplexing links) that forces CFRM to drop out of duplexing. There is no guarantee that it is the structure instance in the surviving site that is kept. It is possible that CFRM keeps the instance in the site that is about to totally fail. In this case, there will not be an instance of the structure in the site that survives the failure.

To summarize, with a Freeze and Stop policy, if there is a surviving, accessible instance of application-related CF structures, this instance will be consistent with the frozen secondary disks. However, depending on the circumstances of the failure, even with structures duplexed across two sites you are not 100% guaranteed to have a surviving, accessible instance of the application structures and therefore you must have the procedures in place to restart your workloads without the structures.

Although a Stop policy can be used to ensure no data loss, if a failure occurs that is a false freeze event, that is, it is a transient failure that did not necessitate recovering using the frozen disks, it results in unnecessarily stopping the systems.

Freeze and Stop last

With this policy, after the Freeze, GDPS checks the status of mirroring on the other replication leg (the leg other than the one that was just frozen) to determine whether the leg that just frozen was the last leg actively replicating data. If the other leg is still actively replicating data, GDPS performs a Go. But if the other leg is already frozen or mirroring status is not OK, GDPS performs a Stop.

When you have only one replication leg defined in your configuration (you have only one secondary copy of your data), using this policy specification is the same as using a Freeze and Stop policy.

Freeze and Stop conditional

Field experience has shown that most of the Freeze triggers are not necessarily the start of a rolling disaster, but are “False Freeze” events that do not necessitate recovery on the secondary disk. Examples of such events include connectivity problems to the secondary disks and secondary disk subsystem failure conditions.

With a COND policy, the action that GDPS takes after it performs the Freeze is conditional. GDPS tries to determine if the mirroring problem was as a result of a permanent or temporary secondary disk subsystem problem:

- ▶ If GDPS can determine that the freeze was triggered as a result of a secondary disk subsystem problem, GDPS performs a Go. That is, it allows production systems to continue to run by using the primary disks. However, updates will not be mirrored until the secondary disk can be fixed and PPRC can be resynchronized.
- ▶ If GDPS cannot ascertain that the cause of the freeze was a secondary disk subsystem, GDPS operates on the assumption that this could still be the beginning of a rolling disaster in the primary site and performs a Stop, resetting all the production systems to guarantee zero data loss. GDPS cannot always detect that a particular freeze trigger was caused by a secondary disk, and that some freeze events that are in fact caused by a secondary disk could still result in a Stop.

For GDPS to determine whether a freeze trigger might have been caused by the secondary disk subsystem, the IBM DS8000 disk subsystems provide a special query capability known as the *Query Storage Controller Status* microcode function. If all disk subsystems in the GDPS managed configuration support this feature, GDPS uses this special function to query the secondary disk subsystems in the configuration to understand the state of the secondaries and if one of these secondaries might have caused the freeze. If you use the COND policy setting but all disks in your configuration do not support this function, GDPS cannot query the secondary disk subsystems, and the resulting action is a Stop.

This option can provide a good compromise where you can minimize the chance that systems would be stopped for a false freeze event and increase the chance of achieving zero data loss for a real disaster event.

Freeze and Stop conditional last

With this policy, after the Freeze, GDPS checks the status of mirroring on the other replication leg (the leg other than the one that was just frozen) to determine if the leg just frozen was the last leg actively replicating data. If the other leg is still actively replicating data, GDPS performs a Go. If the other leg is already frozen or mirroring status is not OK, GDPS performs conditional Stop processing; that is, it queries the secondary disk subsystem and performs a Go if, as a result of the query, it determines that the freeze was caused by the secondary, but performs a Stop if it cannot determine for sure that the problem was caused by the secondary.

When you only have one replication leg defined in your configuration (you only have one secondary copy of your data), using this policy specification is the same as using a Freeze and Stop conditional policy.

PPRCFAILURE policy selection considerations

The PPRCFailure policy option specification directly relates to recovery time and recovery point objectives (RTO and RPO, respectively), which are business objectives. Therefore, the policy option selection is really a business decision rather than an IT decision. If data associated with your transactions is high-value, it might be more important to ensure that no data associated with your transactions is ever lost, so you might decide on a Freeze and Stop policy. If you have huge volumes of relatively low-value transactions, you might be willing to risk some lost data in return for avoiding unnecessary outages with a Freeze and Go policy. The Freeze and Stop Conditional policy attempts to minimize the chance of unnecessary outages and the chance of data loss; however, there is still a risk of either, however small.

The various PPRCFailure policy options, combined with the fact that the policy options are specified on a per replication leg basis (different policies can be specified for different legs), gives you the flexibility to refine your policies to meet your unique business goals.

For example, if your RPO is zero, you can employ the following PPRCFailure policy:

- For RL2, Freeze and Stop (PPRCFailure=STOP)

Since H3 is your disaster recovery copy and you must ensure that you never lose data should you ever have to recover and run on the H3 disk, you must always, unconditionally stop the systems to ensure that no further updates occur to the primary disks that could be lost in a recovery scenario.

- For RL1, Freeze and Stop on last leg only (STOPLAST)

You do not need to take a production outage when PPRC freezes on the high-availability leg if RL2 is still functional and continues to provide disaster recovery protection. However, if RL2 is not functional when PPRC on RL1 suspends, you might want to at least retain the capability to recover on H2 disk with zero data loss if it becomes necessary.

However, if you want to avoid unnecessary outages at the risk of losing data if there is an actual disaster, you can specify Freeze and Go for both of your replication legs.

GDPS HyperSwap function

If there is a problem writing or accessing the *primary* disk because of a failing, failed, or inaccessible primarynon-responsive disk, there is a need to swap from the primary disks to one of the sets of secondary disks.

GDPS/MTMM delivers a powerful function known as HyperSwap. HyperSwap provides the ability to swap from using the primary devices in a mirrored configuration to using what had been one of the sets of secondary devices, in a manner that is transparent to the production systems and applications using these devices. Before the availability of HyperSwap, a transparent disk swap was not possible. All systems using the primary disk would have been shut down (or might have failed, depending on the nature and scope of the failure) and would have been re-IPLed using the secondary disks. Disk failures were often a single point of failure for the entire sysplex.

With HyperSwap, such a switch can be accomplished without IPL and with just a brief hold on application I/O. The HyperSwap function is completely controlled by automation, thus allowing all aspects of the disk configuration switch to be controlled through GDPS.

HyperSwap can be invoked in two ways:

- ▶ **Planned HyperSwap**

A planned HyperSwap is invoked by operator action using GDPS facilities. One example of a planned HyperSwap is where a HyperSwap is initiated in advance of planned disruptive maintenance to a disk subsystem.

- ▶ **Unplanned HyperSwap**

An unplanned HyperSwap is invoked automatically by GDPS, triggered by events that indicate the primary disk problem.

Primary disk problems can be detected as a direct result of an I/O operation to a specific device that fails because of a reason that indicates a primary disk problem such as:

- No paths available to the device
- Permanent error
- I/O timeout

In addition to a disk problem being detected as a result of an I/O operation, it is also possible for a primary disk subsystem to proactively report that it is experiencing an acute problem. The IBM DS8000 provides a special microcode function known as the *Storage Controller Health Message Alert* capability. Problems of different severity are reported by disk subsystems that support this capability. Those problems classified as *acute* are also treated as HyperSwap triggers. After systems are swapped to use the secondary disks, the disk subsystem and operating system can try to perform recovery actions on the former primary without impacting applications since the applications are no longer using those disks.

Planned and unplanned HyperSwap have requirements in terms of the physical configuration, such as having to be symmetrically configured, and so on. While a client's environment meets these requirements, there is no special enablement required to perform planned swaps. Unplanned swaps are not enabled by default and must be enabled explicitly as a policy option. This is described in more detail in "Preferred Swap Leg and HyperSwap (Primary Failure) policy options" on page 199.

When a swap is initiated, GDPS always validates various conditions to ensure that it is safe to swap. For example, if the mirror is not fully duplex on a given leg, that is, not all volume pairs are in a duplex state, a swap cannot be performed on that leg. The way that GDPS reacts to such conditions changes depending on the condition detected and whether the swap is a planned or unplanned swap.

Assuming that there are no show-stoppers and the swap proceeds, for both planned and unplanned HyperSwap, the systems that are using the primary volumes will experience a temporary pause in I/O processing. GDPS blocks I/O both at the channel subsystem level by performing a Freeze which results in all disks going into Extended Long Busy, and also in all systems, where I/O is quiesced at the operating system (UCB) level. This is to ensure that no systems use the disks until the switch is complete. During the time when I/O is paused, the following process is completed:

1. The PPRC configuration is *physically switched*. This includes physically changing the secondary disk status to primary. Secondary disks are protected and cannot be used by applications. Changing their status to primary allows them to come online to systems and be used.
2. The disks will be *logically switched* in each of the systems in the GDPS configuration. This involves switching the internal pointers in the operating system control blocks (UCBs). After the switch, the operating system will point to the former secondary devices which will be the new primary devices.
3. Finally, the systems resume operation using the new, swapped-to primary devices. The applications are not aware of the fact that different devices are now being used.

This brief pause during which systems are locked out of performing I/O is known as the *User Impact Time*. In benchmark measurements at IBM using currently supported releases of GDPS and IBM DS8000 disk subsystems, the User Impact Time to swap 10,000 pairs across 16 systems during an unplanned HyperSwap was less than 10 seconds. Most implementations are actually much smaller than this and typical impact times in a well-configured environment using the most current storage and server hardware are measured in seconds. Although results will depend on your configuration, these numbers give you a high-level idea of what to expect.

HyperSwap can be executed on either replication leg in a GDPS/MTMM environment. For a planned swap, you must specify which leg you want to use for the swap. For an unplanned swap, which leg is chosen depends on many factors, including your HyperSwap policy. This is described in more detail in “Preferred Swap Leg and HyperSwap (Primary Failure) policy options” on page 199.

After a replication leg is selected for the HyperSwap, GDPS swaps all devices on the selected replication leg. Just as the Freeze function applies to the entire consistency group, HyperSwap is for the entire consistency group. For example, if a single mirrored volume fails and HyperSwap is invoked, processing is swapped to one of the sets of secondary devices for *all* primary volumes in the configuration, including those in other, unaffected, disk subsystems. This is to ensure that all primary volumes remain in the same site. If HyperSwap were to swap only the failed LSS, you would then have several primaries in one location, and the remainder in another location. This would make for a significantly complex environment to operate and administer I/O configurations.

Incremental Resynchronization

When a disk switch or recovery on one of the secondaries occurs, MTMM provides for a capability known as “incremental resynchronization” (IR). Assume your H1 disks are the current primaries and the H2 and H3 disks are the current secondaries. If you switch from using H1 to using H2 as your primary disks, to maintain a multi-target configuration, you will need to establish replication on RL1, between H2 and H1, and on RL3, between H2 and H3. A feature of the PPRC copy technology known as Failover/Failback, together with the MTMM IR capability allows you to establish replication for RL1 and RL3 without having to copy all of the data from H2 to H1 or from H2 to H3. Only the changes that occur on B after primary is switched to B are copied in order to resynchronize the two legs.

If there is an unplanned HyperSwap from H1 to H2, because H1 has failed, replication can be established on RL3 between H2 and H3 in order to restore disaster recovery readiness. Again, this is an incremental resynchronization (only changed tracks are copied), so the duration to get to a protected position will be much faster compared to performing an initial copy for the leg.

HyperSwap with less than full channel bandwidth

You may consider enabling unplanned HyperSwap on the cross-site replication leg (RL2), even if you do not have sufficient cross-site channel bandwidth to sustain the full production workload for normal operations. Assuming that a disk failure is likely to cause an outage and that you have to switch to using the H3 disk in the other site (because the H2 disks in the same site are down at the time), the unplanned HyperSwap to H3 might at least present you with the opportunity to perform an orderly shutdown of your systems first. Shutting down your systems cleanly avoids the complications and restart time elongation associated with a crash-restart of application subsystems.

Preferred Swap Leg and HyperSwap (Primary Failure) policy options

Clients might prefer not to immediately enable their environment for unplanned HyperSwap when they first implement GDPS. For this reason, unplanned HyperSwap is not enabled by default. However, we strongly suggest that all GDPS/MTMM clients enable their environment for unplanned HyperSwap, at a minimum, on the local replication leg (RL1). Both copies of disk on the RL1 leg (H1 and H2) are local and therefore distance and connectivity should not be an issue.

You control the actions that GDPS takes for primary disk problems by specifying a Primary Failure policy option. This option is applicable to both replication legs. However, you have the option of overriding this specification at a leg level and request a different action based on which leg is selected by GDPS to act upon. Furthermore, there is the Preferred Swap Leg policy, which is factored in when GDPS decides which leg to act upon as a result of a primary disk problem trigger.

Preferred Swap Leg selection for unplanned HyperSwap

A primary disk problem trigger is common to both replication legs since the primary disk is common to both legs. Before acting on the trigger, GDPS first needs to select which leg to act upon. GDPS provides you with the ability to *influence* this decision by specifying a Preferred Swap Leg policy. GDPS will attempt to select the leg that you have identified as the Preferred Swap Leg first. However, if this leg is not eligible for the action that you specified in your Primary Failure policy, GDPS attempts to select the other active replication leg. These are among the reasons that your Preferred Swap Leg might not be eligible for selection:

- ▶ It is currently the MTIR leg.
- ▶ All pairs for the leg are not in a duplex state.
- ▶ It is currently not HyperSwap enabled.

HyperSwap retry on non-preferred leg

If the preferred leg is viable and selected for an unplanned swap, there is still a possibility (albeit small) that the swap on this leg fails for some reason. When swap on the first leg fails, if the other replication leg is enabled for HyperSwap, GDPS will retry the swap on the other leg. This maximizes the chances of a successful swap.

Primary failure policy options

After GDPS has selected which leg it will act on when a primary disk problem trigger occurs, the first thing it will do will be a Freeze on the selected leg (the same as is performed when a mirroring problem trigger is encountered). GDPS then applies the Primary Failure policy option specified for that leg. The Primary Failure policy for each leg can specify a different action. You can specify the following Primary Failure policy options:

► PRIMARYFAILURE=GO

No swap is performed. The action GDPS takes is the same as for a freeze event with policy option PPRCFailure=GO. A Run action is performed, which will allow systems to continue using the original primary disks. PPRC is suspended and therefore updates are not being replicated to the secondary. Note, however, that depending on the scope of the primary disk problem, it might be that some or all production workloads simply cannot run or cannot sustain required service levels. Such a situation might necessitate restarting the systems on the secondary disks. Because of the freeze, the secondary disks are in a consistent state and can be used for restart. However, any transactions that ran after the Go action will be lost.

► PRIMARYFAILURE=STOP

No swap is performed. The action GDPS takes is the same as for a freeze event with policy option PPRCFailure=STOP. GDPS system-resets all the production systems. This ensures that no further I/O occurs. After performing situation analysis, if it is determined that this was not a transient issue and that the secondaries should be used to IPL the systems again, no data will be lost.

► PRIMARYFAILURE=SWAP,*swap_disabled_action*

The first parameter, SWAP, indicates that after performing the Freeze, GDPS will proceed with performing an unplanned HyperSwap. When the swap is complete, the systems will be running on the new, swapped-to primary disks (former secondaries). Mirroring on the selected leg will be in a suspended state; because the primary disks are known to be in a problematic state, there is no attempt to reverse mirroring. After the problem with the primary disks is fixed, you can instruct GDPS to resynchronize PPRC from the current primaries to the former ones (which are now considered to be secondaries).

The second part of this policy, *swap_disabled_action*, indicates what GDPS should do if HyperSwap had been temporarily disabled by operator action at the time the trigger was encountered. Effectively, an operator action has instructed GDPS not to perform a HyperSwap, even if there is a swap trigger. GDPS has already performed a freeze. The second part of the policy control what action GDPS will take next.

The following options (which are in effect only if HyperSwap is disabled by the operator) are available for the second parameter (remember that the disk is already frozen):

GO	This is the same action as GDPS would have performed if the policy option had been specified as PRIMARYFAILURE=GO.
STOP	This is the same action as GDPS would have performed if the policy option had been specified as PRIMARYFAILURE=STOP.

Preferred Swap Leg and Primary Failure policy selection considerations

For the Preferred Swap Leg policy, consider whether you can tolerate running with disk and systems in opposite sites with no/minimal performance impact. If that is acceptable, you can choose either leg, although it might be better to prefer the RL2 (Site1-Site2) leg. If you cannot tolerate running with disks and systems in opposite sites, choose the RL1, local leg.

For the Primary Failure policy, again we recommend that you specify SWAP for the first part of the policy option to enable HyperSwap, at least on the local replication leg (RL1). If distance and connectivity between your sites is not an issue, consider specifying SWAP for the first part of the policy on the remote replication leg (RL2) also.

For the Stop or Go choice, either as the second part of the policy option or if you will not be using SWAP, similar considerations apply as for the PPRCFailure policy options to Stop or Go. Go carries the risk of data loss if it is necessary to abandon the primary disk and restart systems on the secondary. Stop carries the risk of taking an unnecessary outage if the problem was transient. The key difference is that with a mirroring failure, the primary disks are not broken. When you allow the systems to continue to run on the primary disk with the Go option, other than a disaster (which is low probability), the systems are likely to run with no problems. With a primary disk problem, with the Go option, you are allowing the systems to continue running on what are known to be disks that experienced a problem just seconds ago. If this was a serious problem with widespread impact, such as an entire disk subsystem failure, the applications will experience severe problems. Some transactions might continue to commit data to those disks that are not broken. Other transactions might be failing or experiencing serious service time issues. Also, if there is a decision to restart systems on the secondary because the primary disks are simply not able to support the workloads, there will be data loss. The probability that a primary disk problem is a real problem that will necessitate restart on the secondary disks is much higher when compared to a mirroring problem. A Go specification in the Primary Failure policy increases your risk of data loss.

If the primary failure was of a transient nature, a Stop specification results in an unnecessary outage. However, with primary disk problems, the probability that the problem could necessitate restart on the secondary disks is high, so a Stop specification in the Primary Failure policy avoids data loss and facilitates faster restart.

The considerations relating to CF structures with a PRIMARYFAILURE event are similar to a PPRCFailure event. If there is an actual swap, the systems continue to run and continue to use the same structures as they did before the swap; the swap is transparent. With a Go action, because you continue to update the CF structures along with the primary disks after the Go, if you need to abandon the primary disks and restart on the secondary, the structures are inconsistent with the secondary disks and are not usable for restart purposes. This will prolong the restart, and therefore your recovery time. With Stop, if you decide to restart the systems using the secondary disks, there is no consistency issue with the CF structures because no further updates occurred on either set of disks after the trigger was captured.

GDPS use of DS8000 functions

GDPS strives to use (when it makes sense) enhancements to the IBM DS8000 disk technologies. In this section we provide information about the key DS8000 technologies that GDPS supports and uses.

PPRC Failover/Failback support

When a primary disk failure occurs and the disks are switched to the secondary devices, PPRC Failover/Failback (FO/FB) support eliminates the need to do a full copy when reestablishing replication in the opposite direction. Because the primary and secondary volumes are often in the same state when the freeze occurred, the only differences between the volumes are the updates that occur to the secondary devices after the switch.

Failover processing sets the secondary devices to primary suspended status and starts change recording for any subsequent changes made. When the mirror is reestablished with failback processing, the original primary devices become secondary devices and a resynchronization of changed tracks takes place.

GDPS/MTMM requires PPRC FO/FB capability to be available on all disk subsystems in the managed configuration.

PPRC eXtended Distance (PPRC-XD)

PPRC-XD (also known as Global Copy) is an asynchronous form of the PPRC copy technology. GDPS uses PPRC-XD rather than synchronous PPRC to reduce the performance impact of certain remote copy operations that potentially involve a large amount of data. See 7.6.2, “Reduced impact initial copy and resynchronization” on page 225 for details.

Storage Controller Health Message Alert

This facilitates triggering an unplanned HyperSwap proactively when the disk subsystem reports an acute problem that requires extended recovery time. See “GDPS HyperSwap function” on page 196 for more information about unplanned HyperSwap triggers.

PPRC Summary Event Messages

GDPS supports the DS8000 PPRC Summary Event Messages (PPRCSUM) function which is aimed at reducing the message traffic and the processing of these messages for Freeze events. This is described in “GDPS Freeze function for mirroring failures” on page 192.

Soft Fence

Soft Fence provides the capability to block access to selected devices. As discussed in “Protecting secondary disks from accidental update” on page 203, GDPS uses Soft Fence to avoid write activity on disks that are exposed to accidental update in certain scenarios.

On-demand dump (also known as non-disruptive statesave)

When problems occur with disk subsystems such as those which result in an unplanned HyperSwap, a mirroring suspension or performance issues, a lack of diagnostic data from the time the event occurs can result in difficulties in identifying the root cause of the problem. Taking a full statesave can lead to temporary disruption to host I/O and is often frowned upon by clients for this reason. The on-demand dump (ODD) capability of the disk subsystem facilitates taking a non-disruptive statesave (NDSS) at the time that such an event occurs. The microcode does this automatically for certain events such as taking a dump of the primary disk subsystem that triggers a PPRC freeze event and also allows an NDSS to be requested by an exploiter. This enables first failure data capture (FFDC) and thus ensures that diagnostic data is available to aid problem determination. Be aware that not all information that is contained in a full statesave is contained in an NDSS and therefore there may still be failure situations where a full statesave is requested by the support organization.

GDPS provides support for taking an NDSS using the remote copy panels. In addition to this support, GDPS autonomically takes an NDSS if there is an unplanned Freeze or HyperSwap event.

Query Host Access function

When a PPRC disk pair is being established, the device that is the target (secondary) must not be used by any system. The same is true when establishing a FlashCopy relationship to a target device. If the target is in use, the establishment of the PPRC or FlashCopy relationship fails. When such failures occur, it can be a tedious task to identify which system is holding up the operation.

The Query Host Access disk function provides the means to query and identify what system is using a selected device. GDPS uses this capability and adds usability in several ways:

- ▶ Query Host Access identifies the LPAR that is using the selected device through the CPC serial number and LPAR number. It is still a tedious job for operations staff to translate this information to a system or CPC and LPAR name. GDPS does this translation and presents the operator with more readily usable information, thereby avoiding this additional translation effort.
- ▶ Whenever GDPS is requested to perform a PPRC or FlashCopy establish operation, GDPS first performs Query Host Access to see if the operation is expected to succeed or fail as a result of one or more target devices being in use. GDPS alerts the operator if the operation is expected to fail, and identifies the target devices in use and the LPARs holding them.
- ▶ GDPS continually monitors the target devices defined in the GDPS configuration and alerts operations to the fact that target devices are in use when they should not be. This allows operations to fix the reported problems in a timely manner.
- ▶ GDPS provides the ability for the operator to perform ad hoc Query Host Access to any selected device using the GDPS panels.

Protecting secondary disks from accidental update

A system cannot be IPLed using a disk that is physically a PPRC secondary disk because PPRC secondary disks cannot be brought online to any systems. However, a disk can be secondary from a GDPS (and application use) perspective but physically, from a PPRC perspective, have simplex or primary status.

For both planned and unplanned HyperSwap, and a disk recovery, GDPS changes former secondary disks to primary or simplex state. However, these actions do not modify the state of the former primary devices, which remain in the primary state. Therefore, the former primary devices remain accessible and usable even though they are considered to be the secondary disks from a GDPS perspective. This makes it is possible to accidentally update or IPL from the wrong set of disks. Accidentally using the wrong set of disks can potentially result in a loss of data integrity or data.

GDPS/MTMM provides protection against using the wrong set of disks in different ways:

- ▶ If you attempt to load a system through GDPS (either script or panel) using the wrong set of disks, GDPS rejects the load operation.
- ▶ If you used the HMC rather than GDPS facilities for the load, then early in the IPL process, during initialization of GDPS, if GDPS detects that the system coming up has just been IPLed using the wrong set of disks, GDPS will quiesce that system, preventing any data integrity problems that could be experienced had the applications been started.
- ▶ GDPS uses a DS8000 disk subsystem capability, which is called Soft Fence for configurations where the disks support this function. Soft Fence provides the means to *fence* (that is, block) access to a selected device. GDPS uses Soft Fence when appropriate to fence devices that would otherwise be exposed to accidental update.

7.1.2 Protecting other CKD data

Systems that are fully managed by GDPS are known as *GDPS managed systems* or *GDPS systems*. There are two types of GDPS Systems as follows:

- ▶ z/OS systems in the GDPS sysplex
- ▶ z/VM systems managed by GDPS/MTMM MultiPlatform Resiliency for System z (xDR)

GDPS/MTMM can also manage the disk mirroring of CKD disks used by systems outside of the sysplex: other z/OS systems, Linux on System z, VM, and VSE systems that are not running any GDPS/MTMM or xDR automation. These are known as “foreign systems.”

Because GDPS manages PPRC for the disks used by these systems, these disks will be attached to the GDPS controlling systems. With this setup, GDPS is able to capture mirroring problems and will perform a freeze. All GDPS managed disks belonging to the GDPS systems and these foreign systems are frozen together, regardless of whether the mirroring problem is encountered on the GDPS systems’ disks or the foreign systems’ disks.

GDPS/MTMM is not able to directly communicate with these foreign systems. For this reason, GDPS automation will not be aware of certain other conditions such as a primary disk problem that is detected by these systems. Because GDPS will not be aware of such conditions that would have otherwise driven autonomic actions such as HyperSwap, GDPS will not react to these events.

If an unplanned HyperSwap occurs (because it was triggered on a GDPS managed system), the foreign systems cannot and will not swap to using the secondaries. A setup is prescribed to set a long Extended Long Busy time-out for these systems such that when the GDPS managed systems swap, these systems hang. The ELB prevents these systems from continuing to use the former primary devices. You can then use GDPS automation facilities to reset these systems and re-IPL them using the swapped-to primary disks.

7.2 GDPS/MTMM configurations

At its most basic, a GDPS/MTMM configuration consists of at least one production system, at least one controlling system in a sysplex, primary disks, and secondary disks. The actual configuration depends on your business and availability requirements. The following three configurations are most common:

- Single-site workload configuration

In this configuration, all of the production systems normally run in the same site, referred to as Site1, and the GDPS controlling system runs in Site2. In effect, Site1 is the active site for all production systems. The controlling system in Site2 is running and resources are available to move production to Site2, if necessary, for a planned or unplanned outage of Site1. Although you might also hear this referred to as an *Active/Standby GDPS/MTMM* configuration, we avoid the Active/Standby term to avoid confusion with the same term used in conjunction with the GDPS/Active-Active product.

- Multisite workload configuration

In this configuration, the production systems run in both sites, Site1 and Site2. This configuration typically uses the full benefits of data sharing available with a Parallel Sysplex. Having two GDPS controlling systems, one in each site, is preferable. Although you might also hear this referred to as an *Active/Active GDPS/MTMM* configuration, we avoid the Active/Active term to avoid confusion with the same term used in conjunction with the GDPS/Active-Active product.

- Business Recovery Services (BRS) configuration

In this configuration, the production systems and the controlling system are all in the same site, referred to as Site1. Site2 can be a client site or can be owned by a third-party recovery services provider (thus the name BRS). You might hear this referred to as an *Active/Cold* configuration.

These configuration options are described in more detail in the following sections.

7.2.1 Controlling system

Why does a GDPS/MTMM configuration need a controlling system? At first, you might think this is an additional infrastructure overhead. However, when you have an unplanned outage that affects production systems or the disk subsystems, it is crucial to have a system such as the controlling system that can survive failures that might have impacted other portions of your infrastructure. The controlling system allows you to perform situation analysis after the unplanned event to determine the status of the production systems or the disks, and then to drive automated recovery actions. The controlling system plays a vital role in a GDPS/MTMM configuration.

The controlling system must be in the same sysplex as the production system (or systems) so it can see all the messages from those systems and communicate with those systems. However, it shares an absolute minimum number of resources with the production systems (typically just the couple data sets). By being configured to be as self-contained as possible, the controlling system will be unaffected by errors that can stop the production systems (for example, an Extended Long Busy event on a primary volume).

The controlling system must have connectivity to all the Site1 and Site2 primary and secondary devices that it will manage. If available, it is preferable to isolate the controlling system infrastructure on a disk subsystem that is not housing mirrored disks that are managed by GDPS.

The controlling system is responsible for carrying out all recovery actions following a disaster or potential disaster, for managing the disk mirroring configuration, for initiating a HyperSwap, for initiating a freeze and implementing the freeze/swap policy actions, for reassigning STP roles; for re-IPLing failed systems, and so on.

Note: The availability of the dedicated GDPS controlling system (or systems) in *all* configurations is a fundamental requirement of GDPS. It is not possible to merge the function of the controlling system with any other system that accesses or uses the primary volumes or other production resources.

Configuring GDPS/MTMM with two controlling systems, one in each site is highly recommended. This is because a controlling system is designed to survive a failure in the opposite site of where the primary disks are. Primary disks are normally in Site1 and the controlling system in Site2 is designed to survive if Site1 or the disks in Site1 fail. However, if you reverse the configuration so that primary disks are now in Site2, the controlling system is in the same site as the primary disks. It will certainly not survive a failure in Site2 and might not survive a failure of the disks in Site2 depending on the configuration. Configuring a controlling system in both sites ensures the same level of protection, no matter which site is the primary disk site. When two controlling systems are available, GDPS manages assigning a Master role to the controlling system that is in the same site as the secondary disks and switching the Master role if there is a disk switch.

Improved controlling system availability: Enhanced timer support

Normally, a loss of synchronization with the sysplex timing source will generate a disabled console WTOR that suspends all processing on the LPAR, until a response is made to the WTOR. The WTOR message is IEA394A in STP timing mode.

In a GDPS environment, z/OS is aware that a given system is a GDPS controlling system and will allow a GDPS controlling system to continue processing even when the server it is running on loses its time source and becomes unsynchronized. The controlling system is therefore able to complete any freeze or HyperSwap processing it might have started and is

available for situation analysis and other recovery actions, instead of being in a disabled WTOR state.

In addition, because the controlling system is operational, it can be used to help in problem determination and situation analysis during the outage, thus further reducing the recovery time needed to restart applications.

The controlling system is required to perform GDPS automation in the event of a failure. Actions might include these tasks:

- ▶ Reassigning STP roles
- ▶ Performing the freeze processing to guarantee secondary data consistency
- ▶ Coordinating HyperSwap processing
- ▶ Executing a takeover script
- ▶ Aiding with situation analysis

Because the controlling system needs to run with only a degree of time synchronization that allows it to correctly participate in heartbeat processing with respect to the other systems in the sysplex, this system should be able to run unsynchronized for a period of time (80 minutes) using the local time-of-day (TOD) clock of the server (referred to as *local timing mode*), instead of generating a WTOR.

Automated response to STP sync WTORs

GDPS on the controlling systems, using the BCP Internal Interface, provides automation to reply to WTOR IEA394A when the controlling systems are running in local timing mode. See “Improved controlling system availability: Enhanced timer support” on page 205. A server in an STP network might have recovered from an unsynchronized to a synchronized timing state without client intervention. By automating the response to the WTORs, potential time outs of subsystems and applications in the client’s enterprise might be averted, thus potentially preventing a production outage.

If WTOR IEA394A is posted for production systems, GDPS uses the BCP Internal Interface to automatically reply RETRY to the WTOR. If z/OS determines that the CPC is in a synchronized state, either because STP recovered or the CTN was reconfigured, it will no longer spin and continue processing. If the CPC is still in an unsynchronized state when GDPS automation responded with RETRY to the WTOR, however, the WTOR will be reposted.

The automated reply for any given system is retried for 60 minutes. After 60 minutes, you will need to manually respond to the WTOR.

7.2.2 Single-site workload configuration

A GDPS/MTMM single-site workload environment typically consists of a multisite sysplex, with all production systems running in a single site, normally Site1, and the GDPS controlling system in Site2. The controlling system (or systems, because you may have two in some configurations) will normally run in the site containing the secondary disk volumes.

The multisite sysplex can be a base sysplex or a Parallel Sysplex; a coupling facility is not strictly required. The multisite sysplex must be configured with redundant hardware (for example, a coupling facility and a Sysplex Timer in each site), and the cross-site connections must also be redundant. Instead of using Sysplex Timers to synchronize the servers, you can also use Server Time Protocol (STP) to synchronize the servers.

Figure 7-2 shows a typical GDPS/MTMM single-site workload configuration. LPARs P1 and P2 are in the production sysplex, as are the coupling facilities CF1, CF2, and CF. The primary (H1) disks are in Site1, with a set of secondaries (H2) also in Site1 and another set of secondaries (H3) in Site2. All the production systems are running in Site1, with only the GDPS controlling system (K1) running in Site2. You will notice that system K1's disks (those marked K1) are also in Site2.

The GDPS/MTMM code itself runs under NetView and System Automation, and runs in every system in the GDPS sysplex.

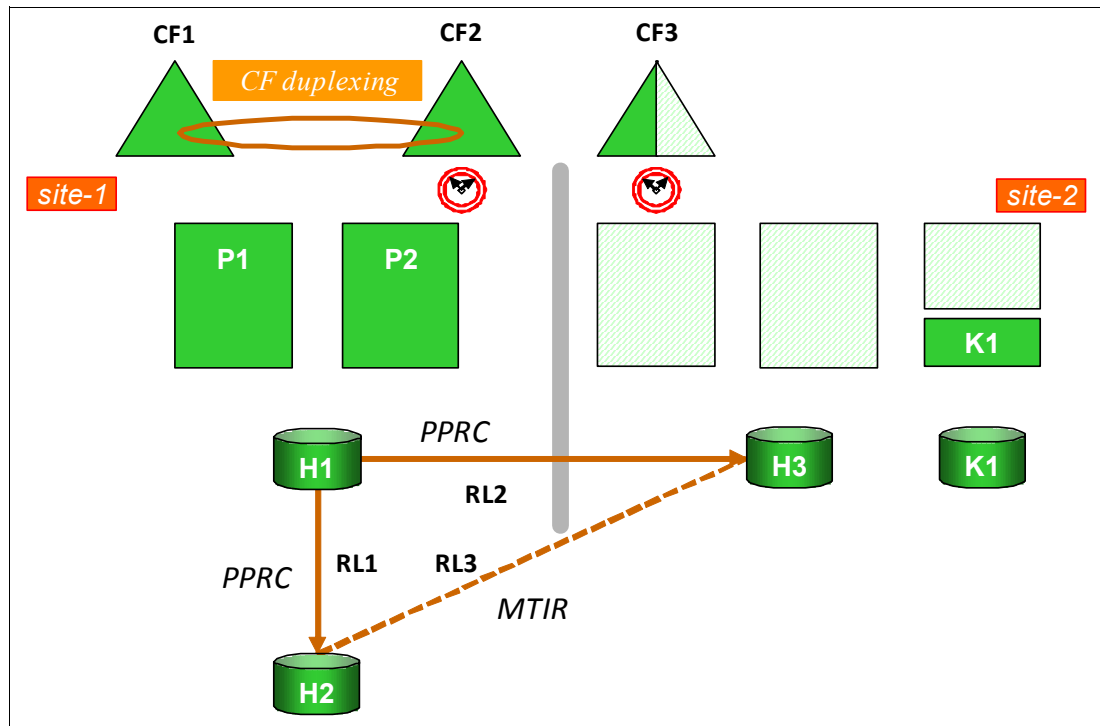


Figure 7-2 GDPS/MTMM single site workload configuration

7.2.3 Multisite workload configuration

A multisite workload configuration, shown in Figure 7-3, differs from a single-site workload in that production systems are running in *both* sites. Although, running a multisite workload as a base sysplex is possible, seeing this configuration as a base sysplex (that is, without coupling facilities) is unusual. This is because a multisite workload is usually a result of higher availability requirements, and Parallel Sysplex and data sharing are core components of such an environment.

Because in this example we have production systems in both sites, we need to provide the capability to recover from a failure in either site. So, in this case, there is also a GDPS controlling system with its own local (not mirrored) disk running in Site1, namely System K2. Therefore, if there is a disaster that disables Site2, there will still be a GDPS controlling system available to decide how to react to that failure and what recovery actions are to be taken.

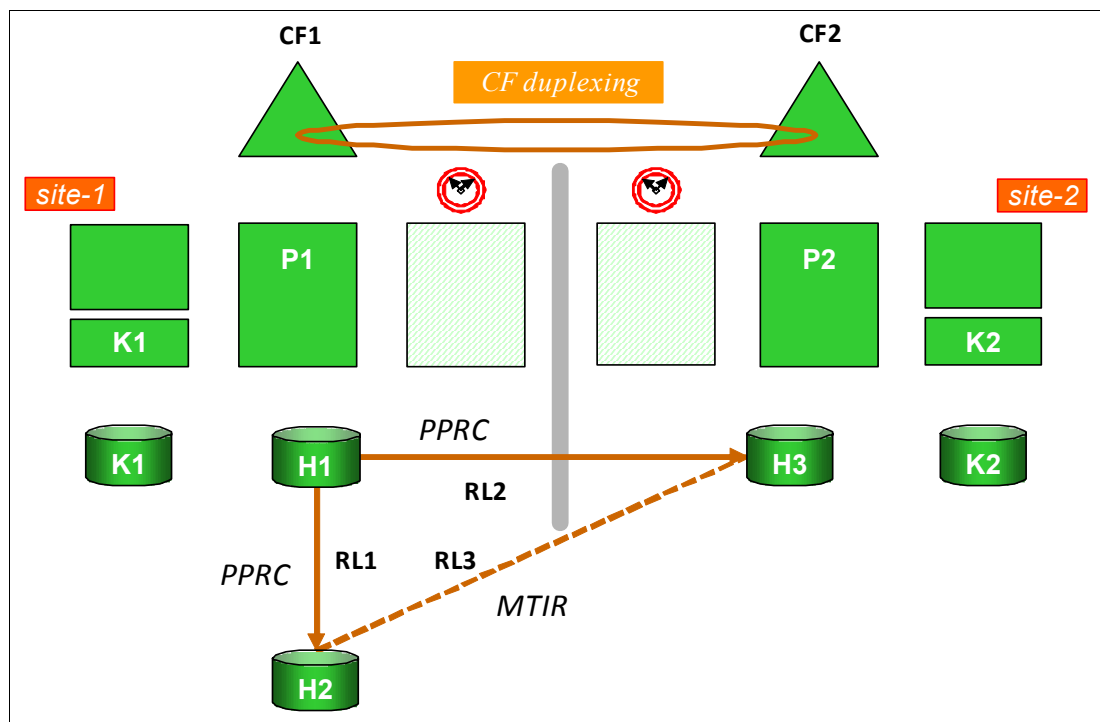


Figure 7-3 GDPS/MTMM multisite workload configuration

7.2.4 Business Recovery Services (BRS) configuration

A third configuration is known as the *BRS configuration*, and is illustrated in Figure 7-4 on page 209. In this configuration, all the systems in the GDPS configuration, including the controlling system, are in a sysplex in the same site, namely Site1. The sysplex does not span the two sites. The second site, Site2, might be a client site or might be owned by a third-party recovery services provider; thus the name BRS.

Site2 will contain the secondary disks and the alternate couple data sets (CDS), and might also contain processors that will be available in case of a disaster, but are not part of the configuration. This configuration can also be used when the distance between the two sites exceeds the distance supported for a multisite sysplex, but is within the maximum distance supported by FICON and Metro Mirror.

Even though there is no need for a multisite sysplex with this configuration, you must have channel connectivity from the GDPS systems to the secondary disk subsystems. Also, as explained in the next paragraph, the controlling system in Site1 will need channel connectivity to its disk devices in Site2. Therefore, FICON link connectivity from Site1 to Site2 will be required. See 2.9.7, “Connectivity options” on page 47, and *IBM z Systems Connectivity Handbook*, SG24-5444, for options available to extend the distance of FICON links between sites.

In the BRS configuration one of the two controlling systems must have its disk devices in Site2. This permits that system to be restarted manually in Site2 after a disaster is declared. After it restarts in Site2, the system runs a GDPS script to recover the secondary disk subsystems, reconfigure the recovery site, and restart the production systems from the disk subsystems in Site2.

If you have only a single controlling system and you have a total cross-site fiber connectivity failure, the controlling system running on Site2 disks might not be able to complete the Freeze operation because it will lose access to its disk in Site2. Having a second controlling system running on Site1 local disks in will guarantee that the freeze operation completes successfully in the event the controlling system running on Site2 disks is down or is unable to function because of a cross-site fiber loss. GDPS will attempt to maintain the current Master system in the controlling system by using the secondary disks.

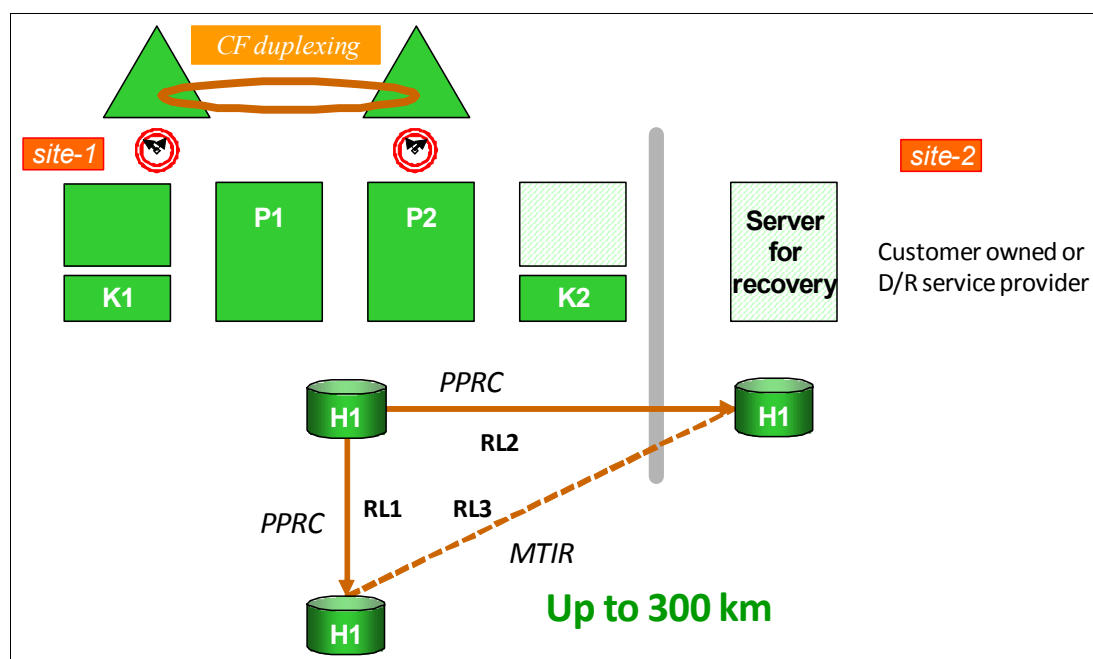


Figure 7-4 GDPS/MTMM BRS configuration

7.2.5 Combining GDPS/MTMM with GDPS/XRC

GDPS/MTMM supports the existence of an additional XRC leg to be configured using the PPRC primary disk. In such a configuration, XRC Incremental Resynchronization (IR) is *not supported*.

If a HyperSwap or a recovery is performed on one of the PPRC legs, you can establish XRC from the new primary, provided that you have connectivity from the new primary devices to the XRC recovery region. However, a full initial copy will be required.

7.2.6 Combining GDPS/MTMM with GDPS/GM in a 4-site configuration

GDPS/MTMM (managing a single synchronous replication leg) can be combined with GDPS/GM in 3-site and 4-site configurations. In such configurations, GDPS/MTMM (when combined with Parallel Sysplex use and HyperSwap) in one region provides continuous availability across a metropolitan area or within the same local site, and GDPS/GM provides disaster recovery capability using a remote site in a different region.

The 4-site environment is configured in a symmetric manner so that there is a GDPS/MTMM-managed replication leg available in both regions to provide continuous availability (CA) within the region, with GDPS/GM to provide cross-region DR, no matter in which region production is running at any time.

This combination is referred to as GDPS/MGM Multi-Target.

See Chapter 11, “Combining local and metro continuous availability with out-of-region disaster recovery” on page 331 for more information about GDPS/MGM configurations.

7.2.7 Other considerations

The availability of the dedicated GDPS controlling system (or systems) in *all* scenarios is a fundamental requirement in GDPS. Merging the function of the controlling system with any other system that accesses or uses the primary volumes is not possible.

Equally important is that certain functions (stopping and restarting systems and changing the couple data set configuration) are done through the scripts and panel interface provided by GDPS. Because events such as systems going down or changes to the couple data set configuration are indicators of a potential disaster, such changes must be initiated using GDPS functions so that GDPS understands that these are planned events.

7.3 Multiplatform Resiliency for System z (also known as xDR)

To reduce IT costs and complexity, many enterprises are consolidating open servers into Linux on System z servers. Linux on System z systems can be implemented either as guests running under z/VM or native Linux on System z systems. Several examples exist of an application server running on Linux on System z and a database server running on z/OS. Two examples are as follows:

- ▶ WebSphere Application Server running on Linux and CICS, DB2 running under z/OS
- ▶ SAP application servers running on Linux and database servers running on z/OS

With a multitiered architecture, there is a need to provide a coordinated near-continuous availability and disaster recovery solution for both z/OS and Linux on System z. The GDPS/MTMM function that provides this capability is called Multiplatform Resiliency for System z, and it can be implemented if the disks being used by z/VM and Linux are CKD disks. For more details about this function, see 10.2, “GDPS/PPRC Multiplatform Resiliency for z Systems” on page 299.

Note that only Linux on System z systems implemented as guests running under z/VM are supported in GDPS/MTMM environments; native Linux on System z systems are not supported.

7.4 Managing the GDPS environment

We have seen how GDPS/MTMM can protect just about any type of data that can reside in a disk subsystem. It can also provide data consistency across multiple platforms. However, as discussed in Chapter 1, “Introduction to business resilience and the role of GDPS” on page 1, the overwhelming majority of System z outages are not disasters. Most are planned outages, with a small percentage of unplanned ones.

In this section, we describe the other aspect of GDPS/MTMM, that is, its ability to monitor and manage the resources in its environment. GDPS provides two mechanisms to help you manage the GDPS sysplex and resources within that sysplex. One mechanism is the NetView interface and the other is support for scripts. We review both of these mechanisms here.

7.4.1 NetView interface

The user interface for GDPS/MTMM is called the NetView 3270 panel interface. An example of the main GDPS/MTMM panel is shown in Figure 7-5.

```
VPCPPNLN                                GDPS - MTMM                                GDPS V3.R11.M3

      ---- GDPS Status Indicators ----

System      = G1C1                      Remote Copy Status = OK
Current Master = G1C2                  Primary Dasd = H1
Debug       = ON

      ---- GDPS Options ----

1      Dasd Remote Copy                7      Sysplex Resource Management
3      Standard Actions                 8      Debug ON/OFF
6      Planned Actions                  9      View Definitions
                                           C      Config Management
                                           M      Run Monitor1/Monitor3

Selection ==> _

      Licensed Materials - Property of IBM
      6942-35B © Copyright IBM Corp. 1998, 2014 All Rights Reserved.
```

Figure 7-5 GDPS/MTMM Main Panel (VPCPPNLN)

This panel has a summary of configuration status at the top, and a menu of selectable choices. As an example, to view the disk mirroring (Dasd Remote Copy) panels enter 1 at the Selection prompt, and then press Enter.

Monitoring function: Status Display Facility

GDPS also provides many monitors to check the status of disks, sysplex resources, and other GDPS-managed resources. Any time there is a configuration change, or something in GDPS that requires manual intervention, GDPS will raise an alert. GDPS uses the Status Display Facility (SDF) provided by System Automation as the primary status feedback mechanism for GDPS. It is the only dynamically updated status display available for GDPS.

GDPS provides a dynamically updated color-coded SDF panel, as shown in Figure 7-6. If something changes in the environment that requires attention, the color of the associated field on the panel will change. At all times, the operators need to have an SDF panel within view so they will immediately become aware of anything requiring intervention or action.

SDF - G1C2GDPS		GDPS Disaster/Recovery Readiness		12/22/14 07:01:33	
Site 1 -----		Site 2 -----			
Automation	1/10(18)	RemoteCopy	1/4(4)	Automation	1/7(7)
G1C1CPC		G1P1&CKDRL1		G1C2S	
G1C1CPC		G1P1&CKDRL2		G1C2S	
G1P1CB		G1C1&CKDRL1		G1C2S	
G1P1S		G1C1&CKDRL2		G1C2S	
G1P1S				G1C2S	
G1P1S				G1C2S	
G1P1S				G1C2S	
G1P1S				G1C2S	
G1P1S				G1C2S	
G1P1S				G1C2S	
G1P1S				G1C2S	
----- Trace Entries -----					
Time	Text	1/11(67)			
17:10:02	CDELPATH 0526F 520D52 5005076306FFD08A 521050 5005076305FFC51B	(			
17:10:02	CDELPATH 0526F RC=7160 REAS=000000008 (17:10:02.411110)				
17:10:02	ANTP0121I DELETE PATHS FAILED- PPRC VOLUMES ARE STILL ACTIVE				
17:10:02	CDELPATH 0540F 540D54 5005076306FFD08A 541052 5005076305FFC51B	(			
17:10:02	CDELPATH 0540F RC=7160 REAS=000000008 (17:10:02.422177)				
20:43:05	CESTPAIR 05301 530D53 00DCD61 01 531051 00FGLP1 01 S--- R0---	(2			
20:43:05	CESTPAIR 05301 RC=7200 REAS=000000008 (20:43:05.046559)				
20:43:05	ANTP0102I ESTAB. PAIR FAILED- DEVICES NOT IN SUSPEND MODE				
20:43:05	CESTPAIR 05302 530D53 00DCD61 02 531051 00FGLP1 02 S--- R0---	(2			
20:43:05	CESTPAIR 05302 RC=7200 REAS=000000008 (20:43:05.052882)				
20:43:05	CESTPAIR 05304 530D53 00DCD61 04 531051 00FGLP1 04 S--- R0---	(2			
====>					
1=Help		2=Detail 3=Return		4=Roll	
				9=Bottom 10=Previous 11=Next 12=Top	
				21=DTRACE 22=DDASD 23=DAUT	

Figure 7-6 GDPS SDF panel

The GDPS SDF panel is divided in two parts: the top part contains status indicators, and the lower part is for trace entries. The status indicators are color-coded, with green meaning that the status is good. Minor problems are indicated by the color pink. And serious problems are shown in red. The goal is to have all status indicators green.

Remote copy panels

The z/OS Advanced Copy Services capabilities are powerful, but the native command-line interface (CLI), z/OS TSO, and ICKDSF interfaces are not as user-friendly as the GDPS DASD remote copy panels are. To more easily check and manage the remote copy environment, you can use the DASD remote copy panels provided by GDPS.

For GDPS to manage the remote copy environment, you must first define the configuration (primary and secondary LSSs, primary and secondary devices, and PPRC links) to GDPS in a file called the GEOPARM file.

After the configuration is known to GDPS, you can use the panels to check that the current configuration matches the one you want. You can start, stop, suspend, and resynchronize mirroring. These actions can be done at the device or LSS level, or both, as appropriate for a selected replication leg.

Figure 7-7 shows the Replication Leg Status and Policies panel.

```

VPCPQSTM      GDPS - MTMM      Replication leg Status and Policies      G1C2

Group =      CKD
Primary dasd:  H1

Multi Target:  OK

Actions:  V iew

      Locations      HyperSwap  Mirror  Preferred      PRIMARYFAILURE  PPRCFailure
      Locations      HyperSwap  Mirror  policies      policy          policy
-  RL1    H1-H2    ENABLED    OK      PREF,RPFC      G0              G0
-  RL2    H1-H3    ENABLED    OK
-  RL3    H2-H3              MTIR          G0              G0

1 Copyopt      2 Policy mgmt      3 View Config/Freeze time
5 Monitor2
Selection ==>
F1=Help      F3=Return      F6=Roll

```

Figure 7-7 DASD Remote Copy Status panel (VPCPQSTM)

The Replication Leg Status and Policies panel, is organized into three sections:

- ▶ The top section displays information related to the entire configuration, including the overall mirroring status and HyperSwap status.
- ▶ The middle section displays the replication legs, along with information related to each replication leg, including the current mirroring status, HyperSwap status, and policy information.
- ▶ The bottom section contains a list of actions that can be accessed by entering the selection number associated with each action.

View SSID Pairs panel

Entering a V (view) line command for a replication leg (on the Replication Leg Status and Policies panel, shown in Figure 7-7 on page 213) presents the panel shown in Figure 7-8.

```
VPCPQSTE  Dasd Mirroring Status =  OK      Monitor2 time = 03:17:46  G1C1
          Soft Fenced =NONE
Actions: Q ueryPath Z QueryReverse V view devices X ceptions D elpath E stpath
          Tot Pairs: 108
          Group: CKD      Replication Leg: RL1      Primary Dasd: H1
          -PRI-SSID-SEC-  UTIL-DEVICES  ----- LINKS 1-4 (LINK-STATUS) -----
_ 520D === 5210  0526F-1526F  01020202-13 01030232-13
_ 530D === 5310  0536F-1536F  01020202-13 01030232-13
_ 540D === 5410  0540F-1540F  01020202-13 01030232-13

1 Epair 2 Dpair 3 Suspend 4 Resynch 5 Monitor2 6 Q Paths 7 Epath 8 Dpath
11 Find 21 FCEp 22 FCEs 23 FCWp 24 FCWs 31 DSS List
Selection ==>
F1=Help  F3=Return  F6=Roll  F7=Up  F8=Down  F9=Toggle  F11=Right
```

Figure 7-8 View Storage Subsystems Status panel (VPCPQSTE)

This panel contains a lot of information and is also the place where many disk-related actions can be initiated for the selected leg. It is entirely possible that everything is working to plan on one replication leg at the same time that another replication leg is experiencing problems.

If you are familiar with using the TSO or ICKDSF interfaces, you might appreciate the ease of use of the DASD remote copy panels.

Remember that these panels provided by GDPS are *not* intended to be a remote copy monitoring tool. Because of the overhead involved in gathering the information for every device to populate the NetView panels, GDPS gathers this data only on a timed basis, or on demand following an operator instruction. The normal interface for finding out about remote copy status or problems is the Status Display Facility (SDF).

Standard Actions

GDPS provides facilities to help manage many common system-related planned actions. There are two reasons to use the GDPS facilities to perform these actions known as *Standard Actions*:

- ▶ They are well tested and based on IBM preferred procedures.
- ▶ Using the GDPS interface lets GDPS know that the changes that it is seeing (for example, a system being partitioned out of the sysplex) are planned changes, and therefore GDPS is not to react to these events.

Standard Actions are single-step actions, or are intended to impact only one resource. Examples are starting a system IPL, maintaining the various IPL address and load parameters that can be used to IPL a system, selecting the IPL address and load parameters to be used the next time a system IPL is performed, or activating/deactivating an LPAR.

If you want to stop a system, change its IPL address, then perform an IPL, you initiate three separate Standard Actions, one after the other. GDPS scripting is a facility that is suited to multi-step, multi-system actions.

The GDPS/MTMM Standard Actions panel is shown in Figure 7-9. It displays all the systems being managed by GDPS/MTMM, and for each one it shows the current status and various IPL information. To perform actions on each system, you simply use a line command letter (L to load, X to reset and so on) next to the selected system.

VPCPSTD1		Standard Actions							G1C2
Actions:		S Stop	R ReIPL	M odify					
		L Load	X Reset	A Activate	D Deactivate	U Dump			
Sysname	IND	Status	IPL type	LPAR	IPLmode	Auto	L-addr	Loadparm	
— SITE1		H1M0P_A							
— G1C1	C	????	NORMAL	S69E		YN			
— G1P1		ACTIVE	NORMAL	S294	NORMAL	YN	5200	520712M	
— CF1		MANUAL	NORMAL	S67A		NN			
— SITE2		H3M0P_2							
— G1C2	C	MASTER	NORMAL	S293		YN			

1 CPC Ops
Selection ==>
F1=Help F3=Return F6=Roll F11=Right

Figure 7-9 GDPS/MTMM Standard Actions panel (VPCPSTD1)

GDPS supports taking a stand-alone dump using the GDPS Standard Actions panel. Clients using GDPS facilities to perform HMC actions no longer need to use the HMC for taking stand-alone dumps.

Sysplex resource management

There are certain resources that are vital to the health and availability of the sysplex. In a multisite sysplex, it can be quite complex trying to manage these resources to provide the required availability while ensuring that any changes do not introduce a single point of failure.

The GDPS/MTMM Sysplex Resource Management panel, shown in Figure 7-10 on page 216, provides you with the ability to manage the resources, with knowledge about where the resources exist. For example, normally you have Primary Couple Datasets (CDS) in Site1, and your alternates in Site2. However, if you will be shutting down Site1, you still want to have a Primary and Secondary set of CDS, but both must be in Site2. The GDPS Sysplex Resource Management panels provide this capability, without you having to know specifically where each CDS is located.

GDPS provides facilities to manage coupling facilities (CFs) in your sysplex. These facilities allow for isolating all of your structures in the CF or CFs in a single site and returning to your normal configuration with structures spread across (and possibly duplexed across) the CFs in the two sites.

Isolating structures into CFs in one site, or returning to normal use with structures spread across CFs in both sites, can be accomplished through the GDPS Sysplex Resource Management panel interface or GDPS scripts. This provides an automated means for managing CFs for planned and unplanned site or disk subsystem outages.

The maintenance mode switch allows you to start or stop maintenance mode on a single CF (or multiple CFs, if all selected CFs are in the same site). DRAIN, ENABLE, and POPULATE function is still available for single CFs.

```

VPCSPM1          Sysplex Resource Management          G1C2
Actions:  View    Modify    EXclude    Include        CTNID: CTRL
                                           CTS:

----- Couple Datasets -----
TYPE      Primary DSN          Alternate DSN
- SYSPLEX  SYS1.G1.XCF.CDSH10   SYS1.G1.XCF.CDSH30
- ARM      SYS1.G1.ARM.CDSH10   SYS1.G1.ARM.CDSH30
- BPXMCDS  SYS1.G1.OMVS.CDSH10  SYS1.G1.OMVS.CDSH30
- CFRM     SYS1.G1.CFRM.CDSH10  SYS1.G1.CFRM.CDSH30
- LOGR     SYS1.G1.LOGR.CDSA2   SYS1.G1.LOGR.CDSA3
- SFM      SYS1.G1.SFM.CDSH10   SYS1.G1.SFM.CDSH30
- WLM      SYS1.G1.WLM.CDSH10   SYS1.G1.WLM.CDSH30

----- Coupling Facilities -----
POLICY     Site1              Site2
- POLA     CF1                CF2
           -                  -
           -                  -
           -                  -

1 Activate New CDSes configuration
4 Use Site1 CF-s      5 USE Site2 CF-s      6 USE Normal CF-s

Selection ==> -
F1=Help          F3=Return          F6=Roll

```

Figure 7-10 Sysplex Resource Management main menu (VPCSPM1)

7.4.2 GDPS scripts

At this point we have shown how GDPS panels provide powerful functions to help you manage GDPS resources. However, using GDPS panels is only one way of accessing this capability. Especially when you need to initiate what might be a complex, compound, multistep procedure involving multiple GDPS resources, it is much simpler to use a script which, in effect, is a workflow.

Nearly all of the main functions that can be initiated through the GDPS panels are also available using GDPS scripts. Scripts also provide additional capabilities that are not available using the panels.

A “script” is simply a procedure recognized by GDPS that pulls together one or more GDPS functions. Scripts can be initiated manually for a planned activity through the GDPS panels (using the Planned Actions interface), automatically by GDPS in response to an event (HyperSwap), or through a batch interface. GDPS performs the first statement in the list, checks the result, and only if it is successful, proceeds to the next statement. If you perform the same steps manually, you would have to check results, which can be time-consuming, and initiate the next action. With scripts, the process is automated.

Scripts can easily be customized to automate the handling of various situations, both to handle planned changes and unplanned situations. This is an extremely important aspect of GDPS. Scripts are powerful because they can access the full capability of GDPS. The ability to invoke all the GDPS functions through a script provides the following benefits:

- Speed

The script will execute the requested actions and check results at machine speeds. Unlike a human, it does not need to search for the latest procedures or the commands manual.

- Consistency

If you were to look into most computer rooms immediately following a system outage, what would you see? Mayhem, with operators frantically scrambling for the latest system programmer instructions. All the phones ringing. Every manager within reach asking when the service will be restored. And every systems programmer with access vying for control of the keyboards. All this results in errors because humans naturally make mistakes when under pressure. But with automation, your well-tested procedures will execute in exactly the same way, time after time, regardless of how much you shout at them.

