GDPS Family - An Introduction to Concepts and Capabilities

Overview of business resilience requirements

Technologies that support IT resilience

Introduction to the GDPS family of offerings

Frank Kyne
Noshir Dhondy
Mike Hrencecin

Redbooks

ibm.com/redbooks

International Technical Support Organization

GDPS Family - An Introduction to Concepts and Capabilities

March 2011

SG24-6374-06

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

Seventh Edition (March 2011)

This edition applies to Version 3, Release 8, Modification 0 of the GDPS family of offerings.

Contents

Notices

This information was developed for products and services offered in the U.S.A.

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 give you any license to these patents. You can send license inquiries, in writing, to:
IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: 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 states 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 Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

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

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.

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 the names and addresses used by an actual business enterprise 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.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at http://www.ibm.com/legal/copytrade.shtml

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

AIX®
CICS®
DB2®
Distributed Relational Database Architecture™
DRDA®
DS8000®
Enterprise Storage Server®
ESCON®
eServer™
FICON®
FlashCopy®
GDPS®
Geographically Dispersed Parallel Sysplex™
HACMP™
HyperSwap®
IBM®
IMS™
MVS™
NetView®
Parallel Sysplex®
RACF®
Redbooks®
Redpaper™
Redbooks (logo) ®
Resource Link™
Sysplex Timer®
System i®
System p®
System Storage®
System x®
System z10®
System z9®
System z®
TDMF®
Tivoli®
TotalStorage®
VTAM®
WebSphere®
z/OS®
z/VM®
z/VSE™
z10™
z9®
zSeries®

The following terms are trademarks of other companies:

InfiniBand, and the InfiniBand design marks are trademarks and/or service marks of the InfiniBand Trade Association.

Disk Magic, and the IntelliMagic logo are trademarks of IntelliMagic BV in the United States, other countries, or both.

Novell, SUSE, the Novell logo, and the N logo are registered trademarks of Novell, Inc. in the United States and other countries.

mySAP, SAP, and SAP logos are trademarks or registered trademarks of SAP AG in Germany and in several other countries.

VMware, the VMware "boxes" logo and design are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions.

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.

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

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

Preface

This IBM® Redbooks® publication presents an overview of the GDPS® family of offerings and the role they play in delivering a business IT resilience solution.

This book begins with a discussion of general concepts of business IT resilience and disaster recovery along with some 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 achieve your defined goals for disaster recover and high availability. Also covered are the features that simplify and enhance data replication activities, the prerequisites for implementing each offering, and some hints for planning for the future as well as immediate business requirements. Tables provide an easy-to-use summary and comparison of the offerings, and the additional planning and implementation services available from IBM are explained.

Finally, a number of practical customer 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, Systems Programmers, and Disaster Recovery Planners. The subsequent chapters provide more technical details about the GDPS offerings and each can be read in isolation for those that are interested. Because of this, if you do read all the chapters, you may note that some information is repeated.

The team who wrote this book

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

Frank Kyne is an Executive IT Specialist at the International Technical Support Organization, Poughkeepsie Center. He has been an author of a number of Parallel Sysplex® Redbooks documents. Before joining the ITSO eleven years ago, Frank worked in IBM Services in Ireland as an MVS™ Systems Programmer.

Noshir Dhondy is a Senior Engineer in the IBM Systems and Technology Group in Poughkeepsie, with over 40 years of experience at IBM. He provided consultation and product planning input to the team that developed Server Time Protocol. He has been involved with various aspects of hardware design and was the lead designer of the 9037 Sysplex Timer®. Noshir is a member of the Parallel Sysplex and GDPS Product Development Teams, and has been providing GDPS technical marketing support since 1999.

Mike Hrencecin is a Development Manager in the IBM Systems and Technology Group (STG). He has nearly 20 years of IT industry experience taking a broad set of development, client consulting, and management roles across IBM hardware, software, and solutions teams. He began work with GDPS in 2003 as an STG product advocate for US clients and continues to stay involved with GDPS, mainframe, and storage technologies as a post-sales client support manager for STG enterprise hardware products.

The authors of the previous editions of this book were:

- David Raften
- Mark Ratte
- Gene Sale

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

George Kozakos
IBM Australia

Thomas Bueche
IBM Germany

Nicholas Clayton
IBM UK

Stephen Anania
Charlie Burger
Alan McClure
David Petersen
Judy Ruby-Brown
Sim Schindel
John Sing
IBM 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 in length, 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 at:

ibm.com/redbooks/residencies.html

Comments welcome

Your comments are important to us!

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

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

  ibm.com/redbooks

- Send your comments in an Internet note to:

  redbook@us.ibm.com

- Mail your comments to:

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 may also include minor corrections and editorial changes that are not identified.

Summary of changes for SG24-6374-06 for GDPS Family - An Introduction to Concepts and Capabilitiesas created or updated on March 30, 2011.

Editor's note: With this, the sixth edition of *GDPS Family - An Introduction to Concepts and Capabilities*, we have departed from our usual practice of including revision bars in the text. This step was taken because after five previous editions the revision bars have become so numerous as to be potentially confusing or misleading. Regular readers of this volume and other titles in the GDPS library published by ITSO are accustomed to seeing the bars to focus their reading on changes from one edition to the next. With that in mind, we have maintained a copy of this document with change bars intact, and can provide the file to any interested reader who relies on revision bars and is having trouble finding information in the current book. To request the alternate version contact:

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

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 System z®.
- Added references to Microsoft® Windows® clusters (particularly in "Integrated configuration of GDPS/GM and VCS clusters" on page 159) 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. Additionally, the overview of the GDPS family of offerings was updated to help this chapter act as a standalone high-level overview.
- Chapter 2, "Infrastructure planning for availability and GDPS" on page 11 has been modified, moving some of the more technical details to subsequent chapters, while 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" on page 44 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 79.

- References to *Metro Mirror* were replaced by *PPRC* when discussing IBM’s 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® (ESS) 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.
- RCMF offerings have been stabilized and details moved to new appendices.
- Removed Peer-to-Peer tape from “DR in two data centers, global distance” on page 197 as the configuration of this legacy hardware (with immediate mode) would not be 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 remote copy technologies for CA and DR” on page 30 has been added to Chapter 2, “Infrastructure planning for availability and GDPS” on page 11.
- A new section “Improved controlling system availability - enhanced timer support” on page 54 has been added to Chapter 3, “GDPS/PPRC” on page 43, and to Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 79.
- A new section “GDPS/PPRC in a three-site configuration” has been added to Chapter 3, “GDPS/PPRC” on page 43 and to Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 79.
- A new section “GDPS/PPRC management of distributed systems and data” has been added to Chapter 3, “GDPS/PPRC” on page 43.
- A new section “GDPS/PPRC monitoring and alerting” has been added to Chapter 3, “GDPS/PPRC” on page 43, to Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 79, to Chapter 5, “GDPS/XRC” on page 107, and to Chapter 6, “GDPS/Global Mirror” on page 123.
- A new section “Other GDPS-related facilities” 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 GDPS/PPRC reduced impact initial copy and resynchronization and GDPS/PPRC Query Services.
- A new section, “Support for two GDPS Controlling systems” on page 151 has been added to Chapter 7, “GDPS extensions for heterogeneous systems and data” on page 145.
- A new section, “Disk and LSS sharing” on page 152has been added to Chapter 7, “GDPS extensions for heterogeneous systems and data” on page 145.
- A new section, “Integrated configuration of GDPS/GM and VCS clusters” on page 159 has been added to Chapter 7, “GDPS extensions for heterogeneous systems and data” on page 145.
- A new section, “DCM support for SA AppMan” on page 164 has been added to Chapter 7, “GDPS extensions for heterogeneous systems and data” on page 145.

Changed information

- 2.1.2, “Multi-site sysplex considerations” on page 12 was updated to change the maximum fiber distance from 100 km to 200 km (with RPQ).
- 2.4.3, “Global Mirror” on page 28 was rewritten.
- 2.8.1, “Capacity Backup Upgrade” on page 35 and 2.8.2, “On/Off Capacity on Demand” on page 35 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 36 to reflect the recently available Parallel Sysplex InfiniBand technology for coupling and STP and HMC connectivity requirements for STP.
- 2.9.8, “Connectivity options” on page 39 was updated.
- 8.3, “GDPS Metro/Global Mirror solution” on page 179 and 8.4, “GDPS Metro z/OS Global Mirror solution” on page 183 have been updated to include the Incremental Resynchronization function.
- Chapter 3, “GDPS/PPRC” on page 43 and Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 79 were restructured to introduce concepts in 3.1.1, “Protecting data integrity and data availability” on page 44 prior to the discussion of configurations in 3.2, “GDPS/PPRC configurations” on page 53 and 4.2, “GDPS/PPRC HM configurations” on page 89.
- “HyperSwap policy options” on page 48 was added.
- Tables provided in Chapter 3, “GDPS/PPRC” on page 43, Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 79, Chapter 5, “GDPS/XRC” on page 107, and Chapter 6, “GDPS/Global Mirror” on page 123 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 17, was added to explain the business impact of using Synchronous and Asynchronous remote copy technologies.
- A new chapter, Chapter 8, “Combining Local/Metro continuous availability with out-of-region disaster recovery” on page 173, discusses the GDPS/MGM and GDPS/MzGM offerings.
- Multiplatform for System z now supports native zLinux LPARs.
- GDPS/PPRC and GDPS HyperSwap® Manager have been enhanced to provide coexistence support for HyperSwap and TDMF®.
- IMS™ XRF coexistence support added to GDPS/PPRC and GDPS HyperSwap Manager.
- GDPS/PPRC provides for transparent exploitation 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 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 stagging data sets added in z/OS 1.7.

Changed information

- The GDPS/PPRC BRS configuration has moved to Chapter 3, “GDPS/PPRC” on page 43.
- 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

This revision reflects the addition, deletion, or modification of new and changed information described below.

New information

- Information about the GDPS/GM offering has been added

1

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 (which we refer to as IT resilience for brevity) from a technical perspective. This chapter includes a general discussion, not only specific to mainframe platforms although the topics are covered from an enterprise systems and mainframe perspective. Finally, we introduce the members of the Geographically Dispersed Parallel Sysplex™ (GDPS) family of offerings and provide a brief description of the aspects of an IT resilience solution 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 some industries now take the decision about whether to have an IT resilience capability out of your hands.

This document 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 document

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 due to a steady increase in the complexity of applications, the increasingly advanced capabilities of available technology, competitive business environments, and goverment regulations.

In the following chapter we briefly describe the available technologies typically leveraged in a GDPS solution to achieve IT resilience goals. Additionally, in order to understand the positioning and capabilities of the various offerings (which encompass hardware, software, and services), it is good 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 43, we describe the capabilities and prerequisites of each offering in the GDPS family of offerings. Because each offering addresses fundamently 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 a variety of hardware and software platforms. After covering the GDPS family of offerings, Chapter 7, “GDPS extensions for heterogeneous systems and data” on page 145 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 some examples of how the various GDPS offerings can satisfy your requirements for IT resilience and disaster recovery.

As a side note, the descriptions of the legacy Remote Copy Management Facility (RCMF) offerings are now included in Appendixes because these offerings are generally replaced by more full featured GDPS offering peers.

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*, which concentrates solely on recovering from an *unplanned* event.

When investigating IT resilience options, two things that must be at the forefront of your thinking are:

1. **Recovery Time Objective (RTO)**. This is 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 you had a significant service outage. How long was that outage, and how much pain did your company suffer as a result? This will help you get a feel for whether your RTO should be measured in days, hours, or minutes.

2. **Recovery Point Objective (RPO)**. This is how much data your company is willing to recreate *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 once daily full volume tape dumps, your RPO is 24 to 48 hours depending on when the tapes are taken off site. 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 mirroring instead of relying on these tapes alone.

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

1.3.1 Disaster Recovery

As mentioned earlier, 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. Some options rely on off-site or even outsourced locations that are contracted to provide data protection or even servers in the event of 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 right answer for which approach is better for every business, but 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 is not comprehensive of all possible DR offerings and approaches, 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	N/A - all data lost	N/A
Tape vaulting	Measured in days since last stored backup	Days
Electronic vaulting	Hours	Hours (hot remote location) to days
Active replication to remote site (w/o 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)	1 or more hours (dependent on automation)

Generally some form of real-time software or hardware replication will be required to achieve an RPO of minutes or less, but the only technologies that can provide an RPO of zero are synchronous replication technologies (described more in 2.3, "Synchronous versus asynchronous data transfer” on page 17) coupled with automation to ensure 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 as well as 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 once very 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 neither practiced nor accurate. This is where automation comes in to mitigate the point of failure introduced by the human element and to ensure you actually meet the RTO required of the business.

Finally, you may decide that one DR option is not appropriate for all aspects of the business. Some applications may tolerate a much greater loss of data and may not have an RPO as low as others. At the same time, some applications may not require recovery within hours whereas others most certainly do. While there is obvious flexibility in choosing different DR solutions for each application, the approach supported by GDPS is to provide a single optimized solution for the enterprise. This generally leads to a simpler solution and, since less infrastructure and software may need to be duplicated, often a more cost-effective solution too.

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 document, we concentrate on two aspects of IT resilience: disaster recovery, as discussed previously, and *Continuous Availability (CA)*, which encompasses not only recovering from disasters, but 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 may be sufficient. Others may need to go a step further and provide near-continuous application availability.

The market drivers behind the need for IT resilience are as follows:

- High and constantly increasing customer and market requirements for continuous availability of IT processes
- Financial loss, due to 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 impact a company like having an IT outage – even for a matter of minutes – and then finding the incident splashed across the newspapers and the evening news. Today, your customers, employees, and suppliers expect to be able to do business with you around the clock, and from all corners of the globe.

To help keep business operations running 24x365, 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:

- Key business applications and data are protected and available
- Should a disaster occur, business operations continue with a minimal impact

Regulations

In some countries, government regulations lay down specific rules for how an organization must handle its 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 health care 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 U.S. 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. While a set of recommendations published in Singapore (like the S 540-2008 Standard on Business Continuity Management)[2] may only be directly addressing businesses in a relatively small city state, it is common for companies to do business in many countries around the world, where these may be requirements for ongoing business operations of any kind.

Business requirements

It is important to understand 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 would suffer as a result of a outage.

A solution must be identified that balances the costs of the solution with the financial impact of an outage. A number of 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 needs to calculate the impact in your specific case. If you have not already conducted such an exercise, you may 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 would be far smaller 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.

1.3.3 Other considerations

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

[1] http://www.sec.gov/news/studies/34-47638.htm

[2] http://www.ss540.org/

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 not only the mainframe data, but also perhaps data accessed by multiple flavors 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. Should a disaster occur, your processes must be designed to recover this data in a consistent manner. Just as you would not consider recovering half an application's DB2® data to 8:00 AM and the other half to 5:00 PM, the data touched by these distributed applications must be managed to ensure that *all* 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 comprise only a small part of the equipment in that room. How confident are you that all those other platforms could be recovered? And if they can be recovered, will it be to the same point in time as your mainframe systems?

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 the end users can access them. Therefore, part of your IT resilience solution must include not only addressing the non-mainframe parts of your infrastructure, but also ensuring that recovery is integrated with the mainframe plan.

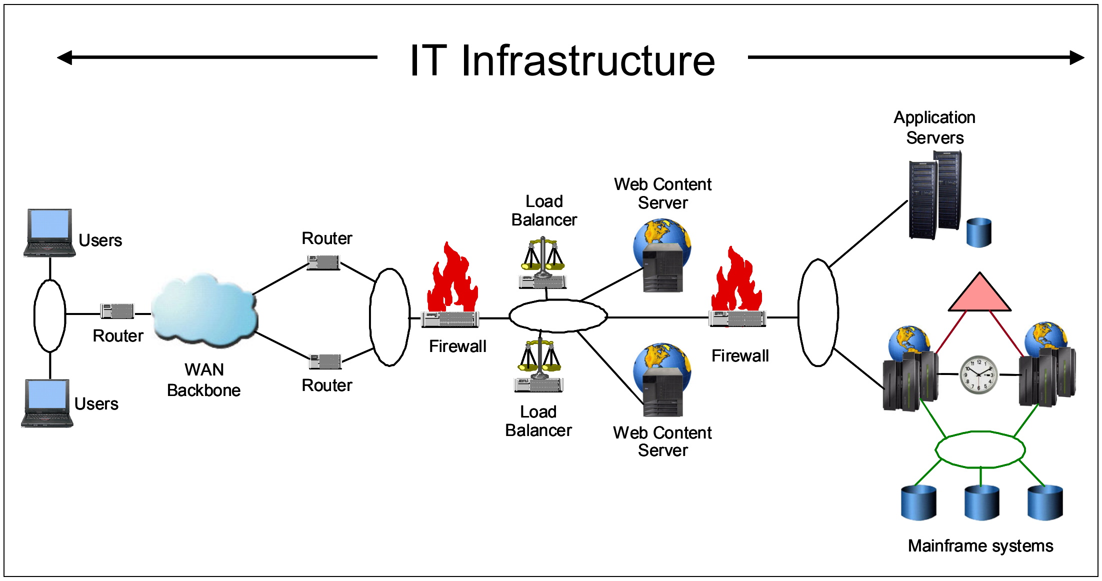


Figure 1-1 Typical IT infrastructure

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 could 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 may take minutes or even hours.

The majority of outages are indeed planned, and even among unplanned outages, the majority 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. Some of these outages may be quickly recovered from, but others may 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 demonstrated, IT resilience encompasses a lot 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, you should bear the following things in mind:

- **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 remote copy environment.** Does the solution provide an easy-to-use interface for monitoring and managing the remote copy environment? Will it automatically react to connectivity or other failures in the remote copy configuration?
- **Support for data consistency**. Does the solution provide data consistency across all remote copied volumes and disk subsystems? Does it provide support for protecting the consistency of the secondary volumes if it is necessary to resynchronize the primary and secondary volumes?
- **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 / re-IPL) 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 case of 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, in the event of 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 via 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 exploit 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 actually 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 leverages server and storage hardware as well as automation and clustering software technologies, many of which are described in more detail in Chapter 2, "Infrastructure planning for availability and GDPS" on page 11. 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 the solution meets and fulfils your business objectives.

The following list provides a short description of each offering with a view of what IT resiliency objectives it is intended to address. Additional details are included in separate chapters later in this book:

- GDPS/PPRC
  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
  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 systems automation capability that makes GDPS/PPRC a more complete DR solution.
- GDPS/XRC
  DR solution across two sites separated by virtually unlimited distance between sites. The solution is based on the IBM XRC asynchronous disk mirroring technology (also branded by IBM as z/OS Global Mirror).
- GDPS/Global Mirror
  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 subsystems based asynchronous form of remote copy.

- GDPS Metro/Global Mirror
  A three-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 cascading mirroring technology that combines PPRC and Global Mirror.
- GDPS Metro/z/OS Global Mirror
  A three-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).

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

- GDPS automation code: The code has been developed and enhanced over a number of years to exploit new hardware and software capabilities, to reflect best practices based on the IBM experience with GDPS customers since its inception in 1998, and to address the constantly changing requirements of our customers.
- Capabilities to exploit underlying hardware and software capabilities: IBM software and hardware products have support to surface problems that may affect the availability of those components, and to facilitate repair actions.
- Services: There is perhaps only one thing in common across all the GDPS implementations: every one has some 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: as an example, more function-rich offerings like 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. Each chapter can be read in isolation; that is, it is not necessary to read all chapters if you are only interested in a specific offering. If you do read all the chapters, you may note that some information is repeated in each chapter.

1.6 A vendor neutral solution

The GDPS automation code relies on the runtime capabilities of Tivoli® NetView® and Tivoli System Automation (SA). While these products provide tremendous first level automation capabilities in and of themselves, there are alternative solutions you may already have from other vendors. GDPS continues to deliver features and functions that take advantage of properties unique to the Tivoli products (like support for alert management through Tivoli IOM), but Tivoli NetView and Tivoli SA also work very well alongside other first level automation solutions. In other words, while there are indeed advantages to a comprehensive solution from IBM, you do not have to replace your current automation investments before moving forward with a GDPS solution.

Additionally, each of the GDPS solutions relies on IBM-developed disk replication technologies: PPRC for GDPS/PPRC, XRC for GDPS/XRC, and Global Mirror for GDPS/GM. These architectures are, of course, 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 (ESS) 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, XRC, GM, and FlashCopy, too) have also been licensed by many major enterprise storage vendors. This allows customers 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. Indeed, while most GDPS installations do rely on IBM storage products, there are several production installations of GDPS around the world that rely on non-IBM storage vendor products.

Finally, IBM has a GDPS Qualification Program[3] 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 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 the test results.

Recognize that this qualification program does not imply that IBM provides defect or troubleshooting support for a qualified vendor's products. This 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.

You will want to 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.

1.7 Summary

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

In the next chapter we introduce key infrastructure technologies related to IT resilience focused on the mainframe platform. After that, we describe how the various GDPS offerings exploit those technologies. And finally, we position the various GDPS offerings against typical business scenarios and requirements.

It is our intent to update this document as new GDPS capabilities are delivered.

[3] http://www.ibm.com/systems/z/gdps/qualification.html

2

Infrastructure planning for availability and GDPS

In this chapter, we discuss a number of technologies that are available to help you achieve your goals related to IT resilience, recovery time, and recovery point objectives. To understand how the GDPS offerings described in this document 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 we 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 System z[1] 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 balancing not only provides higher levels of availability now, it also provides a foundation to achieve greater resiliency should you implement GDPS too.

In the following sections we briefly discuss Parallel Sysplex, the benefits you can derive by exploiting the technology, and the points to consider if you should decide to implement GDPS/PPRC. 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.

2.1.1 Maximizing application availability

There is only one way to protect applications from the loss of a single component (such as a CICS® region or a z/OS system) — that is to run multiple, failure-isolated copies. This infers an ability to share data at the record level, with integrity, and to *dynamically* balance 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 at the same time 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.

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 exploiting Parallel Sysplex is available in other IBM documentation and it will not be covered in this document.

2.1.2 Multi-site sysplex considerations

The considerations for a multi-site 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 multi-site workload configurations. Configurations where the production systems run together in one site or the other (but not split between 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

[1] In this document we use the term System z to refer generically to the IBM System z and zSeries® ranges of processors. If something only applies to System z or zSeries processors, we point that out at the time.

at one site, but where partially active systems or systems only enabled for queries are running at the other site, are still multi-site workloads.

There are several phrases often used to describe variations of multi-site workload. Short definitions are included here for some of the more commonly implemented variations.

Active/Active This refers to a multi-site workload configuration where z/OS systems are actively running 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 balancing 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 once contending locks and communications timeouts clear. When combined with HyperSwap, an Active-Active configuration has the potential to provide near-continuos availability for applications even in the case of a site outage.

Active/Warm This refers to a multi-site workload configuration that is very similar to the traditional GDPS 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 merely IPLed without subsystems or other resources active. This configuration is intended to save IPL time when moving workload between sites. This can be most effective for supporting planned movement of workload because in many unplanned scenarios the “warm” systems might not survive.

Active/Query This refers to a multi-site workload configuration that is very close to a traditional GDPS 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 due to serialization, thereby protecting shared resources when delays due to 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 may only provide value for planned scenarios because in many unplanned scenarios the “query” or “hot standby” subsystems might not survive.

There are potentially many more configuration variations that you can come up with, but from a Parallel Sysplex and GDPS perspective, all of these fall into either the single-site or the multi-site workload category.

To simplify this discussion, only the single-site (Active/Standby) and fully multi-site (Active/Active) configurations are discussed further in this book.

Active/standby or active/active configuration

Traditionally, Parallel Sysplexes have been 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. Also, even if the systems sharing the data are spread across more than one site, all the primary disk subsystems would normally be contained in the same site, so a failure affecting the primary disks would impact the systems in both sites. As a result, an active/active workload configuration does not, in itself, provide significantly greater availability than an active/standby workload configuration.

More specifically, be careful when applying a multi-site workload configuration when the underlying Parallel Sysplex cannot be configured to spread the important Coupling Facility structures across the sites while still achieving the required performance. As discussed later in this chapter and illustrated in Table 2-1 on page 37, the Coupling Link technology can support links upwards of 100 km with qualified DWDM. However, this does not mean that your workload will tolerate even 1 km of distance. Individual coupling operations will be delayed by tens of microseconds per kilometer and while this time can be calculated, there is no good way to validate the queuing effects created by workload intensity and the amount of sharing that is unique to each environment. In other words, you will need to actually run your workload with connections at distance to evaluate the tolerance and impacts of distance.

Finally, the benefits of Active/Active (as well as the other variations on multi-site workload configurations mentioned here) come with additional complexity. This must be taken into account when weighing the benefits of these multi-site workload configurations.

CF Structure Duplexing

There are two mechanisms for duplexing CF structures. The first of these, User-Managed Structure Duplexing is only supported for use 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 recommended that the GBP structures used by a production DB2 data sharing group should always be duplexed.

The other type of structure duplexing, System-Managed Coupling Facility Structure Duplexing (hereafter referred to as SM Duplexing), provides a general purpose, hardware-assisted, easy-to-exploit 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 may seem that you are guaranteed to be able to use the duplexed instance of your structures in the event you have to recover and restart your workload with the frozen secondary copy of your disks. However, this is not always the case! There could be rolling disaster scenarios where prior to, following, or during the freeze event, there is some sort of 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, if there is a surviving, accessible instance of application-related 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.

For more information, refer to the white paper entitled *System-Managed CF Structure Duplexing*, GM13-0103, available on the Internet at:

http://www.ibm.com/servers/eserver/zseries/library/techpapers/gm130103.html

2.2 Data consistency

The ability to do a database restart, rather than a database recovery, is essential in order to meet the Recovery Time Objective (RTO) of many businesses—typically less than an hour. Database restart allows starting a database application (as you would do following a database manager, 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 involves restoring the last set of image copies and applying log changes to bring the database up to the point of failure.

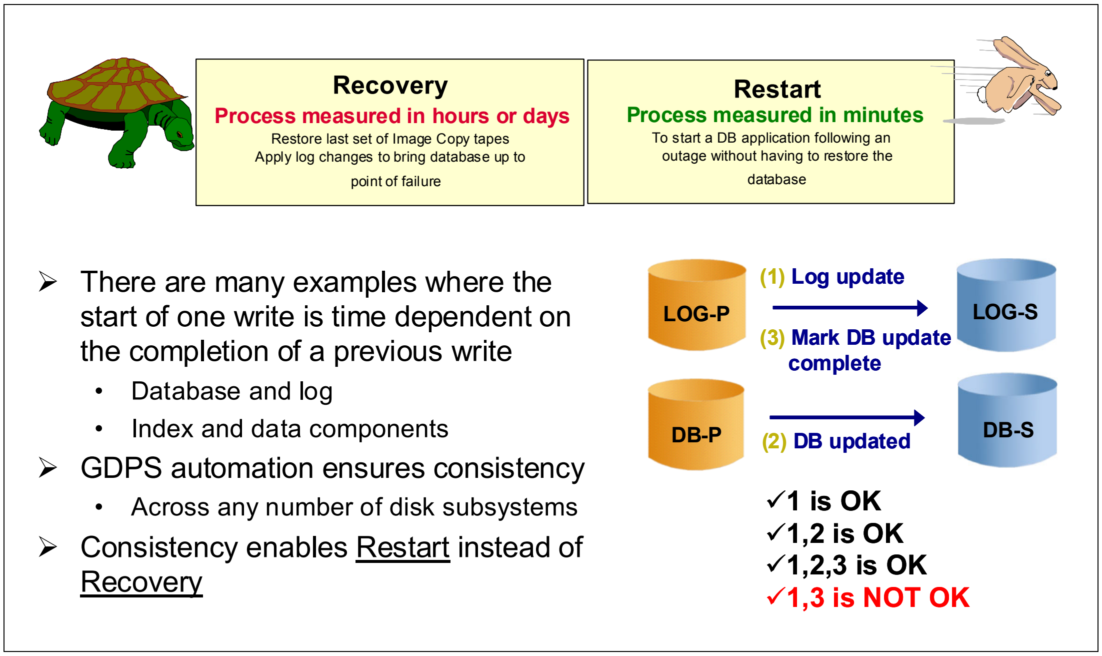
But there is more to consider than simply the data for one data manager. What if you have an application that updates data in IMS, DB2, and VSAM? If you need to do a recover for these, will your recovery tools not only allow you to recover them to the same point in time, but also to the level of granularity 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 primary and secondary volumes spread across any number of storage (both tape and disk) subsystems is essential to provide not only data integrity, but also the ability to do a normal database restart in the event of a disaster.

2.2.1 Dependent write logic

Database applications commonly ensure the consistency of their data by using *dependent write logic* irrespective of whether remote data mirroring is being used. Dependent write logic states that if `I/O B` must logically follow `I/O A`, then `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 itself, such as databases, catalog/VTOC, and VSAM file updates, where dependent writes are issued.

As an example, in Figure 2-1 on page 16, 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 makes a database update it will (1) write an entry to the log indicating the update intent, (2) make the database update, and then (3) write another entry to the log indicating the update was made. If you are going to remote copy these volumes, it is vital that *all* the updates are mirrored to the secondary disks.



Recovery

Process measured in hours or days

Restore last set of Image Copy tapes
Apply log changes to bring database up to
point of failure

Restart

Process measured in minutes

To start a DB application following an
outage without having to restore the
database

- There are many examples where the start of one write is time dependent on the completion of a previous write
  - Database and log
  - Index and data components
- GDPS automation ensures consistency
  - Across any number of disk subsystems
- Consistency enables Restart instead of Recovery

✓1 is OK
✓1,2 is OK
✓1,2,3 is OK
✓1,3 is NOT OK

Figure 2-1 The need for data consistency

It is unlikely that all the components in a data center fail at the same instant even in the rare case of a full data center outage. The networks may 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 such that cross-volume and subsystem data consistency is preserved during intermittent and staged failures that may 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 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.[2]

We discuss data consistency for synchronous remote copy in more detail in “PPRC data consistency” on page 21.

[2] 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.

For the two IBM asynchronous remote copy offerings, the consistency of the volumes in the recovery site is guaranteed because of the way these offerings work. This is discussed further in 2.4.3, “Global Mirror” on page 28 and “XRC data consistency” on page 25.

2.3 Synchronous versus asynchronous data transfer

Synchronous 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.

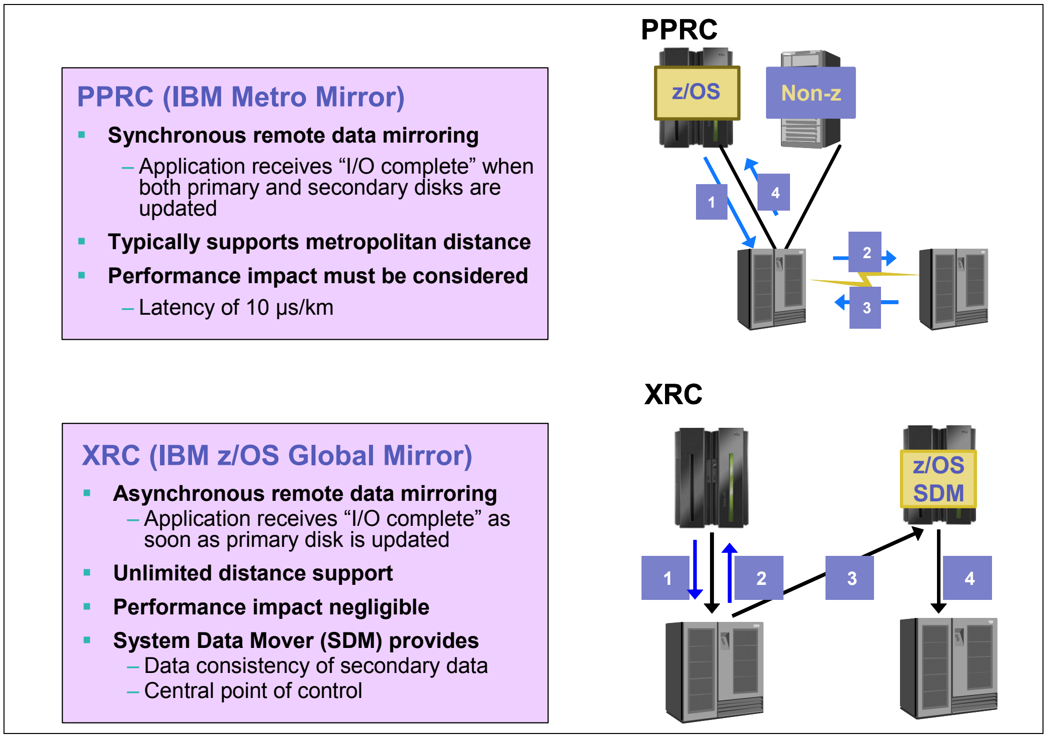
Terminology note: In this book we will continue to use the term Peer-to-Peer Remote Copy (PPRC) when referring to the synchronous disk replication architecture. The rebranded name of the IBM implementation of this architecture is IBM Metro Mirror, which will be used when specifically referring to the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

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

When using synchronous data transfer, illustrated in Figure 2-2 on page 18 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 acknowledgement that the write is complete (4) is sent to the application. Because the application must wait until it receives the acknowledgement 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 due to signal latency[3].

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

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



PPRC (IBM Metro Mirror)

- **Synchronous remote data mirroring**
  - Application receives “I/O complete” when both primary and secondary disks are updated
- **Typically supports metropolitan distance**
- **Performance impact must be considered**
  - Latency of 10 µs/km

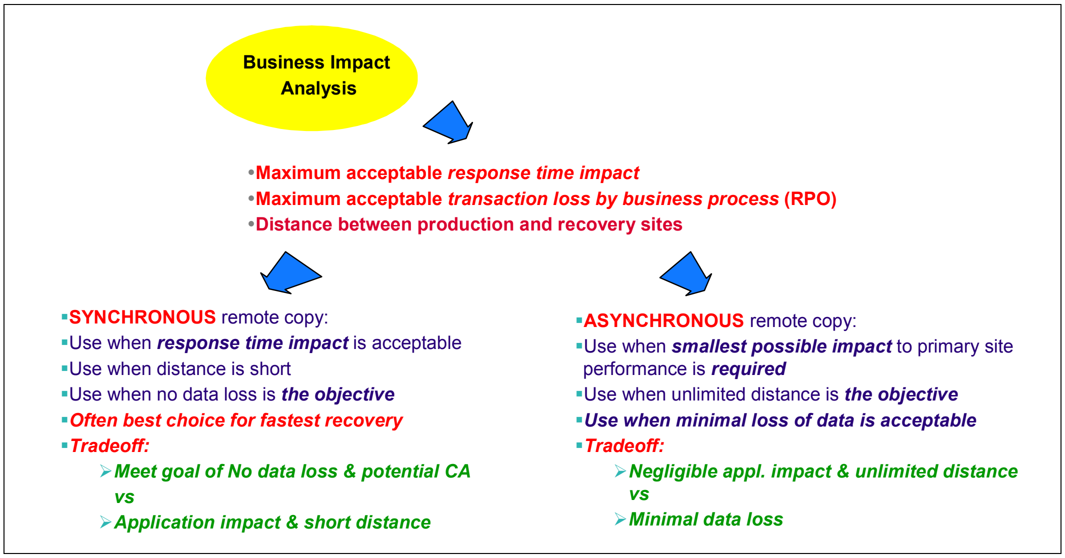
XRC (IBM z/OS Global Mirror)

- **Asynchronous remote data mirroring**
  - Application receives “I/O complete” as soon as primary disk is updated
- **Unlimited distance support**
- **Performance impact negligible**
- **System Data Mover (SDM) provides**
  - Data consistency of secondary data
  - Central point of control

Figure 2-2 Synchronous (PPRC) versus asynchronous (XRC)

With asynchronous replication, illustrated in Figure 2-2 by XRC (IBM z/OS Global Mirror), the application writes to the primary disk subsystem (1) and receives an acknowledgement 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 (3) (4). 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.

When selecting a remote copy solution, a business impact analysis should be performed to determine which solution will meet the businesses requirements while ensuring your service delivery objectives continue to be met (Figure 2-3 on page 19). 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 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 applications ability to tolerate the increased response times must factored into the decision process.



Business Impact Analysis

- **Maximum acceptable *response time impact***
- **Maximum acceptable *transaction loss by business process* (RPO)**
- **Distance between production and recovery sites**

- **SYNCHRONOUS** remote copy:
- Use when ***response time impact*** is acceptable
- Use when distance is short
- Use when no data loss is ***the objective***
- ***Often best choice for fastest recovery***
- ***Tradeoff:***
  - ***Meet goal of No data loss & potential CA***
    vs
  - ***Application impact & short distance***

- **ASYNCHRONOUS** remote copy:
- Use when ***smallest possible impact*** to primary site performance is ***required***
- Use when unlimited distance is ***the objective***
- ***Use when minimal loss of data is acceptable***
- ***Tradeoff:***
  - ***Negligible appl. impact & unlimited distance***
    vs
  - ***Minimal data loss***

Figure 2-3 Business impact analysis

Many enterprises have requirements, both business and regulatory requirements, to provide near continuous data availability, without loss of transactional data, while protecting critical business data in the event of a wide-scale disruption. This can be achieved by implementing three-copy, sometimes referred to as three-site, mirroring solutions that exploit both synchronous and asynchronous replication technologies. Synchronous solutions are used to protect against the day-to-day disruptions with no loss of transactional data. While 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 additional information about three-copy replication solutions, refer to Chapter 8, “Combining Local/Metro continuous availability with out-of-region disaster recovery” on page 173.

2.4 Remote copy technologies

There are two primary ways to make your data available following a disaster:

- Use some form of tape-based backup
- Remote copy the data to a recovery site

For companies with an RTO of a small number of hours, 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 document you already have, or are planning to implement, some form of remote copy.

Remotely copying your data not only eliminates the time that would be required to restore the volumes, it also 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 carried out between the disk subsystems.
- XRC: The task of retrieving the updates from the primary disk subsystem and applying those changes to the secondary volumes is carried out by a z/OS component called the System Data Mover (SDM).
- IBM Global Mirror: This offering mirrors the data asynchronously; however, unlike XRC, all interactions are carried out 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 any of the remote copy technologies described in the following sections, refer to *IBM System Storage DS8000: Copy Services with IBM System z,* SG24-6787.

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 even greater distances after a 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 36 for more details.
>
> More significantly, you must 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 to support the workloads to 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

If you have a Recovery Point Objective of zero (0), meaning zero data loss, PPRC is the only IBM remote copy option.

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 zero data loss is to immediately stop all update activity to the primary disks should the remote copy relationship cease to exist (if you lose connectivity between the primary and secondary devices, for example).

So, 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 could be stopped for a reason other than a real disaster—for example, if the failure were to be 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, should reduce the likelihood of such events to an acceptably low level.

Supported platforms

PPRC remote copying is supported for any IBM or non-IBM disk subsystem that supports the PPRC architecture, specifically `CGROUP FREEZE/RUN` support. PPRC can mirror both Fixed Block Architecture (FBA) disks typically used by non System z platforms and CKD disks traditional used by mainframe operating systems such as z/OS, z/VM®, and z/VSE™. Not all operating systems necessarily support an interface to control the remote copy function; however, the PPRC function can be controlled from a connected z/OS system as long as there are sufficient CKD formatted volumes defined in the storage subsystem (as described more fully in 3.1.3, “Protecting distributed (FBA) data” on page 51 for GDPS/PPRC and 4.1.2, “Protecting distributed (FBA) data” on page 87 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, which can help with migrations and technology upgrades.

Distance

The maximum distance supported for IBM Metro Mirror is 300 km (without an RPQ). Note that typical GDPS/PPRC and GDPS/HyperSwap manager configurations are limited to distances less than this due to Coupling Link or timer configurations. See 2.9.3, “Coupling links” on page 37 for more details about the supported distances for these Parallel Sysplex connections. Additionally, you will need to contact other storage vendors (if required) to understand the maximum distances supported by their PPRC compatible mirroring implementations.

Performance

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 may 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 utilization also will increase. Depending on the type and number of PPRC links you configured, additional links and the use of Parallel Access Volumes (PAVs) may 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.

Connectivity

Connectivity between the primary and secondary disk subsystems can be provided by direct connections between the primary and secondary disk subsystems, by FICON® switches, by DWDMs, and by channel extenders. The type of inter-site connection (dark fiber or telecommunications link) available determines the type of connectivity you use: telecommunication links can be used by channel extenders, the other types of connectivity require dark fiber.

You can find information about connectivity options and considerations for System z in the latest 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 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 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. At this point, the application program can continue its processing, and move on to any dependent writes that may have been waiting for this one to complete.

However, PPRC on its own only provides this consistency 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 for GDPS/PPRC in 3.1.1, "Protecting data integrity and data availability" on page 44 and for GDPS/PPRC HyperSwap Manager in 4.1.1, "Protecting data integrity and data availability" on page 80.

Addressing z/OS device limits

As clients implement IT resiliency solutions that rely on multiple copies of data, more and 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 datasets to fewer larger volumes, but even with that, there are times when this may 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)

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

Support is included with GDPS release 3.7 for GDPS/PPRC and GDPS/PPRC HyperSwap Manager to allow definition, system IPL automation, and HyperSwap capability for configurations where PPRC secondary devices are in an alternate subchannel set. With this feature, GDPS can now support PPRC configurations with nearly 64K device pairs. The prerequisites for this support include:

- Systems are z10™ or z9® running z/OS 1.10 or later with relevant z/OS maintenance
- GDPS release 3.7 on all systems
- A disk configuration that meets all requirements to be HyperSwap ready
- Disk control units all support PPRC Failover/Failback

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

- Some devices cannot be included in an alternate subchannel set even though they are mirrored in the configuration. In particular, the following must be defined in the base subchannel set 0 (MSS0):
  - IPL volumes (including Stand Alone Dump volumes)
  - IODF devices
  - zLinux devices (as part of a GDPS xDR configuration)
  - Fixed Block Architecture (FBA) open-LUN devices managed by GDPS
  - GDPS managed FlashCopy target devices
- The four-digit device ranges for both the primary and secondary devices must be the same, so you may need to move definitions of PAV aliases to make room for the secondaries in the alternate subchannel set.
- Tools and processes that rely on the following z/OS commands will require update because they are not applicable for devices in an alternate subchannel set.

  ```
  DISPLAY UCB,,,xxxx
  VARY xxxx,ONLINE
  VARY xxxx,OFFLINE
  ```
- The legacy PPRC Dynamic Address Switching (P/DAS) function is not compatible with secondary devices in an alternate subchannel set.

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 and GDPS/PPRC HyperSwap Manager offerings are 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. It involves a System Data Mover (SDM) that is found only in z/OS, with supporting code in the primary disk subsystems. XRC maintains a copy of the data asynchronously at a remote location. 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 document.

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 some amount of data that has not yet been collected when a disaster hits. As a result, XRC can only be used when your Recovery Point Objective (RPO) 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, so if there is insufficient capacity (MIPS, storage, and I/O configuration) available to the SDM, this will result in longer delays collecting the updates from the primary disk subsystems, 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 could cause the secondary volumes to drift further behind at peak times.

- The use of device blocking

  Enabling blocking for devices will result in I/O write activity to be paused for devices with very 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 very high write rates will result in delays being inserted into the application's I/O response to prevent the secondary disk from falling behind. This option will slow the I/O activity resulting in a smaller RPO and is less pervasive than device blocking.

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

- Unexpected peaks in the workload
- An under-configured 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, refer to the chapter entitled "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 remote copied by an SDM. This capability is provided in the following operating systems:

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

Note that XRC does *not* support FBA devices.

It is also possible to use XRC to remote copy volumes being used by System z 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 be in the same LSS to provide consistency. For more information, refer to the section entitled "Understanding the Importance of Timestamped Writes" in the latest revision of *z/OS DFSMS Advanced Copy Services,* SC35-0428.

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 impact 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 have to use channel extenders and telecommunication lines.

XRC data consistency

XRC uses time stamps and consistency groups to ensure your data is consistent across the copy operation. When an XRC pair is established, the primary disk subsystem notifies all systems in the 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 across multiple LSSs.

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 (refer to 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 signalled 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 timestamped copy of the update is kept in the primary disk subsystems cache. Dependent writes can proceed at this point.

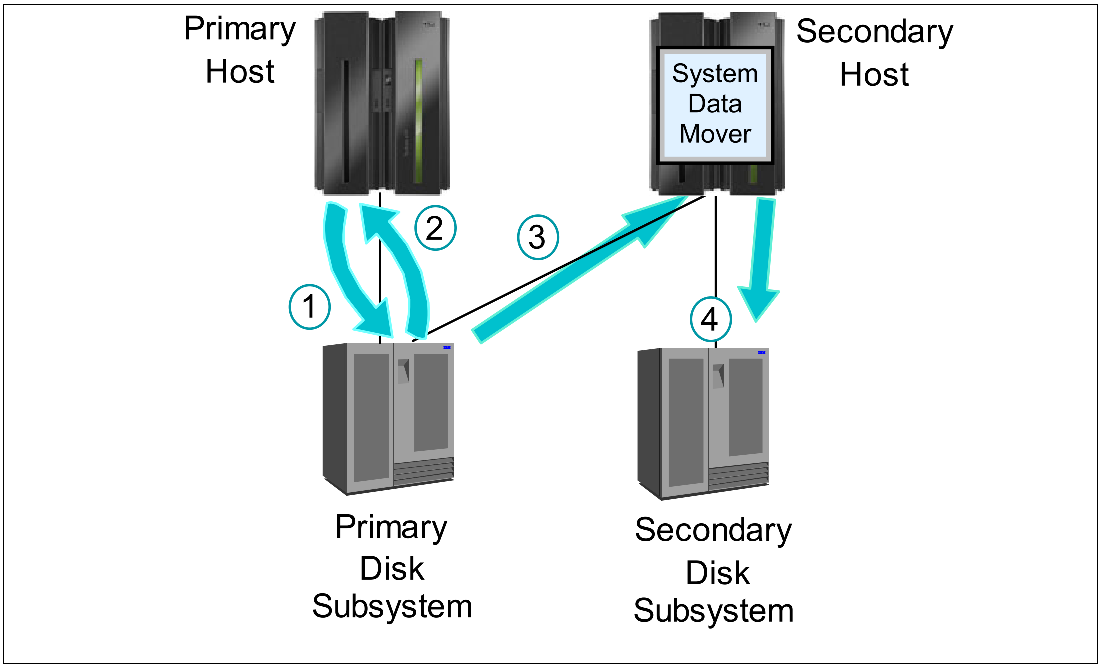


Figure 2-4 Data flow when using z/OS Global Mirror

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 off-loaded 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 will apply 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 hundreds of volumes. However, if the number of volumes to be mirrored exceeds the number that can be effectively managed by a single SDM instance, you can use the Coupled XRC support to extend the number of volumes that can be managed into the thousands. CXRC provides the ability to couple multiple SDMs in different LPARs together into a master session. CXRC coordinates the consistency of data for 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 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). Whereas Coupled XRC links SDMs running in different LPARs, MXRC allows you to have up to 20 SDMs in a single LPAR, of which 13 can be coupled together into a cluster. Up to 14 SDM clusters can then be coupled together, allowing for an architectural limit 182 SDMs.

Multiple Reader

XRC Multiple Reader allows automatic load balancing over multiple readers in an XRC environment. The function can provide increased parallelism through multiple SDM readers and improved throughput for XRC remote mirroring configurations.

It can allow XRC to:

- 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

Prior to the introduction of Multiple Readers, you needed to plan carefully to balance the primary volume update rate versus the rate that the SDM could “drain” the data. If the drain rate could not keep up with the update rate, there was a potential to impact application I/O performance.

GDPS/XRC has been enhanced to exploit this multireader function, and thus provide these benefits.

Extended Distance FICON

Extended Distance FICON is an improvement focused on providing XRC customers a choice of selecting simpler, 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 requires is supported only on the IBM system z10 and higher servers, and the IBM System Storage DS8000 disk subsystems.

SDM Offload to zIIP

Another recent enhancement is the capability to run the System Data Mover (SDM) on one of the specialty engines referred to as IBM System z9® or IBM System z10® Integrated Information Processor (zIIP). By offloading some of the SDM workload to a zIIP, better price performance and improved utilization 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 utilization at the mirrored site. Another benefit is that with an investment of a zIIP specialty processor at the mirrored site, you may now be able to cost justify the investment in and implementation of a disaster recovery solution that can lower server utilization at the mirrored site, while at the same time reduce software and hardware fees.

Hardware prerequisites

XRC requires 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 have at least as large as the primary device.

XRC is also supported on disk subsystems from other vendors who 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 in time you may have to remote copy from the recovery site back to the production site. To accomplish this, the IBM z/OS Global Mirror feature code must be installed in the secondary disk subsystems, and the primary device must be at least as large as the secondary device. Therefore, it makes sense to maintain a symmetrical configuration across both primary and secondary devices.

An additional requirement is that all the systems writing to the primary volumes must be connected to the same timer 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 provides a proven disk mirroring foundation for an enterprise Disaster Recovery solution that provides large scalability and good performance.

XRC is a recommended 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
- The ability to 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 two-site disaster recovery and backup solution for the System z and open systems environments. Using asynchronous technology, Global Mirror operates over Fibre Channel Protocol (FCP) communication links and is designed to maintain a consistent and restartable copy of data at a remote site that can be located at virtually unlimited distances from the local site.

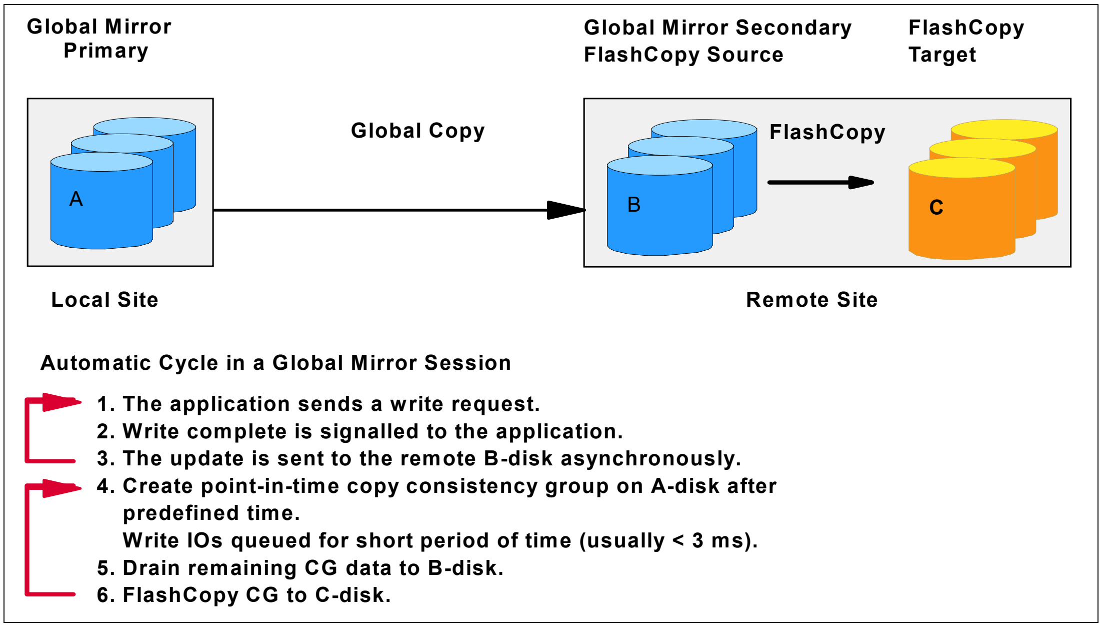
Global Mirror works by using three sets of disks, as shown in Figure 2-5 on page 29. Global Copy (PPRC Extended Distance (PPRC-XD)) is used to continually transmit data from the primary (A) to secondary (B) volumes, using the out-of-sync bitmap to determine what needs to be transmitted. Note that Global 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.

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 are residing on the Global Mirror primary for the next consistency group will be drained to the secondaries. After all of the primary devices have been drained, an inband 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. Immediately after the FlashCopy process is logically complete, the primary disk systems are notified to continue with the Global Copy process. For more information about FlashCopy, refer to 2.6, “FlashCopy” on page 31.

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 code in the disk subsystems. You can have up to 17 mirrored pairs (one Master primary and secondary pair, and 16 Subordinate primary and secondary pairs) in a pool.



Automatic Cycle in a Global Mirror Session

1. The application sends a write request.
2. Write complete is signalled to the application.
3. The update is sent to the remote B-disk asynchronously.
4. Create point-in-time copy consistency group on A-disk after predefined time.
   Write IOs queued for short period of time (usually < 3 ms).
5. Drain remaining CG data to B-disk.
6. FlashCopy CG to C-disk.

Figure 2-5 Global Mirror: How it works

Recovery Point Objective

Because Global Mirror is an asynchronous remote copy solution, there will always be some amount of data that will need to be recreated following a disaster. As a result, Global Mirror can only be used when your 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 could 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 five to ten 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 due to unexpected peaks in the workload or an under-configured 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, refer to *IBM System Storage DS8000: Copy Services with IBM System z*, SG24-6787.

Supported platforms

Because Global Mirror is built on Global Copy (PPRC Extended Distance function), any platform that is supported by IBM Metro Mirror is also supported by Global Mirror. Currently this support is limited to the IBM Enterprise Storage Server (ESS) and the IBM DS8000 family of products.

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 impact 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 subsystem. 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.

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 FBA devices and removes the requirement for timestamped updates that is imposed by XRC.

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

2.4.4 Combining 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, refer to Chapter 8, “Combining Local/Metro continuous availability with out-of-region disaster recovery” on page 173. Combining the technologies of Metro Mirror with either Global Mirror or XRC (also referred to as z/OS Global Mirror in this section) allows customers 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 may also allow customers 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 located in a recovery site typically out of the local metropolitan region. As described in 2.4.3, “Global Mirror” on page 28, a fourth set of disks, also located 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 FBA devices can be mirrored to the recovery site protecting both System z and open system data.

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

Metro z/OS Global Mirror

GDPS Metro/z/OS Global Mirror (MzGM) is a multitarget 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 located in a recovery site, typically out of the local metropolitan region. Because XRC only supports CKD devices, only System z data can be mirrored to the recovery site. However, since both PPRC and XRC are supported by multiple storage vendors, this solution provides flexibility that MGM cannot.

For enterprises looking to protect System z data, MzGM delivers a three-copy replication strategy to provide continuous availability for day-to-day disruptions, while protecting critical business data and functions in the event of a wide-scale disruption.

2.5 Tape resident data

Operational data, that is, data that is used directly by applications supporting end 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 only resides 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, as well you can mirror your tape-resident data. While not specifically integrated into GDPS, the IBM Virtualization Engine TS7700 provides comprehensive support for replication of tape data. See *IBM Virtualization Engine TS7740 R1.5 and TS7720: New Virtualization Options for Mainframe Servers*, SG24-7712 for more information on 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. Additionally, there is a Dataset 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 under the covers by the disk subsystem, with minimal impact on other disk subsystem activities. The volumes created when you FlashCopy 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 no use to you either.

So, how do you protect yourself from such a scenario? One way (our recommended 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 some number of hours behind the primary volumes, but even data a few hours old of consistent data is better than unusable current data.

An additional benefit of FlashCopy is that it provides the ability to do 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 a number of different 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 only contains 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(s) will be terminated.

INCREMENTAL 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, giving you the option to perform a FlashCopy on a more frequent basis.

Zero Suspend Creates a recoverable set of tertiary disks for recovery testing without suspending the XRC operation at all. This allows Disaster Recovery (DR) testing to be performed without ever losing the DR capability. Before this support, in order to produce a consistent tertiary copy, you needed to suspend XRC for all volumes, FlashCopy secondary volumes, and resynchronize XRC sessions.

Taking FlashCopies with the COPY option is only supported with the full GDPS/PPRC or GDPS/XRC offerings. GDPS/PPRC HyperSwap Manager takes all FlashCopies with the NOCOPY option. Details of how the specific FlashCopy options are used by each offering are discussed in the specific chapter of the offering.

If you plan on exploiting FlashCopy, remember that the source and target volumes must be within the same physical disk subsystem. So, if you want to be able to FlashCopy all your secondary volumes, it means that only half the volumes in the subsystem can be allocated as secondaries, with the other half reserved for use as FlashCopy targets. If you use all the volumes as secondaries and subsequently decide to exploit FlashCopy, you are presented with the task of having to move half the secondary volumes to another subsystem to free up volumes for FlashCopy use.

Also, and this is particularly relevant to GDPS/PPRC customers, remember that if you did a site switch to run in the recovery site, at some point you will want to resynchronize to move back to the production site. To protect those volumes, you might also want to provide a FlashCopy capability (along with spare volumes) in the primary subsystems as well.

User-initiated FlashCopy

User-initiated FlashCopy supports FlashCopy of all defined FlashCopy volumes using panel commands and GDPS script keywords. Data consistency when doing a User-initiated FlashCopy is a user responsibility. If a FlashCopy has already been taken, a prompt will be displayed to allow the user to decide whether to overwrite it or not.

For more information about FlashCopy, refer to *Implementing ESS Copy Services with IBM eServer zSeries,* SG24-5680.

FlashCopy Space Efficient (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 could be used like any other DS8000 volume. However, the intended and only *recommended* 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 (in other words, 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

Details of how FlashCopy SE is used by each offering is discussed in the specific chapter of the offering.

2.7 Automation

If you have challenging Recovery Time and Recovery Point objectives, implementing disk remote copy, 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, it is important to realize that they are only enabling technologies. In order 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 RTO is a “challenge” to you. If you have done 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, elongate the whole process. Operators must react to events as they happen, while consulting recovery documentation. On the other hand, 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 out of date. 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. It does not matter how many people watch it, 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 programmed to behave in. You do not need to worry about out of date procedures being used. The operators can concentrate on handing 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. As long as 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 life saver.

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 may 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 recovery 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 exploiting the GDPS-provided automation, you save all the effort to design and develop this code yourself, and also benefit from the IBM experience with hundreds of customers across your industry as well as 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 could be used to help meet customer's temporary peak workload requirements.

The only capabilities described in this section are the ones exploited by GDPS. There are other capabilities to upgrade the server capacity, either on a temporary or permanent basis, but they are not discussed in this section.

2.8.1 Capacity Backup Upgrade

Capacity Backup Upgrade (CBU) for System z processors provides reserved emergency backup server capacity for situations in which you lose capacity in another part of your establishment. It helps you to recover by adding reserved capacity on a designated System z system. A CBU system normally operates with a base server configuration and with a preconfigured number of additional central processors (CPs) 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 very quickly enable 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 CPs 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 an appropriate number of Reserved CPs were defined in the LPAR Image profiles).

Before CBU can be activated manually or by GDPS, authentication needs to be performed to confirm that the customer has a valid agreement in place. The System z Remote Service Facility (RSF) was required for the authentication process. New function is now available with GDPS V3.5 and higher to allow you to specify one of the following:

- Enter the keyword (password) yourself as part of the script statement.
- Prompt the operator to enter the keyword (password).
- Continue using the RSF process to obtain the keyword.

CBU is available for all PU types on IBM System z. On previous generations, it was only available for CPs. IFLs, ICFs, and zAAPs were not supported for CBU. The CBU contract allows for an agreed-upon number of tests over the period of the contract.

For more information about CBU, refer to *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. On/Off CoD can be used for customer peak workload requirements, for any length of time, and has a daily hardware and software charge.

On/Off CoD helps customers whose business conditions do not justify a permanent upgrade in capacity to contain workload spikes that may exceed permanent capacity such 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, and it is restricted to double the currently-installed capacity.

New function is now available with GDPS V3.5 and higher for new keywords in GDPS scripts to support activation and deactivation of the On/Off CoD function.

2.9 Cross-site connectivity considerations

When setting up a recovery site may entail a sizeable capital investment to get started, you may 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 impact the recovery capability you can provide. Conversely, the type of recovery capability you want to provide will impact the types of connectivity you can use. So, in this section, we list the connections that must be provided, from a simple remote copy configuration right through to an active/active workload configuration.

This section briefly reviews the types of cross-site connections that you must provide for the different GDPS solutions, and the technology that has to be used to provide that connectivity. All these discussions relate solely to cross-site connectivity. We assume that you already have whatever intra-site 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 based solutions

For PPRC with GDPS, the secondary disks must be defined to and also be channel-accessible to the production systems for GDPS to be able to manage those devices. If you only need to manage those devices, minimal connectivity is required. But if you foresee a situation where systems in the production site will be running off the disks in the recovery site, then you need to provide connectivity equivalent to that provided to the corresponding primary volumes in the production site.

Depending on your director/switch configuration, you may be able to share the director-to-director links between channel and PPRC connections; refer to *IBM System z Connectivity Handbook,* SG24-5444, for more information.

XRC 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.

2.9.2 Remote copy links

You will obviously need connectivity for your remote copy activity. This will either be between storage subsystems (for PPRC or Global Mirror), or from the SDM system to the primary disks (for XRC).

PPRC and Global Mirror based solutions

The IBM Metro Mirror and Global Mirror implementations leverage Fiber 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 ESCON® connections for PPRC, we strongly recommend the use of FCP links for best performance over distance.

XRC based solutions

If you are using XRC, the System Data Movers (SDMs) will typically be located 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 27, an enhancement to the industry standard FICON architecture (FC-SB-3) helps avoid degradation of performance at extended distances and this may also benefit XRC applications within 300 km where channel extension technology had previously been required to obtain adequate performance.

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 multi-site Parallel Sysplex, you will need to provide coupling link connectivity between sites.

While there are some older coupling link technologies that are still in some client environments, we strongly recommend that either Parallel Sysplex InfiniBand (PSIFB) Long Reach or ISC-3 peer mode links be used to provide this connectivity. Using these coupling links offers the following advantages:

- Support for connections across metropolitan distances when used in conjunction with DWDMs
- Better performance than legacy “compatibility mode” ISC-3 links
- Support for STP message exchanges

Table 2-1 presents a summary of 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 (peer mode)	2 Gbps 1 Gbps[a]	10 km 20 km[b]	100 km
PSIFB Long Reach (Feature 0168)	5.0 Gbps 2.5 Gbps[c]	10 km	100 km
PSIFB (Feature 0163 - for use within a data center)	6 GBytes/s 3 GBytes/s[d]	150 meters	NA

a. RPQ 8P2197 provides an ISC-3 Daughter Card that clocks at 1 Gbps.
b. Requires RPQ 8P2197 and 8P2263 (System z Extended Distance).
c. 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.
d. The PSIFB links negotiate to 12x IB-SDR link data rate of 3 GBytes/sec when connected to System z9 servers.

2.9.4 IBM 9037 Sysplex Timer

The IBM 9037 Sysplex Timer provides the synchronization for the time of day (TOD) clocks of multiple servers, and thereby allows events occurring on different servers to be properly sequenced in time. The Sysplex Timer has two types of links:

- ETR links used to connect the Sysplex Timer to the servers it is synchronizing
- CLO links used to connect the two Sysplex Timers in an Expanded Availability configuration so as to maintain synchronization between them

When using a DWDM, the maximum supported distance between a Sysplex Timer and a server it is connected to is 100 km. However, the maximum supported distance, again when using a DWDM, for the CLO link is 40 km. This means that if the two sites are more than 40 km apart, it is recommended that you provide an intermediary site for one of the Sysplex Timers that is within 40 km of the site containing the other Sysplex Timer in order to avoid a single point of failure that affects both Sysplex Timers.

2.9.5 Server Time Protocol (STP)

Server Time Protocol (STP) is a server-wide facility that is implemented in the Licensed Internal Code (LIC) of the IBM System z10, System z9, zSeries z990, and z890 servers. It is designed to provide the capability for multiple z10, z9 EC, z9 BC, z990, and z890 servers to maintain time synchronization with each other. STP is the follow-on to the 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), as well as 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 that can be used to transport STP messages are the Inter System Channel-3 (ISC-3) links configured in peer mode, Integrated Cluster Bus-3 (ICB-3) links, Integrated Cluster Bus-4 (ICB-4) links, and Parallel Sysplex using InfiniBand (PSIFB) links.

If you are configuring a sysplex across two or more sites, you will need to synchronize servers in multiple sites. In this case, STP requires either ISC-3 links running in peer mode or PSIFB Long Reach Links. For more information about Server Time Protocol, refer to *Server Time Protocol Planning Guide*, SG24-7280, and to *Server Time Protocol Implementation Guide*, SG24-7281.

2.9.6 XCF signalling

One of the requirements of 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 signalling structures in a CF and channel-to-channel adaptors. 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 recommend providing *both* XCF signalling structures and CTCs for XCF cross-site communication.

2.9.7 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 could use SNA consoles; however, be aware that they cannot be used until VTAM® is started, so they cannot be used for initial system loading.

2.9.8 Connectivity options

Having discussed what you need to connect across the two sites, we now briefly review the most common options for providing that connectivity. There are a number of ways to provide all this connectivity, from direct channel connection through to DWDMs. Table 2-2 shows the different options.