- Thoroughly tested procedures

Because they behave in a consistent manner, you can test your procedures over and over until you are sure they do everything that you want, in exactly the manner that you want. Also, because you need to code everything and cannot assume a level of knowledge (as you might with instructions intended for a human), you are forced to thoroughly think out every aspect of the action the script is intended to undertake. And because of the repeatability and ease of use of the scripts, they lend themselves more easily to frequent testing than manual procedures.

Planned Actions

As mentioned earlier, GDPS scripts are simply procedures that pull together into a list one or more GDPS functions. For the scripted procedures that you might use for a planned change, these scripts can be initiated from the panels called *Planned Actions* (option 6 on the main GDPS panel as shown in Figure 7-5 on page 211).

As one example, you can have a short script that stops a system and then re-IPLs it in an alternate LPAR location, as shown in Example 7-1. The sample also handles deactivating the original LPAR after the system is stopped and activating the alternate LPAR before the system is IPLed in this location.

Example 7-1 Sample script to re-IPL a system

```
COMM='Example script to re-IPL system SYS1 on alternate ABNORMAL LPAR location'
SYSPLEX='STOP SYS1'
SYSPLEX='DEACTIVATE SYS1'
IPLTYPE='SYS1 ABNORMAL'
SYSPLEX='ACTIVATE SYS1 LPAR'
SYSPLEX='LOAD SYS1'
```

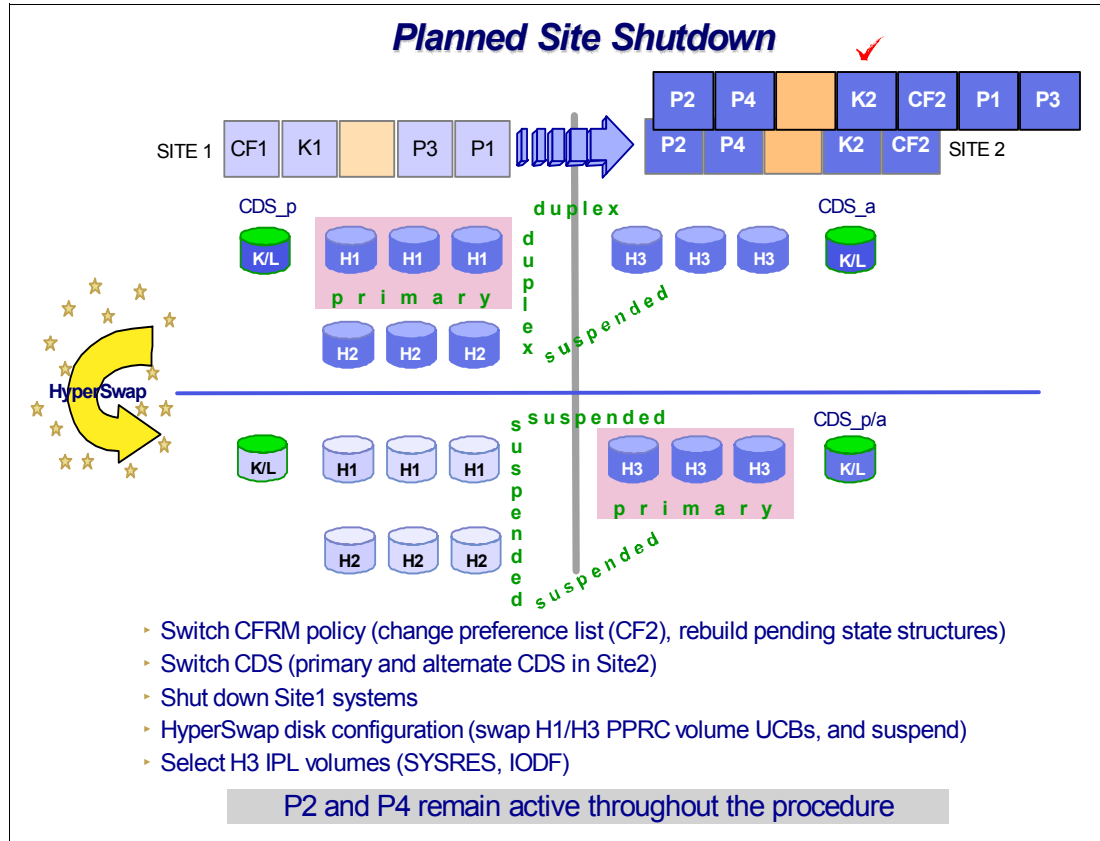


Figure 7-11 GDPS/MTMM Planned Action

A more complex example of a Planned Action is shown in Figure 7-11. In this example, a single action in GDPS executing a planned script of only a few lines results in a complete planned site switch. Specifically, the following actions are done by GDPS:

- ▶ The systems in Site1, P1 and P3, are stopped (P2 and P4 remain active in this example).
- ▶ The sysplex resources (CDS and CF) are switched to use only those in Site2.
- ▶ A HyperSwap is executed to use the disk in Site2 (H3 disk). As a result of the swap GDPS automatically switches the IPL parameters (IPL address and load parameters) to reflect the new configuration.
- ▶ The IPL location for the P1 and P3 systems are changed to the backup LPAR location in Site2.
- ▶ The backup LPAR locations for P1 and P3 systems are activated.
- ▶ P1 and P3 are IPLed in Site2 using the disk in Site2.

Using GDPS removes the reliance on out-of-date documentation, provides a single repository for information about IPL addresses and load parameters, and ensures that the process is done the same way every time with no vital steps accidentally overlooked.

STP CTN role reassignments: Planned operations

GDPS provides a script statement that allows you to reconfigure an STP-only CTN by reassigning the STP-only CTN server roles. In an STP CTN servers (CPCs) are assigned special roles to identify which CPC is preferred to be the clock source (Preferred Time Server, or PTS), which CPC is able to take over as the clock source for planned and unplanned events (Backup Time Server, or BTS), which CPC is the active clock source (Current Time Server, or CTS), and which CPC assists in STP recovery (Arbiter).

It is strongly recommended that the server roles be reassigned before performing planned disruptive actions on any of these special role servers. Examples of planned disruptive actions are power-on reset (POR) and Activate/Deactivate. The script statement can be integrated as part of your existing control scripts to perform these planned disruptive actions.

For example, if you are planning to deactivate the CPC that is the PTS/CTS, you can now execute a script to perform the following tasks:

- ▶ Reassign the PTS/CTS role to a different CPC in the CTN
- ▶ Optionally also reassign the BTS and Arbiter roles if required
- ▶ Execute script statements you might already have in place today to deactivate the PTS/CTS CPC

After the disruptive action is completed you can execute a second script to restore the STP roles to their normal operational state, as listed here:

- ▶ Script statement to activate the CPC
- ▶ Reassign the STP server roles to their normal operational state
- ▶ Statements you might already have in existing scripts to perform IPLs and so on

Post swap scripts

These are scripts, also known as Takeover Scripts, that are intended to define actions that GDPS will execute after an unplanned HyperSwap. There are a number of specific unplanned HyperSwap scenarios and for each one, there is a reserved name for the associated Takeover script. In the case of an unplanned HyperSwap trigger, GDPS/MTMM will immediately and automatically execute an unplanned HyperSwap. Following the HyperSwap operation, GDPS will then execute the appropriate takeover script if it has been defined.

The post swap Takeover scripts have reserved names which helps GDPS determine the applicability of the script for the given unplanned swap situation. For example, if there is an unplanned swap from H1 to H3, GDPS will, if you have defined it, automatically schedule a script named SWAPSITE13.

As previously mentioned, these scripts provide you with the facility to automatically perform actions that you may want to take following an unplanned HyperSwap. Typical actions you may want to perform following an unplanned HyperSwap include resynchronizing mirroring for the MTIR replication leg and changing the couple data set configuration. For HyperSwap operations that swap production from one site to another, you might want to reconfigure STP to keep the CTS role on the CPC that is in the same site as the swapped-to, new primary devices.

Scripts for other unplanned events

GDPS monitors data-related events and also performs system-related monitoring. When GDPS detects that a z/OS system is no longer active, it verifies whether the policy definition indicates that Auto IPL has been enabled, that the threshold of the number of IPLs in the predefined time window has not been exceeded, and that no planned action is active. If these conditions are met, GDPS can automatically re-IPL the system in place, bring it back into the Parallel Sysplex, and restart the application workload (Figure 7-12).

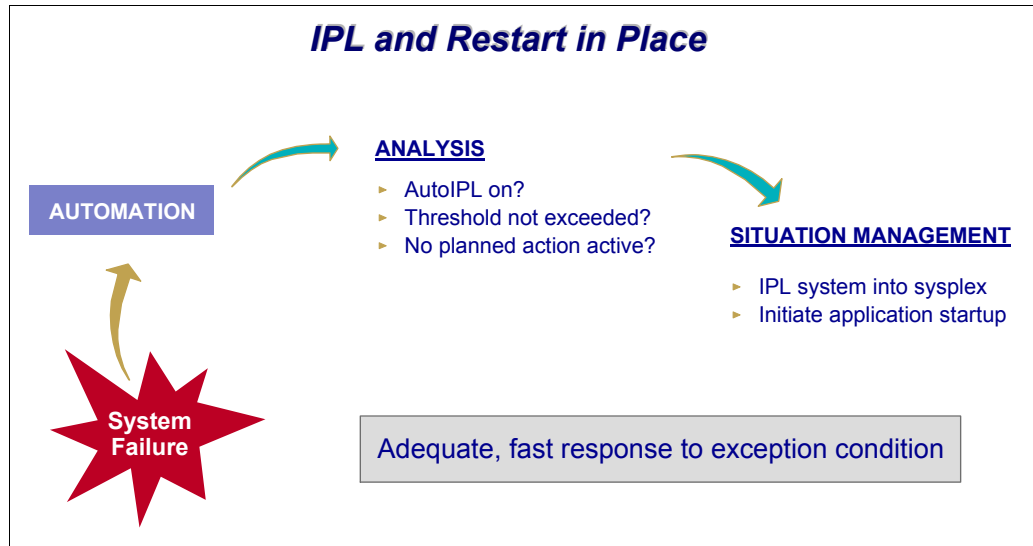


Figure 7-12 Recovering a failed image

Although Auto IPL processing takes place automatically based on policy and does not require a script, you can have scripts prepared to provide recovery for similar events, such as a complete processor failure. In such a script, you would want to activate backup partitions for all the systems on that processor, activate CBU if appropriate, and IPL these systems. You could have one such script prepared in advance for every server in your configuration.

STP CTN role reassignments: Unplanned failure

If a failure condition has resulted in the PTS, BTS, or Arbiter no longer being an operational synchronized CPC in the CTN, a suggestion is that after the failure and possible STP recovery action, the STP roles be reassigned to operational CPCs in the CTN. The reassignment reduces the potential for a sysplex outage in the event a second failure or planned action affects one of the remaining special role CPCs.

The script statement capability described in "STP CTN role reassignments: Planned operations" on page 219 can be used to integrate the STP role reassignment as part of an existing script and eliminate the requirement for the operator to perform the STP reconfiguration task manually at the HMC.

STP WTOR IEA394A response: Unplanned failure

As described in "Improved controlling system availability: Enhanced timer support" on page 205, a loss of synchronization with the sysplex timing source will generate a disabled console WTOR. This suspends all processing on the LPAR until a response to the WTOR is provided. The WTOR message is IEA394A if the CPC is in STP timing mode (either in an STP Mixed CTN or STP-only CTN).

GDPS, using scripts, can reply (either ABORT or RETRY) to the IEA394A sync WTOR for STP on systems that are spinning because of a loss of synchronization with their Current Time Source. As described in “Automated response to STP sync WTORs” on page 206, an autonomic function exists to reply RETRY automatically for 60 minutes on any GDPS systems that have posted this WTOR.

The script statement complements and extends this function, as described:

- ▶ It provides the means to reply to the message after the 60-minute automatic reply window expires.
- ▶ It can reply to the WTOR on systems that are not GDPS systems (foreign systems) that are defined to GDPS; the autonomic function only replies on GDPS systems.
- ▶ It provides the ability to reply ABORT on any systems you do not want to restart for a given failure scenario before reconfiguration and synchronization of STP.

Batch scripts

GDPS also provides a flexible batch interface to invoke planned action scripts. These scripts can be invoked:

- ▶ As a REXX program from a user terminal
- ▶ By using the IBM MVS MODIFY command to the NetView task
- ▶ From timers in NetView
- ▶ Triggered through the SA automation tables

This capability, along with the Query Services interface described in ????, provides a rich framework for user-customizable systems management procedures.

7.4.3 System Management actions

Most of the GDPS Standard Actions require actions to be done on the HMC. The interface between GDPS and the HMC is through the BCP Internal Interface (BCPii), and this allows GDPS to communicate directly with the hardware for automation of HMC actions such as Load, Stop (graceful shutdown), Reset, Activate LPAR, and Deactivate LPAR. GDPS can also perform ACTIVATE (power-on reset), CBU ACTIVATE/UNDO, OOCOD ACTIVATE/UNDO, and STP role reassignment actions against an HMC object that represents a CPC.

The GDPS LOAD and RESET Standard Actions (available through the Standard Actions panel or the SYSPLEX script statement) allow specification of a *CLEAR* or *NOCLEAR* operand. This provides operational flexibility to accommodate client procedures, thus eliminating the requirement to use the HMC to perform specific LOAD and RESET actions.

Furthermore, when you LOAD a system using GDPS (panels or scripts), GDPS can listen for operator prompts from the system being IPLed and reply to such prompts. GDPS provides support for optionally replying to such IPL-time prompts automatically, removing reliance on operator skills and eliminating operator error for selected messages that require replies.

SYSRES Management

Today many clients maintain multiple alternate z/OS SYSRES devices (also known as *IPLSETs*) as part of their maintenance methodology. GDPS provides special support to allow clients to identify IPLSETs. This removes the requirement for clients to manage and maintain their own procedures when IPLing a system on a different alternate SYSRES device.

GDPS can automatically update the IPL pointers after any disk switch or disk recovery action that changes the GDPS primary disk location indicator for PPRC disks. This removes the requirement for clients to perform additional script actions to switch IPL pointers after disk switches, and greatly simplifies operations for managing alternate SYSRES “sets.”

7.5 GDPS/MTMM monitoring and alerting

The GDPS SDF panel, discussed in “Monitoring function: Status Display Facility” on page 211, is where GDPS dynamically displays color-coded alerts.

Alerts can be posted as a result of an unsolicited error situation that GDPS listens for. For example, if one of the multiple PPRC links that provide the path over which PPRC operations take place is broken, there is an unsolicited error message issued. GDPS listens for this condition and will raise an alert on the SDF panel, notifying the operator of the fact that a PPRC link is not operational. Clients run with multiple PPRC links and if one is broken, PPRC continues over any remaining links. However, it is important for operations to be aware that a link is broken and fix this situation because a reduced number of links results in reduced PPRC bandwidth and reduced redundancy. If this problem is not fixed in a timely manner and more links fail, it can result in production impact because of insufficient mirroring bandwidth or total loss of PPRC connectivity (which results in a freeze).

Alerts can also be posted as a result of GDPS periodically monitoring key resources and indicators that relate to the GDPS/MTMM environment. If any of these monitoring items are found to be in a state deemed to be *not normal* by GDPS, an alert is posted on SDF.

Various GDPS monitoring functions are executed on the GDPS controlling systems and on the production systems. This is because, from a software perspective, it is possible that different production systems have different views of some of the resources in the environment, and although status can be normal in one production system, it can be not normal in another. All GDPS alerts generated on one system in the GDPS sysplex are propagated to all other systems in the GDPS. This propagation of alerts provides for a single focal point of control. It is sufficient for the operator to monitor SDF on the master controlling system to be aware of all alerts generated in the entire GDPS complex.

When an alert is posted, the operator will have to investigate (or escalate, as appropriate) and corrective action will need to be taken for the reported problem as soon as possible. After the problem is corrected, this is detected during the next monitoring cycle and the alert is cleared by GDPS automatically.

GDPS/MTMM monitoring and alerting capability is intended to ensure that operations are notified of and can take corrective action for any problems in their environment that can affect the ability of GDPS/MTMM to do recovery operations. This will maximize the chance of achieving your availability and RPO/RTO commitments.

7.5.1 GDPS/MTMM health checks

In addition to the GDPS/MTMM monitoring described, GDPS provides health checks. These health checks are provided as a plug-in to the z/OS Health Checker infrastructure to check that certain settings related to GDPS adhere to preferred practices.

The z/OS Health Checker infrastructure is intended to check a variety of settings to determine whether these settings adhere to z/OS optimum values. For settings found to be not in line with preferred practices, exceptions are raised in the Spool Display and Search Facility (SDSF). If these settings do not adhere to recommendations, this can hamper the ability of GDPS to perform critical functions in a timely manner.

Often, if there are changes in the client environment, this might necessitate adjustment of various parameter settings associated with z/OS, GDPS, and other products. It is possible that you can miss making these adjustments, which can affect GDPS. The GDPS health checks are intended to detect such situations and avoid incidents where GDPS is unable to perform its job because of a setting that is perhaps less than ideal.

For example, GDPS/MTMM provides facilities for management of the couple data sets (CDS) for the GDPS sysplex. One of the health checks provided by GDPS/MTMM checks that the couple data sets are allocated and defined to GDPS in line with the GDPS preferred practices recommendations.

Similar to z/OS and other products that provide health checks, GDPS health checks are optional. Several optimum values that are checked and the frequency of the checks can be customized to cater to unique client environments and requirements.

There are a few z/OS preferred practices that conflict with GDPS preferred practices. The related z/OS and GDPS health checks result in conflicting exceptions being raised. For such health check items, to avoid conflicting exceptions, z/OS provides the ability to define a *coexistence policy* where you can indicate which practice is to take precedence; GDPS or z/OS. GDPS provides sample coexistence policy definitions for the GDPS checks that are known to be conflicting with z/OS.

GDPS also provides a convenient interface for managing the health checks using the GDPS panels. You can use it to perform actions such as activate/deactivate or run any selected health check, view the customer overrides in effect for any optimum values, and so on. Figure 7-13 shows a sample of the GDPS Health Checks Information Management panel. In this example you see that all the health checks are enabled. The status of the last run is also shown, indicating that some were successful and some resulted in raising a medium exception. The exceptions can also be viewed using other options on the panel.

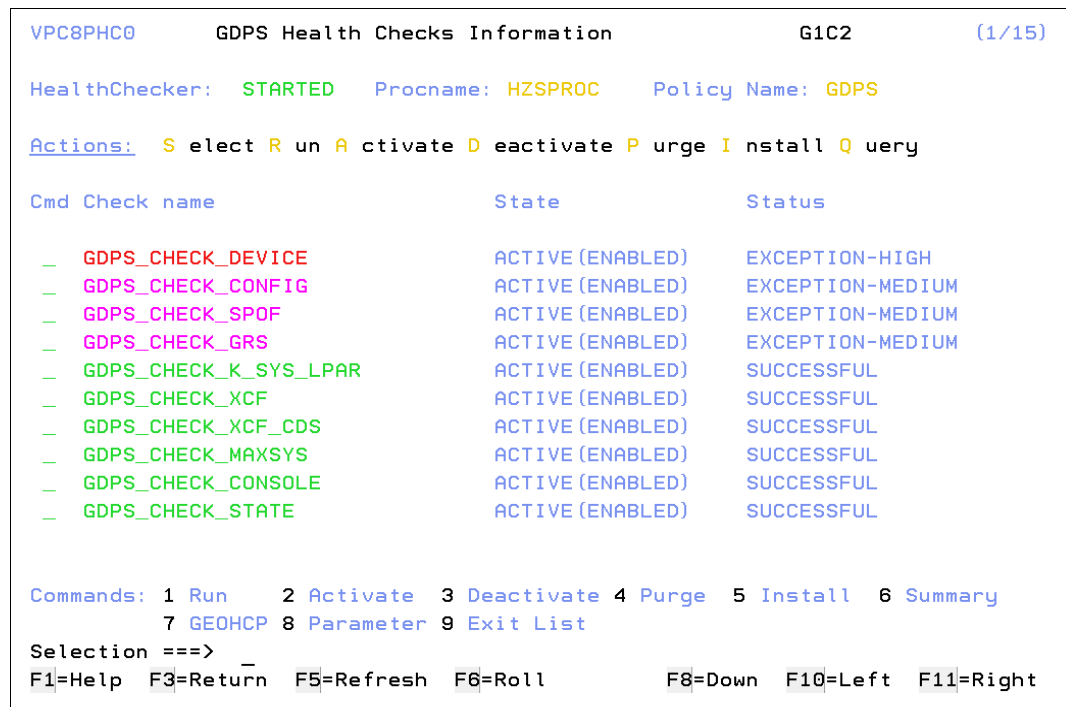


Figure 7-13 GDPS/MTMM Health Checks Information Management panel (VPC8PHC0)

7.6 Other facilities related to GDPS

Miscellaneous facilities that GDPS/MTMM provides can assist in various ways, such as reducing the window during which disaster recovery capability is not available.

7.6.1 HyperSwap and TDMF coexistence

To minimize disruption to production workloads and service levels, many enterprises use IBM’s Transparent Data Migration Facility (TDMF) for storage subsystem migrations and other disk relocation activities. The migration process is transparent to the application, and the data is continuously available for read and write activities throughout the migration process.

However, the HyperSwap function is mutually exclusive with software that moves volumes around by switching UCB pointers. The currently supported versions of TDMF and GDPS allow operational coexistence. With this support, TDMF automatically temporarily disables HyperSwap as part of the disk migration process only during the brief time when it switches UCB pointers. Manual operator interaction is not required. Without this support, through operator intervention, HyperSwap is disabled for the entire disk migration, including the lengthy data copy phase.

7.6.2 Reduced impact initial copy and resynchronization

Performing PPRC copy of a large amount of data across a large number of devices while the same devices are used in production by application workloads can potentially affect production I/O service times if such copy operations are performed synchronously. Your disk subsystems and PPRC link capacity are typically sized for steady state update activity, but not for bulk, synchronous replication. Initial copy of disks and resynchronization of disks are examples of bulk copy operations that can affect production if performed synchronously.

There is no need to perform initial copy or resynchronizations using synchronous copy, because the secondary disks cannot be made consistent until all disks in the configuration have reached duplex state.

GDPS supports initial copy and resynchronization using asynchronous PPRC-XD (also known as Global Copy). When GDPS initiates copy operations in asynchronous copy mode, GDPS monitors progress of the copy operation and when the volumes are near full duplex state, GDPS converts the replication from the asynchronous copy mode to synchronous PPRC. Initial copy or resynchronization using PPRC-XD eliminates the performance impact of synchronous mirroring on production workloads.

Without asynchronous copy it might be necessary to defer these operations or reduce the number of volumes being copied at any given time. This would delay the mirror from reaching a duplex state, thus impacting a client's ability to recovery. Use of the XD-mode asynchronous copy allows clients to establish or resynchronize mirroring during periods of high production workload, and can potentially reduce the time during which the configuration is exposed.

7.6.3 Concurrent Copy cleanup

The DFSMS Concurrent Copy (CC) function uses a "sidefile" that is kept in the disk subsystem cache to maintain a copy of changed tracks that have not yet been copied. For a PPRCed disk, this sidefile is not mirrored to the secondary subsystem. If you perform a HyperSwap while a Concurrent Copy operation is in progress, this will result in the job performing the copy failing after the swap. GDPS will not allow a planned swap when a Concurrent Copy session exists against your primary PPRC devices. However, unplanned swaps will still be allowed.

If you plan to use HyperSwap for primary disk subsystem failures (unplanned HyperSwap), try to eliminate any use of Concurrent Copy because you cannot plan when a failure will occur. If you choose to run Concurrent Copy operations while enabled for unplanned HyperSwaps, and a swap occurs when a Concurrent Copy operation is in progress, the job performing the Concurrent Copy operation is expected to fail.

Checking for CC is performed by GDPS immediately before performing a planned HyperSwap. SDF trace entries are generated if one or more CC sessions exist, and the swap command will end with no PPRC device pairs being swapped. You must identify and terminate any CC and XRC sessions against the PPRC primary devices before the swap.

When attempting to resynchronize your disks, checking is performed to ensure that the secondary devices do not retain CC status from the time when they were primary devices. These are not supported as PPRC secondary devices. Therefore, GDPS will not attempt to establish a duplex pair with secondary devices if it detects a CC session.

GDPS provides a function to discover and terminate Concurrent Copy sessions that would otherwise cause errors during a resync operation. The function is controlled by a keyword that provides options to disable, to conditionally enable, or to unconditionally enable the cleanup of Concurrent Copy sessions on the target disks. This capability eliminates the manual task of identifying and cleaning up orphaned Concurrent Copy sessions before resynchronizing a suspended PPRC mirror.

7.6.4 Easy Tier Heat Map Transfer

IBM DS8000 Easy Tier optimizes data placement (placement of logical volumes) across the various physical tiers of storage within a disk subsystem to optimize application performance. The placement decisions are based on learning the data access patterns, and can be changed dynamically and transparently using this data.

PPRC mirrors the data from the primary to the secondary disk subsystem. However, the Easy Tier learning information is not included in PPRC scope. The secondary disk subsystems are optimized according to the workload on these subsystems, which is different than the activity on the primary (there is only write workload on the secondary whereas there is read/write activity on the primary). As a result of this difference, during a disk switch or disk recovery, the secondary disks that you switch to are likely to display different performance characteristics compared to the former primary.

Easy Tier Heat Map Transfer is the DS8000 capability to transfer the Easy Tier learning from a PPRC primary to the secondary disk subsystems so that the secondary disk subsystems can also be optimized, based on this learning, and will have similar performance characteristics if it is promoted to become the primary.

GDPS integrates support for Heat Map Transfer. In a Multi-Target PPRC environment, Heat Map Transfer is established for both secondary targets. The appropriate Heat Map Transfer actions (such as start/stop of the processing and reversing transfer direction) are incorporated into the GDPS managed processes. For example, if PPRC is temporarily suspended on one leg by GDPS for a planned or unplanned secondary disk outage, Heat Map Transfer is also suspended on that leg, or if PPRC direction is reversed as a result of a HyperSwap, Heat Map Transfer direction is also reversed.

7.7 GDPS/MTMM flexible testing and resync protection

Configuring point-in-time copy (FlashCopy) capacity in your MTMM environment provides two significant benefits:

- ▶ You can conduct regular DR drills or other tests using a copy of production data while production continues to run.
- ▶ You can save a consistent, “golden” copy of the PPRC secondary data, which can be used if the primary disk or site is lost during a PPRC resynchronization operation.

FlashCopy and the various options related to FlashCopy are discussed in 2.6, “FlashCopy” on page 38. GDPS/MTMM supports taking a FlashCopy of the current primary or either of the current secondary disks sets. The COPY, NOCOPY, NOCOPY2COPY, and INCREMENTAL options are supported. CONSISTENT FlashCopy is supported in conjunction with COPY, NOCOPY, and INCREMENTAL FlashCopy.

FlashCopy can also be used, for example, to back up data without the need for extended outages to production systems; to provide data for data mining applications; for batch reporting, and so on.

7.7.1 Use of space-efficient FlashCopy volumes

As discussed in “Space-efficient FlashCopy (FlashCopy SE)” on page 40, by using space-efficient (SE) volumes, you might be able to lower the amount of physical storage needed and thereby reduce the cost associated with providing a tertiary copy of the data. GDPS provides support allowing space-efficient FlashCopy volumes to be used as FlashCopy target disk volumes. Whether a target device is space-efficient or not is transparent to GDPS; if any of the FlashCopy target devices defined to GDPS are space-efficient volumes, GDPS will simply use them. All GDPS FlashCopy operations with the NOCOPY option, whether through GDPS scripts, panels, or FlashCopies automatically taken by GDPS, can use space-efficient targets.

Space-efficient volumes are ideally suited for FlashCopy targets when used for resync protection. The FlashCopy is taken before the resync and can be withdrawn as soon as the resync operation is complete. As changed tracks are sent to the secondary for resync, the time zero (T0) copy of this data is moved from the secondary to the FlashCopy target device. This means that the total space requirement for the targets is equal to the number of tracks that were out of sync, which typically will be significantly less than a full set of fully provisioned disks.

Another potential use of space-efficient volumes is if you want to use the data for limited disaster recovery testing.

Understanding the characteristics of FlashCopy SE is important to determine whether this method of creating a point-in-time copy will satisfy your business requirements. For example, will it be acceptable to your business if, because of an unexpected workload condition, the repository on the disk subsystem for the space-efficient devices becomes full and your FlashCopy is invalidated so that you are unable to use it? If your business requirements dictate that the copy must always be guaranteed to be usable, space-efficient might not be the best option and you can consider using standard FlashCopy instead.

7.8 GDPS tools for GDPS/MTMM

GDPS/MTMM includes tools that provide function that is complementary to GDPS function. The tools represent the kind of function that all or many clients are likely to develop themselves to complement GDPS. Using these tools eliminates the need for you to develop similar function yourself. The tools are provided in source code format which means that if the tool does not exactly meet your requirements, you can modify the code to suit your needs.

The following tools are available with GDPS/MTMM:

- **GDPS XML Conversion (GeoXML) Tool**

This tool helps you to convert an existing GDPS/PPRC (or GDPS/PPRC HyperSwap Manager - GDPS/HM) GEOPARM configuration definition file for a single replication leg to GDPS/MTMM XML format GEOPARM definitions. This simplifies the task of defining the MTMM configuration for existing GDPS/PPRC (or GDPS/HM) clients that will be moving to using GDPS/MTMM.

- **GDPS EasyLog Tool**

This is a Microsoft Windows-based tool to help you extract and easily download the MVS Syslog and NetView log from a z/OS environment. It also helps in analyzing the Netlog after it is downloaded to a workstation.

7.9 Services component

As you have learned, GDPS affect much more than simply remote copy. It also includes system, server hardware and sysplex management, automation, testing processes, disaster recovery processes, and so on.

Most installations do not have skills in all these areas readily available. It is also extremely rare to find a team that has this range of skills across many implementations. However, the GDPS/MTMM offering includes exactly that: access to a global team of specialists in all the disciplines you need to ensure a successful GDPS/MTMM implementation.

Specifically, the Services component includes several or all of the following services:

- ▶ Planning to determine availability requirements, configuration recommendations, and implementation and testing plans
- ▶ Installation and necessary customization of NetView and System Automation
- ▶ Remote copy implementation
- ▶ GDPS/MTMM automation code installation and policy customization
- ▶ Assistance in defining Recovery Point and Recovery Time objectives
- ▶ Education and training on GDPS/MTMM setup and operations
- ▶ Onsite implementation assistance
- ▶ Project management and support throughout the engagement

The sizing of the Services component of each project is tailored for that project, based on many factors including what automation is already in place, whether remote copy is already in place, whether the two centers are already in place with a multisite sysplex, and so on. This means that the skills provided are tailored to the specific needs of each particular implementation.

7.10 GDPS/MTMM prerequisites

See the following web page for the latest GDPS/MTMM prerequisite information:

<http://www.ibm.com/systems/z/advantages/gdps/getstarted/gdpspprc.html>

7.11 Comparison of GDPS/MTMM versus other GDPS offerings

So many features and functions are available in the various members of the GDPS family that recalling them all and remembering which offerings support them is sometimes difficult. To position the offerings, Table 7-1 lists the key features and functions and indicates which ones are delivered by the various GDPS offerings.

Table 7-1 Supported features matrix

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/MTMM	GDPS Virtual Appliance	GDPS/XRC	GDPS/GM
Continuous availability	Yes	Yes	Yes	Yes	No	No
Disaster recovery	Yes	Yes	Yes	Yes	Yes	Yes

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/MTMM	GDPS Virtual Appliance	GDPS/XRC	GDPS/GM
CA/DR protection against multiple failures	No	No	Yes	No	No	No
Continuous Availability for foreign z/OS systems	Yes with z/OS Proxy	No	No	No	No	No
Supported distance	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	Virtually unlimited	Virtually unlimited
Zero Suspend FlashCopy support	Yes, using CONSISTENT	Yes, using CONSISTENT for secondary only	Yes, using CONSISTENT	No	Yes, using Zero Suspend FlashCopy	Yes, using CGPause
Reduced impact initial copy/resync	Yes	Yes	Yes	Yes	Not applicable	Not applicable
Tape replication support	Yes	No	No	No	No	No
Production sysplex automation	Yes	No	Yes	Not applicable	No	No
Span of control	Both sites	Both sites (disk only)	Both sites	Both sites	Recovery site	Disk at both sites; recovery site (CBU or LPARs)
GDPS scripting	Yes	No	Yes	Yes	Yes	Yes
Monitoring, alerting and health checks	Yes	Yes	Yes	Yes (except health checks)	Yes	Yes
Query Services	Yes	Yes	No	No	Yes	Yes
MSS support for added scalability	Yes (secondary in MSS1)	Yes (secondary in MSS1)	Yes (H2 in MSS1, H3 in MSS2)	No	No	Yes (GM FC and Primary for MGM in MSS1)
MGM 3-site and 4-site	Yes (all configurations)	Yes (3-site only and non-IR only)	Yes (all configurations)	No	Not applicable	Yes (all configurations)
MzGM	Yes	Yes	Yes (non-IR only)	No	Yes	Not applicable
Open LUN	Yes	Yes	No	No	No	Yes
z/OS equivalent function for Linux for IBM z Systems	Yes	No	Yes (Linux for IBM z Systems running as a z/VM guest only)	Yes (Linux for IBM z Systems running as a z/VM guest only)	Yes	Yes
Heterogeneous support through DCM	Yes (VCS and SA AppMan)	No	No	No	Yes (VCS only)	Yes (VCS and SA AppMan)

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/MTMM	GDPS Virtual Appliance	GDPS/XRC	GDPS/GM
z/BX hardware management	Yes	No	No	No	No	No
GDPS GUI	Yes	Yes	No	Yes	No	Yes

7.12 Summary

GDPS/MTMM is a powerful offering that provides disaster recovery, continuous availability, and system/sysplex resource management capabilities. HyperSwap, available with GDPS/MTMM, provides the ability to transparently swap disks between disk locations. The power of automation allows you to test and perfect the actions to be taken, either for planned or unplanned changes, thus minimizing or eliminating the risk of human error.

This offering is one of the offerings in the GDPS family, along with GDPS/PPRC, GDPS/HM, and GDPS Virtual Appliance, that offers the potential of zero data loss, and that can achieve the shortest recovery time objective, typically less than one hour following a complete site failure.

It is also one of the only members of the GDPS family, along with GDPS/PPRC and GDPS Virtual Appliance, that is based on hardware replication and that provides the capability to manage the production LPARs. Although GDPS/XRC and GDPS/GM offer LPAR management, their scope for system management only includes the systems in the recovery site, and not the production systems running in Site1.

GDPS/MTMM is the only GDPS offering that can provide zero-data-loss disaster recovery protection, even *after* a primary disk failure.

In addition to the disaster recovery and planned reconfiguration capabilities, GDPS/MTMM also provides a user-friendly interface for monitoring and managing the various elements of the GDPS configuration.



GDPS/Active-Active solution

In this chapter, we introduce the GDPS/Active-Active solution. This solution aims to significantly reduce the time spent to recover systems in a disaster recovery situation, and enable planned and unplanned switching of workloads between sites. The chapter includes sections that discuss the following aspects of GDPS/Active-Active:

- ▶ Concepts
- ▶ Products
- ▶ Environment
- ▶ Functions and features
- ▶ Testing
- ▶ Services

8.1 Overview of GDPS/Active-Active

In this section, we provide a high-level description of the GDPS/Active-Active solution and explain where it fits in with the other GDPS products.

8.1.1 Positioning GDPS/Active-Active

The key metrics in business continuity are as follows:

- ▶ Recovery time objective (RTO): How long can you afford to be without your systems?
- ▶ Recovery point objective (RPO): How much data can you afford to lose or re-create?
- ▶ Network recovery objective (NRO): How long does it take to switch over the network?

There are multiple offerings in the GDPS family, all of which are covered in this book. The GDPS products other than GDPS/Active-Active are continuous availability (CA) and disaster recovery (DR) solutions that are based on synchronous or asynchronous disk hardware replication.

To achieve the highest levels of availability and minimize the recovery for planned and unplanned outages, various clients have deployed GDPS/PPRC *Active/Active* configurations, which have the following requirements:

- ▶ All critical data must be PPRCed and HyperSwap enabled.
- ▶ All critical CF structures must be duplexed.
- ▶ Applications must be Parallel Sysplex enabled.

However, the signal latency between sites will potentially affect online workload throughput and batch duration. This results in sites typically being separated by no more than approximately 20 km fiber distance¹.

Consequently, the GDPS/PPRC *Active/Active* configuration, which can provide an RPO of zero (0) and an RTO as low as a few minutes, does not provide a solution if an enterprise requires that the distance between the active sites is much greater than 20 to 30 km.

The GDPS products based on asynchronous hardware replication, GDPS/XRC and GDPS/GM, provide for virtually unlimited site separation. However, they do require that the workload from the failed site be restarted in the recovery site and this typically will take 30 to 60 minutes. Thus, GDPS/XRC and GDPS/GM are not able to achieve the RTO of seconds required by various enterprises for their most critical workloads.

In summary, when using the GDPS products based on hardware replication, it is not possible to achieve aggressive RPO and RTO goals while providing the sufficient site separation that is being required by some enterprises.

For these reasons, the Active/Active sites concept was conceived.

¹ The distance between sites in a GDPS/PPRC *Active/Active* configuration that any client can tolerate will depend on the client's application workloads and service level requirements. Each client must test with its own applications and workloads to determine the distance it can achieve. Nearly all clients running GDPS/PPRC *Active/Active* workloads are running their two sites at a 20 km distance or less. However, this does not necessarily mean that somewhat larger distances are not possible.

8.1.2 GDPS/Active-Active sites concept

The Active/Active sites concept consists of having two sites, separated by virtually unlimited distances, running the same applications and having the same data to provide cross-site workload balancing and continuous availability and disaster recovery. This is a fundamental paradigm shift from a *failover* model to a *continuous availability* model.

GDPS/Active-Active (GDPS/A-A) does not use any of the infrastructure-based data replication techniques that other GDPS products rely on, such as Metro Mirror (PPRC), Global Mirror (GM), or z/OS Global Mirror (XRC).

Instead, GDPS/Active-Active relies on both of the following methods:

- ▶ Software-based asynchronous replication techniques for copying the data between sites
- ▶ Automation, primarily operating at a workload level, to manage the availability of selected workloads and the routing of transactions for these workloads

The GDPS/Active-Active product, which is a component of the GDPS/Active-Active solution, acts primarily as the coordination point or controller for these activities. It is a focal point for operating and monitoring the solution and readiness for recovery.

Note: For simplicity, in this chapter we refer to both the solution and the product as GDPS/Active-Active. We might also refer to the environment managed by the solution, and the solution itself, as Active-Active.

What is a workload

A *workload* is defined as the aggregation of the following components:

- ▶ Software
User-written applications such as COBOL programs, and the middleware runtime environment (for example, CICS regions, InfoSphere Replication Server instances and DB2 subsystems)
- ▶ Data
A related set of objects that must preserve transactional consistency and optionally referential integrity constraints (for example, DB2 tables and IMS databases)
- ▶ Network connectivity
One or more TCP/IP addresses and ports (for example, 10.10.10.1:80)

Two workload types are supported and managed in a GDPS/Active-Active environment:

- ▶ Update or read/write workloads
These run in what is known as the *Active/Standby configuration*. In this case, a workload managed by GDPS/Active-Active will be active in one sysplex and receiving transactions routed to it by the workload distribution mechanism that is managed by the IBM Multi-site Workload Lifeline. The workload will also be using software replication to copy changed data to another instance of the workload running in a second sysplex, where all the infrastructure components (LPARs, systems, middleware and so on) and even the application are ready to receive work in what is termed a standby mode. The updated data from the active instance of the workload is applied in real time to the database subsystem instance running in standby mode.

► Query or read-only workloads

These workloads are associated with update workloads, but they can be actively running in both sites at the same time. Workload distribution between the sites is based on policy options, and takes into account environmental factors such as the latency for replication that determines the age (or currency) of the data in the standby site. There is no data replication associated with the query workload as there are no updates to the data. You can associate up to two query workloads with a single update workload.

Figure 8-1 shows these concepts for an update workload at a high level (not all redundant components are shown in detail). Transactions arrive at the workload distributor, also known as the *load balancer*. Depending on the current situation, the transactions are routed to what is termed the currently active sysplex in the configuration for that particular workload.

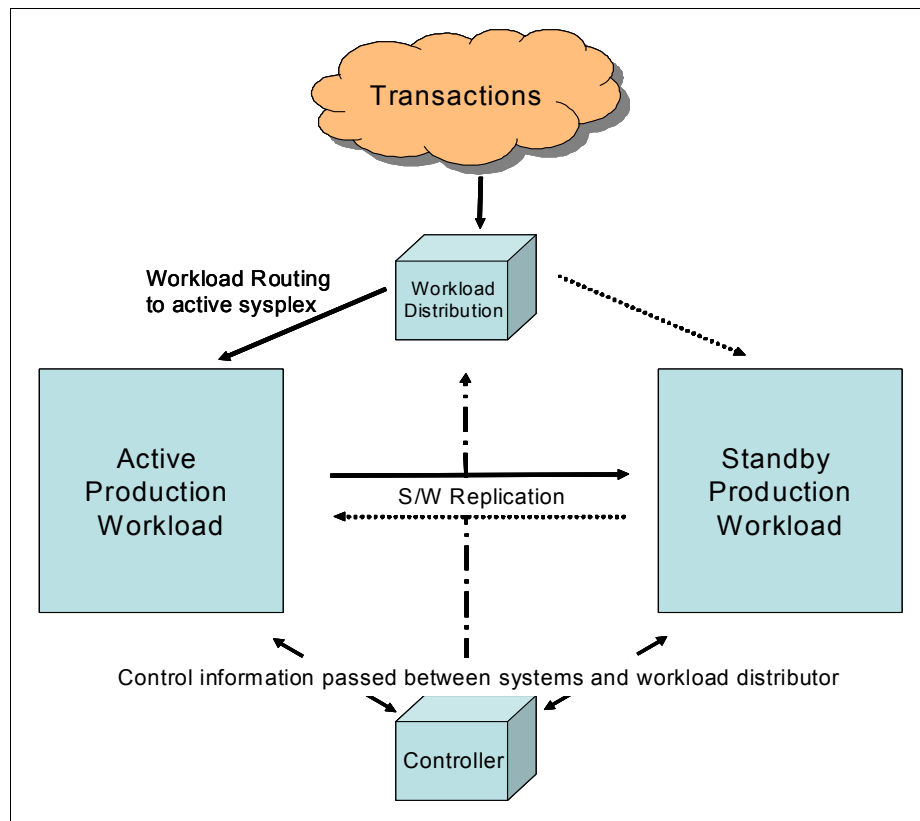


Figure 8-1 GDPS/Active-Active concept

The environment is constantly being monitored to ensure that workload is being processed in the active sysplex. If GDPS/Active-Active detects that workload is not processing normally, then a policy-based decision is made to either automatically start routing work to the standby sysplex (rather than the currently active sysplex), or to prompt the operator to take some action. In a similar way, for query workloads, a policy using the latency of replication as thresholds that will trigger GDPS/Active-Active or other products in the solution to take some action.

Information is constantly being exchanged by the systems in the active and standby sysplexes, the GDPS controllers (one in each location), and the workload distribution mechanism to ensure that an accurate picture of the health of the environment is maintained to enable appropriate decisions from the automation.

It is also possible, in a planned manner, to switch each workload from the currently active to the standby sysplex if the need arises, such as for routine maintenance and so on.

Note: In this chapter we sometimes refer to workloads managed by GDPS/Active-Active as *Active-Active workloads*.

In your environment you are likely to have some applications and data which you do not want to manage with or simply cannot be managed by GDPS/Active-Active. For example, you may have an application that uses a data type for which software data replication is not available or is not supported by GDPS/Active-Active. You will still need to provide high availability and disaster recovery for such applications and data. For this, GDPS/Active-Active provides for integration and co-operation with other GDPS products which rely on hardware replication and are independent of application and data type. Specifically, special coordination is provided with GDPS/PPRC which we describe in 8.5, “GDPS/Active-Active co-operation with GDPS/PPRC or GDPS/MTMM” on page 264 and GDPS/MGM which we describe in 8.6, “GDPS/Active-Active disk replication integration” on page 267.

8.2 GDPS/Active-Active solution products

The GDPS/Active-Active architecture, shown at a conceptual level in Figure 8-2, consists of several products coordinating the monitoring and managing of the various aspects of the environment.

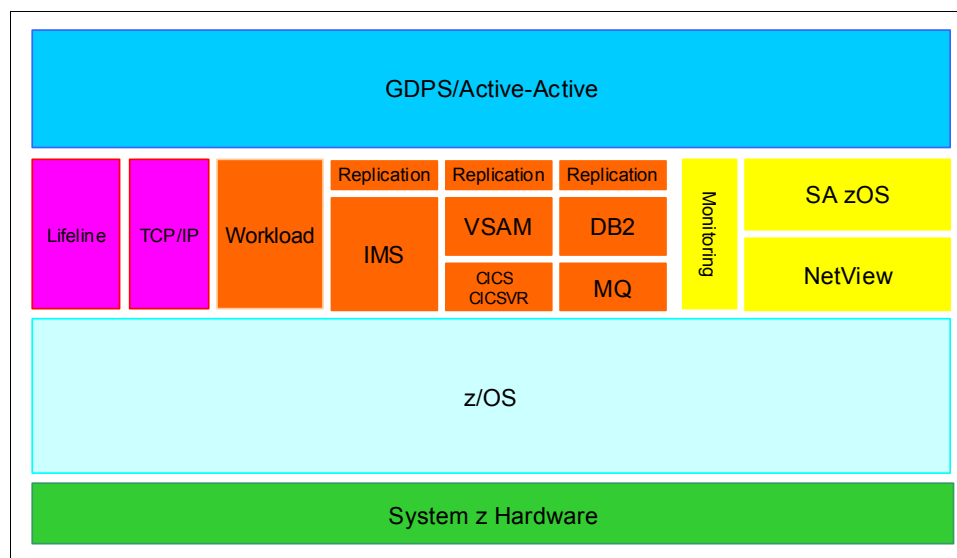


Figure 8-2 GDPS/Active-Active architecture

This section describes, at a high level, the various products required for GDPS/Active-Active and their role or function within the overall framework. The following products are briefly discussed:

- ▶ GDPS/Active-Active
- ▶ IBM Tivoli NetView for z/OS
 - IBM Tivoli NetView for z/OS Enterprise Management Agent (NetView agent)
- ▶ IBM Tivoli NetView Monitoring for GDPS
- ▶ IBM Tivoli Monitoring

- ▶ IBM Tivoli System Automation for z/OS
- ▶ IBM Multi-site Workload Lifeline for z/OS
- ▶ Middleware such as CICS, IMS, DB2, C/PSM to run the workloads
- ▶ Replication Software
 - IBM InfoSphere Data Replication for DB2 for z/OS
WebSphere MQ is required for DB2 data replication
 - IBM InfoSphere Data Replication for VSAM for z/OS
CICS Transaction Server, CICS VSAM Recovery, or both, are required for VSAM replication
 - IBM InfoSphere IMS Replication for z/OS
- ▶ Other optional components
 - IBM Tivoli OMEGAMON® XE family of monitoring products for monitoring the various parts of the solution

In 8.3, “GDPS/Active-Active environment” on page 240, we provide a solution view that illustrates how the products are used in the various systems in which they run.

8.2.1 The GDPS/Active-Active product

The GDPS/Active-Active product provides automation code that is an extension of many of the techniques that have been tried and tested in other GDPS products and with many client environments around the world for management of their mainframe continuous availability and disaster recovery requirements. The GDPS/Active-Active control code runs only on the Controller systems in the Active-Active environment.

The key functions provided by GDPS/Active-Active code are as follows:

- ▶ Workload management, such as starting or stopping all components of a workload in a given sysplex.
- ▶ Replication management, such as starting or stopping replication for a given workload from one sysplex to the other.
- ▶ Routing management, such as stopping or starting routing of transactions to one sysplex or the other for a given workload.
- ▶ System and Server management, such as STOP (graceful shutdown) of a system, LOAD, RESET, ACTIVATE, DEACTIVATE the LPAR for a system and capacity on-demand actions such as CBU/OOCOD activation.
- ▶ Monitoring the environment and alerting for unexpected situations.
- ▶ Planned/Unplanned situation management and control, such as planned or unplanned site or workload switches.
 - Autonomic actions, such as automatic workload switch (policy-dependent).
- ▶ Powerful scripting capability for complex/compound scenario automation.
- ▶ Co-operation with GDPS/PPRC to provide continuous data availability in the Active-Active sysplexes.
- ▶ Single point of control for managing disk replication functions when running GDPS/MGM together with GDPS/Active-Active to protect non Active-Active data.
- ▶ Easy-to-use graphical user interface.

8.2.2 Tivoli NetView for z/OS

The NetView product is a prerequisite for GDPS/Active-Active automation and management code. In addition to being the operating environment for GDPS, the NetView product provides additional monitoring and automation functions associated with the GDPS/Active-Active solution.

Monitoring capability using the NetView agent is provided for the following items:

- ▶ IBM Multi-site Workload Lifeline for z/OS
- ▶ IBM InfoSphere Data Replication for DB2 for z/OS
- ▶ IBM InfoSphere Data Replication for VSAM for z/OS
- ▶ IBM InfoSphere Data Replication for IMS for z/OS

NetView Agent

The Tivoli NetView for z/OS Enterprise Management Agent (also known as TEMA) is used in the solution to pass information from the z/OS NetView environment to the Tivoli Enterprise Portal, which is used to provide a view of your enterprise from which you can drill down to more closely examine components of each system being monitored. The NetView agent requires IBM Tivoli Monitoring.

8.2.3 IBM Tivoli Monitoring

IBM Tivoli Monitoring is a suite of monitoring components to monitor and report on various aspects of a client's IT environment. Several of the IBM Tivoli Monitoring components are used in the overall monitoring of aspects (such as monitoring the workload) within the GDPS/Active-Active environment.

The specific components required for GDPS/Active-Active are listed here.

Tivoli Enterprise Portal

Tivoli Enterprise Portal (portal client or portal) is a Java-based interface for viewing and monitoring your enterprise. Tivoli Enterprise Portal offers two modes of operation: desktop and browser.

Tivoli Enterprise Portal Server

Tivoli Enterprise Portal Server (portal server) provides the core presentation layer for retrieval, manipulation, analysis, and preformatting of data. The portal server retrieves data from the hub monitoring server in response to user actions at the portal client, and sends the data back to the portal client for presentation. The portal server also provides presentation information to the portal client so that it can render the user interface views suitably.

Tivoli Enterprise Monitoring Server

The Tivoli Enterprise Monitoring Server (monitoring server) is the collection and control point for performance and availability data and alerts received from monitoring agents (for example, the NetView agent). It is also responsible for tracking the online or offline status of monitoring agents.

The portal server communicates with the monitoring server, which in turn controls the remote servers and any monitoring agents that might be connected to it directly.

8.2.4 System Automation for z/OS

IBM Tivoli System Automation for z/OS is a cornerstone of all members of the GDPS family of products. In GDPS/Active-Active it provides the critical policy repository function, in addition to managing the automation of the workload and systems elements. System Automation for z/OS also provides the capability for GDPS to manage and monitor systems in multiple sysplexes.

System Automation for z/OS is required on the Controllers and all production systems running Active-Active workloads. If you use an automation product other than System Automation for z/OS to manage your applications, you do not need to replace your entire automation with System Automation. Your existing automation can coexist with System Automation and an interface is provided to ensure proper coordination takes place.

8.2.5 IBM Multi-site Workload Lifeline for z/OS

This product provides intelligent routing recommendations to external load balancers for server instances that can span two sysplexes/sites. The IBM Multi-site Workload Lifeline for z/OS product consists of Advisors and Agents.

There is one Lifeline Advisor that is active in the same z/OS image as the GDPS Primary Controller and assumes the role of primary Advisor, and at most one other Lifeline Advisor that is active on the Backup Controller and assumes the role of secondary Advisor.

The two Advisors exchange state information so that the secondary Advisor can take over the primary Advisor role if that the current primary Advisor is terminated or there is a failure on the system where the primary Advisor was active.

In addition, there is a Lifeline Agent that is active on all z/OS images where workloads can run. All Lifeline Agents monitor the health of the images they are running on, and the health of the workload. These Agents communicate this information back to the primary Lifeline Advisor, which then calculates routing recommendations.

Finally, external load balancers establish a connection with the primary Lifeline Advisor and receive routing recommendations through the open-standard Server/Application State Protocol (SASP) API that is documented in RFC 4678.

The Lifeline Advisor also establishes a Network Management Interface (NMI) to allow network management applications (such as NetView) to retrieve internal data that the Advisor uses to calculate routing recommendations.

The Lifeline Advisors and Agents use configuration information stored in text files to determine what workloads need to be monitored and how to connect to each other and external load balancers.

GDPS/Active-Active provides high-level control capabilities to start and stop replication for a given workload to one sysplex/site or the other using either GDPS automation scripts or panel actions in the GDPS GUI.

8.2.6 Middleware

Middleware components such as CICS regions or DB2 subsystems form a fundamental part of the Active/Active environment because they provide the application services required to process the workload. To maximize the availability characteristics of the GDPS/Active-Active environment, applications and middleware need to be replicated across multiple images in the active and standby Parallel Sysplexes to cater for local high availability in case of component failure. Automation needs to be in place to ensure clean start, shutdown, and local recovery of these critical components. CICS/DB2 workloads managed by CPSM derive additional benefits in a GDPS/Active-Active environment.

8.2.7 Replication software

Unlike in other GDPS solutions where the replication is based on mirroring the disk-based data at the block level using hardware (such as Metro Mirror or Global Mirror) or a combination of hardware and software (z/OS Global Mirror, also known as XRC) as previously mentioned, replication in GDPS/Active-Active is managed by software only. The following products are supported in GDPS/Active-Active:

- ▶ IBM InfoSphere Data Replication for DB2 for z/OS

This product, also known widely as Q-rep, uses underlying IBM WebSphere MQ as the transport infrastructure for moving the DB2 data from the source to the target copy of the database. Transaction data is *captured* at the source site and placed in IBM MQ queues for transmission to a destination queue at the target location, where the updates are then *applied* in real time to a running copy of the database.

For very large scale and update-intensive DB2 replication environments, a single pair of capture/apply engines may not be able to keep up with the replication. Q-rep provides a facility known as Multiple Consistency Groups (MCG) where the replication work is spread across multiple capture/apply engines, yet the time order (consistency) for the workload across all of the capture/apply engines is preserved in the target database. GDPS supports and provides specific facilities for workloads using MCG with DB2 replication.

- ▶ IBM InfoSphere Data Replication for IMS for z/OS

IBM InfoSphere IMS Replication for z/OS is the product that provides IMS data replication and uses a similar capture and apply technique to that outlined for DB2 data. However, IMS Replication does not use MQ as the transport infrastructure to connect the source and target copies. Instead, TCP/IP is used in place of MQ through the specification of host name and port number to identify the target to the source and similarly to define the source to the target.

- ▶ IBM InfoSphere Data Replication for VSAM for z/OS

IBM InfoSphere Data Replication for VSAM for z/OS is similar in structure to the IMS replication product except that it is for replicating VSAM data. For CICS VSAM data, the sources for capture are CICS log streams. For non-CICS VSAM data, CICS VSAM Recovery (CICS VR) is required for logging, and will be the source for replicating such data. Similar to IMS replication, TCP/IP is used as the transport for VSAM replication.

GDPS/Active-Active provides high-level control capabilities to start and stop replication between identified source and target instances through both scripts and panel actions in the GDPS GUI. GDPS also monitors replication latency and uses this information when deciding whether Query workloads can be routed to the standby site or not.

8.2.8 Other optional components

Other components can optionally be used to provide specific monitoring, as described here.

Tivoli OMEGAMON XE family

Additional products such as Tivoli OMEGAMON XE on z/OS, Tivoli OMEGAMON XE for DB2, and Tivoli OMEGAMON XE for IMS, can optionally be deployed to provide specific monitoring of products that are part of the Active/Active sites solution.

8.3 GDPS/Active-Active environment

In this section we provide a conceptual view of a GDPS/Active-Active environment, plugging in the products that run on the various systems in the environment. We then take a closer look at how GDPS/Active-Active works. Finally, we briefly discuss environments where Active-Active and other workloads coexist on the same sysplex.

Figure 8-3 shows the key components of a GDPS/Active-Active environment.

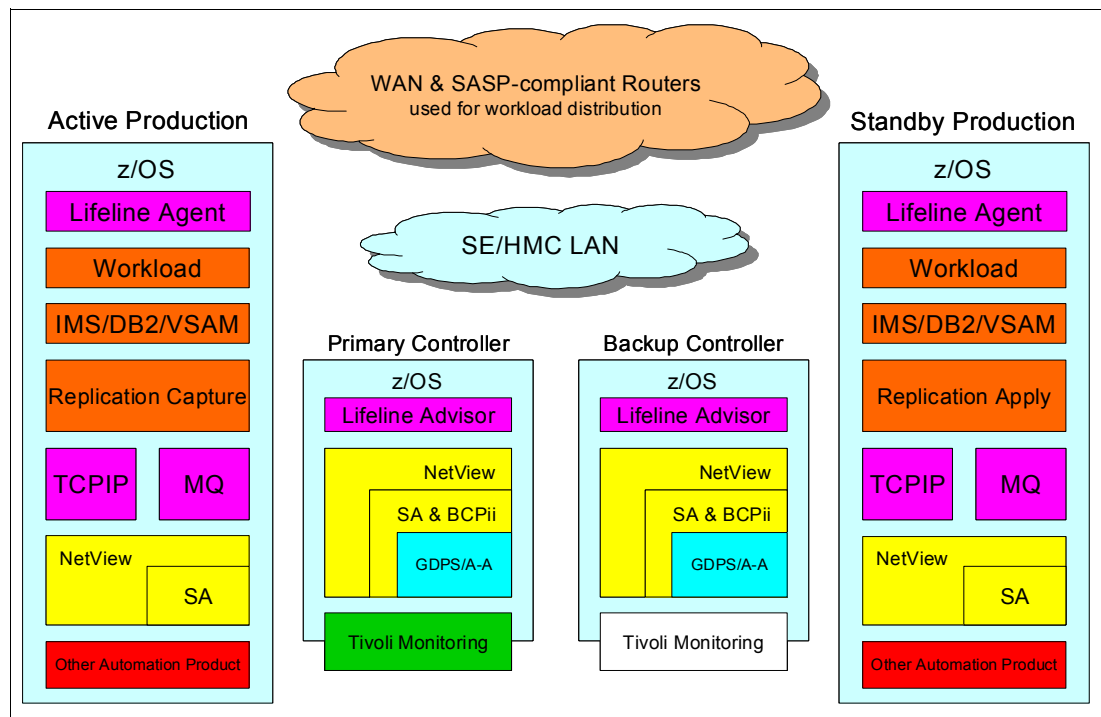


Figure 8-3 GDPS/Active-Active environment functional overview

The GDPS/Active-Active environment consists of two production sysplexes (also referred to as *sites*) in different locations. For each update workload that is to be managed by GDPS/Active-Active, at any given point in time, one of the sysplexes will be the *active* sysplex and the other will act as *standby*. In the figure we have one workload and only one active production system running this workload in one sysplex, and one production system that is standby for this workload. However, there can be multiple cloned instances of the active and the standby production systems in the two sysplexes.

When there are multiple workloads managed by GDPS, a given sysplex can be the active sysplex for one update workload, while it is standby for another. It is the routing for each update workload that determines which sysplex is active and which sysplex is standby for a given workload. As such, in environments where there are multiple workloads, there is no concept as an active sysplex. There is a sysplex that is the currently active one *for a given update workload*.

The production systems, both the active and the standby instances, are actively running the workload managed by GDPS. What makes a sysplex (and therefore the systems in that sysplex) active or standby is whether update transactions are currently being routed to that sysplex. The SASP routers in the network, which are shown in the figure as the cloud under GDPS and LifeLine Advisor, control routing of transactions for a given workload to one sysplex or the other. Although a single router is the minimum requirement, we expect that you will configure multiple routers for resiliency.

The workload is actively running on the z/OS system in both sysplexes. The workload on the system that is active for that workload is actually processing update transactions because update transactions are being routed to this sysplex.

The workload on the standby sysplex is actively running but is not processing any update transactions because update transactions are not being routed to it. It is waiting for work, and is able to process work at any time if there is a planned or unplanned workload switch resulting in transactions being routed to this sysplex. If there is a workload switch, the standby sysplex will become the active sysplex for the given workload.