Table 2-2 Cross-site connectivity options

Connection type	Direct (Unrepeated)	Switch/Director or Cascaded Directors	DWDM	Channel extender
Server to disk	Yes	Yes	Yes	Yes
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 signalling	Yes	Yes (CTC) No (coupling links)	Yes	Yes (CTC only) No (coupling links)
HMC/consoles	Yes	Yes	Yes	Yes

Note that the distance supported varies both by device type and also by the connectivity method.

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

FICON switches/directors

For information about System z qualified FICON and Fibre Channel Protocol (FCP) products, as well as products that support intermixing of FICON and FCP within the same physical FC switch or FICON Director, go to:

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

The maximum unrepeated 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 may 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, refer to *IBM System z Connectivity Handbook,* SG24-5444.

Wavelength Division Multiplexing (WDM)

A 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 bitstreams over a single fiber link, thereby making optimal use of the available bandwidth.

WDM solutions that support the protocols described in this document 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, there are several qualified WDM solutions that support the following key protocols used in a GDPS solution:

- Enterprise Systems Connection (ESCON)
- IBM 9037 Sysplex Timer (ETR/CLO)
- 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 other, non-System z, 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 Redpaper™ publications for each product, is available at the Resource Link™ website at:

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

Refer to “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 unlimited[4].

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.

On the other hand, channel extenders typically do not support the same range of protocols as DWDMs. In a System z 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, refer to *IBM System z Connectivity Handbook,* SG24-5444.

More information about channel extenders that have been qualified to work with IBM storage is available on the following website:

http://www.ibm.com/systems/storage/disk/ds8000/interop.pdf

2.9.9 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

[4] For information about the impact of distance on response times when using channel extenders, contact your IBM representative to obtain the white paper entitled “*The effect of IU pacing on XRC FICON performance at distance*” which is available on the IBM intranet at: http://w3.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP100440

recovery actions you take are very different depending on whether the failure you just detected is a connectivity failure or a real site failure.

If you only have 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. In addition 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 enterprises ability to recover in the event of a real 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.

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 could also be used to create a copy of the secondary devices to ensure a consistent copy of the data is available should a disaster type event occur 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, could result in compromising the enterprise's ability to recover from a disastrous event, or in extended recovery times and exposure to additional data loss.

2.11 Summary

We have now covered the major building blocks in an IT resilience solution. We have 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 subjects 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.

3

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 GPDS/PPRC fall into two categories: protecting your data, and controlling the GDPS-managed resources. Some of the provided functions are:

- Protecting your data:
  - Ensuring the consistency of the secondary 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
  - Management of the remote copy configuration for System z and non-System z platform data
- Controlling the GDPS-managed resources 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 System z 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, the majority of outages are planned, and even among unplanned outages, the majority are not disasters. GDPS/PPRC provides capabilities to help provide the required levels of availability across these outages as well as in a disaster scenario. These capabilities are described in this chapter.

3.1.1 Protecting data integrity and data availability

2.2, "Data consistency" on page 15 mentions 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 of how GDPS automation in GDPS/PPRC is designed to provide both data consistency in the event of mirroring problems and data availability in the event of disk problems.

There are two types of disk problems that trigger a GDPS automated reaction:

- 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 discussed in the section "Freeze function."

- 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 is not responsive. This is discussed in "HyperSwap function" on page 46.

Freeze function

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 `CGROUP FREEZE` and `CGROUP RUN` commands, which have been implemented as part of Metro Mirror and also by other enterprise disk vendors. In this way, as long as the disk hardware supports `CGROUP FREEZE/RUN` commands, GDPS can ensure consistency across all data in the Sysplex (consistency group) irrespective of disk hardware type. This preferred approach differs from proprietary hardware approaches that only work for one type of disk hardware. For a related introduction to data consistency with synchronous disk mirroring, refer to "PPRC data consistency" on page 21.

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

- Remote copy is suspended for all device pairs in the configuration.
- All paths between the indicated pair of logical subsystems (LSS) are removed.
- 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 will be issued to the affected devices until the ELB is reset with a `CGROUP RUN` command or until it times out (the consistency group timer setting commonly defaults to 120 seconds or 2 minutes).

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 will issue a `CGROUP FREEZE` for all LSS pairs containing devices in the GDPS configuration. A very important point is that 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 would not be unusual to see short gaps between the times when the `FREEZE` command is issued to each disk subsystem, but because of the ELB, this gap is not a problem.

After GDPS detects that all remote copy sessions have been suspended, and the consistency of the remote disks is protected, GDPS will either issue a `CGROUP RUN` command to the LSSs, allowing them to continue operation in remote copy-suspended mode, or system reset all production systems (thereby ensuring that no non-mirrored updates can be applied), depending on your GDPS FREEZE policy as described in “Freeze policies” on page 45.

GDPS/PPRC uses a combination of storage subsystem and sysplex triggers to capture, at the first indication of a potential disaster, a data-consistent secondary site copy of your data using the `CGROUP FREEZE` function. In this way, the consistent image of the data is ensured on the secondary copy at the very first sign of a disaster, even before production applications are aware of I/O errors. Ensuring the data consistency of the secondary copy ensures that a normal restart can be performed, instead of having to perform DBMS recovery actions. This is the essential design element of GDPS to minimize the time to recover the critical workload in the event of a disaster at the primary site.

For more information about the implementation of PPRC and IBM’s Metro Mirror refer to *DS8000 Copy Services for IBM System z*, SG24-6787.

Freeze policies

Following a Freeze event, GDPS always performs a `CGROUP FREEZE` to create a consistent set of secondary volumes. The action that GDPS takes subsequent to the `CGROUP FREEZE` is specified by the installation in the GDPS freeze policy. Prior to GDPS/PPRC V3.7, you basically had two options:

- GO - GDPS will allow the production systems to continue operation after mirroring has been suspended.
- STOP - GDPS will reset the production systems while I/O is suspended.

GDPS V3.7 (with APAR) has introduced enhanced Freeze and Swap policy options. The next two sections FREEZE and STOP and FREEZE and GO are provided for completeness for those customers using V3.6 and for V3.7 and V3.8 customers who prefer not to change their Freeze and Swap policies. If you are installing GDPS/PPRC for the first time or an existing customer running V3.7 or higher, it is recommended that you change your Freeze and Swap policies to the ones described in “Enhanced Freeze and Swap policy options (V3.7 and V3.8)” on page 48.

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 policy choice 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 down before continuing.

If you are using duplexed structures along with a FREEZE and STOP policy, it may seem that you are guaranteed to be able to use the duplexed instance of your structures in the event you have to recover and restart your workload with the frozen secondary copy of your disks. However, this is not always the case! There could be rolling disaster scenarios where prior to, following, or during the freeze event, there is some sort of 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 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.

Note: If FREEZE and STOP is used, and the event that caused GDPS to take action was a transient event rather than a real disaster, you will have stopped all production systems unnecessarily.

FREEZE and GO

If you can accept an RPO that is *not necessarily* zero, you may decide to let the production systems *continue operation* after the secondary volumes have been protected. A FREEZE and GO policy is recommended for this case. This way you avoid an unnecessary outage if the trigger were to be only a transient event.

On the other hand, if the trigger is the first sign of an actual disaster, you could continue operating for some amount of time before all systems actually fail. Any updates made to the primary volumes during this time will not have been remote copied to the secondary disk, and therefore are lost. In addition, because the 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 recovery must be used to restart applications, resulting in elongated recovery times.

Considerations

The decision of whether to implement a STOP or GO freeze policy is really a business decision, rather than an IT decision. If your transactions are very high value, it may be more important to ensure that no transactions are ever lost, so you may decide on FREEZE and STOP. If you have huge volumes of relatively low value transactions, you may be willing to risk some lost data in return for avoiding unneeded outages with a FREEZE and GO policy.

Most installations start with a FREEZE and GO policy. Companies that have an RPO of zero will typically then move on and implement a FREEZE and STOP policy after the implementation is proven stable.

HyperSwap function

In the event that there is a problem writing or accessing the *primary* disk because of either a primary disk hard failure or because the disk subsystem is not accessible or not responsive, then there is a need to swap from the primary disk subsystems to the secondary disk subsystems.

GDPS/PPRC delivers a powerful function known as “HyperSwap.” HyperSwap provides the ability to non-disruptively swap from using the primary volume of a mirrored pair to using what had been the secondary volume. Prior to the availability of HyperSwap, an IPL was required on every system if you wanted to switch and run from the secondary volumes, meaning that it was not possible to maintain application availability across a switch from primary to secondary volumes.

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

There are two ways that HyperSwap can be invoked:

- Planned HyperSwap

  A planned HyperSwap is invoked manually by operator action using GDPS facilities. One example of a planned HyperSwap would be 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 failure of a primary disk device.

In both cases, the systems that are using the primary volumes will experience a temporary pause in processing. During this pause, the disk mirroring configuration is changed to allow use of the secondary volumes (and mirroring may be established in the opposite direction, depending on the option selected), the UCBs for the primary devices are updated to point to the formerly secondary volumes, and then the systems resume operation.

In benchmark measurements at IBM using currently supported releases of GDPS, the I/O hold time for an unplanned HyperSwap is generally less than 30 seconds for even very large configurations (for example, a 10-way Sysplex with approximately 20,000 mirrored volume pairs, or even a 30-way Sysplex with a more moderately sized disk configuration). Most implementations in the world are actually much smaller than this and typical I/O hold times using the most current storage and server hardware are measured in seconds. While results will obviously depend on your configuration, these numbers give you a high-end figure for what to expect.

A HyperSwap in GDPS/PPRC affects *all* mirrored LSSs with devices in the configuration (single consistency group). For example, if one single mirrored volume were to fail, and HyperSwap is invoked, processing would be swapped to the secondary copy of *all* mirrored volumes in the configuration, including those in other, unaffected, subsystems. The reason for this is that to maintain disaster readiness, all primary volumes *must* be in the same site. If HyperSwap were to only swap the failed LSS, you would have some primaries in one site, and the remainder in the other site.

Why is this necessary?

Consider the configuration in Figure 3-1 on page 48. This is what might happen if only the volumes of a single LSS or subsystem were HyperSwapped without swapping the whole consistency group. What would happen if there were a remote copy failure at 15:00? The secondary disks in both sites would be frozen at 15:00 and the primary disks (in the case of a FREEZE and GO policy) would 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 the volumes in that site would be consistent and if you had the disaster at 15:10, you would lose 10 minutes worth of data with the GO policy, but at least all the data in Site2 is usable. Using a FREEZE and STOP policy is not any better for this partial swap scenario because with a mix of primary disks in either site, you would have to maintain I/O configurations that could match *every* possible combination simply to IPL any systems. More likely, you would 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.

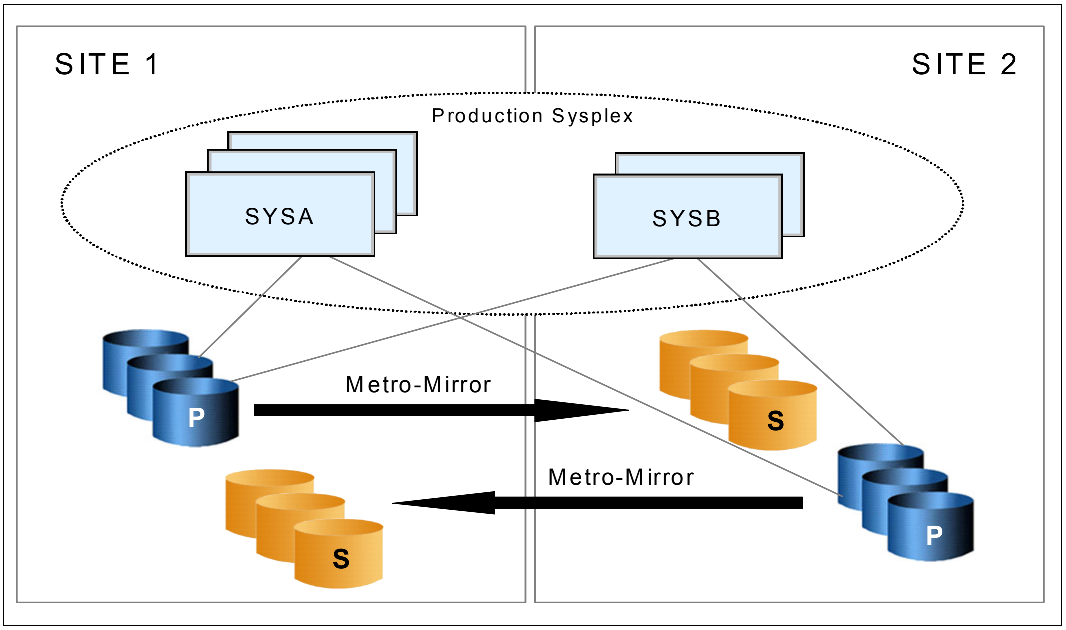


Figure 3-1 Unworkable Metro Mirror disk configuration

HyperSwap policy options

If you elect to use HyperSwap, there are additional options you can specify in your freeze policy that indicate a HyperSwap should be attempted following a HyperSwap trigger. Instead of FREEZE and STOP, you would specify SWAP,STOP, and instead of FREEZE and GO, you would specify SWAP,GO.

- If SWAP,STOP or SWAP,GO is specified, and you receive a HyperSwap trigger, any systems that successfully process the HyperSwap will be swapped and any systems that *cannot* process the HyperSwap will be system reset. Remember that the HyperSwap trigger indicates an error on a primary device, so any systems that continue to use that device would soon fail anyway.
- If you receive a freeze trigger, I/O to the systems will either be resumed after the secondary is consistent (if GO is specified) or the systems will be reset (if STOP is specified).

Enhanced Freeze and Swap policy options (V3.7 and V3.8)

The single Freeze policy option specified by SWAP,GO or SWAP,STOP was confusing and restrictive because it attempted to specify with a single specification what actions should be taken for mirroring problems and primary disk failures. It was not clear that the second parameter, GO or STOP, referred to actions to be taken if a mirroring problem was detected. It also did not allow the customer the flexibility to have different combinations, such as GO for mirroring problems and STOP for disk failures. Last but not the least, it resulted in inconsistent handling of certain triggers. The same trigger was sometimes handled as a mirroring problem and sometimes as a disk failure.

An enhancement in V3.8 base code and V3.7 code with an APAR allows you to split the policy options and separately specify what actions GDPS should take when

- Mirroring problems (Freeze triggers) are detected
- Disk failures (HyperSwap triggers) are detected

Mirroring Problems

Actions to be taken when Freeze triggers are detected can now be specified as either:

- PPRCFAILURE = GO

  After performing the Freeze, the production systems continue to update the primary disks. This is similar to Freeze and Go. Recovering on the secondary disks is safe since the secondary disk was frozen in a consistent state. However, recovering on secondary disks results in loss of any data that was written to the primary disk after the freeze.

- PPRCFAILURE = STOP

  After performing the Freeze, the production systems will be quiesced, resulting in all the work that is updating the primary PPRC devices being stopped. This is similar to Freeze and Stop. It is the only option that guarantees no data loss in case recovery on the secondary disk would be required.

- PPRCFAILURE = COND

  Field experience has shown that most of the 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. Some examples of these events are connectivity problems to the secondary disks or secondary disk subsystem failure conditions.

  If PPRCFAILURE = COND is specified, the action after the Freeze is **conditional** on the results of a new disk subsystem query. GDPS working in synergy with a new disk subsystem query function can determine the status of the secondary disk subsystems when a Freeze trigger is detected. If GDPS detects a secondary disk subsystem failure, then the actions are the same as GO, the production systems continue to update the primary disks; otherwise processing is the same as STOP.

  This option can improve the capability of zero data loss, and reduce the potential of loss of availability due to “False Freeze” events. It requires that all disk sub systems managed by GDPS support the *Query Storage Controller Status feature.*

Primary disk failures

Actions to be taken when HyperSwap triggers are detected can now be specified as either:

- PRIMARYFAILURE = GO

  After performing the Freeze, the disks are not swapped, and the production systems continue to update the primary disks. This is similar to PPRCFAILURE = GO.

- PRIMARYFAILURE = STOP

  After performing the Freeze, the disks are not swapped, and the production systems will be quiesced, resulting in all the work that is updating the primary PPRC devices being stopped. This is similar to PPRCFAILURE = STOP.

- PRIMARYFAILURE = SWAP,GO or SWAP,STOP

  The first parameter SWAP indicates that after performing the Freeze, GDPS should perform an unplanned HyperSwap if a primary disk problem is detected for one of the GDPS managed PPRCed disks. This is done only if the environment is enabled for HyperSwap at the time the failure occurs. GDPS transparently swaps to using the secondary disks instead of the primary, and applications do not incur any downtime.

With this option, you must also decide on a policy for what GDPS should do if the environment is not enabled for HyperSwap at the time the failure occurs.

The options for the second parameter are:

- GO

  After performing the Freeze, the disks are not swapped, and the production systems continue to update the primary disks.

- STOP

  After performing the Freeze, the disks are not swapped, and the production systems will be quiesced, resulting in all the work that is updating the primary PPRC devices being stopped.

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 re-established with failback processing, the original primary devices become secondary devices and a resynchronization of changed tracks takes place.

GDPS/PPRC transparently exploits the PPRC FO/FB capability if it is installed on the disk subsystems. This support mitigates RPO exposures by reducing the amount of time needed to resynchronize mirroring after a HyperSwap. Of course, the resync time will depend on how long mirroring was suspended and the number of changed tracks that must be transferred.

All disk subsystems in your GDPS configuration, in both Site1 and Site2, must support PPRC Failover/Failback for GDPS to exploit this capability.

Protection during IPL

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 usage) perspective but physically from a PPRC perspective have simplex or primary status.

For both planned and unplanned HyperSwap, GDPS points to the new set of primary disks. However, for both these actions, if the former primaries are accessible, although they are considered to be the secondary disks from a GDPS perspective, they are still usable by applications since their actual PPRC status is not secondary. That is why it is possible to accidentally IPL from the wrong set of disks. Accidentally using the wrong set of disks could result in a potential data integrity or data loss problem.

Prior to V3.8, GDPS/PPRC provided IPL protection early in the IPL process, during initialization of GDPS, by not allowing a system IPLed on the wrong set of disks to continue running. This protection has been enhanced with V3.8 by GDPS rejecting the IPL if it is attempted using GDPS (3270 panel, Web GUI or a GDPS script) against the incorrect set of disks.

3.1.2 Protecting tape data

Although most of your critical data will be resident on disk, it is possible that some of the data you require following a disaster may reside 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 a single integrated recovery process when using the legacy IBM TotalStorage® 3494-based Virtual Tape Subsystem in a Peer-to-Peer configuration. Additionally, although not specifically integrated into GDPS, the IBM Virtualization Engine TS7700 provides comprehensive support for replication of tape data. See *IBM Virtualization Engine TS7740 R1.5 and TS7720: New Virtualization Options for Mainframe Servers*, SG24-7712 for more information on the TS7700 technology that complements GDPS for tape data.

3.1.3 Protecting distributed (FBA) data

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

- System z or zSeries disks

  Prior to Open LUN support, GDPS only managed disks that were used by System z systems, although disks could be z/VM, VSE, or Linux on System z disks. All these disks are formatted as Count-Key-Data (CKD) disks, the traditional mainframe format. There was no specific name used for these subsystems.

  However, with the introduction of GDPS support for disks used by platforms other than System z, we have had to expand the terminology in some places to differentiate between disks used by systems running on a System z server (CKD), and those used by the non-System z server.

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

- Open LUN disks

  Disks that are used by systems other than those running on zSeries are traditionally formatted as Fixed Block Architecture (FBA). In this book, we generally use the term “Open LUN disks” to refer to such devices. However, the term “FBA disks” is used sometimes as well; both terms are used interchangeably.

  GDPS/PPRC V3.8 has added support for SCSI attached FB disk used by native Linux for System z systems under GDPS xDR[a] control. There is a need to differentiate between FB disk used by other distributed systems (including Linux on System z systems not under xDR control) and the FB disk used by native Linux on System z xDR.

[a] Refer to 3.3.1, “Multiplatform Resiliency for System z (also known as xDR)” on page 58 for more details.

GDPS/PPRC can manage the mirroring of FBA devices (also known as Open LUNs) in use by non-mainframe operating systems; this also includes SCSI disks written by Linux for System z. The FBA 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. Note that CKD and xDR managed FB disks are always in the same consistency group - they are always frozen and swapped together.

GDPS requires CKD utility devices in the disk subsystem in order for it to send commands to monitor and control the mirroring of the FBA devices. More specifically, GDPS will need at least one CKD utility device in each hardware cluster of the storage subsystem. A sample configuration of this function called Open LUN Management is shown in Figure 3-2 on page 52.

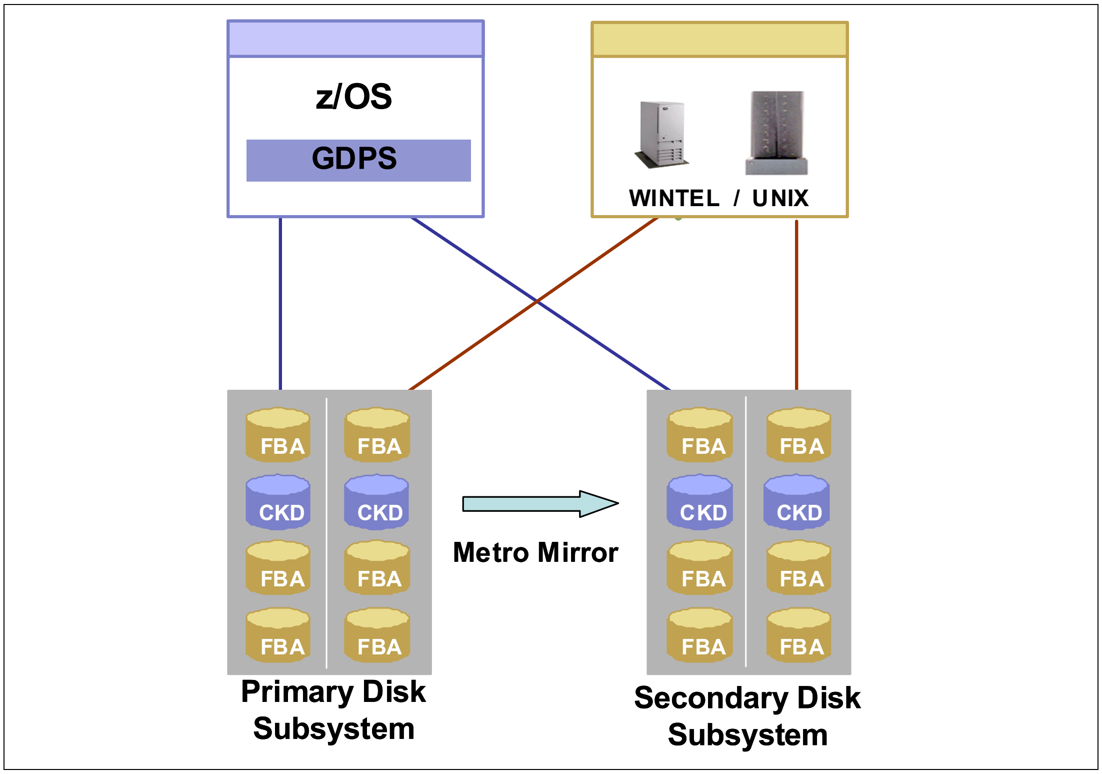


Figure 3-2 GDPS/PPRC Open LUN Management

For a more detailed discussion of Open LUN management refer to 7.1, “Open LUN Management function” on page 146

3.1.4 Protecting other CKD data

GDPS/PPRC can also manage the disk mirroring of CKD format disks used by non-GDPS System z images: z/OS, Linux on System z, VM, and VSE LPARs that are not running GDPS/PPRC or System Automation for Multiplatforms (SA MP). Because GDPS/PPRC is not able to communicate with these images, any errors on those devices will not be recognized by GDPS. Therefore, although you can use GDPS/PPRC to manage the non-GDPS CKD disks, you cannot have the same guarantee of consistency that you do for systems running the GDPS/PPRC code.

In the case of a freeze event on systems running GDPS/PPRC (or SA MP) the disks belonging to the non-GDPS images will also be frozen. If you have a FREEZE and STOP policy, and GDPS/PPRC has access to the HMC for any non-GDPS System z images, GDPS/PPRC has the ability to system reset the associated LPARs.

An additional consideration is in relation to HyperSwap. As one example, because VSE does not support HyperSwap and there is no GDPS code running on the non-GDPS System z images, should an unplanned HyperSwap occur, these systems will not swap to the secondaries. It is possible that, in some circumstances, the non-GDPS System z systems could continue to update the old primaries, resulting in a data integrity exposure. For this reason, if you plan to exploit HyperSwap, and you also want to manage CKD volumes on non-GDPS System z images, you should be aware of this potential exposure. There is no problem doing a HyperSwap as long as you set the consistency group timer value high enough (hours to days) or you manually shut down the non-GDPS System z images prior to a planned HyperSwap.

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, primary disks, and secondary disks. The actual configuration depends on your business and availability requirements. We list the three most common configurations here.

- Active/Standby 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.

- Active/Active configuration

  In this configuration, the production systems run in both sites, Site1 and Site2. This configuration typically exploits the full benefits of data sharing available with a Parallel Sysplex. It is recommended to have two GDPS controlling systems, one in each site.

- 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 could be a customer site or could be owned by a third-party recovery services provider (thus the name BRS).

3.2.1 Controlling system

Why does a GDPS/PPRC configuration need a controlling system? At first, you may 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 may 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 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 will be unaffected by errors that may 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 GDPS-managed mirrored disks.

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; and for initiating a freeze and implementing the freeze policy actions following a freeze event. Refer to 3.1.1, “Protecting data integrity and data availability” on page 44 for more details about Freeze policies and HyperSwap processing.

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.

Improved controlling system availability - enhanced timer support

Enhancements in GDPS V3.6 and z/OS V1.9 (with enabling software updates) help improve GDPS recovery times for events that impact the primary time source for the sysplex, whether the time source is Server Time Protocol (STP) or External Time Reference (ETR) based. These enhancements allow the 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 may have started, instead of being in a disabled WTOR state. 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 IEA015A if the CPC that z/OS is running on is in ETR timing mode (either in an ETR network or in an STP Mixed Coordinated Timing Network (CTN)), and it is IEA394A if the CPC is in STP timing mode (either in an STP Mixed CTN or STP-only CTN).

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 in the event of a failure. Actions may include:

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

Because the controlling system only needs to run 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 TOD clock of the server (referred to as local timing mode), instead of generating a WTOR.

Automated response to ETR or STP sync WTORs

GDPS on the Controlling systems, using the BCP Internal Interface, provides automation to reply to WTORs IEA015A or IEA394A when the Controlling systems are running in local timing mode. Refer to “Improved controlling system availability - enhanced timer support” on page 54. A server in an ETR or STP network may have recovered from an unsynchronized to a synchronized timing state without customer intervention. By automating the response to the WTORs, potential time outs of subsystems and applications in the customer’s enterprise may 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, on the other hand, the CPC is still in an unsynchronized state when GDPS automation responded with RETRY to the WTOR. 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 Active/standby configuration

A GDPS/PPRC active/standby workload environment typically consists of a multi-site sysplex, with all production systems normally running in 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 multi-site sysplex can be a base sysplex or a Parallel Sysplex; a Coupling Facility is not strictly required. The multi-site 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 3-3 shows a typical GDPS/PPRC active/standby 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 just 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.

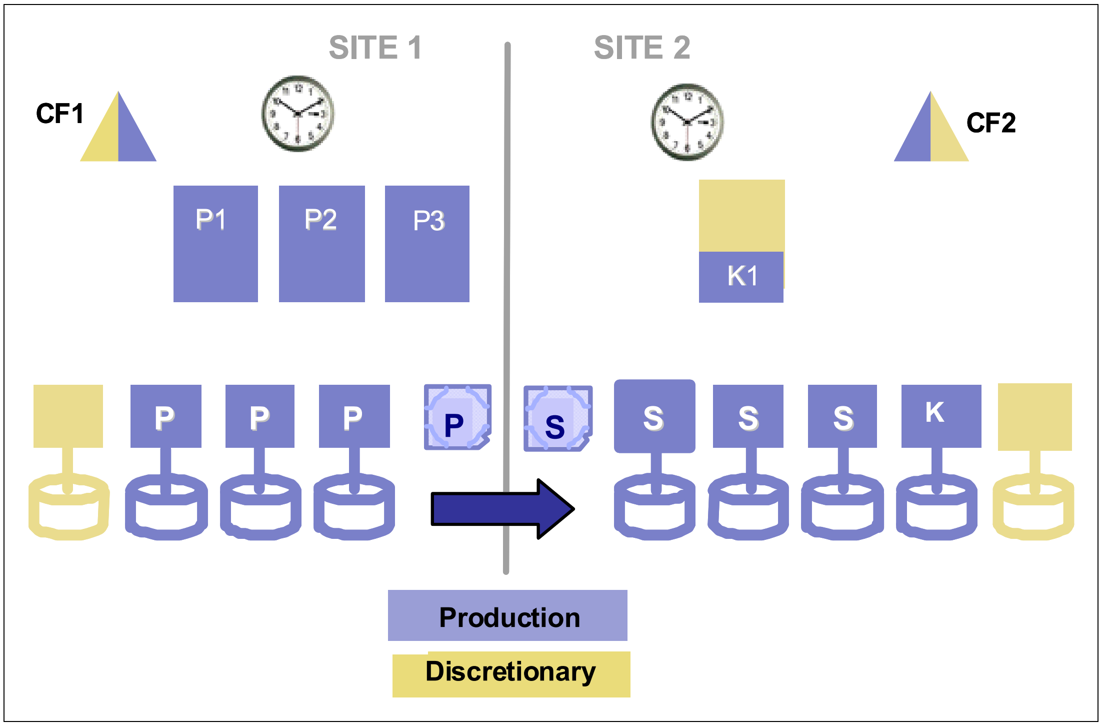


Figure 3-3 GDPS/PPRC active/standby workload configuration

3.2.3 Active/active configuration

An active/active workload configuration, shown in Figure 3-4 on page 56, differs from an active/standby workload in that production systems are running in *both* sites. Also, although it is possible to run an active/standby workload as a base sysplex, it would be unusual to see an active/active workload using a base sysplex (that is, without coupling facilities). This is because an active/active 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 should be taken.

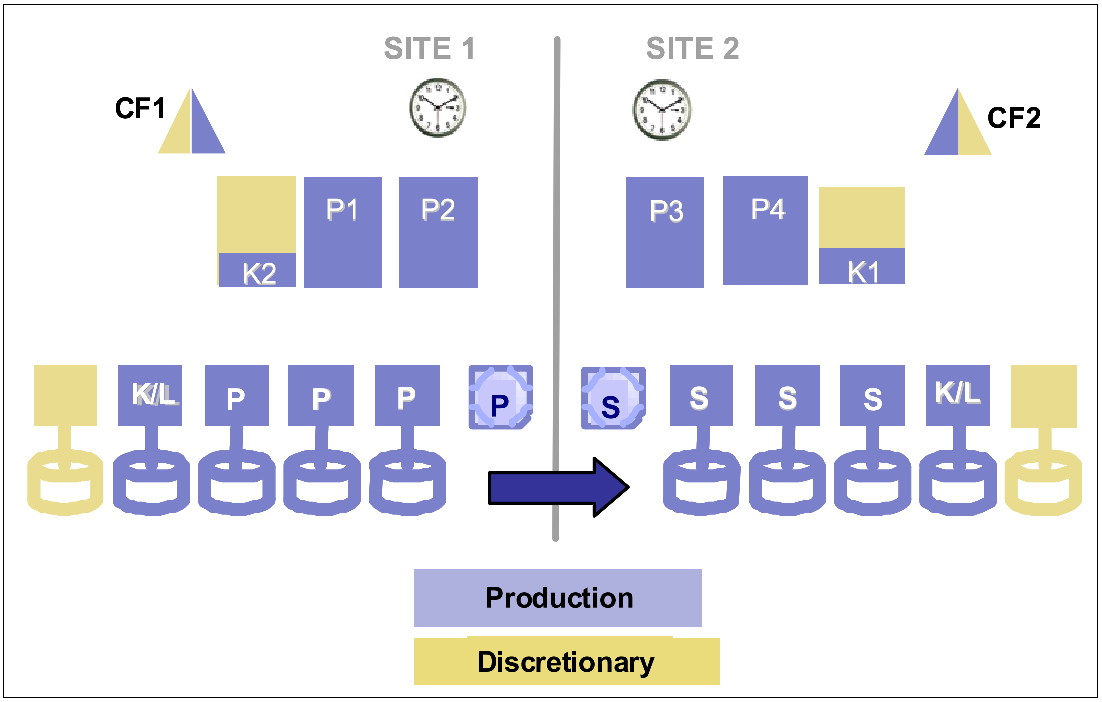


Figure 3-4 GDPS/PPRC active/active workload configuration

3.2.4 Business Recovery Services (BRS) configuration

A third configuration is what is known as the "BRS configuration", illustrated in Figure 3-5 on page 57. 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 could be a customer site or could 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 may also contain processors that would 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 multi-site sysplex, but is within the maximum distance supported by FICON and Metro Mirror.

Even though there is no need for a multi-site 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. Refer to 2.9.8, "Connectivity options" on page 39, and *IBM System z Connectivity Handbook,* SG24-5444, for options available to extend the distance of FICON links between sites.

In the BRS configuration, the Controlling system (K1) should have its disk devices in Site2. This permits the K1 system to be restarted manually in Site2 after a disaster has been declared. The K1 system will then be used to recover the secondary disk subsystems to a simplex state when necessary and then, using a GDPS control script, 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 K1 system 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 in Site1 (K2 in Figure 3-5) on local disks in Site1 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 due to a cross-site fiber loss. GDPS will attempt to maintain the current Master system in the Controlling system by using the secondary disks.

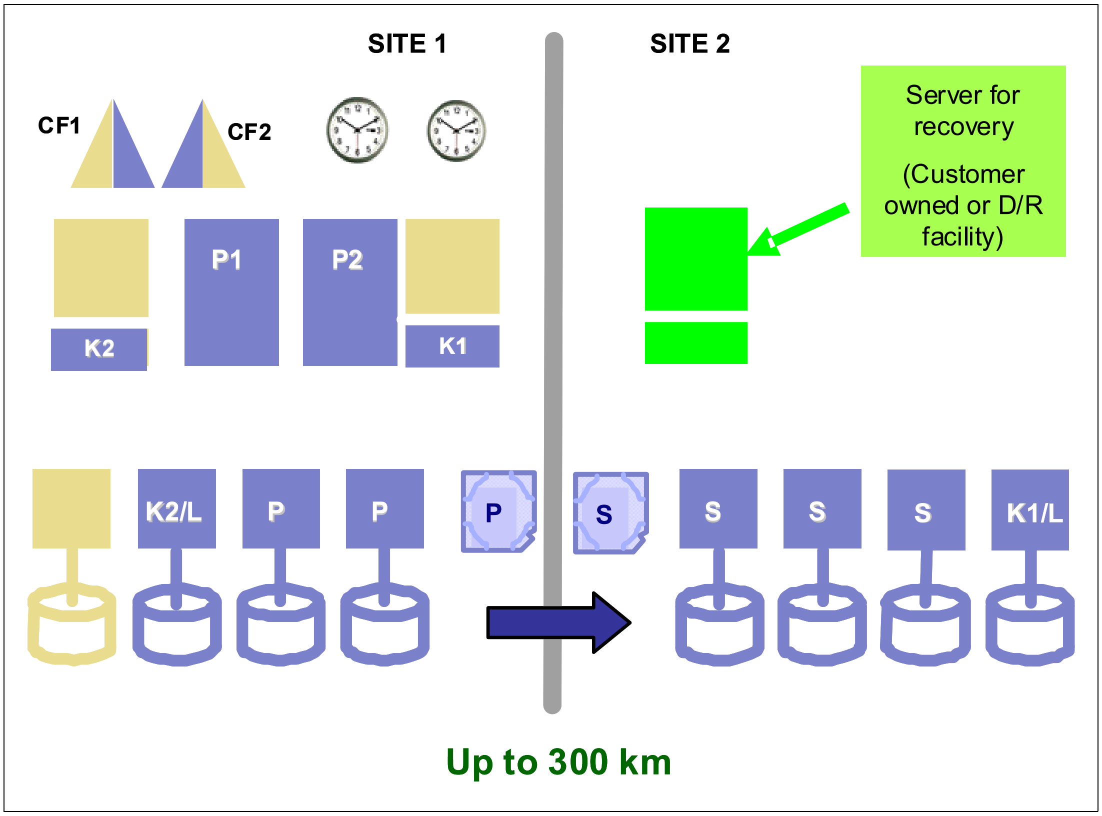


Figure 3-5 GDPS/PPRC BRS configuration

3.2.5 GDPS/PPRC in a three-site configuration

GDPS/PPRC can be combined with GDPS/XRC or GDPS/GM in a three-site configuration. In this configuration, GDPS/PPRC (when combined with Parallel Sysplex exploitation and HyperSwap) 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.

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 some additional automation capabilities.

Refer to Chapter 8, “Combining Local/Metro continuous availability with out-of-region disaster recovery” on page 173 for a more detailed discussion of 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 might implement GDPS/PPRC HyperSwap Manager (GDPS/PPRC 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 discussed in Chapter 4, “GDPS/PPRC HyperSwap Manager” on page 79.

3.2.7 Other 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.

Equally important is that certain functions (stopping and restarting systems and changing the Couple Data Set configuration) are carried out 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 (FBA) data” on page 51, it is possible for GDPS/PPRC to manage FBA disks on behalf of distributed systems either in the same session as System z CKD disks or in a separate session. However, for these distributed systems, although GDPS/PPRC manages the remote copy and recovery of the disks, it is not able to perform any system recovery actions in the recovery site.

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

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

3.3.1 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 either be implemented as guests running under z/VM, or native Linux on System z systems. There are several examples of an application server running on Linux on System z and a database server running on z/OS, such as:

- 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 multi-tiered architecture there is a need to provide a coordinated near Continuous Availability/Disaster Recovery solution for both z/OS and Linux on System z. The GDPS/PPRC function that provides this capability is called Multiplatform Resiliency for System z, and it can be implemented as long as the disks being used by z/VM and Linux are CKD disks. For more details about this function, refer to 7.2, “GDPS/PPRC Multiplatform Resiliency for System z” on page 147.

3.3.2 Distributed Cluster Management

Distributed Cluster Management (DCM) is a new GDPS capability which allows the management and coordination of disaster recovery across distributed servers that may be clustered using clustering solutions, and the System z 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 will provide advisory and coordination functions between GDPS and one or more VCS clusters.

For a more detailed discussion of the DCM function, refer to 7.3, “Distributed Cluster Management” on page 153.

3.4 Managing the GDPS environment

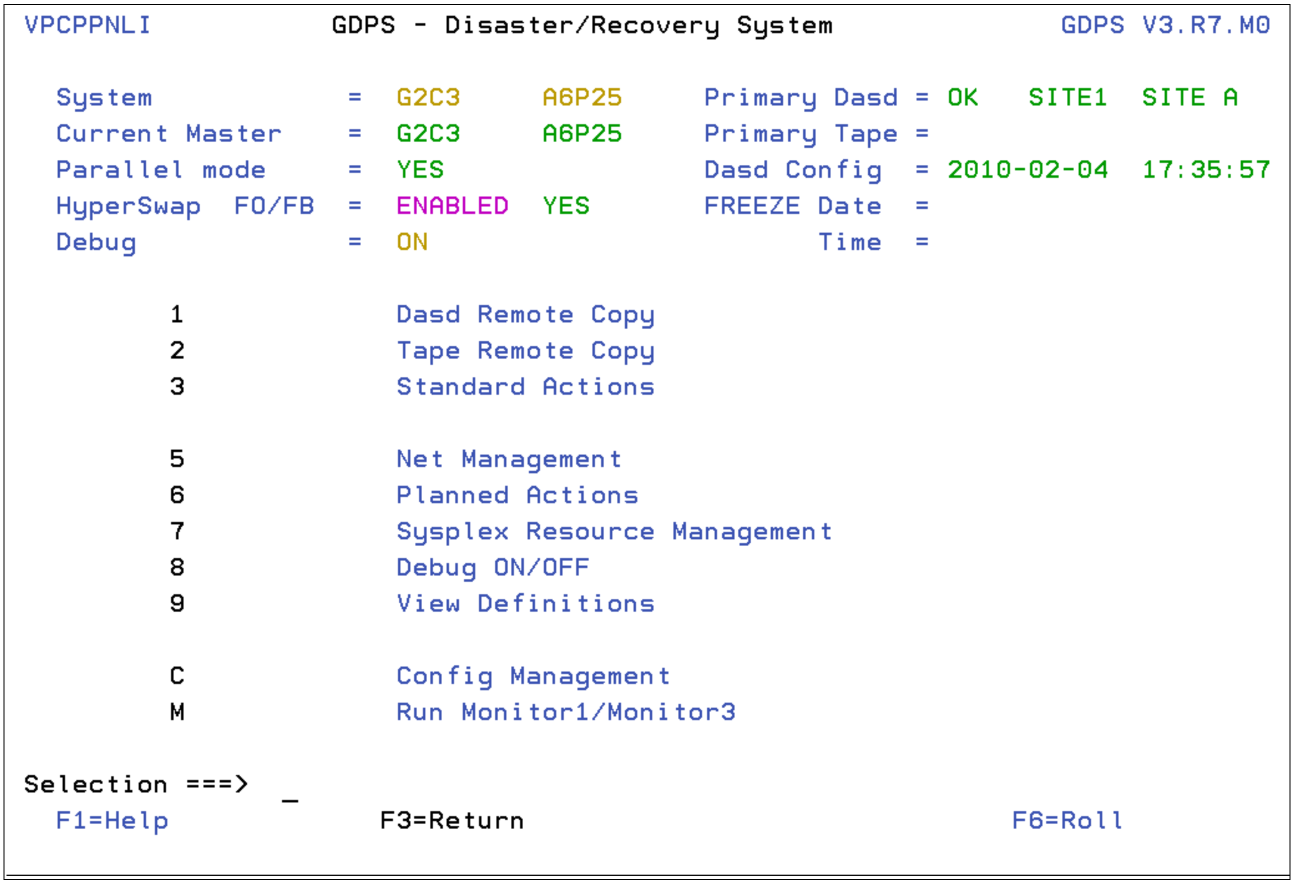
We have seen how GDPS/PPRC can protect just about any type of data that can reside in a disk subsystem. Further, it can provide data consistency across all platforms that are sharing the disk subsystem. However, as discussed in Chapter 1, “Introduction to Business Resilience and the role of GDPS” on 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/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.

3.4.1 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 “web interface” in this document).

An example of the main GDPS/PPRC 3270-based panel is shown in Figure 3-6 on page 60.



```
VPCPPNLI           GDPS - Disaster/Recovery System             GDPS V3.R7.M0

  System            =  G2C3     A6P25    Primary Dasd = OK   SITE1   SITE A
  Current Master    =  G2C3     A6P25    Primary Tape =
  Parallel mode     =  YES               Dasd Config  = 2010-02-04  17:35:57
  HyperSwap  FO/FB  =  ENABLED  YES      FREEZE Date  =
  Debug             =  ON                       Time  =

         1             Dasd Remote Copy
         2             Tape Remote Copy
         3             Standard Actions

         5             Net Management
         6             Planned Actions
         7             Sysplex Resource Management
         8             Debug ON/OFF
         9             View Definitions

         C             Config Management
         M             Run Monitor1/Monitor3

Selection ===>
                                                                              _
  F1=Help             F3=Return                                  F6=Roll
```

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

This panel is relatively simple to use, with a summary of configuration status at the top of the panel and a menu of choices that a user can select from below. As an example, a user would simply type a `1` at the `Selection ===>` prompt and press Enter to view the disk mirroring ("Dasd Remote Copy") panels.

GDPS web interface

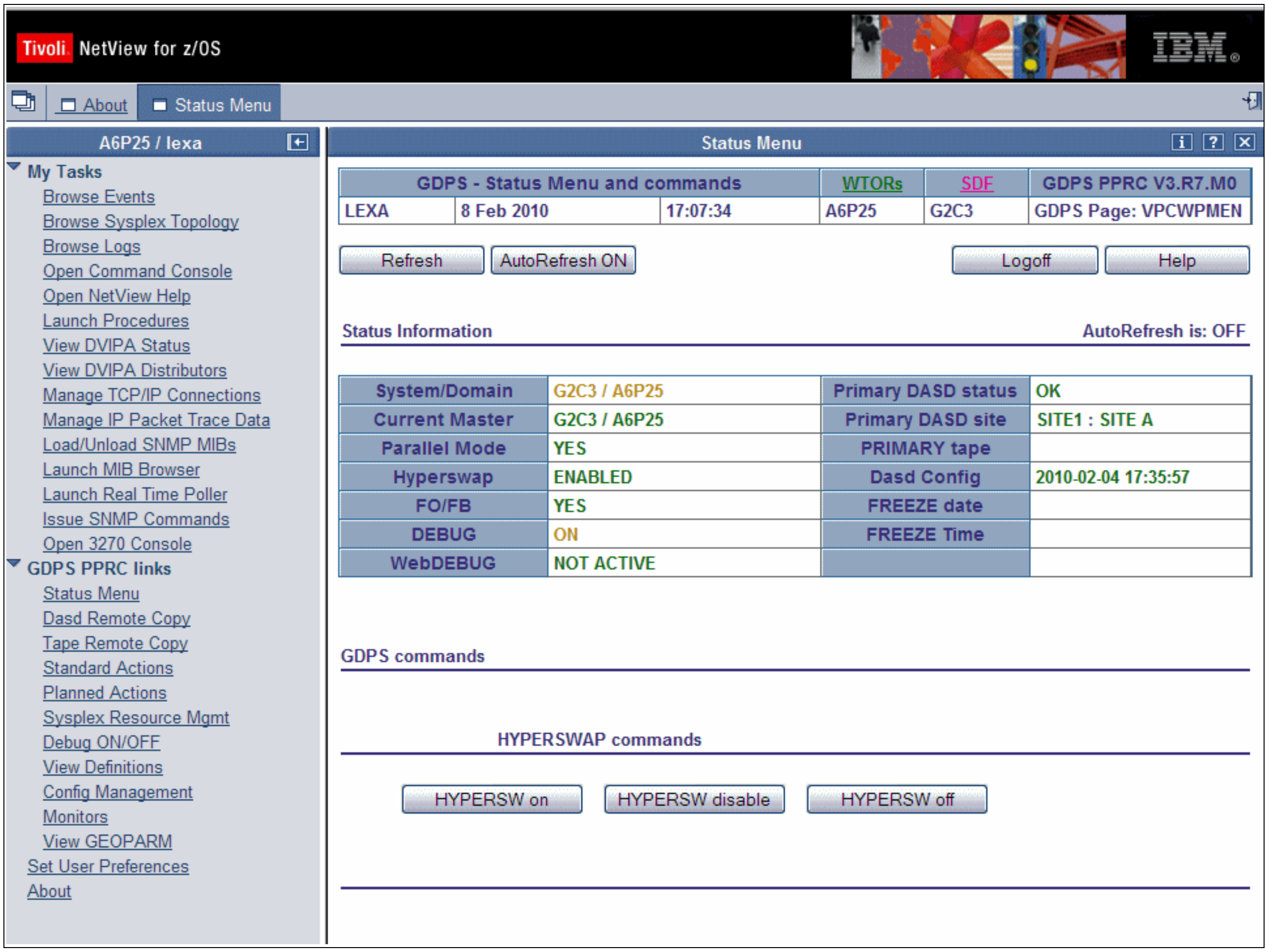
The web interface is a browser-based interface designed to improve operator productivity. The web interface provides the same functional capability as the 3270-based panel, such as providing management capabilities for Remote Copy Management, Standard Actions, Sysplex Resource Management, and SDF Monitoring using simple point-and-click procedures. In addition, users can open multiple windows to allow for continuous status monitoring, while performing other GDPS/PPRC management functions.

The web interface display is split into 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/PPRC web interface is shown in Figure 3-7 on page 61. The left frame, shown below `GDPS PPRC links`, allows you to select the menu options. These options can be displayed at all times, or you can optionally collapse the frame.

Note: For the remainder of section 3.4, “Managing the GDPS environment” on page 59, only the web interface is shown to illustrate the various GDPS management functions. The equivalent traditional 3270 panels are not shown here.



GDPS - Status Menu and commands			WTORs	SDF	GDPS PPRC V3.R7.M0
LEXA	8 Feb 2010	17:07:34	A6P25	G2C3	GDPS Page: VPCWPMEN

Status Information AutoRefresh is: OFF

System/Domain	G2C3 / A6P25	Primary DASD status	OK
Current Master	G2C3 / A6P25	Primary DASD site	SITE1 : SITE A
Parallel Mode	YES	PRIMARY tape	
Hyperswap	ENABLED	Dasd Config	2010-02-04 17:35:57
FO/FB	YES	FREEZE date	
DEBUG	ON	FREEZE Time	
WebDEBUG	NOT ACTIVE		

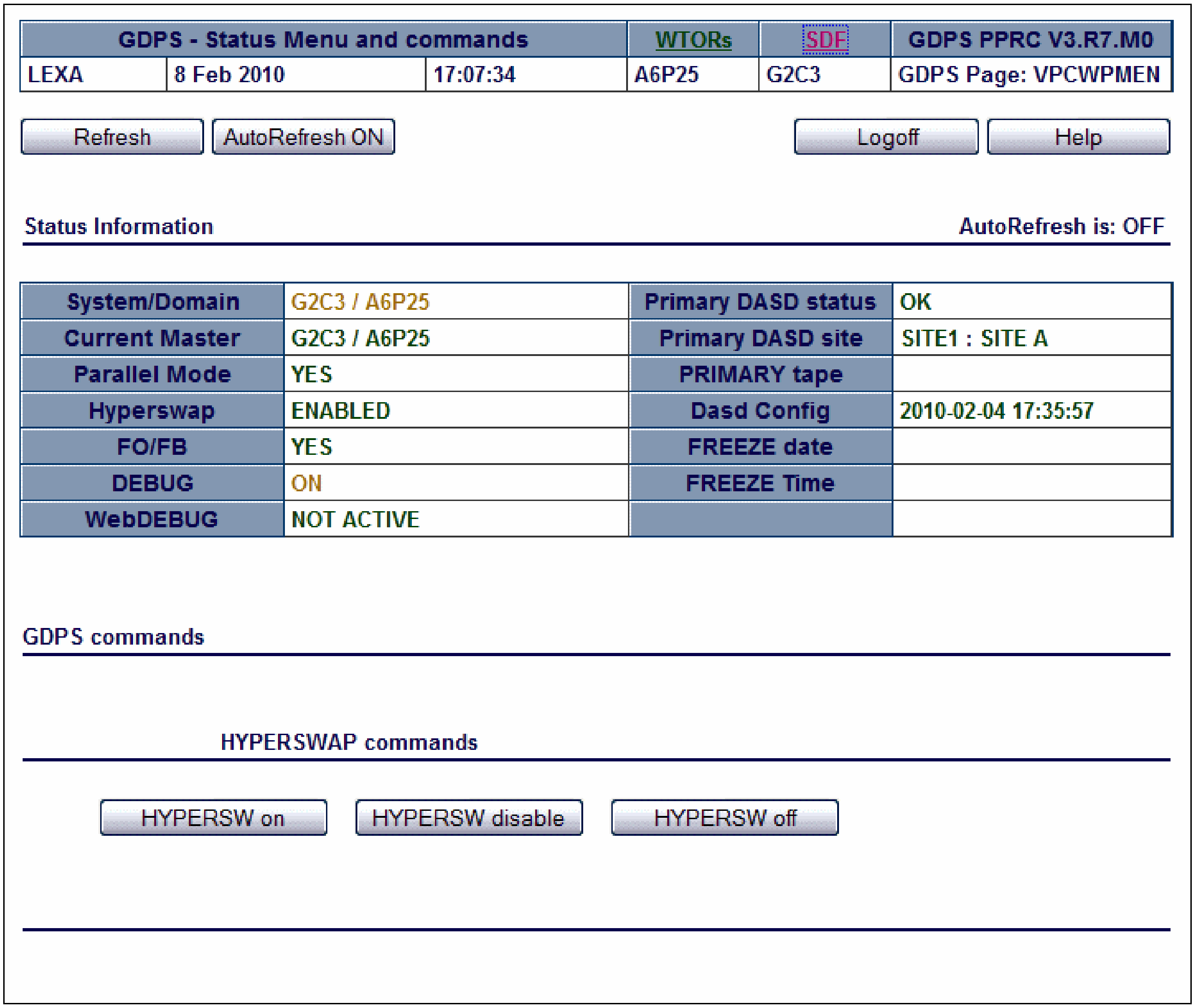
GDPS commands

HYPERSWAP commands

Figure 3-7 Full view of GDPS main panel with task bar and status information

Main Status panel

The GDPS web interface status frame shown in Figure 3-8 is the equivalent to the main GDPS panel. The information on this frame is what is found on the top portion of the 3270 GDPS Main panel.



GDPS - Status Menu and commands			WTORs	SDF	GDPS PPRC V3.R7.M0
LEXA	8 Feb 2010	17:07:34	A6P25	G2C3	GDPS Page: VPCWPMEN

Status Information

AutoRefresh is: OFF

System/Domain	G2C3 / A6P25	Primary DASD status	OK
Current Master	G2C3 / A6P25	Primary DASD site	SITE1 : SITE A
Parallel Mode	YES	PRIMARY tape	
Hyperswap	ENABLED	Dasd Config	2010-02-04 17:35:57
FO/FB	YES	FREEZE date	
DEBUG	ON	FREEZE Time	
WebDEBUG	NOT ACTIVE		

GDPS commands

HYPERSWAP commands

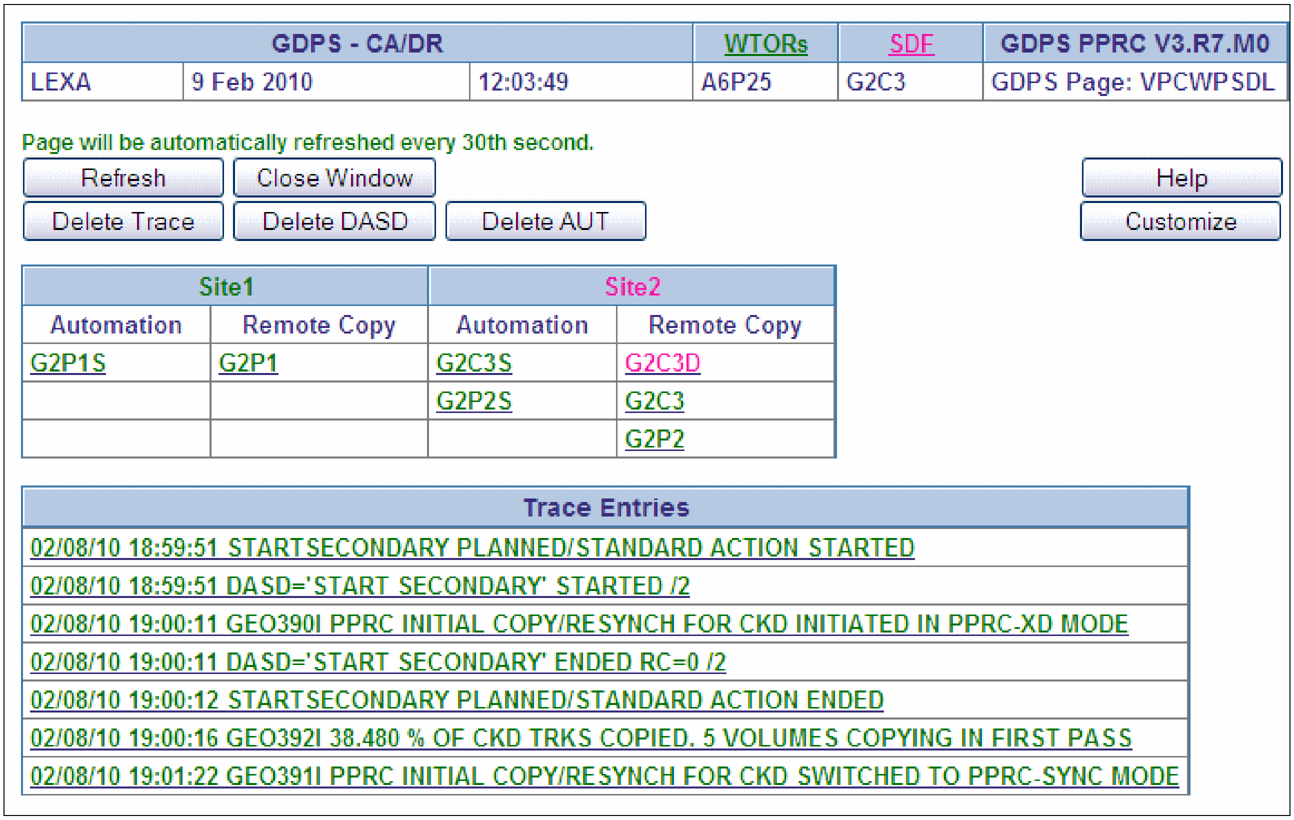
Figure 3-8 GDPS web interface: Main status frame

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. It is the only dynamically updated status display available for GDPS.

SDF provides a dynamically-updated color-coded panel, as shown in Figure 3-9 on page 63. 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 should have an SDF panel within view so they will immediately become aware of anything requiring intervention or action.

The web interface can be set up to automatically refresh every 30 seconds. 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.



GDPS - CA/DR			WTORs	SDF	GDPS PPRC V3.R7.M0
LEXA	9 Feb 2010	12:03:49	A6P25	G2C3	GDPS Page: VPCWPSDL

Page will be automatically refreshed every 30th second.

Site1		Site2	
Automation	Remote Copy	Automation	Remote Copy
G2P1S	G2P1	G2C3S	G2C3D
		G2P2S	G2C3
			G2P2

Trace Entries
02/08/10 18:59:51 STARTSECONDARY PLANNED/STANDARD ACTION STARTED
02/08/10 18:59:51 DASD='START SECONDARY' STARTED /2
02/08/10 19:00:11 GEO390I PPRC INITIAL COPY/RESYNCH FOR CKD INITIATED IN PPRC-XD MODE
02/08/10 19:00:11 DASD='START SECONDARY' ENDED RC=0 /2
02/08/10 19:00:12 STARTSECONDARY PLANNED/STANDARD ACTION ENDED
02/08/10 19:00:16 GEO392I 38.480 % OF CKD TRKS COPIED. 5 VOLUMES COPYING IN FIRST PASS
02/08/10 19:01:22 GEO391I PPRC INITIAL COPY/RESYNCH FOR CKD SWITCHED TO PPRC-SYNC MODE

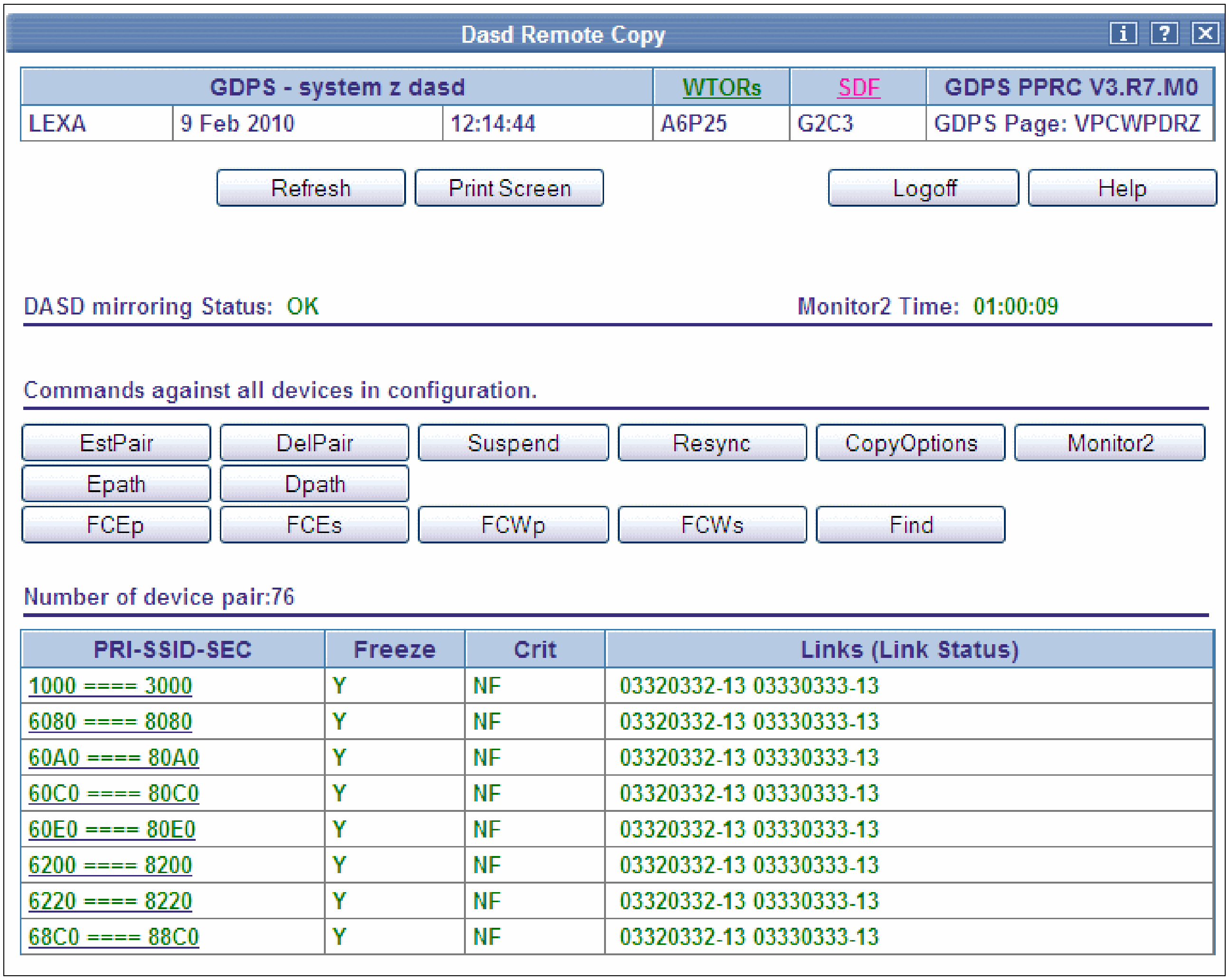
Figure 3-9 NetView SDF web interface

Remote copy panels

The z/OS Advanced Copy Services capabilities are very powerful, but the native z/OS TSO and ICKDSF interfaces are not very user friendly. To make it easier for operators to check and manage the remote copy environment, you can (and should) use the GDPS-provided DASD Remote Copy panels.

In order 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 desired one. You can start, stop, suspend, and resynchronize mirroring at the volume or LSS level. You can initiate a FlashCopy and you can reverse the direction of mirroring. These actions can be carried out at the device or LSS level, or both, as appropriate. Figure 3-10 on page 64 shows the mirroring panel for CKD devices.


Dasd Remote Copy

GDPS - system z dasd			WTORs	SDF	GDPS PPRC V3.R7.M0
LEXA	9 Feb 2010	12:14:44	A6P25	G2C3	GDPS Page: VPCWPDRZ

DASD mirroring Status: OK Monitor2 Time: 01:00:09

Commands against all devices in configuration.