The workload on the standby sysplex can be actively processing query transactions for the query workload that is associated with an update workload. Replication latency at any given point in time, in conjunction with thresholds you specify in the GDPS policy, determines whether query transactions are routed to the standby sysplex or not. The GDPS policy indicates when the latency or the replication lag is considered to be too high (that is, the data in the standby sysplex is considered to be too far behind) to the extent that query transactions should no longer be routed there, but should be routed to the active sysplex instead. When query transactions are no longer being routed to the standby sysplex because the latency threshold was exceeded, there is another threshold that you specify in the GDPS policy which indicates when it is OK to route query transactions to the standby sysplex once again.

For example, your policy might indicate that query transactions for a given workload should not be routed to the standby sysplex if latency exceeds 7 seconds and that it is OK to route to the standby sysplex after latency falls below 4 seconds. Latency is continually monitored to understand whether query transactions can be routed to the standby sysplex or not. In addition to the latency control, you can specify a policy to indicate what percentage of the incoming query transactions should be routed to the standby site or whether you simply want the conditions such as latency and workload health to dictate a dynamic decision on which of the two sysplexes query transactions would be routed to at any given point in time.

The workload itself is any subsystem receiving and processing update and or query transactions through the routing mechanism and using the replicated databases.

On the active system, you see a replication *capture engine*. One or more such engines can exist, depending on the data being replicated. This is the software replication component that captures all updates to the databases used by the workload managed by GDPS and forwards them to the standby sysplex.

On the standby sysplex, there is the counterpart of the capture engine, which is the *apply engine*. The apply engine receives the updates sent by the capture engine and immediately applies them to the database for the standby sysplex. The data replication in a GDPS environment is asynchronous. This means that the workload can perform a database update and this write operation can complete, independent of the replication process. Replication will require sufficient bandwidth for transmission of the data being replicated. IBM has services that can help you determine the bandwidth requirements based on your workload. If replication is disrupted for any reason, the replication engines, when restored, have logic to know where they left off and are able to transmit only those changes made after the disruption.

Because the replication is asynchronous, no performance impact is associated with replication. For a planned workload switch, the switch can take place after all updates are drained from the sending side and applied on the receiving side. For DB2 replication, GDPS provides additional automation to determine whether all updates have drained. This allows planned switch of workloads using DB2 replication to be completely automated.

For an unplanned switch, because replication is asynchronous, there will typically be some data captured but not yet transmitted and therefore not yet applied on the target sysplex. The amount of this data effectively translates to RPO. With a correctly-sized, robust transmission network, the RPO, during normal operations, is expected to be as low as just a few seconds. You might also hear the term *latency* used in conjunction with replication. Latency is simply another term that is used for the replication lag or RPO.

Although we talk about RPO, data is lost only if the original active site or the disks in this site where some updates were stranded are physically damaged so that they cannot be restored with the data intact. Following an unplanned switch to the standby site, if the former active site is restored with its data intact, any stranded updates can be replicated to the new active site at that time and no data will have been lost.

Additionally, a specialized implementation of GDPS/Active-Active, known as a Zero Data Loss or ZDL configuration, can be employed in some environments to provide an RPO of zero, even when the disks in the site that failed have been physically damaged, provided that the two GDPS/Active-Active sites are within supported Metro Mirror distances². For more information about the ZDL configuration, see 8.7, “Zero Data Loss Configuration” on page 268.

MQ is shown on production systems. MQ is required for DB2 replication. Either CICS or CICS VR is required on the production systems for VSAM replication.

On the production systems on both the active and standby sysplexes, you will also see the monitoring and management products. NetView, System Automation, and the LifeLine Agent run on all production systems, monitoring the system, the workload on the system and also replication latency and providing information to the active-active Controllers.

TCP/IP on the production systems is required in support of several functions related to GDPS/Active-Active.

Finally, on the production systems we show that you might have a product other than System Automation to manage your applications. In such an environment, as previously described, System Automation is still required for GDPS/Active-Active workload management. However, it is not necessary to replace your existing automation to use System Automation. A simple process for enabling the coexistence of System Automation and other automation products is available.

² The maximum supported Metro Mirror replication distance without RPQ is 300km.

Not shown in Figure 8-3 on page 240 is the possibility of running other workloads not managed by GDPS/Active-Active on the same production systems that run Active-Active workloads. We discuss other, non-Active-Active workloads in 8.3.2, “Considerations for other non-Active-Active workloads” on page 247.

Figure 8-3 on page 240 shows two GDPS Controller systems. At any point in time, one is the Primary Controller and the other is the Backup. These will typically be in each of the production sysplex locations, but there is no requirement that they are co-located in this way. GDPS/Active-Active introduces the term *Controller*, as opposed to the Controlling System term used within other GDPS solutions. The function of the Primary Controller is to provide a point of control for the systems and workloads participating in the GDPS/Active-Active environment for both planned actions (such as IPL and directing which is the active sysplex for a given workload) and for recovery from unplanned outages. The Primary Controller is also where the data collected by the monitoring aspects of the solution can be accessed.

Both Controllers run NetView, System Automation and GDPS/Active-Active control code, and the LifeLine Advisor. The Tivoli Monitoring components Tivoli Enterprise Monitoring Server and Tivoli Enterprise Management Agent run on the Controllers. Figure 8-3 on page 240 shows that there is a portion of Tivoli Monitoring not running on z/OS. The Tivoli Enterprise Portal Server component can run either on Linux on z Systems or on a distributed server.

Together with System Automation on the Controllers you see the BCP Internal Interface (BCPii). GDPS, on the Controller, uses this interface to perform hardware actions against the LPAR of production systems or the LPAR of the other Controller system such as LOAD, RESET, and so on, and for performing hardware actions for capacity on demand such as CBU or OOCOD activation.

Figure 8-3 on page 240 also shows the Support Element/Hardware Management Console (SE/HMC) local area network (LAN). This is a key element of the GDPS/Active-Active solution. The SE/HMC LAN spans the z Systems servers for both sysplexes in the two sites. This allows for a Controller in one site to act on hardware resources in the other site. To provide a LAN over large distance, the SE/HMC LANs in each site are bridged over the WAN.

It is desirable to isolate the SE/HMC LAN on a network other than the client’s WAN, which is the network used for the Active-Active application environment and connecting systems to each other. When isolated on a separate network, Lifeline Advisor (which is responsible for detecting failures and determining whether a sysplex has failed altogether or not) can try to access the site that appears to have failed over both the WAN and the SE/HMC LAN.

If the site is accessible through the SE/HMC LAN but not the WAN, then Lifeline can conclude that only the WAN failed, and not the target sysplex. Thus, isolating the SE/HMC LAN from the WAN provides an additional check when deciding whether the entire sysplex has failed and therefore, whether or not a workload switch is to be performed.

8.3.1 GDPS/Active-Active: A closer look

We described how GDPS/Active-Active works at a conceptual level, and how the various products that comprise the solution fit into the Active-Active framework.

In this section, we examine more closely how GDPS/Active-Active works, using an example of a GDPS/Active-Active environment with multiple workloads; see Figure 8-4. In this example, we consider only update workloads. Extending this example with query workloads corresponding to one or more of the update workloads can be a simple matter.

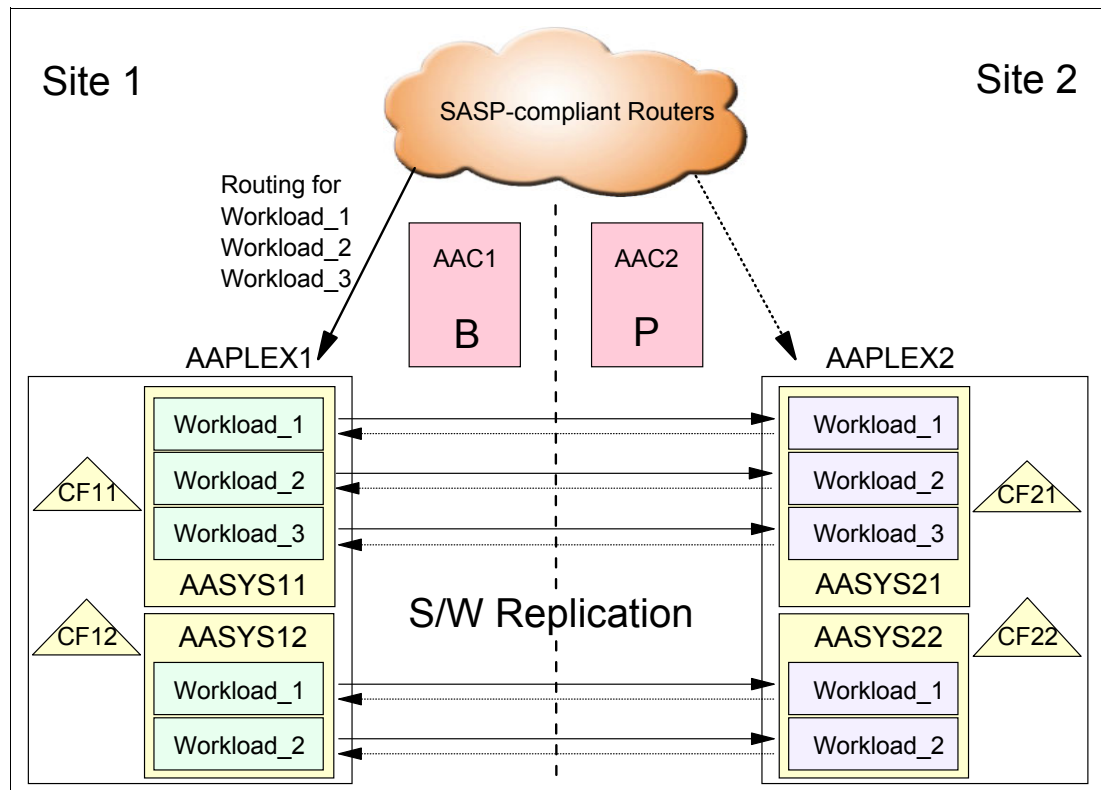


Figure 8-4 GDPS/Active-Active environment with multiple workloads: All active in one site

The figure shows two sites, Site1 and Site2, and a Parallel Sysplex in each site: AAPLEX1 runs in Site1 and AAPLEX2 runs in Site2. Coupling facilities CF11 and CF12 serve AAPLEX1 structures. CF21 and CF22 serve AAPLEX2 structures.

Each sysplex consists of two z/OS images. The z/OS images in AAPLEX1 are named AASYS11 and AASYS12. The images in AAPLEX2 are named AASYS21 and AASYS22. There are also two GDPS Controller systems: AAC1 is in Site1, and AAC2 is in Site2.

Three workloads are managed by GDPS in this environment: Workload_1, Workload_2, and Workload_3. As you can see, Workload_1 and Workload_2 are cloned, Parallel Sysplex-enabled applications that run on both z/OS images of the sysplexes. Workload_3 runs only in a single image in the two sysplexes.

At this time, the transactions for all three workloads are being routed to AAPLEX1. The workloads are running in AAPLEX2, but they are not processing transactions because no transactions are being routed to AAPLEX2.

AAPLEX1 is the source for data replication for all three workloads, and AAPLEX2 is the target. Also shown are reverse replication links from AAPLEX2 towards AAPLEX1. This indicates that if the workload is switched, the direction of replication can and will be switched.

If AASYS12 incurs an unplanned z/OS outage, then all three workloads would continue to run in AASYS11. It is possible, depending on the sizing of the systems, that AASYS11 does not have sufficient capacity to run the entire workload. Also, AASYS11 is now a single point of failure for all three workloads. In such a case, where no workload has failed but there is a possible degradation of performance and availability levels, you need to decide whether you want to continue running all three workloads in AASYS11 until AASYS12 can be restarted or whether you switch one or more (or possibly all three) workloads to run in AAPLEX2 systems. These are decisions you will prepare in advance, that is, a so-called, pre-planned unplanned scenario.

If you decide to switch one or more workloads to run actively in AAPLEX2, you typically will use a pre-coded planned action GDPS script to perform the switch of the workloads you want. Switching a workload in this case requires the following actions, all of which can be performed in a single script:

1. Stop the routing of transactions for the selected workloads to AAPLEX1.
2. Wait until all updates for the selected workloads on AAPLEX1 are replicated to AAPLEX2.
3. Stop replication for the selected workloads from AAPLEX1 to AAPLEX2.
4. Start replication for the selected workloads from AAPLEX2 to AAPLEX1.
5. Start the routing of transactions for the selected workloads to AAPLEX2.

Such a planned action script, after it is initiated, can complete the requested switching of the workloads in a matter of seconds.

As you can see, we do not stop the selected workloads in AAPLEX1. There is no need to stop the workload for this particular scenario where we simply toggled the subject workloads to the other site to temporarily provide more capacity, remove a temporary single point of failure, or both. We did assume in this case that AAPLEX2 had sufficient capacity available to run the workloads being switched. If AAPLEX2 did not have sufficient capacity, GDPS can have additionally activated On/Off Capacity on Demand (OOCOD) on one or more servers in Site2 running the AAPLEX2 systems before routing transactions there.

Now, assume that you decide to switch Workload_2 to Site2 but you keep Site1/AAPLEX1 as the primary for the other two workloads. When the switch is complete, the resulting position is depicted in Figure 8-5. In this picture, we assume that you have also restarted in place the failed image, AASYS12.

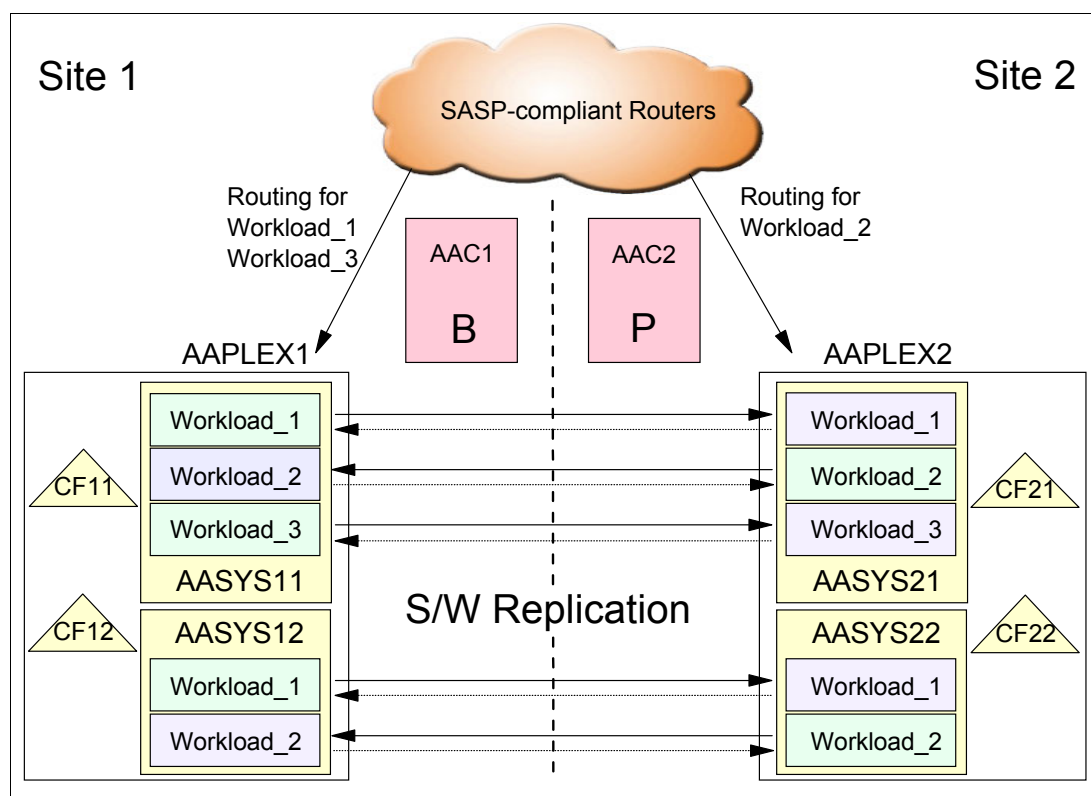


Figure 8-5 GDPS/Active-Active environment with different workloads active in different sites

The router cloud shows to which site the transactions for each of the workloads is being routed. Based on routing, AAPLEX2 is now the active sysplex for Workload_2. AAPLEX1 remains the active sysplex for Workload_1 and Workload_3. Replication for the data for Workload_2 is from AAPLEX2 to AAPLEX1. Replication for the other two workloads is still from AAPLEX1 to AAPLEX2. You might hear the term *dual Active/Active* being used to describe this kind of an environment where both sites or sysplexes are actively running different workloads, but each workload is Active/Standby.

The example that we discussed was an outage of AASYS12 that runs only cloned instances of the applications for Workload_1 and Workload_2. In contrast, Workload_3 has no cloned instances and runs only on AASYS11. An unplanned outage of AASYS11 will result in an actual failure of Workload_3 in its current sysplex. This is a failure that is detected and, based on your workload failure policy, can trigger an automatic switch of the failed workload to the sysplex that is standby for that workload.

However, if you do not want GDPS to perform automatic workload switch for failed workloads, you can select the option of an operator prompt. The operator is prompted as to whether GDPS is to switch the failed workload or not. If the operator accepts switching of the workload, then GDPS will perform the necessary actions to switch the workload. For this kind of switch resulting from a workload failure, whether automatic or operator confirmed, no pre-coded scripts are necessary. GDPS understands the environment and performs all the required actions to switch the workload.

In this particular example, all components of Workload_3 were already running in AAPLEX2 and were ready to receive transactions. If Workload_3 was not running at the time a switch is triggered, then GDPS cannot perform the switch. In this case, the operator is notified that the standby sysplex is not ready to accept transactions for the given workload. The operator can now fix whatever is missing (for example, the operator can use the GDPS GUI to start the subject workload in the target sysplex) and then respond to the prompt, allowing GDPS to proceed with the switch.

Continuing with the same example where AASYS11 has failed, resulting in failure of Workload_3 in AAPLEX1, when GDPS performs the workload switch, then AAPLEX2 becomes the active sysplex and AAPLEX1 should be the standby. However, AAPLEX1 can serve only as standby when AASYS11 is restarted and Workload_3 is started on it.

Meanwhile, transactions are running in AAPLEX2 and updating the data for Workload_3. Until replication components of Workload_3 are restarted in AAPLEX1, the updates are not replicated from AAPLEX2 to AAPLEX1. When replication components are restored on AAPLEX1, replication must be started for Workload_3 from AAPLEX2 to AAPLEX1. The replication components for Workload_3 on AAPLEX1 will now resynchronize, and the delta updates that occurred while replication was down will be sent across. When this is complete, AAPLEX1 can be considered to be ready as the standby sysplex for Workload_3.

For an entire site/sysplex failure, GDPS provides similar capabilities as those for individual workload failure. In this particular case, multiple workloads might be affected. Similar to workload failure, there is policy that determines whether GDPS is to automatically switch workloads that fail as a result of a site failure or perform a prompted switch. The only difference here is that the policy is for workloads that fail as a result of an entire site failure whereas in the previous example, we discussed the policy for individual workload failure. For each workload you can specify individually whether GDPS is to perform an automatic switch or prompt the operator. Furthermore, for each workload you can select a different option (automatic or prompt) for individual workload failure versus site failure.

For entire site or sysplex failures where multiple workloads are affected and switched, GDPS provides parallelization. This means that the RTO for switching multiple workloads is much the same as switching a single workload. Unplanned workload switches are expected to take slightly longer than planned switches. This is because GDPS must wait an amount of time to make sure that the unresponsive condition of the systems/workloads is not because of a temporary stall that can soon clear itself (that is, a false alarm). This is a safety mechanism similar to the failure detection interval for systems running in a sysplex where in the Active-Active case, the aim is to avoid unnecessary switches because of a false alert. However, after the failure detection interval expires and the systems/workloads continue to be unresponsive, the workload switches are fast, and as mentioned previously, are performed in parallel for all workloads being switched.

In summary, GDPS/Active-Active manages individual workloads. Different workloads can be active in different sites. What is not allowed is for a particular workload to be actively receiving and running transactions in more than one site at any given point in time.

8.3.2 Considerations for other non-Active-Active workloads

In the same sysplex where Active-Active workloads are running, you might have other workloads that are not managed by GDPS/Active-Active.

In such an environment, where Active-Active and non-Active-Active workloads coexist, it is important to provide the necessary level of isolation for the Active-Active workloads and data. The data belonging to the Active-Active workloads is replicated under GDPS/Active-Active control and must not be used by non-managed applications.

So, assume you have a workload active in Site1 and standby in Site2. And assume you have a non-managed application in Site1 that uses the same data that is used by your managed workload. If you now switch your managed workload to Site2, the non-managed workload that is not included in the Active-Active solution scope will continue to update the data in Site1 while the managed workload has started to update the database instance in Site2. Such use of data belonging to Active-Active workloads by non-managed applications can result in data loss, potential data corruption, and serious operational issues.

For this reason, the data belonging to Active-Active workloads must not be used by other applications. The simplest way to provide this isolation is to run Active-Active workloads and other workloads in different sysplexes.

We understand that it might not be easy or possible to provide sysplex-level isolation. In this case, if you have isolated the Active-Active workloads and data, you might have other non-managed workloads and the data for such workloads coexisting in the same sysplex with Active-Active. However, another technique (in addition to GDPS/Active-Active for the Active-Active workloads), perhaps hardware replication together with a solution such as GDPS/PPRC, GDPS/GM, or GDPS/XRC, needs to be employed to protect the data and to manage the recovery process for the non-Active-Active workloads.

GDPS/Active-Active has specific function to cooperate and coordinate actions with GDPS/PPRC running on the same sysplex. GDPS/PPRC can protect the entire sysplex, not just the systems running the Active-Active workloads. See 8.5, “GDPS/Active-Active co-operation with GDPS/PPRC or GDPS/MTMM” on page 264 for a more detailed description of this capability.

GDPS/Active-Active also provides for integration of disk replication functions for a GDPS/MGM configuration so that the GDPS/Active-Active Controllers can act as a single point of management and control for both GDPS/Active-Active workloads and GDPS/MGM replication. All data for all systems, for both Active-Active and non Active-Active workloads, can be covered with GDPS/MGM. 8.6, “GDPS/Active-Active disk replication integration” on page 267 provides a high-level overview of this facility.

Because client environments and requirements vary, there is no “one size fits all” type of recommendation that we can make here. Suffice it to say that it is possible to combine GDPS/Active-Active with various other hardware-replication-based GDPS products to provide a total recovery solution for a sysplex that houses both Active-Active and other workloads. If you are unable to isolate your Active-Active workloads into a separate sysplex, discuss this with your IBM GDPS specialist, who can provide you with guidance based on your specific environment and requirements.

8.4 GDPS/Active-Active functions and features

In this section we provide a brief overview of the functions and capabilities provided by the GDPS/Active-Active product:

- ▶ GDPS web graphical user interface
- ▶ Standard Actions for system/hardware automation
- ▶ Monitoring and Alerting
- ▶ GDPS scripts
- ▶ GDPS Query Services

8.4.1 GDPS/Active-Active web interface

GDPS/Active-Active is operated on the Controller systems using an operator interface provided through a web-based browser session. The interface is intuitive and easy to use. Unlike other predecessor GDPS products, there is no 3270-based user interface available with GDPS/Active-Active.

The web interface display, as shown in Figure 8-6, has three sections:

- ▶ A portfolio or menu bar on the left with links to the main GDPS options.
- ▶ A window list on top allowing switching between multiple open frames.
- ▶ An active task frame (work area) where the relevant information is displayed and activities are performed for a selected option. The active task frames for different tasks have a common “look and feel” to the layout.

Nearly all frames have a Help button to provide extensive help text associated with the information displayed and the selections available on that specific frame.

Note: Some panels provided as samples may not be the latest version of the panel. They are intended to give you an idea of the capabilities available using the web interface.

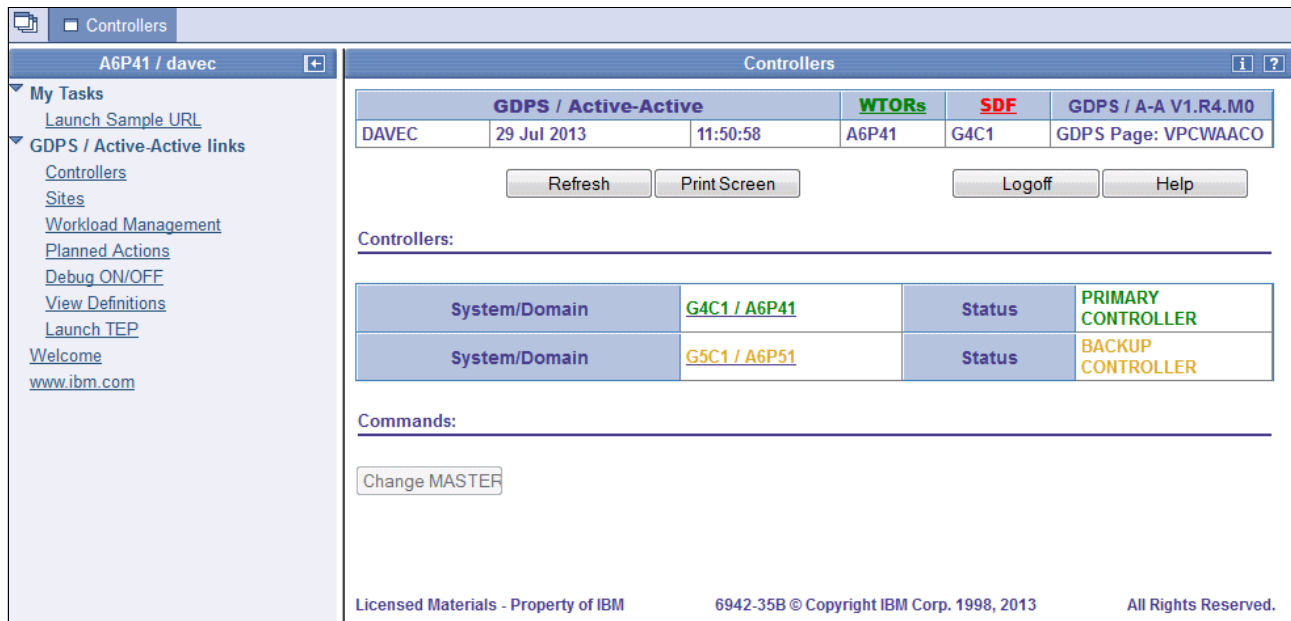


Figure 8-6 GDPS user interface: Initial panel

Controllers panels and functions

When an operator accesses the GDPS/Active-Active web interface, the initial panel that is displayed is the “Controllers” panel; see Figure 8-6. This panel identifies the Controller systems for this GDPS/Active-Active environment. In this example they are the systems named G4C1 (NetView Domain ID A6P41), which is the Primary Controller (or the Master) system, and G5C1 (NetView Domain ID A6P51), which is the Backup Controller.

At the top of the menu bar at the left, you can see that the operator is currently logged on to the Controller system with a domain ID of A6P41, which happens to be the Primary Controller. In this position, the operator can perform only actions such as STOP, which is a graceful shutdown, LOAD or RESET the LPAR, and so on, against the other Controller. GDPS does not allow disruptive actions to be performed against the system that the operator is logged on to.

At the bottom of the panel, you see a disabled Change MASTER button. When this button is selectable (it is selectable only when you are logged on to the Backup Controller), you click it to make the current Backup Controller the new Master Controller (that is, you can perform a Controller switch).

GDPS Standard Actions

Because the operator is normally logged on to the Primary Controller, the operator is allowed to perform actions against only the Backup Controller. When the Backup Controller is selected, the frame shown in Figure 8-7 is displayed. On this frame you see that GDPS Standard Actions can be performed against the other Controller system, which in this case is the Backup Controller.

The screenshot shows a web interface titled "Controllers" with a sub-header "GDPS / Active-Active". It displays system information for DAVEC, including dates and times. Below this are several buttons: Return, Refresh, Print Screen, Logoff, Help, Load, Stop, Dump, Reset, Activate LPAR, and Deactivate LPAR. A section titled "Selected IPLtype/mode" contains a table with one row for system G5C1, showing it is in NORMAL mode with an ACTIVE status. Below this is a "Modify IPLtype/IPLmode" section with buttons for Select, Modify, New, and Delete. Finally, an "Optional IPLtypes/modes for system: G5C1" section shows a table with one row for NORMAL mode, with LoadAddr 2820 and LoadParm 2827G5.

System	IPLtype	IPLmode	IPLset	Loadaddr	LoadParm	Status
G5C1	NORMAL	NORMAL		2820	2827G5	ACTIVE

IPLtype	IPLmode	IPLset	LoadAddr	LoadParm	DumpAddr	DumpParm
NORMAL	NORMAL		2820	2827G5		

Figure 8-7 Web interface frame with GDPS Standard Actions buttons

Figure 8-7 shows the following GDPS Standard Actions that can be performed against the selected target system, available as buttons in the frame:

- ▶ LOAD
- ▶ STOP (graceful shutdown)
- ▶ RESET
- ▶ Activate LPAR
- ▶ Deactivate LPAR
- ▶ Modification and selection of Load Address and Load Parameters to be used during a subsequent LOAD operation

Most of the GDPS Standard Actions require actions to be done on the HMC. The interface between GDPS and the HMC is through the BCP Internal Interface (BCPii). GDPS uses the BCPii interface provided by System Automation for z/OS.

When a specific Standard Action is selected by clicking the button for that action, there are further prompts and windows for operator action such as confirming that they really want to perform the subject operation.

Although in this example we have shown using GDPS Standard Actions to perform operations against the other Controller, in an Active-Active environment, you will also use the same set of Standard Actions to operate against production systems in the environment.

If certain actions are performed as part of a compound workflow (such as planned shutdown of an entire site where multiple systems will be stopped, the LPARs for multiple systems RESET and Deactivated, and so on), then the operator will typically not use the web interface but instead perform the same actions through the GDPS scripting interface. GDPS scripts are discussed in detail in 8.4.3, “GDPS/Active-Active scripts” on page 259.

The GDPS LOAD and RESET Standard Actions (available through the Standard Actions panel or the SYSPLEX script statement) allow specification of a CLEAR or NOCLEAR operand. This provides operational flexibility to accommodate customer procedures, eliminating the requirement to use the HMC to perform specific LOAD and RESET actions.

GDPS supports stand-alone dumps using the GDPS Standard Actions panel. The stand-alone Dump option can be used against any z Systems operating system defined to GDPS. Customers using GDPS facilities to perform HMC actions no longer need to use the HMC for stand-alone dumps.

Sites panels and functions

The Sites task, when selected from the menu on the left side of every web interface window, allows you to perform GDPS Standard Actions against the production systems within your GDPS/Active-Active environment. Two examples of the frame that is displayed when you select this task are shown in Figure 8-8. As shown, the information provides a view of the status of the systems within the sites. The upper panel in the display shows normal status, with all systems active. The lower panel gives clear indication of a problem in Site G5, where neither of the two expected systems are active.

The figure shows two screenshots of the 'Sites' window. Both windows have a header bar with 'Sites' and icons for info, help, and close. Below the header is a table with columns: 'GDPS / Active-Active', 'WTORs', 'SDF', and 'GDPS / A-A V1.R1.M0'. The first row of data shows 'DAVEC', '17 Feb 2011', '11:54:15', 'A6P41', 'G4C1', and 'GDPS Page: VPCWAASI'. Below this table are buttons for 'Refresh', 'Print Screen', 'Logoff', and 'Help'.

The 'Sites' section below the buttons shows a table with two rows:

Site	Status
G4	ACTIVE 2 out of 2 Systems
G5	ACTIVE 2 out of 2 Systems

The second screenshot is identical but the status for Site G5 is 'ACTIVE 0 out of 2 Systems'.

Figure 8-8 Sites window

Essentially, apart from the standard header information, this panel allows you to select which of the sites you want to interact with. You simply click the site name. Figure 8-9 shows the frame displayed when, in our example, **G4** is selected.

The screenshot shows the 'Sites' window with 'Selected Site: G4' and the message 'GDPS/PPRC is setup in target sysplex'. Below this is a table with columns: 'Sysname', 'CA', 'Status', 'IPLtype', 'Lpar', 'IPLmode', 'Auto', 'Loadaddr', and 'Loadparm'.

Sysname	CA	Status	IPLtype	Lpar	IPLmode	Auto	Loadaddr	Loadparm
G4C2	CA	ACTIVE	NORMAL	S636	NORMAL	YY	4800	4807G4
G4P1	A	ACTIVE	NORMAL	S235	SITE1	YY	2900	2908G4
G4P2	A	ACTIVE	NORMAL	S236	SITE1	YY	2900	2908G4
CF1		MANUAL	NORMAL	S238	NORMAL	NN		

Figure 8-9 Sites window for site/sysplex G4 selected

You can then select the specific system you want to use as a target for a GDPS Standard Actions operation. Performing Standard Actions, such as STOP, LOAD, RESET and so on against a production system is identical to performing such actions against a Controller as shown on Figure 8-7 on page 250 and described in “GDPS Standard Actions” on page 250.

Workload Management panels and functions

The Workload Management task, selected from the menu bar, displays the Workload Management window. An example of this frame is shown in Figure 8-10. This frame shows and provides “at a glance” high-level status summary information for all workloads, updates, and queries, that are defined for this GDPS environment.

Workload Management

?

?

×

GDPS / Active-Active

WTORs

SDF

GDPS / A-A V1.R4.M0

DAVEC

5 Aug 2013

11:16:30

A6P41

G4C1

GDPS Page: VPCWAWOM

Refresh

Logoff

Help

ACTIVEQUERY workload configurations

Type	Workload	Site 1	Site 2	Routing	Latency
Update	CICS WORKLOAD 1	UP / G4	UP / G5	QUIESCED	N/A
Query	CICS WORKLOAD 2	UP / G4	UP / G5	QUIESCED	N/A
Update/MCG	CICS WORKLOAD 3	UP / G4	UP / G5	G4	N/A
Query	CICS WORKLOAD 4	UP / G4	UP / G5	G4	N/A
Query	CICS WORKLOAD 5	UP / G4	UP / G5	QUIESCED	N/A
Update/MCG	CICS WORKLOAD 6	UP / G4	UP / G5	QUIESCED	N/A
Query	CICS WORKLOAD 7	UP / G4	UP / G5	QUIESCED	N/A
Query	CICS WORKLOAD 8	UP / G4	UP / G5	QUIESCED	N/A
Update	IMS WORKLOAD 1	DOWN / G4	DOWN / G5	QUIESCED	N/A
Query	IMS WORKLOAD 2	DOWN / G4	DOWN / G5	QUIESCED	N/A

ACTIVESTANDBY workload configurations

Type	Workload	Site 1	Site 2	Routing	Latency
Update	CICS_WORKLOAD_9	DOWN / G4	DOWN / G5	QUIESCED	N/A

Figure 8-10 Workload Management frame

The status shown in each of the sites is based on information from GDPS Monitoring and from System Automation running in the production systems in that site.

You can click any of the workload names to select the details frame for that workload. An example of this Workload details frame is shown in Figure 8-11.

Workload Management

GDPS / Active-Active

WTORs

SDF

GDPS / A-A V1.R4.M0

DAVEC

5 Aug 2013

11:52:02

A6P41

G4C1

GDPS Page: VPCWAWO1

Return

Refresh

Print Screen

Logoff

Help

Commands against workload: CICS_WORKLOAD_4

Start Workload

Stop Workload

Start Workload G4

Stop Workload G4

Start Workload G5

Stop Workload G5

Start Routing

Stop Routing

Start Routing to G4

Stop Routing to G4

Start Routing to G5

Stop Routing to G5

Workload	CICS_WORKLOAD_4		
Site 1	G4	Status	UP
Site 2	G5	Status	UP
Routing direction	G4		
Associated with	CICS_WORKLOAD_3		
Max Latency Exceeded	N/A	Latency	*
Workload Type	ACTIVEQUERY		
Workload Failure	AUTOMATIC		
Site Failure	AUTOMATIC		
Assigned Auto Op	AUTWWQ2		
Query Routing	DYNAMIC		

Figure 8-11 Workload details frame

The Workload details frame allows you to perform the following operations against the selected workload such as Start/Stop of the workload or Start/Stop of routing for that workload to one site or the other.

In addition to these operations, the frame provides further status detail associated with the selected workload.

Similar to Standard Actions, there are GDPS script statements that perform these same operations and typically, a script is used to perform these actions and Standard Actions for a compound/complex scenario such as an entire site shutdown. See 8.4.3, “GDPS/Active-Active scripts” on page 259 for details about using the GDPS scripting capability.

Planned Actions panels and functions

GDPS Planned Actions are initiated from the Planned Actions frame within the GDPS user interface. When you select the Planned Actions task from the menu bar, you see a Planned Actions frame similar to that shown in Figure 8-12.

Planned Actions					
GDPS / Active-Active			WTORs	SDF	GDPS / A-A V1.R4.M0
DAVEC	29 Jul 2013	12:21:57	A6P41	G4C1	GDPS Page: VPCWPFUS
Help Refresh Print Screen Logoff					
Scriptname	Comment:				
ACT_LOAD_G5C1	ACTIVATE & LOAD G5C1				
ACTCF1NORMAL	ACTIVATE AND ENABLE CF1 USING INGCFL				
ACTCPC	POR CP23 USING SCLP				
ACTIVATE_CF_G5	ACTIVATE CF2 IN G5				
ACTIVATE_GXP1	ACTIVATE G4P1 (SITE1) AND G5P1 (SITE2)				
ACTIVATE_G4P1	ACTIVATE G4P1 ON SITE1				
ACTIVATE_G5C1	ACTIVATE G5C1 ON SITE2				
ACTIVATE_G5P1	ACTIVATE G5P1 ON SITE2				
ACTIVATE_G5P2	ACTIVATE G5P2 ON SITE2				
ALL_CAPACITY	CAPACITY STATEMENT : SYNTAX CHECK				
ALL_CBU	JUST FOR SYNTAX CHECKING - DON				
ALL_OOCOD	JUST A SYNTAX CHECKING				
ASSIST	SEND A WTOR AND WAIT FOR ANSWER				
BADROUTING	START ROUTING OF WORKLOAD CICS NUM 1 ON G6				

Figure 8-12 Sample Planned Actions frame

Planned Actions allows you to view and execute scripts for planned scenarios such as site shutdown, site start, or CEC shutdown and start. You are presented with a list of scripts that you have already coded, anticipating a given planned scenario. Along with the name of the script, you are also presented with a comment that describes what a given script is intended for. You can then select a script for viewing and execution on this panel.

When you select a script from the list, you are presented with a panel that displays the actual script content as shown in Figure 8-13. On this panel, after you view the actual script content, you can execute it. If you have selected the wrong script, you can return.

Planned Actions					
GDPS / Active-Active			WTORs	SDF	GDPS / A-A V1.R4.M0
DAVEC	29 Jul 2013	12:23:48	A6P41	G4C1	GDPS Page: VPCWPFU2
Total no of steps :3 Script statements in order					
Execute Script Return Refresh Print Screen Logoff Help					
COMM='ACTIVATE & LOAD G5C1' SYSPLEX='ACTIVATE G5C1 LPAR' SYSPLEX='LOAD G5C1'					

Figure 8-13 Planned Action script example

Launching Tivoli Enterprise Portal from the GDPS web interface

You can use the Launch Tivoli Enterprise Portal link on the menu bar to view information available through the Tivoli Enterprise Portal. The Tivoli Enterprise Portal provides views and levels of detail that pertain to the GDPS/Active-Active environment other than what is available through the GDPS web interface. Therefore, when investigating a problem (for example, because of an alert that is raised in GDPS), it can be quite useful to simply launch Tivoli Enterprise Portal directly from the GDPS web interface and drill down into the views of your environment that are available through Tivoli Enterprise Portal.

After Tivoli Enterprise Portal is launched, you can get to Active-Active frames to view details pertaining to the Active-Active Load Balancers, Replication Servers, Workload Lifeline Advisors, and Workloads. Figure 8-14 shows Tivoli Enterprise Portal views of Replication Servers. The bottom of the view contains summary information for the replicator associated with each of the workloads managed by GDPS/Active-Active. The graph at the top shows details about the breakdown of latency for each of the replicators in the environment.

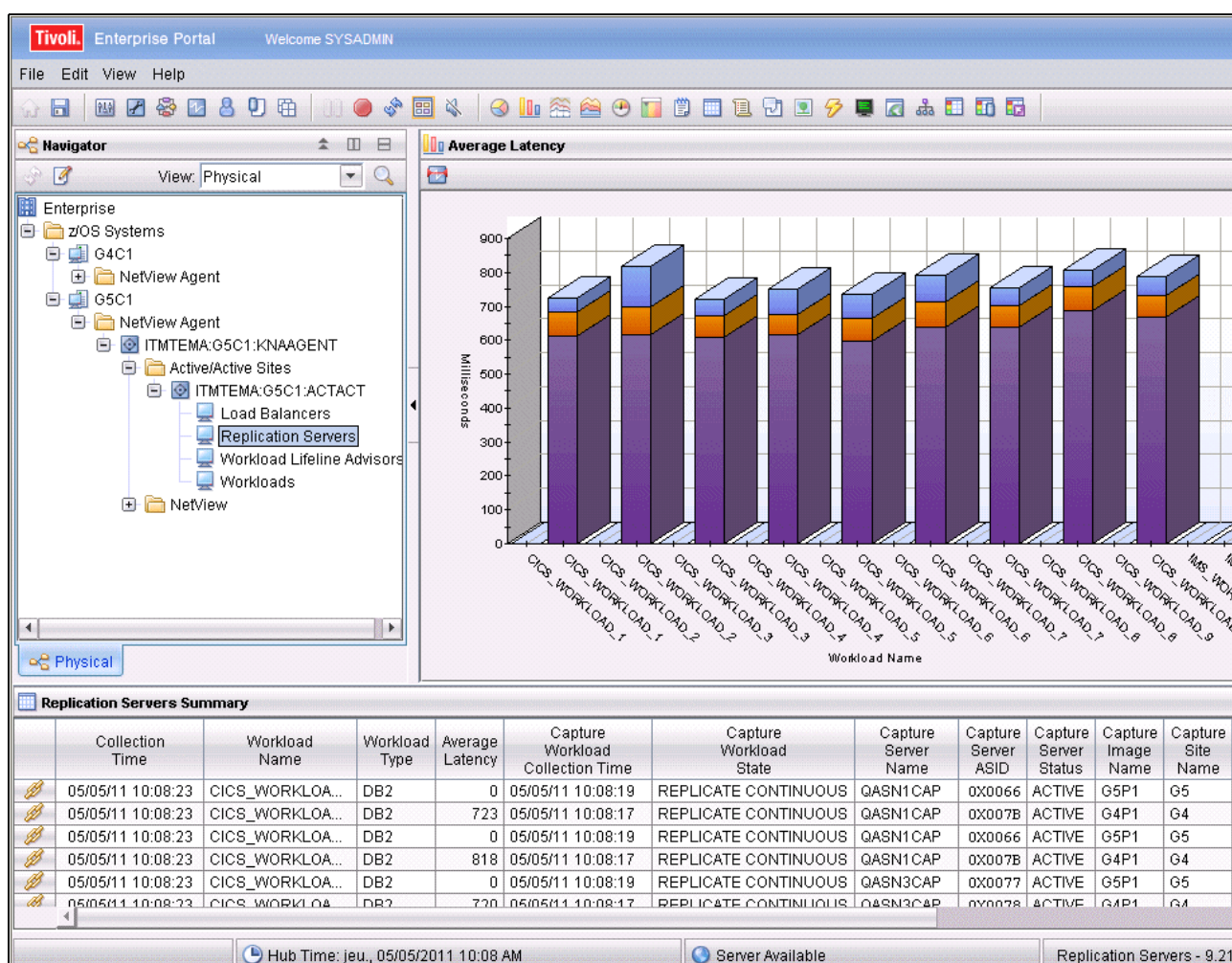


Figure 8-14 Launch Tivoli Enterprise Portal: Replication Servers frame

The Tivoli Enterprise Portal, in addition to providing a monitoring interface to the overall solution, allows you to set up specific situations for alerting of conditions such as the replication latency exceeding a certain threshold. The workload-related workspaces can also show quickly such things as the number of servers active in both sites and where the routing is active to. This information can be useful to correlate against that shown in the GDPS web interface to confirm the status of any particular resources.

Other web interface options

Other options are available to the operator through the web interface.

Status Display Facility

Status Display Facility (SDF) is the focal point for monitoring the GDPS/Active-Active environment. A link to SDF is available on the top portion of every web interface frame. SDF is an important component of GDPS and is described in 8.4.2, “GDPS/Active-Active monitoring and alerting” on page 257.

WTORs

Similar to SDF, the WTORs function is selectable on the top portion of every web interface frame. The WTORs function opens a new window to display any write to operator with reply (WTOR) messages that are outstanding and provides the option to reply to any selected message.

Debug On/Off

As a NetView based automation application, GDPS/Active-Active uses the NetView log as the main repository for information logging. In addition to the NetView log, selected critical GDPS messages are also sent to the z/OS system log.

The GDPS Debug facility enables logging in the NetView log, providing more detailed trace entries pertaining to the operations that GDPS is performing. If you encounter a problem, you might want to collect debug information for problem determination purposes. If directed by IBM support, you might need to trace execution of specific modules. The GDPS debug facility also allows you to select the modules to be traced. The Debug frame is displayed when you select **Debug On/Off** task on the menu bar.

View definitions

The view definitions option, also selected through the menu bar, allows you to view the various definitions and options related to GDPS that are in effect. The bulk of the GDPS definitions are made in the System Automation policy database. If you modify some of these definitions or for any other reason want to check what definitions GDPS is using, you can use this facility.

8.4.2 GDPS/Active-Active monitoring and alerting

GDPS/Active-Active Controller systems perform periodic monitoring of resources and conditions that are critical or important for the healthy operation of the environment. For example, GDPS checks whether the workloads that are managed by GDPS are running on both the active and the standby sysplexes, whether the BCP Internal Interface is functional, whether the connectivity from the Controller to the production systems is intact, current replication latency and so on. If GDPS discovers any exception situations, it raises Status Display Facility (SDF) alerts.

In addition to any exception condition that might be discovered through monitoring, GDPS also captures messages from other components in the environment that can be indicative of a problem and raises alerts.

The Status Display Facility (SDF) is a facility provided by System Automation and is used as the primary status feedback mechanism for GDPS. SDF can be viewed by selecting the SDF link, which is available on the top portion of every GDPS web interface frame.

If all is well and there are no alerts indicative of a potential issue with the environment, the SDF link on the GDPS web interface frames are displayed in green. If any SDF entry is displayed in a color other than green, it is indicative that there is an alert. For example, pink is used to report a problem that is not catastrophic and red is used for a serious exception condition.

No matter which frame operators view, they can click SDF on the top portion of the frame to view the SDF window and check the alert.

In addition to using SDF to monitor the GDPS status, when Standard Actions or scripts are executing, each step is displayed in the trace portion of the SDF window. This allows the operator to follow script execution.

When you select the SDF view, it opens in a new browser window, as shown in Figure 8-15. The windows shows several sections:

- ▶ Site alerts, split into two categories:
 - Workload alerts
 - Site or system alerts
- ▶ GDPS controller alerts
- ▶ Trace entries alerts

GDPS / Active-Active			WTOBs	SDF	GDPS / A-A V1.R4.M0
DAVEC	29 Jul 2013	12:30:57	A6P41	G4C1	GDPS Page: VPCWASDL
Refresh			Close Window		Help
Delete Trace					Customize
Site1: G4			Site2: G5		
WorkLoad	System/Site		WorkLoad	System/Site	
G4_GEO1131	G4P1_GEO1210		G5_GEO1131	G5P1_GEO1210	
G4_GEO1131	G4C2_GEO1210		G5_GEO1131	G5P2_GEO1210	
G4_GEO1131	G4P1_GEO1218		G5_GEO1131	G5C2_GEO1210	
	G4P2_GEO1210			G5C2_GEO1218	
	G4P2_GEO1218			G5P1_GEO1218	
				G5P2_GEO1218	
Controller alerts					
GEO1199W GDPS/A-A INITIALIZATION IS COMPLETE					
Trace Entries					
07/29/13 11:54:30 LOAD PLANNED/STANDARD ACTION STARTED					
07/28/13 13:33:05 MESSAGE:GEO1174A,RECOVERY NOT INITIATED, OPERATOR REPLIED NO					
07/29/13 11:54:31 SYSPLEX=LOAD G5C1 NORMAL NORMAL NOCLEAR STARTED /1					
07/29/13 11:59:38 SYSPLEX=LOAD G5C1 NORMAL NORMAL NOCLEAR ENDED RC=0 /1					
07/29/13 11:59:38 LOAD PLANNED/STANDARD ACTION ENDED					

Figure 8-15 SDF window

To see further details about any alert, simply click the alert. A new window is displayed with the details for the selected alert. For example, if you click the first alert at the upper left (G4_GEO1131), you are presented with the window shown in Figure 8-16.

GDPS			WTOs	SDF	GDPS / A-A V1.R4.M0
DAVEC	29 Jul 2013	12:31:37	A6P41	G4C1	GDPS Page: VPCWPSDD

Component	WK	System	G4C1
Color	RED	Priority	140
Date	07/29/13	Time	12:26:52
Reporter	AUTGEO5	Node	A6P41
Ref.Value	G4_GEO1131		

Alert Text

GEO1131 WORKLOAD CICS_WORKLOAD_2 IN G4PXPlexGRP IS :DOWN

Prev. Trace Entry Next Trace Entry

Close Window Delete Alert

Figure 8-16 SDF Alert detail display

8.4.3 GDPS/Active-Active scripts

We have already reviewed the GDPS web interface, which provides powerful functions to help you manage your workloads and systems in the sites where they are running. However, the GDPS web interface is not the only means for performing these functions. Nearly all of the functions that can be manually initiated by the operator through the web interface are also available through GDPS scripts. There are other actions not available through the web interface, such as activating capacity on demand (CBU or OOCOD), that are only possible using GDPS scripts. In addition to the set of script commands that are supplied by GDPS, you can integrate your own REXX procedures and execute them as part of a GDPS script.

A *script* is simply a procedure recognized by GDPS that pulls together into a workflow (or a list, if you will) one or more GDPS functions to be executed one after the other. GDPS checks the result of each command and proceeds with the next command only if the previous command executed successfully. Scripts can be initiated manually through the GDPS panels (using the Planned Actions interface), automatically by GDPS in response to an event (Unplanned Actions), or through a batch interface.

Scripts are easy to code. Using scripts forces you to properly plan for the actions you need to take for various planned and unplanned outage scenarios, and how to bring the environment back to normal. In this sense, when you use scripts, you properly plan even for an unplanned event and will not be caught unprepared. This is an extremely important aspect of GDPS. Scripts are powerful because they can use the full capability of GDPS.

The ability to plan and script your scenarios and invoke all the GDPS functions provides the following benefits:

- **Speed**

A script will execute the requested actions as quickly as possible. Unlike a human, it does not need to search for the latest procedures or the commands manual. It can check results fast and continue with the next statement immediately when one statement is complete.

- **Consistency**

If you were to look into most computer rooms immediately following a system outage, what would you see? Mayhem! Operators frantically scrambling for the latest system programmer instructions. All the phones ringing. Every manager within reach asking when the service will be restored. And every systems programmer with access vying for control of the keyboards. All this results in errors because humans naturally make mistakes when under pressure. But with automation, your well-tested procedures will execute in exactly the same way, time after time, regardless of how much you shout at them.

- **Thoroughly thought-out and tested procedures**

Because they behave in a consistent manner, you can test your procedures over and over until you are sure they do everything that you want, in exactly the manner that you want. Also, because you need to code everything and cannot assume a level of knowledge (as you might with instructions intended for a human), you are forced to thoroughly think out every aspect of the action the script is intended to undertake. And because of the repeatability and ease of use of scripts, they lend themselves more easily to frequent testing than manual procedures.

- **Reduction of requirement for onsite skills**

How many times have you seen disaster recovery tests with large numbers of people onsite for the test and many more standing by for a call? How realistic is this? Can all of these people actually be onsite on short notice if there really was a catastrophic failure? Using GDPS automation and scripts removes the need for the numbers and the range of skills that enterprises traditionally needed to do complex or compound reconfiguration and recovery actions.

Planned Actions

As mentioned, GDPS scripts are simply procedures that pull together into a list one or more GDPS functions to be executed sequentially. For the scripted procedures that you can use for planned changes to the environment, these scripts can be initiated from the Planned Actions frame, as described in “Planned Actions panels and functions” on page 255.

As a simple example, you can have a script that recycles a z/OS system. This is an action that you would perform if you apply maintenance to the software that required a re-IPL of the system. The script executes the STOP standard action, which performs an orderly shutdown of the target system followed by a LOAD of the same system.

However, it is possible that in your environment you use alternate system volumes. While your system runs on one set of system volumes, you perform maintenance on the other set. So assuming you are running on alternate SYSRES1 and you apply this maintenance to SYSRES2, your script also needs to point to SYSRES2 before it performs the LOAD operation.

As part of the customization you perform when you install GDPS, you can define entries with names of your choice for the load address and load parameters associated with the alternate SYSRES volumes for each system. When you want to LOAD a system, you simply use a script statement to point to one of these pre-customized entries using the entry name that you used when defining them to GDPS.

Example 8-1 shows a sample script to perform this action. In this example, MODE=ALTRES2 points to the load address and load parameters associated with alternative SYSRES2 where you applied your maintenance.

Example 8-1 Sample script to re-IPL a system on an alternate SYSRES

```
COMM='Re-IPL system AASYS11 on alternate SYSRES2'  
SYSPLEX='STOP AASYS11'  
IPLTYPE='AASYS11 MODE=ALTRES2'  
SYSPLEX='LOAD AASYS11'
```

Example 8-2 shows a sample script to switch a workload that is using DB2 or IMS replication from its current active site to its standby site.

Example 8-2 Sample script to switch a workload between sites

```
COMM='Switch WORKLOAD_1'  
ROUTING 'SWITCH WORKLOAD=WORKLOAD_1'
```

No target site is specified in the ROUTING SWITCH statement. This is because GDPS has awareness of where WORKLOAD_1 is currently active and GDPS simply switches it to the other site. The single ROUTING SWITCH statement³ performs the following actions:

- ▶ Stops routing of update transactions to the original active site.
- ▶ Waits for replication of the final updates in the current active site to drain.
- ▶ Starts routing update transactions to the former standby site which now becomes the new active site for this workload.
- ▶ Also, if a query workload is associated with this update workload, and if, for example, 70% of queries were being routed to the original standby site, after the switch, the routing for the query workload is changed to send 70% of queries to the new standby site.

All of these actions are done as a result of executing a single script with a single command. This demonstrates the simplicity and power of GDPS scripts.

³ At the time of writing, the ability to switch workloads with a single command is possible only for workloads that use either DB2 or IMS data exclusively. The final site shutdown example shows how other workloads would be switched in the context of a site shutdown scenario.

Our final example for using a script can be for the purpose of shutting down an entire site, perhaps in preparation for disruptive power maintenance at that site. For this example, we use the configuration previously described with three workloads, all active in Site1, as shown in Figure 8-17.

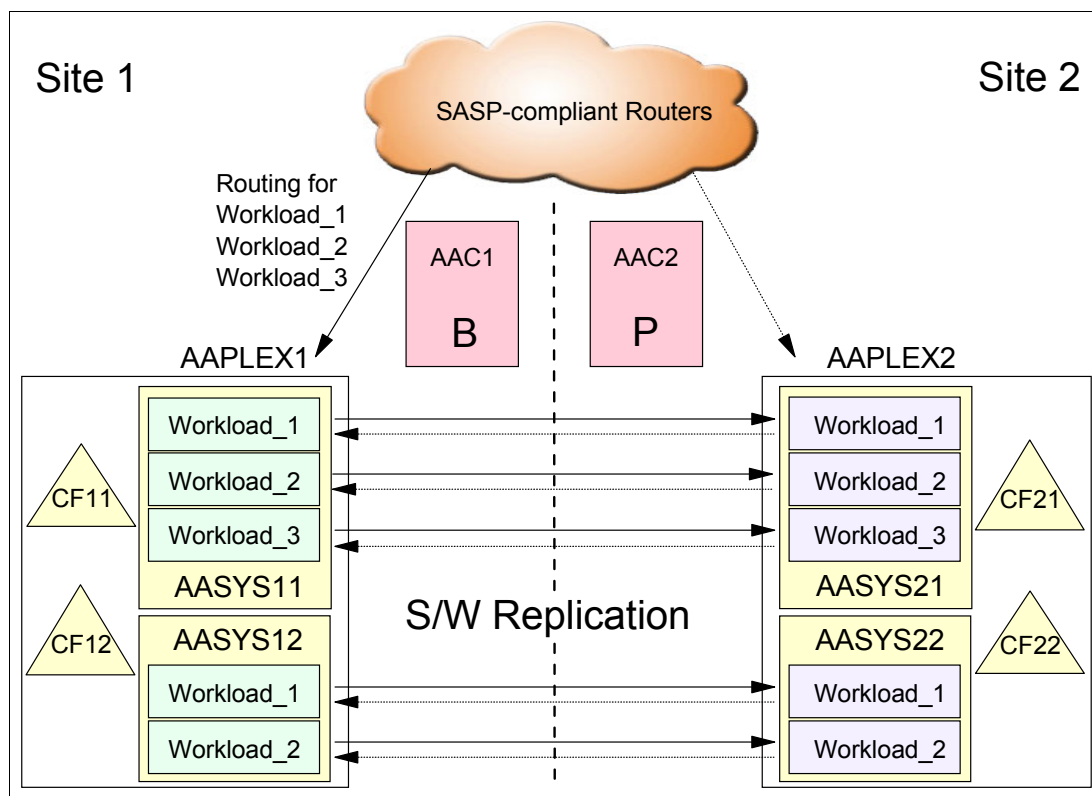


Figure 8-17 GDPS/Active-Active environment sample for Site1 shutdown script

The sequence of events to completely shut down Site1 is as follows:

1. Stop routing transactions for all workloads to AAPLEX1.
2. Wait until all updates on AAPLEX1 are replicated to AAPLEX2. In this example, we are assuming these workloads do not support the ROUTING SWITCH function (non-DB2 or IMS replication based). Therefore, full automation for controlling whether data has drained or not isn't available yet.
3. Stop replication from AAPLEX1 to AAPLEX2.
4. Activate On/Off Capacity on Demand (OOCOD) on the CECs running the AAPLEX2 systems and CFs (although not shown in the diagram, for this example we assume the CECs are named CPC21 and CPC22).
5. Start routing transactions for all workloads to AAPLEX2.
6. Stop the AASYS11 and AASYS12 systems.
7. Deactivate the system and CF LPARs in Site1.

The planned action script to accomplish the Site1 shutdown for this environment is shown in Example 8-3.

Example 8-3 Sample Site1 shutdown script

```
COMM='Switch all workloads to Site2 and Stop Site1'
ROUTING='STOP WORKLOAD=ALL SITE=AAPLEX1'
ASSIST='WAIT UNTIL ALL UPDATES HAVE DRAINED - REPLY OK WHEN DONE'
REPLICATION='STOP WORKLOAD=ALL FROM=AAPLEX1 TO=AAPLEX2'
OOCOD='ACTIVATE CPC=CPC21 ORDER=order#'
OOCOD='ACTIVATE CPC=CPC22 ORDER=order#'
ROUTING='START WORKLOAD=ALL SITE=AAPLEX2'
SYSPLEX='STOP SYSTEM=(AASYS11,AASYS12)'
SYSPLEX='DEACTIVATE AASYS11'
SYSPLEX='DEACTIVATE AASYS12'
SYSPLEX='DEACTIVATE CF11'
SYSPLEX='DEACTIVATE CF12'
```

These sample scripts demonstrate the power of the GDPS scripting facility. Simple, self-documenting script statements drive compound and complex actions. A single script statement can operate against multiple workloads or multiple systems. A complex procedure can be described in a script by coding just a handful of statements.

Another benefit of such a facility is the reduction in skill requirements to perform the necessary actions to accomplish the task at hand. For example, in the workload switch and the site shutdown scenarios, depending on your organizational structure within the IT department, you might have required database, application/automation, system, and network skills to be available to perform all of the required steps in a coordinated fashion.

Batch scripts

GDPS also provides a flexible batch interface to invoke planned action scripts. These scripts are not (and cannot be) invoked from the GDPS web interface, but are invoked from some other planned event external to GDPS. The initiating event can be, for example, a job, or messages triggered by a job scheduling application.

This capability, along with the Query Services described in 8.4.4, “GDPS/Active-Active Query Services” on page 264, provides a rich framework for user-customizable automation and systems management procedures.

Switch scripts

As described in 8.3.1, “GDPS/Active-Active: A closer look” on page 244, in the event of workload or entire site failure, GDPS performs the necessary steps to switch one or more workloads to the standby site. This switching, based on the selected policy, can be completely automatic with no operator intervention or can occur after operator confirmation. However, in either case, the steps required to switch any workload are performed by GDPS and no scripts are required for this.

Although GDPS performs the basic steps to accomplish switching of affected workloads, there might be additional actions specific to your environment that you want GDPS to perform along with the workload switch steps. One such example can be activating CBU for additional capacity in the standby site.

Switch scripts are unplanned actions that run as a result of a workload failure or site failure detected by GDPS. These scripts cannot be activated manually. They are initiated automatically, if you have coded them, as a result of an automatic or prompted workload or site switch action initiated by GDPS. The intent of Switch scripts is to complement the standard workload/site switch processing that is performed by GDPS.

8.4.4 GDPS/Active-Active Query Services

GDPS maintains configuration information and status information in NetView variables for the various elements of the configuration that it manages. GDPS Query Services is a capability that allows client-written NetView REXX programs to query the value for numerous GDPS internal variables. The variables that can be queried pertain to the GDPS environment itself (such as the version and release level of the GDPS control code), the sites, the sysplexes, and the workloads managed by GDPS/Active-Active.

Query Services allows clients to complement GDPS automation with their own automation code. In addition to the Query Services function, which is part of the base GDPS product, GDPS provides several samples in the GDPS SAMPLIB library to demonstrate how Query Services can be used in client-written code.

8.5 GDPS/Active-Active co-operation with GDPS/PPRC or GDPS/MTMM

In an Active-Active environment, it is essential that each of the sysplexes running the Active-Active workloads is as highly available as possible. As such, we suggest that the Active-Active workloads are Parallel Sysplex enabled, data sharing applications. Although this eliminates a planned or unplanned system outage from being a single point of failure, disk data within each local sysplex is not protected by Parallel Sysplex alone.

To protect the data for each of the two sysplexes comprising the GDPS/Active-Active environment, these sysplexes can be running GDPS/PPRC or GDPS/MTMM with PPRC replication and HyperSwap, which complement and enhance local high, continuous availability for the sysplex or sysplexes. We describe the various capabilities available with GDPS/PPRC in Chapter 3, “GDPS/PPRC” on page 53 and in Chapter 7, “GDPS/MTMM” on page 189 for GDPS/MTMM.

With GDPS/Active-Active and GDPS/PPRC (or GDPS/MTMM) monitoring and managing the same production systems for a particular sysplex, certain actions must be coordinated. This is so that the GDPS controlling systems for the two environments do not interfere with each other or that one environment does not misinterpret actions taken by the other environment.

For example, it is possible that one of the systems in the sysplex needs to be re-IPLed for a software maintenance action. The re-IPL of the system can be performed either from a GDPS/Active-Active Controller or using GDPS/PPRC (or GDPS/MTMM) that is running on all systems in the same sysplex. Assume that you initiate the re-IPL from the GDPS/Active-Active Controller. GDPS/PPRC (or GDPS/MTMM) will detect that this system is no longer active. It will interpret what was a planned re-IPL of a system as a system failure and issue a takeover prompt.

The GDPS/Active-Active co-operation with GDPS/PPRC (or GDPS/MTMM) provides coordination and serialization of actions across the two environments to avoid issues that can stem from certain common resources being managed from multiple control points. In our example, when you initiate the re-IPL from the Active-Active Controller, it communicates this action to the GDPS/PPRC (or GDPS/MTMM) controlling system.

The GDPS/PPRC (or GDPS/MTMM) controlling system then locks this system as a resource so that no actions can be performed against it until the Active-Active Controller signals completion of the action. This same type of coordination takes place regardless of whether the action is initiated by GDPS/Active-Active or GDPS/PPRC (or GDPS/MTMM).

GDPS/Active-Active can support coordination with GDPS/PPRC (or GDPS/MTMM) running in either or both of the Active-Active sites.

In Figure 8-18 on page 266, we show a GDPS/Active-Active environment across two regions, Region A and Region B. SYSPLEXA in Region A and SYSPLEXB in Region B comprise the two sysplexes managed by GDPS Active-Active. Systems AAC1 and AAC2 are the GDPS/Active-Active Controller systems. Also, each of these sysplexes is managed by an instance of GDPS/PPRC, with systems KP1A/KP2A being the GDPS/PPRC controlling systems for SYSPLEXA and KP1B/KP2B being the GDPS/PPRC controlling systems in SYSPLEXB. The GDPS Active-Active Controllers communicate to each of the GDPS/PPRC controlling systems in both regions. It is this communication that makes the cooperation possible.

SYSPLEXA contains the data for the Active-Active workloads, and also other data for applications running in the same sysplex but not managed by Active-Active, plus the various system infrastructure data which is also not managed by Active-Active. All of this data that belongs to SYSPLEXA is replicated within Region A using PPRC and is HyperSwap-protected and managed by GDPS/PPRC. The Active-Active data is replicated through software to SYSPLEXB.

Similarly, there is another instance of GDPS/PPRC managing SYSPLEXB with the Active-Active data and also any non Active-Active data belonging to SYSPLEXB being replicated through PPRC and HyperSwap protected within Region B.

Each of SYSPLEXA and SYSPLEXB could be running in a single physical site or across two physical sites within their respective regions.

All of the data within both sysplexes is HyperSwap protected, meaning that disk within a region is not a single point of failure and the sysplex can continue to function during planned or unplanned disk outages. HyperSwap is transparent to all applications running in the sysplex (assuming that the data for all applications is replicated with PPRC). This means that it is also transparent to all of the subsystems in charge of running the Active-Active workloads, replicating the Active-Active data, and monitoring the Active-Active environment. HyperSwap of disks within a region is transparent to the cross-region software replication process. Software replication will know about and capture data only from the logs on the current primary PPRC volumes. If there is a HyperSwap, software replication simply continues capturing data from the logs which are now on the new primary volumes.

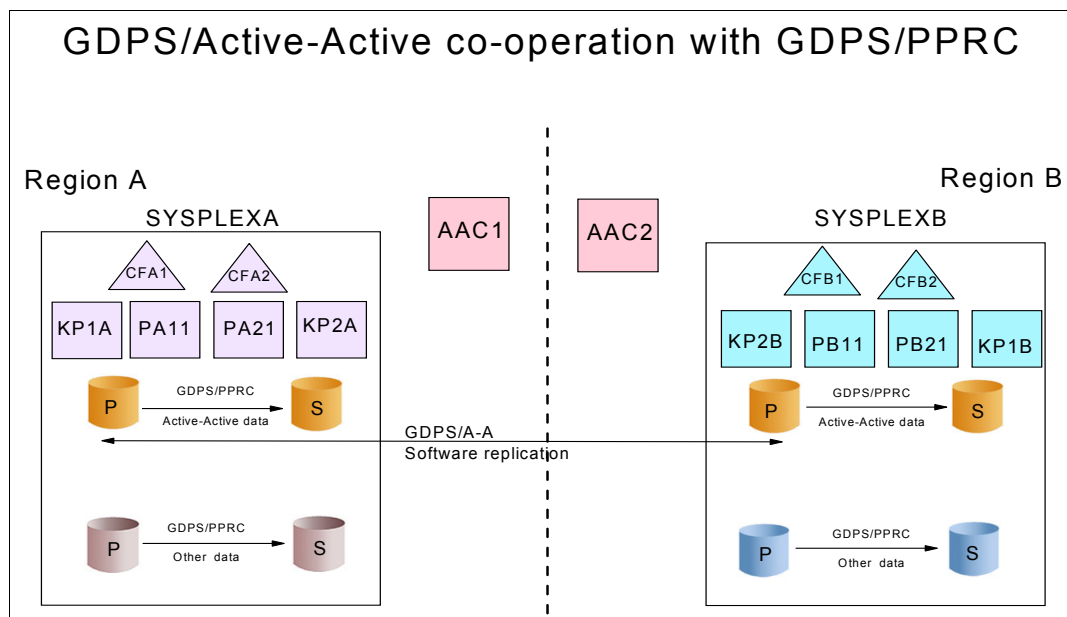


Figure 8-18 GDPS/Active-Active co-operation with GDPS/PPRC

In addition to the HyperSwap protection, GDPS/PPRC (or GDPS/MTMM) provides several other facilities and benefits which are described in Chapter 3, “GDPS/PPRC” on page 53 (or Chapter 7, “GDPS/MTMM” on page 189). Given the capabilities of GDPS/PPRC and GDPS/MTMM, we would expect that clients would perform most of the day-to-day system, sysplex, and PPRC management activities for each of the two sysplexes using the respective GDPS/PPRC or GDPS/MTMM facilities. However, GDPS/Active-Active must be used for management and switching of the Active-Active workloads and replication. Finally, management (actions such as STOP and IPL) of the Active-Active controllers can be performed only using GDPS/Active-Active because these systems are outside of the respective sysplexes and GDPS/PPRC (or GDPS/MTMM) scope of control is limited to the systems in the sysplex.

In summary, GDPS/Active-Active and GDPS/PPRC (or GDPS/MTMM) can be deployed in a complementary fashion and these products provide the necessary controls to facilitate any coordination that would be required when operating on common resources.

8.6 GDPS/Active-Active disk replication integration

The primary focus for GDPS/Active-Active is to provide near-continuous availability for your Active-Active workloads. The GDPS/Active-Active disk replication integration functions are designed to *complement* the GDPS/Active-Active functions and provide disaster recovery provision for your entire production sysplexes under the control of GDPS/Active-Active. This is done by integrating disk-based replication control into GDPS/Active-Active so that you can manage and control aspects of your disaster recovery provision for these sysplexes from a single point. Currently, this support is provided with the GDPS/MGM solution that would be implemented for protection of the entire production sysplexes, which are running both Active-Active and non-Active-Active workloads.

Figure 8-19 shows a GDPS/Active-Active environment where you have two Active-Active workloads that are both active in Sysplex A within Region A. In addition, other workloads such as batch or processing other non-Active-Active work are running in Sysplex A. In Sysplex B (in Region B), there are the standby instances of the Active-Active workloads plus other non-Active-Active work. The data for the Active-Active workloads is software-replicated between Sysplex A and Sysplex B. Both Sysplex A and Sysplex B have local disk resilience for the entire data (data used by Active-Active applications, the system infrastructure data and data used by other, non-Active-Active applications) provided by PPRC and HyperSwap managed by GDPS/PPRC plus an out of region DR copy for the entire data provided by Global Mirror (managed by GDPS/GM). Both the PPRC and GM copies, although not shown in this diagram, are being managed using GDPS/MGM.

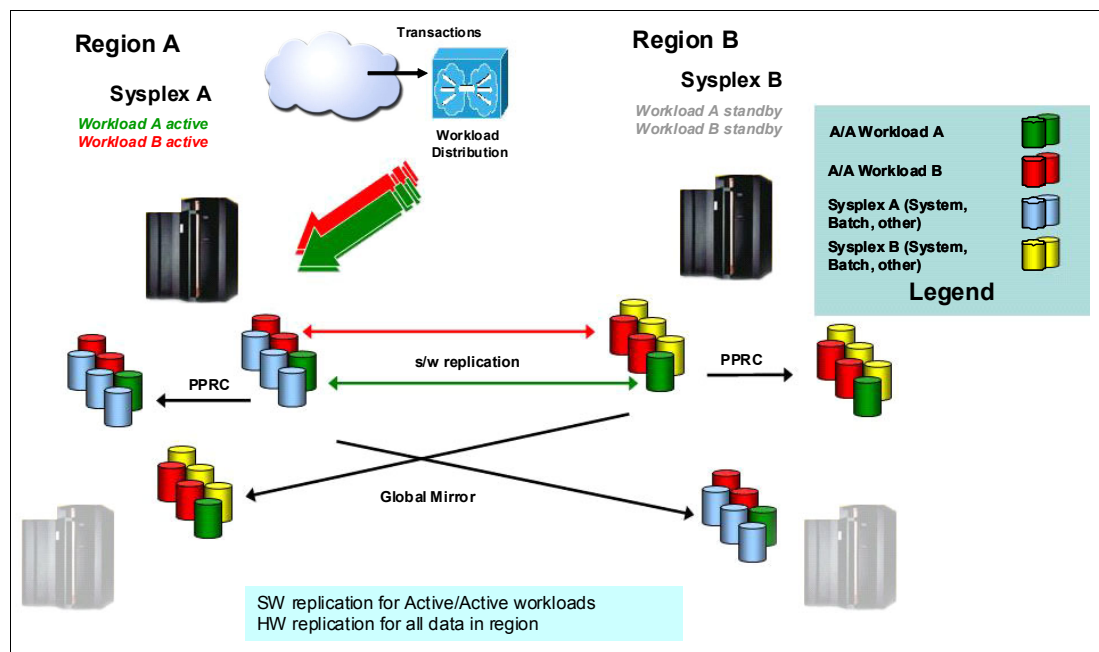


Figure 8-19 Integration of hardware and software replication

Several scenarios are supported through disk replication integration:

- ▶ Start replication.
- ▶ Prepare the DR copy (for example the GDPS/GM practice FlashCopy) for a DR test.
- ▶ Initiate a planned switch to the DR copy for a sysplex.
- ▶ Initiate disaster recovery for a sysplex.
- ▶ Return to normal configuration following a DR invocation or planned switch.

Using the GDPS/Active-Active Controller to perform GDPS/Active-Active related operations and also the hardware replication actions greatly simplifies operations because all of these actions are performed from a single point of control. Without this integration, it would be necessary to run different steps of a complex operation (such as a planned region switch of an entire sysplex) by executing different steps on multiple different controlling systems.

The disk integration function provides simple, high-level primitives whereby a single script statement coordinates performing of multiple disk replication-related operations across the GDPS/PPRC and GDPS/GM environments that comprise the GDPS/MGM configuration. For example, assume that you want to perform a planned region-switch of Sysplex A to its DR hardware replica in Region B: first, you switch the Active-Active workloads to run in Sysplex B, then you must stop all systems in Sysplex A (both of these actions can be performed using GDPS/Active-Active functions).

Next, to start Sysplex A on the hardware replica in Region B, multiple disk replication-related steps are necessary. We do not provide details of these steps here, but some of them must be performed by the GDPS/PPRC controlling system, and others need to be performed by the GDPS/GM controlling system.

With the GDPS/Active-Active disk integration capability, *all* of these disk replication steps are reduced to a single script statement (for this example, `DISKREPLICATION SWITCH SYSPLEX=SYSPLEXA PLANNED`), which is initiated from and coordinated by the GDPS/Active-Active Controller.

In summary, GDPS/Active-Active in conjunction with GDPS/MGM provides a comprehensive set of capabilities to automate out-of-region, near-continuous availability (for Active-Active workloads), and disaster recovery (for all workloads) for your production sysplexes using GDPS/Active-Active as the single point of control.

8.7 Zero Data Loss Configuration

The ZDL configuration is a specialized implementation of GDPS/Active-Active that, with GDPS/MTMM, can support zero data loss for an unplanned outage in the active site when the two GDPS/Active-Active sites are within supported Metro Mirror distances⁴.

At a high level, zero data loss is achieved by placing a secondary copy of the primary disk for the active workloads in the standby site and performing both the software replication capture and apply process in the standby site. By using this configuration, should the active site suffer an outage, the latest updates are available on disk in the standby site and therefore not lost as any such updates would be in the 'normal' or non-ZDL model⁵.

Restriction: At the time of writing, the ZDL configuration is supported only for workloads that are using DB2 data.

Also, the ZDL configuration support is provided only for workloads currently active in Site1 and standby in Site2. This configuration is known, and referred to as an asymmetric configuration as ZDL is only set up in one direction; that is from Site1 to Site2, not in the opposite direction from Site2 to Site1.

⁴ The maximum supported Metro Mirror replication distance without RPQ is 300 km.

⁵ To achieve zero data loss in site outage scenarios, a STOP policy must be in effect for PPRC mirroring and primary DASD failure events. That is, no update can be made to the primary copy of the data if it cannot be synchronously replicated to the secondary copy being used for the replication capture process.

The ZDL configuration is a priced feature of GDPS/Active-Active. It is also defined at a workload level, which allows ZDL and non-ZDL workloads to be operated in the same GDPS/Active-Active environment.

Figure 8-20 shows a high-level view of the normal software replication model (non-ZDL configuration) in GDPS/Active-Active where the capture process runs in the same site as the active workload. Given that software replication is asynchronous in nature, an unplanned loss of the active site (Site1) not only results in loss of both the workload and the capture process but is highly likely to leave so called “stranded transactions” that have not yet been sent to the standby site (Site2) to be applied. These stranded transactions could become “lost” transactions if the unplanned outage is catastrophic and the data cannot be retrieved later.

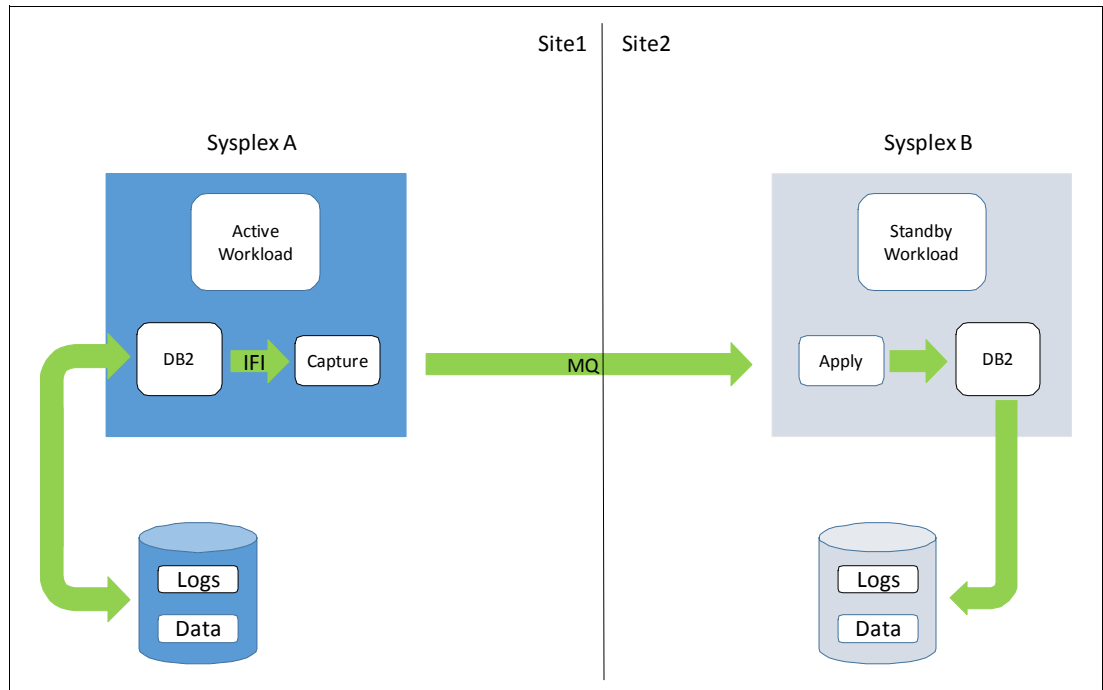


Figure 8-20 High-level view of DB2 replication for a non-ZDL environment

For workloads where such potential for data loss is unacceptable, if the two GDPS/Active-Active sites are within Metro Mirror distances, there is a solution that can deliver zero data loss. Figure 8-21 shows a high-level view of the ZDL configuration.

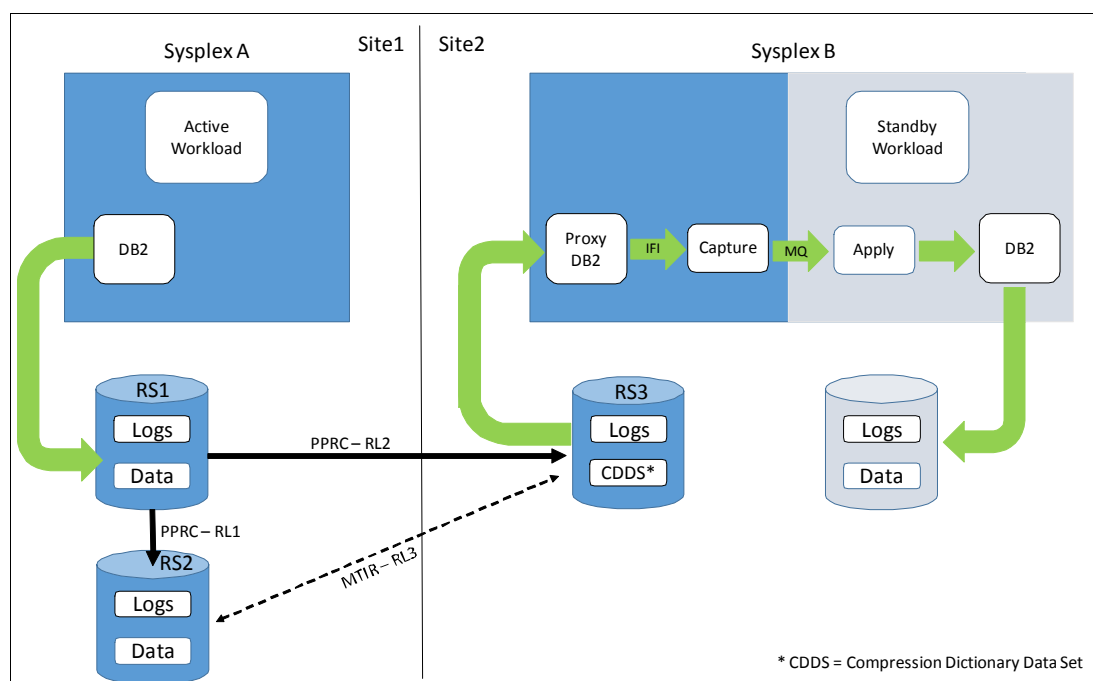


Figure 8-21 High level view of ZDL configuration

In Figure 8-21, RS3, which is located in the standby site (Site2), is a PPRC secondary copy of the primary data (RS1) from the active site (Site1). A PPRC secondary copy (RS2) of the active site data in Site1 is also present for high availability purposes (HyperSwap). RS3 is not a copy of all the data on RS1; only specific data required for the software replication capture process needs to be replicated using Metro Mirror to Site2. The DB2 logs and the DB2 compression dictionary are the subset of data required. All three copies of data are managed using the GDPS/MTMM solution.

Running in Site2 are several systems (at least two for high availability) that are driving the software replication capture process. These systems are part of the same sysplex (shown as SysplexB in Figure 8-21) as the systems running the apply process and the standby workloads. The systems running the capture process are using specialized capabilities to read from the Metro Mirror secondary volumes to access the DB2 log information required for the capture process to determine the changes that need to be sent (still over IBM MQ as in a normal DB2 replication implementation) to the apply process for writing to the standby copy of the data. Because Metro Mirror is a synchronous mirroring solution, all of the updates that have been applied to the data in the active site are available on the Metro Mirror secondary volumes and can be picked up by the capture process running in the standby site and applied to the standby copy of data, regardless of the state of the disk in the active site.

8.8 Flexible testing with GDPS/Active-Active

The best testing possible to understand if a workload is able to run in the recovery location is to actually run it there. As the saying goes, “The test of the pudding is in the tasting.” GDPS/Active-Active is well-positioned for this task because the application is already running in both locations and providing some level of confidence that the infrastructure in either site is able to sustain the workload. For complete confidence, you will also need to switch the workload so that the previously standby instance becomes active and actually processes transactions for some period of time.

Toggling a workload between sites in a GDPS/Active-Active setup can be easy. The workload can be periodically switched to run in one site or other site (in a planned manner) in a matter of seconds, with no data loss. Running the workload live in the other site with transactions being routed to it will give you the best opportunity to assess whether adjustments are needed to your infrastructure or operating procedures and ensure that both of your sites are ready to assume live workloads.

Creating workload failures to test unplanned workload switching can also be simple. However, we know that not all clients are willing to do such tests live in their production environments, no matter how small the expected service disruption might be.

A preferred practice is to maintain a sandbox testing environment that closely represents the production environment. If you already have a sandbox testing environment for Parallel Sysplex, this can be extended so you have a test GDPS/Active-Active environment by adding another sandbox sysplex in the other site and a couple of Controllers for the sandbox GDPS.

If you do not have a sandbox sysplex but have, for example, a development sysplex, this can be extended to serve as a testing environment. With such a test environment, you can test new levels of software components or maintenance to these components before you introduce such changes into production. The test GDPS/Active-Active will also let you test both planned and at least some portion of unplanned outage and switch scenarios before they are tested in production. And as previously mentioned, various scenarios might never get tested in production, in which case the testing in the test environment can still provide an indication regarding whether the solution is set up properly and can be expected to work.

8.9 GDPS/Active-Active services

As explained, GDPS/Active-Active touches on much more than simply data replication. It also touches many other aspects in your environment such as sysplex, automation, network, workload routing, workload management, testing processes, planned and unplanned outage scenario testing and so on.

Most installations do not have all these skills readily available, and it is extremely rare to find a team with this range of skills across many implementations. However, the GDPS/Active-Active offering includes just that: access to a global team of specialists in all the disciplines you need to ensure a successful GDPS implementation.

Having said that, the most successful GDPS projects are those in which IBM and client skills form a unified team to perform the implementation.

Specifically, the Services component of GDPS/Active-Active includes some or all of the following tasks:

- ▶ Planning to determine availability requirements, configuration recommendations, implementation, and testing plans
- ▶ Installation and necessary customization of:
 - NetView
 - System Automation
 - Customization for coexistence with other automation product
 - Multisite Workload Lifeline Advisor
 - Tivoli Monitoring
- ▶ Data replication implementation:
 - Bandwidth analysis
 - Installation and necessary customization of
 - InfoSphere Data Replication Server for z/OS (DB2, IMS, or VSAM)
- ▶ Setup of SASP-compliant routers and switches
- ▶ Assistance with cross-site connectivity for WAN and SE/HMC LAN
- ▶ GDPS/Active-Active automation code installation and customization
 - Training on GDPS/Active-Active setup and operations
 - Assistance with planning, coding and testing GDPS scripts and scenarios
 - Assistance with planning and implementing GDPS/Active-Active cooperation, integration with GDPS/PPRC, GDPS/MGM, or all of these
- ▶ Project management and support throughout the engagement

The services that IBM can provide in conjunction with a high availability and disaster recovery project are not restricted to those that we have listed. Here we have provided a list of services that specifically relate to GDPS/Active-Active implementation. The sizing of the services component of each project is tailored for that project based on many factors, including what automation or replication is already in place, which of the prerequisite products are already installed, and so on. This means that the services and skills provided as part of those services are tailored to the specific needs of each particular client and implementation.

8.10 GDPS/Active-Active prerequisites

See the GDPS web page to find the most current list of prerequisites for GDPS/Active-Active:

<http://www.ibm.com/systems/z/advantages/gdps/getstarted/gdpsaa.html>

8.11 GDPS/Active-Active comparison to other GDPS offerings

In each of the chapters that describe the other GDPS products that are based on hardware replication, we provide a table to compare the characteristics of these solutions against each other at a high level. We do not include GDPS/Active-Active in these comparisons because it is a somewhat “apples to oranges” comparison.

You have seen that GDPS/Active-Active is fundamentally different than the other GDPS products. It is based on software replication rather than hardware, and it is workload-level management and switch rather than system-level management and restart.

Furthermore, we have discussed how GDPS/Active-Active is not necessarily mutually exclusive with other GDPS products and how GDPS/PPRC or GDPS/MGM can be combined with GDPS/Active-Active to provide a comprehensive, robust near-continuous availability and disaster recovery solution for your enterprise.

Basic positioning and comparison for GDPS/Active-Active against the other GDPS products is described in 8.1.1, “Positioning GDPS/Active-Active” on page 232.

8.12 Summary

GDPS/Active-Active is a powerful offering that facilitates near-instantaneous switching of workloads between two sites that can be separated by virtually unlimited distances. Based on asynchronous software replication, planned switches can be accomplished with no data loss (RPO 0). When sufficient replication bandwidth is provided, the RPO can be as low as a few seconds for an unplanned workload switch.

GDPS/Active-Active provides a range of capabilities, through an intuitive web interface or using simple and powerful scripting, for workload management, workload routing, data replication management, management of system and hardware resources for planned and unplanned events. Through extensive monitoring and failure detection mechanisms, unplanned workload switches can be completely automated, removing human intervention and optimizing RPO.

For enterprises that require high levels of protection with near zero RPO and RTO at distances beyond the practical reach of a GDPS/PPRC Active/Active configuration, GDPS/Active-Active is uniquely positioned to meet these requirements for critical workloads.



GDPS Virtual Appliance

In this chapter, we provide an overview of the GDPS Virtual Appliance offering. The GDPS Virtual Appliance supports both planned and unplanned situations, which helps to maximize application availability and provide business continuity. In particular, a GDPS Virtual Appliance solution can deliver the following capabilities:

- ▶ Near-continuous availability solution
- ▶ Disaster recovery (DR) solution across metropolitan distances
- ▶ Recovery time objective (RTO) less than an hour
- ▶ Recovery point objective (RPO) of zero

The main objective of the GDPS Virtual Appliance is to provide the above mentioned capabilities to clients using z/VM and Linux on IBM z Systems and do not have z/OS in their environments¹. The virtual appliance model employed by this offering results in a solution that is easily managed and operated without requiring z/OS skills.

The functions provided by the GDPS Virtual Appliance fall into two categories: protecting your data and controlling the resources managed by GDPS. These functions include the following:

- ▶ Protecting your data:
 - Ensures the consistency of the secondary data if there is a disaster or suspected disaster, including the option to also ensure zero data loss
 - Transparent switching to the secondary disk using HyperSwap
 - Management of the remote copy configuration
- ▶ Controlling the resources managed by GDPS during normal operations, planned changes, and following a disaster:
 - Monitoring and managing the state of the production Linux for z Systems guest images and LPARs (shutdown, activating, deactivating, IPL, and automated recovery)
 - Support for switching your disk, or systems, or both, to another site
 - User-customizable scripts that control the GDPS Virtual Appliance action workflow for planned and unplanned outage scenarios

¹ For clients who run z/OS and have z/OS skills, equivalent capabilities exist by using the GDPS/PPRC Multiplatform Resiliency for z Systems described in 10.2, “GDPS/PPRC Multiplatform Resiliency for z Systems” on page 299.

9.1 Introduction to the GDPS Virtual Appliance

The GDPS Virtual Appliance is a continuous availability and disaster recovery solution that handles many types of planned and unplanned outages. As mentioned in Chapter 1, “Introduction to business resilience and the role of GDPS” on page 1, most outages are planned, and even among unplanned outages, most are not disasters. The GDPS Virtual Appliance provides capabilities to help provide the required levels of availability across these outages and in a disaster scenario. This chapter describes the data integrity and availability protection as well as the systems management capabilities provided by the GDPS Virtual Appliance.

The term *production system* is used throughout this chapter and it refers to any z/VM images together with the Linux on z Systems guests that are being managed by this instance of the GDPS Virtual Appliance.

9.2 GDPS Virtual Appliance configuration components

This section contains a high-level description of the components in a GDPS Virtual Appliance configuration. The components consist of both hardware and software. The hardware includes the disk subsystems that contain the production data and the remote copy services that perform the data replication. The software components include GDPS and other automated management code that runs on the GDPS Virtual Appliance, as well as GDPS Multiplatform Resiliency for z Systems (also known as xDR), which runs on the z/VM systems that are managed by the GDPS Virtual Appliance. Figure 9-1 shows an example of a GDPS Virtual Appliance environment.

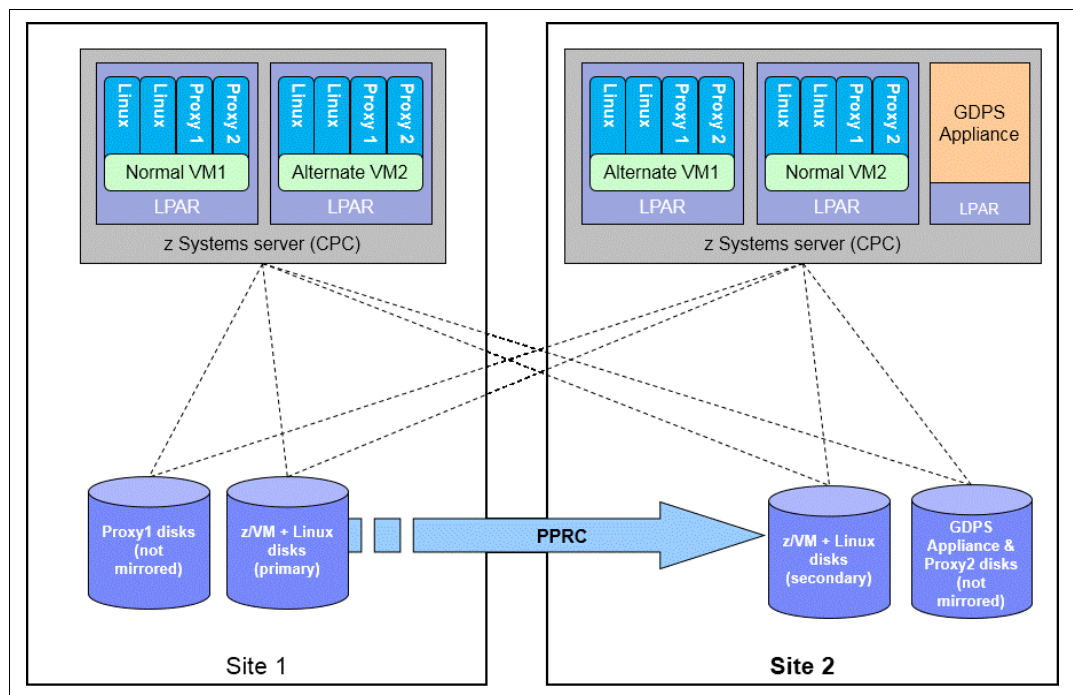


Figure 9-1 GDPS Virtual Appliance Environment

9.2.1 GDPS Virtual Appliance

The GDPS Virtual Appliance is a self-contained system which includes the GDPS/PPRC software that provides monitoring and management of PPRC replication of production disk as well as monitoring and management of the z/VM systems which are using the production disk. The GDPS Appliance allows you to initiate planned events and to perform situation analysis after an unplanned event to determine the status of the production systems or the disks, and then to drive automated recovery actions. The GDPS Virtual Appliance is responsible for carrying out all actions during a planned event or following a disaster or potential disaster; actions such as managing the disk mirroring configuration, initiating a HyperSwap, initiating a freeze and implementing the freeze/swap policy actions, re-IPLing failed systems, and so on.

A GDPS Virtual Appliance environment is typically spread across two data centers (Site1 and Site2) where the primary copy of the production disk is normally in Site1. The GDPS Appliance must have connectivity to all the Site1 and Site2 primary and secondary devices that it will manage. For availability reasons, the GDPS Virtual Appliance runs in Site2 on local disk that is not mirrored with PPRC. This provides failure isolation for the appliance system to ensure that it is not impacted by failures that affect the production systems and remains available to automate any recovery action.

9.2.2 Multiplatform Resiliency for z Systems

The GDPS Virtual Appliance provides automated management of z/VM systems with a function called “Multiplatform Resiliency for z Systems (also known as xDR)”. To provide these capabilities, the GDPS Virtual Appliance communicates and coordinates with System Automation for Multiplatforms (SA MP) running on Linux on IBM z Systems.

In each GDPS xDR-managed z/VM system you must configure two special Linux guests, which are known as the *proxy* guests as shown in Figure 9-1 on page 276. One proxy node is configured on Site1 disk and the other is configured on Site2 disk. The proxies are guests that are dedicated to providing communication and coordination with the GDPS Virtual Appliance. They must run System Automation for Multiplatforms with the separately licensed xDR feature.

The proxy guests serve as the middleman for GDPS. They communicate commands from GDPS to z/VM, monitor the z/VM environment, and communicate status information and failure information (such as a HyperSwap triggers affecting the z/VM disk) back to the GDPS Virtual Appliance. At any given time, the proxy node running on disk in the PPRC secondary site is the *Master proxy*, and this is the proxy node that the GDPS Virtual Appliance coordinates actions with. The proxy node Master role is switched automatically when PPRC disk is switched (or recovered) or when the Master proxy fails.

The disks being used by z/VM, the guest machines, and the proxy guest in this configuration must be CKD disks.

z/VM provides a HyperSwap function. With this capability, the virtual device associated with one real disk can be swapped transparently to another disk. GDPS coordinates planned and unplanned HyperSwap for z/VM disks, providing continuous data availability. For site failures, GDPS provides a coordinated Freeze for data consistency across all z/VM systems.

GDPS can perform a graceful shutdown of z/VM and its guests and perform hardware actions such as LOAD and RESET against the z/VM system's partition. GDPS supports taking a PSW restart dump of a z/VM system. Also, GDPS can manage CBU/OOCOD for IFLs and CPs on which z/VM systems are running.

9.3 Protecting data integrity and data availability with the GDPS Virtual Appliance

In 2.2, “Data consistency” on page 17, we point out that data integrity across primary and secondary volumes of data is essential to perform a database restart and accomplish an RTO of less than hour. This section provides details about how GDPS automation in the GDPS Virtual Appliance provides both data consistency if there are mirroring problems and data availability if there are disk problems.

Two types of disk problems trigger a GDPS automated reaction:

- ▶ PPRC Mirroring problems (Freeze triggers)
No problem exists writing to the primary disk subsystem, but a problem exists mirroring the data to the secondary disk subsystem. This is described in 9.3.1, “GDPS Freeze function for mirroring failures” on page 278.”
- ▶ Primary disk problems (HyperSwap triggers)
There is a problem writing to the primary disk: Either a hard failure, or the disk subsystem is not accessible or not responsive. This is described in 9.3.2, “GDPS HyperSwap function” on page 279.

9.3.1 GDPS Freeze function for mirroring failures

GDPS uses automation to stop all mirroring when a remote copy failure occurs. In particular, the GDPS automation uses the IBM PPRC Freeze/Run architecture, which is implemented as part of Metro Mirror on IBM disk subsystems and also by other enterprise disk vendors. In this way, if the disk hardware supports the Freeze/Run architecture, GDPS can ensure consistency across all data for the managed systems (consistency group) regardless of disk hardware type. This preferred approach differs from proprietary hardware approaches that work only for one type of disk hardware. For a related introduction to data consistency with synchronous disk mirroring, see “PPRC data consistency” on page 24.

When a mirroring failure occurs, this problem is classified as a Freeze trigger and GDPS stops activity across *all* disk subsystems at the time the initial failure is detected, thus ensuring that the dependant write consistency of the remote disks is maintained. This is what happens when a GDPS performs a Freeze:

- ▶ Remote copy is suspended for all device pairs in the configuration.
- ▶ While the suspend command is being processed, each device goes into a long busy state.
- ▶ No I/Os can be issued to the affected devices until the long busy state is thawed with the PPRC Run action or until it times out. The consistency group timer setting commonly defaults to 120 seconds, although for most configurations a longer or extended long busy (ELB) setting is preferred.
- ▶ All paths between the PPRCed disks are removed, preventing further I/O to the secondary disks if PPRC is accidentally restarted.

Because no I/Os are processed for a remote-copied volume during the ELB, dependent write logic ensures the consistency of the remote disks. GDPS performs a Freeze for all PPRCed devices in the GDPS managed configuration.

Important: Because of the dependent write logic, it is not necessary for all devices to be frozen at the same instant. In a large configuration with many thousands of remote copy pairs, it is not unusual to see short gaps between the times when the Freeze command is issued to each disk subsystem. Because of the ELB, however, such gaps are not a problem.

After GDPS automation performs the Freeze and the consistency of the remote disks is protected, the GDPS Virtual Appliance will perform a Run action against all LSSs. This will remove the ELB and allow production systems to continue using these devices. The devices will be in remote copy-suspended mode, meaning that any further writes to these devices are no longer being mirrored. However, changes are being tracked by the hardware so that only the changed data will be resynchronized to the secondary disks later.

If the Freeze trigger turns out to be the first sign of an actual disaster, your z/VM systems might continue operating for an amount of time before those systems actually fail. Any updates made to the primary volumes during this time will not have been replicated to the secondary disk, and are therefore lost.

The GDPS Virtual Appliance uses a combination of storage subsystem and production system triggers to automatically secure, at the first indication of a potential disaster, a data-consistent secondary site copy of your data using the Freeze function. In this way, the secondary copy of the data is preserved in a consistent state, perhaps even before production applications are aware of any issues. Ensuring the data consistency of the secondary copy ensures that a normal system restart can be performed instead of having to perform DBMS forward recovery actions. This is an essential design element of GDPS to minimize the time to recover the critical workloads if there is a disaster at the primary site.

You can appreciate why such a process must be automated. When a device suspends, there is not enough time to launch a manual investigation process.

In summary, freeze is triggered as a result of a PPRC suspension event for any primary disk in the GDPS Virtual Appliance configuration, that is, at the first sign that a duplex mirror that is going out of the duplex state. When a device suspends, all attached systems are sent a “State Change Interrupt” (SCI). A message is issued in all of those systems and then each VM system must issue multiple I/Os to investigate the reason for the suspension event.

When GDPS performs a freeze, all primary devices in the PPRC configuration suspend. This can result in significant SCI traffic and many messages in all of the systems. GDPS, in conjunction with z/VM and microcode on the DS8000 disk subsystems, supports reporting suspensions in a summary message per LSS instead of at the individual device level. When compared to reporting suspensions on a per devices basis, the Summary Event Notification for PPRC Suspends (PPRCSUM) dramatically reduces the message traffic and extraneous processing associated with PPRC suspension events and freeze processing.

9.3.2 GDPS HyperSwap function

If there is a problem writing or accessing the *primary* disk because of a failing, failed, or non-responsive primary disk, then there is a need to swap from the primary disks to the secondary disks.

The GDPS Virtual Appliance delivers a powerful function known as HyperSwap. HyperSwap provides the ability to swap from using the primary devices in a mirrored configuration to using what had been the secondary devices, transparent to the production systems and applications using these devices.

Without HyperSwap, a transparent disk swap is not possible. All systems using the primary disk would need to be shut down (or might have failed, depending on the nature and scope of the failure) and would have to be re-IPLed using the secondary disks. Disk failures are often a single point of failure for the entire production environment.

With HyperSwap, such a switch can be accomplished without IPL and with just a brief hold on application I/O. The HyperSwap function is completely controlled by automation, thus allowing all aspects of the disk configuration switch to be controlled through GDPS.

HyperSwap can be invoked in two ways:

- ▶ **Planned HyperSwap**

A planned HyperSwap is invoked by operator action using GDPS facilities. One example of a planned HyperSwap is where a HyperSwap is initiated in advance of planned disruptive maintenance to a disk subsystem.

- ▶ **Unplanned HyperSwap**

An unplanned HyperSwap is invoked automatically by GDPS, triggered by events that indicate the primary disk problem.

Primary disk problems can be detected as a direct result of an I/O operation to a specific device that fails because of a reason that indicates a primary disk problem such as:

- No paths available to the device
- Permanent error
- I/O timeout

In addition to a disk problem being detected as a result of an I/O operation, it is also possible for a primary disk subsystem to proactively report that it is experiencing an acute problem. The IBM DS8000 family has a special microcode function known as the *Storage Controller Health Message Alert* capability. Problems of different severity are reported by disk subsystems that support this capability. Those problems classified as acute are also treated as HyperSwap triggers. After systems are swapped to use the secondary disks, the disk subsystem and operating system can try to perform recovery actions on the former primary without impacting the applications using those disks.

Planned and unplanned HyperSwap have requirements in terms of the physical configuration, such as having it be symmetrically configured, and so on.

When a swap is initiated, GDPS always validates various conditions to ensure that it is safe to swap. For example, if the mirror is not fully duplex, that is, not all volume pairs are in a duplex state, a swap cannot be performed. The way that GDPS reacts to such conditions changes depending on the condition detected and whether the swap is a planned or unplanned swap.

Assuming that there are no show-stoppers and the swap proceeds, for both planned and unplanned HyperSwap, the systems that are using the primary volumes will experience a temporary pause in I/O processing. GDPS blocks I/O both at the channel subsystem level by performing a Freeze which results in all disks going into Extended Long Busy, and also in all systems, where I/O is quiesced at the operating system (UCB) level. This is to ensure that no systems use the disks until the switch is complete.

During the time when I/O is paused, the following process is completed:

1. The PPRC configuration is *physically switched*. This includes physically changing the secondary disk status to primary. Secondary disks are protected and cannot be used by applications. Changing their status to primary allows them to come online to systems and be used.

2. The disks will be *logically switched* in each of the systems in the GDPS configuration. This involves switching the internal pointers in the operating system control blocks. After the switch, the operating system will point to the former secondary devices instead of the current primary devices.
3. For planned swaps, optionally, the mirroring direction can be reversed.
4. Finally, the systems resume operation using the new, swapped-to primary devices. The applications are not aware of the fact that different devices are now being used.

This brief pause during which systems are locked out of performing I/O is known as the *User Impact Time*. In benchmark measurements at IBM using currently supported releases of GDPS and IBM DS8000 disk subsystems, the User Impact Time to swap 10,000 pairs across 16 systems during an unplanned HyperSwap was less than 10 seconds. Most implementations are actually much smaller than this and typical impact times using the most current storage and server hardware are measured in seconds. Although results will depend on your configuration, these numbers give you a high-level idea of what to expect.

The GDPS Virtual Appliance HyperSwaps all devices in the managed configuration. Just as the Freeze function applies to the entire consistency group, similarly HyperSwap is for the entire consistency group. For example, if a single mirrored volume fails and HyperSwap is invoked, processing is swapped to the secondary copy of *all* mirrored volumes for all managed systems in the configuration, including volumes in unaffected subsystems. This is because, to maintain disaster readiness, all primary volumes *must* be in the same site. If HyperSwap were to swap the only failed devices, you would then have several primaries in one site, and the remainder in the other site. This would also make for a significantly complex environment to operate and administer I/O configurations.

HyperSwap with less than full channel bandwidth

You may consider enabling unplanned HyperSwap even if you do not have sufficient cross-site channel bandwidth to sustain the full production workload for normal operations. Assuming that a disk failure is likely to cause an outage and you will need to switch to using disk in the other site, the unplanned HyperSwap might at least present you with the opportunity to perform an orderly shutdown of your systems first. Shutting down your systems cleanly avoids the complications and restart time elongation associated with a crash-restart of application subsystems.

9.3.3 GDPS use of DS8000 functions

GDPS strives to use (when it makes sense) enhancements to the IBM DS8000 disk technologies. In this section we provide information about the key DS8000 technologies that the GDPS Virtual Appliance supports and uses.

PPRC Failover/Failback support

When a primary disk failure occurs and the disks are switched to the secondary devices, PPRC Failover/Failback (FO/FB) support eliminates the need to do a full copy when reestablishing replication in the opposite direction. Because the primary and secondary volumes are often in the same state when the freeze occurred, the only differences between the volumes are the updates that occur to the secondary devices after the switch. Failover processing sets the secondary devices to primary suspended status and starts change recording for any subsequent changes made. When the mirror is reestablished with failback processing, the original primary devices become secondary devices and a resynchronization of changed tracks takes place.

The GDPS Virtual Appliance requires PPRC FO/FB capability to be available on all disk subsystems in the managed configuration.

PPRC eXtended Distance (PPRC-XD)

PPRC-XD (also known as Global Copy) is an asynchronous form of the PPRC copy technology. GDPS uses PPRC-XD rather than synchronous PPRC to reduce the performance impact of certain remote copy operations that potentially involve a large amount of data. See 3.7.2, “Reduced impact initial copy and resynchronization” on page 94 for details.

Storage Controller Health Message Alert

This facilitates triggering an unplanned HyperSwap proactively when the disk subsystem reports an acute problem that requires extended recovery time. See 9.3.2, “GDPS HyperSwap function” on page 279 for more information about unplanned HyperSwap triggers.

PPRCS Summary Event Messages

GDPS supports the DS8000 PPRC Summary Event Messages (PPRCSUM) function, which is aimed at reducing the message traffic and the processing of these messages for Freeze events. This is described in 9.3.1, “GDPS Freeze function for mirroring failures” on page 278.

Soft Fence

Soft Fence provides the capability to block access to selected devices. As discussed in 9.3.4, “Protecting secondary disks from accidental update” on page 283, GDPS uses Soft Fence to avoid write activity on disks that are exposed to accidental update in certain scenarios.

On-demand dump (also known as non-disruptive statesave)

When problems occur with disk subsystems such as those which result in an unplanned HyperSwap, a mirroring suspension or performance issues, a lack of diagnostic data from the time the event occurs can result in difficulties in identifying the root cause of the problem. Taking a full statesave can lead to temporary disruption to host I/O and is often frowned upon by clients for this reason. The on-demand dump (ODD) capability of the disk subsystem facilitates taking a non-disruptive statesave (NDSS) when such an event occurs. The microcode does this automatically for certain events, such as taking a dump of the primary disk subsystem that triggers a PPRC freeze event. It also allows an NDSS to be requested. This enables first failure data capture (FFDC) and thus ensures that diagnostic data is available to aid problem determination. Be aware that not all information that is contained in a full statesave is contained in an NDSS and therefore there may still be failure situations where a full statesave is requested by the support organization.

GDPS provides support for taking an NDSS using the GDPS GUI. In addition to this support, GDPS autonomically takes an NDSS if there is an unplanned Freeze or HyperSwap event.

9.3.4 Protecting secondary disks from accidental update

A system cannot be IPLed using a disk that is physically a PPRC secondary disk because PPRC secondary disks cannot be brought online to any systems. However, a disk can be secondary from a GDPS (and application use) perspective but physically have a simplex or primary status from a PPRC perspective.

For both planned and unplanned HyperSwap, and a disk recovery, GDPS changes former secondary disks to primary or simplex state. However, these actions do not modify the state of the former primary devices, which remain in the primary state. Therefore, the former primary devices remain accessible and usable even though they are considered to be the secondary disks from a GDPS perspective. This makes it possible to accidentally update or IPL from the wrong set of disks. Accidentally using the wrong set of disks can result in a potential data integrity or data loss problem.

The GDPS Virtual Appliance provides protection against using the wrong set of disks in different ways:

- ▶ If you attempt to load a system through GDPS (either script or panel or GUI) using the wrong set of disks, GDPS rejects the load operation.
- ▶ If you used the HMC rather than GDPS facilities for the load, then early in the IPL process, during initialization of GDPS, if GDPS detects that the system coming up has just been IPLed using the wrong set of disks, GDPS will quiesce that system, preventing any data integrity problems that could be experienced had the applications been started.
- ▶ GDPS uses a DS8000 disk subsystem capability, which is called Soft Fence for configurations where the disks support this function. Soft Fence provides the means to *fence*, which means block access to a selected device. GDPS uses Soft Fence when appropriate to fence devices that would otherwise be exposed to accidental update.

9.4 Managing the GDPS environment

You have seen how the GDPS Virtual Appliance can protect your data during unplanned outages. However, as discussed in Chapter 1, “Introduction to business resilience and the role of GDPS” on page 1, the majority of z Systems outages are not disasters. Most are planned outages, with a small percentage of unplanned ones.

In this section, we describe other aspects of the GDPS Virtual Appliance, that is, its ability to monitor and manage the resources in its environment.

9.4.1 GDPS graphic user interface

The user interface that is used for managing the GDPS Virtual Appliance environment is known as the *GDPS graphic user interface* or *GDPS GUI*. Figure 9-2 shows the GDPS GUI home page.



Figure 9-2 GDPS GUI home page

As you can see, there are four distinct areas of the page:

1. Page Header. This area allows you to initiate certain GDPS actions on demand. These actions including actions such as these:
 - Executing a GDPS monitoring process. Additional information about GDPS monitors can be found in “GDPS monitoring and alerting” on page 290.
 - Temporarily disabling or reenabling HyperSwap.
2. Navigation menu. This area contains icon links to the other panels that are available. Clicking an icon link opens a new tab in the Main Window and displays the corresponding panel in the new tab. Examples:
 - Dashboard. This panel is described in “Dashboard panel” on page 285.
 - Standard Actions. This panel is described in “Standard Actions panel” on page 285.
 - Planned Actions. This panel is described in “Planned Actions Panel” on page 286.
 - SDF Alerts. This panel is described in “SDF Panel” on page 287.
3. Main Window. This area is a tabbed workspace where the various GDPS panels are displayed. The Dashboard panel is displayed by default and is described in “Dashboard panel” on page 285. Other tabs will be added to this area as additional panels are displayed. Inactive/hidden tabs can be brought to the foreground by clicking the associated tab.
4. Status Summary. This area contains a graphical summary of the status and health of the GDPS managed environment, including the HyperSwap status, the disk mirroring status, the number of alerts of each severity currently displayed on the appliance, and the number of outstanding operator replies currently displayed on the appliance.

Dashboard panel

The Dashboard panel is the anchor content for the main window. This panel tab is always available to be made active. It shows at a glance the status of the components in your GDPS environment. Figure 9-2 on page 284 shows an example of the Dashboard panel. It includes icons that can be selected for the processors and disk in both Site1 and Site2. It also graphically shows the current direction and the status of the PPRC mirror, plus the percentage of volume pairs that are in duplex state.

Clicking on the arrow indicating the status and direction of the mirror opens the LSS Pairs panel. This panel is described in “LSS Pairs panel” on page 288. Clicking on Site1 or Site2 processor icon opens the Standard Actions panel. This panel is described in Standard Actions panel, below.

Standard Actions panel

GDPS provides facilities to help manage many common system-related actions. There are two reasons to use the GDPS facilities to perform these *Standard Actions*:

- ▶ They are well tested and based on IBM recommended procedures.
- ▶ Using the GDPS interface lets GDPS know that the changes that it is seeing are planned changes, and therefore GDPS is not to react to these events.

Standard actions are performed using the Standard Actions panel, which is shown in Figure 9-3. This panel is displayed by clicking one of the processor icons displayed on the Dashboard panel (as described in “Dashboard panel” on page 285), or by clicking the Standard Actions icon on the navigation menu (as described in 9.4.1, “GDPS graphic user interface” on page 284).

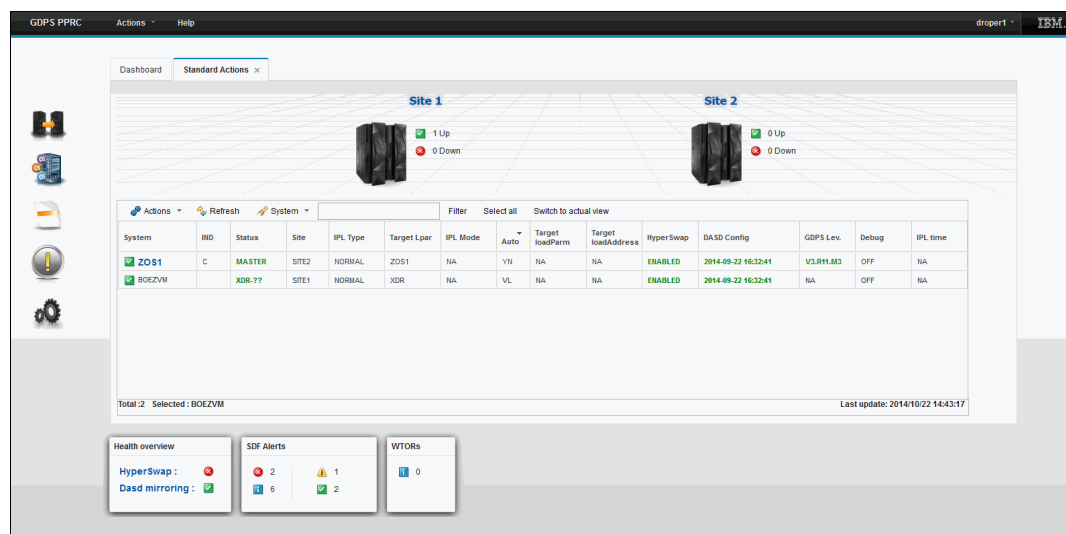


Figure 9-3 GDPS GUI Standard Actions panel

The panel displays a list of all of the systems defined to GDPS. The upper portion of the panel contains site icons with a summary count of the number of systems up and down in each Site.

Above the system list header is a toolbar that allows you to perform actions such as stopping, loading, and resetting systems and activating and deactivating LPARs.

If you double-click a z/VM system in the list presented, another panel opens where you can operate at the *cluster* or Linux on IBM z Systems *guest* level within that z/VM image.

Planned Actions Panel

The Planned Actions panel shown in Figure 9-4 is displayed by clicking the Planned Actions icon on the Navigation menu as described in 9.4.1, “GDPS graphic user interface” on page 284.

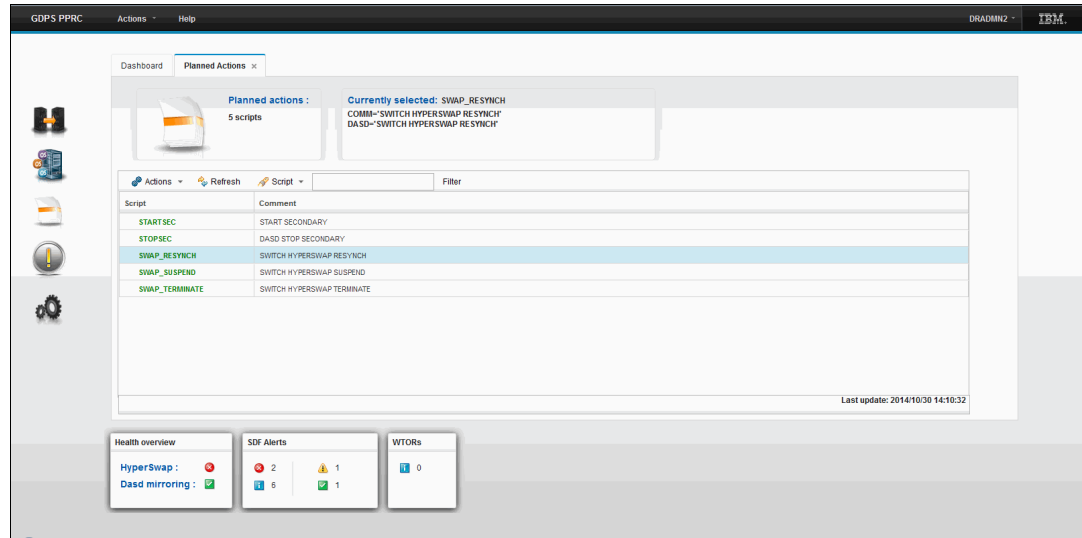


Figure 9-4 GDPS GUI Planned Actions panel

The panel displays a list of all of the *Control scripts* that have been defined to GDPS. A Control script is simply a procedure recognized by GDPS that pulls together one or more GDPS functions. Control scripts allow you to perform complex, multistep operations without having to execute each step individually, using various panel options. See section 9.4.2, “GDPS scripts” on page 289, for more information on Control scripts.

The upper portion of the Planned Actions panel contains a display box which contains the statements defined for any script that is selected. A script can be executed by double-clicking the script.

SDF Panel

The SDF panel is the main panel for monitoring the status of GDPS managed resources. You navigate to this panel by clicking on the SDF alert icon displayed on the Navigation menu as described in 9.4.1, “GDPS graphic user interface” on page 284. An example of the SDF panel is shown in Figure 9-5.

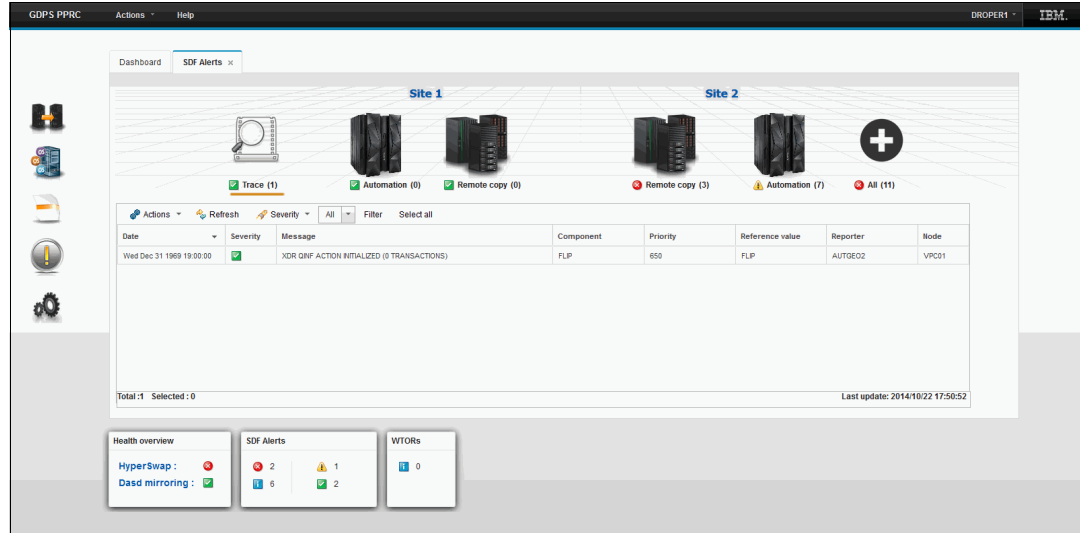


Figure 9-5 GDPS GUI SDF panel

The panel is divided horizontally into two sections. The upper section contains clickable icons for filtering the SDF entry list displayed in the lower section based on the type of alert. The filtering icon labels indicate how many alerts of that type and location in parenthesis.