Number of device pair:76

PRI-SSID-SEC	Freeze	Crit	Links (Link Status)
1000 ==== 3000	Y	NF	03320332-13 03330333-13
6080 ==== 8080	Y	NF	03320332-13 03330333-13
60A0 ==== 80A0	Y	NF	03320332-13 03330333-13
60C0 ==== 80C0	Y	NF	03320332-13 03330333-13
60E0 ==== 80E0	Y	NF	03320332-13 03330333-13
6200 ==== 8200	Y	NF	03320332-13 03330333-13
6220 ==== 8220	Y	NF	03320332-13 03330333-13
68C0 ==== 88C0	Y	NF	03320332-13 03330333-13

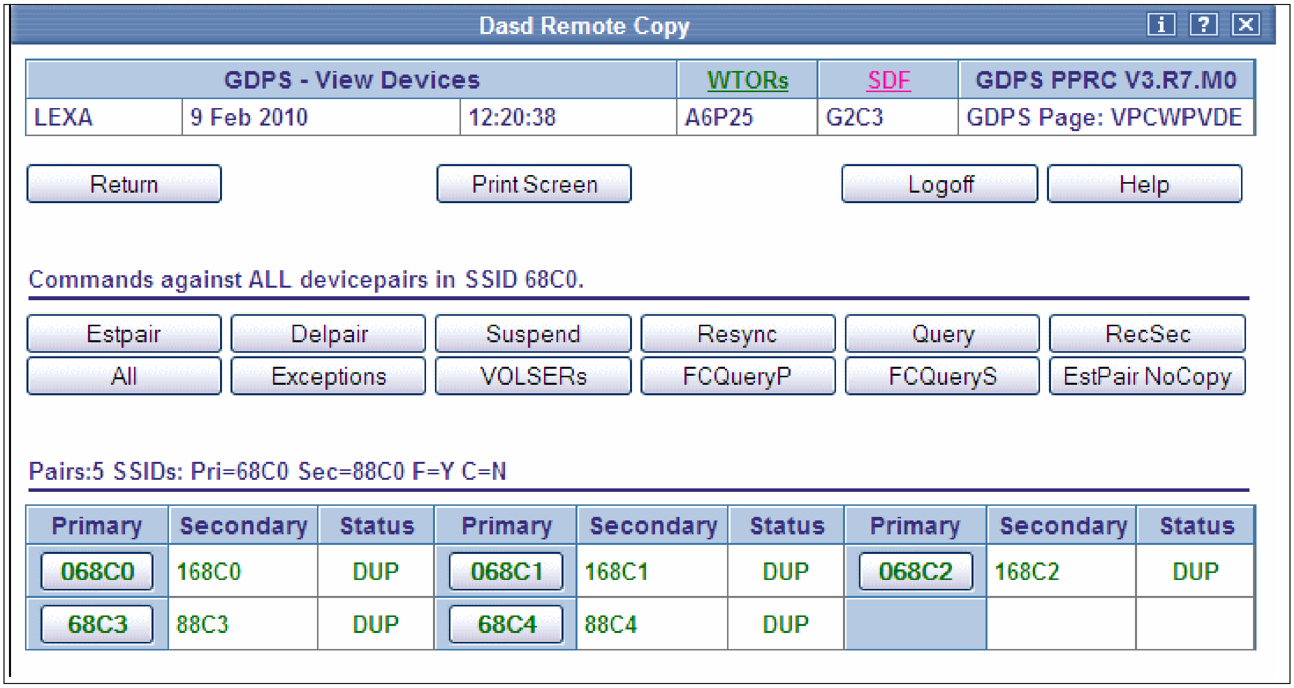
Figure 3-10 DASD Remote Copy SSID web interface

The Dasd Remote Copy frame is organized into three sections:

- A top section displays the mirroring status, and buttons for Return, Refresh and Help.
- A middle section displays actions against all the panel-displayed SSID-pairs (similar to the bottom section of the equivalent 3270 panel).
- A bottom section displays the list of all the SSID-pairs.

To perform an action on a single SSID-pair, click the actual pair. This brings you to a panel 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 has been selected, the frame shown in Figure 3-11 on page 65 is displayed. The bottom of 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 (indicated as DUP on the panel). Also note that the secondary devices for some of these pairs are in an alternate subchannel set (MSS1 in this case). Additional details can be viewed for each pair by clicking the button for the primary device in the pair.



GDPS - View Devices			WTORs	SDF	GDPS PPRC V3.R7.M0
LEXA	9 Feb 2010	12:20:38	A6P25	G2C3	GDPS Page: VPCWPVDE

Commands against ALL devicepairs in SSID 68C0.

Pairs:5 SSIDs: Pri=68C0 Sec=88C0 F=Y C=N

Primary	Secondary	Status	Primary	Secondary	Status	Primary	Secondary	Status
068C0	168C0	DUP	068C1	168C1	DUP	068C2	168C2	DUP
68C3	88C3	DUP	68C4	88C4	DUP			

Figure 3-11 Web Interface Dasd Remote Copy "View Devices" detail frame

If you are familiar with using the TSO or ICKDSF interfaces you will appreciate how much more user friendly these panels are.

Remember that these GDPS-provided panels are *not* intended to be a remote copy monitoring tool. Because of the overhead involved in gathering the information to populate the NetView panels, GDPS only gathers this information 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.

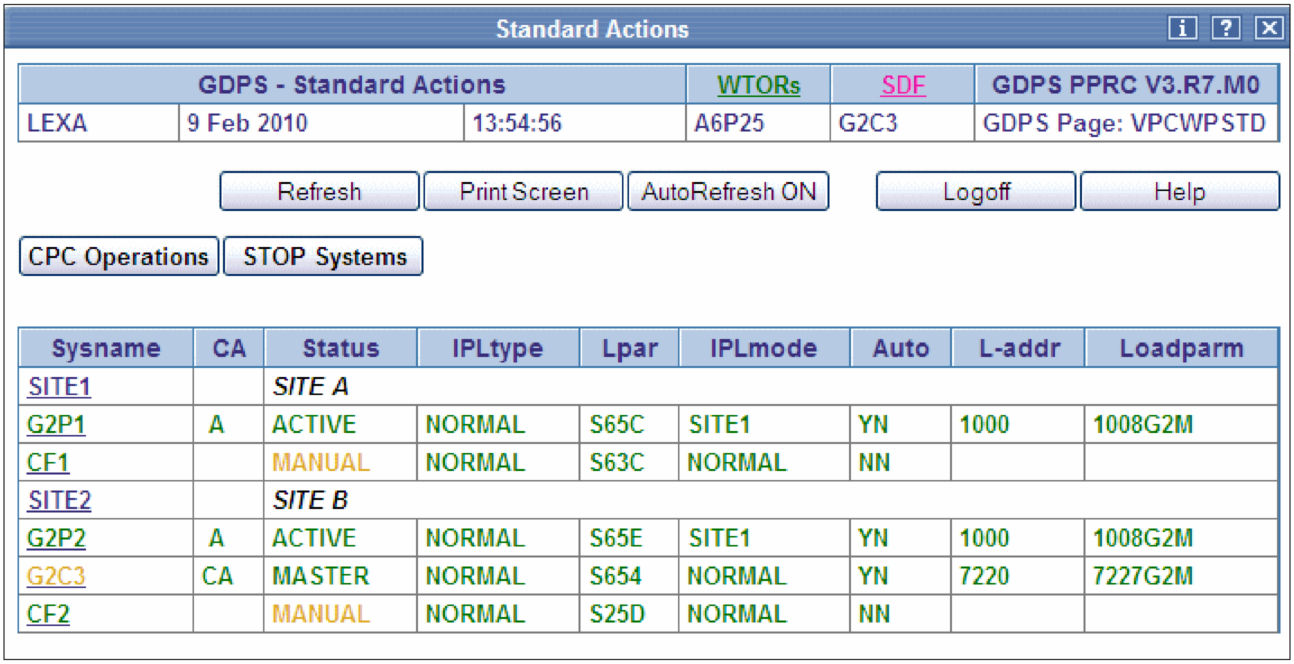
Standard Actions

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 recommended 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 should not 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 really single steps, or are intended to impact just one resource. Examples are starting a system IPL, updating the IPL address or the load parameters to be used the next time a system IPL is started, or activating an LPAR. So, if you wanted to stop a system, change its IPL address, then perform an IPL, you would initiate three separate *Standard Actions*.

The GDPS/PPRC Standard Actions web interface panel is shown in Figure 3-12 on page 66. It displays all the LPARs being managed by GDPS/PPRC and for each one it shows the current status and various IPL information. To perform actions on each system, the web Interface requires you to click the specific system where the action is to be performed. Once the specific LPAR has been selected, the appropriate action can be selected, including Load, Stop, Reset, ReIPL, Activate LPAR, and Deactivate LPAR.



GDPS - Standard Actions			WTORs	SDF	GDPS PPRC V3.R7.M0
LEXA	9 Feb 2010	13:54:56	A6P25	G2C3	GDPS Page: VPCWPSTD

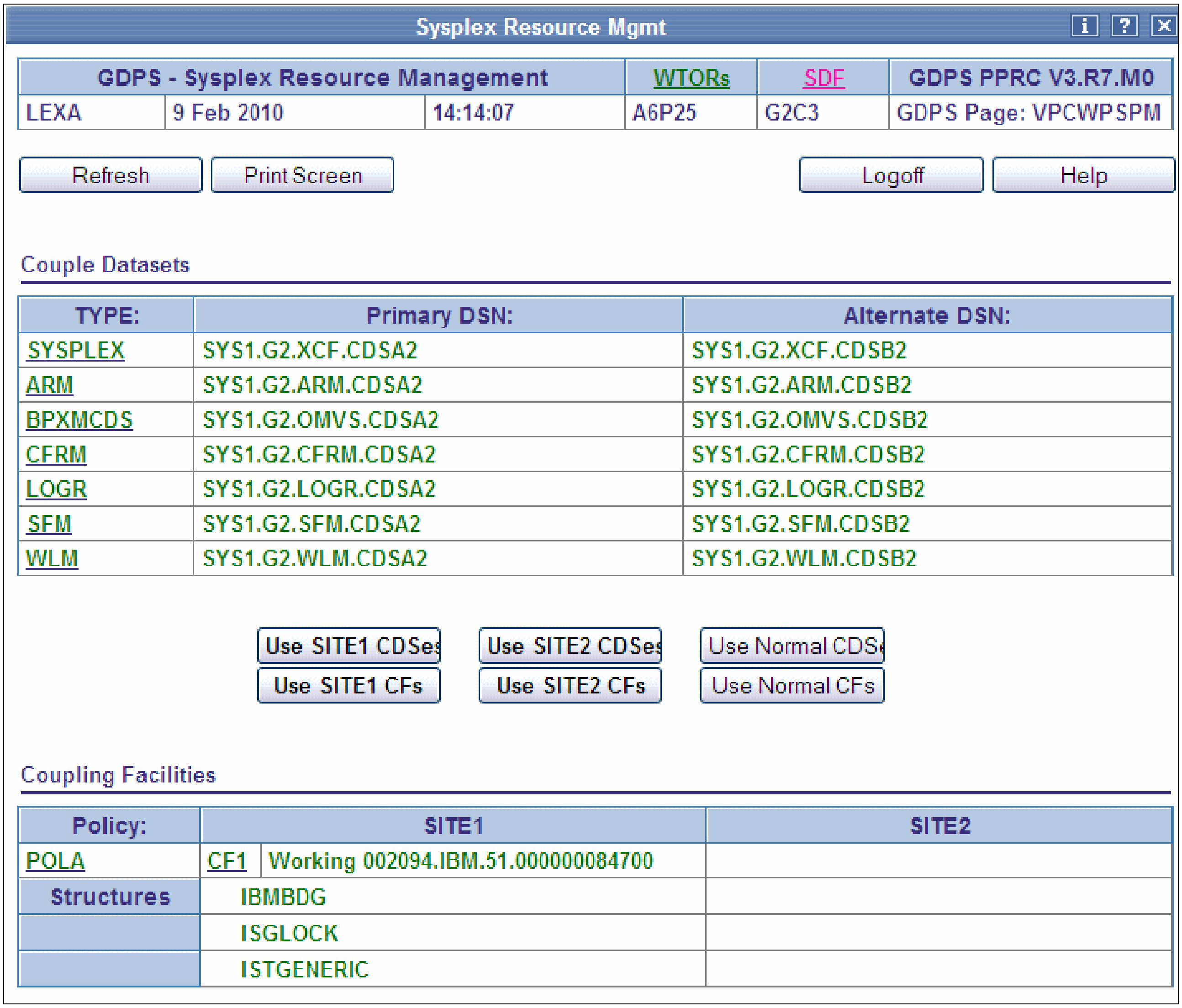
Sysname	CA	Status	IPLtype	Lpar	IPLmode	Auto	L-addr	Loadparm
SITE1		*SITE A*						
G2P1	A	ACTIVE	NORMAL	S65C	SITE1	YN	1000	1008G2M
CF1		MANUAL	NORMAL	S63C	NORMAL	NN		
SITE2		*SITE B*						
G2P2	A	ACTIVE	NORMAL	S65E	SITE1	YN	1000	1008G2M
G2C3	CA	MASTER	NORMAL	S654	NORMAL	YN	7220	7227G2M
CF2		MANUAL	NORMAL	S25D	NORMAL	NN		

Figure 3-12 GDPS/PPRC Standard Actions panel

Sysplex resource management

There are certain resources that are vital to the health and availability of the sysplex. In a multi-site 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 web interface, shown in Figure 3-13 on page 67, provides you with the ability to manage the resources, with knowledge about where the resources exist. For example, normally you would 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 - Sysplex Resource Management			WTORs	SDF	GDPS PPRC V3.R7.M0
LEXA	9 Feb 2010	14:14:07	A6P25	G2C3	GDPS Page: VPCWPSPM

Couple Datasets

TYPE:	Primary DSN:	Alternate DSN:
SYSPLEX	SYS1.G2.XCF.CDSA2	SYS1.G2.XCF.CDSB2
ARM	SYS1.G2.ARM.CDSA2	SYS1.G2.ARM.CDSB2
BPXMCDS	SYS1.G2.OMVS.CDSA2	SYS1.G2.OMVS.CDSB2
CFRM	SYS1.G2.CFRM.CDSA2	SYS1.G2.CFRM.CDSB2
LOGR	SYS1.G2.LOGR.CDSA2	SYS1.G2.LOGR.CDSB2
SFM	SYS1.G2.SFM.CDSA2	SYS1.G2.SFM.CDSB2
WLM	SYS1.G2.WLM.CDSA2	SYS1.G2.WLM.CDSB2

Coupling Facilities

Policy:	SITE1		SITE2
POLA	CF1	Working 002094.IBM.51.000000084700	
Structures	IBMBDG		
	ISGLOCK		
	ISTGENERIC		

Figure 3-13 GDPS/PPRC Sysplex Resource Management web interface

3.4.2 GDPS scripts

We have now seen how the GDPS panels provide powerful functions to help you manage your data and systems. However, the GDPS panels are only one way of accessing this capability. All of the functions that can be initiated via the panels are also accessible from GDPS scripts. A script is simply a procedure recognized by GDPS that pulls together into a list one or more GDPS functions. 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.

There are several standard scripts included as part of a GDPS implementation, but scripts can easily be customized or written by you to automate the handling of certain situations – both to handle planned changes and unplanned situations. This is an extremely important aspect of GDPS. Scripts are very powerful because they can access the full capability of GDPS. The ability to invoke all the GDPS functions through a script provides:

- 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 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 System 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

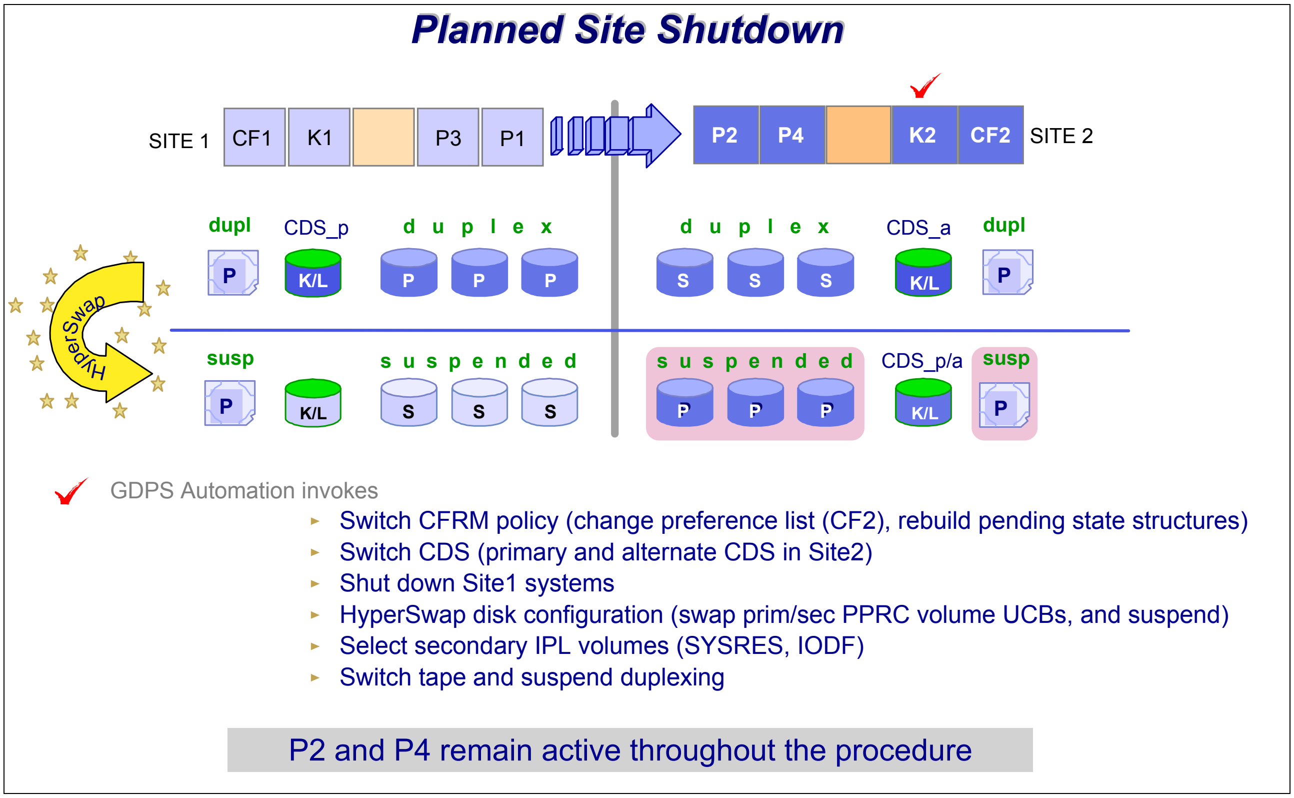
  Because they behave in a consistent manner, you can test your procedures over and over until you are sure they do everything that you want, in exactly the manner that you want. Also, because you need to code everything and cannot assume a level of knowledge (as you might with instructions intended for 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 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 60). As one example, you could have a script that would stop an LPAR and then re-IPL it in an alternate location – all in a very short script, 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'
```



Planned Site Shutdown

GDPS Automation invokes

- Switch CFRM policy (change preference list (CF2), rebuild pending state structures)
- Switch CDS (primary and alternate CDS in Site2)
- Shut down Site1 systems
- HyperSwap disk configuration (swap prim/sec PPRC volume UCBs, and suspend)
- Select secondary IPL volumes (SYSRES, IODF)
- Switch tape and suspend duplexing

P2 and P4 remain active throughout the procedure

Figure 3-14 GDPS/PPRC Planned Action

A more complex example of a Planned Action is shown in Figure 3-14. In this example, a single action in GDPS executing a planned script of only six lines, results in a complete planned site switch. Specifically, the following actions are carried out by GDPS:

- The sysplex resources (CDS and CF) are switched to only use those in Site2.
- The systems in Site1 are stopped (note that P2 and P4 remain active in this example).
- A HyperSwap is executed to use the disk in Site2.
- The IPL parameters (IPL address and load parameters) are updated to reflect the new configuration.
- The systems are restarted 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 carried out the same way every time with no vital steps accidentally overlooked.

STP CTN role reassignments - planned operations

A new GDPS script statement has been introduced with GDPS V3.8 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 - PTS), which CPC is able to take over as the clock source for planned and unplanned events (Backup Time Server - BTS), which CPC is the active clock source (Current Time Server - CTS), and which CPC assists in STP recovery (Arbiter).

It is strongly recommended that the server roles be reassigned prior to performing planned disruptive actions on any of these special role servers. Examples of planned disruptive actions are Power on Reset (POR), Activate/Deactivate. The new script statement can now 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:

– 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 may 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

– Script statement to Activate the CPC
– Reassign the STP server roles to their normal operational state
– Statements you may already have in existing scripts to perform IPLs etc.

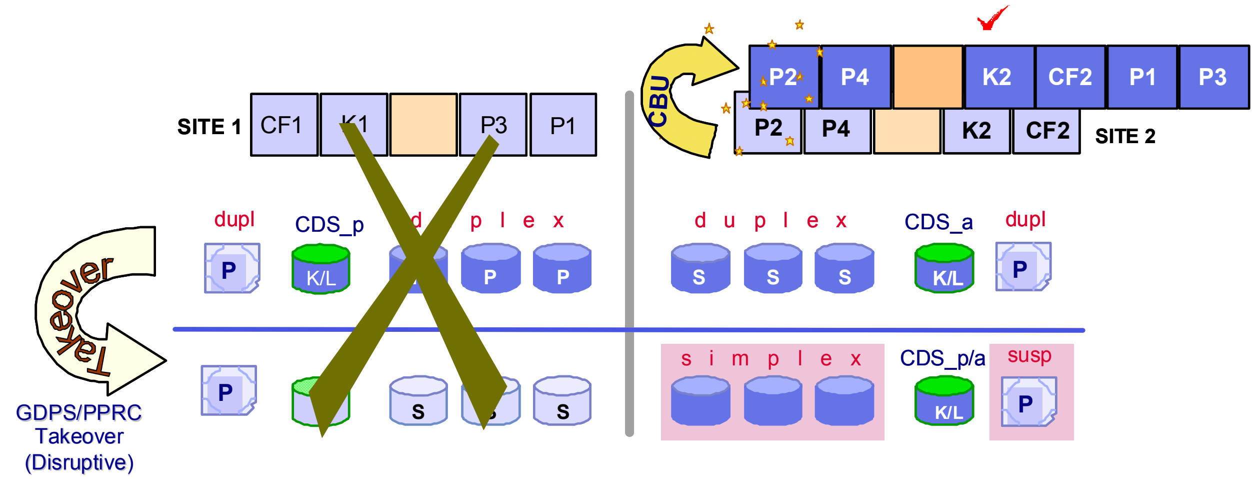
Recovery scripts

There are 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 to protect the integrity of the secondary data.

After the freeze and the action indicated in the freeze policy (STOP or GO) has completed, GDPS will present the operator with a prompt listing the scripts that can be executed at this time. The operator simply has to select the appropriate action and GDPS does the rest.

An example is contained in Figure 3-15 on page 71. In this example, the operator has selected to bring everything up in Site2. GDPS will adjust the CDS and policies to only use those in Site2, it will update all the IPL information to point to what were the secondary volumes, stop any 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 of this is carried out with just a single operator instruction.

Site1 Failure - Freeze trigger



✓ GDPS Automation invokes

- Freeze secondary disk configuration and recover
- Switch Peer-to-Peer VTS configuration and suspend duplexing
- Switch CFRM policy; switch CDS configuration
- Reset Site1 and Site2 systems (except K2)
- Select secondary IPL volumes (SYSRES, IODF)
- Stop expendable systems and/or perform Capacity Backup (CBU)
- Restart production systems in Site2

Figure 3-15 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 as well, again using a GDPS script. The actions in this case 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 will need to be resynchronized before the move. 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 do not contain all the data in the primary disks, and therefore the data is inconsistent. It is like you backed up a disk onto two tapes, and lost the second tape.

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 take a FlashCopy of those volumes 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.

GDPS not only monitors data-related events: scripts can also be provided to drive recovery from a system failure. As seen in Figure 3-16 on page 72, 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 automatically re-IPLs the system in place, brings it back into the Parallel Sysplex, and restarts the workload.

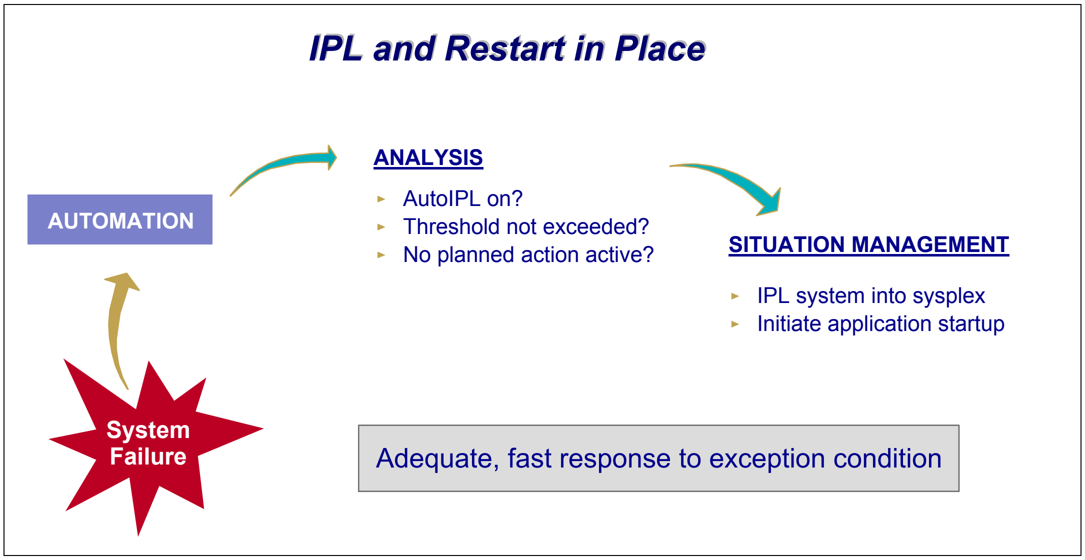


Figure 3-16 Recovering a failed image

Similarly, if a complete processor fails, GDPS will provide recovery support for all the GDPS-managed LPARs on that box.

STP CTN role reassignments - unplanned failure

If a failure condition has resulted in either the PTS, BTS or Arbiter to be no longer an operational synchronized CPC in the CTN, it is recommended 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 new script statement capability described in “STP CTN role reassignments - planned operations” on page 69 can be used to integrate the STP role reassignment as part of an existing script such as a Site takeover script and eliminates 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 54, 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 if the CPC is in STP timing mode (either in an STP Mixed CTN or STP-only CTN).

A new script statement is added to GDPS/PPRC V3.8 which allows GDPS scripts to reply (either ABORT or RETRY) to the IEA394A sync WTOR for STP on systems that are spinning due to loss of synchronization with their Current Time Source. As described in “Automated response to ETR or STP sync WTORs” on page 54, autonomic function exists to reply RETRY automatically for 60 minutes on any GDPS systems that have posted this WTOR.

The new script statement complements and extends this function.

- It provides the means to reply to the message after the 60 minutes automatic reply window is up.

- It can reply to the WTOR on non-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 wish to restart for a given failure scenario prior to 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 MVS MODIFY command to the NetView task
- From timers in NetView
- Triggered via the SA automation tables

This capability, along with the query services interface described in 3.6.3, “GDPS/PPRC Query Services” on page 75 provides a rich framework for user customizable systems management procedures.

3.5 GDPS/PPRC monitoring and alerting

The GDPS SDF panel, discussed in “Monitoring function - Status Display Facility” on page 62, 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. Customers run with multiple PPRC links and if one is broken, PPRC still continues over any remaining links. However, it is important that operations are 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 could result in production impact due to insufficient mirroring bandwidth or total loss of PPRC connectivity (which would result 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.

Some GDPS monitoring functions are executed, not only on the GDPS controlling systems, but also 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 could 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 of and can take corrective action for any problems in their environment that can impact the ability of GDPS/PPRC to carry out recovery operations. This will maximize the chance of achieving your availability and RPO/RTO commitments.

3.5.1 GDPS/PPRC and the z/OS Health Checker

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 GDPS-related settings adhere to best practices.

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

Often, if there are changes in the customer environment, this might necessitate adjustment of some parameter settings associated with z/OS, GDPS, and other products. It is possible that you could miss making these adjustments, which could affect GDPS. The GDPS health checks are intended to detect such situations and avoid incidents where GDPS is unable to perform its job due to 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 CDS are allocated and defined to GDPS in line with the GDPS best practices recommendations.

Similar to z/OS and other products that provide health checks, GDPS health checks are optional. Additionally, a number of the best practices values that are checked, as well as the frequency of the checks, are customer-customizable to cater to unique customer environments and requirements.

3.6 Other GDPS-related facilities

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.

3.6.1 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 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 reliable propagated by z/OS during the HyperSwap to the new primary volumes. For HyperSwap, all RESERVEs must be converted to GRS global enqueue via 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 via GRS RNL processing. This coexistence problem has also been resolved so that GDPS is now informed prior to 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 will re-enable HyperSwap.

3.6.2 GDPS/PPRC reduced impact initial copy and resynchronization

Previously, in a GDPS/PPRC configuration, some customers can defer the initial copy or resynchronization of the disks to a period of low workload activity to mitigate any possible performance impact on production workloads. They can also manually pace the number of volumes that are concurrently initial copied or resynchronized.

Starting with GDPS V3.6, the GDPS/PPRC `DASD START SECONDARY` script statement is enhanced to initiate the initial copy and resynchronization using asynchronous PPRC-XD (known as Global Copy on IBM Enterprise Disk subsystems). GDPS then monitors progress of the copy operation and when the volumes are near full duplex state, GDPS will convert the replication from the asynchronous copy mode to synchronous PPRC. Initial copy or resynchronization using PPRC-XD eliminates the performance impact of synchronous mirroring for production workload, allowing customers to establish or re-synchronize mirroring during periods of high production workload. In this way you can avoid exposing the configuration to a lack of HyperSwap capability by resynchronizing any time, instead of potentially waiting many hours or even days until there is a period of low activity.

This function requires that all disk subsystems in the GDPS configuration support PPRC-XD operation as implemented by IBM Global Copy.

3.6.3 GDPS/PPRC 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 available for GDPS 3.6 and higher that allows customer-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, as well as other GDPS facilities such as HyperSwap and GDPS Monitors.

Query Services allows customers 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 a number of samples in the GDPS SAMPLIB library to demonstrate how Query Services can be used in customer-written code. GDPS also makes available to customers 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 customers with an excellent example of how to write programs to exploit Query Services.

3.7 Flexible testing

It is highly recommended that you conduct regular DR drills or full production tests in your production 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 consistent copy of the data available if a disaster-type event should occur while testing your site failover and failback procedures.

In addition, FlashCopy could 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 could 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.

3.7.1 Usage of FlashCopy Space Efficient volumes

As discussed in “FlashCopy Space Efficient (FlashCopy SE)” on page 33, 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. Support is added to allow FlashCopy Space Efficient 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, panels or FlashCopies automatically taken by GDPS, can use Space Efficient targets.

One of the potential uses of Space Efficient volumes would be when mirrored pairs are suspended, and additional updates are being “recorded” at the local site. These updates must be sent to the remote site during the resynchronization process. A consistent copy of the data is not available until the resynchronization process completes, which is why it is recommended that a FlashCopy of the consistent data be taken before the resync begins.

Another potential use of Space Efficient volumes would be if you want to use the data for limited D/R testing.

You should have an understanding of some of the characteristics of Space Efficient FlashCopy to determine if this method of creating a point in time copy will satisfy your business requirements. For example, will it be acceptable to your business if, due to some unexpected workload condition, the repository on the disk subsystem for the Space Efficient devices gets full and your FlashCopy is invalidated such 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 should consider using standard FlashCopy instead.

3.8 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 would be extremely rare to find a team that had this range of skills across many implementations. However, the GDPS/PPRC offering includes just 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 some or all of the following:

- Planning to determine availability requirements, configuration recommendations, implementation and testing plans
- Installation and necessary customization of NetView and System Automation
- Remote copy implementation
- Peer-to-Peer VTS 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 things including what automation is already in place, whether remote copy is already in place, if the two centers are already in place with a multi-site sysplex, and so on. This means that the skills provided are tailored to the specific needs of each particular implementation.

3.9 GDPS/PPRC prerequisites

Refer to the following website for the latest GDPS/PPRC prerequisite information:

http://www.ibm.com/systems/z/advantages/gdps/getstarted/gdpspprc.html

3.10 Comparison of GDPS/PPRC versus other GDPS offerings

There are so many features and functions in the various members of the GDPS family that it is sometimes difficult to recall them all and to remember which offerings support them. 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/XRC	GDPS/GM
Continuous availability	Yes	Yes	No	No
Disaster recovery	Yes	Yes	Yes	Yes
Supported distance	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	Virtually unlimited	Virtually unlimited
FlashCopy support	Yes	Yes	Yes	Yes
Reduced impact initial copy/resynch	Yes	Yes	N/A	N/A
PtP VTS support	Yes	No	Yes	No

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/XRC	GDPS/GM
Production sysplex automation	Yes	No	No	No
Span of control	Both sites	Both sites (disk only)	Recovery site	Disk at both sites. Recovery site (CBU / LPARs)
GDPS scripting	Yes	No	Yes	Yes
Monitoring, alerting and health checks	Yes	Yes	Yes	Yes
Query Services	Yes	Yes	No	No
MGM	Yes (IR or non-IR)	Yes (Non-IR only)	N/A	Yes (IR or non-IR)
MzGM	Yes	Yes	Yes	N/A
Open LUN	Yes	Yes	No	Yes
z/OS equivalent functionality for Linux for System z	Yes	No	Yes	Yes
Heterogeneous support via DCM	Yes (VCS and SA AppMan)	No	Yes (VCS only)	Yes (VCS only)
Web interface	Yes	Yes	No	Yes

3.11 Summary

GDPS/PPRC is a powerful offering that provides disaster recovery, continuous availability, and sysplex resource management capabilities. HyperSwap, available with GDPS/PPRC (or GDPS/PPRC HyperSwap Manager), provides the ability to dynamically swap between disks across two sites. The power of automation allows you to test and perfect the actions to be taken, either for planned or unplanned changes, eliminating the risk of human error.

This is the one offering in the GDPS family that offers the potential of zero data loss, and which can achieve the shortest Recovery Time Objective, typically less than one hour following a complete site failure.

It is also the only member of the GDPS family that provides the capability to manage the production LPARs. Although GDPS/XRC offers LPAR management, its scope for system management only includes the systems in the recovery site, not the production systems running in Site1.

In addition to the disaster recovery and planned reconfiguration capabilities, GDPS/PPRC also provides a much more user-friendly interface for monitoring and managing the remote copy configuration.

4

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 is designed to extend 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 functions for protecting data provided by GDPS/PPRC HM include the following functions:

- Ensuring the consistency of the secondary 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
- Management of the remote copy configuration for System z and non-System z platform data

Since GDPS/PPRC HM is a subset of the GDPS/PPRC offering, you may want to simply skip to the comparison presented in Table 4-1 on page 104 if you have already read Chapter 3, “GDPS/PPRC” on page 43.

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 multi-site 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

2.2, “Data consistency” on page 15 mentions that data integrity across primary and secondary volumes of data is essential to perform a database restart instead of extensive recoveries. This section provides details of how automation in GDPS/PPRC HM is designed to provide both data consistency in the event of mirroring problems and data availability in the event of disk problems.

There are two different types of disk problems that trigger a GDPS automated reaction:

- 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 discussed in the section “Freeze function.”

- 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 is not responsive. This is discussed in “HyperSwap function” on page 82.

Freeze function

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 `CGROUP FREEZE` and `CGROUP RUN` commands which have been implemented as part of Metro Mirror and also by other enterprise disk vendors. In this way, so long as the disk hardware supports `CGROUP FREEZE/RUN` commands, GDPS can ensure consistency across all data in the Sysplex (consistency group) irrespective of disk hardware type. This preferred approach differs from proprietary hardware approaches that only work for one type of disk hardware. For a related introduction to data consistency with synchronous disk mirroring, refer to “PPRC data consistency” on page 21.

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

- Remote copy is suspended for each device in that pair.
- All paths between the indicated pair of logical subsystems (LSS) are removed.

- 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 will be issued to the affected devices until the ELB is reset with a `CGROUP RUN` command or until it times out (the consistency group timer setting commonly defaults to 120 seconds or 2 minutes).

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 will issue a `CGROUP FREEZE` for all LSS pairs containing devices in the GDPS configuration. A very important point is that 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 would not be unusual to see a short gap when the `FREEZE` command is issued to each disk subsystem, but because of the ELB, this gap is not a problem.

After GDPS detects that all remote copy sessions have been suspended, and the consistency of the remote disks is protected, GDPS will either issue a `CGROUP RUN` command to the LSSs, allowing them to continue operation in remote copy-suspended mode, or system reset all production systems (thereby ensuring that no non-mirrored updates can be applied), depending on your GDPS FREEZE policy as described in “Freeze policies” on page 81.

GDPS/PPRC HM uses a combination of storage subsystem and sysplex triggers to capture, at the first indication of a potential disaster, a data-consistent secondary site copy of your data using the `CGROUP FREEZE` function. In this way, the consistent image of the data is ensured on the secondary copy at the very first sign of a disaster, even before production applications are aware of I/O errors. Ensuring the data consistency of the secondary copy ensures that a normal restart can be performed, instead of having to perform DBMS recovery actions. This is the essential design element of GDPS to minimize the time to recover the critical workload in the event of a disaster at the primary site.

For more information about the implementation of PPRC and IBM Metro Mirror refer to *DS8000 Copy Services for IBM System z*, SG24-6787.

Freeze policies

Following a Freeze event, GDPS always performs a `CGROUP FREEZE` to create a consistent set of secondary volumes. The action that GDPS takes subsequent to the `CGROUP FREEZE` is specified by the installation in the GDPS freeze policy. Prior to GDPS/PPRC V3.7, you basically had two options:

- GO - GDPS will allow the production systems to continue operation after mirroring has been suspended.
- STOP - GDPS will reset the production systems while I/O is suspended.

GDPS V3.7 (with APAR) has introduced enhanced Freeze and Swap policy options. The next two sections FREEZE and STOP and FREEZE and GO are provided for completeness for those customers using V3.6 and for V3.7 and V3.8 customers who prefer not to change their Freeze and Swap policies. If you are installing GDPS/PPRC HM for the first time or an existing customer running V3.7 or higher, it is recommended that you change your Freeze and Swap policies to the ones described in “Enhanced Freeze and Swap policy options (V3.7 and V3.8)” on page 84.

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 policy choice 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 down before continuing.

If you are using duplexed structures along with a FREEZE and STOP policy, it may seem that you are guaranteed to be able to use the duplexed instance of your structures in the event you have to recover and restart your workload with the frozen secondary copy of your disks. However, this is not always the case! There could be rolling disaster scenarios where prior to, following, or during the freeze event, there is some sort of 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 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.

Note: If FREEZE and STOP is used, and the event that caused GDPS to take action was a transient event rather than a real disaster, you will have stopped all production systems unnecessarily.

FREEZE and GO

If you can accept an RPO that is *not necessarily* zero, you may decide to let the production systems *continue operation* after the secondary volumes have been protected. A FREEZE and GO policy is recommended for this case. This way you avoid an unnecessary outage if the trigger were to be only a transient event.

On the other hand, if the trigger is the first sign of an actual disaster, you could continue operating for some amount of time before all systems actually fail. Any updates made to the primary volumes during this time will not have been remote copied to the secondary disk, and therefore are lost. In addition, because the 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 recovery must be used to restart applications, resulting in elongated recovery times.

Considerations

The decision of whether to implement a STOP or GO freeze policy is really a business decision, rather than an IT decision. If your transactions are very high value, it may be more important to ensure that no transactions are ever lost, so you may decide on FREEZE and STOP. If you have huge volumes of relatively low value transactions, you may be willing to risk some lost data in return for avoiding unneeded outages with a FREEZE and GO policy.

Most installations start with a FREEZE and GO policy. Companies that have an RPO of zero will typically then move on and implement a FREEZE and STOP policy after the implementation is proven stable.

HyperSwap function

In the event that there is a problem writing or accessing the *primary* disk because of either a primary disk hard failure or because the disk subsystem is not accessible or not responsive, then there is a need to swap from the primary disk subsystems to the secondary disk subsystems.

GDPS/PPRC and GDPS/PPRC HM deliver a powerful function known as “HyperSwap.” HyperSwap provides the ability to non-disruptively swap from using the primary volume of a mirrored pair to using what had been the secondary volume. Prior to the availability of HyperSwap, an IPL was required on every system if you wanted to switch and run from the secondary volumes, meaning that it was not possible to maintain application availability across a switch from primary to secondary volumes.

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

There are two ways that HyperSwap can be invoked:

- Planned HyperSwap

  A planned HyperSwap is invoked manually by operator action using GDPS facilities. One example of a planned HyperSwap would be 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 failure of a primary disk device.

In both cases, the systems that are using the primary volumes will experience a temporary pause in processing. During this pause, the disk mirroring configuration is changed to allow use of the secondary volumes (and mirroring can be established in the opposite direction, depending on the option selected), the UCBs for the primary devices are updated to point to the formerly secondary volumes, and then the systems resume operation.

In benchmark measurements at IBM using currently supported releases of GDPS, the I/O hold time for an unplanned HyperSwap is generally less than 30 seconds for even very large configurations (for example, a 10-way Sysplex with approximately 20,000 mirrored volume pairs, or even a 30-way Sysplex with a more moderately sized disk configuration). Most implementations in the world are actually much smaller than this and typical I/O hold times using the most current storage and server hardware are generally measured in seconds. While results will obviously depend on your configuration, these numbers give you a high-end figure for what to expect.

A HyperSwap in GDPS/PPRC HM affects *all* mirrored LSSs with devices in the configuration (single consistency group). For example, if one single mirrored volume were to fail, and HyperSwap is invoked, processing would be swapped to the secondary copy of *all* mirrored volumes in the configuration, including those in other, unaffected, subsystems. The reason for this is that to maintain disaster readiness, all primary volumes *must* be in the same site. If HyperSwap were to only swap the failed LSS, you would have some primaries in one site, and the remainder in the other site.

Why is this necessary?

Consider the configuration in Figure 4-1 on page 84. This is what might happen if only the volumes of a single LSS or subsystem were HyperSwapped without swapping the whole consistency group. What would happen if there were a remote copy failure at 15:00? The secondary disks in both sites would be frozen at 15:00 and the primary disks (in the case of a FREEZE and GO policy) would 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 the volumes in that site would be consistent and if you had the disaster at 15:10, you would lose 10 minutes worth of data with the GO policy, but at least all the data in Site2 is usable. Using a FREEZE and STOP policy is not any better for this partial swap scenario because with a mix of primary disks in either site, you would have to maintain I/O configurations that could match *every* possible combination simply to IPL any systems. More likely, you would 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.

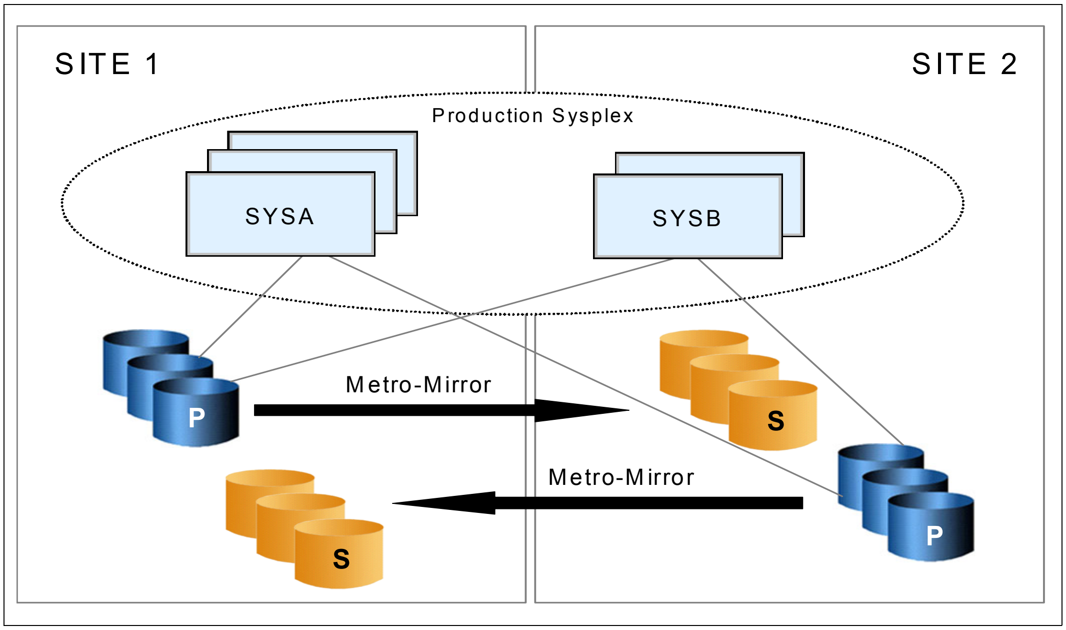


Figure 4-1 Unworkable Metro Mirror disk configuration

HyperSwap policy options

If you elect to use HyperSwap, there are additional options you can specify in your freeze policy that indicate a HyperSwap should be attempted following a HyperSwap trigger. Instead of FREEZE and STOP, you would specify SWAP,STOP**,** and instead of FREEZE and GO, you would specify SWAP,GO.

- If SWAP,STOP or SWAP,GO is specified, and you receive a HyperSwap trigger, any systems that successfully process the HyperSwap will be swapped and any systems that *cannot* process the HyperSwap will be system reset. Remember that the HyperSwap trigger indicates an error on a primary device, so any systems that continue to use that device would soon fail anyway.
- If you receive a freeze trigger, I/O to the systems will either be resumed after the secondary is consistent (if GO is specified) or the systems will be reset (if STOP is specified).

Enhanced Freeze and Swap policy options (V3.7 and V3.8)

The single Freeze policy option specified by SWAP,GO or SWAP,STOP was confusing and restrictive because it attempted to specify with a single specification what actions should be taken for mirroring problems and primary disk failures. It was not clear that the second parameter, GO or STOP, referred to actions to be taken if a mirroring problem was detected. It also did not allow the customer the flexibility to have different combinations, such as GO for mirroring problems and STOP for disk failures. Last but not the least, it resulted in inconsistent handling of certain triggers. The same trigger was sometimes handled as a mirroring problem and sometimes as a disk failure.

An enhancement in V3.8 base code and V3.7 code with an APAR allows you to split the policy options and separately specify what actions GDPS should take when

- Mirroring problems (Freeze triggers) are detected
- Disk failures (HyperSwap triggers) are detected

Mirroring Problems

Actions to be taken when Freeze triggers are detected can now be specified as either:

- PPRCFAILURE = GO

  After performing the Freeze, the production systems continue to update the primary disks. This is similar to Freeze and Go. Recovering on the secondary disks is safe since the secondary disk was frozen in a consistent state. However, recovering on secondary disks results in loss of any data that was written to the primary disk after the freeze.

- PPRCFAILURE = STOP

  After performing the Freeze, the production systems will be quiesced, resulting in all the work that is updating the primary PPRC devices being stopped. This is similar to Freeze and Stop. It is the only option that guarantees no data loss in case recovery on the secondary disk would be required.

- PPRCFAILURE = COND

  Field experience has shown that most of the 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. Some examples of these events are connectivity problems to the secondary disks or secondary disk subsystem failure conditions.

  If PPRCFAILURE = COND is specified, the action after the Freeze is **conditional** on the results of a new disk subsystem query. GDPS working in synergy with a new disk subsystem query function can determine the status of the secondary disk subsystems when a Freeze trigger is detected. If GDPS detects a secondary disk subsystem failure, then the actions are the same as GO, the production systems continue to update the primary disks; otherwise processing is the same as STOP.

  This option can improve the capability of zero data loss, and reduce the potential of loss of availability due to “False Freeze” events. It requires that all disk sub systems managed by GDPS support the *Query Storage Controller Status feature.*

Primary disk failures

Actions to be taken when HyperSwap triggers are detected can now be specified as either:

- PRIMARYFAILURE = GO

  After performing the Freeze, the disks are not swapped, and the production systems continue to update the primary disks. This is similar to PPRCFAILURE = GO.

- PRIMARYFAILURE = STOP

  After performing the Freeze, the disks are not swapped, and the production systems will be quiesced, resulting in all the work that is updating the primary PPRC devices being stopped. This is similar to PPRCFAILURE = STOP.

- PRIMARYFAILURE = SWAP,GO or SWAP,STOP

  The first parameter SWAP indicates that after performing the Freeze, GDPS should perform an unplanned HyperSwap if a primary disk problem is detected for one of the GDPS managed PPRCed disks. This is done only if the environment is enabled for HyperSwap at the time the failure occurs. GDPS transparently swaps to using the secondary disks instead of the primary, and applications do not incur any downtime.

  With this option, you must also decide on a policy for what GDPS should do if the environment is not enabled for HyperSwap at the time the failure occurs.

  The options for the second parameter are:

  – GO

    After performing the Freeze, the disks are not swapped, and the production systems continue to update the primary disks.

  – STOP

    After performing the Freeze, the disks are not swapped, and the production systems will be quiesced, resulting in all the work that is updating the primary PPRC devices being stopped.

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 difference 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 re-established with failback processing, the original primary devices become secondary devices and a resynchronization of changed tracks takes place.

GDPS/PPRC and GDPS/PPRC HM transparently exploit the PPRC FO/FB capability if it is installed on the disk subsystems. This support mitigates RPO exposures by reducing the amount of time needed to resynchronize mirroring after a HyperSwap. Of course, the resync time will depend on how long mirroring was suspended and the number of changed tracks that must be transferred.

All disk subsystems in your GDPS configuration, in both Site1 and Site2, must support PPRC Failover/Failback for GDPS to exploit this capability.

Protection during IPL

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 usage) perspective but physically from a PPRC perspective have simplex or primary status.

For both planned and unplanned HyperSwap, GDPS points to the new set of primary disks. However, for both these actions, if the former primaries are accessible, although they are considered to be the secondary disks from a GDPS perspective, they are still usable by applications since their actual PPRC status is not secondary. That is why it is possible to accidentally IPL from the wrong set of disks. Accidentally using the wrong set of disks could result in a potential data integrity or data loss problem.

GDPS/PPRC HM V3.8 has been enhanced to provide IPL protection early in the IPL process. During initialization of GDPS, a system IPLed on the wrong set of disks will not be allowed to continue running.

4.1.2 Protecting distributed (FBA) data

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

- System z or zSeries disks

  Prior to Open LUN support, GDPS only managed disks that were used by System z systems, although disks could be z/VM, VSE, or Linux on System z disks. All these disks are formatted as Count-Key-Data (CKD) disks, the traditional mainframe format. There was no specific name used for these subsystems.

  However, with the introduction of GDPS support for disks used by platforms other than System z, we have had to expand the terminology in some places to differentiate between disks used by systems running on a System z server (CKD), and those used by the non-System z server.

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

- Open LUN disks

  Disks that are used by systems other than those running on zSeries are traditionally formatted as Fixed Block Architecture (FBA). In this book, we generally use the term “Open LUN disks” to refer to such devices. However, the term “FBA disks” is used sometimes as well; both terms are used interchangeably.

GDPS/PPRC HM can manage the mirroring of FBA devices (also known as Open LUNs) in use by non-mainframe operating systems; this also includes SCSI disks written by Linux for System z. The FBA 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.

GDPS requires CKD utility devices in the disk subsystem in order for it to send commands to monitor and control the mirroring of the FBA devices. More specifically, GDPS will need at least one CKD utility device in each hardware cluster of the storage subsystem. A sample configuration of this function, called Open LUN Management, is shown in Figure 4-2 on page 88.

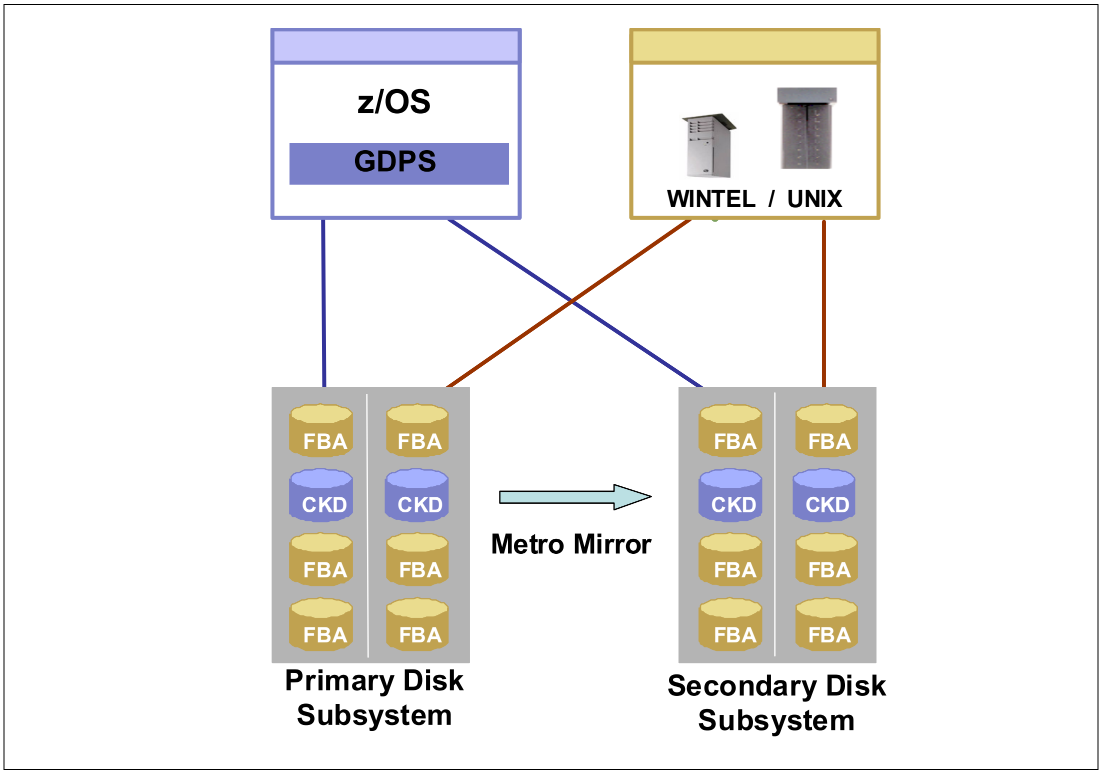


Figure 4-2 GDPS/PPRC HM Open LUN Management

4.1.3 Protecting other CKD data

GDPS/PPRC HM can also manage the disk mirroring of CKD format disks used by non-GDPS System z images: z/OS, Linux on System z, VM, and VSE LPARs that are not running GDPS/PPRC HM. Because GDPS/PPRC HM is not able to communicate with these images, any errors on those devices will not be recognized by GDPS. Therefore, although you can use GDPS/PPRC HM to manage the non-GDPS CKD disks, you cannot have the same guarantee of consistency that you do for systems running the GDPS code.

In the case of a freeze event on systems running GDPS/PPRC HM, the disks belonging to the non-GDPS images will also be frozen. If you have a FREEZE and STOP policy, and GDPS/PPRC HM has access to the HMC for any non-GDPS System z images, GDPS has the ability to system reset the associated LPARs.

An additional consideration is in relation to HyperSwap. As one example, because VSE does not support HyperSwap and there is no GDPS code running on the non-GDPS System z images, should an unplanned HyperSwap occur, these systems will not swap to the secondaries. It is possible that, in some circumstances, the non-GDPS System z systems could continue to update the old primaries, resulting in a data integrity exposure. For this reason, if you plan to exploit HyperSwap, and you also want to manage CKD volumes on non-GDPS System z images, you should be aware of this potential exposure. There is no problem doing a HyperSwap as long as you set the consistency group timer value high enough (hours to days) or you manually shut down the non-GDPS System z images prior to a planned HyperSwap.

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 located 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 may 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 may 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 just the Sysplex couple data sets). By being configured to be as self-contained as possible, the controlling system will be unaffected by errors that may 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 GDPS-managed mirrored disks.

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; and for initiating a freeze and implementing the freeze policy actions following a freeze event. Refer to 4.1.1, “Protecting data integrity and data availability” on page 80 for more details about freeze policies and HyperSwap processing.

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.

Improved controlling system availability - enhanced timer support

Enhancements in GDPS V3.6 and z/OS V1.9 (with enabling software updates) help improve GDPS recovery times for events that impact the primary time source for the sysplex, whether the time source is Server Time Protocol (STP) or External Time Reference (ETR) based. These enhancements allow the 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 may have started, instead of being in a disabled WTOR state. 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 IEA015A if the CPC that z/OS is running on is in ETR timing mode (either in an ETR network or in an STP Mixed Coordinated Timing Network (CTN)), and it is IEA394A if the CPC is in STP timing mode (either in an STP Mixed CTN or STP-only CTN).

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 in the event of a failure. Actions may include:

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

Because the controlling system only needs to run 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 TOD clock of the server (referred to as local timing mode), instead of generating a WTOR.

Automated response to ETR or STP sync WTORs

GDPS on the Controlling systems, using the BCP Internal Interface, provides automation to reply to WTORs IEA015A or IEA394A when the Controlling systems are running in local timing mode. Refer to “Improved controlling system availability - enhanced timer support” on page 89. A server in an ETR or STP network may have recovered from an unsynchronized to a synchronized timing state without customer intervention. By automating the response to the WTORs, potential time outs of subsystems and applications in the customer’s enterprise may 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, on the other hand, the CPC is still in an unsynchronized state when GDPS automation responded with RETRY to the WTOR. 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 4-3 on page 91. 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. Note that 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 will 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 may be acceptable, we recommend 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.

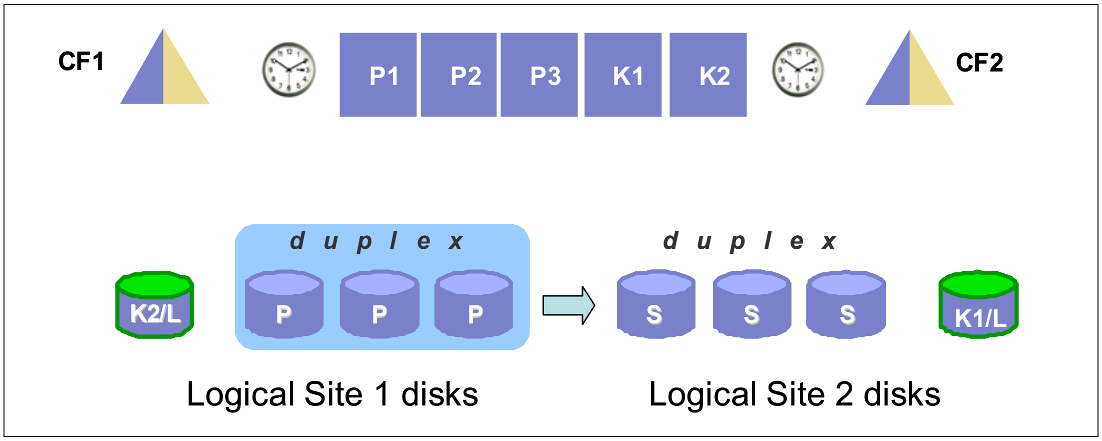


Figure 4-3 GDPS/PPRC HM single-site configuration

4.2.3 GDPS/PPRC HM in a two-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 4-4. This configuration *does* provide the foundation for disaster recovery since the secondary copy of disk is in a separate site protected from a disaster in Site. Additionally, GDPS/PPRC HM delivers the freeze capability ensuring a consistent set of secondary disk in case of a disaster.

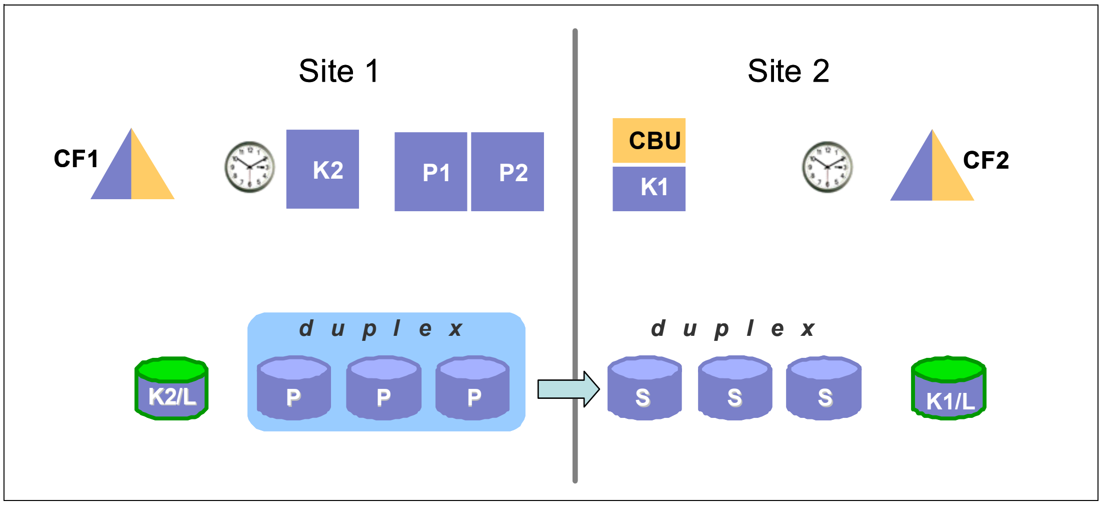


Figure 4-4 GDPS/PPRC HM two-site configuration

If you have a two-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 multi-site 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 56.

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 may not be acceptable.

4.2.4 GDPS/PPRC HM in a three-site configuration

GDPS/PPRC HM can be combined with GDPS/XRC or GDPS/GM in a three-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 some additional automation capabilities.

Refer to Chapter 8, “Combining Local/Metro continuous availability with out-of-region disaster recovery” on page 173 for a more detailed discussion of the capabilities and limitations of using GDPS/PPRC HM in a GDPS/MGM and GDPS/MzGM solution.

4.2.5 Other 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 environment

The bulk of the functionality delivered with GDPS/PPRC HM relates to maintaining the integrity of the secondary disks, and being able to non-disruptively 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

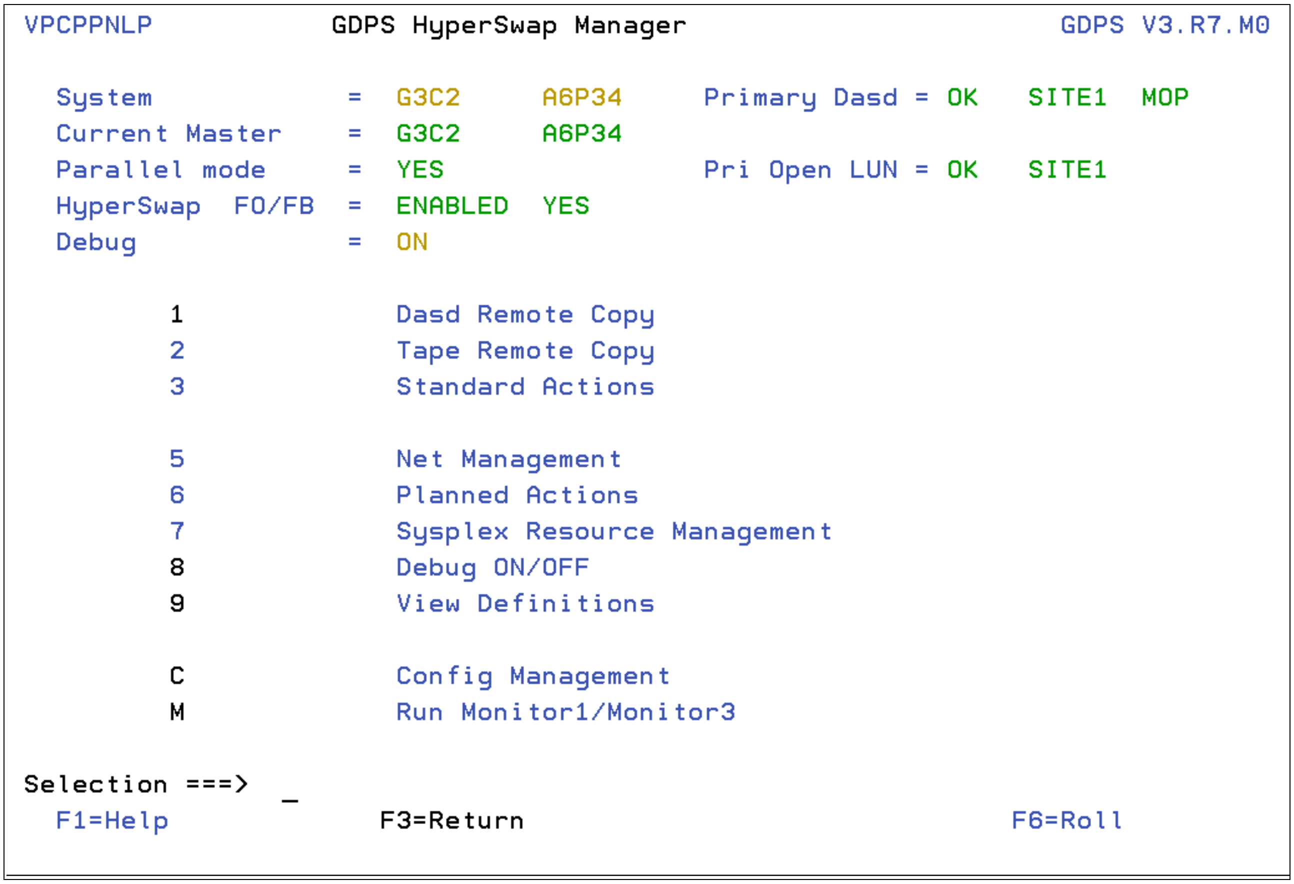
Note that GDPS/PPRC HM does not provide script support, so all these functions are only available through the GDPS NetView interfaces.

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 “web interface” in this document).

An example of the main GDPS/PPRC 3270-based panel is shown in Figure 4-5 on page 93.

Notice that a number of the 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.