Any SDF alerts that pass the applied filtering will be displayed in the SDF entry list at the bottom of the panel.

Above the entry list header is a toolbar that allows you to delete alerts, display help associated with alerts, and so on.

Remote Copy management panels

To manage the remote copy environment using the GDPS Virtual Appliance, you first define your entire remote copy configuration, including your primary and secondary LSSs, your primary and secondary devices, and your PPRC links, to GDPS in a file called the GEOPARM file. This enables GDPS to provide you with the capability to perform actions against all of the devices/pairs in your environment with a single action, rather than having to execute an action against each device/pair.

This section describes panel options provided by GDPS to enable you to manage your Remote Copy environment.

LSS Pairs panel

The initial panel for Remote Copy management is the LSS pairs panel. You navigate to this panel by clicking on the mirroring status and direction arrow that is displayed on the Dashboard panel as described in “Dashboard panel” on page 285. An example of the LSS Pairs panel is shown in Figure 9-6.

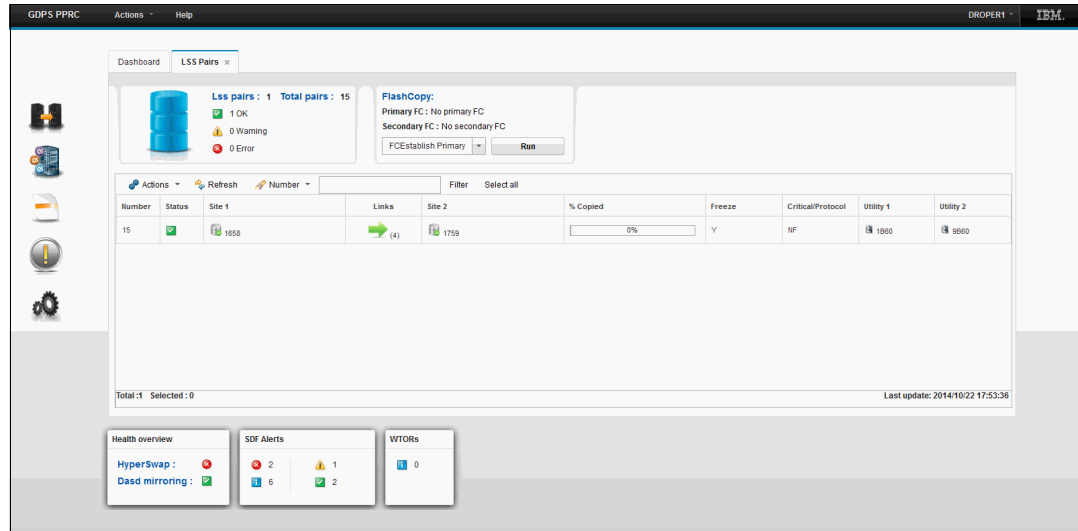


Figure 9-6 GDPS GUI LSS Pairs panel

The panel displays a list of all of the LSS pairs defined in the GDPS mirror. The upper left corner contains a summary count of the total number of LSS pairs and the number of LSS pairs by status severity. Double-clicking an LSS pair opens the Pairs panel for the LSS pair, as described in “Pairs panel” on page 289.

Above the LSS pair list header is a toolbar that enables you to perform various functions against all of the volume pairs in the selected LSS pairs. Examples of the functions you can perform using the toolbar include querying the status of the pairs, suspending mirroring for the pairs, restarting mirroring for the pairs, and recovering the secondary devices for the pairs.

Pairs panel

The Pair panel provides the ability to perform Remote Copy management at the volume pair level, rather than at the LSS pair level. An example of the Pairs panel is shown in Figure 9-7.

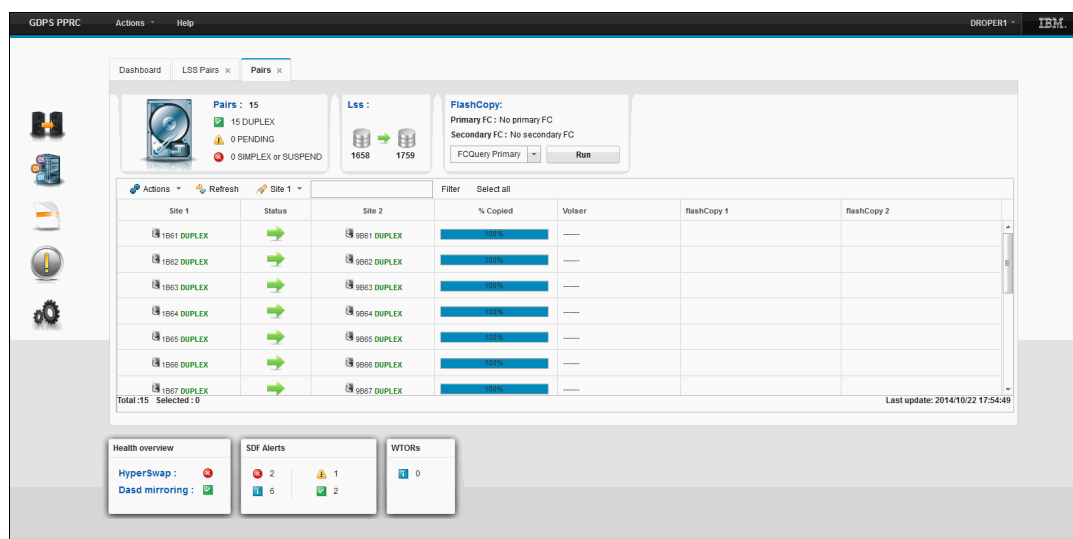


Figure 9-7 GDPS GUI Pairs panel

The panel displays a list of all of the volume pairs defined in the selected LSS. The upper left corner contains a summary count of the total number of volume pairs and the number of volume pairs by status severity. Double-clicking on a volume pair will issue a query for the pair and display the resulting output in a dialog box.

Above the volume pair list header is a toolbar that provides you the ability to perform various functions against all of the selected volume pairs. Examples include querying the status of the pairs, suspending mirroring for the pairs, restarting mirroring for the pairs, and recovering the secondary devices for the pairs.

9.4.2 GDPS scripts

As previously mentioned, GDPS provides the ability for you to automate complex, multistep planned operations against your Remote Copy environment and against the production systems in your environment through the use of Control scripts.

Again, a script is a procedure recognized by GDPS that pulls together one or more GDPS functions. When executing a script, GDPS performs the first statement in the list, checks the result, and only if it is successful, proceeds to the next statement. If you perform the same steps manually, you would have to check results, which can be time-consuming, and then initiate the next action. With scripts, the process is automated.

Scripts are powerful because they can access the full capability of GDPS. The ability to invoke GDPS functions through a script provides the following benefits:

- Speed

The script will execute the requested actions and check results at machine speeds. Unlike a human, it does not need to search for the latest procedures or the commands manual.

- Consistency

With automation, your procedures will execute in exactly the same way, time after time.

- Thoroughly tested procedures

Because they behave in a consistent manner, you can test your procedures over and over until you are sure they do everything that you want, in exactly the manner that you want. Also, because you need to code everything and cannot assume a level of knowledge (as you might with instructions intended for a human), you are forced to thoroughly think out every aspect of the action the script is intended to undertake. And because of the repeatability and ease of use of the scripts, they lend themselves more easily to frequent testing than manual procedures.

9.4.3 System Management actions

Most of the GDPS Standard Actions require actions to be done on the HMC. The interface between the GDPS Virtual Appliance and the HMC is through a facility called the BCP Internal Interface (BCPii), and this allows GDPS to communicate directly with the hardware for automation of HMC actions such as LOAD, RESET, Activate LPAR, and Deactivate LPAR. GDPS can also perform ACTIVATE (power-on reset), CBU ACTIVATE/UNDO, and OOCOD ACTIVATE/UNDO.

Furthermore, when you LOAD a z/VM system using GDPS (panels or scripts), GDPS will listen for certain² operator prompts from the system being IPLed and reply to the prompts. This support for replying to these IPL-time prompts automatically, helps to remove reliance on operator skills and eliminating operator error for any messages that require replies.

SYSRES Management

Today many clients maintain multiple alternate z/VM SYSRES devices (also known as *IPLSETs*) as part of their maintenance methodology. GDPS provides special support to allow clients to identify IPLSETs. This removes the requirement for clients to manage and maintain their own procedures when IPLing a system on a different alternate SYSRES device.

GDPS can automatically update the IPL pointers after any disk switch or disk recovery action that changes the GDPS primary site indicator for PPRC disks. This removes the requirement for clients to perform additional script actions to switch IPL pointers after disk switches, and greatly simplifies operations for managing alternate SYSRES “sets.”

9.5 GDPS monitoring and alerting

The GDPS SDF panel, discussed in “SDF Panel” on page 287, is where GDPS dynamically displays color-coded alerts.

Alerts can be posted as a result of an unsolicited error situation that GDPS listens for. For example, if one of the multiple PPRC links that provide the path over which PPRC operations take place is broken, there is an unsolicited error message issued.

GDPS listens for this condition and will raise an alert on the SDF panel, notifying the operator of the fact that a PPRC link is not operational. Clients run with multiple PPRC links and if one is broken, PPRC continues over any remaining links.

² Only operator prompts that can be safely replied to in a consistent manner are candidates for automatic replies.

However, it is important for operations to be aware that a link is broken and fix this situation because a reduced number of links results in reduced PPRC bandwidth and reduced redundancy. If this problem is not fixed in a timely manner and more links have a failure, it can result in production impact because of insufficient mirroring bandwidth or total loss of PPRC connectivity (which results in a freeze).

Alerts can also be posted as a result of GDPS periodically monitoring key resources and indicators that relate to the GDPS Virtual Appliance environment. If any of these monitoring items are found to be in a state deemed to be not normal by GDPS, an alert is posted which can be viewed using the GDPS GUI on the appliance system.

When an alert is posted, the operator will have to investigate (or escalate, as appropriate) and corrective action will need to be taken for the reported problem as soon as possible. After the problem is corrected, this is detected during the next monitoring cycle and the alert is cleared by GDPS automatically.

The GDPS Virtual Appliance monitoring and alerting capability is intended to ensure that operations are notified of and can take corrective action for any problems in their environment that can affect the ability of the appliance to do recovery operations. This maximizes the chance of achieving your availability and RPO and RTO commitments.

9.6 Services component

As you have learned, GDPS touches on much more than simply remote copy. It also includes automation, database management and recovery, testing processes, disaster recovery processes, and other areas.

Most installations do not have skills in all these areas readily available. It is extremely rare to find a team that has this range of skills across many implementations. However, the GDPS Virtual Appliance offering includes exactly that: Access to a global team of specialists in all the disciplines you need to ensure a successful GDPS Virtual Appliance implementation.

Specifically, the Services component includes several or all of the following services:

- ▶ Planning to determine availability requirements, configuration recommendations, and implementation and testing plans
- ▶ Remote copy implementation
- ▶ GDPS Virtual Appliance installation and policy customization
- ▶ Assistance in defining Recovery Point and Recovery Time objectives
- ▶ Education and training on the GDPS Virtual Appliance setup and operations
- ▶ Onsite implementation assistance
- ▶ Project management and support throughout the engagement

The sizing of the Services component of each project is tailored for that project, based on many factors including what automation is already in place, whether remote copy is already in place, the cross-site connectivity in place, and so on. This means that the skills provided are tailored to the specific needs of each particular implementation.

9.7 GDPS Virtual Appliance prerequisites

For more information about GDPS Virtual Appliance prerequisites, see this website:

<http://www.ibm.com/systems/z/advantages/gdps/getstarted/gdpspprc.html>

9.8 GDPS Virtual Appliance compared to other GDPS offerings

So many features and functions are available in the various members of the GDPS family that recalling them all and remembering which offerings support them is sometimes difficult. To position the offerings, Table 9-1 lists the key features and functions and indicates which ones are delivered by the various GDPS offerings.

Table 9-1 Supported features matrix

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/MTMM	GDPS Virtual Appliance	GDPS/XRC	GDPS/GM
Continuous availability	Yes	Yes	Yes	Yes	No	No
Disaster recovery	Yes	Yes	Yes	Yes	Yes	Yes
CA/DR protection against multiple failures	No	No	Yes	No	No	No
Continuous Availability for foreign z/OS systems	Yes with z/OS Proxy	No	No	No	No	No
Supported distance	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	Virtually unlimited	Virtually unlimited
Zero Suspend FlashCopy support	Yes, using CONSISTENT	Yes, using CONSISTENT for secondary only	Yes, using CONSISTENT	No	Yes, using Zero Suspend FlashCopy	Yes, using CGPause
Reduced impact initial copy/resync	Yes	Yes	Yes	Yes	Not applicable	Not applicable
Tape replication support	Yes	No	No	No	No	No
Production sysplex automation	Yes	No	Yes	Not applicable	No	No
Span of control	Both sites	Both sites (disk only)	Both sites	Both sites	Recovery site	Disk at both sites; recovery site (CBU or LPARs)
GDPS scripting	Yes	No	Yes	Yes	Yes	Yes

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/MTMM	GDPS Virtual Appliance	GDPS/XRC	GDPS/GM
Monitoring, alerting and health checks	Yes	Yes	Yes	Yes (except health checks)	Yes	Yes
Query Services	Yes	Yes	No	No	Yes	Yes
MSS support for added scalability	Yes (secondary in MSS1)	Yes (secondary in MSS1)	Yes (H2 in MSS1, H3 in MSS2)	No	No	Yes (GM FC and Primary for MGM in MSS1)
MGM 3-site and 4-site	Yes (all configurations)	Yes (3-site only and non-IR only)	Yes (4-site only)	No	Not applicable	Yes (all configurations)
MzGM	Yes	Yes	Yes (non-IR only)	No	Yes	Not applicable
Open LUN	Yes	Yes	No	No	No	Yes
z/OS equivalent function for Linux for IBM z Systems	Yes	No	Yes (Linux for IBM z Systems running as a z/VM guest only)	Yes (Linux for IBM z Systems running as a z/VM guest only)	Yes	Yes
Heterogeneous support through DCM	Yes (VCS and SA AppMan)	No	No	No	Yes (VCS only)	Yes (VCS and SA AppMan)
z/BX hardware management	Yes	No	No	No	No	No
GDPS GUI	Yes	Yes	No	Yes	No	Yes

9.9 Summary

The GDPS Virtual Appliance is a powerful offering that provides disaster recovery, continuous availability, and system resource management capabilities for z/VM and Linux on z Systems. GDPS Appliance is the only GDPS offering that is packaged in a virtual appliance, eliminating the necessity for z/OS and sysplex skills in order to manage and operate the solution.

HyperSwap, available with the GDPS Virtual Appliance, provides the ability to transparently swap disks between two sites. The power of automation allows you to test and perfect the actions to be taken, either for planned or unplanned changes, thus minimizing or eliminating the risk of human error.

This is one of the offerings in the GDPS family, along with GDPS/PPRC, GDPS/HM and GDPS/MTMM, that offers the potential of zero data loss, and that can achieve the shortest recovery time objective, typically less than one hour following a complete site failure.

It is also one of the only members of the GDPS family, again along with GDPS/PPRC and GDPS/MTMM, that is based on hardware replication and that provides the capability to manage the production LPARs. Although GDPS/XRC and GDPS/GM offer LPAR management, their scope for system management is limited, and only includes the systems in the recovery site, and not the production systems running in Site1.

In addition to the disaster recovery and planned reconfiguration capabilities, the GDPS Virtual Appliance also provides a user-friendly interface for monitoring and managing the various elements of the GDPS configuration.



GDPS extensions for heterogeneous systems and data

Most enterprises today have a heterogeneous IT environment where the applications and data reside on a variety of hardware and software platforms, such as IBM z Systems, IBM System p, UNIX, Windows, and Linux. Such an environment can benefit greatly if there is a single point of control to manage the data across all the platforms, and for the disaster recovery solution to coordinate the recovery across multiple platforms.

In this chapter we describe the various GDPS extensions that are available for clients to manage data and coordinate disaster recovery across multiple platforms. The various extensions are available in one or more of the GDPS offerings. The following extensions are described in this chapter:

- ▶ Open LUN Management function
Available for GDPS/PPRC, GDPS/PPRC HyperSwap Manager, and GDPS/Global Mirror.
- ▶ GDPS/PPRC Multiplatform Resiliency for z Systems (also known as xDR)
Available for GDPS/PPRC.

Note that the GDPS Virtual Appliance offering is also based on xDR technology. In this chapter we discuss the GDPS/PPRC implementation of xDR. The xDR technology as applicable to the GDPS Virtual Appliance is described in Chapter 9, “GDPS Virtual Appliance” on page 275.
- ▶ Distributed Cluster Management support for Veritas Cluster Servers (VCS)
Available for GDPS/PPRC, GDPS/XRC, and GDPS/GM.
- ▶ Distributed Cluster Management support for Tivoli System Automation Application Manager (SA AppMan)
Available for GDPS/PPRC and GDPS/GM.

10.1 Open LUN Management function

As discussed in 3.1.3, “Protecting distributed (FB) data” on page 66, many enterprises today run applications that update data across multiple platforms. For these enterprises, there is a need to manage and protect not just the z Systems data but also the data residing on servers that are not z Systems servers. GDPS/PPRC, GDPS/PPRC HyperSwap Manager, and GDPS/Global Mirror have been extended to manage a heterogeneous environment of z Systems and distributed systems data by providing a function called Open LUN Management.

The Open LUN Management function allows GDPS to be a single point of control to manage business resiliency across multiple tiers in the infrastructure, thus improving cross-platform system management and business processes. If installations share their disk subsystems between the z Systems and distributed systems platforms, GDPS/PPRC and GDPS/Global Mirror can manage the Metro Mirror or Global Mirror remote copy configurations and the FlashCopy for distributed systems storage.

Open LUN support extends the GDPS/PPRC Freeze capability to Open LUN (FB) devices that are in supported disk subsystems to provide data consistency for the z Systems data and the data on Open LUNs. If you are using the GDPS/PPRC function known as xDR (described in 10.2, “GDPS/PPRC Multiplatform Resiliency for z Systems” on page 299), GDPS also supports native Linux for z Systems running on SCSI-attached Fixed Block (FB) disks. In addition to providing the Freeze capability for FB disks, GDPS supports planned and unplanned HyperSwap for the FB disks used by xDR-managed native Linux systems.

In an Open LUN configuration, you can select one of the following two options to specify how Freeze and HyperSwap processing are to be handled for Open LUN (FB disks) and z Systems (CKD disks), when mirroring or primary disk problems are detected:

- ▶ You can select to Freeze all devices managed by GDPS.
If this option is used, both the CKD and FB devices are in a single consistency group. Any Freeze trigger, either for the z Systems or Open LUN devices, will result in both the Open LUN and the z Systems LSSs managed by GDPS being frozen. This option allows you to have consistent data across heterogeneous platforms in the case of a disaster, thus allowing you to restart systems in the site where secondary disks are located. This option is especially appropriate when there are distributed units of work on z Systems and distributed servers that update the same data, for example using IBM DB2 DRDA, the IBM Distributed Relational Database Architecture™.
- ▶ You can select to Freeze devices by group.
If this option is selected, the CKD and xDR controlled FB devices are in a separate consistency group from the non-xDR FB devices. The Freeze will be performed only on the group for which the Freeze trigger was received. If the Freeze trigger occurs for a z Systems disk device, only the CKD and xDR controlled FB devices will be frozen. If the trigger occurs for a non-xDR controlled FB disks, only those disks will be frozen.

Note: Regardless of the option you select, CKD and xDR FB disks are always in the same consistency group; they are always frozen and swapped together. The freeze and swap policy options that you select are applied to the CKD and xDR FB disks together.

The Global Mirror remote copy technology, described in 2.4.3, “Global Mirror” on page 32, inherently provides data consistency for both z Systems and distributed systems data.

Open LUN (FB disk) management prerequisites

GDPS requires the disk subsystems containing the FB devices to support specific architectural features. These architectural features are supported by all IBM disk subsystems:

- The ability to manage FB devices through a CKD utility device

GDPS runs on z/OS and can therefore communicate directly over a channel connection to CKD devices only. To communicate commands to the FB LSS and devices, the PPRC architecture allows using a CKD utility device in the same disk subsystem as a go-between to send commands and to monitor and control the mirroring of the FB devices. GDPS will need at least one CKD utility device in each hardware cluster of the storage subsystem where FB devices are located. A sample example of this function called Open LUN Management is shown in Figure 10-1 on page 298.

- The ability to send SNMP traps to report certain errors

The FB LSS and devices must communicate certain error conditions back to GDPS (for example, suspension of an FB device pair in GDPS/PPRC or GDPS/PPRC HM or an abnormal state of a Global Mirror session in GDPS/GM). This status is communicated to the z/OS host that is running the GDPS controlling system through an IP connection using SNMP traps. GDPS captures these traps and drives autonomic action such as performing a freeze for a mirroring failure.

A sample Open LUN GDPS/PPRC configuration is shown in Figure 10-1. Not shown in this picture are the IP connections from the attached disks to the z/OS host where GDPS is running.

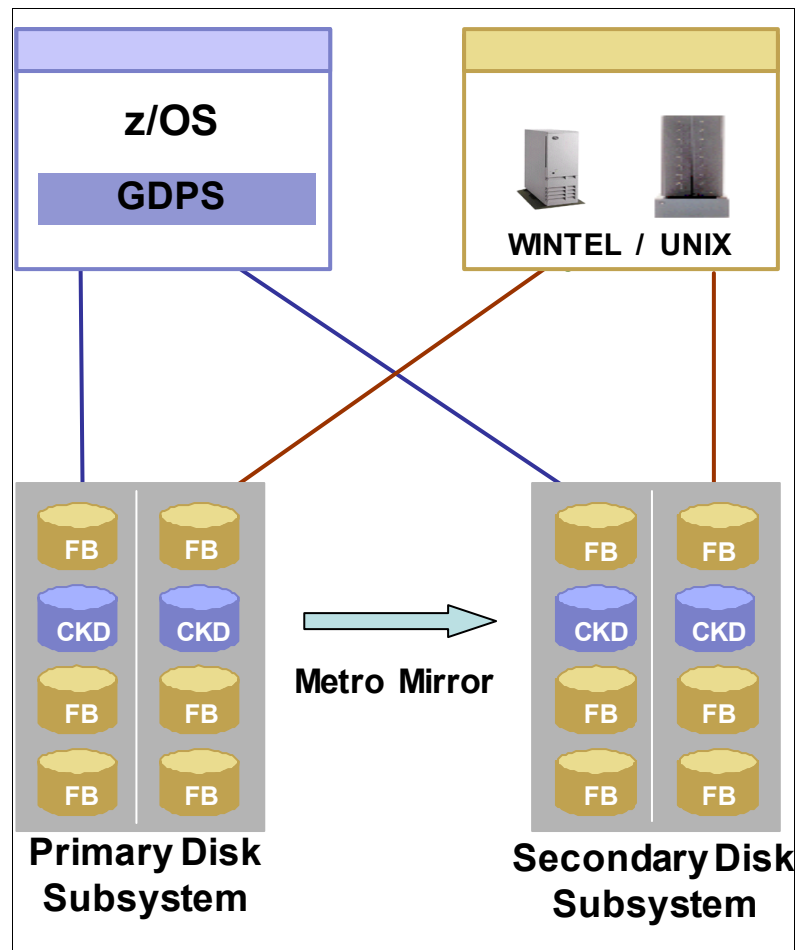


Figure 10-1 Open LUN support

For more information about managing replication and the systems and recovery of the systems that use FB disks, see 10.3, “Distributed Cluster Management” on page 307.

Also, GDPS/PPRC (or GDPS/PPRC HM) HyperSwap involves changing information at the control block level within the operating system on any system using the disks being swapped. GDPS/PPRC supports HyperSwap for FB devices used by native Linux on System z systems managed by xDR. See 10.2.2, “Native Linux on z Systems” on page 303. Otherwise, HyperSwap for FB disks is not supported.

10.2 GDPS/PPRC Multiplatform Resiliency for z Systems

Note: For the remainder of this section, Linux on z Systems might also be referred to as just *Linux*. The terms are used interchangeably.

As described in 3.3.1, “Multiplatform Resiliency for z Systems (also known as xDR)” on page 75, when clients implement a multitiered architecture, with application servers that run on Linux on z Systems and database servers that run on z/OS, there is a need to provide a coordinated near-continuous availability and disaster recovery solution for both z/OS and Linux. GDPS/PPRC provides this capability with a function called “Multiplatform Resiliency for z Systems (also known as xDR).” To provide these capabilities, GDPS/PPRC communicates and coordinates with System Automation for Multiplatforms (SA MP) running on Linux on z Systems.

Linux can run in either of the following ways:

- ▶ As a guest under z/VM
- ▶ Native in a z Systems partition

From a GDPS perspective, this is not an either/or proposition. The same instance of GDPS/PPRC can manage one or more z/VM systems *and* one or more native Linux on z Systems in addition to z/OS systems.

It is not mandatory to have any z/OS production systems managed by GDPS. The only z/OS systems that are mandatory are the GDPS controlling systems. Using one or preferably two z/OS systems, you can manage a production environment that consists only of z/VM, native Linux on z Systems, or both.

Most xDR functions can be performed only by a GDPS controlling system. We strongly suggest that GDPS/PPRC environments managing xDR systems be configured with two controlling systems.

10.2.1 Guest Linux under z/VM

In a GDPS xDR-managed z/VM system you must configure a special Linux guest, which is known as the *proxy* guest. The proxy is a guest that is dedicated to providing communication and coordination with the GDPS/PPRC controlling system. It must run System Automation for Multiplatforms with the separately licensed xDR feature.

The proxy guest serves as the middleware for GDPS. It communicates commands from GDPS to z/VM, monitors the z/VM environment, and communicates status information and failure information, such as a HyperSwap trigger affecting z/VM disk back to the GDPS/PPRC controlling system. GDPS/PPRC uses SA MP to pass commands to z/VM and Linux guests.

GDPS xDR supports definition of two proxy nodes for each z/VM host; one proxy node running on Site1 disk and the other running on Site2 disk. This support extends the two-controlling system model to the xDR proxy nodes, so it provides a high-availability proxy design. At any particular time, the proxy node running on disk in the PPRC secondary site is the Master proxy, and this is the proxy node that the GDPS Master controlling system coordinates actions with. Similar to the controlling system Master role, the proxy node Master role is switched automatically when PPRC disk is switched (or recovered) or when the Master proxy fails.

The disks being used by z/VM, the guest machines, and the proxy guest in this configuration must be CKD disks. GDPS xDR support allows for the definition of the PPRC secondary devices configured to z/VM and the guest machines to be defined in an alternate subchannel set¹. This can simplify definitions and provide high scalability for your disk configuration. For additional information, see “Addressing z/OS device limits in GDPS/PPRC and GDPS/MTMM environments” on page 25.

Note: Most xDR functions, including HyperSwap, benefit non-Linux guests of z/VM also. In fact, having no “production” Linux guests at all is possible. The only requirement for Linux guests is for the xDR proxy nodes, which must be dedicated Linux guests.

Be aware, however, that a z/VM host running z/OS guests is not supported by xDR.

z/VM provides a HyperSwap function. With this capability, the virtual device associated with one real disk can be swapped transparently to another disk. GDPS/PPRC coordinates planned and unplanned HyperSwap for both z/OS and z/VM disks, providing continuous data availability spanning the multitiered application. It does not matter whether the first disk failure is detected for a z/VM disk or a z/OS disk; all are swapped together.

For site failures, GDPS/PPRC provides a coordinated Freeze for data consistency across z/VM and z/OS. Again, it does not matter whether the first freeze trigger is captured on a z/OS disk or a z/VM disk; all will be frozen together.

System and hardware management capabilities similar to those available for z/OS systems are also available for z/VM systems. GDPS xDR can perform a graceful shutdown of z/VM and its guests and perform hardware actions, such as LOAD and RESET against the z/VM system’s partition. GDPS supports taking a stand-alone dump of a z/VM system and, in the event of a HyperSwap, it automatically switches the pointers of the z/VM dump volumes to the swapped to site. GDPS can manage CBU/OOCOD for IFLs and CPs on which z/VM systems are running.

¹ Only alternate subchannel set 1 (MSS1) is currently supported for defining the PPRC secondary devices.

Figure 10-2 shows an example of a configuration where several Linux nodes are running as guests under z/VM. One of the Linux guests is the proxy. The non-proxy SAP Linux guests are shown as also running SA MP. This is not mandatory. If you do run SA MP in the production Linux guest systems, GDPS provides additional capabilities for such guests. The figure also illustrates the actions taken if a disk failure is detected and HyperSwap is invoked by GDPS/PPRC.

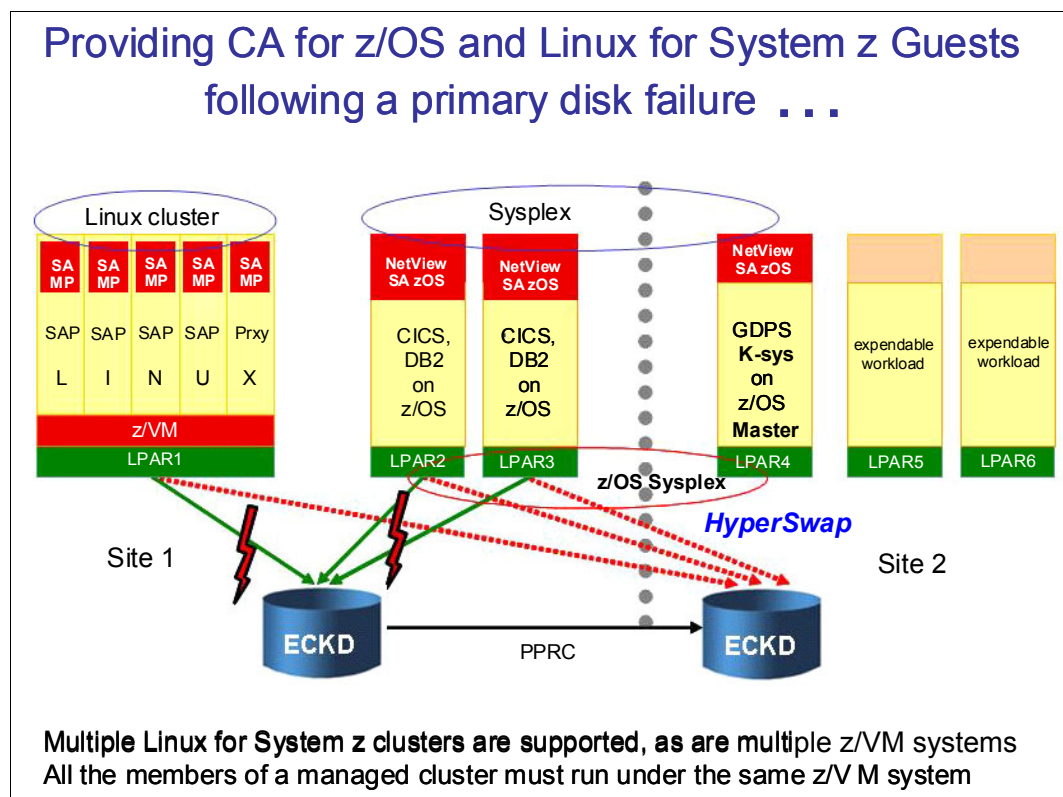


Figure 10-2 HyperSwap example: Linux running as a guest under z/VM

GDPS controlled shutdown of z/VM

Graceful shutdown of a z/VM system involves multiple virtual servers. This is a rather complex process and GDPS has special automation to control this shutdown. The GDPS automated process occurs in multiple phases:

- During the first phase, all the SA MP clusters with all the nodes for these clusters are stopped. The master proxy is the only guest running SA MP that is not stopped. When all clusters and nodes running SA MP have been successfully stopped, GDPS proceeds to the next phase.
- During the second phase all remaining guests that are capable of processing the *shutdown signal* are stopped.
- In phase three, the master proxy server and z/VM are shut down.

When an xDR-managed z/VM system is shut down using the GDPS Stop Standard Action (or equivalent script statement), all xDR-managed guests are stopped in parallel. GDPS provides the ability to control the sequence in which you stop guest systems during a z/VM shutdown.

GDPS/PPRC xDR support for z/VM Single System Image clustering

z/VM introduced Single System Image (SSI) clustering whereby four z/VM systems can be clustered to provide more effective resource sharing and other capabilities.

GDPS xDR supports z/VM systems that are members of an SSI cluster. GDPS will be aware of the fact that a z/VM system is a member of an SSI. This allows GDPS to perform certain system control actions for these z/VM systems correctly, observing SSI rules.

Linux guests can be transparently moved from one z/VM system in an SSI cluster to another, that is, without requiring the guest to be stopped. This capability, which is called *Live Guest Relocation*, provides continuous availability for Linux guests of z/VM in planned outage situations. If a z/VM system is going to be shut down, for disruptive software maintenance for example, the relocatable Linux guests can first be moved to other z/VM systems in the cluster to avoid an outage to these guests. Similarly, for an entire site shutdown, the guests under all z/VM systems in the site to be shut down can first be moved to z/VM systems in the other site.

GDPS provides support for performing Live Guest Relocation for xDR-managed z/VM systems. GDPS provides a relocation test capability that tries to assess whether a particular relocation action is likely to be successful or not. For example, the target z/VM system might not have sufficient resources to host the guest to be moved. Such a test function is quite useful because it will allow you to rectify potential problems before they are encountered. GDPS management for CPs and IFLs using On/Off Capacity on Demand is complementary to this function. You can use GDPS to first increase IFL capacity on the target CEC before performing the actual move.

GDPS/PPRC xDR support for z/VSE guests of z/VM

GDPS provides specific support for z/VSE guest systems. GDPS monitoring of z/VSE guests requires z/VSE 5.1 with the GDPS Connector (also known as the GDPS Client) enabled for GDPS monitoring. z/VSE guests of xDR-managed z/VM systems can be enabled for special GDPS xDR monitoring and management:

- ▶ GDPS can detect the failure of a z/VSE guest and automatically restart it.
- ▶ z/VSE guests can be gracefully shut down as part of the graceful shutdown of the hosting z/VM system initiated by GDPS.

Disk and LSS sharing

GDPS xDR supports the sharing of a logical disk control unit (LSS) by multiple z/VM systems. This facilitates the efficient sharing of resources, provides configuration flexibility, and simplifies the setup that would be required to keep the LSSs separate. It also enables xDR environments to use the z/VM Cross System Extension (CSE) capability.

For example, suppose you have more than one z/VM system and want to do these tasks:

- ▶ Share the IBM RACF database across your systems.
- ▶ Manage one VM Directory for all the systems.
- ▶ Ensure that a minidisk is linked only RW on one guest on one system, and have all the systems enforce that.
- ▶ Share the z/VM System Residence volumes.

z/VM CSE can do all that for you. The z/VM CSE enables you to treat separate VM systems as a single system image, thereby lowering your system management workload and providing higher availability. See *z/VM CP Planning and Administration*, SC24-6083, for CSE details.

If you want to share LSSs and disks, consider this information:

- In one LSS, you may place disks for as many xDR-managed z/VM systems as you want.
- If you want, any z/VM disk managed by GDPS can be shared by multiple xDR-managed z/VM systems. This requires that you also implement z/VM CSE.

Serialization for disk is supported through the Reserve/Release mechanism for minidisks under z/VM control.

In addition to various z/VMs sharing an LSS, having z/OS and z/VM disks in the same LSS is also possible. Here, the LSS capacity is being split between z/OS and z/VM. Any individual disk must not be shared by z/VM and z/OS systems.

10.2.2 Native Linux on z Systems

In this configuration, Linux runs natively in its own partition on z Systems. System Automation for Multiplatform, along with the separately licensed xDR feature, must be running in each xDR-managed Linux system. SA MP on each system monitors that system and reports status information or disk errors encountered by that system to the GDPS/PPRC controlling system. The controlling system communicates commands for a managed Linux system through SA MP on that system.

The disks being used by native Linux in this configuration can be either CKD or SCSI-attached Fixed Block (FB) disks. This support builds on the existing Open LUN management capability that provides PPRC and Freeze for FB disks. The xDR support also provides planned and unplanned HyperSwap for the FB disks that are used by xDR-managed native Linux systems. Any given Linux system must be running with either all CKD or all FB disks. A system running with a mixture of CKD and FB disks is not supported.

Similar to the xDR-managed z/VM systems, GDPS provides coordinated HyperSwap and coordinated freeze across z/OS and native Linux systems. Regardless of whether a swap trigger is first detected on z/OS or Linux, GDPS coordinates a swap for the entire GDPS/PPRC configuration for these systems. The same is true for freeze. The entire configuration for z/OS and native Linux is frozen together, providing consistency across these environments. If z/VM systems are also included in the GDPS managed configuration, freeze and swap are coordinated across all three environments.

Similar system and hardware management capabilities available for z/OS systems are also available for native Linux systems. GDPS xDR can perform a graceful shutdown of a native Linux system, perform hardware actions including LOAD, RESET, ACTIVATE, and DEACTIVATE against the Linux system's partition, manage CBU/OOCod for IFLs, and so on.

Figure 10-3 shows an example of a configuration where several Linux on z Systems nodes are running natively in their own partitions and all of them are under GDPS xDR control. The Linux systems running in LPAR1 and LPAR2 are using CKD disks. The Linux systems running in LPAR8 and LPAR9 are using SCSI-attached FB disks.

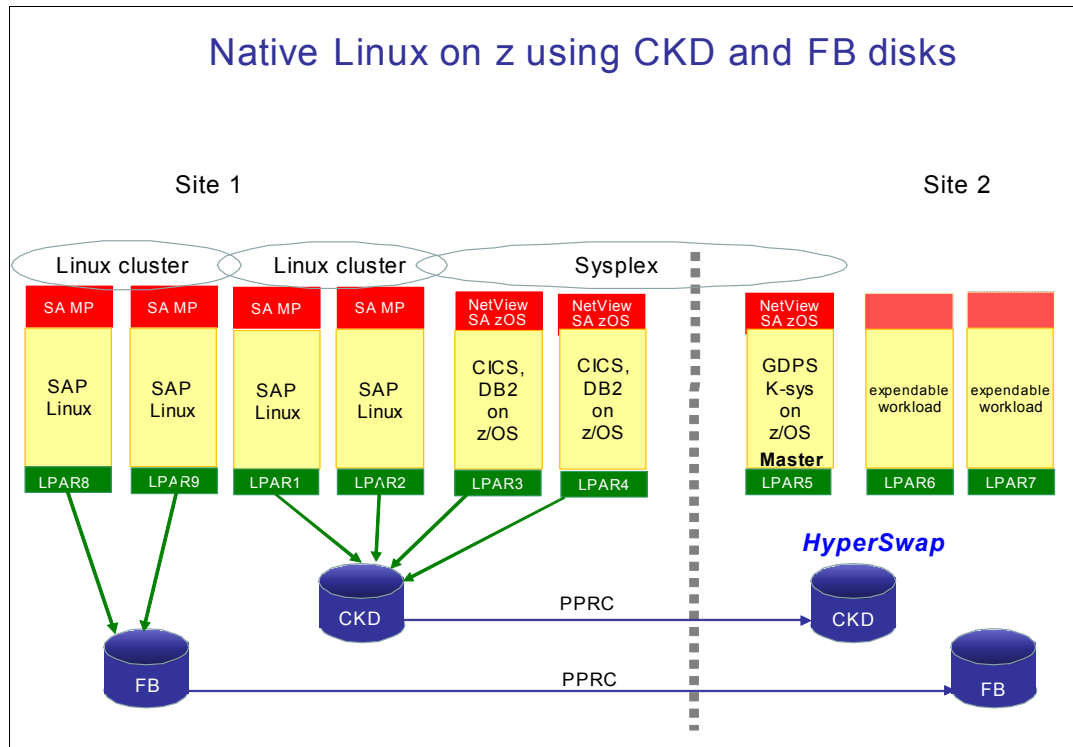


Figure 10-3 Native Linux on z Systems: LPARs using CKD and FB disks

In this configuration, when a primary disk problem is detected for either a CKD or xDR FB disk and the environment is enabled for HyperSwap, when the trigger occurs, a HyperSwap is performed for all of the CKD disks and xDR FB disks. Figure 10-4 illustrates the actions taken in such a configuration. Even though the disk failure was associated with the CKD disks, both the CKD and FB disks are swapped to the secondary copy of the disks.

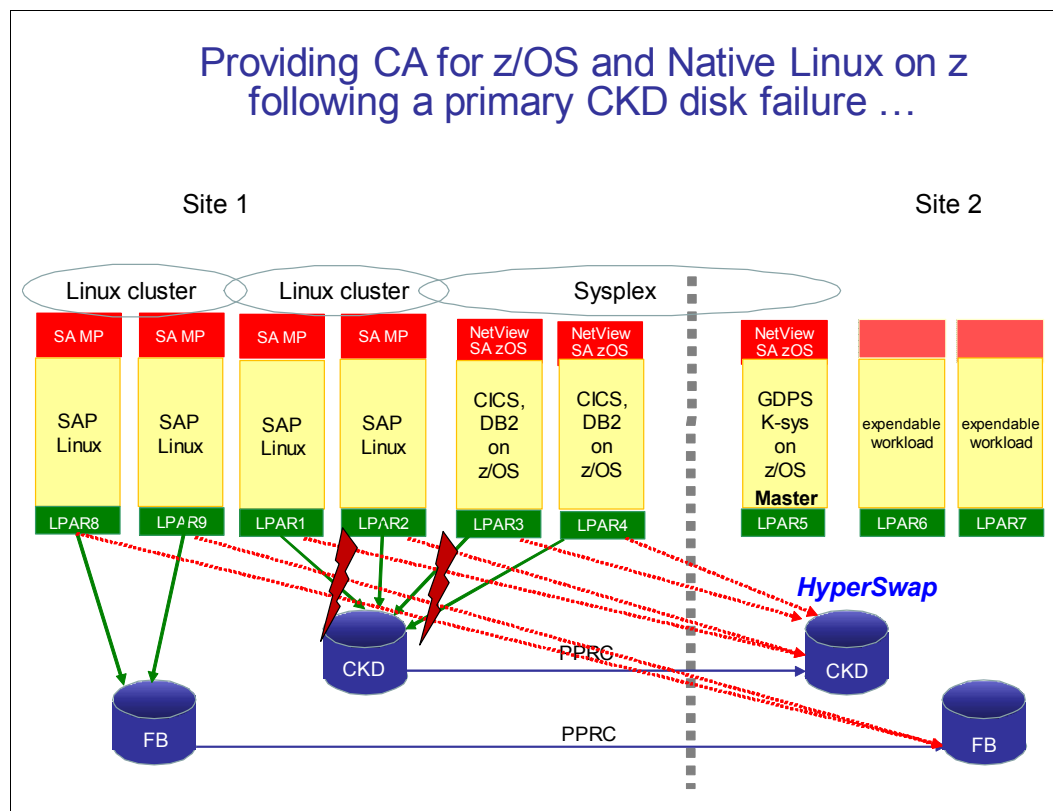


Figure 10-4 HyperSwap example following a CKD disk failure

10.2.3 Support for two GDPS Controlling systems

Originally, xDR supported only one GDPS Controlling system (also referred to as the GDPS Master K-sys), as illustrated in Figure 10-2 on page 301 and Figure 10-4. xDR functions were able to be processed only by the single GDPS Master K-sys. In the event of a planned or unplanned outage of the GDPS Master K-sys, the current Master function switched to a production system but xDR processing was interrupted, because production systems cannot perform xDR functions.

If your SA MP xDR environment is configured to support two GDPS Controlling systems, xDR processing is protected in the event of a planned or unplanned outage of the Controlling system that is the current Master. This is because the alternate Controlling system will take over the current Master responsibility and the alternate Controlling system is able to perform xDR functions. Also, in the event of an autonomic Master switch as a result of a disk swap, xDR functions are protected because the alternate Master is a Controlling system and can manage xDR resources.

Figure 10-5 on page 306 shows an xDR configuration with two GDPS Controlling systems following a HyperSwap of the primary disks from Site1 to Site2. The Master K-sys has been moved to K2-sys in Site1. xDR functions can still be performed by K2-sys; for example, a subsequent disk failure in Site2.

During cluster initialization, the proxy and non-proxy nodes send their initialization signal to both GDPS Controlling systems. Only the GDPS system that is the current Master will respond to the initialization signal, and this is how the Linux nodes will know which of the Controlling systems is the current Master. Certain events (such as heartbeating and communication of an I/O error) will be sent to the current Master; certain other events (such as initialization) are communicated to both Controlling systems.

In the event of a Master K-sys switch, GDPS will inform the Linux nodes of the switch and the Linux nodes then resume relevant communications with the new Master K-sys.

We suggest that you run GDPS with two Controlling systems and enable xDR to support two Controlling systems.

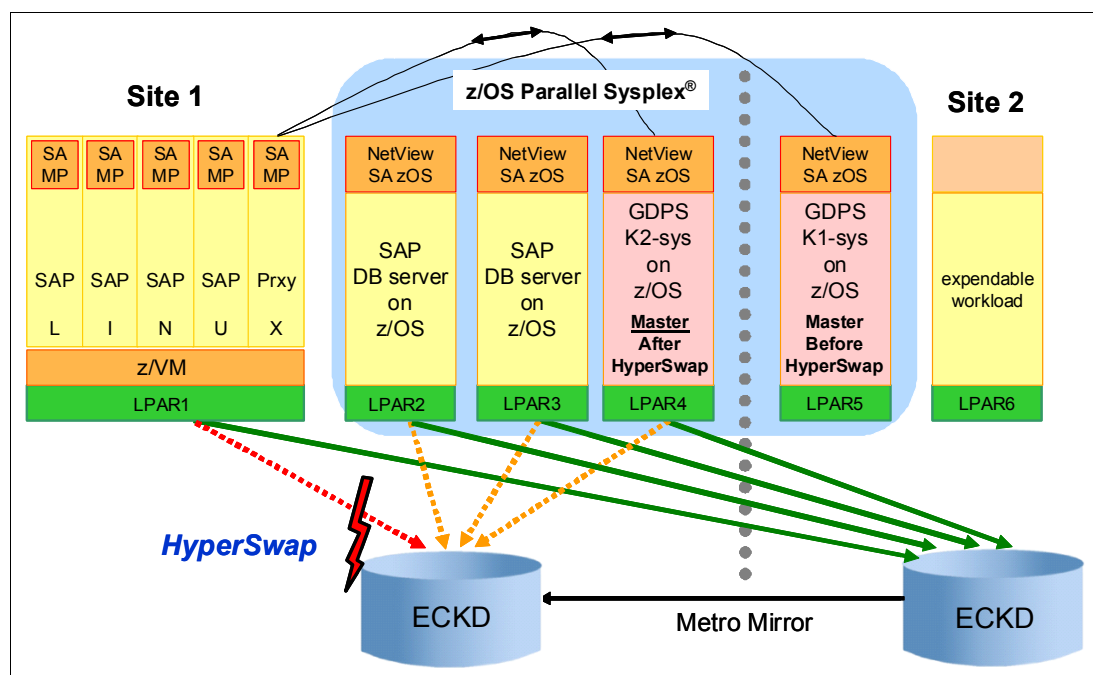


Figure 10-5 xDR configuration with two Controlling systems after a HyperSwap

10.2.4 Customization Verification Program

The xDR Customization Verification Program (CVP) verifies that installation and customization activities have been done correctly for both xDR native and guest Linux on z Systems environments. This helps identify any issues with the customization of the environment where many components exist with quite specific setup and customization requirements. It also helps identify aspects of the xDR customization that do not adhere to preferred practices recommendations.

CVP is an operator-initiated program that can be used after initial setup, and periodically thereafter, to ensure that changes to the environment have not broken the xDR setup. Two separate programs are provided: One to run on the controlling systems and another to run on the Linux server to ensure that both ends of the implementation are verified.

10.2.5 xDR Extended Monitoring

The GDPS HyperSwap Monitor provides checking for z/OS systems to ascertain whether the z/OS systems managed by GDPS meet required conditions. Any system that does not meet the required conditions is marked as “not HyperSwap-ready.” A planned HyperSwap is not allowed to execute unless all systems are HyperSwap-ready. If an unplanned swap is triggered, systems that are not HyperSwap-ready are reset and the swap is performed with the participation of only those systems that are HyperSwap-ready.

GDPS also performs similar HyperSwap monitoring for xDR systems (native and z/VM guest environments). Several environmental conditions required for HyperSwap for xDR systems are checked and if an xDR system does not meet one or more environmental conditions, GDPS attempts to autonomically fix the detected issue. If it is not possible to autonomically fix the issue, alerts will be raised.

Also, any such xDR system that does not meet all environmental conditions that are monitored are marked as “not HyperSwap-ready.” Raising alerts during monitoring allows an installation to act on the alert and to fix the reported problems in a timely manner to avoid having the system reset if an unplanned swap is triggered.

10.3 Distributed Cluster Management

Distributed Cluster Management (DCM) allows the management and coordination of planned and unplanned outages across non z Systems distributed clusters in coordination with the z Systems workloads that GDPS is responsible for.

As discussed in 10.1, “Open LUN Management function” on page 296 and 10.2, “GDPS/PPRC Multiplatform Resiliency for z Systems” on page 299, many enterprises have requirements to provide automated failover and rapid recovery of business-critical applications on z Systems and on other platforms, such as UNIX, IBM AIX, Microsoft Windows, and Linux. In addition, when you have a multitiered architecture, there is a need to provide a coordinated near-continuous availability and disaster recovery solution for applications that might be residing on servers that are not z Systems servers and those residing on z Systems.

In addition to Open LUN management and the Multiplatform Resiliency functions, GDPS/PPRC, GDPS/XRC, and GDPS/GM also include DCM. The DCM support is provided in GDPS/PPRC for both Symantec Veritas Cluster Server (VCS) clusters and IBM Tivoli System Automation Application Manager (SA AppMan) and both these distributed cluster servers can be managed concurrently by a single GDPS/PPRC. For GDPS/XRC and GDPS/GM, the DCM support is available only for VCS clusters.

DCM provides advisory and coordination functions between GDPS and distributed servers managed by VCS or SA AppMan. See “DCM functions (VCS)” on page 315 and “DCM functions (SA AppMan)” on page 324 for more details.

10.3.1 Distributed Cluster Management terminology

This section presents terminology and a brief description for each DCM term that is common to both the DCM support for VCS and SA AppMan. For terminology that is applicable only to the DCM support provided by VCS, see “VCS terminology” on page 309. For terminology applicable only to the DCM support for SA AppMan, see “SA AppMan terminology” on page 319.

Distributed Cluster Management (DCM)

This is the GDPS capability to manage and coordinate disaster recovery across distributed servers that are clustered using high availability clustering solutions alongside the z Systems workloads that GDPS is responsible for.

Application site This is a site in which the applications (both distributed applications and z Systems applications) normally reside. This site is also referred to as Site1 by GDPS with DCM.

Recovery site This is a site into which the applications that normally reside in the application site are recovered (unplanned) or moved (planned). This site is referred to as Site2 in GDPS with DCM. It is also where the GDPS/PPRC controlling system is typically located, where the GDPS/GM R-sys runs, and where the DCM agents on the distributed systems typically run.

Cluster This is a group of servers and other resources that act like a single system and enable high availability and, in some cases, load balancing and parallel processing.

K-sys This is the GDPS controlling system.

R-sys This is the GDPS Remote controlling system in a GDPS/GM configuration. It is in the Recovery site.

Geographically Dispersed Open Clusters (GDOC)

This is an IBM services offering to help clients plan for and implement Veritas Global Clusters (VCS) or System Automation Application Manager (SA AppMan) to provide high availability and disaster recovery for distributed server workloads.

If you do not already have a VCS GCO or SA AppMan implementation, consider combining GDOC services and GDPS services to engage IBM in assisting you with the integrated, end-to-end implementation of VCS or SA AppMan and GDPS with DCM.

10.3.2 DCM support for VCS

This section describes how the functions available with GDPS/PPRC (see Chapter 3, “GDPS/PPRC” on page 53), GDPS/XRC (see Chapter 5, “GDPS/XRC” on page 137), and GDPS/GM (see Chapter 6, “GDPS/Global Mirror” on page 163) have been integrated with functions provided by the Symantec cross-platform clustering solution Veritas Cluster Server (VCS).

Note: In the context of this section, the subject high-availability clusters are Veritas Cluster Server clusters. However, the DCM technology is extensible to other high-availability clustering solutions.

VCS terminology

This section presents terminology and provides a brief description for each term that is applicable to the DCM support for VCS:

Veritas Symantec delivers a suite of products under the Veritas brand.

Veritas Cluster Server (VCS)

This refers to a high availability and disaster recovery solution for cluster configurations. VCS monitors systems and application services, and restarts services when hardware or software fails.

Global Cluster Option (GCO)

This refers to functions included in the Veritas Cluster Server HA/DR bundle. The Global Cluster Option for VCS enables a collection of VCS clusters to work together to provide wide area disaster recovery.

Global cluster This term denotes the pair of VCS clusters that are linked together using VCS GCO. (In this section, you might also see this referred to as Veritas Global Cluster.)

GDPS agent This term refers to the logic residing on the VCS cluster that communicates global cluster status and events to GDPS, and accepts commands on behalf of the cluster from GDPS DCM code for VCS resources that are managed by GDPS. There is one GDPS agent per VCS global cluster, normally running in Site2. We also refer to this as the DCM agent.

Service group This term is the name that is used to denote an application running on the global cluster.

Symantec Cluster Server overview

Symantec Cluster Server powered by Veritas is a clustering solution that can be used to reduce the RTO for both planned and unplanned events. VCS can gracefully shut down applications and restart them on an available server. The failover can be to a local server in the same site or, for disaster recovery, the failover can be to a remote cluster located several thousand miles away.

VCS supports multiple operating systems, such as IBM AIX, Oracle Solaris, HP-UX, and Linux. It also supports multiple hardware, software, and database replication technologies. For more information, see the Symantec Cluster Server web page:

http://www.symantec.com/business/products/overview.jsp?pcid=2247&pvid=20_1

Figure 10-6 on page 310 shows examples of different VCS configurations:

- ▶ The high-availability clustering (LAN) configuration shown on the left side of the figure is an example of a high-availability cluster without any mirroring. It does not have disaster recovery capabilities.
- ▶ The two configurations in the middle of the figure show high-availability clusters using synchronous data replication within a metropolitan area network (MAN), with two separate clusters: A production cluster and a failover (backup) cluster. The Global Cluster Option (GCO) provides the failover control to the backup cluster. As shown in these examples, you can use either remote mirroring or replication technologies.
- ▶ The high-availability clusters with extended distance disaster recovery (WAN) configuration on the right side of the figure is the same as the metropolitan area examples, except that it has an extended distance environment on which you use an asynchronous data replication technology across an unlimited distance. Again, GCO is required.

The configurations shown in the red and blue circles in Figure 10-6 are the ones that have been integrated with GDPS. We will describe in more detail the integration provided by GDPS/PPRC for VCS clusters using GCO (examples shown with red circles) and the integration provided by GDPS/XRC and GDPS/GM for VCS clusters across an extended distance (example shown with blue circle).

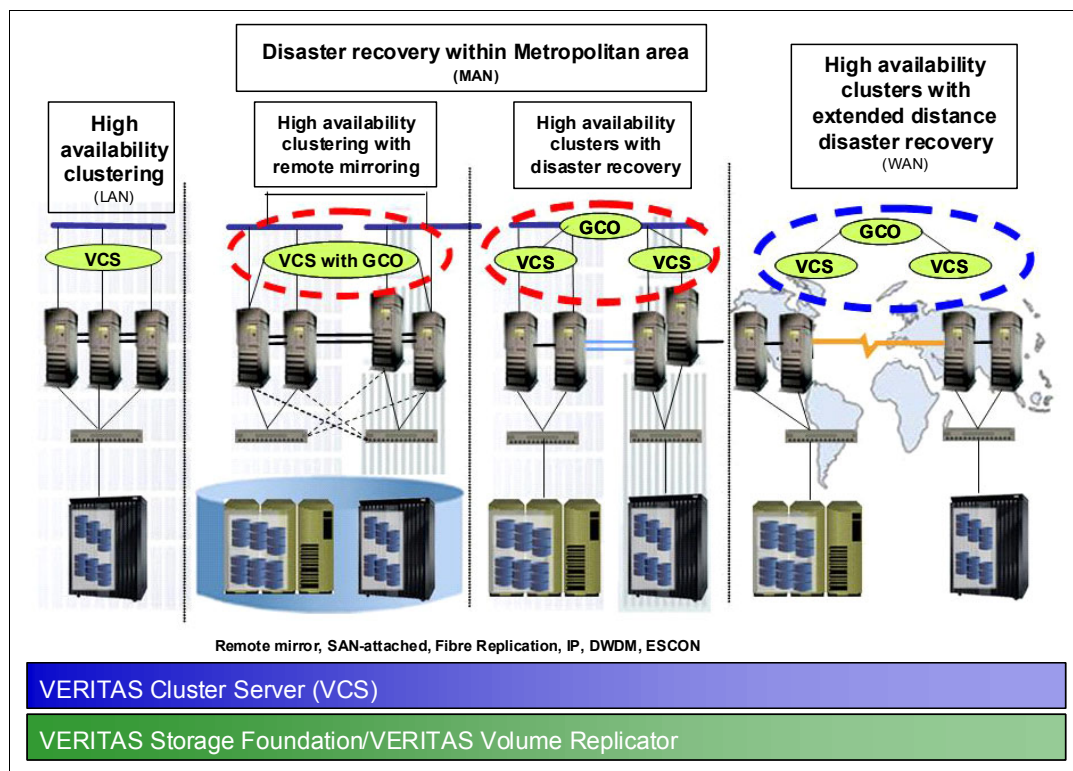


Figure 10-6 Veritas Cluster Server configurations

Integrated configuration of GDPS/PPRC and VCS clusters

Figure 10-7 on page 311 illustrates the various components of a GDPS/PPRC configuration integrated with a VCS configuration that has the GCO function implemented. The GDPS/PPRC configuration consists of a multisite Parallel Sysplex with a set of primary disks in Site1 being mirrored to a set of secondary disks in Site2 by using Metro Mirror. The disk mirroring is managed by GDPS/PPRC as described in Chapter 3, "GDPS/PPRC" on page 53. There is a minimum of one GDPS K-sys in Site2, and optionally there can be a second K-sys in Site1.

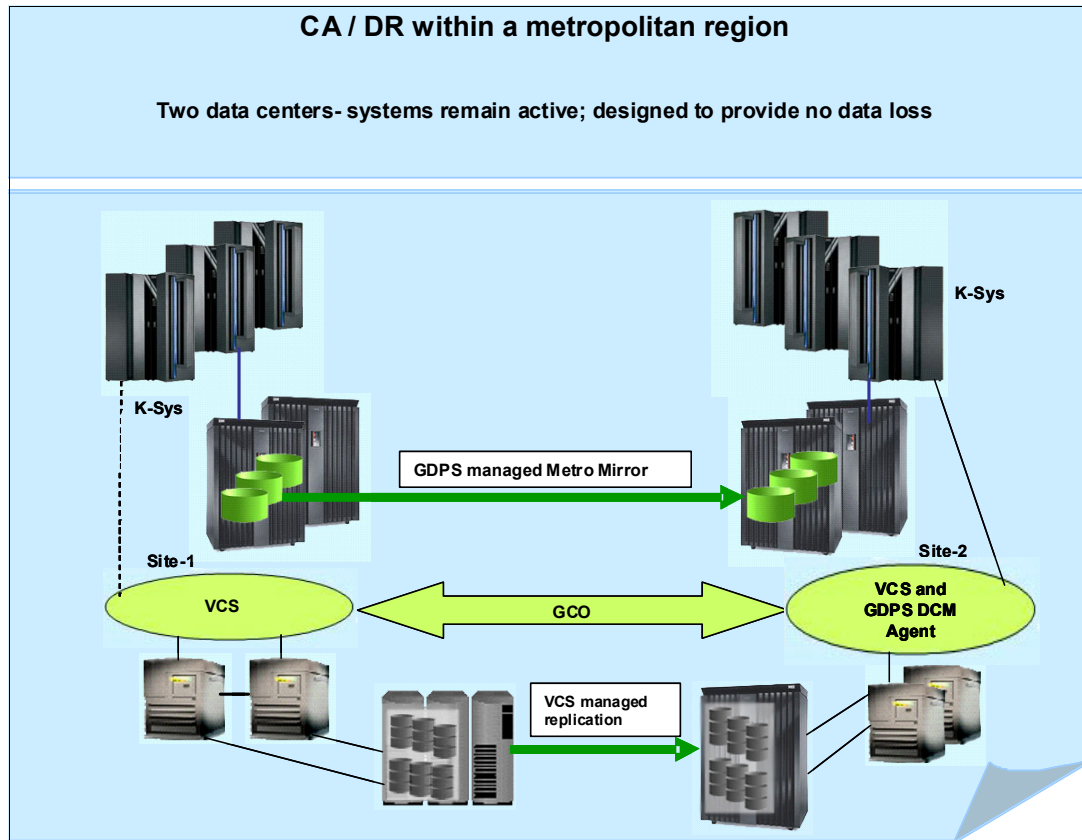


Figure 10-7 GDPS/PPRC integration with VCS using GCO: metropolitan distance

The VCS configuration in this example is composed of two clusters: the production cluster in Site1, and the failover cluster in Site2. This is also referred to as the *Active/Standby* configuration.

The GCO option in each cluster allows the two clusters to work together as one global cluster to provide failover capability if there is a disaster in the site with the production cluster. VCS manages the data replication of the distributed systems data from Site1 to Site2.

The VCS configuration can also be an Active/Active configuration, in which case both Site1 and Site2 have production clusters and have their corresponding failover clusters in the opposite site. For example, the Site2 failover cluster backs up the Site1 production cluster and vice versa.

For each cluster pair, a GDPS agent (also referred to as the *DCM agent*) resides in each cluster (that is, in Site1 and in Site2). At any time, only one of the GDPS agents will be active. Typically, the GDPS/PPRC K-sys in Site2 will have the master role and the GDPS agent in Site2 will be the active agent. A heartbeat is sent from the GDPS agent to the GDPS/PPRC K-sys.

The main objective of the DCM function is to provide a disaster recovery solution between a local and a remote site across both z/OS (using GDPS) and distributed systems applications running on Microsoft Windows, UNIX, IBM AIX, and Linux. DCM can also be used for planned site switches from local to remote sites for clients that have sufficient resources in the recovery site to support this function.

Integrated configuration of GDPS/XRC and VCS clusters

Figure 10-8 illustrates the various components of a GDPS/XRC configuration integrated with a VCS configuration that has the GCO option. The GDPS/XRC configuration consists of one or more System Data Movers (SDMs) and a GDPS K-sys in a sysplex or Parallel Sysplex in Site2, the recovery site.

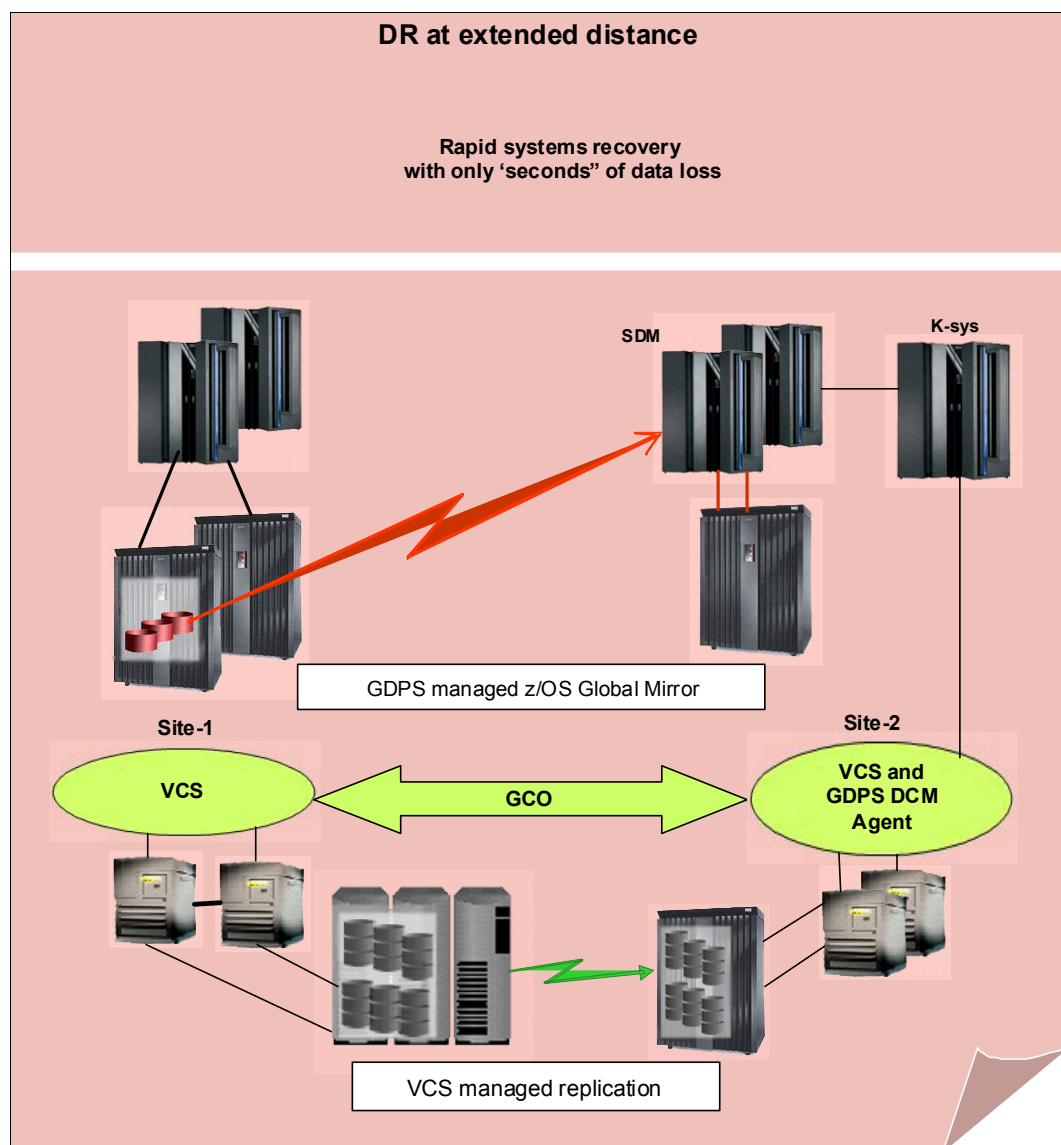


Figure 10-8 GDPS/XRC integration with VCS using GCO: Unlimited distance

The SDMs copy data from a set of primary disks in Site1 and form consistency groups before mirroring the data to a set of secondary disks in Site2 using z/OS Global Mirror (XRC). The disk mirroring is managed by GDPS/XRC as described in Chapter 5, "GDPS/XRC" on page 137.

The VCS configuration in this example is composed of two clusters: the production cluster in Site1, and the failover cluster in Site2. The GCO option in each cluster allows the two clusters to work together as one global cluster to provide failover capability if there is a disaster in the site with the production cluster. VCS manages the data replication of the distributed systems data from Site1 to Site2.

For each cluster pair, there is a GDPS agent (also referred to as the *DCM agent*) in each cluster (that is, in Site1 and in Site2). At any time, only one of the GDPS agents will be active. Typically, the GDPS agent in Site2 will be the active agent. A heartbeat is sent from the GDPS agent to the GDPS/XRC K-sys.

The main objective of the DCM function is to provide a disaster recovery solution between a local site and a remote site across both z/OS (using GDPS) and distributed systems applications running on Microsoft Windows, UNIX, IBM AIX, and Linux. DCM can also be used for planned site switches from local to remote sites for clients that have sufficient resources in the recovery site to support this function.

Integrated configuration of GDPS/GM and VCS clusters

Figure 10-9 illustrates the various components of a GDPS/GM configuration integrated with a VCS configuration that has the GCO option. The GDPS/GM configuration in this example is a Parallel Sysplex configuration in the application site (Site1), an application site controlling system (Kg-sys), a recovery site controlling system (Kr-sys), primary disks, and two sets of disks in the recovery site. The disk mirroring of the primary disks to the recovery site (Site2) is managed by GDPS/GM as described in Chapter 6, “GDPS/Global Mirror” on page 163.

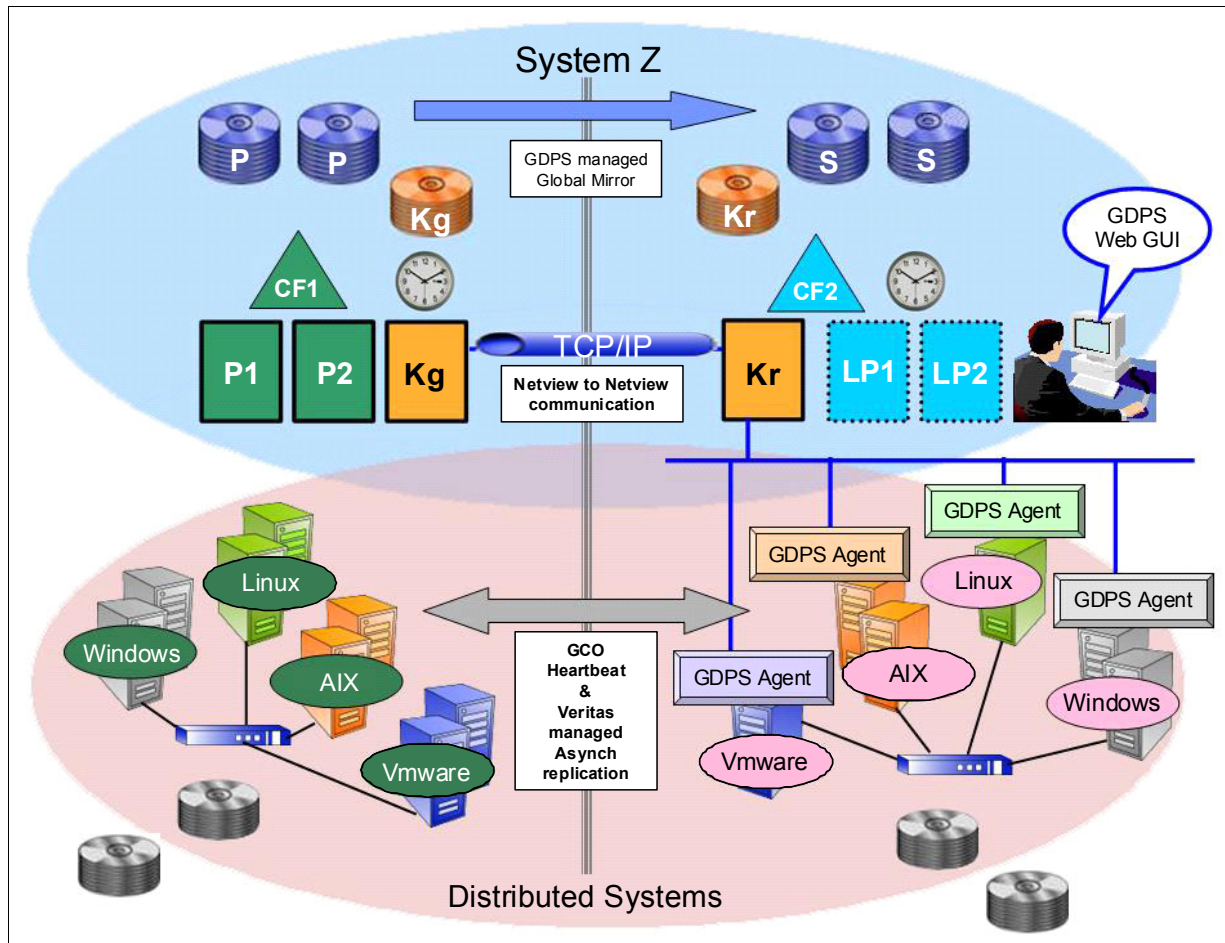


Figure 10-9 GDPS/GM integration with VCS using GCO: Unlimited distance

The VCS configuration in this example is composed of four VCS global clusters:

- ▶ A Microsoft Windows production cluster in Site1, and its failover cluster in Site2.
- ▶ A Linux production cluster in Site1, and its failover cluster in Site2.
- ▶ An AIX production cluster in Site1, and its failover cluster in Site2.
- ▶ A VMware production cluster in Site1, and its failover cluster in Site2.

The GCO option in each cluster allows the two clusters to work together as one global cluster to provide failover capability if a disaster occurs in the site with the production cluster. VCS manages the data replication of the distributed systems data from Site1 to Site2.

For each cluster pair a GDPS agent (also referred to as the *DCM agent*) resides in each cluster (that is, in Site1 and in Site2). At any time, only one of the GDPS agents will be active. Typically, the GDPS agent in Site2 will be the active agent. Similarly, the GDPS DCM functions are active in either the Kr-sys or Kg-sys.

If both Kr-sys and Kg-sys are active, GDPS DCM code is active only in the Kr-sys. A heartbeat is sent from the GDPS agent to both Kg-sys and Kr-sys. However, only the K-sys with DCM active (typically the Kr-sys) will establish communications with the agent.

The main objective of the DCM function is to provide a disaster recovery solution between a local site and a remote site across both z/OS (using GDPS) and distributed systems applications running on Microsoft Windows, UNIX, IBM AIX, and Linux. DCM can also be used for planned site switches from local to remote sites for clients that have sufficient resources in the recovery site to support this function.

Multiple VCS cluster configurations

More than one VCS cluster can exist in an enterprise. For example, assume an SAP application spans multiple platforms: IBM System x, IBM System p, and IBM z Systems. In this case, there will be a System x VCS cluster, a System p VCS cluster, and either GDPS/PPRC, GDPS/XRC, or GDPS/GM running a z Systems workload.

Consider the following points in this scenario:

- ▶ Each global cluster runs one instance of the GDPS agent.
- ▶ Each global cluster must be composed of servers of the same server type (AIX, Linux, Oracle Solaris, and so on).

There can be multiple global clusters of different server types, as in this example:

- AIX VCS cluster 1 in Site1, AIX cluster 2 in Site2: Comprises one global cluster.
- AIX VCS cluster 3 in Site1, AIX cluster 4 in Site2: A second global cluster.
- Linux VCS cluster 1 in Site1, Linux VCS cluster in Site2: A third global cluster.

Figure 10-10 shows a sample configuration with multiple global clusters managed by DCM. As can be seen from this figure, each GDPS agent sends its own heartbeat to the GDPS K-sys in Site2.

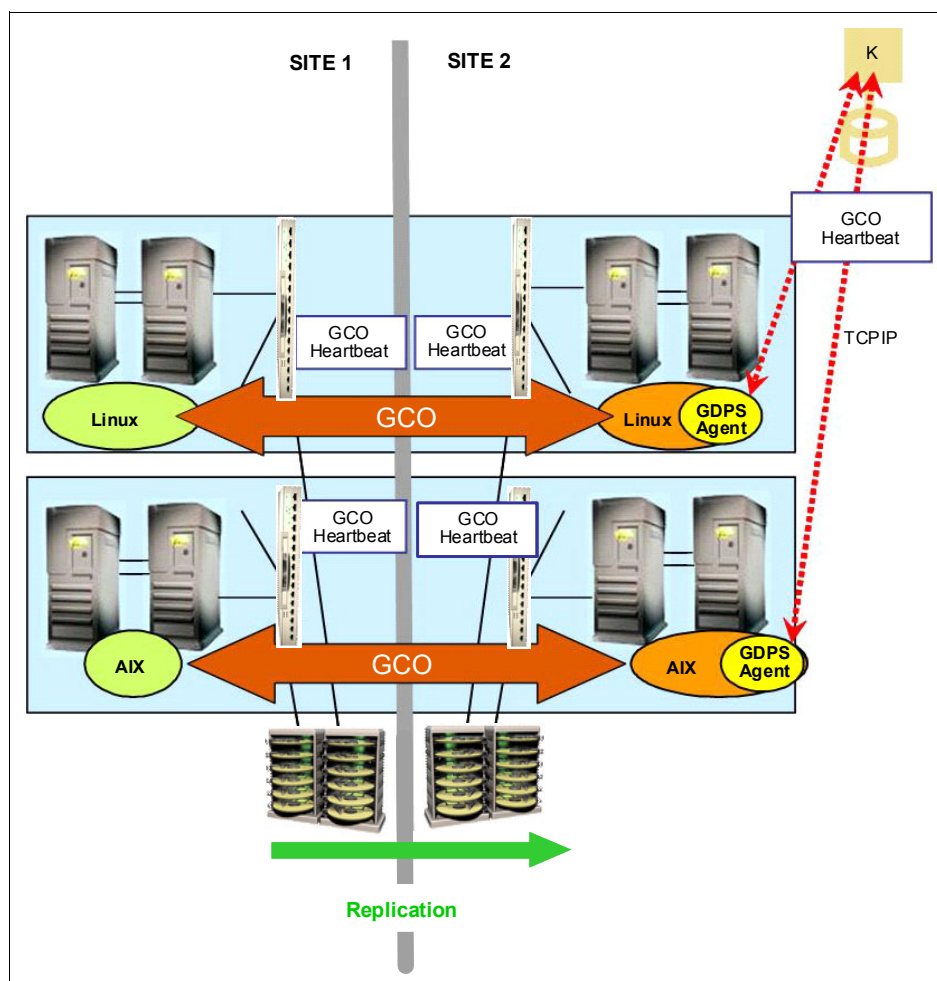


Figure 10-10 Multiple VCS cluster support

DCM functions (VCS)

The DCM support in GDPS/PPRC, GDPS/XRC, and GDPS/GM provide advisory and coordination functions between GDPS and one or more VCS clusters. The advisory functions will provide the capability of continuous heartbeat and status gathering to alert the support staff about any events that might prevent recovery at the time of an outage.

The coordination functions will allow workflow integration for takeover and recovery testing, cross-platform monitoring to maintain recovery capability, and cross-platform recovery management to provide an automated enterprise-level rapid recovery in the case of an outage.

The integration between GDPS and Veritas Clusters provides the following functions:

- Monitoring

GDPS monitors DCM-related resources and generates SDF alerts for resources in an abnormal state.

- ▶ Manual operations

The GDPS panels include an option to query and view the status of DCM resources, and perform planned operations on individual DCM resources.

- ▶ Automation

GDPS issues the takeover prompt and suggests possible scripts to run when it detects various failures associated with DCM resources.

- ▶ Scripting

The scripting capability in GDPS provides workflow integration for actions taken on distributed servers and z Systems servers during a planned or unplanned event. GDPS script statements are provided to control planned and unplanned actions associated with VCS resources:

- Starting the applications for a single cluster or service group.
- Stopping the resources (agent or applications) for a single cluster or service group.
- Switching the applications for a single cluster or service group to the opposite site.
- Planned site switch (either Site1 to Site2, or Site2 back to Site1) of VCS resources. Either all or a selected subset of the VCS resources can be failed over.
- Unplanned failover of the VCS resources (all or a selected subset) from Site1 to Site2.

Sample takeover script: Site1 failure (GDPS/PPRC and VCS)

Figure 10-11 shows an example of a GDPS/PPRC configuration with production systems in Site1 and the GDPS K-sys in Site2. Also in the configuration are two VCS global clusters: an AIX production cluster in Site1 and its failover cluster in Site2; and similarly, a Linux production cluster in Site1 with its failover cluster in Site2. A GDPS agent in each cluster sends a heartbeat to the GDPS K-sys through communication links, as shown in the Figure 10-11.

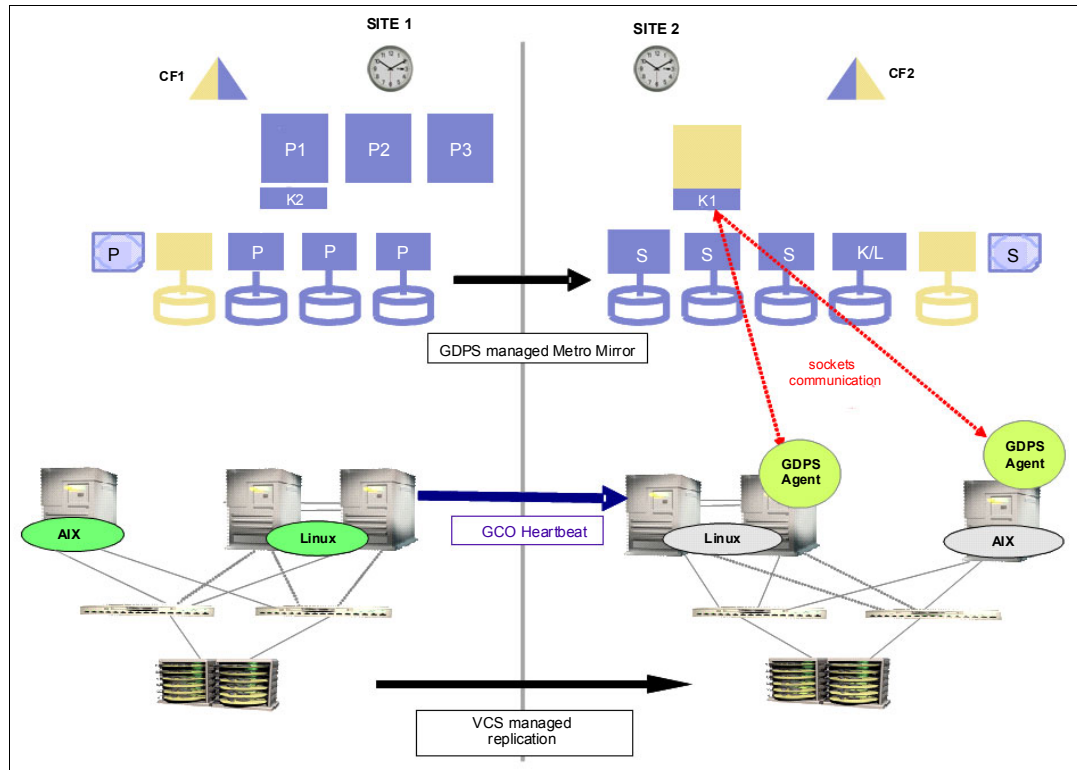


Figure 10-11 Example of GDPS/PPRC configuration with VCS clusters

The failure scenario shown in Figure 10-12 is a Site1 failure when one or more failures occurred in Site1, which can include a mirroring failure, the loss of one or more production z/OS systems, or the loss of one or more VCS clusters in Site1. A disaster is declared which includes the decision to recover all processing in Site2.

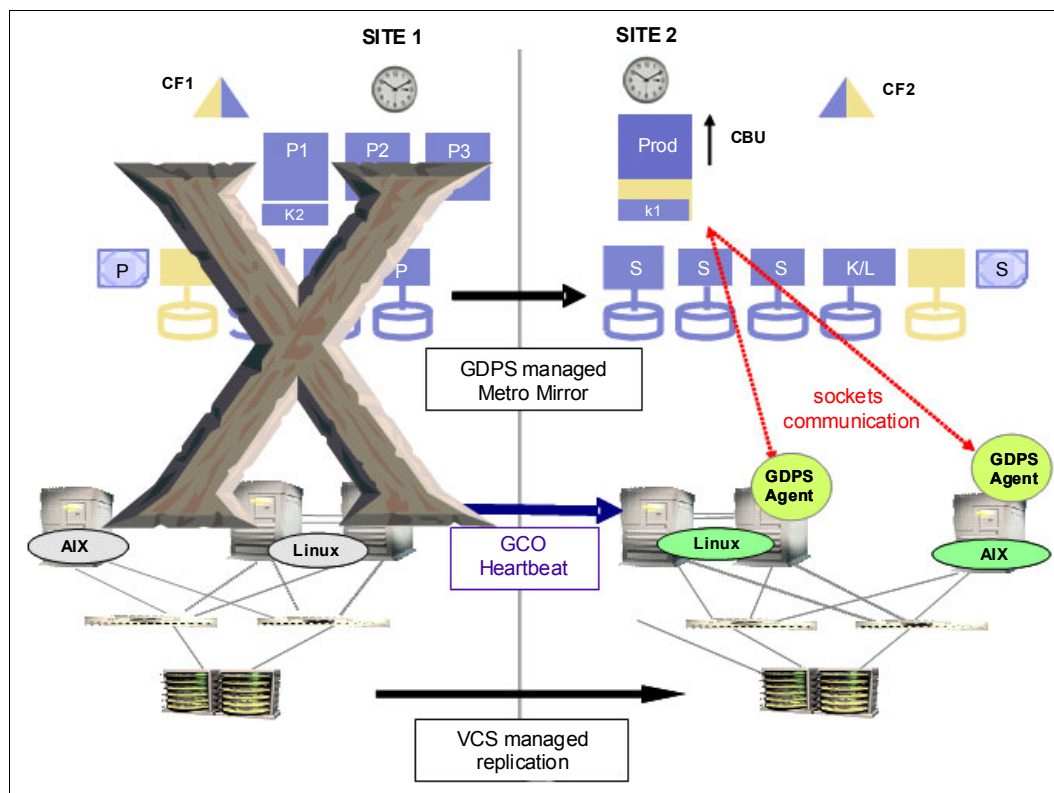


Figure 10-12 Example of GDPS/PPRC configuration with VCS clusters: Site1 failure

With GDPS DCM support providing the ability to manage VCS clusters from GDPS, you have the ability to switch z/OS systems and data and VCS systems and data to Site2 in a coordinated manner.

An existing GDPS script that previously performed failover and restart of z Systems resources can be extended to also include statements to automate the failover of VCS clusters to Site2. When the additional script statements are executed as part of the site takeover script, GDPS automation does the following actions to move z Systems resources from Site1 to Site2:

1. Resets production systems in Site1
2. Reconfigures the secondary disks
3. Activates CBU for servers in Site2
4. Switches couple data sets to those in Site2
5. Activates partitions in Site2
6. Re-IPLs P1, P2, and P3 in Site2

GDPS does the following actions to move VCS clusters from Site1 to Site2:

- Forces a switch of the service group for the AIX cluster from Site1 to Site2.
- Forces a switch of the service group for the Linux cluster from Site1 to Site2.

This example demonstrates the coordinated recovery that can be accomplished across both z Systems resources and VCS clusters when an unplanned outage affects Site1.

10.3.3 DCM support for SA AppMan

This section describes how the functions available with GDPS/PPRC (see Chapter 3, “GDPS/PPRC” on page 53) and GDPS/GM (see Chapter 6, “GDPS/Global Mirror” on page 163) have been integrated with functions provided by IBM Tivoli System Automation Application Manager (SA AppMan).

SA AppMan terminology

This section presents terminology and provides a brief description for each term that is applicable to the DCM support for SA AppMan:

GDPS agent This refers to the logic residing on the cluster that communicates cluster status and events to GDPS and accepts commands on behalf of the cluster from GDPS DCM code for cluster resources that are managed by GDPS. There is one GDPS agent for all SA AppMan cluster sets, running in Site2.

The GDPS agent is available only if you have enabled the Distributed Disaster Recovery (DDR) function of the System Automation Application Manager which is an additional feature license.

System Automation Application Manager (SA AppMan)

IBM Tivoli System Automation Application Manager is designed for high availability and disaster recovery solutions, providing the ability to automate applications across multitiered, heterogeneous environments. It was previously known as the End-to-End Automation Management Component of Tivoli System Automation for Multiplatforms.

Distributed Disaster Recovery (DDR)

This refers to the SA AppMan feature that provides the interaction with GDPS. As mentioned, the GDPS agent is available only if you have enabled DDR.

First Level Automation (FLA) domain

This is used for automation back-end hosting resources that are managed by an automation management product; for example, a Linux cluster on which the applications are automated by IBM Tivoli System Automation for Multiplatforms.

Agentless Adapter (ALA)

The local or remote ALA is used to manage non-clustered domains without an FLA manager.

Domain

This term refers to the automation scope of an automation product instance such as SA MP, IBM High Availability Cluster Multi-Processing (IBM HACMP™), Microsoft Cluster Server (MSCS) and so on. From the FLA perspective, a domain is a cluster. From an SA AppMan perspective, the end-to-end domain automates a whole set of clusters.

Cluster

This refers to a group of servers and other resources that act like a single system and enable high availability and, in some cases, load balancing and parallel processing.

Stretched cluster

This refers to a cluster that is dispersed across two sites.

Cluster set	This refers to the set of one or more clusters that constitute alternatives on different sites supporting the same workload. A cluster set can have a maximum of one local cluster per site. The cluster set is the granularity for which planned/unplanned actions can be performed. For stretched clusters, a cluster set has exactly one cluster, which is stretched across two sites.
Business-critical workload	This refers to applications that are critical to the business (such as databases and web servers).
Discretionary workload	This refers to applications that are not business-critical (for example, development and test applications). Such applications are expected to be shut down to provide backup capacity for business-critical workload applications during planned/unplanned site switch processing.

Note: You define whether an application is business-critical or discretionary when you define your applications in the SA AppMan policy.

SA AppMan overview

SA AppMan uses advanced, policy-based automation to initiate, execute and coordinate starting, stopping, restarting and failing over across entire composite applications in complex cluster environments. Through a single point of control, the software helps you ease management of cross-cluster resource dependencies and improve IT operating efficiency by curtailing manual tasks and maximizing application availability across your enterprise. SA AppMan helps you easily coordinate and manage availability across cluster technologies, and stand-alone servers, so you can better control your enterprise business services.

In the example shown in Figure 10-13 on page 321, the resource *Web*, which is defined on a Windows cluster, has a *startAfter* relationship to the group *Enterprise Service*, which consists of resources that are running on an AIX or Linux cluster, and on a z/OS sysplex. In end-to-end automation management, the resources *App* and *DB2* can have relationships between each other although they are running on different clusters and on different platforms. SA AppMan will make sure, when the applications are started, that *Web* will not be started unless *Enterprise Service* is up and running.

The scope of first-level automation domains is to ensure the high availability of resources as specified in their local (first-level) automation policy. The scope of end-to-end automation is to control the relationships these resources have that span the first-level automation cluster boundary. End-to-end automation does not replace the first-level automation products. Rather, it sends requests to the first-level automation domains to accomplish the goals specified in the end-to-end automation policy as shown in Figure 10-13.

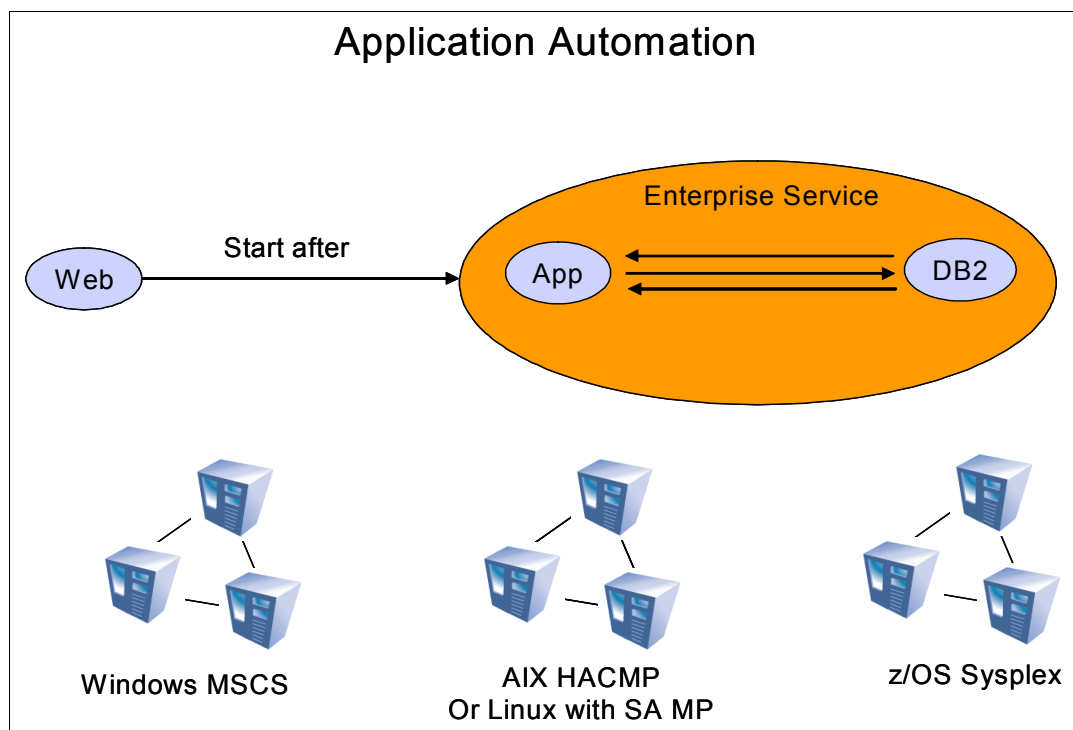


Figure 10-13 Sample end-to-end automation

SA AppMan provides adapters to help manage any combination of the following major clustering technologies:

- ▶ Veritas Cluster Server (VCS) on Solaris
- ▶ High Availability Cluster Multi-Processing (HACMP) on IBM AIX
- ▶ Microsoft Cluster Server (MSCS) on Windows
- ▶ IBM Tivoli System Automation for Multiplatforms (Linux, AIX, Windows, Solaris)
- ▶ IBM Tivoli System Automation for z/OS

For non-clustered servers SA AppMan manages by using *agentless adapters* (ALA).

Integrated configuration of GDPS/PPRC and SA AppMan

Figure 10-14 on page 322 illustrates the various components of a GDPS/PPRC configuration, integrated with a set of different clusters managed by SA AppMan. The GDPS/PPRC configuration, shown at the top of the figure, consists of a multisite Parallel Sysplex with a set of primary disks in Site1 being mirrored to a set of secondary disks in Site2 using Metro Mirror. There is a minimum of one GDPS K-sys in Site2, and optionally there can be a second K-sys in Site1.

The SA AppMan managed configuration is shown at the bottom of the figure. You see different cluster sets that are individually controlled, from an application point of view, by their first-level automation product. For applications having cross-cluster dependencies, SA AppMan provides end-to-end coordination across the cluster sets.

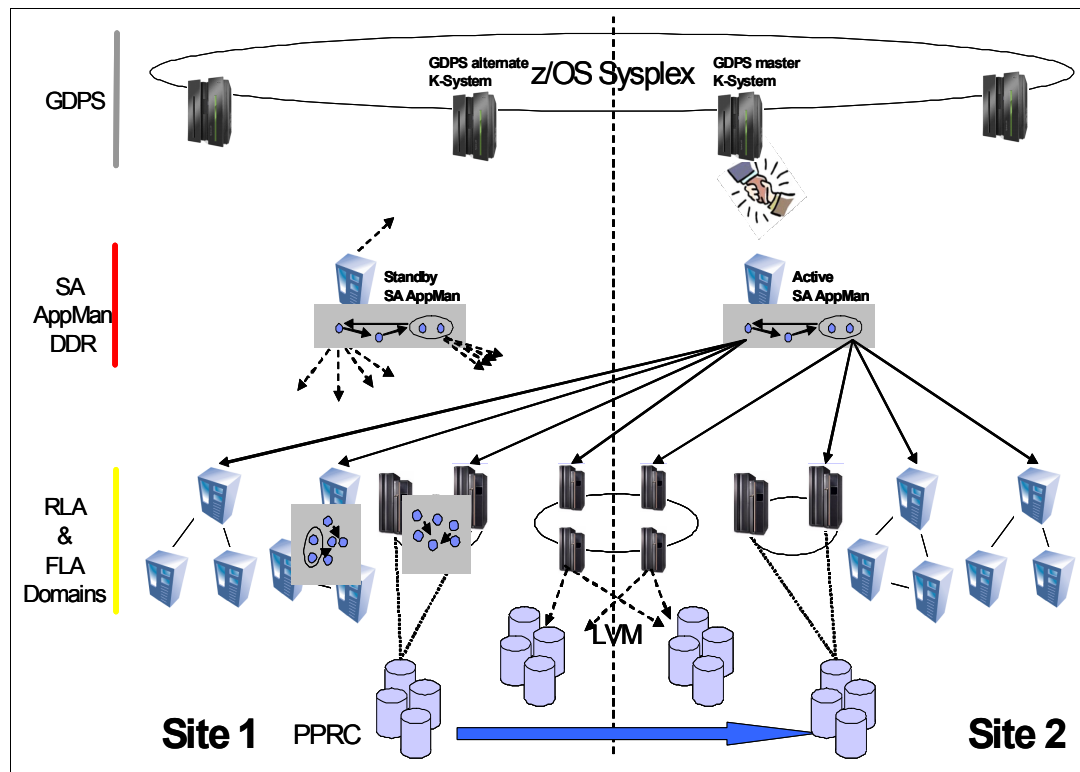


Figure 10-14 SA Application Manager: Distributed Disaster Recovery configuration

In the middle of the figure, one of the distributed servers has the active SA AppMan feature called Distributed Disaster Recovery (DDR). Because the DDR feature includes the GDPS agent function for clusters that support it, DDR integrates availability and disaster recovery features in GDPS/PPRC with advanced automation capabilities delivered with SA AppMan for management of complex, heterogeneous application environments. The environment portrayed is configured for high availability with two instances of both the GDPS controlling system and SA AppMan.

SA AppMan code that communicates with GDPS is implemented in its own server or servers isolated from the GDPS K-sys and the cluster sets that are automated. A policy is defined to describe the topology for sites, cluster sets, and applications controlled by SA AppMan. GDPS does not know the configuration of end-to-end resources, resource groups, clusters or nodes. The GDPS K-sys communicates with the GDPS agent within SA AppMan. The agent provides to GDPS information about any “Site1 workload”, “Site2 workload”, “business-critical workload”, or “discretionary workload” on a per-cluster set basis.

GDPS can then send commands like **start**, **stop**, and so on to cluster sets for these cluster sets. Thus, SA AppMan topology information is not defined in GDPS. Instead, GDPS discovers the SA AppMan resources it will be managing through its communication with the SA AppMan agent, and GDPS presents high-level status for these resources on the GDPS 3270 panel and the GDPS web GUI interface.

Cluster sets

The DCM-supported configuration consists of one or more cluster sets. Each cluster consists of one or multiple systems (nodes) of a single supported platform type (IBM System p, IBM System x, IBM System i®, and so on), and multiple applications can be running. As defined in “SA AppMan terminology” on page 319, a *cluster set* is a set of one or more clusters that constitute alternatives on different sites supporting the same workload. A cluster set can have a maximum of one local cluster per site.

From a DCM perspective, SA AppMan and GDPS work on the cluster set level. GDPS has no awareness of the individual clusters but only of cluster sets. GDPS is also aware of whether a given application is active in Site1 or Site2. There is only one SA AppMan agent, but it controls multiple cluster sets, as shown in Figure 10-14 on page 322.

The following cluster configurations are supported:

- ▶ **Non-stretched cluster active-passive**
This is the simplest configuration, in which all application groups are available in one site and servers in the other site are either running discretionary workload, or are idle. The secondary site is effectively a standby environment in case of a failure in the primary site.
- ▶ **Stretched cluster active-passive**
This configuration looks like the non-stretched cluster, because all applications run in one of the sites (usually Site1).
- ▶ **Stretched cluster active-active**
In this configuration, all nodes in the cluster are active in both sites.

Note: There are two major differences between stretched and non-stretched clusters:

- ▶ For stretched clusters, the application data is replicated with LVM and disk errors are dealt with by LVM.
- ▶ For stretched clusters, GDPS might not be involved in a switch of workload from one site to the other because it might be accomplished completely by first-level automation (FLA).

Data replication

GDPS DCM and SA AppMan do not interface with each other for data replication-related events. SA AppMan expects local disk is available for the workload when this workload is started in a site. Data for the SA AppMan-managed workloads for non-stretched cluster sets can be replicated using Metro Mirror, and this can be managed using the Open LUN support provided by GDPS; see 10.1, “Open LUN Management function” on page 296 for more information. In this way, z/OS and distributed cluster data can be controlled from one point, and a planned switch or unplanned failover for both z/OS and distributed data can be managed from a single control point.

Other data replication technologies, such as software replication in AIX Logical Volume Manager (LVM), can be used for the distributed data. However, SA AppMan will still assume local data is available when the associated workload is started in a site. Mirroring with LVM is not controlled by GDPS or SA AppMan, but is assumed to be managed by the automation product (for example, HACMP) managing the stretched cluster FLA domain.

Data replication for stretched clusters must be performed through LVM so that a data failover can be performed without interruptions to the servers. For a site failure in any of the sites, a stretched cluster with LVM provides availability without any assistance from GDPS.

DCM functions (SA AppMan)

The DCM support in GDPS/PPRC provides advisory and coordination functions between GDPS and one or more SA AppMan cluster sets. The advisory functions provide the capability of continuous heartbeat and status gathering to alert the support staff about any events that might prevent recovery at the time of an outage.

The coordination functions allow workflow integration for takeover and recovery testing, cross-platform monitoring to maintain recovery capability, and cross-platform recovery management to provide an automated enterprise-level rapid recovery in case of an outage.

The integration between GDPS and SA AppMan provides the following functions:

- ▶ **Monitoring**
GDPS monitors DCM-related resources and generates SDF alerts for resources in an abnormal state, and takeover prompts for cluster faults.
- ▶ **Manual operations**
The GDPS panels include an option to query and view the status of DCM resources, and perform planned operations on individual DCM resources.
- ▶ **Automation**
GDPS issues the takeover prompt and suggests possible scripts to run when it detects various failures associated with DCM resources.
- ▶ **High availability for SA AppMan**
Toggle support is provided to switch to the alternate SA AppMan instance under certain scenarios such as SA AppMan server failure or GDPS/PPRC controlling system Master switch.
- ▶ **Scripting**
The scripting capability in GDPS provides workflow integration for actions taken on distributed servers and z Systems servers during a planned or unplanned event. GDPS script statements are provided to control planned and unplanned actions associated with SA AppMan resources:
 - Power On/Off for nodes within one or more cluster sets per site
 - Starting or stopping the applications for one or more cluster sets per site
 - Planned site switch (either Site1 to Site2 or Site2 back to Site1) of SA AppMan resources
 - Unplanned failover of the SA AppMan resources from Site1 to Site2

Sample takeover script: Site1 failure (GDPS/PPRC and SA AppMan)

Figure 10-15 on page 325 shows a sample configuration that consists of the following components:

- ▶ A z/OS sysplex controlled by GDPS configured as a single-site workload (all application systems in Site1 in normal running mode) and the corresponding Metro Mirror configuration.
- ▶ A Linux cluster set (non-stretched cluster) with the active cluster in Site1 and a standby cluster in Site2. Data for the workload in the cluster set is mirrored using Metro Mirror under control by GDPS using the Open LUN support.
- ▶ An AIX stretched cluster with active-active application in both sites. Data replication is performed through LVM.

The DDR feature that includes the GDPS agent communicates with the GDPS K-sys through communication links as shown in Figure 10-15.

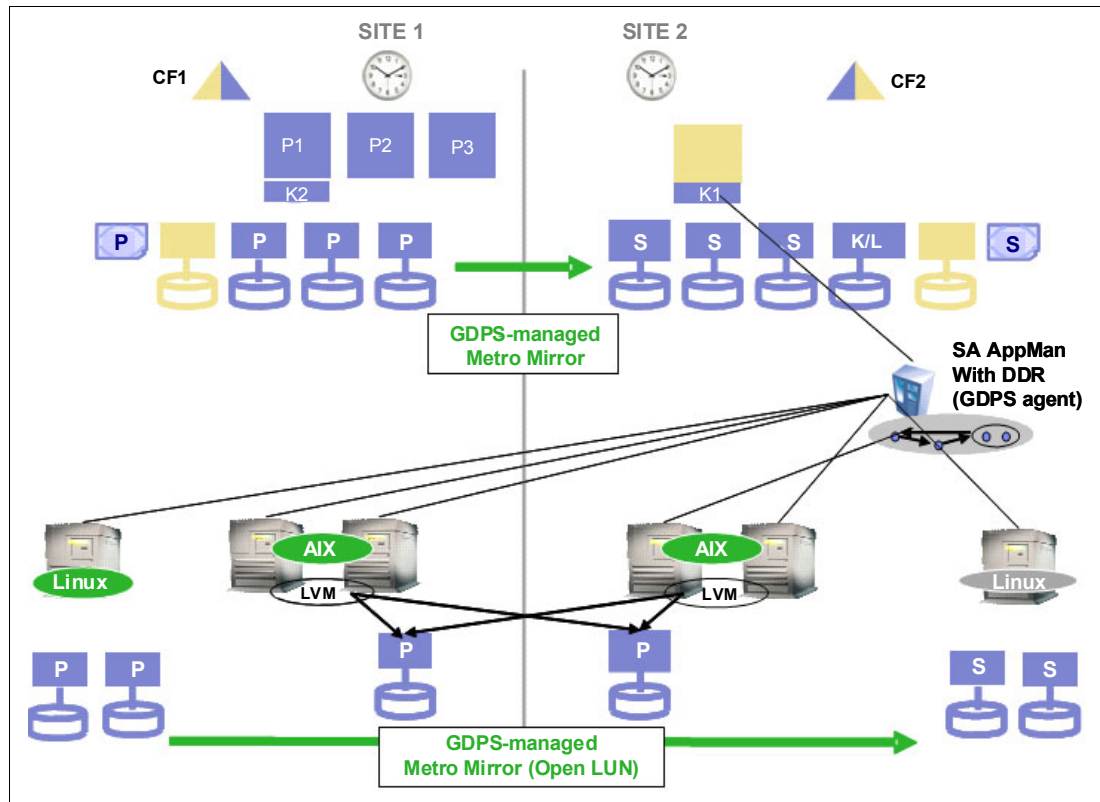


Figure 10-15 Sample GDPS/PPRC DCM configuration

Figure 10-16 on page 326 represents a failure scenario in which one or more failures occur in Site1. This can include a Metro Mirror mirroring failure, the loss of one or more production z/OS systems, or the loss of one or more SA AppMan clusters in Site1. A disaster is declared that includes the decision to recover all processing in Site2.

With GDPS DCM support providing the ability to manage SA AppMan clusters from GDPS, you have the ability to switch z/OS systems and data and SA AppMan systems and data to Site2 in a coordinated manner.

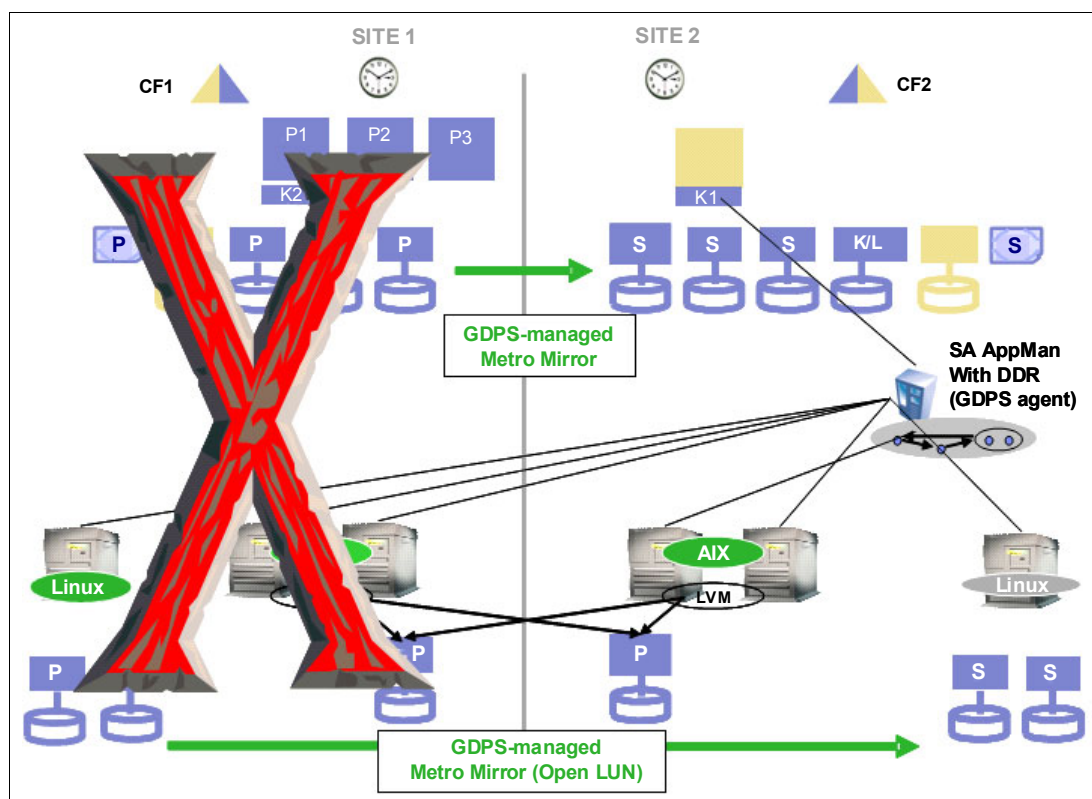


Figure 10-16 Sample GDPS/PPRC DCM configuration: Site1 failure

An existing GDPS script that previously performed failover and restart of z Systems resources can be extended to also include statements to automate the failover of SA AppMan clusters to Site2. When the additional script statements are run as part of the site takeover script, GDPS automation performs the following actions to move z Systems resources from Site1 to Site2:

- ▶ Resets production systems in Site1
- ▶ Reconfigures the secondary disks
- ▶ Activates CBU for servers in Site2
- ▶ Switches couple data sets to those in Site2
- ▶ Activates partitions in Site2
- ▶ Re-IPLs P1, P2, and P3 in Site2

GDPS performs the following actions to move the non-stretched SA AppMan Linux cluster from Site1 to Site2:

- ▶ Makes available those Metro Mirror secondary devices in Site2 (managed by Open LUN) that are associated with the Linux cluster.
- ▶ Sends a Reset Site1 command to SA AppMan to power off all distributed production systems in Site1.
- ▶ Sends a command to SA AppMan to start workload associated with the Linux cluster in Site2.

The AIX active-active stretched cluster managed by its automation product (for example HACMP) might have had production workload running in Site2 all along. If the production workload was not already running in Site2 at the time of the failure, the GDPS script statement to start workload in Site2 will see to it that the AIX workload is started.

This example demonstrates the coordinated recovery that can be accomplished across both the z Systems resources and the AppMan clusters, when there is an unplanned outage that affects Site1.

Integrated configuration of GDPS/GM and SA AppMan

The concepts and capabilities provided by Distributed Cluster Management (DCM) in conjunction with SA AppMan in a GDPS/GM environment are similar to those described in “Integrated configuration of GDPS/PPRC and SA AppMan” on page 321. In this section, we do not repeat the same concepts. Instead, we provide an overview of the solution and explain the differences with the GDPS/PPRC implementation.

The major difference between the GDPS/PPRC and GDPS/GM implementations stems from the difference between these two solution models:

- ▶ GDPS/PPRC is a high availability and disaster recovery solution implemented across two sites separated by a relatively short distance.
- ▶ GDPS/GM, because of the asynchronous nature of the copy technology and the typical distances between the two sites, is only a disaster recovery solution.

This difference reflects on the DCM AppMan implementation. Whereas supporting stretched clusters in a GDPS/PPRC environment might be possible, stretching clusters do not make sense in a “failover” recovery model. Similarly, in a GDPS/PPRC environment, using LVM instead of PPRC for replication is possible. In a GDPS/GM environment, it is unlikely that LVM can be supported over large distances².

² Consult with your vendor to understand their support position.

Figure 10-17 provides a high-level depiction of how GDPS/GM and DCM AppMan integrate.

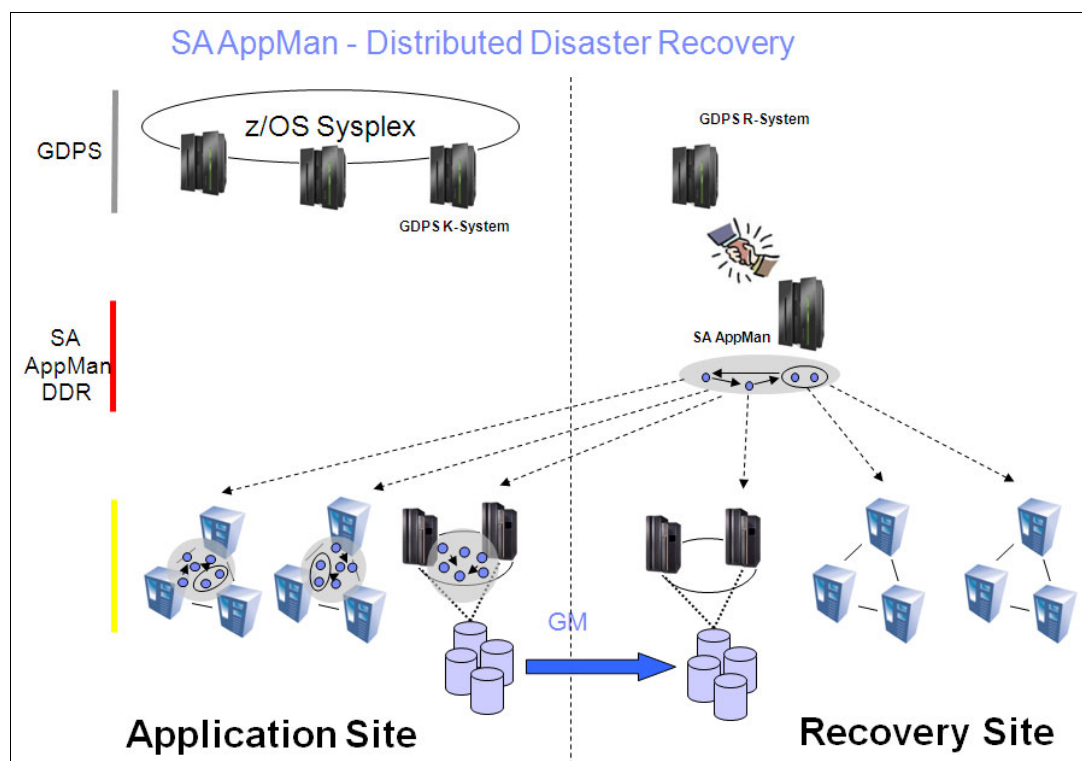


Figure 10-17 SA Application Manager: Distributed Disaster Recovery configuration for GDPS/GM

GDPS/GM DCM AppMan features the following general characteristics:

- ▶ The AppMan server runs on Linux for z Systems (guest or native).
- ▶ The AppMan server communicates with the GDPS/GM remote controlling system.
The AppMan server and the GDPS/GM remote controlling system are expected to be co-located in the recovery site, although this is not mandatory.
- ▶ The Appman server and the clusters in the application site that it manages can be separated by unlimited distances.
- ▶ The clusters in the application site are running business-critical workloads. The clusters in the recovery site can be standby or running discretionary workloads.
- ▶ High availability for SA AppMan delivers toggle support to switch to the alternate SA AppMan instance under certain scenarios such as SA AppMan server failure.
- ▶ GDPS manages Global Mirror replication on behalf of the AppMan clusters:
 - Distributed data for one or more AppMan clusters can be in the same Global Mirror session together with the data for one or more z Systems if cross-system consistency is required. In this case, the z Systems and distributed clusters having their data mirrored in the same session are recovered together.
 - Distributed data for different clusters can be managed as one or more independent Global Mirror sessions isolated from z Systems data. A Global Mirror session and the clusters having their data mirrored in that session will be the scope of recovery.

- If a recovery action is initiated, GDPS automation performs the Global Mirror data recovery and coordinates, with the AppMan server, the recovery actions for the distributed clusters. GDPS instructs the AppMan server to kill any discretionary workload on clusters that will be used to house critical workloads, and to start the critical workloads in the recovery site. All of this is accomplished with a single GDPS automation script.

10.3.4 GDPS/PPRC Support for IBM zEnterprise BladeCenter Extension (zBX)

GDPS provides the capability to manage *workloads* running on zBX hardware using the functions described in 10.3.2, “DCM support for VCS” on page 308 and 10.3.3, “DCM support for SA AppMan” on page 319. The GDPS/PPRC hardware management support for zBX compliments this support to enable direct control of the zBX hardware (blades) and virtual server resources. GDPS builds on support provided by system automation for z/OS processor operations to enable you to activate and deactivate the blades and virtual servers that run on them. You can manage these resources through GDPS panels or through GDPS scripts. This enables you to combine recovery of z/OS-based workloads and workloads running on the zBX in a single automated workflow. Although this support is available without the need for VCS or SA AppMan workload management, these solutions can be combined to provide both hardware and workload management.

10.3.5 Summary

The Distributed Cluster Management function of GDPS can provide a single point of control to monitor and manage both z Systems resources and distributed server resources. DCM can also provide coordinated failover for planned and unplanned events that can affect either the z Systems resources or the distributed server resources or both. In short, you can attain business resiliency across your entire enterprise.



Combining local and metro continuous availability with out-of-region disaster recovery

In this chapter, we discuss the capabilities and considerations for implementing GDPS/Metro Global Mirror (GDPS/MGM) and GDPS/Metro z/OS Global Mirror (GDPS/MzGM). It is of interest to clients that have requirements for both continuous availability locally and regional disaster recovery protection.

GDPS/MGM and GDPS/MzGM combine the continuous availability attributes of GDPS/PPRC HM, GDPS/PPRC, or GDPS/MTMM with the out-of-region disaster recovery capabilities of GDPS/GM or GDPS/XRC to protect critical business data during a wide-scale disruption. They also provide for fast automated recovery under various failure conditions.

Note: Both GDPS/PPRC and GDPS/PPRC HyperSwap Manager can be combined with Global Mirror or z/OS Global Mirror, as described in this chapter. To aid in readability, only GDPS/PPRC is used in the text for most of the descriptions. If a particular function is not supported by GDPS/PPRC HyperSwap Manager, it is specifically mentioned.

Also, GDPS/MTMM can only be combined with Global Mirror, it cannot be combined with z/OS Global Mirror.

The following functions are provided by GDPS/MGM and GDPS/MzGM:

- ▶ Three-copy disk mirroring using GDPS/PPRC or GDPS/MTMM to support zero data loss for day-to-day disruptions at metropolitan distances, and GDPS/GM or GDPS/XRC for long distance, out-of-region data protection, with limited data loss during a wide-scale disruption.

- ▶ Four-copy¹ disk mirroring combining GDPS/PPRC or GDPS/MTMM in the production region to support zero data loss for day-to-day disruptions at metropolitan distances, GDPS/GM or GDPS/XRC between the two regions, and another instance of GDPS/PPRC or GDPS/MTMM in the recovery region to manage Global Copy (PPRC-XD) that can be switched to synchronous-mode while moving production to the recovery region in a planned or unplanned manner.
- ▶ Multisite management of the remote copy environment to maintain data integrity and data consistency across all disk copies.
- ▶ Support for transparent switching to secondary disks if there is a primary disk storage subsystem failure by using GDPS/PPRC or GDPS/MTMM with HyperSwap.
This support offers the ability to incrementally resynchronize the GDPS/GM² or GDPS/XRC mirror after a PPRC HyperSwap.
- ▶ Fast, automated recovery for RTO of less than an hour for site and regional disasters.
- ▶ Zero data loss protection for both open systems and z Systems by using GDPS/PPRC and GDPS/GM, assuming that only one site is lost during a disaster.
- ▶ Use of FlashCopy to facilitate nondisruptive functions (such as backups, data mining, application testing, disaster recovery testing), and to provide a consistent copy of the data during remote copy synchronization to ensure disaster readiness is maintained at all times.
- ▶ Planned switch to run production in the recovery region and then return home.

11.1 Introduction

Enterprises running highly critical applications have an increasing need to improve the overall resilience of their business services and functions. Enterprises already doing synchronous replication have become accustomed to the availability benefits of relatively short distance synchronous replication. This is especially true in mainframe environments where the capabilities of HyperSwap provide the ability to handle disk subsystem failures without an outage and to use server capacity in both sites.

Regulatory bodies (both governmental and industry-based) in various countries are requiring enterprises to maintain a significant distance between their primary and disaster locations to protect against wide-scale disruptions. For some organizations, this can result in a requirement to establish backup facilities well outside the range of synchronous replication capabilities, thus driving the need to implement asynchronous disk mirroring solutions.

From a business perspective, this could mean compromising continuous availability to comply with regulatory requirements. With a three-copy disk mirroring solution, the availability benefits of synchronous replication can be combined with the distance allowed by asynchronous replication to meet both the availability expectations of the business and the requirements of the regulator. Further extension to four-copy configurations allows for equivalent high availability characteristics when running in either region.

¹ Incremental Resynchronization of GDPS/GM and management of four copy configurations are not supported in conjunction with GDPS/PPRC HM.

² Incremental Resynchronization of GDPS/GM and management of four copy configurations are not supported in conjunction with GDPS/PPRC HM.

11.2 Design considerations

In the following sections we describe design considerations including three-copy solutions versus 3-site solutions, multi-target and cascading topologies, four-copy solutions, and cost considerations.

11.2.1 Three-copy solutions versus 3-site solutions

It is not always the case that clients implementing a three-copy mirroring solution will have three independent data centers (as shown in Figure 11-1), each with the capability to run production workloads.