```
VPCPPNLP            GDPS HyperSwap Manager                       GDPS V3.R7.M0

  System             =  G3C2     A6P34     Primary Dasd = OK    SITE1   MOP
  Current Master     =  G3C2     A6P34
  Parallel mode      =  YES                Pri Open LUN = OK    SITE1
  HyperSwap  FO/FB   =  ENABLED  YES
  Debug              =  ON

         1              Dasd Remote Copy
         2              Tape Remote Copy
         3              Standard Actions

         5              Net Management
         6              Planned Actions
         7              Sysplex Resource Management
         8              Debug ON/OFF
         9              View Definitions

         C              Config Management
         M              Run Monitor1/Monitor3

Selection ===>
  F1=Help      _       F3=Return                                 F6=Roll
```

Figure 4-5 Main GDPS/PPRC HM 3270-based panel

This panel is relatively simple to use, with a summary of configuration status at the top of the screen and a menu of choices that a user can select from below. As an example, a user would simply type a `1` at the `Selection ===>` prompt and press Enter to view the disk mirroring ("Dasd Remote Copy") panels.

GDPS web interface

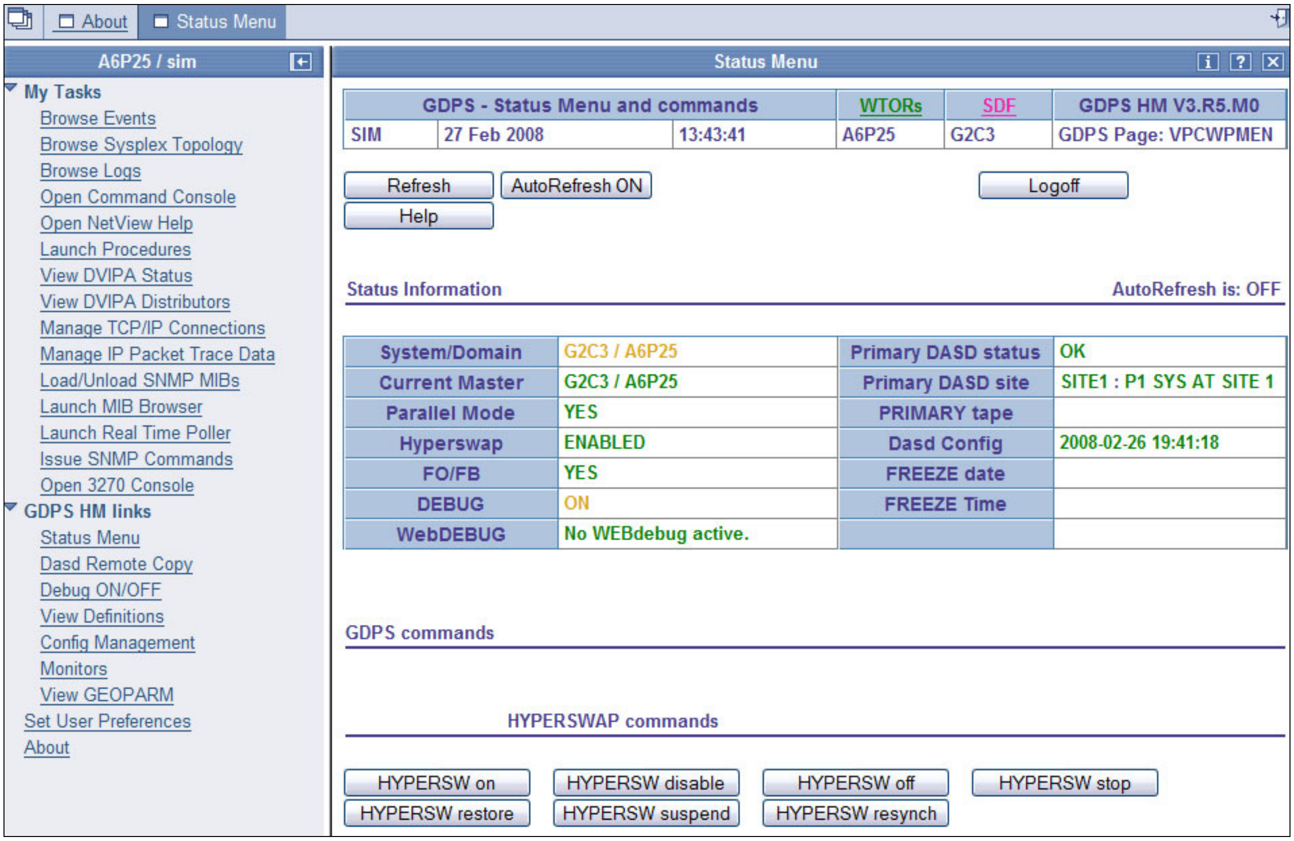
The web interface is a browser-based interface alternative to the traditional 3270 panels designed to improve operator productivity. The web interface provides the same functional capability as the 3270-based panels, such as providing management capabilities for Remote Copy Management and SDF Monitoring. In addition, users can open multiple windows to allow for continuous status monitoring, while performing other GDPS/PPRC management functions.

The web interface display is split into 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 panel of the GDPS/PPRC HM web interface is shown in Figure 4-6. The left frame, shown below `GDPS HM links`, allows you to select the menu options. These options can be displayed at all times, or you can optionally collapse the frame.

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



GDPS - Status Menu and commands			WTORs	SDF	GDPS HM V3.R5.M0
SIM	27 Feb 2008	13:43:41	A6P25	G2C3	GDPS Page: VPCWPMEN

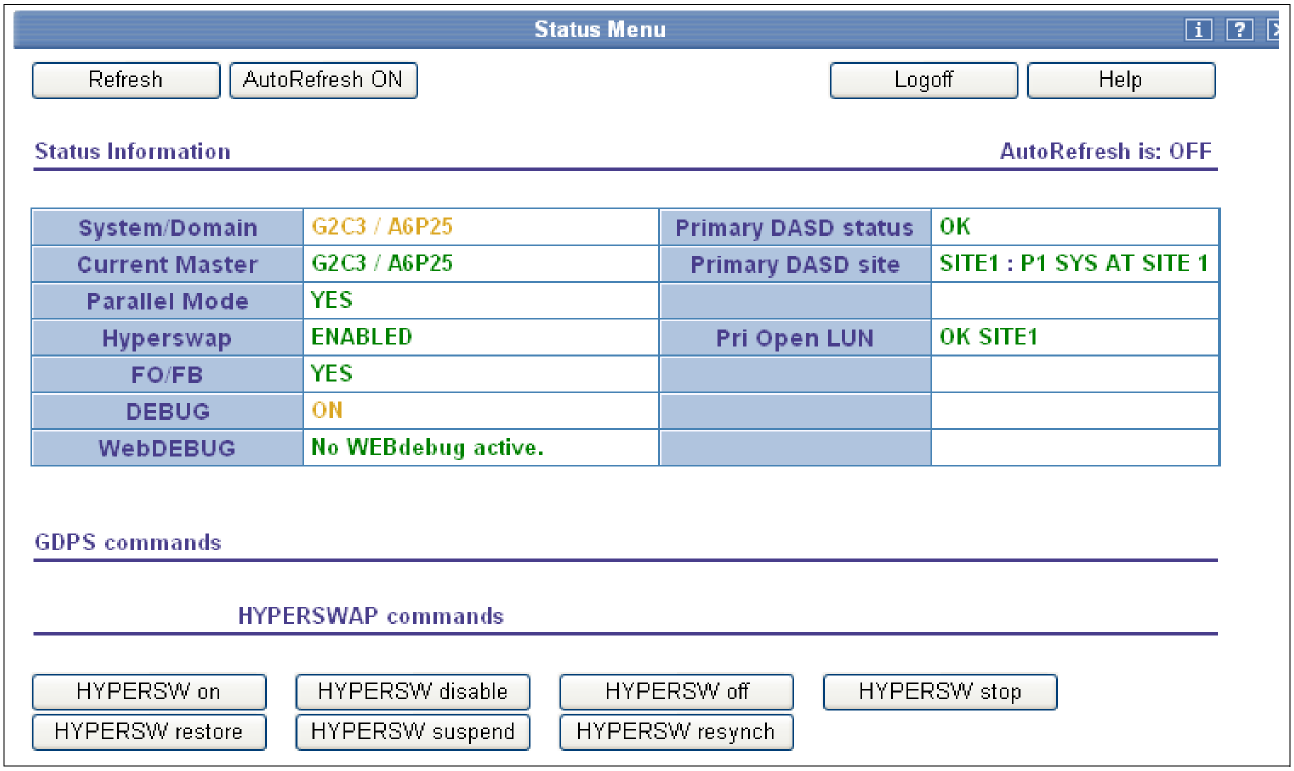
Status Information AutoRefresh is: OFF

System/Domain	G2C3 / A6P25	Primary DASD status	OK
Current Master	G2C3 / A6P25	Primary DASD site	SITE1 : P1 SYS AT SITE 1
Parallel Mode	YES	PRIMARY tape	
Hyperswap	ENABLED	Dasd Config	2008-02-26 19:41:18
FO/FB	YES	FREEZE date	
DEBUG	ON	FREEZE Time	
WebDEBUG	No WEBdebug active.		

Figure 4-6 Full view of GDPS main panel with task bar and status information

Main Status panel

The GDPS Web Interface status panel shown in Figure 4-7 on page 95 is the equivalent to the main GDPS panel. The information on this frame is what is found on the top portion of the 3270 GDPS Main panel. Additionally, this screen includes the `HYPERSW` commands that could be entered at a 3270 prompt in buttons at the bottom of the screen.



Status Information AutoRefresh is: OFF

System/Domain	G2C3 / A6P25	Primary DASD status	OK
Current Master	G2C3 / A6P25	Primary DASD site	SITE1 : P1 SYS AT SITE 1
Parallel Mode	YES		
Hyperswap	ENABLED	Pri Open LUN	OK SITE1
FO/FB	YES		
DEBUG	ON		
WebDEBUG	No WEBdebug active.		

GDPS commands

HYPERSWAP commands

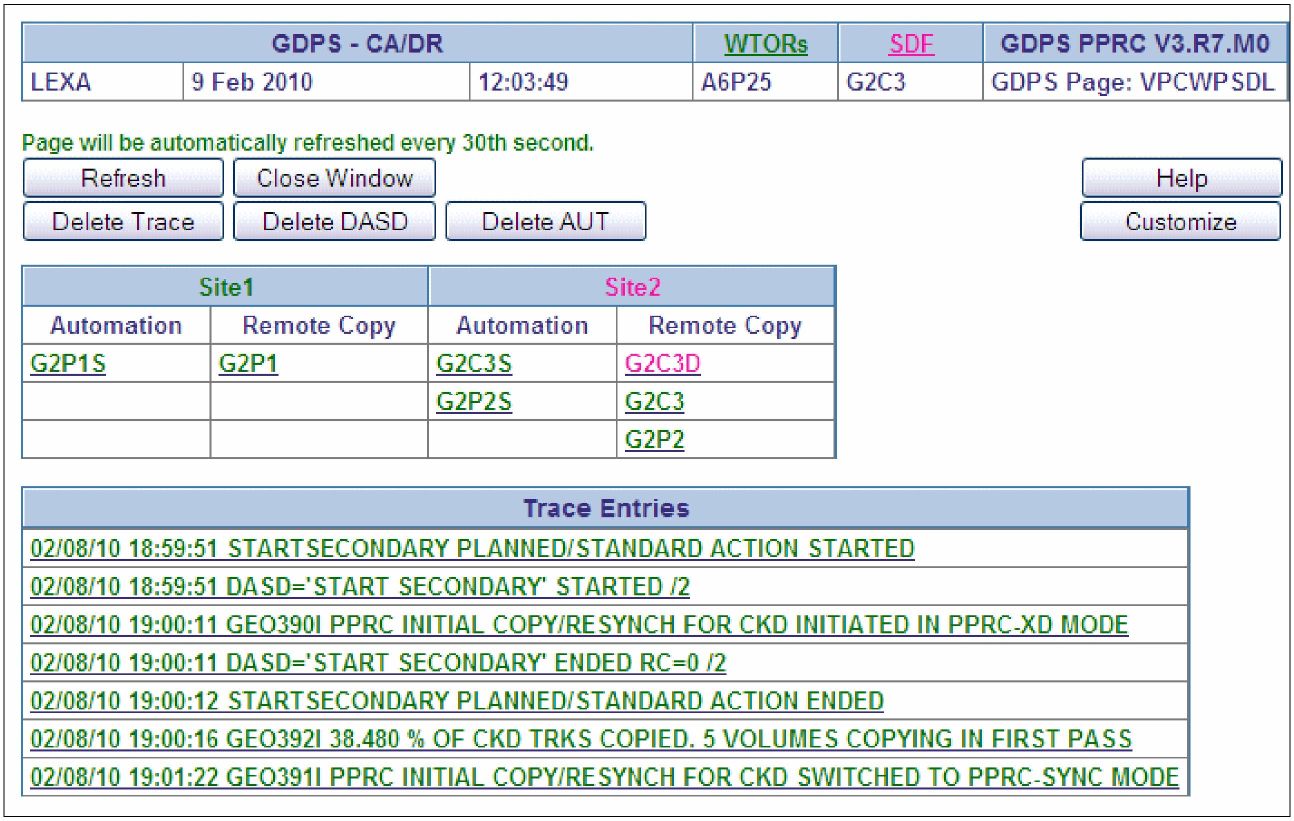
Figure 4-7 GDPS web interface - Main Status panel

Monitoring function - Status Display Facility (SDF)

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. It is the only dynamically updated status display available for GDPS.

SDF provides a dynamically-updated color-coded panel, as shown in Figure 4-8 on page 96. 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 should have an SDF panel within view so they will immediately become aware of anything requiring intervention or action.

The web interface can be set up to automatically refresh every 30 seconds. 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 and clicking any of the highlighted fields, you can obtain detailed information regarding the alert.



GDPS - CA/DR			WTORs	SDF	GDPS PPRC V3.R7.M0
LEXA	9 Feb 2010	12:03:49	A6P25	G2C3	GDPS Page: VPCWPSDL

Page will be automatically refreshed every 30th second.

Site1		Site2	
Automation	Remote Copy	Automation	Remote Copy
G2P1S	G2P1	G2C3S	G2C3D
		G2P2S	G2C3
			G2P2

Trace Entries
02/08/10 18:59:51 STARTSECONDARY PLANNED/STANDARD ACTION STARTED
02/08/10 18:59:51 DASD='START SECONDARY' STARTED /2
02/08/10 19:00:11 GEO390I PPRC INITIAL COPY/RESYNCH FOR CKD INITIATED IN PPRC-XD MODE
02/08/10 19:00:11 DASD='START SECONDARY' ENDED RC=0 /2
02/08/10 19:00:12 STARTSECONDARY PLANNED/STANDARD ACTION ENDED
02/08/10 19:00:16 GEO392I 38.480 % OF CKD TRKS COPIED. 5 VOLUMES COPYING IN FIRST PASS
02/08/10 19:01:22 GEO391I PPRC INITIAL COPY/RESYNCH FOR CKD SWITCHED TO PPRC-SYNC MODE

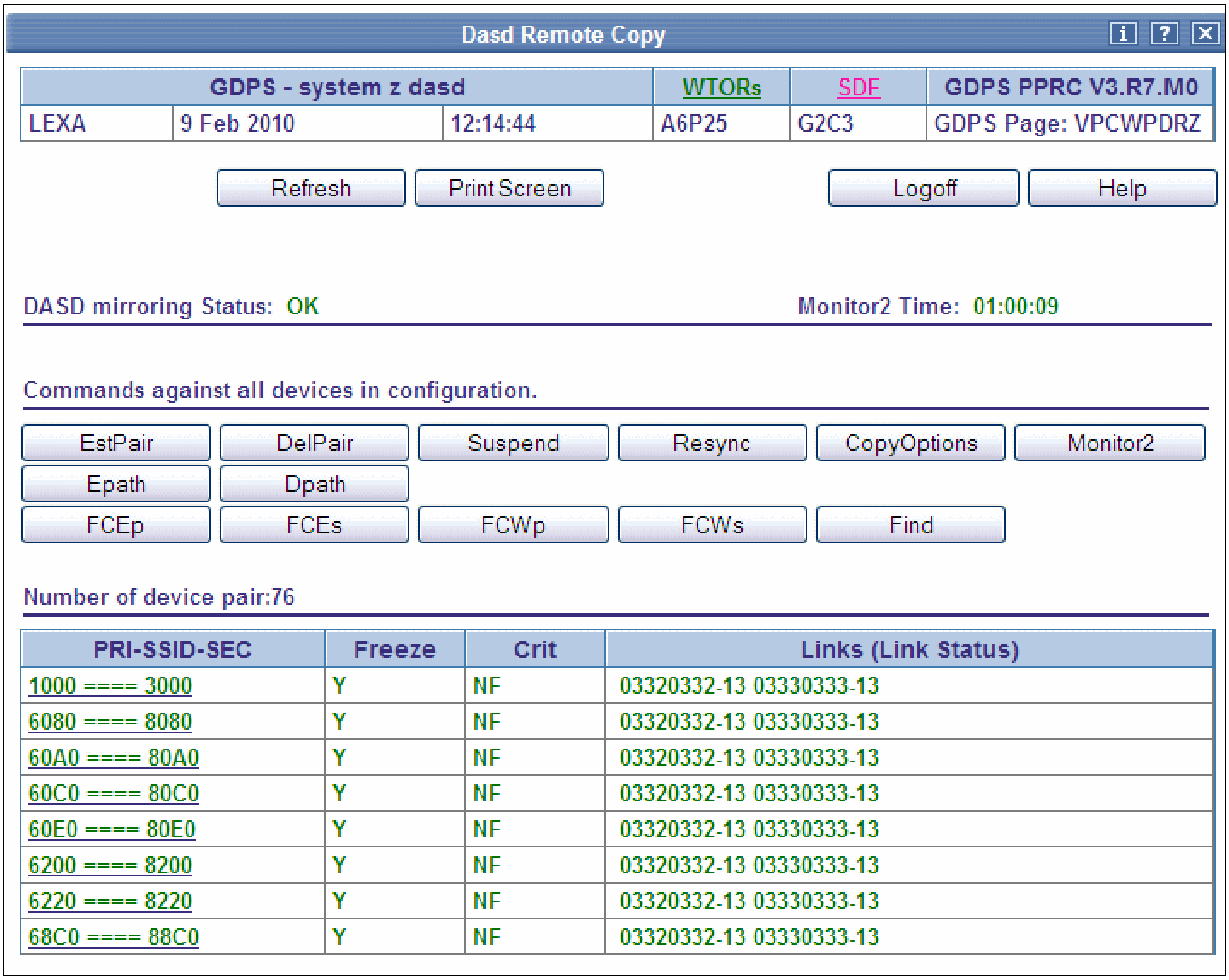
Figure 4-8 NetView SDF web interface

Remote copy panels

The z/OS Advanced Copy Services capabilities are very powerful, but the native z/OS TSO and ICKDSF interfaces are not very user friendly. To make it easier for operators to check and manage the remote copy environment, you can (and should) use the GDPS-provided `DASD Remote Copy` panels.

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 desired one. You can start, stop, suspend, and resynchronize mirroring at the volume or LSS level. You can initiate a FlashCopy, and you can reverse the direction of mirroring. These actions can be carried out at the device or LSS level, or both, as appropriate. Figure 4-9 on page 97 shows the mirroring panel for CKD devices.



GDPS - system z dasd			WTORs	SDF	GDPS PPRC V3.R7.M0
LEXA	9 Feb 2010	12:14:44	A6P25	G2C3	GDPS Page: VPCWPDRZ

DASD mirroring Status: OK Monitor2 Time: 01:00:09

Commands against all devices in configuration.

Number of device pair:76

PRI-SSID-SEC	Freeze	Crit	Links (Link Status)
1000 ==== 3000	Y	NF	03320332-13 03330333-13
6080 ==== 8080	Y	NF	03320332-13 03330333-13
60A0 ==== 80A0	Y	NF	03320332-13 03330333-13
60C0 ==== 80C0	Y	NF	03320332-13 03330333-13
60E0 ==== 80E0	Y	NF	03320332-13 03330333-13
6200 ==== 8200	Y	NF	03320332-13 03330333-13
6220 ==== 8220	Y	NF	03320332-13 03330333-13
68C0 ==== 88C0	Y	NF	03320332-13 03330333-13

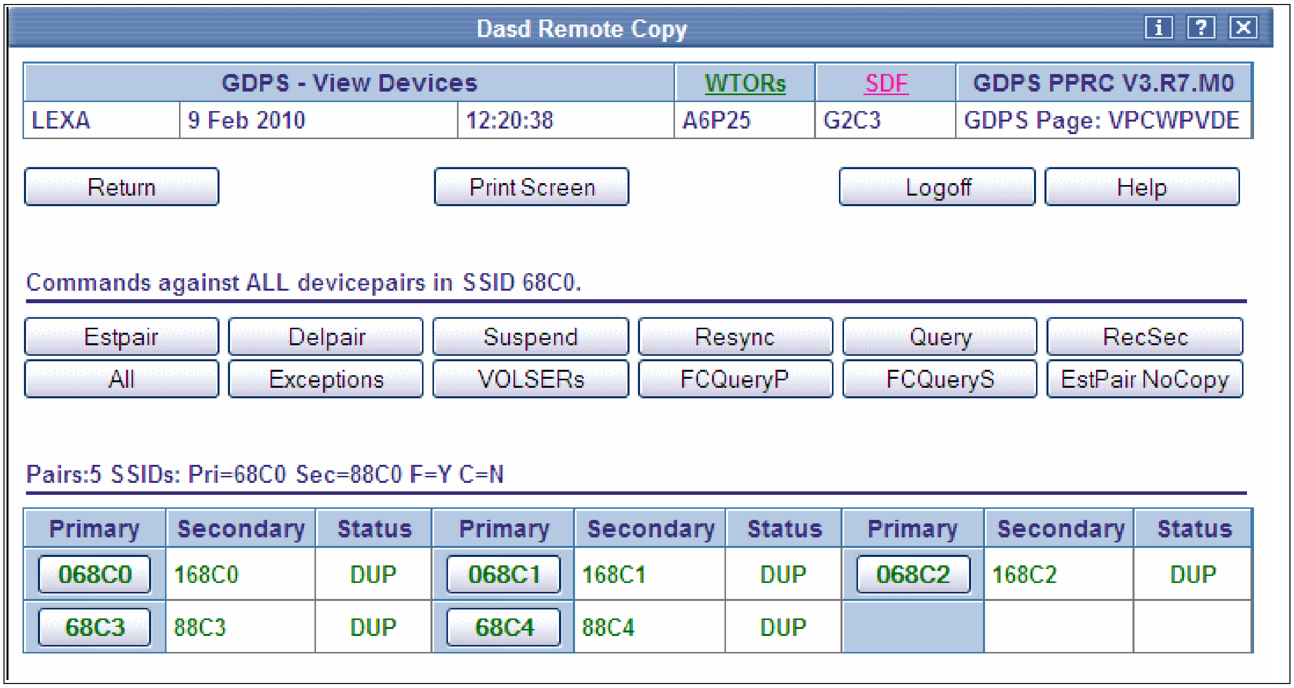
Figure 4-9 DASD Remote Copy SSID web interface

The DASD Remote Copy frame is organized into three sections:

- A top section displays the mirroring status, and buttons for Return, Refresh and Help.
- A middle section displays actions against all the panel-displayed SSID-pairs (similar to the bottom section of the equivalent 3270 panel).
- A bottom section displays the list of all the SSID-pairs.

To perform an action on a single SSID-pair, click the actual pair. This brings you to a panel 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 has been selected, the frame shown in Figure 4-10 on page 98 is displayed. The bottom of 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 (indicated as DUP on the panel). Also note that the secondary devices for some of these pairs are in an alternate subchannel set (MSS1 in this case). Additional details can be viewed for each pair by clicking the button for the primary device in the pair.



GDPS - View Devices			WTORs	SDF	GDPS PPRC V3.R7.M0
LEXA	9 Feb 2010	12:20:38	A6P25	G2C3	GDPS Page: VPCWPVDE

Commands against ALL devicepairs in SSID 68C0.

Pairs:5 SSIDs: Pri=68C0 Sec=88C0 F=Y C=N

Primary	Secondary	Status	Primary	Secondary	Status	Primary	Secondary	Status
068C0	168C0	DUP	068C1	168C1	DUP	068C2	168C2	DUP
68C3	88C3	DUP	68C4	88C4	DUP			

Figure 4-10 Web Interface Dasd Remote Copy "View Devices" detail frame

If you are familiar with using the TSO or ICKDSF interfaces, you will appreciate how much more user friendly the panel is.

Remember that these GDPS-provided panels are *not* intended to be a remote copy monitoring tool. Because of the overhead involved in gathering the information to populate the NetView panels, GDPS only gathers this information 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. These commands are entered at a NetView command prompt. Some examples are described in this section. Note that this is not a comprehensive list, a number of other GDPS Netview commands are available.

HYPERSW

The HYPERSW command can be used for the following purposes:

- 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 via a PPRC failover or recover action.
- Restore PPRC mirroring to a duplex state.

GDPSTIME

GDPSTIME is a new GDPS Netview command which has been introduced with GDPS V3.8 to allow 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 - PTS), which CPC is able to take over as the clock source for planned and unplanned events (Backup Time Server - BTS), which CPC is the active clock source (Current Time Server - CTS), and which CPC assists in STP recovery (Arbiter).

Planned Operations

It is strongly recommended that the server roles be reassigned prior to performing planned disruptive actions on any of these special role servers. Examples of planned disruptive actions are Power on Reset (POR), Activate/Deactivate. GDPSTIME provides you with a single point of control to perform the STP CTN reconfiguration prior to performing planned disruptive actions, instead of managing this from the STP panels of the HMC.

Unplanned Failure

If a failure condition has resulted in either the PTS, BTS or Arbiter to be no longer an operational synchronized CPC in the CTN, it is recommended 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 new GDPSTIME command can be used to perform the reassignment, which eliminates the requirement for the operator to perform the STP reconfiguration task manually at the HMC.

4.4 GDPS/PPRC HM monitoring and alerting

The GDPS SDF panel, discussed in “Monitoring function - Status Display Facility (SDF)” on page 95, 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. Customers run with multiple PPRC links and if one is broken, PPRC still continues over any remaining links. However, it is important that operations are 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 could result in production impact due to insufficient mirroring bandwidth or total loss of PPRC connectivity (which would result 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.

Some GDPS monitoring functions are executed not only on the GDPS controlling systems, but also 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 could 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 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 impact the ability of GDPS/PPRC HM to carry out recovery operations. This will maximize the chance of achieving your IT resilience commitments.

4.4.1 GDPS/PPRC and the z/OS Health Checker

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 GDPS-related settings adhere to best practices.

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

Often, if there are changes in the customer environment, this might necessitate adjustment of some parameter settings associated with z/OS, GDPS, and other products. It is possible that you could 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 due to 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 CDS are allocated and defined to GDPS in line with the GDPS best practices recommendations.

Similar to z/OS and other products that provide health checks, GDPS health checks are optional. Additionally, a number of the best practices values that are checked, as well as the frequency of the checks, are customer-customizable to cater to unique customer environments and requirements.

4.5 Other GDPS-related facilities

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 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 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 via 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 via GRS RNL processing. This coexistence problem has also been resolved so that GDPS is now informed prior to 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 will reenable HyperSwap.

4.5.2 GDPS/PPRC HM reduced impact initial copy and resynchronization

Previously, in a GDPS/PPRC HM configuration, some customers could defer the initial copy or resynchronization of the disks to a period of low workload activity to mitigate any possible performance impact on production workloads. They could also manually pace the number of volumes that are concurrently initial copied or resynchronized.

Starting with GDPS V3.6, the `HYPERSW RESTORE` command is enhanced to initiate the initial copy and resynchronization using asynchronous PPRC-XD (known as Global Copy on IBM Enterprise Disk subsystems). GDPS then monitors progress of the copy operation and when the volumes are near full duplex state, GDPS will convert the replication from the asynchronous copy mode to synchronous PPRC. Initial copy or resynchronization using PPRC-XD eliminates the performance impact of synchronous mirroring for production workload, thereby allowing customers to establish or re-synchronize mirroring during periods of high production workload. In this way you can avoid exposing the configuration to a lack of HyperSwap capability by resynchronizing any time instead of potentially waiting many hours, or even days later, when there is a period of low activity.

This function requires that all disk subsystems in the GDPS configuration support PPRC-XD operation as implemented by IBM Global Copy.

4.5.3 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 available for GDPS 3.6 and higher that allows customer-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, as well as other GDPS facilities such as HyperSwap and GDPS Monitors.

Query Services allows customers 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 a number of samples in the GDPS SAMPLIB library to demonstrate how Query Services can be used in customer-written code. GDPS also makes available to customers 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 customers with an excellent example of how to write programs to exploit Query Services.

4.6 Flexible testing

It is highly recommended that you conduct regular DR drills or full production tests in your production 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 consistent copy of the data available if a disaster-type event should occur while testing your site failover and failback procedures.

In addition, FlashCopy could 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 could 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.

4.6.1 Usage of FlashCopy Space Efficient volumes

As discussed in “FlashCopy Space Efficient (FlashCopy SE)” on page 33, by using Space Efficient (SE) volumes, you may be able to lower the amount of physical storage needed and thereby reduce the cost associated with providing a tertiary copy of the data. Support is added to allow FlashCopy Space Efficient 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, panels or FlashCopies automatically taken by GDPS, can use Space Efficient targets.

One of the potential uses of Space Efficient volumes would be when mirrored pairs are suspended, and additional updates are being “recorded” at the local site. These updates must be sent to the remote site during the resynchronization process. A consistent copy of the data is not available until the resynchronization process completes, which is why it is recommended that a FlashCopy of the consistent data be taken before the resync begins.

Another potential use of Space Efficient volumes would be if you want to use the data for limited D/R testing.

You should have an understanding of some of the characteristics of Space Efficient FlashCopy to determine if this method of creating a point in time copy will satisfy your business requirements. For example, will it be acceptable to your business if due to some unexpected workload condition, the repository on the disk subsystem for the Space Efficient devices gets full and your FlashCopy is invalidated such 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 should consider using standard FlashCopy instead.

4.7 Services component

As you have seen, GDPS touches on much more than simply remote copy. It also includes automation, 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 would be extremely rare to find a team that had this range of skills across many implementations. However, the GDPS/PPRC HM offering includes just 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. Still, the sizing of the services component of each project can be tailored for that project, based on many things, including what automation is already in place, whether remote copy is already in place, if the two centers are already in place with a multi-site sysplex if required, and so on. This means that the skills provided are tailored to the specific needs of each particular implementation.

4.8 GDPS/PPRC HM prerequisites

Refer to the following website for the latest GDPS/PPRC HM prerequisite information:

http://www.ibm.com/systems/z/advantages/gdps/getstarted/gdpspprc_hsm.html

4.9 Comparison of GDPS/PPRC HM versus other GDPS offerings

There are so many features and functions in the various members of the GDPS family that it is sometimes difficult to recall them all and to remember which offerings support them. To position the offerings, Table 4-1 on page 104 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/XRC	GDPS/GM
Continuous availability	Yes	Yes	No	No
Disaster recovery	Yes	Yes	Yes	Yes
Supported distance	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	Virtually unlimited	Virtually unlimited
FlashCopy support	Yes	Yes (NOCOPY only)	Yes (Zero Suspend)	Yes (No UCB)
Reduced impact initial copy/resync	Yes	Yes	N/A	N/A
PtP VTS support	Yes	No	Yes	No
Production sysplex automation	Yes	No	No	No
Span of control	Both sites	Both sites (disk only)	Recovery site	Disk at both sites. Recovery Site (CBU / LPARs)
GDPS scripting	Yes	No	Yes	Yes
Monitoring, alerting, and health checks	Yes	Yes	Yes	Yes
Query Services	Yes	Yes	No	No
MGM	Yes (IR or non-IR)	Yes (Non-IR only)	N/A	Yes (IR or non-IR)
MzGM	Yes	Yes	Yes	N/A
Open LUN	Yes	Yes	No	Yes
z/OS equivalent functionality for Linux for System z	Yes	No	Yes	Yes
Heterogeneous support via DCM	Yes (VCS and SA AppMan)	No	Yes (VCS only)	Yes (VCS only)
Web interface	Yes	Yes	No	Yes

4.10 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 multi-site 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 as well.

In addition to disaster recovery and continuous availability capabilities, GDPS/PPRC HM also provides a much more user-friendly interface for monitoring and managing the remote copy configuration.

5

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 just GDPS/PPRC with a longer distance between the sites; additional differences are discussed in this chapter.

This chapter describes the following capabilities of GDPS/XRC:

- Protecting your data
  - Protecting the integrity of the secondary data (both disk and tape) 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 non-z/OS data, coordinated with management of the z/OS data
- Controlling the GDPS-managed resources 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 via 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 would be acceptable for PPRC. Channel extender technology can be used to place the secondary disk 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 remote copy pairs and automates the process of recovering the production environment with limited manual intervention, including invocation of CBU[1], thus providing significant value in reducing the duration of the recovery window and requiring less operator interaction.

In order to get the best performance and availability from your z/OS Global Mirror configuration, it is very important to have a sound understanding of how z/OS Global Mirror works and how to tune it. This information can be obtained from the following documents:

- *DFSMS Extended Remote Copy Installation Planning Guide*, GC35-0481
- *DFSMS Extended Remote Copy Reference Information for Advanced Users*, GC35-0482
- XRC CPU Utilization, available from your IBM representative[2]

These documents are very readable, comprehensive guides that provide all the information you need to effectively plan and manage your configuration.

Whereas GDPS/PPRC is a high availability and disaster recovery solution within a single multi-site sysplex, GDPS/XRC is specifically an automated disaster recovery solution. GDPS/XRC controls the remote mirroring and automates disaster recovery for the primary site. The systems running GDPS/XRC are remote from the production systems and are not members of the sysplex at the primary site. Additionally, unlike GDPS/PPRC, GDPS/XRC has *no* knowledge of what is happening in the production systems. The only systems GDPS/XRC has any knowledge of are those 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 mechanism, it is not possible to have zero data loss when using XRC, so the Recovery Point Objective when using XRC must be more than zero. In a typical XRC configuration, an RPO of one minute should be achievable.

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 process of recovering 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

[1] Where available.

[2] This paper is available on the IBM internal network at: http://w3.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP100331

integrity of the secondary disk data. The role of GDPS, from a remote copy perspective, is to manage the remote copy configuration and to drive the recovery process.

Traditional System z data

There are two perspectives on XRC:

1. The systems whose data is being remote copied using XRC.
2. The system (or systems) where the SDM function will run.

The following systems support time stamping of I/Os when the target volume has been defined as a primary XRC volume:

- Any supported release of z/OS.
- System z for Linux drivers support time stamping of writes and also contain changes to support device blocking.
- z/VM guests with appropriate z/VM updates for time stamping.

Therefore, volumes used by any of these systems can be managed by GDPS/XRC. All the volumes will be managed to a single point of consistency, as long as they are managed by the same SDM or SDMs that are in the same Master session. For more information, refer to “XRC data consistency” on page 25.

It is also possible to use XRC to remote copy volumes being used by System z operating systems that do *not* time stamp their I/Os (for example, z/VM without timestamping code and z/VSE). However, in this case it is not possible to provide consistency across multiple LSSs. For more information, refer to the section entitled “Understanding the Importance of Timestamped Writes” in the latest revision of *z/OS DFSMS Advanced Copy Services,* SC35-0428.

Even though all these operating systems can use the primary volumes, z/OS is the only operating system that supports running the System Data Mover function. Therefore, in a GDPS/XRC configuration, you need a minimum of one z/OS system to provide the SDM function. And because the GDPS Controlling system should be dedicated for GDPS use, the z/OS system running the SDM should be separate from the GDPS controlling system.

5.1.2 Protecting tape data

Although most of your critical data will be resident on disk, it is possible that some of the data you require following a disaster will reside on tape. Just as you mirror your disk-resident data to protect it, equally you can mirror your tape-resident data. GDPS/XRC provides support for a single integrated recovery process when using the legacy IBM TotalStorage 3494-based Virtual Tape Subsystem in a Peer-to-Peer configuration. Additionally, while not specifically integrated into GDPS, the IBM Virtualization Engine TS7700 provides comprehensive support for replication of tape data. See *IBM Virtualization Engine TS7740 R1.5 and TS7720: New Virtualization Options for Mainframe Servers*, SG24-7712 for more information about the TS7700 technology that complements GDPS for tape data.

5.2 GDPS/XRC configuration

At its most basic, a GDPS/XRC configuration consists of a single production system updating the primary volumes in the production site, one SDM system in the recovery site, and one GDPS Controlling system (K-sys), also in the recovery site. The SDM system and the Controlling system must both be in the same sysplex. There is no requirement for the production system to be in a sysplex; however, all the systems updating the primary volumes must be connected to the same Sysplex Timers or use Server Time Protocol (STP). Figure 5-1 shows a simplified illustration of the physical topology of a GDPS/XRC implementation.

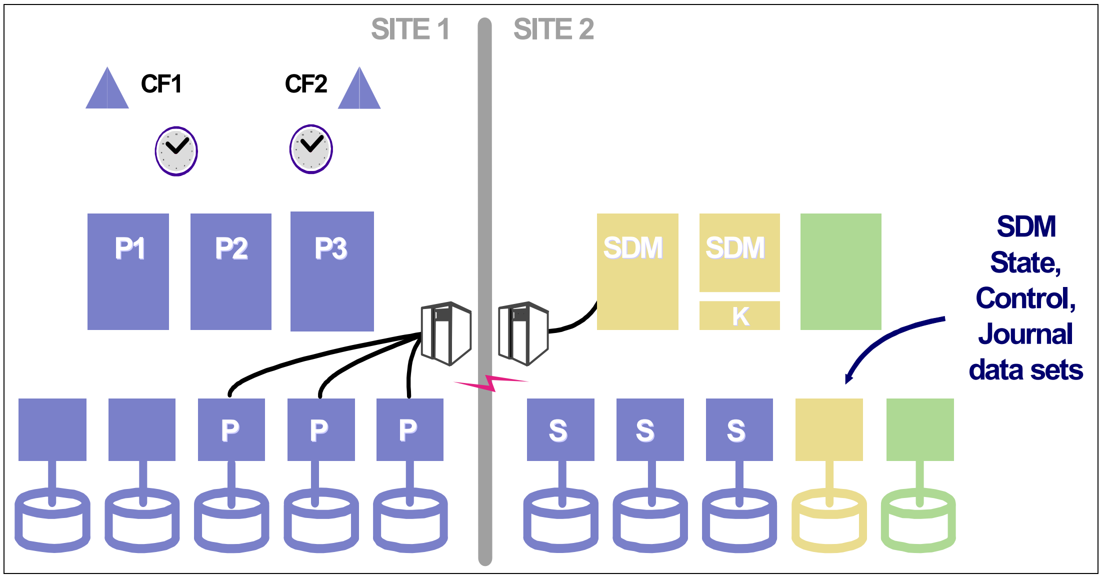


Figure 5-1 GDPS/XRC - topology

Just as for GDPS/PPRC, 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. However, unlike a GDPS/PPRC configuration, it is not necessary to have the controlling system disks completely isolated from the other systems in the GDPS sysplex (the SDM systems).

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, with most people using channel extension technology to provide this connectivity. Unlike GDPS/PPRC, there is no requirement for dark fiber between the sites.

In a more complex configuration, where you have more primary volumes, you might exploit 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 could 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 three-site configuration

GDPS/XRC can be combined with GDPS/PPRC in a three-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 exploitation 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 remote copy technologies for CA and DR” on page 30, refer to Chapter 8, “Combining Local/Metro continuous availability with out-of-region disaster recovery” on page 173 for a more detailed discussion of 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 disks (this is managed by VCS); therefore, it is not possible to have a common consistency point between the System z CKD data and the distributed data. However, for environments where a common consistency point is not a requirement, DCM together with VCS does provide some key availability and recovery capabilities which may be of interest. DCM is discussed in further detail in 7.3.2, “DCM support for VCS” on page 154.

5.4 Managing the GDPS environment

It is important to understand that the world, as GDPS/XRC sees it, only consists of those systems in the SDM sysplex in the recovery site. This is shown in Figure 5-2. The GDPS Controlling system can only see the systems in the recovery site. If all the systems in the production site were to go down, GDPS/XRC has no automatic knowledge of this event. However, GDPS/XRC is aware of all the resources in the recovery site that you defined to it, and has the ability to monitor and manage those resources.

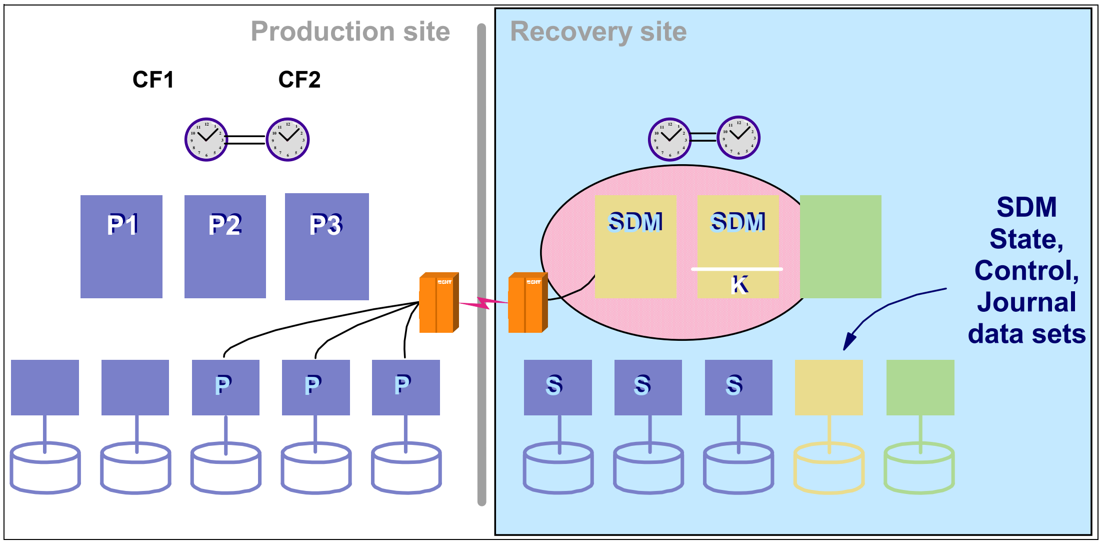
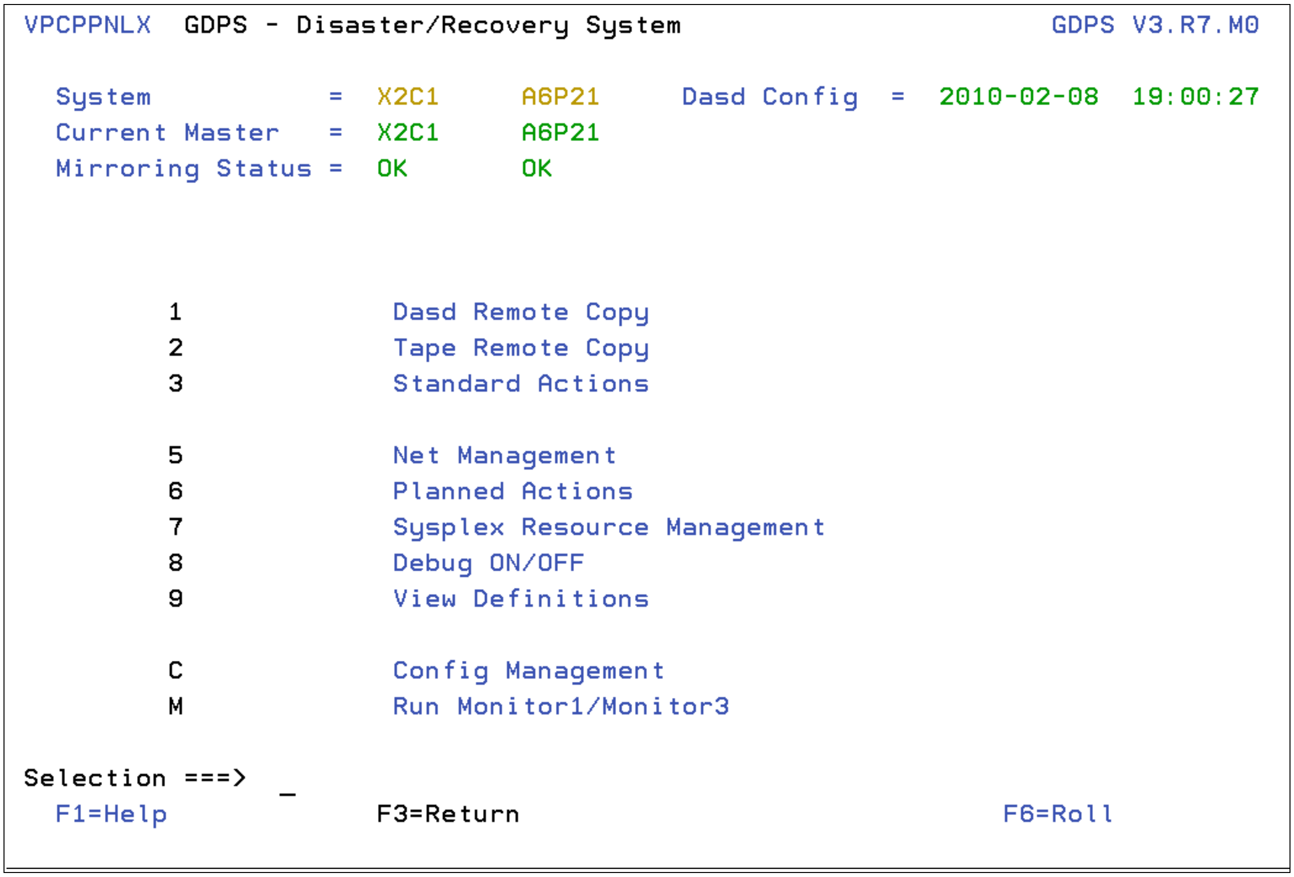


Figure 5-2 GDPS/XRC span of control

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 is comprised 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.R7.M0

  System          =  X2C1     A6P21     Dasd Config  =  2010-02-08  19:00:27
  Current Master  =  X2C1     A6P21
  Mirroring Status =  OK       OK




         1              Dasd Remote Copy
         2              Tape Remote Copy
         3              Standard Actions

         5              Net Management
         6              Planned Actions
         7              Sysplex Resource Management
         8              Debug ON/OFF
         9              View Definitions

         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:

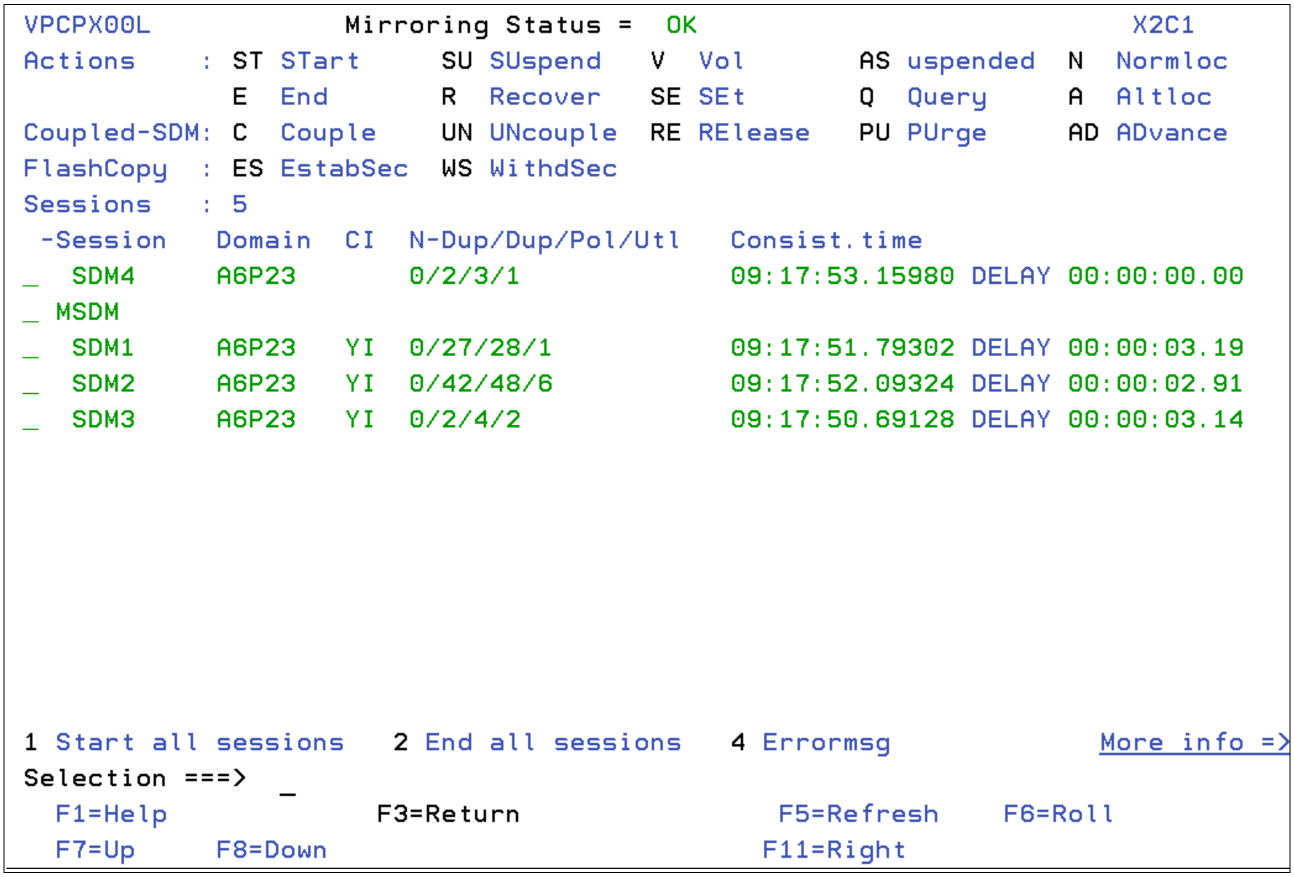
- Query and control the disk and tape remote copy configuration
- Initiate GDPS-provided standard actions against GDPS-managed LPARs (like IPL, LPAR Deactivate, and so on)
- Initiate GDPS scripts that you create
- Manage Coupling Facilities and Couple Data Sets relating to the SDM sysplex
- Refresh the definitions of the remote copy configuration

Remote copy panels

Although z/OS Global Mirror (XRC) provides powerful mirroring capabilities, the operator interface to it is not particularly user-friendly. To make it easier for operators to check and manage the remote copy environment, you can (and should) use the GDPS-provided Disk Remote Copy panels.

In order for GDPS to manage the remote copy environment, you must first define the configuration (primary and secondary volsers or device numbers, and information about the 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 desired one. You can start, stop, suspend, and resynch mirroring at the volume or LSS level, and you can initiate a FlashCopy of the secondary volumes. These actions can be carried out at the device or LSS level, or both, as appropriate. Figure 5-4 shows the mirroring panel for GDPS/XRC.