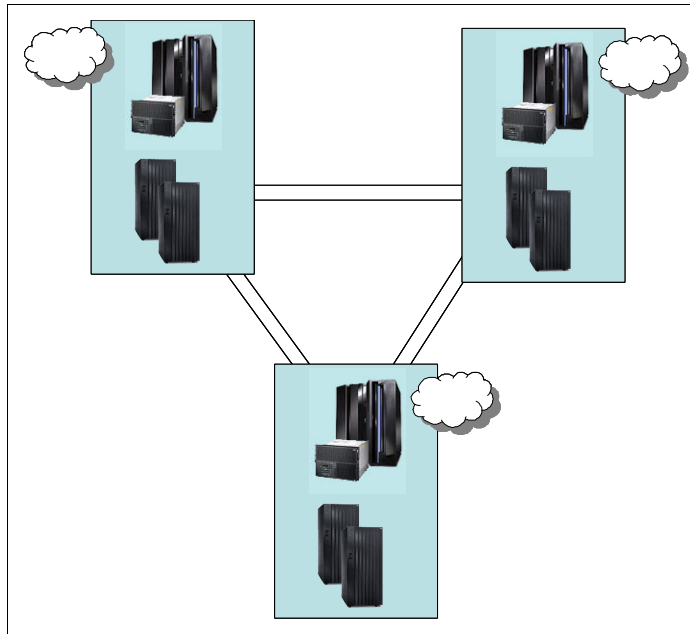


Figure 11-1 Three-site solution

Having three distinct locations with both the connectivity required for the replication and connectivity for user access is expensive and might not provide sufficient cost justification. As the distance between the locations connected with synchronous mirroring increases, the ability to provide continuous availability features such as cross-site disk access, HyperSwap, or CF duplexing diminishes.

Having a production location with two copies of data within a single data center (shown in Figure 11-2), along with a third copy of the data at a remote recovery location, provides you with many of the benefits of a full 3-site solution while allowing for a reduced overall cost. Disk subsystem failures are handled as local failures and if the single site has some degree of internal resilience, then even minor “disaster-type” events can perhaps be handled within the single location.

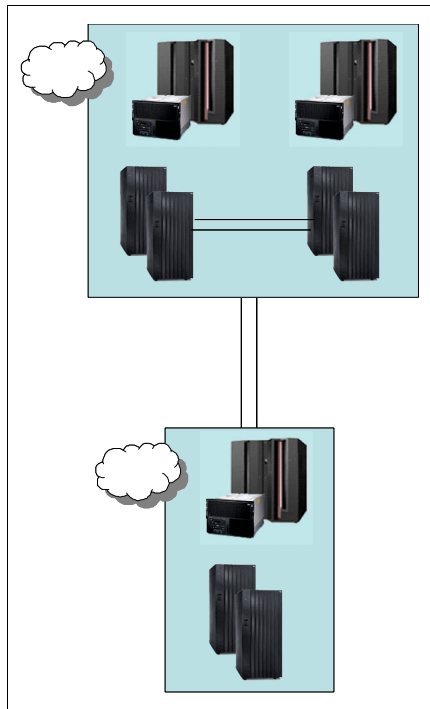


Figure 11-2 A 2-site solution

Another benefit of the two-data center solution, especially in a z Systems environment, is that you can realize the full benefit of features such as HyperSwap and coupling facility duplexing to provide continuous availability features without provisioning significant additional and expensive cross-site connectivity, or having concerns regarding the impact of extended distance on production workloads.

Figure 11-3 illustrates another variation of this scenario, in which the primary data center is a campus location with separate machine rooms or buildings, each with the ability to run production workloads.

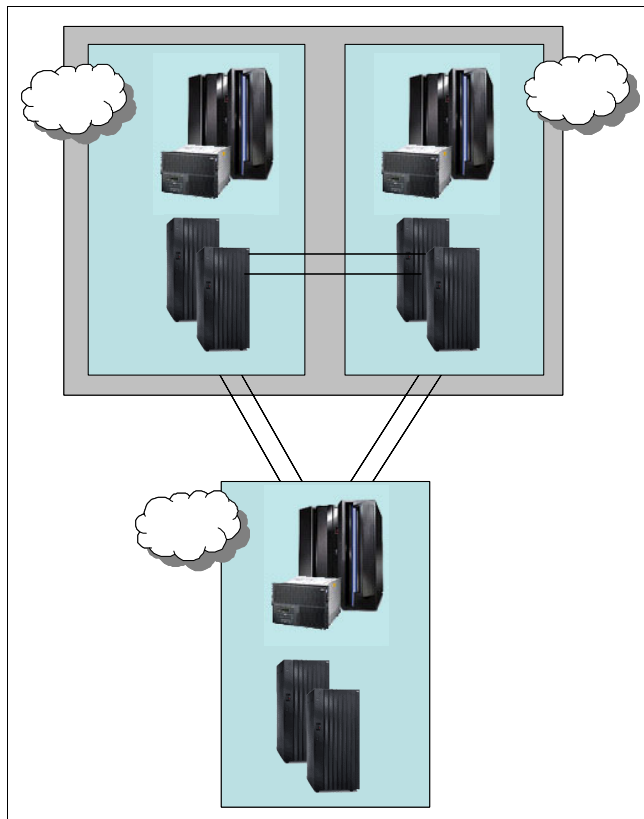


Figure 11-3 A 2-site solution: Campus and Recovery site

In the past, clients often used the bunker topology (as shown in Figure 11-4) to create a solution that could provide mirroring at extended distances, but still handle a primary site failure without data loss.

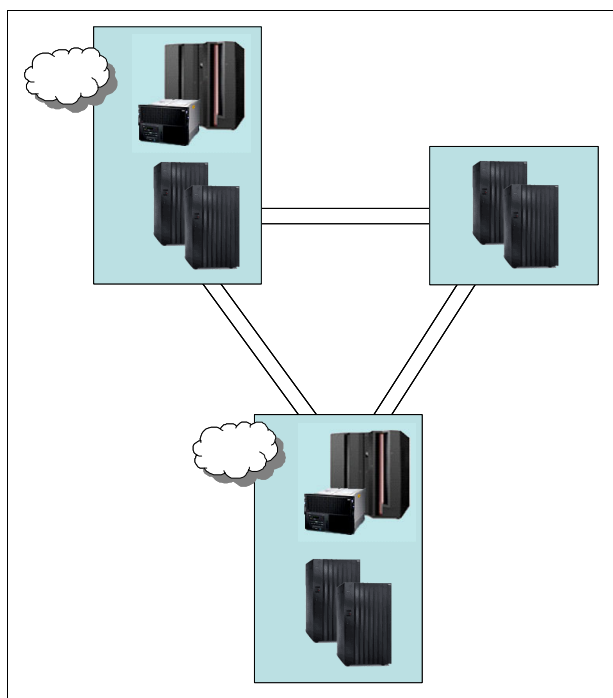


Figure 11-4 Two sites and an intermediate bunker

There are several arguments against this approach:

- ▶ For guaranteed zero data loss you need a policy in which, if the mirroring stops, the production applications are also stopped. There are clients who have implemented such a policy, but it is not a common policy. If production is allowed to continue after a local mirroring failure, then zero data loss cannot be guaranteed in all situations.
- ▶ If the disaster event also affects the bunker site or affects the bunker site first, then zero data loss is again not guaranteed. If the reason for the extended distance to the recovery site was to handle regional events, then this possibility cannot be excluded.
- ▶ The networking and hardware costs of the bunker site are probably still considerable despite there being no servers present. Further investment in the availability characteristics of the primary location or in a campus-type solution in which the synchronous secondary disk subsystems can be used for production services might provide a greater return on investment for the business.

11.2.2 Multi-target and cascading topologies

Multi-target and cascading topologies are similar in terms of capabilities in that both provide a synchronous and an asynchronous copy of the production data. Certain failure scenarios are handled more simply by multi-target solutions and other scenarios by cascading solutions.

The key requirements for either topology are as follows:

- ▶ A viable recovery copy and recovery capability is available at all times in a location other than where production is running. It is possible that regulatory requirements will demand this. This requirement implies that no situations exist in which both offsite copies are compromised.

- Any single site failure will result in only at most a short outage of the replication capability between the surviving sites to ensure minimal exposure where there might be increased data loss for a second failure.

With this requirement, being able to do incremental resynchronization between any two copies of the data is extremely desirable. Not having this can result in an extended period of exposure to additional data loss in case of a second failure.

11.2.3 Four-copy solutions

Many parallels can be drawn between 4-site solutions and the 3-site solutions presented in the previous section. That is, clients are unlikely to have four discrete physical data centers, because of cost implications. In all probability, the four-copy solution is most likely to be implemented in two physical locations where each location has two “hardened” data center facilities in the one location. For example, in two adjacent buildings on a campus, or even two separate data center halls within a single building where the halls are separated by fire resistant barriers and are independently provided with power and so on.

11.2.4 Cost considerations

The third and potentially fourth locations are, in many situations, regarded as an insurance copy and as mainly providing regulatory compliance. This might imply that costs for this location are kept to an absolute minimum.

Reducing the network bandwidth to remote locations can provide significant cost savings for the overall cost of the solution. Given that a synchronous copy is already available ‘locally’, trading off the RPO versus the cost of the network might be a useful compromise especially if the times of increased RPO are during periods of batch processing or database maintenance where the transactional data loss would be smaller.

Using a disaster recovery service provider such as IBM BCRS is one method of reducing the costs of the third location, fourth location, or both. Shared hardware assets and the removal of the requirement to invest in additional physical locations can provide significant cost benefits, and with the majority of events expected to be handled in the two main locations, the disadvantages of a shared facility are reduced.

11.2.5 Operational considerations

When running in multiple locations and combining different techniques together to provide an overall solution, there can be the requirement to do synchronized actions in both regions. To facilitate this from an operational standpoint, GDPS provides a Remote Script Execution function, so that from a single point of control you are able to initiate actions in any of the individual GDPS environments that make up the overall solution.

11.3 GDPS Metro/Global Mirror 3-site solution

This section describes the capabilities and requirements of the GDPS Metro/Global Mirror 3-site (GDPS/MGM 3-site) solution.

GDPS provides two configurations for the GDPS/MGM 3-site solution:

- ▶ The first configuration is a cascading data replication solution that combines the capabilities of GDPS/PPRC and GDPS/GM. This variation is referred to as a *cascaded-only GDPS/MGM 3-Site* configuration.
- ▶ The second configuration is an extension of the cascaded-only GDPS/MGM 3-Site configuration that can dynamically switch between a cascaded topology and a multi-target topology as necessary to optimize recovery scenarios such as HyperSwap. This configuration combines GDPS/MTMM with GDPS/GM, and is referred to as a *GDPS/MGM Multi-Target 3-Site* configuration.

For both configurations, synchronous replication between a primary and secondary disk subsystem that is within a single data center, or between two data centers within metropolitan distances, is implemented with GDPS/PPRC or GDPS/MTMM.

GDPS/GM is used to asynchronously replicate data to a third disk subsystem in a recovery site that is typically out of the local metropolitan region.

Because both Metro Mirror and Global Mirror are hardware-based remote copy technologies, CKD and FB devices can be mirrored to the recovery site, which protects both z Systems and distributed system data³. For enterprises that require consistency across both distributed systems and z Systems data, GDPS/MGM provides a comprehensive three-copy data replication strategy to protect against day-to-day disruptions, while protecting critical business data and functions if there is a wide-scale disruption.

11.3.1 GDPS/MGM 3-site overview

The GDPS/MGM 3-site configuration that is shown in Figure 11-5 on page 339 is a 3-site continuous availability and DR solution. In this example, Site1 and Site2 are running an Active/Active workload (for more information see 3.2.3, “Multisite workload configuration” on page 72) and are within metropolitan distances to ensure optimal application performance. All data that is required to recover critical workloads is on disk and is mirrored. Each site is configured with sufficient spare capacity to handle failed-over workloads during a site outage.

The third site, or recovery site, can be at virtually unlimited distance from Site1 to Site2 to protect against regional disasters. Asynchronous replication is running between Site2 and the recovery site. Redundant network connectivity is installed between Site1 and the recovery site to provide continued disaster recovery protection during a Site2 disaster or a failure of the disk subsystems in Site2. For more information, see “Incremental resynchronization for GDPS/MGM 3-site” on page 339.

³ This capability is available with the *cascaded-only* GDPS/MGM 3-Site configuration only.

There is sufficient CPU capacity installed to support the R-sys. CBU is installed and GDPS will invoke CBU on z Systems to provide the additional capacity needed to support production workloads if disaster recovery is invoked.

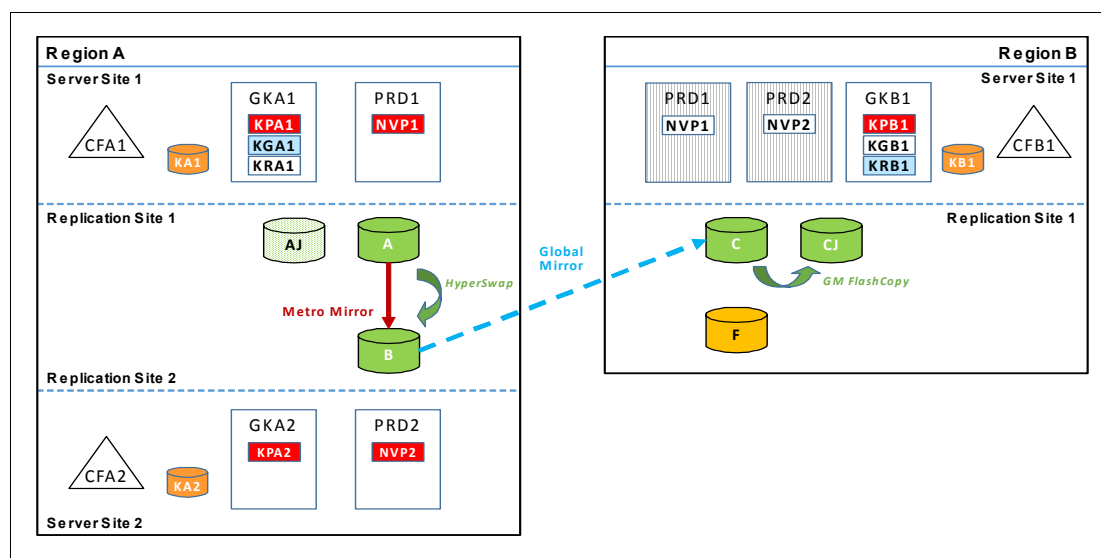


Figure 11-5 GDPS Metro Global Mirror cascaded configuration

The A disks are synchronously mirrored to the B disks in Site2 using Metro Mirror. The B disks are then asynchronously mirrored to a third (C) set of disks in the recovery site using Global Mirror. A fourth set of disks (CJ), also in the recovery site, are the FlashCopy targets used to provide the consistent data (“journal”) for disaster recovery. A fifth (F) and optional set of disks are used for stand-alone disaster recovery testing or, in the event of a real disaster, to create a “golden” or insurance copy of the data. For more detailed information about Global Mirror, see Chapter 6, “GDPS/Global Mirror” on page 163.

Because some distance is likely to exist between the local sites, Site1 and Site2, running the PPRC leg of MGM, and the remote recovery site that is the GM recovery site, we also distinguish between the local sites and the remote site using *region* terminology. Site1 and Site2 are in one region, Region A, and the remote recovery site is in another region, Region B.

Incremental resynchronization for GDPS/MGM 3-site

The incremental resynchronization function of Metro Global Mirror enables incremental resynchronization between Site1 and the recovery site when the intermediate site, Site2, or the disk subsystems in the intermediate site are not available.

Without this capability, if the intermediate site becomes unavailable, the data at the recovery site starts to age because data could no longer be replicated. Instead of requiring a new Global Mirror session from the production site to the recovery site (and a full copy), incremental resynchronization capability with GDPS/MGM 3-site supports a configuration where only incremental changes must be copied from Site1 to the recovery site.

Figure 11-6 shows how GDPS/MGM 3-site can establish a Global Mirror session between the production site, Site1, and the recovery site when it detects that the intermediate site is unavailable.

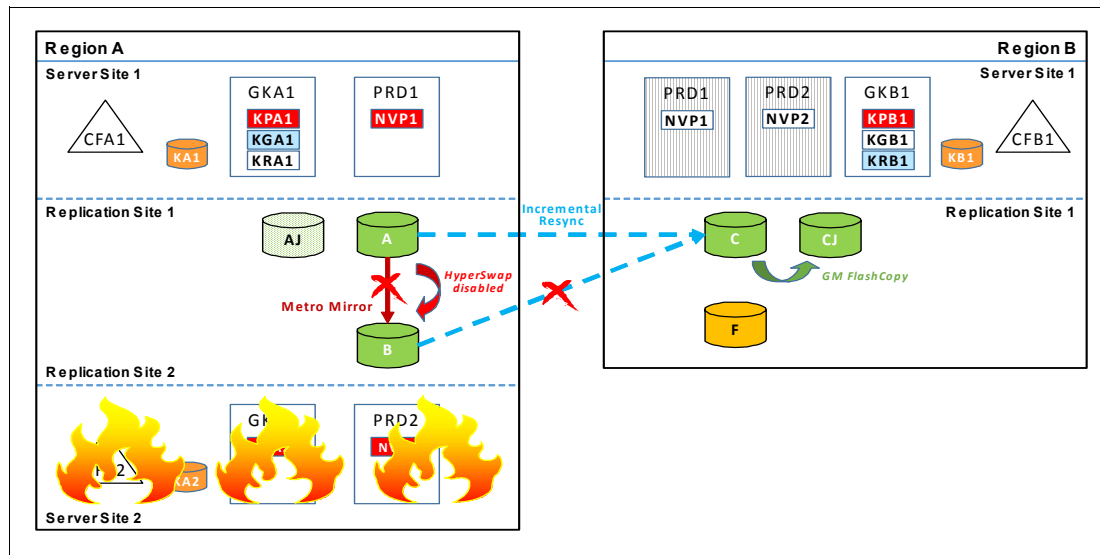


Figure 11-6 GDPS Metro Global Mirror configuration after Site2 outage

After the session is established, only an incremental resynchronization of the changed data needs to be performed, which allows the disaster recovery capability to be restored in minutes, instead of hours, when the intermediate site is not available.

Figure 11-7 shows a GDPS/MGM Multi-Target 3-site configuration when it is in a multi-target topology. GDPS/MGM Multi-Target 3-site configurations can dynamically switch between a cascaded topology and a multi-target topology to optimize processing of various recovery scenarios.

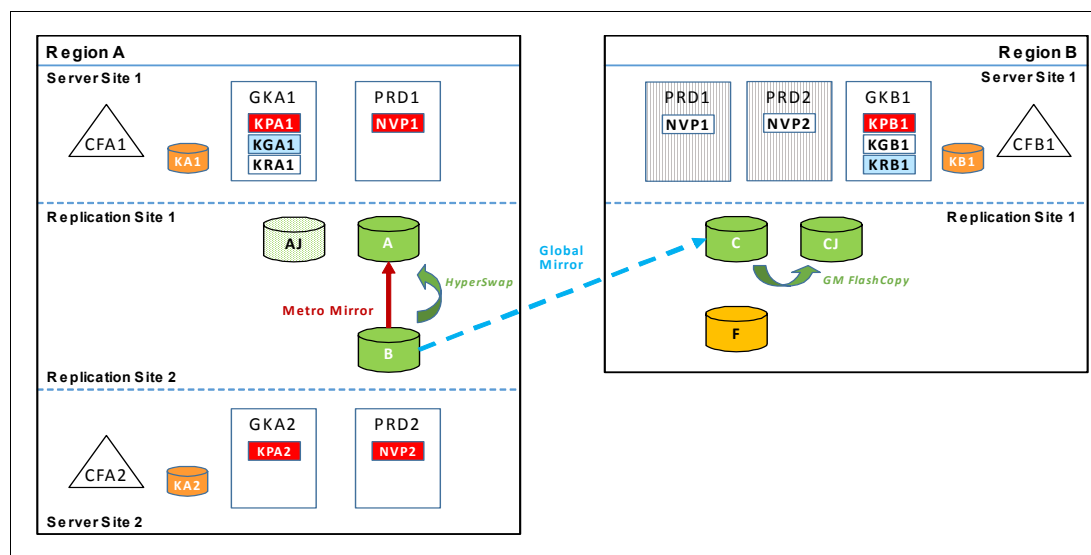


Figure 11-7 GDPS/MGM Multi-Target 3-site configuration

Assume that your GDPS/MGM Multi-Site 3-Site configuration started out in a cascaded topology, as shown in Figure 11-5 on page 339. If you execute a planned HyperSwap to the B disk, followed by a reverse and resynchronization of Metro Mirror from the B disk back to the A disk, you would wind up in the multi-target topology that is shown in Figure 11-7. As shown in Figure 11-7, the B disk is now the primary copy of data that application systems are currently accessing and the A disk is the Metro Mirror secondary disk to the B disk. HyperSwap has been reenabled to provide high availability for the Region A data. This synchronous relationship is managed by GDPS/MTMM in Region A. The B disk is also the Global Mirror primary disk, being copied to the C disk that is the Global Mirror secondary disk. This asynchronous relationship is managed using GDPS/GM. Incremental resynchronization is still enabled from the A disk to the C disk to protect from a failure of the B disk and allow the Global Mirror copy to be re-established without the need for a full copy.

The advantage of the multi-target capability in this scenario is that, following the HyperSwap, Global Mirror from the B disk to the C disk can remain active and maintain your DR position, while Metro Mirror in Region A is being resynchronized from the B disk back to the A disk. In the same situation with cascaded-only MGM 3-Site, Global Mirror from the B disk to the C disk must be suspended while Metro Mirror in Region A is being resynchronized from the B disk, back to the A disk, which results in your DR position aging while Metro Mirror is being resynchronized.

GDPS/MGM Procedure Handler

The GDPS/MGM Procedure Handler is a fully integrated component of GDPS for use in 3-site IR or 4-site configurations. The Procedure Handler, along with the provided procedures, can be used to drive several complex scenarios with a single script invocation, as shown in the following examples:

- ▶ To incrementally reintroduce the Site2 intermediate disk if GM had been incrementally resynchronized from Site1 to the recovery site.

A supplied procedure provides the ability to return to an A-B-C configuration when running in an A-C configuration. Without the procedure, returning to an A-B-C configuration would have required full initial copy for both PPRC (A-disk to B-disk) and for GM (B-disk to C-disk). Thus, the procedure provides significant availability and disaster recovery benefit for IR environments.

The procedure can be used for this purpose only if the B-disk is returned “intact,” meaning that metadata on the disk subsystem pertaining to its status as a PPRC secondary and GM primary disk is still available. If you need to introduce a new disk subsystem into the configuration, this is going to require full initial copy of all the data.
- ▶ To perform a planned toggle between the A-disk and the B-disk.

If you intend to perform periodic “flip/flops” of Site1 and Site2 (or A-disk and B-disk), another procedure allows you to go from an A-B-C configuration to a B-A-C configuration and then back to an A-B-C configuration in conjunction with A-disk to B-disk planned HyperSwap and B-disk to A-disk planned HyperSwap.
- ▶ To incrementally “return home” after recovering production on C-disk, or after you have switched production to C-disk by reintroducing both the A-disk and the B-disk.

This is a C to A-B-C (or B-A-C) transformation. It assumes that both the A-disk and the B-disk are returned intact. Although the MGM mirror can be incrementally reinstated, a production outage is necessary to move production from running on the C-disk in the recovery site back to either the A-disk or the B-disk in one of the local sites.

The Procedure Handler supports only CKD disks. Incremental Resynchronization is not supported with GDPS/PPRC HM.

11.3.2 GDPS/MGM Site1 failures

The primary role of GDPS is to protect the integrity of the B copy of the data. At the first indication of a failure in Site1, GDPS/PPRC or GDPS/MTMM will freeze all B disks, both CKD and FB, to prevent logical contamination of data that is on the B devices. For more information about GDPS/PPRC processing, see Chapter 3, “GDPS/PPRC” on page 53.

At this point, the GDPS/GM session between Site2 and the recovery site is still running, and both locations most likely will have the same set of data after a brief amount of time. The business focus is now on restarting the production systems in either Site2 or the recovery site, depending on the failure scenario. If the systems are started in Site2, the GDPS/GM solution is already in place.

11.3.3 GDPS/MGM Site2 failures

In this situation, the production systems are still running, so the business requirement is to ensure that disaster recovery capabilities are restored as fast as possible. The GDPS/GM session should be restarted as soon as possible between Site1 and the recovery site using incremental resynchronization. See “Incremental resynchronization for GDPS/MGM 3-site” on page 339 for more details. If incremental resynchronization is not configured, a full copy is required.

This scenario has possibly less impact to the business than a failure of the production site, but this depends on the specific environment.

11.3.4 GDPS/MGM region switch and return home

It is possible to switch production from running in Region A (in either Site1 or Site2) to Region B. Many GDPS/MGM 3-site customers run Site1 and Site2 in the same physical site or on a campus where these two sites are separated by little distance. In such configurations, there might be planned outage events, such as complete power maintenance, that are likely to affect both sites.

Similarly, an unplanned event that impacts both sites will force recovery in Region B.

While production runs in Region B, the disk subsystems in this region track the updates that are made. When Region A is available again, assuming that all disks configured in the region come back intact, it is possible to return production back to Region A using the appropriate supplied procedure without requiring fully copying the data back. Because the updates have been tracked, only the data that changed while Region A was down are sent back to the Region A disks to bring them up to date. Then production is shut down in Region B. The final updates are allowed to be drained to Region A and production can then be restarted in Region A.

Because Region A and Region B are not symmetrically configured, the capabilities and levels of protection offered when production runs in Region B will be different. Most notably, because there is no PPRC of the production data in Region B, there is no HyperSwap protection to provide continuous data access. For the same reason, the various operational procedures for GDPS will also be different when running in Region B. However, even if no outage is planned for Region A, switching production to Region B periodically (for example, once or twice a year) and running live production there for a brief period of time is the best form of disaster testing because it will provide the best indication of whether Region B is properly configured to sustain real, live production workloads.

11.3.5 Scalability in a GDPS/MGM 3-site environment

As described in “Addressing z/OS device limits in GDPS/PPRC and GDPS/MTMM environments” on page 25, GDPS/PPRC allows defining the PPRC secondary devices in alternate subchannel set 1 (MSS1), which allows up to nearly 64 K devices to be mirrored in a GDPS/PPRC configuration. The definitions of these devices are in the application site I/O definitions.

Similarly, “Addressing z/OS device limits in a GDPS/GM environment” on page 34 describes how GM allows defining the GM FlashCopy target devices in alternate MSS1 in the recovery site I/O definitions and not defining the practice FlashCopy target devices at all to the GDPS/GM R-sys, again, allowing up to nearly 64 K devices to be mirrored in a GDPS/GM configuration.

In a GDPS/MGM 3-site environment where the PPRC secondary devices defined in MSS1 are the GM primary devices, there is additional support in GDPS/GM that allows the GM primary devices to be defined in MSS1. With the combined alternate subchannel set support in GDPS/PPRC (or GDPS/MTMM) and GDPS/GM, up to nearly 64 K devices can be replicated using the MGM technology.

11.3.6 Other considerations in a GDPS/MGM 3-site environment

With Global Mirror, deliberately underconfiguring the bandwidth provided to reduce the total cost of the solution is possible. If significant peaks exist, then this cost savings might be considerable because the network costs are often a significant portion of ongoing costs. The drawback with under-configuring bandwidth can be that this could affect the recovery point that can be achieved. If a disaster affects the entire production region, both Site1 and Site2, during any peak when the GM mirror is running behind, there is likely to be more data loss.

11.3.7 Managing the GDPS/MGM 3-site environment

GDPS provides a range of solutions for disaster recovery and continuous availability in a z Systems-centric environment. GDPS/MGM 3-site provides support for Metro Global Mirror within a GDPS environment. GDPS builds on facilities provided by System Automation and NetView and uses inband connectivity to manage the Metro Global Mirror relationships.

GDPS/MGM 3-site runs two services to manage Metro Global Mirror, both of which run on z/OS systems, as explained here:

- ▶ GDPS/PPRC (or GDPS/MTMM) services run on every z/OS image in the production sysplex and the controlling systems, K1 and K2, in Site1 and Site2. Each controlling system is allocated on its own non-mirrored disk and has access to the primary and secondary disk subsystems.

During normal operations, the master function runs in the controlling system located where the secondary disks reside. This is where the day-to-day management and recovery of the PPRC mirroring environment is performed. If Site1 or Site2 fails, the Master system manages the recovery of the PPRC disks and production systems.

- ▶ The second controlling system is an alternate that takes over the master function if the Master controlling system becomes unavailable or a Master switch takes place as result of, for example, a HyperSwap.

The GDPS/GM services run in the Kg and R-sys controlling systems. Kg runs in the production sysplex and is responsible for controlling the Global Mirror environment and sending information to the R-sys running in the recovery site. The R-sys is responsible for carrying out all recovery actions during a wide-scale disruption that affect both Site1 and Site2.

In addition to managing the operational aspects of Global Mirror, GDPS/GM provides facilities to restart z Systems production systems in the recovery site. By providing scripting facilities, it provides a complete solution for the restart of a z Systems environment in a disaster situation without requiring expert manual intervention to manage the recovery process.

GDPS supports both z Systems and distributed systems devices in a cascaded-only GDPS/MGM 3-Site environment. Only z systems devices are supported in a GDPS/MGM Multi-Target 3-site environment.

11.3.8 Flexible testing in a GDPS/MGM 3-site environment

To facilitate testing of site failover and failback processing, consider installing additional disk capacity to support FlashCopy in Site1 and Site2. The FlashCopy can be used at both Site1 and Site2 to maintain disaster recovery checkpoints during remote copy resynchronization. This ensures there is a consistent copy of the data available if a disaster-type event should occur while testing your site failover and failback procedures. In addition, the FlashCopy could be used to provide a copy to be used for testing or backing up data without the need for extended outages to production systems.

GDPS/MGM 3-site supports an additional FlashCopy disk device, referred to as *F disks*. These are additional FlashCopy target devices that might optionally be created in the recovery site. The F disks might be used to facilitate stand-alone testing of your disaster recovery procedures while the Global Mirror environment is running. This ensures that a consistent and current copy of the data is available at all times. In addition, the F disk can be used to create a “gold” or insurance copy of the data if a disaster situation occurs.

Currently, GDPS/MGM 3-site supports the definition and management of a single F device for each Metro-Global Mirror *triplet* (B, C, and CJ disk combinations) in the configuration. To reduce management and operational complexity, support exists in GDPS/GM to support the F disk without adding a requirement for these disks to be defined to the I/O configurations of the GDPS systems managing them. Known as “No UCB” FlashCopy, this support allows for the definition of F disks without the need to define additional UCBs to the GDPS management systems.

In addition to the ability to test on the F disks, GDPS/MGM 3-site configurations also support testing using X-disk support in GDPS/GM as described in 6.7.2, “Creating a test copy using GM CGPause and testing on isolated disks” on page 184.

11.3.9 GDPS Query Services in a GDPS/MGM 3-site environment

GDPS/PPRC provides Query Services, allowing you to query various aspects of the PPRC leg of a GDPS/MGM 3-site environment. Similarly, GDPS/GM provides Query Services, allowing you to query various aspects of the GM leg of a GDPS/MGM 3-site environment.

The GDPS/GM query services also have awareness of the fact that a particular environment is a GDPS/MGM 3-site environment enabled for Incremental Resynchronization (IR) and returns additional information pertaining to the IR aspects of the environment. In a GM environment, at any time, the GM session could be running from Site2 to the recovery site (B disk to C disk) or from Site1 to the recovery site (A disk to C disk). If GM is currently running B to C, this is the Active GM relationship and the A to C relationship is the Standby GM relationship. The GM query services in an MGM 3-site IR environment return information about both the active and the standby relationships for the physical and logical control units in the configuration and the devices in the configuration.

11.3.10 Prerequisites for GDPS/MGM 3-site

GDPS/MGM 3-site has the following prerequisites:

- ▶ GDPS/PPRC or GDPS/PPRC HM is required for cascaded-only GDPS/MGM 3-Site. If GDPS/PPRC HM is used, the Incremental Resynchronization function is not available. GDPS/MTMM is required for GDPS/MGM Multi-Target 3-Site.
- ▶ GDPS/GM is required and the GDPS/GM prerequisites must be met.
- ▶ Consult with your storage vendor to ensure that the required features and functions are supported on your disk subsystems.

Important: For the latest GDPS prerequisite information, see the GDPS product website:

<http://www.ibm.com/systems/z/advantages/gdps/getstarted>

11.3.11 GDPS/Active-Active disk replication integration with GDPS/MGM

It is possible to use cascaded-only GDPS/MGM 3-Site to provide local high availability and remote disaster recover for the sysplexes that run GDPS/Active-Active workloads. For such an environment, GDPS/Active-Active provides special facilities to manage the disk replication aspects of specific planned and unplanned scenarios using the GDPS/Active-Active Controller system as the single point of control. For more information, see 8.6, “GDPS/Active-Active disk replication integration” on page 267 for details.

This support is not provided with GDPS/MGM Multi-Target 3-Site.

11.4 GDPS Metro/Global Mirror 4-site solution

GDPS provides two configurations for the GDPS/MGM 4-site solution:

- ▶ The first configuration is an extension to the cascaded-only 3-site configuration described in “GDPS Metro/Global Mirror 3-site solution” on page 338, in that it is a cascaded configuration made up of a combination of GDPS/PPRC and GDPS/GM. This variation is referred to as a *cascaded-only GDPS/MGM 4-Site* configuration.
- ▶ The second configuration is an extension of the multi-target 3-Site configuration also described in “GDPS Metro/Global Mirror 3-site solution” on page 338, in that it can dynamically switch between a cascaded topology and a multi-target topology as necessary to optimize recovery scenarios such as HyperSwap. This configuration combines GDPS/MTMM with GDPS/GM, and is referred to as a *GDPS/MGM Multi-Target 4-site* configuration.

The critical difference between the 3-site configurations and the 4-site configurations, is that with the two GDPS/MGM 4-site configurations, a second copy of data is available in the recovery region that can provide a high-availability (HA) copy if you perform either a planned or unplanned switch of production to the recovery region. These 4-site configurations can also be described as *symmetrical 4-site configurations* because the same capabilities, from a data high-availability perspective, are available whether you are running your production services in Region A or Region B.

This fourth copy of data is created using non-synchronous Global Copy (also known as PPRC-XD) that can be switched to synchronous-mode during a planned or unplanned region switch, thus providing the HA copy in that region.

Figure 11-8 shows a cascaded-only MGM 4-site configuration that consists of the four copies of data, labeled A, B, C, and D. The Global Mirror FlashCopy target device (or “journal device”) is shown in Figure 11-8 as CJ.

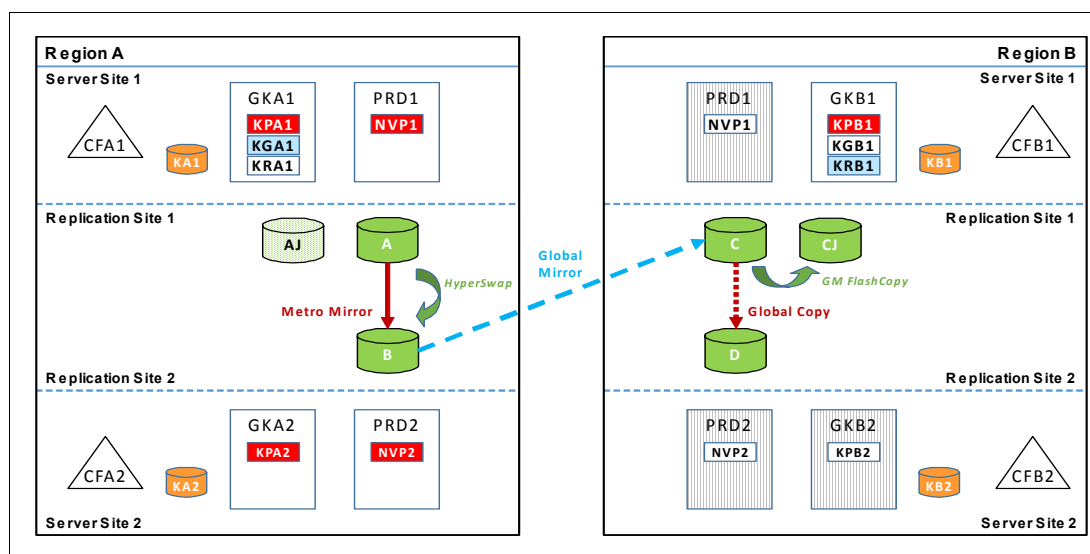


Figure 11-8 Cascaded-only GDPS/MGM 4-site configuration

In Figure 11-8, which shows a *steady state* when running in Region A, the A disk is the primary copy of data that application systems are currently accessing. The B disk is the Metro Mirror secondary disk to the A disk, and HyperSwap is enabled to provide high availability for the Region A data. This relationship is managed by GDPS/PPRC in Region A. The B disk is also the Global Mirror primary disk, being copied to the C disk that is the Global Mirror secondary disk. This is managed using GDPS/GM. Incremental resynchronization is also enabled from the A disk to the C disk to protect from a failure of the B disk and allow the Global Mirror copy to be re-established without the need for a full copy. This, as you see, is the same as a 3-site configuration. Where it differs is that the D disk is present and is a Global Copy secondary to the C disk. This relationship, which is managed by GDPS/PPRC running in Region B, can be converted to fully synchronous Metro Mirror when you perform a switch of production to Region B for whatever reason. This is referred to as an *A-B-C-D configuration*.

If you switch production to Region B, you then use the C disk as the primary copy, with the D disk now being the Metro Mirror secondary, and GM primary disk and the A disk are the GM secondaries. The B disks are then the Global Copy secondary disk to the A disk. This is referred to as a *C-D-A-B configuration*.

Figure 11-9 depicts a GDPS/MGM Multi-Target 4-site configuration when it is in a multi-target topology. As previously stated, GDPS/MGM Multi-Target 4-site configurations can dynamically switch between a cascaded topology and a multi-target topology to optimize processing of various recovery scenarios.

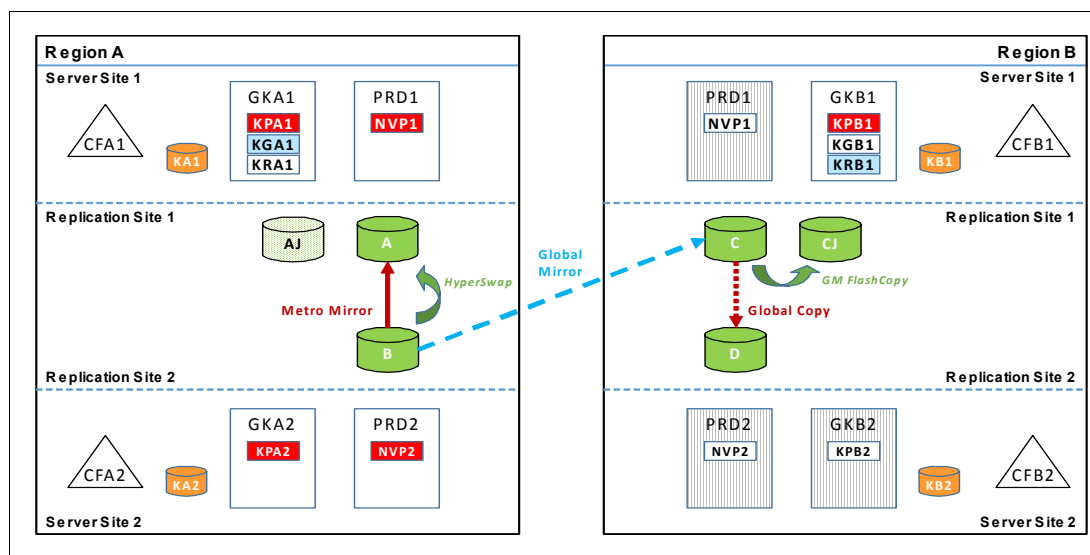


Figure 11-9 GDPS/MGM Multi-Target 4-site configuration

Assume that your GDPS/MGM Multi-Site 4-Site configuration started out in a cascaded topology as depicted in Figure 11-8 on page 347. If you execute a planned HyperSwap to the B disk, followed by a reverse and resynchronization of Metro Mirror from the B disk back to the A disk, you would wind up in the multi-target topology depicted in Figure 11-9. In the figure, the B disk is now the primary copy of data that application systems are currently accessing and the A disk is the Metro Mirror secondary disk to the B disk. HyperSwap has been reenabled to provide high availability for the Region A data. This relationship is managed by GDPS/MTMM in Region A. The B disk is also the Global Mirror primary disk, being copied to the C disk that is the Global Mirror secondary disk. This is managed using GDPS/GM. Incremental resynchronization is still enabled from the A disk to the C disk to protect from a failure of the B disk and allow the Global Mirror copy to be re-established without the need for a full copy. Finally, in Region B, the D disk is a Global Copy secondary to the C disk. Again, this relationship, which is managed by GDPS/MTMM running in Region B, can be converted to fully synchronous Metro Mirror when you perform a switch of production to Region B for whatever reason.

The advantage of the multi-target capability in this scenario is that, following the HyperSwap, Global Mirror from the B disk to the C disk can remain active, maintaining your DR position, while Metro Mirror in Region A is being resynchronized from the B disk back to the A disk. In the same situation with cascaded-only MGM 4-Site, Global Mirror from the B disk to the C disk must be suspended while Metro Mirror in Region A is being resynchronized from the B disk, back to the A disk, resulting in your DR position aging while Metro Mirror is being resynchronized.

The MGM 4-site configurations, as mentioned, remove the single point of failure of disk when you switch to the recovery region. As with GDPS/MGM 3-site, precoded procedures are provided by GDPS to manage the following scenarios in the 4-site environments:

- ▶ Moving the GM session if there is an intermediate disk subsystem failure.
- ▶ Reintroduction of the intermediate disk subsystem.
- ▶ Planned Region switch to move production to the opposite region.

However, several additional considerations exist for an MGM 4-site configuration over those previously mentioned for MGM 3-site configurations:

- ▶ DR testing can be done on the D disk (when production is in Region A) without affecting your DR position.
- ▶ Managing devices defined in an alternate subchannel set is not supported.
- ▶ The use of asymmetric devices in the remote copy configuration is not supported.
- ▶ The use of X-disk for creating a test copy is not supported (it is not required, because testing can be done on the D disk or B disk, depending on the region where production is currently running).
- ▶ Use of GDPS/PPRC HM is not supported in a 4-site configuration because the Incremental Resynchronization function is required.

11.4.1 Benefits of a GDPS/MGM 4-site configuration

You can probably see that in effect, a 4-site configuration is managed as two somewhat separate 3-site MGM configurations, where the fourth copy is most relevant when you perform a region switch, or when you want to perform a DR test. The key advantages of a 4-site MGM configuration can be summarized as follows:

- ▶ HA capability when running production in either region.
- ▶ Retention of DR capability following a region switch. In a 3-site MGM configuration your DR position ages while running on the C-disk.
- ▶ Nearly identical operational procedures when running in either region.

11.5 GDPS Metro z/OS Global Mirror 3-site solution

This section describes the capabilities and requirements of the GDPS Metro/z/OS Global Mirror (GDPS/MzGM) solution.

GDPS Metro/z/OS Global Mirror is a multi-target data replication solution that combines the capabilities of GDPS/PPRC and GDPS/XRC.

GDPS/PPRC or GDPS/PPRC HyperSwap Manager is used to manage the synchronous replication between a primary and secondary disk subsystem located either within a single data center, or between two data centers located within metropolitan distances.

GDPS/XRC is used to asynchronously replicate data from the primary disks to a third disk system in a recovery site, typically out of the local metropolitan region. Because z/OS Global Mirror (XRC) supports only CKD devices, only z Systems data can be mirrored to the recovery site.

For enterprises that want to protect z Systems data, GDPS/MzGM delivers a three-copy replication strategy to provide continuous availability for day-to-day disruptions, while protecting critical business data and functions during a wide-scale disruption.

11.5.1 GDPS/MzGM overview

The solution that is shown in Figure 11-10 is an example of a 3-site GDPS/MzGM continuous availability and DR implementation. In this example, Site1 and Site2 are running an Active/Active workload (see 3.2.3, “Multisite workload configuration” on page 72) and located within metropolitan distances to ensure optimal application performance. All data that is required to recover critical workloads is resident on disk and mirrored. Each site is configured with sufficient spare capacity to handle failover workloads if there is a site outage.

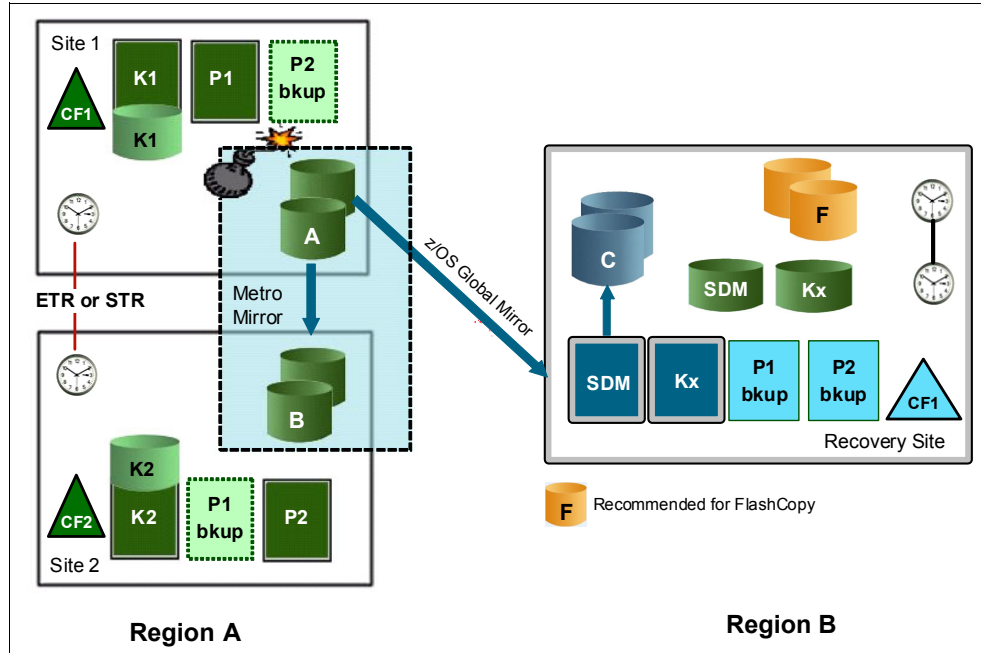


Figure 11-10 GDPS z/OS Metro Global Mirror

The third or recovery site can be located at a virtually unlimited distance from Site1 and Site2 locations to protect against regional disasters. Because of the extended distance, GDPS/XRC is used to asynchronously replicate between Site1 and the recovery site.

Redundant network connectivity is installed between Site2 and the recovery site to provide continued data protection and DR capabilities during a Site1 disaster or a failure of the disk subsystems in Site1. For more information, see “Incremental resynchronization for GDPS/MzGM” on page 351. Sufficient mainframe resources are allocated to support the SDMs and GDPS/XRC controlling system. In a disaster situation, GDPS invokes CBU to provide the additional capacity needed for production workloads.

The A disks are synchronously mirrored to the B disks in Site2 using Metro Mirror. In addition, A disks are asynchronously mirrored to a third (C) set of disks in the recovery site using z/OS Global Mirror (XRC). An optional, and highly recommended, fourth (F) set of disks in the recovery site is used to create FlashCopy of the C disks. These disks can then be used for stand-alone disaster recovery testing, or in a real disaster, to create a “gold” or insurance copy of the data. For more information about z/OS Global Mirror, see Chapter 5, “GDPS/XRC” on page 137.

Because some distance is likely to exist between the local sites, Site1 and Site2, running the PPRC leg of MzGM, and the remote recovery site, which is the XRC recovery site, we also distinguish between the local sites and the remote site using *Region* terminology. Site1 and Site2 are in one region, Region A and the remote recovery site is in another region, Region B.

Incremental resynchronization for GDPS/MzGM

With the incremental resynchronization (IR) function of Metro z/OS Global Mirror, you should be able to move the z/OS Global Mirror (XRC) primary disk location from Site1 to Site2 or vice versa without having to perform a full initial copy of all data.

Without incremental resynchronization, if Site1 becomes unavailable and the PPRC primary disk is swapped to Site2, the data at the recovery site starts to age because updates are no longer being replicated. The disaster recovery capability can be restored by establishing a new XRC session from Site2 to the recovery site. However, without incremental resynchronization, a full copy is required and this could take several hours or even days for significantly large configurations. The incremental resynch allows restoring the XRC mirror using the Site2 disks as primary and sending to the recovery site only the changes that have occurred since the PPRC disk switch.

Figure 11-11 shows how GDPS/MzGM can establish a z/OS Global Mirror session between Site2 and the recovery site when it detects that Site1 is unavailable.

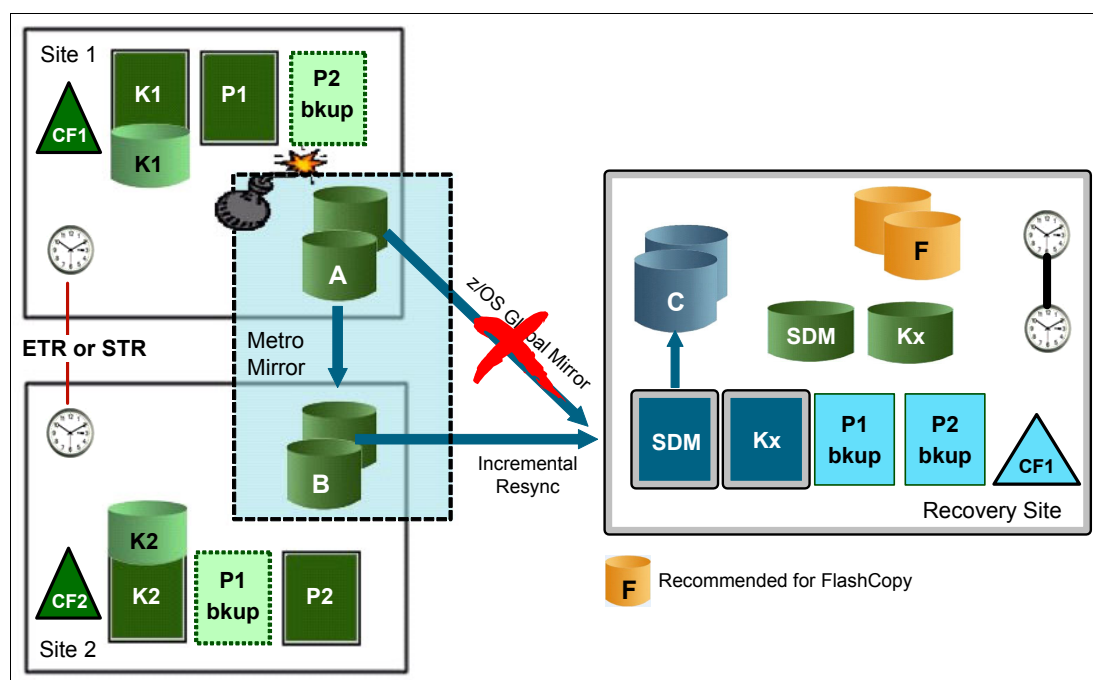


Figure 11-11 GDPS Metro z/OS Global Mirror configuration after Site1 outage

After the session is established, only an incremental resynchronization of the changed data needs to be performed, thus allowing the disaster recovery capability to be restored in minutes, instead of hours, when the intermediate site is not available. GDPS can optionally perform this resynchronization of the XRC session using the swapped-to disks totally automatically, requiring no operator intervention.

11.5.2 GDPS/MzGM Site1 failures

At the first indication of a failure, GDPS will issue a freeze command to protect the integrity of the B copy of the disk. For a more detailed description of GDPS/PPRC processing, see Chapter 3, "GDPS/PPRC" on page 53.

If the freeze event were part of a larger problem in which you could no longer use the A-disk or Site1, you must recover the B-disk and restart production applications using the B-disk. After the production systems are restarted, the business focus will be on establishing z/OS Global (XRC) mirroring between Site2 and the recovery site as soon as possible. You can perform incremental resynchronization from the B-disk to the C-disk and maintain disaster recovery readiness.

If the failure was caused by a primary disk subsystem failure, and Site1 systems are not impacted, GDPS/PPRC will use HyperSwap to transparently switch all systems in the production sysplex to the secondary disks in Site2, and the production systems will continue to run. In this case also, GDPS can perform incremental resynchronization from the B-disk to the C-disk and maintain disaster recovery readiness.

11.5.3 GDPS/MzGM Site2 failures

In this situation, the production systems in Site1 will continue to run and replication to the remote site is still running. GDPS, based on user-defined actions, will restart Site2 production systems in Site1. No action is required from an application or disaster recovery solution perspective. This scenario has less impact to the business than a failure of the Site1 location. When Site2 is recovered, if the disks have survived, an incremental resynchronization can be initiated to resynchronize the A and B disks.

11.5.4 GDPS/MzGM region switch and return home

It is possible to switch production from running in Region A (in either Site2 or Site2) to Region B. Many GDPS/MzGM customers run Site1 and Site2 in the same physical site or on a campus where these two sites are separated by little distance. In such configurations there could be planned outage events, such as complete power maintenance, that is likely to affect both sites.

Similarly, an unplanned event that impacts both sites will force recovery in Region B.

When Region A is available again, assuming that all disks configured in the region come back intact, it is possible to return production to Region A using a step-by-step procedure provided by GDPS to accomplish this *return home* operation.

To move data back to Region A, the z/OS Global Mirror (XRC) remote copy environment must be designed to allow the mirroring session to be reversed. Production will be running in Region B, and Region A will need to run the GDPS/XRC SDM systems. This means you need to ensure that the proper connectivity and resources are configured in both regions to allow them to assume the recovery region role.

Because Region A and Region B are not symmetrically configured, the capabilities and levels of protection offered when production runs in Region B will be different. Most notably, because there is no PPRC of the production data in Region B, there is no HyperSwap protection to provide continuous data access. For the same reason, the various operational procedures for GDPS will also be different when running in Region B.

However, even if no outage is planned for Region A, switching production to Region B periodically (for example, once or twice a year) and running live production there for a brief period of time is the best form of disaster testing because it will provide the best indication of whether Region B is properly configured to sustain real, live production workloads.

11.5.5 Management of the GDPS/MzGM environment

GDPS/MzGM provides management functions for a Metro/z/OS Global Mirror in a GDPS environment. The GDPS/PPRC management functions described in 11.3.7, “Managing the GDPS/MGM 3-site environment” on page 344, are also provided by GDPS/MzGM.

GDPS/XRC services run on the Kx controlling system in the recovery site along with the SDM systems. The SDM and Kx systems must be in the same sysplex. The Kx controlling system is responsible for managing the z/OS Global Mirror (XRC) remote copy process and recovering the production systems if a disaster occurs. It does not detect what is happening in Site1 and Site2.

If a wide-scale disruption that impacts both Site1 and Site2 occurs, the operator must initiate the recovery action to restart production systems in the recovery site. At this point, the Kx system activates the production LPARs and coupling facilities, and is able to respond to certain z/OS initialization messages. However, it cannot automate the complete start of the production systems. For this, the K1 or K2 systems can be used to automate the application start and recovery process in the production sysplex.

11.5.6 Flexible testing of the GDPS/MzGM environment

To facilitate testing of site failover and failback processing, consider installing additional disk capacity to support FlashCopy in Site1 and Site2. The FlashCopy can be used at both sites to maintain disaster recovery checkpoints during remote copy resynchronization. This ensures that a consistent copy of the data will be available if a disaster-type event should occur while testing your site failover and failback procedures. In addition, the FlashCopy could be used to provide a copy to be used for testing or backing up data without the need for extended outages to production systems.

By combining z/OS Global Mirror with FlashCopy, you can create a consistent point-in-time tertiary copy of the z/OS Global Mirror (XRC) data sets and secondary disks at your recovery site. The tertiary devices can then be used to test your disaster recovery and restart procedures while the GDPS/XRC sessions between Site1 and the recovery site are running, which ensures that disaster readiness is maintained at all times. In addition, these devices can be used for purposes other than DR testing; for example, nondisruptive data backup, data mining, or application testing.

With the addition of GDPS/XRC Zero Suspend FlashCopy, enterprises are able to create the tertiary copy of the z/OS Global Mirror (XRC) data sets and secondary disks without having to suspend the z/OS Global Mirror (XRC) mirroring sessions. This GDPS function prevents the SDM from writing new consistency groups to the secondary disks while FlashCopy is used to create the tertiary copy of the disks.

The time to establish the FlashCopies will depend on the number of secondary SSDs involved, the largest number of devices in any SSID, and the speed of the processor. Zero Suspend FlashCopy will normally be executed on the GDPS K-system in the recovery site, where there should be limited competition for CPU resources.

Because SDM processing is suspended while FlashCopy processing is occurring, performance problems in your production environment might occur if the SDM is suspended too long. For this reason, Zero Suspend FlashCopy should be evaluated by testing on your configuration, under different load conditions, to determine whether this facility can be used in your environment.

For enterprises that have requirements to test their recovery capabilities and maintain the currency of the replication environment, you will need to provide additional disk capacity to support FlashCopy. By providing an additional usable copy of the data, you have the flexibility to perform on-demand DR testing and other nondisruptive activities, while maintaining up-to-date DR readiness.

11.5.7 Prerequisites for GDPS/MzGM

GDPS MzGM has the following prerequisites:

- ▶ GDPS/PPRC or GDPS/PPRC HM is required.
- ▶ GDPS/XRC is required and the GDPS/XRC prerequisites must be satisfied.
- ▶ Consult with your storage vendor to ensure required features and functions are supported on your disk subsystems.

Important: For more information about the latest GDPS prerequisites, see the following GDPS product website:

<http://www.ibm.com/systems/z/advantages/gdps/getstarted>

11.6 GDPS Metro z/OS Global Mirror 4-site solution

A GDPS/MzGM 4-site configuration is an extension to the 3-site configuration described in the previous section. The critical difference from the 3-site configuration is that in the recovery region, a second copy of data is available that can provide a high-availability (HA) copy if you perform either a planned or unplanned switch of production to the recovery region. This can also be described as a *symmetrical 4-site configuration* because the same capabilities, from a data high-availability perspective, are available whether you are running your production services in Region A or Region B.

This fourth copy of data is created using synchronous PPRC and managed by GDPS/PPRC, thus providing the HA copy in that region.

Figure 11-12 shows a 4-site configuration consisting of the four copies of data, labeled A1, A2, B1, and B2.

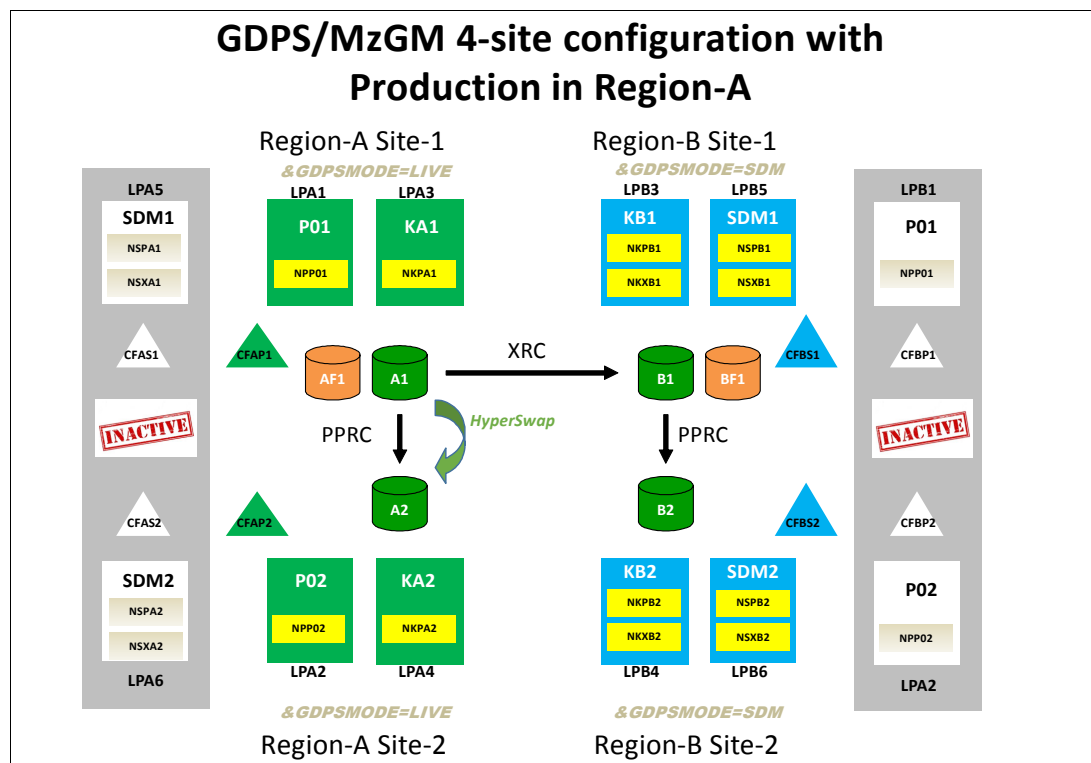


Figure 11-12 GDPS/MzGM 4-site configuration

In the figure, which shows a *steady state* when running in Region A, the A1 disk is the primary copy of data that application systems are currently accessing. The A2 disk is the Metro Mirror secondary disk to the A1 disk, and HyperSwap is enabled to provide high availability for the Region A data. This relationship is managed by GDPS/PPRC in Region A. The A1 disk is also the XRC primary disk, being copied to the B1 disk that is the XRC secondary disk. This is managed using GDPS/XRC. Incremental resynchronization is also enabled from the A2 disk to the B1 disk to allow the XRC session to be re-established, without the need for a full copy, in the event of a planned or unplanned HyperSwap in Region A.

This, as you see, is the same as a 3-site configuration. Where it differs is that the B2 disk is present and is a PPRC (aka Metro Mirror) secondary disk to the B1 disk. This relationship, which is managed by GDPS/PPRC running in Region B, is kept in a fully synchronous Metro Mirror state so that when you perform a switch of production to Region B for whatever reason, you are immediately protected by HyperSwap.

If you switch production to Region B, you then use the B1 disk as the primary copy, with the B2 disk now being the Metro Mirror secondary. The B1 disk is also the XRC primary disk with the A1 disk being the XRC secondary disk. The A2 disks are then the Metro Mirror secondary disk to the A1 disk.

Several other considerations exist for an MzGM 4-site configuration over those previously mentioned for MzGM 3-site configurations:

- ▶ DR testing can be done on the B2 disk (when production is in Region A) without affecting your DR position.
- ▶ Use of GDPS/PPRC HM is not supported in a 4-site configuration.
- ▶ xDR management for Native Linux on z Systems is not supported in a 4-site configuration.
- ▶ Managing XRC on behalf of multiple GDPS/PPRC environments is not supported in a 4-site configuration.

11.6.1 Benefits of a GDPS/MzGM 4-site configuration

You can probably see that, in effect, a 4-site configuration is managed as two somewhat separate 3-site MzGM configurations, where the fourth copy is most relevant when you perform a region switch, or when you want to perform a DR test. The key advantages of a 4-site MzGM configuration can be summarized as follows:

- ▶ HA capability when running production in either region
- ▶ Nearly identical operational procedures when running in either region



Sample continuous availability and disaster recovery scenarios

In this chapter, we describe several common client scenarios and requirements, and what we believe to be the most suitable solution for each case.

The following scenarios are described:

- ▶ A client with a single data center that has already implemented IBM Parallel Sysplex with data sharing and workload balancing wants to move to the next level of availability.
- ▶ A client with two centers needs a disaster recovery capability that will permit application restart in the remote site following a disaster.
- ▶ A client with two sites (but all production systems running in the primary site) needs a proven disaster recovery capability *and* a near-continuous availability solution.
- ▶ A client with two sites at continental distance needs to provide a disaster recovery capability.
- ▶ A client with two sites at relatively long metropolitan distance needs to provide local continuous availability and remote disaster recovery with zero data loss.
- ▶ A client who runs only IBM z/VM with Linux on z Systems guests (no z/OS in their environment) with two sites at metropolitan distance require an automated disaster recovery and near continuous availability solution.

The scenarios described in this chapter pertain to using the GDPS products that are based on hardware disk replication. The scenarios for GDPS/Active-Active using software data replication are described in Chapter 8, “GDPS/Active-Active solution” on page 231.

12.1 Introduction

In the following sections, we describe how the various GDPS service offerings can address different continuous availability (CA) and disaster recovery (DR) requirements. Because every business is unique, the following sections do not completely list all the ways the offerings can address the specific needs of your business, but they do serve to illustrate key capabilities.

In the figures accompanying the text we show minimal configurations for clarity. Many client configurations are more complex than this, but both configurations are supported.

12.2 Continuous availability in a single data center

In the first scenario, the client has only one data center, but wants to have higher availability. The client has already implemented data sharing for their critical applications, and uses dynamic workload balancing to mask the impact of outages. They already mirror all their disks within the same site but have to take planned outages when they want to switch from the primary to secondary volumes in preparation for a disk subsystem upgrade or application of a disruptive microcode patch. They are concerned that their disk is their only remaining resource whose failure can take down all their applications. The configuration is shown in Figure 12-1.

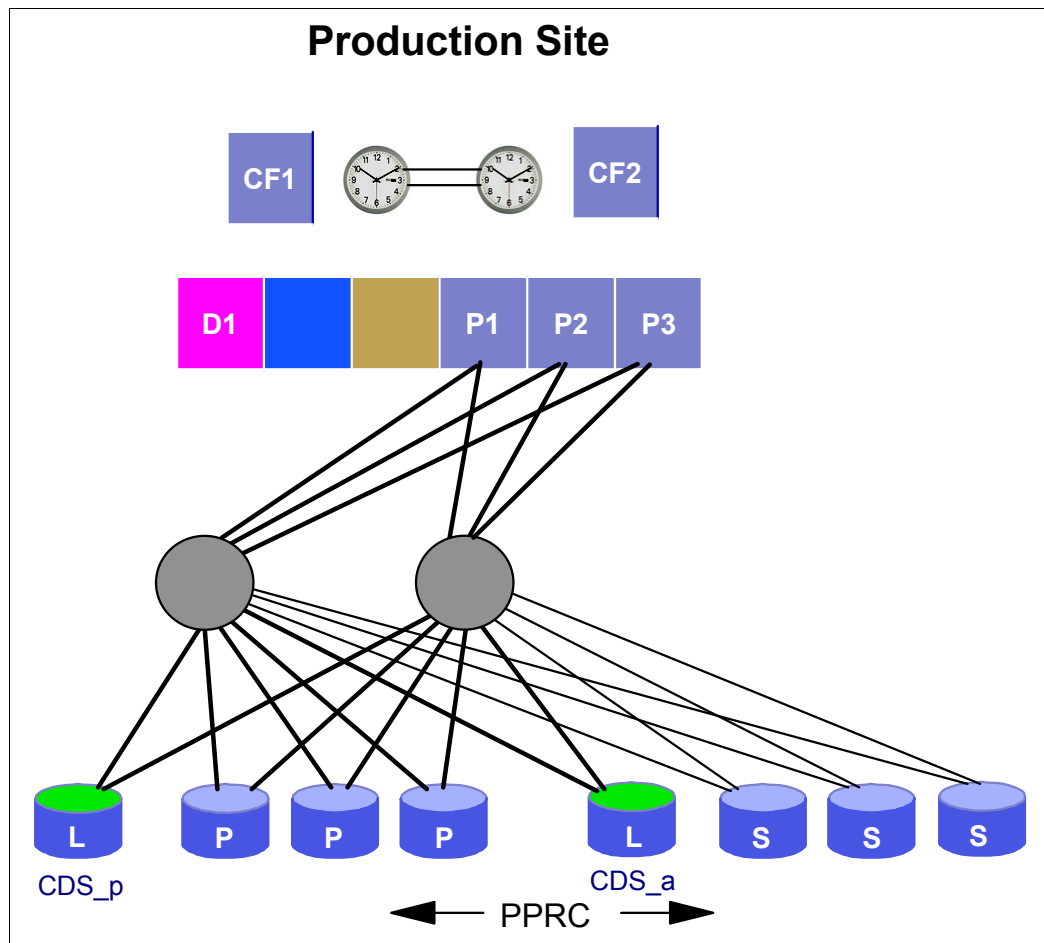


Figure 12-1 Data sharing, workload balancing, mirroring: Single site

From a disaster recovery perspective, the client relies on full volume dumps. Finding a window of time that is long enough to create a consistent set of backups is becoming a challenge. In the future, they plan to have a second data center, to protect them in case of a disaster. And in the interim, they want to investigate the use of FlashCopy to create a consistent set of volumes that they can then dump in parallel with their batch work. But their current focus is on improved resiliency within their existing single center.

Table 12-1 lists the client's situation and requirements, and shows which of those requirements can be addressed by the most suitable GDPS offering for this client's requirements, namely GDPS/PPRC HyperSwap Manager.

Table 12-1 Mapping client requirements to GDPS/PPRC HyperSwap Manager attributes

Attribute	Supported by GDPS/PPRC HM
Single site	Y
Synchronous remote copy support	Y (PPRC)
Transparent swap to secondary disks	Y (HyperSwap)
Ability to create a set of consistent tape backups	Y ^a
Ability to easily move to GDPS/PPRC in the future	Y

a. To create a consistent source of volumes for the FlashCopy in GDPS/PPRC HyperSwap Manager, you must create a freeze-inducing event and be running with a Freeze and Go policy

This client has a primary short-term objective to be able to provide near-continuous availability, but wants to ensure that they address that in a strategic way.

In the near term, they need the ability to transparently swap to their secondary devices in case of a planned or unplanned disk outage. Because they have only a single site, do not currently have a TS7700, and do not currently have the time to fully implement GDPS system and resource management, the full GDPS/PPRC offering is more than they currently need.

By implementing GDPS/PPRC HyperSwap Manager, they can achieve their near-term objectives in a manner that positions them for a move to full GDPS/PPRC in the future.

Figure 12-2 shows the client configuration after implementing GDPS/PPRC HyperSwap Manager. Now, if they have a failure on the primary disk subsystem, the controlling system will initiate a HyperSwap, transparently switching all of the systems in the GDPS sysplex over to what were previously the secondary volumes. The darker lines connecting the secondary volumes in the figure indicate that the processor-to-control unit channel capacity is now similar to that used for the primary volumes.

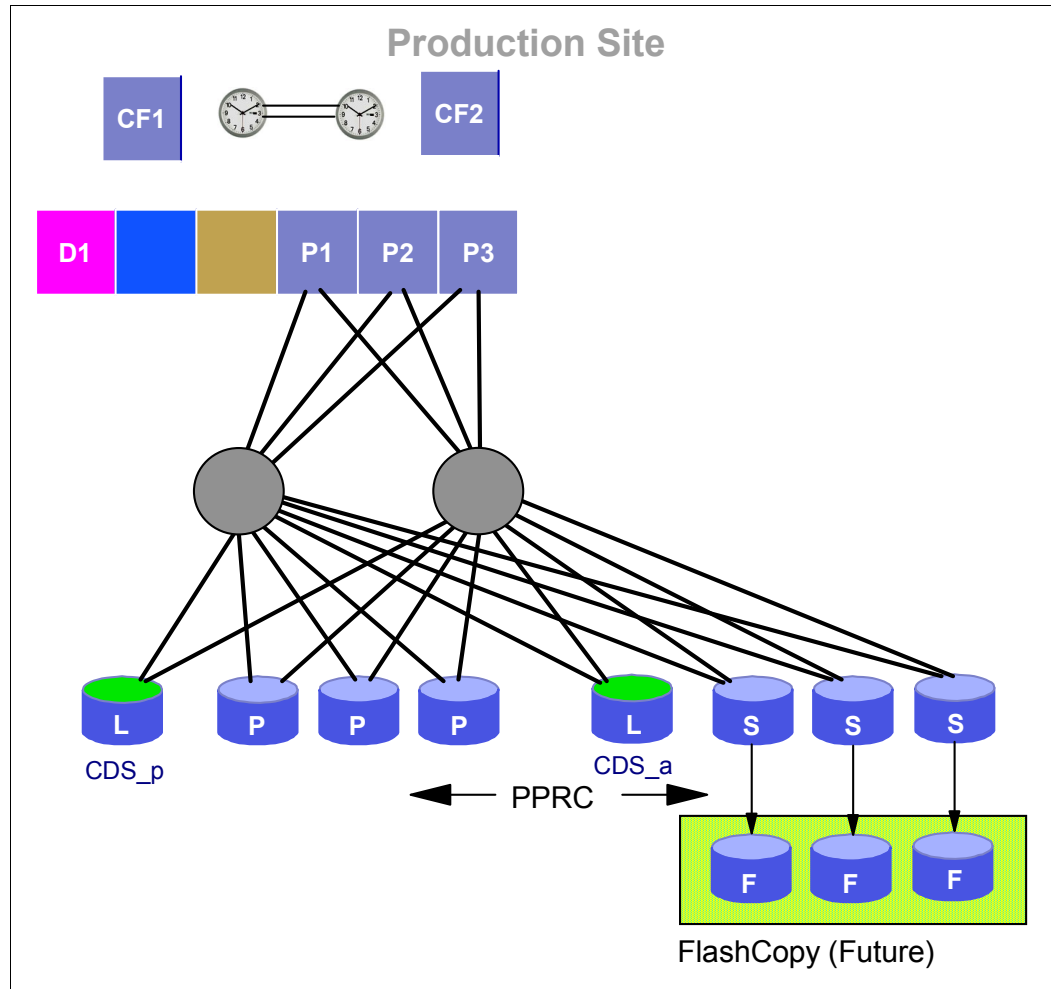


Figure 12-2 Continuous availability within a single data center

After the client has implemented GDPS and enabled the HyperSwap function, their next move will be to install the additional disk capacity so it can use FlashCopy. The client will then be able to use the Freeze function to create a consistent view that can be flash-copied to create a set of volumes that can then be full-volume dumped for disaster recovery. This will create a more consistent set of backup tapes than the client has today (because today it is backing up a running system) and the backup window will now be only a few seconds rather than the hours it takes that it currently takes. This enables the client to make more frequent backups.

12.3 DR across two data centers at metro distance

The next scenario relates to a client that is under pressure to provide a disaster recovery capability in a short time frame, perhaps for regulatory reasons. The client has a second data center within metropolitan distances and suitable for synchronous mirroring, but has not yet implemented mirroring between the sites. Before moving to a full GDPS/PPRC environment, the client was going to complete their project to implement data sharing and workload balancing. However, events have overtaken them and they now need to provide the disaster recovery capability sooner than they had expected.

The client can select between the full GDPS/PPRC offering, as they had planned to do in the long term, or to install GDPS/PPRC HyperSwap Manager now. Because they will not be using the additional capabilities delivered by GDPS/PPRC in the immediate future, the client decides to implement the lower-cost GDPS/PPRC HyperSwap Manager option. Table 12-2 summarizes the client's situation and requirements and shows how those requirements can be addressed by GDPS/PPRC HyperSwap Manager.

Table 12-2 Mapping client requirements to GDPS/PPRC HyperSwap Manager attributes

Attribute	Supported by GDPS/PPRC HM
Two sites, 12 km apart	Y
Synchronous remote copy support	Y (PPRC)
Maintain consistency of secondary volumes	Y (Freeze)
Maintain consistency of secondary volumes during PPRC resynch	Y ^a (FlashCopy)
Ability to move to GDPS/PPRC in the future	Y

a. FlashCopy is used to create a consistent set of secondary volumes before a resynchronization, following a suspension of remote copy sessions.

This client needs to be able to quickly provide a disaster recovery capability. The primary focus in the near term, therefore, is to be able to restart its systems at the remote site as though it was restarting off the primary disks following a power failure. Longer term, however, the RTO (which is the time to get the systems up and running again in the remote site) will be reduced to the point that it can no longer be achieved without the use of automation (this will be addressed by a move to GDPS/PPRC). The client also has a requirement to have a consistent restart point at *all* times (even during DR testing).

This client will implement GDPS/PPRC HyperSwap Manager, with the controlling system in the primary site and the secondary disks in the remote site. The auxiliary storage subsystems are configured with sufficient capacity to be able to use FlashCopy for the secondary devices; this will allow the client to run DR tests without impacting its mirroring configuration.

GDPS/PPRC HyperSwap Manager will be installed and the Freeze capability enabled. After the Freeze capability is enabled and tested, the client will install the additional intersite channel bandwidth required to be able to HyperSwap between the sites. This configuration is shown in Figure 12-3. Later, in preparation for a move to full GDPS/PPRC, the client will move the controlling system (and its disks) to the remote site.

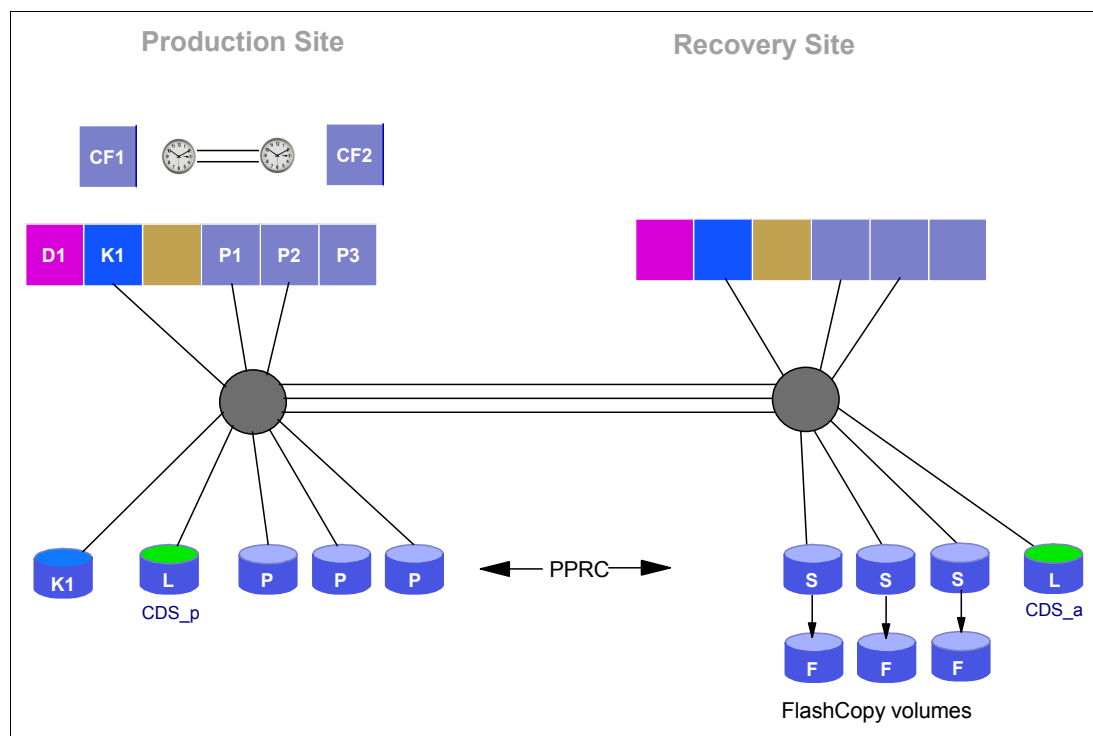


Figure 12-3 GDPS/PPRC 2-site HM configuration

12.4 DR and CA across two data centers at metro distance

The client in this scenario has two centers within metro distance of each other. The client already uses PPRC to remote copy the primary disks (both CKD and FB) to the second site. They also have the infrastructure in place for a cross-site sysplex; however, all production work still runs in the systems in the primary site.

The client is currently implementing data sharing, along with dynamic workload balancing, across their production applications. In parallel with the completion of this project, they want to start looking at how the two sites and their current infrastructure can be maximized to provide disaster recovery and continuous or near-continuous availability in planned and unplanned outage situations, including the ability to dynamically switch the primary disks back and forth between the two sites.

Because the client is already doing remote mirroring, their first priority is to ensure that the secondary disks provide the consistency to allow restart, rather than recovery, in case of a disaster. Because of pressure from their business, the client wants to move to a zero (0) data loss configuration as quickly as possible, and also wants to investigate other ways to reduce the time required to recover from a disaster.

After the disaster recovery capability has been tested and tuned, the client's next area of focus will be continuous availability, across both planned and unplanned outages of applications, systems, and complete sites.

This client is also investigating the use of z/VM and Linux on z Systems to consolidate several of their thousands of PC servers onto the mainframe. However, this is currently a lower priority than their other tasks.

Because of the disaster recovery and continuous availability requirements of this client, together with the work they have already done and the infrastructure in place, the GDPS offering for them is GDPS/PPRC. Table 12-3 shows how this offering addresses this client's needs.

Table 12-3 Mapping client requirements to GDPS/PPRC attributes

Attribute	Supported by GDPS/PPRC
Two sites, 9 km apart	Y
Zero data loss	Y (PPRC with Freeze policy of SWAP, STOP)
Maintain consistency of secondary volumes	Y (Freeze)
Maintain consistency of secondary volumes during PPRC resynch	Y ^a (FlashCopy)
Remote copy and remote consistency support for FB devices	Y (Open LUN support)
Ability to conduct DR tests without impacting DR readiness	Y (FlashCopy)
Automated recovery of disks and systems following a disaster	Y (GDPS script support)
Ability to transparently swap z/OS disks between sites transparently	Y (HyperSwap)
DR and CA support for Linux guests under z/VM	Y

a. FlashCopy is used to create a consistent set of secondary volumes before a resynchronization, following a suspension of remote copy sessions.

Although this client has performed a significant amount of useful work already, fully benefiting from the capabilities of GDPS/PPRC will take a significant amount of time, so the project is divided into the following steps:

1. Install GDPS/PPRC, define the remote copy configuration to GDPS, and start using GDPS to manage and monitor the configuration.

This will make it significantly easier to implement changes to the remote copy configuration. Rather than issuing many PPRC commands, the GDPS configuration definition simply needs to be updated and activated, and the GDPS panels then used to start the new remote copy sessions.

Similarly, any errors in the remote copy configuration will be brought to the operator's attention using the NetView SDF facility. Changes to the configuration, to stop or restart sessions, or to initiate a FlashCopy, are far easier using the NetView interface.

2. After the staff becomes familiar with the remote copy management facilities of GDPS/PPRC, enable the Freeze capability, initially as PPRCFAILURE=GO and then moving to PPRCFAILURE=COND or STOP when the client is confident with the stability of the remote copy infrastructure. Because HyperSwap will not be implemented immediately, they will specify a PRIMARYFAILURE=STOP policy to avoid data loss if recovery on the secondary disks becomes necessary after a primary disk problem.

Although the client has PPRC today, they do not have the consistency on the remote disks that is required to perform a restart rather than a recovery following a disaster. The GDPS Freeze capability will add this consistency, and enhance it with the ability to ensure zero (0) data loss following a disaster when a PPRCFAILURE=STOP policy is implemented.

3. Add the FB disks to the GDPS/PPRC configuration, including those devices in the Freeze group, so that all mirrored devices will be frozen in case of a potential disaster. As part of adding the FB disks, a second controlling system will be set up¹.

Although the client does not currently have distributed units of work that update both the z/OS *and* FB disks, the ability to Freeze all disks at the same point in time makes cross-platform recovery significantly simpler.

In the future, if the client implements applications that update data across multiple platforms inside the scope of a single transaction, the ability to have consistency across *all* disks will move from being “nice to have” to a necessity.

4. Implement GDPS Sysplex Resource Management to manage the sysplex resources within the GDPS, and start using the GDPS Standard actions panels.

GDPS system and sysplex management capabilities are an important aspect of GDPS. They ensure that all changes to the configuration conform to previously prepared and tested rules, and that everyone can check at any time to see the current configuration, that is, which sysplex data sets and IPL volumes are in use. These capabilities provide the logical equivalent of the whiteboard used in many computer rooms to track this type of information.

5. Implement the GDPS Planned and Unplanned scripts to drive down the RTO following a disaster.

The GDPS scripting capability is key to recovering the systems in the shortest possible time following a disaster. Scripts run at machine speeds, rather than at human speeds. They can be tested over and over until they do precisely what you require. And they will always behave in exactly the same way, providing a level of consistency that is not possible when relying on humans.

However, the scripts are not limited to disaster recovery. This client sometimes has outages as a result of planned maintenance to its primary site. Using the scripts, they can use HyperSwap to keep its applications available as it moves its systems one by one to the recovery site in preparation for site maintenance, and then back to the normal locations after maintenance is complete.

Because all production applications will still be running in the production site at this time, the processor in the second site is much smaller. However, to enable additional capacity to quickly be made available in case of a disaster, the processor has the CBU feature installed. The GDPS scripts can be used to automatically enable the additional CBU engines as part of the process of moving the production systems to the recovery processor.

6. After the disaster recovery aspect has been addressed, HyperSwap will be implemented to provide a near-continuous availability capability for the z/OS systems. A controlling system should be set up in each site when using HyperSwap to ensure a system is always available to initiate a HyperSwap regardless of where the primary disks might be at that time. In the case of this client, they had already set up the second controlling system when they added the FB devices to the GDPS configuration.

The client will use both planned HyperSwap (to move their primary disks before planned maintenance on the primary subsystems) and unplanned HyperSwap (allowing the client to continue processing across a primary subsystem failure). They will test planned HyperSwap while their Primary Failure policy option is still set to STOP. However, when they are comfortable and ready, they will change to running with a PRIMARYFAILURE=SWAP, STOP policy to enable unplanned HyperSwap.

¹ Only the GDPS controlling systems can see the FB disks. Therefore, a second controlling system is recommended to ensure the FB disks can always be managed even if a controlling system is down for some reason.

7. Finally, and assuming that the consolidation onto Linux on z Systems has proceeded, the heterogeneous disaster recovery capability will be implemented to manage z/VM systems and its guests and to add planned and unplanned HyperSwap support for z/VM and the Linux guests.

Although the ability to transparently swap FB devices using HyperSwap is not available for z/VM guest Linux systems using FB disks, it is still possible to manage PPRC for these disks. GDPS will provide data consistency and will perform the physical swap, and can manage the re-IPL on the swapped-to disks.

z/VM systems hosting Linux guests using CKD disks will be placed under GDPS xDR control, providing them with near-equivalent management to what is provided for z/OS systems in the sysplex, including planned and unplanned HyperSwap.

And because it is all managed by the same GDPS, the swap can be initiated as a result of a problem on a z/OS disk, meaning that you do not have to wait for the problem to spread to the Linux disks before the swap is initiated. Equally, a problem on a CKD Linux disk can result in a HyperSwap of the Linux disks and the z/OS disks.

The projected final configuration is shown in Figure 12-4 (for clarity, we have not included the Linux components in the figure).

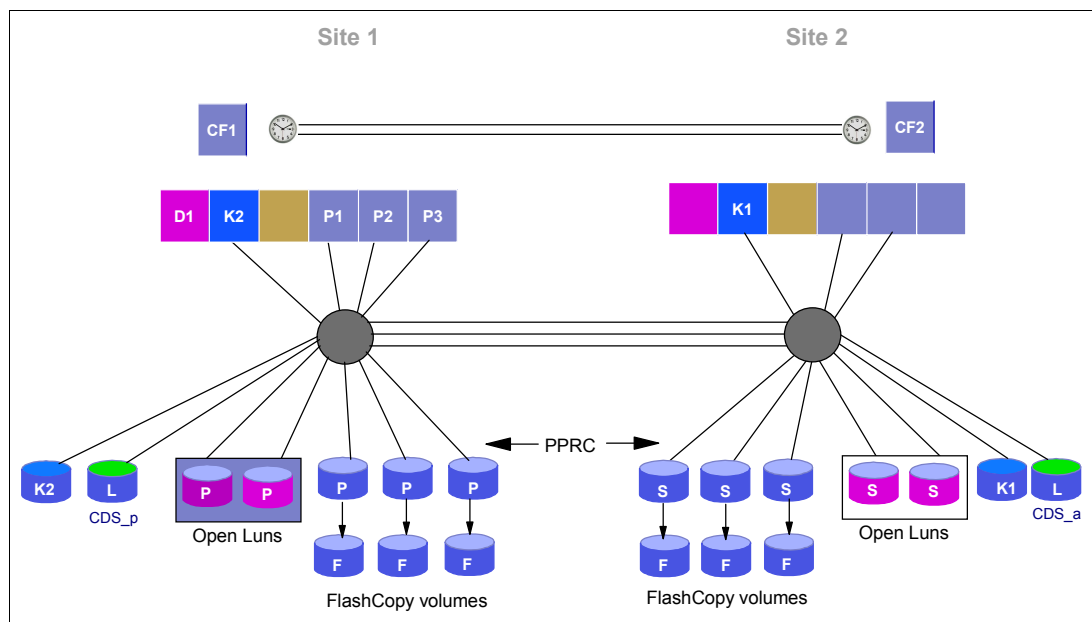


Figure 12-4 Active/standby workload GDPS/PPRC configuration

12.4.1 Active/active workload

As mentioned, this client is in the process of enabling all its applications for data sharing and dynamic workload balancing. This project will proceed in parallel with the GDPS project. When the critical applications have been enabled for data sharing, the client plans to move to an Active/Active workload configuration, with several production systems in the primary site and others in the recovery site.

To derive the maximum benefit from this configuration, it most likely is possible to transparently swap from the primary to secondary disks. Therefore, it is expected that the move to an Active/Active workload will not take place until after HyperSwap is enabled.

The combination of multisite data sharing and HyperSwap means that the client's applications will remain available across outages affecting a software subsystem (DB2, for example), an operating system, a processor, a coupling facility, or a disk subsystem (primary or secondary). The only event that can potentially result in a temporary application outage is an instantaneous outage of all resources in the primary site; this can result in the database managers in the recovery site having to be restarted.

The move to an Active/Active workload might require creating minor changes to the GDPS definitions, several new GDPS scripts, and modifications to existing ones, depending on whether new systems will be added or some of the existing ones moved to the other site. Apart from that, however, there is no fundamental change in the way GDPS is set up or operated.

12.5 DR and CA across two data centers at metro distance for z/VM and Linux on z Systems only

The client in this scenario runs their main production work on Linux on z Systems, which run as z/VM guests. The production data resides on CKD disks. The critical workloads are running on four z/VM systems. Two of the z/VM systems run in one site and the other two in the other site. They also have a couple of other, less important production z/VM systems running Linux guests. The z Systems server in each site is configured with IFL engines only (no general-purpose CPs), and the client has no z/OS systems or skills. They have two centers within metro distance of each other. The client already uses PPRC to remote copy the primary disks to the second site. They also have the infrastructure and connectivity in place for the SSI cluster.

The disk environment is well-structured. Although the various z/VM systems share a physical disk subsystem, the disks for each of the z/VM systems are isolated at an LSS level.

Because the client is already doing remote mirroring, their first priority is to ensure that the secondary disks provide the consistency to allow restart in case of a disaster, rather than recovery. Because of pressure from their business, the client wants to move to a zero (0) data loss configuration as quickly as possible, and also wants to investigate ways to reduce the time required to recover from a disaster.

There are also regulatory pressures that force the client to periodically demonstrate that they can run their production workload in either site for an extended period of time. Therefore, they also need to have processes to perform planned workload moves between sites as automatically and as fast as possible with minimum operator intervention.

Because of the disaster recovery and continuous availability requirements of this client, together with the work that they have already done and the infrastructure that is in place, the GDPS offering for them is the GDPS Virtual Appliance. Table 12-4 shows how this offering addresses this client's needs.

Table 12-4 Mapping client requirements to GDPS/PPRC attributes

Attribute	Supported by GDPS Virtual Appliance
Two sites, 9 km apart	Y
Zero data loss	Y (PPRC with Freeze policy of SWAP, STOP)
Maintain consistency of secondary volumes	Y (Freeze)
Maintain consistency of secondary volumes during PPRC resynch	Y ^a (FlashCopy)
Remote copy and remote consistency support for FB devices	Y (Open LUN support)
Ability to conduct DR tests without impacting DR readiness	Y (FlashCopy)
Automated recovery of disks and systems following a disaster	Y (GDPS script support)
Ability to transparently swap z/VM (and guest) disks between sites transparently	Y (HyperSwap)
DR and CA support for Linux guests under z/VM	Y
Ability to automate planned move of systems between sites	Y (Script support)
z/OS skills not required	Y

a. FlashCopy is used to create a consistent set of secondary volumes prior to a resynchronization, following a suspension of remote copy sessions.

Although this client has performed a significant amount of useful work already, fully benefiting from the capabilities of the GDPS Virtual Appliance, they are concerned about enabling appliance management for their entire production environment all at once. Because they have their disks isolated in separate LSSes for the SSI and the stand-alone z/VM systems, the following phasing-in of the function is possible:

1. Install a general-purpose CP engine on the Site2 z Systems server to run the GDPS Virtual Appliance².
2. Install the GDPS Virtual Appliance to initially manage one of the stand-alone z/VM systems and the data for this system. Start with the least critical system.
 Define the remote copy configuration to GDPS, and start by using GDPS to manage and monitor the configuration for the first z/VM system.
 In this limited implementation, the client can test all aspects of the GDPS Virtual Appliance, isolated from their more important systems. They can code and test scripts, exercise Freeze, planned and unplanned HyperSwap, refine their operational procedures and prepare for cutover of their more important z/VM systems.
3. After the staff becomes familiar with the appliance, the client can then put the second z/VM system and the disks of this system under appliance management. They can perform more tests in this environment to understand how the appliance works when there are multiple systems under its control and make final preparations for moving the SSI environment to be under appliance control.

² The option to purchase a z Systems general-purpose CP engine for clients that require one is included in the GDPS Virtual Appliance deal.

4. Finally, the client will add the 4-way SSI into the appliance managed environment, perform some more tests and finalize their implementation.

Although the client does not currently have distributed units of work that update both the z/OS *and* FB disks, the ability to Freeze all disks at the same point in time makes cross-platform recovery significantly simpler.

In the future, if the client implements applications that update data across multiple platforms within the scope of a single transaction, the ability to have consistency across *all* disks will move from being “nice to have” to a necessity.

5. After all systems are under GDPS control, the client can schedule a test to move their workload to all run in Site2 using the Site2 disks. Primary disk role will be swapped to Site2 using planned HyperSwap thus making the move transparent to the systems that were already running in Site2. The PPRC mirror will be reversed to run from Site2 disks toward the Site1 disks in order to retain unplanned HyperSwap capability while the workload is running in Site2. The systems running in Site1 will be stopped and re-IPLed in Site2 after the disks are swapped. A single planned action script will be used to perform this move, minimizing operator intervention and the time required to execute the entire process.

Similarly, a planned action script will be used to move the systems back to their “normal” locations.

The first time that the client does this exercise, they will isolate to run production in Site2 over a weekend period, returning to normal before Monday morning. However, using the same process and scripts, they will eventually schedule moves where they remain in Site2 for a longer period of time.

12.6 Local CA and remote DR across two data centers at long metropolitan distance

The client has two datacenters (Site1 and Site2) at 100 km distance. They run all their systems in Site1 and Site2 is their disaster recovery location.

They already use PPRC to mirror their data to Site2 and have GDPS/PPRC (single-site workload) implemented to manage the environment. They use GDPS/PPRC with a Freeze and Stop policy because they have a requirement for zero data loss (RPO=0). However, they have enabled this environment for unplanned swaps because of the long distance between the sites. Also, because they do not have sufficient cross-site channel bandwidth between, they cannot run production with their systems in Site1 using the disks in Site2. The reason that they have HyperSwap enabled is so they can do a graceful shutdown of their systems. After the systems are shut down, they move production to Site2.

The client has a large number of mirrored devices and defines their PPRC secondary devices in an alternate subchannel that is set to mitigate their UCB constraint. They have FlashCopy devices in Site2, which they use for periodic DR validation testing.

The fact that they are unable to fully benefit from HyperSwap means that disk failure is a single point of failure for their sysplex, so they need to invoke DR for a disk failure, a single component failure. They have a requirement to eliminate this single point of failure by providing a local PPRC mirrored copy of the data, which will give them the full benefit of HyperSwap.

They are due for a disk technology refresh and would like to take advantage of this activity to add a local copy of the disk for CA.

Whatever solution they chose, the client must not be exposed, from a DR risk perspective, while implementing the solution.

Given their requirement for local PPRC and HyperSwap, they need to decide how to also protect their data for DR purposes. Although using XRC or GM in conjunction with the local PPRC mirror in an MGM or MzGM 3-site configuration could be an option, with XRC or GM, they cannot achieve zero data loss that is an absolute requirement for their business. MTMM can provide them with a synchronous mirror, both locally and in the remote data center, and meet their zero data loss requirement.

Another key consideration the client has is the skills that they have already built in using GDPS/PPRC as their DR solution. Although they understand that a new topology with an extra copy of data will necessitate changes, they would like to avoid reinventing the wheel and using a radically different solution that voids their investment in the GDPS technology. They would like the solution to be phased in.

GDPS/MTMM is the ideal solution for this client. The MTMM copy technology meets their requirements for local CA and remote DR with minor additional skill requirement and their existing PPRC mirror can remain functional during the upgrade from GDPS/PPRC to GDPS/MTMM. In Table 12-5, we show how GDPS/MTMM can meet the client's requirements. The client is already using GDPS/PPRC, so they need to select a solution that provides all of the benefits of GDPS/PPRC and meets their additional requirements.

Table 12-5 Mapping client requirements to GDPS/MTMM attributes

Attribute	Supported by GDPS/MTMM
Two sites, 100 km apart	Y
Zero data loss	Y (Freeze policy with STOP)
Maintain consistency of secondary volumes	Y (Freeze)
Local CA and remote DR	Y (MTMM technology)
Ability to conduct DR tests without impacting DR readiness	Y (FlashCopy)
Automated recovery of disks and systems following a disaster	Y (GDPS script support)
Ability to transparently swap z/OS disks between the local copies of data transparently	Y (HyperSwap, preferred leg)
Ability to transparently swap z/OS disks between the one of the Site1 copies and the Site2 copy transparently to facilitate orderly shutdown	Y (HyperSwap, non-preferred leg)
Support for a single PPRC leg (Site1-Site2) to facilitate a phased migration to the new topology	Y
Protect investment in GDPS/PPRC and GDPS/PPRC skills	Y
Maintain existing Site1-Site2 mirror while adding local mirror	Y

The client can plan for the following high-level steps when moving their GDPS/PPRC environment to a GDPS/MTMM environment:

1. Refresh the existing GDPS/PPRC Site1 and Site2 disks with new technology disks that support the MTMM technology. This is a process that clients are fairly familiar with already. Often, it can be achieved nondisruptively using HyperSwap or TDMF technologies. At this time, the client will also acquire the third set of disks that will be installed locally.

2. Upgrade GDPS/PPRC to GDPS/MTMM. This will initially be a GDPS/MTMM with a single replication leg which is the client's existing GDPS/PPRC mirror. GDPS/MTMM in a single leg configuration will function very similar to GDPS/PPRC with some minor differences. At this point, the client has the same protection and capabilities that they had with GDPS/PPRC. The procedural changes required to accomplish this implementation step are quite minor since the overall topology of their mirror has not changed. The client will have to adjust some of their GDPS scripts and operational procedures, but this will not be a major change.
3. Finalize the implementation by adding the second, local replication leg to the GDPS/MTMM configuration. This step will, again, require some modifications to the client's existing GDPS automation scripts as well as addition of some new scripts since the new topology with two replication legs can now cater to additional planned and unplanned outage scenarios. The operational procedures will also need to be changed in parallel. Because the client has somewhat familiarized themselves with the high-level differences between GDPS/PPRC and GDPS/MTMM while running in the single-leg configuration, this second step will not be a radical change from a skills perspective. With the accomplishment of this step, the client will meet all of their requirements.

12.7 DR in two data centers, global distance

The client in this scenario has a data center in Asia and another in Europe. Following the tsunami disaster in 2004, the client decides to remote copy their production sysplex data to their data center in Europe. The client is willing to accept the small data loss that will result from the use of asynchronous remote copy.

However, there is a requirement that the data in the remote site is consistent, to allow application restart. In addition, to minimize the restart time, the solution must provide the ability to automatically recover the secondary disks and restart all the systems. The client has about 10000 primary volumes that they want to mirror. The disks in the Asian data center are IBM, but those in the European center that will be used as the secondary volumes are currently non-IBM.

The most suitable GDPS offering for this client is GDPS/XRC. Because of the long distance between the two sites (approaching 15000 km), using a synchronous remote copy method is out of the question. Because the disks in the two data centers are from a different vendor, GDPS/GM is also out of the question. Table 12-6 shows how the client's configuration and requirements map to the capabilities of GDPS/XRC.

Table 12-6 Mapping client requirements to GDPS/XRC attributes

Attribute	Supported by GDPS/XRC
Two sites, separated by thousands of km	Y
Willing to accept small data loss	Y (actual amount of data loss will depend on several factors, most notably the available bandwidth)
Maintain consistency of secondary volumes	Y
Maintain consistency of secondary volumes during resynch	Y ^a (FlashCopy)
Over 10000 volumes	Y (use coupled SDM support)
Requirement for data replication for and between multiple storage vendors products	Y

Attribute	Supported by GDPS/XRC
Only z/OS disks need to be mirrored	Y
Automated recovery of disks and systems following a disaster	Y(GDPS script support)

- a. FlashCopy is used to create a consistent set of secondary volumes before a resynchronization, following a suspension of remote copy sessions.

The first step for the client is to size the required bandwidth for the XRC links. This information will be used in the tenders for the remote connectivity. Assuming the cost of the remote links is acceptable, the client will start installing GDPS/XRC concurrently with setting up the remote connectivity.

Pending the availability of the remote connectivity, three LPARs will be set up for XRC testing (two SDM LPARs, plus the GDPS controlling system LPAR). This will allow the systems programmers and operators to become familiar with XRC and GDPS operations and control. The addressing of the SDM disks can be defined and agreed to and added to the GDPS configuration in preparation for the connectivity being available.

The final configuration is shown in Figure 12-5. The GDPS systems are in the same sysplex and reside on the same processor as the European production systems. In case of a disaster, additional CBU engines on that processor will automatically be activated by a GDPS script during the recovery process.

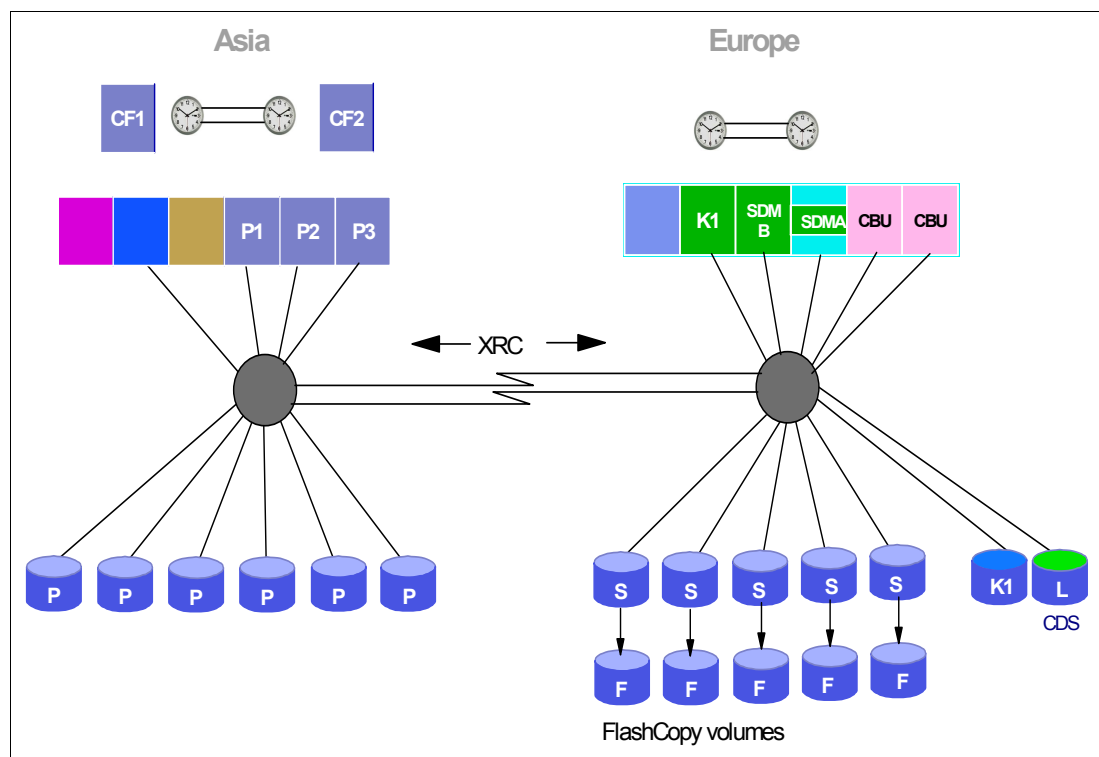


Figure 12-5 Final GDPS/XRC configuration

12.8 Other configurations

There are many other combinations of configurations. However, we believe that the examples provided here cover the options of one or two sites, short and long distance, and continuous

availability and disaster recovery requirements. If you feel that your configuration does not fit into one of the scenarios described here, contact your IBM representative for more information about how GDPS can address your needs.

Glossary

A

AOM. asynchronous operations manager.

application system. A system made up of one or more host systems that perform the main set of functions for an establishment. This is the system that updates the primary disk volumes that are being copied by a copy services function.

asynchronous operation. A type of operation in which the remote copy XRC function copies updates to the secondary volume of an XRC pair at some time after the primary volume is updated. Contrast with synchronous operation.

B

backup. The process of creating a copy of data to ensure against accidental loss.

C

cache. A random access electronic storage in selected storage controls used to retain frequently used data for faster access by the channel.

central processor complex (CPC). The unit within a cluster that provides the management function for the storage server. It consists of cluster processors, cluster memory, and related logic.

channel connection address (CCA). The input/output (I/O) address that uniquely identifies an I/O device to the channel during an I/O operation.

channel interface. The circuitry in a storage control that attaches storage paths to a host channel.

consistency group time. The time, expressed as a primary application system time-of-day (TOD) value, to which XRC secondary volumes have been updated. This term was previously referred to as *consistency time*.

consistent copy. A copy of a data entity (for example a logical volume) that contains the contents of the entire data entity from a single instant in time.

control unit address. The high-order bits of the storage control address, used to identify the storage control to the host system.

D

dark fibre. A dedicated fibre link between two sites that is dedicated to use by one client.

DASD. direct access storage device.

data in transit. The update data on application system DASD volumes that is being sent to the recovery system for writing to DASD volumes on the recovery system.

data mover. See system data mover.

device address. The ESA/390 term for the field of an ESCON device-level frame that selects a specific device on a control unit image. The one or two leftmost digits are the address of the channel to which the device is attached. The two rightmost digits represent the unit address.

device number. The ESA/390 term for a four-hexadecimal-character identifier, for example 13A0, that you associate with a device to facilitate communication between the program and the host operator. The device number that you associate with a subchannel.

Device Support Facilities program (ICKDSF). A program used to initialize DASD at installation and perform media maintenance.

DFDSS. Data Facility Data Set Services is an IBM licensed program to copy, move, dump, and restore data sets and volumes.

DFSMSdss. A functional component of DFSMS/MVS used to copy, dump, move, and restore data sets and volumes.

disaster recovery (DR). Recovery after a disaster, such as a fire, that destroys or otherwise disables a system. Disaster recovery techniques typically involve restoring data to a second (recovery) system, then using the recovery system in place of the destroyed or disabled application system. Also see *recovery*, *backup*, and *recovery system*.

dual copy. A high availability function made possible by the nonvolatile storage in cached IBM storage controls. Dual copy maintains two functionally identical copies of designated DASD volumes in the logical storage subsystem, and automatically updates both copies every time a write operation is issued to the dual copy logical volume.

duplex pair. A volume composed of two physical devices within the same or different storage subsystems that are defined as a pair by a dual copy, PPRC, or XRC operation, and are not in suspended or pending state. The operation records the same data onto each volume.

DWDM. Dense Wavelength Division Multiplexor. A technique used to transmit several independent bit streams over a single fiber link.

E

extended remote copy (XRC). A hardware-based and software-based remote copy service option that provides an asynchronous volume copy across storage subsystems for disaster recovery device migration, and workload migration.

F

fixed utility volume. A simplex volume assigned by the storage administrator to a logical storage subsystem to serve as working storage for XRC functions on that storage subsystem.

FlashCopy. A point-in-time copy services function that can quickly copy data from a source location to a target location.

floating utility volume. Any volume of a pool of simplex volumes assigned by the storage administrator to a logical storage subsystem to serve as dynamic storage for XRC functions on that storage subsystem.

J

journal. A checkpoint data set that contains work to be done. For XRC, the work to be done consists of all changed records from the primary volumes. Changed records are collected and formed into a “consistency group”, and then the group of updates is applied to the secondary volumes.

K

km. kilometer.

L

Licensed Internal Code (LIC). Microcode that IBM does not sell as part of a machine, but licenses to the customer. LIC is implemented in a part of storage that is not addressable by user programs. Some IBM products use it to implement functions as an alternative to hard-wired circuitry.

link address. On an ESCON interface, the portion of a source or destination address in a frame that ESCON uses to route a frame through an ESCON director. ESCON associates the link address with a specific switch port that is on the ESCON director. Equivalently, it associates the link address with the channel subsystem or controller link-level functions that are attached to the switch port.

logical partition (LPAR). The ESA/390 term for a set of functions that create the programming environment that is defined by the ESA/390 architecture. ESA/390 architecture uses this term when more than one LPAR is established on a processor. An LPAR is conceptually similar to a virtual machine environment except that the LPAR is a function of the processor. Also, the LPAR does not depend on an operating system to create the virtual machine environment.

logical subsystem (LSS). The logical functions of a storage controller that allow one or more host I/O interfaces to access a set of devices. The controller aggregates the devices according to the addressing mechanisms of the associated I/O interfaces. One or more logical subsystems exist on a storage controller. In general, the controller associates a given set of devices with only one logical subsystem.

O

orphan data. Data that occurs between the last, safe backup for a recovery system and the time when the application system experiences a disaster. This data is lost when either the application system becomes available for use or when the recovery system is used in place of the application system.

P

peer-to-peer remote copy (PPRC). A hardware-based remote copy option that provides a synchronous volume copy across storage subsystems for disaster recovery, device migration, and workload migration.

pending. The initial state of a defined volume pair, before it becomes a duplex pair. During this state, the contents of the primary volume are copied to the secondary volume.

PPRC. See peer-to-peer remote copy.

PPRC dynamic address switching (P/DAS). A software function that provides the ability to dynamically redirect all application I/O from one PPRC volume to another PPRC volume.

primary device. One device of a dual copy or remote copy volume pair. All channel commands to the copy logical volume are directed to the primary device. The data on the primary device is duplicated on the secondary device. See also secondary device.

PTF. program temporary fix.

R

RACF. Resource Access Control Facility.

recovery system. A system that is used in place of a primary application system that is no longer available for use. Data from the application system must be available for use on the recovery system. This is usually accomplished through backup and recovery techniques, or through various DASD copying techniques, such as remote copy.

remote copy. A storage-based disaster recovery and workload migration function that can copy data in real time to a remote location. Two options of remote copy are available. See peer-to-peer remote copy and extended remote copy.

resynchronization. A track image copy from the primary volume to the secondary volume of only the tracks that have changed since the volume was last in duplex mode.

S

secondary device. One of the devices in a dual copy or remote copy logical volume pair that contains a duplicate of the data on the primary device. Unlike the primary device, the secondary device may accept only a limited subset of channel commands.

sidefile. A storage area used to maintain copies of tracks within a concurrent copy domain. A concurrent copy operation maintains a sidefile in storage control cache and another in processor storage.

simplex state. A volume is in the simplex state if it is not part of a dual copy or a remote copy volume pair. Ending a volume pair returns the two devices to the simplex state. In this case, there is no longer any capability for either automatic updates of the secondary device or for logging changes, as would be the case in a suspended state.

site table. Entity within GDPS that is created from information in the GEOPLEX DOMAINS. It contains a list of all of the systems in the GDPS environment.

suspended state. When only one of the devices in a dual copy or remote copy volume pair is being updated because of either a permanent error condition or an authorized user command. All writes to the remaining functional device are logged. This allows for automatic resynchronization of both volumes when the volume pair is reset to the active duplex state.

synchronization. An initial volume copy. This is a track image copy of each primary track on the volume to the secondary volume.

synchronous operation. A type of operation in which the remote copy PPRC function copies updates to the secondary volume of a PPRC pair at the same time that the primary volume is updated. Contrast with asynchronous operation.

system data mover. A system that interacts with storage controls that have attached XRC primary volumes. The system data mover copies updates made to the XRC primary volumes to a set of XRC-managed secondary volumes.

T

timeout. The time in seconds that the storage control remains in a "long busy" condition before physical sessions are ended.

U

utility volume. A volume that is available to be used by the extended remote copy function to perform data mover I/O for a primary site storage control's XRC-related data. A device that is used to gather information about the environment for configuration setup. It is also used to issue PPRC Freeze commands to the SSID-pair.

X

XRC. See *extended remote copy*.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this book.

IBM Redbooks publications

The following IBM Redbooks publications provide additional information about the topic in this book. Some publications referenced in this list might be available in softcopy only.

- ▶ *IBM System Storage Business Continuity: Part 1 Planning Guide*, SG24-6547
- ▶ *IBM DS8870 Copy Services for IBM z Systems*, SG24-6787
- ▶ *IBM z Systems Connectivity Handbook*, SG24-5444
- ▶ *IBM TotalStorage Enterprise Storage Server Implementing ESS Copy Services with IBM eServer zSeries*, SG24-5680
- ▶ *IBM Virtualization Engine TS7700 with R 2.0*, SG24-7975
- ▶ *Server Time Protocol Implementation Guide*, SG24-7281
- ▶ *Server Time Protocol Planning Guide*, SG24-7280

The following IBM Redpaper publications contain information about the DWDMs that are qualified for use with GDPS:

- ▶ *IBM System z Qualified WDM: Adva FSP 2000 at Release Level 6.2*, REDP-3903
- ▶ *IBM System z Qualified WDM: Nortel Optical Metro 5200 at Release Level 10.0*, REDP-3904
- ▶ *zSeries Qualified WDM Vendor: Cisco Systems*, REDP-3905
- ▶ *zSeries Qualified WDM Vendor: Lucent Technologies*, REDP-3906

You can search for, view, download or order these documents and other Redbooks publications, Redpaper publications, Web Docs, draft and additional materials, from here:

ibm.com/redbooks

Other publications

These publications are also relevant as further information sources:

- ▶ *Advanced Copy Services*, SC35-0428
- ▶ *DFSMS Extended Remote Copy Installation Planning Guide*, GC35-0481
- ▶ *DFSMS Extended Remote Copy Reference Information for Advanced Users*, GC35-0482
- ▶ *System-Managed CF Structure Duplexing Implementation Summary*, GM13-0540
- ▶ *System z Capacity on Demand User's Guide*, SC28-6846
- ▶ *Tivoli NetView for OS/390 Installation and Administration Guide*, SC31-8236
- ▶ *z/VM CP Planning and Administration*, SC24-6083

Online resources

These web pages are also relevant as further information sources:

- ▶ The following page, on the IBM ResourceLink website, contains a list of the qualified Dense Wavelength Division Multiplexing (DWDM) vendors:
<http://ibm.co/1Jia5AJ>
- ▶ Several related web pages provide more information about aspects of disaster recovery and business resilience:
- ▶ Speech by SEC Staff: Disaster Recovery and Business Continuity Planning:
<http://www.sec.gov/news/speech/spch050103mag.htm>
- ▶ Interagency Paper on Sound Practices to Strengthen the Resilience of the U.S. Financial System:
<http://www.sec.gov/news/studies/34-47638.htm>
- ▶ U.S.-EU & U.S.-Swiss Safe Harbor Frameworks, export.gov website:
<http://www.export.gov/safeharbor/>

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

A

- application availability
 - maximizing 14
 - role of HyperSwap 58, 109, 197, 280
- asynchronous PPRC, see Global Mirror
- automation
 - role in a disaster recovery solution 41

B

- Basel II 5
- Batch scripts 90, 221, 290

C

- Capacity Backup Upgrade 42
- Capacity Upgrade on Demand 43
- CF Hint support xxvii
- connectivity
 - what devices must be catered for 44
- connectivity requirements for GDPS/XRC 140
- continuous availability
 - role of HyperSwap 58, 108, 197, 280
- Controlling system 69, 118, 205
 - role in GDPS/GM 165
 - role in GDPS/PPRC 69, 118, 205
 - role in GDPS/XRC 140
- Coupled XRC 30
- Coupling Facility
 - connectivity requirements 45
 - considerations relating to distances 16
- cross-platform considerations 6

D

- data consistency
 - in XRC 28
- data consistency across multiple data managers 18
- database recovery
 - comparison to database restart 17
- database restart
 - comparison to database recovery 17
 - in GDPS/PPRC 55, 106, 192, 279
- dependent write logic
 - definition 18
 - in GDPS/PPRC 55, 105, 192, 278
- Disaster Recovery
 - SHARE tiers 3
- Disk Magic 24
- distance
 - considerations for duplexed CFs 17
- DWDMs 48

F

- FCP PPRC links 45

FlashCopy

- considerations for control unit capacity planning 40
- COPY mode 39
- description 38
- modes of operation 39
- NOCOPY mode 39
- role in a disaster recovery solution 38
- support for Open LUN devices xxvii
- target volume contents 39
- user-initiated 40
- using to create consistent DR tape backups 360

- Freeze policies 56, 106, 193

G

- GDPS
 - GDPS utility device requirements 164
 - support of CBU 42
- GDPS offerings
 - common components 10
- GDPS scripts 146, 176
 - benefits of 146, 177
- GDPS/GM
 - Controlling systems 165, 168
 - introduction 164
 - summary of functions 163
 - typical configuration 165
- GDPS/PPRC
 - alerting functions 80, 123, 211
 - Controlling system 72, 74, 121, 208, 210
 - introduction 54, 190, 276
 - managing the remote copy configuration 81, 124, 212
 - multi-site workload 72, 208
 - requirement for a Parallel Sysplex 72, 208
 - sample scripts 88
 - services component 98, 133, 228, 291
 - single-site workload 70, 206
 - Standard Actions 84, 215
 - support for Linux guests 299
 - Sysplex Resource Management 84, 215
 - typical configurations 68, 117, 204
- GDPS/PPRC HM
 - Controlling system 69, 118, 205
 - Controlling system requirements 68, 117, 205
 - description 104
 - HyperSwap connectivity considerations 120
 - summary of features 103
 - supported distances 74, 121, 210
- GDPS/XRC
 - configuration example 140
 - introduction 138
 - managing multiple production sites 140
 - Recovery Point Objective 138
 - Recovery Time Objective 138
 - relationship to primary systems 138

- support for Coupled XRC 140
- support for Multi-SDM 140
- support of CBU 138
- supported systems 139

Global Mirror

- connectivity requirements 34
- introduction 32
- Recovery Point Objective 33

H

Health Insurance Portability and Accountability Act 5

HyperSwap xxvii

- benefits of 58, 108, 197, 279
- types 59, 109, 197, 280

I

IT Resilience

- definition 2

IT Resilience solution

- characteristics of 7

M

Metro Mirror, see PPRC

Multiple XRC 30

O

Online resources 378

Open LUNs

- prerequisites 296

P

Parallel Access Volumes

- using with PPRC 24

Parallel Sysplex

- as a prerequisite for GDPS offerings 14
- multi-site considerations 16
- role in providing IT Resilience capability 14
- role in relation to GDPS/PPRC 14
- role in relation to GDPS/XRC 14

Planned Action

- sample 86, 218

Planned Actions 146–147, 178

planned outage 7

PPRC

- introduction 23
- maximum supported distance 23
- performance considerations 24

R

Recovery Point Objective

- definition 3
- for GDPS/PPRC 23, 57, 107
- for GDPS/XRC 27, 138
- for Global Mirror 33

Recovery Time Objective

- definition 2

- for GDPS/PPRC HM 104

- for GDPS/XRC 138

- role of automation 41

Redbooks Web site

- Contact us xv

S

scenarios

- CA and DR in two sites, metro distance 362

- CA in a single site 358

- DR in two site, metro distance 361

scripts 88, 219

Standard Actions

- description of 83, 214

sysplex considerations for GDPS/XRC 140

System-Managed Coupling Facility Structure Duplexing 17

U

Unplanned Action 146, 149

User-initiated FlashCopy 40

X

XRC

- data consistency 28

- data flow 29

- distance limitation 138

- factors affecting amount of data loss 27

- hardware prerequisites 31

- Recovery Point Objective 27

- support for systems that do not timestamp I/Os 28, 139

- supported distances 28

- supported platforms 28

Z

z/OS Global Mirror, see XRC

zero data loss 23

- GDPS/PPRC options 56, 106

- remote copy options 27

Redbooks

IBM GDPS Family: An Introduction to Concepts and Capabilities

SG24-6374-12

ISBN 0738441880



(0.5" spine)

0.475" <-> 0.873"

250 <-> 459 pages



SG24-6374-12

ISBN 0738442534

Printed in U.S.A.

Get connected