```
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
Sessions   : 5
 -Session   Domain  CI  N-Dup/Dup/Pol/Utl   Consist.time
_  SDM4     A6P23       0/2/3/1             09:17:53.15980 DELAY 00:00:00.00
_ MSDM
_  SDM1     A6P23   YI  0/27/28/1           09:17:51.79302 DELAY 00:00:03.19
_  SDM2     A6P23   YI  0/42/48/6           09:17:52.09324 DELAY 00:00:02.91
_  SDM3     A6P23   YI  0/2/4/2             09:17:50.69128 DELAY 00:00:03.14







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 Disk Mirroring panel for GDPS/XRC

If you are familiar with using the TSO interface to XRC, you will appreciate how much more user friendly the panel is.

Remember that these GDPS-provided panels are *not* intended to be a remote copy monitoring tool. Because of the overhead involved in gathering the information to populate the NetView panels, GDPS only gathers this information 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

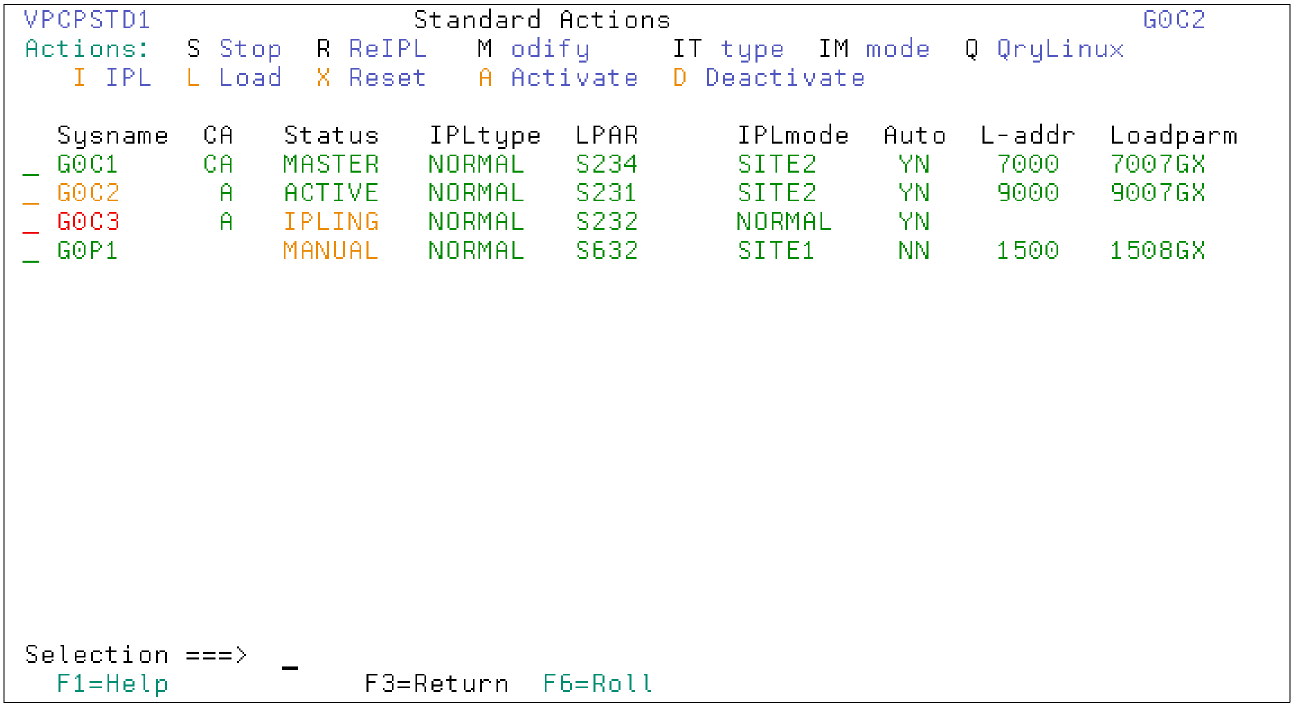
You will recall that we previously discussed how the overwhelming majority of System z outages are planned outages. Even though GDPS/XRC only manages the SDM systems in the recovery site, it is still important that those systems are available and are correctly managed. GDPS provides facilities to help manage any outages affecting these systems. There are two reasons to use the GDPS facilities:

- They are well-tested and based on IBM recommended procedures.

- Using the GDPS interface lets GDPS know that the changes that it is seeing (CDSs being deallocated or systems going out of the sysplex, for example) are planned changes, and therefore it should not 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-step, or are intended to impact just one resource. Examples are IPLing a system, updating the IPL address or the Loadparm to be used the next time a system is IPLed, or activating an LPAR. So, if you wanted to stop a system, change its IPL address, then IPL it, that would be three separate Standard Actions that you would 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 carried out on each system, including Stop, re-IPL (stop followed by IPL), Activate, Deactivate, and so on.



```
VPCPSTD1                    Standard Actions                            G0C2
Actions:  S Stop  R ReIPL   M odify     IT type  IM mode  Q QryLinux
   I IPL  L Load  X Reset   A Activate  D Deactivate

  Sysname  CA   Status   IPLtype  LPAR      IPLmode  Auto  L-addr  Loadparm
_ G0C1     CA   MASTER   NORMAL   S234      SITE2    YN     7000   7007GX
_ G0C2      A   ACTIVE   NORMAL   S231      SITE2    YN     9000   9007GX
_ G0C3      A   IPLING   NORMAL   S232      NORMAL   YN
_ G0P1          MANUAL   NORMAL   S632      SITE1    NN     1500   1508GX









Selection ===> _
  F1=Help          F3=Return  F6=Roll
```

Figure 5-5 GDPS/XRC Standard Actions panel

GDPS scripts

All of the functions that can be initiated via the panels (and more) are also accessible from GDPS scripts. A script is a “program” consisting of one or more GDPS functions. Scripts can be initiated manually through the GDPS panels (Planned Actions), automatically by GDPS in response to an event (referred to as an Unplanned Action), or through a batch job.

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 very powerful because they can access the full capability of GDPS. The ability to invoke all the GDPS functions through a script provides:

- 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 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 System 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 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 112). A Planned Action script might consist of a number of tasks. For example, you could have a script that would stop an LPAR, change its IPL address to the alternate SYSRES, and then re-IPL it, all from a single script.

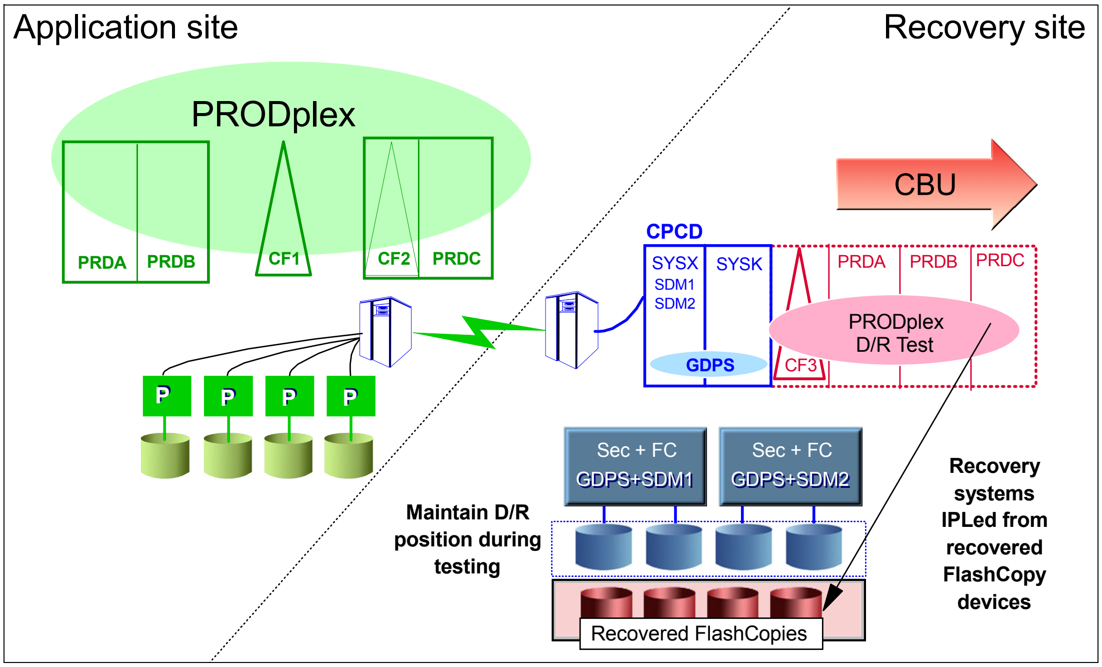


Figure 5-6 GDPS/XRC Planned Action

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 procedures while XRC sessions between the primary to the secondary volumes is running, and you maintain up-to-date disaster readiness. Specifically, the following actions are carried out by GDPS in this script:

- Zero Suspend FlashCopy is initiated:
  - This prevents the SDMs from writing new consistency groups to the secondary disks.
  - A FlashCopy is taken of all XRC secondary devices, as well as the devices that house 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 run.
- The CBU capacity reserved PUs on CPCD is activated.
- Any test systems whose LPARs would be used for a production system in case of a disaster are deactivated.
- The CF LPARs and the LPARs that will contain the recovered production systems are activated, followed by an IPL of those systems.

So, the result of a single action on your behalf (initiating the Planned Action) is that you have created a copy of your production system that can be used for DR testing, you have brought up recovery versions of your production systems, and you have maintained disaster readiness by still having synchronization between the primary and secondary volumes.

The use of this capability removes the reliance on out-of-date documentation, provides a single repository for information about IPL addresses and Loadparms, and ensures that the process is carried out the same way every time, with no vital steps accidentally overlooked.

Unplanned Actions

Unplanned Actions are GDPS scripts, just like Planned Actions. However, they are used in a different way. While Planned Actions are initiated from the GDPS panels, Unplanned Actions are initiated by GDPS in response to a predefined event.

You will recall that in a GDPS/XRC environment, GDPS only has knowledge about what is happening in the SDM sysplex. As a result, the recovery script for a disaster in the production site would actually be defined as a Planned Action, because GDPS will not see any messages indicating that the production systems have failed.

In GDPS/XRC, Unplanned Actions are only used to react to failures in the SDM systems 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).

Batch scripts

Because the full range of GDPS functions is available to you, you can have scripts that will carry out 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 customer with high availability requirements is creating a set of consistent tape backups. Backing up tens of terabytes to tape would involve stopping the applications for many minutes - time that is not available in most installations. However, using a combination of GDPS batch scripts and FlashCopy, you *can* achieve this.

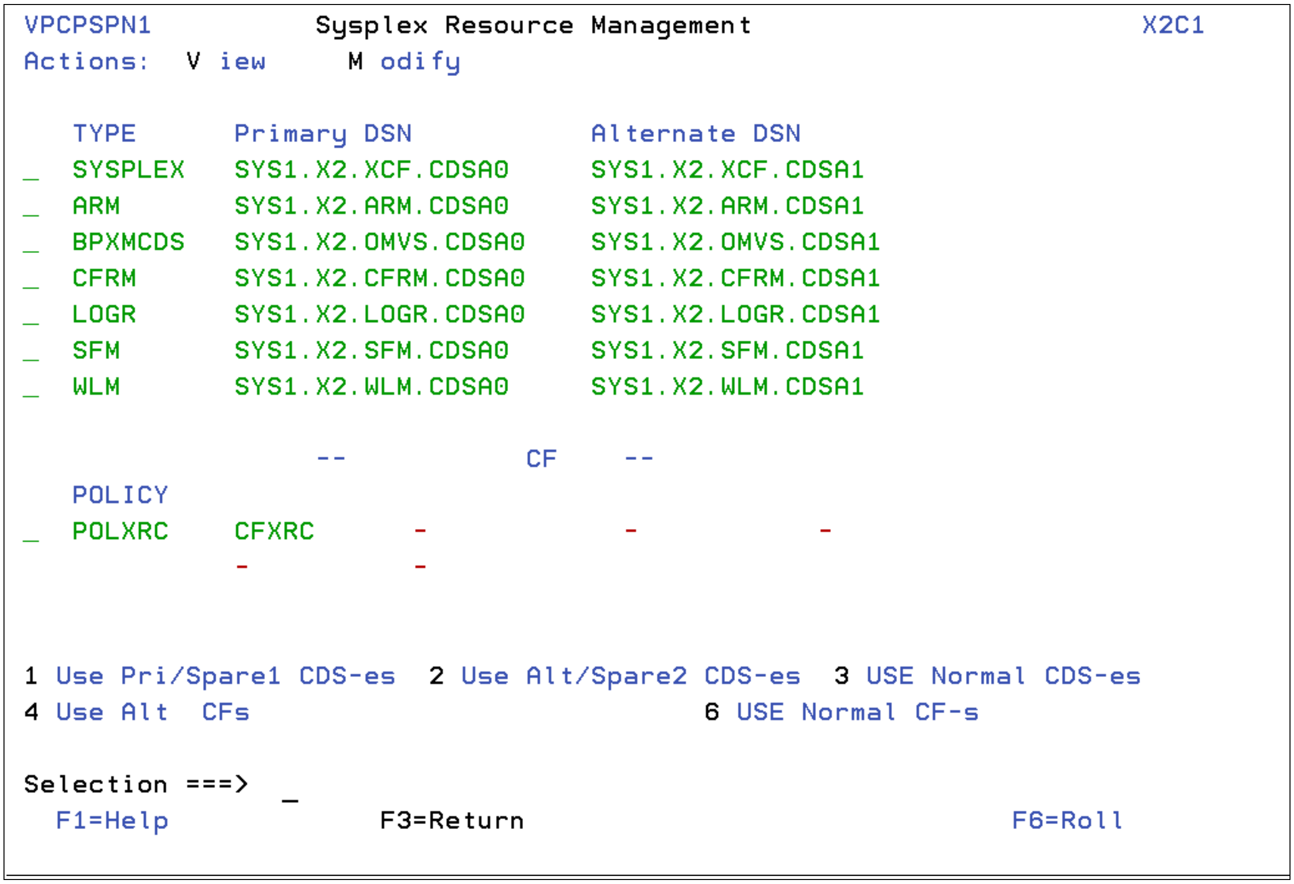
Just as you can have a Planned Action to create a set of tertiary volumes for a DR test, you could have a similar script that would create the tertiary volumes, and then take tape backups of those volumes. The net effect is basically the same as if you had stopped all work in your primary site for the 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 were to switch to a new Primary sysplex CDS using the SETXCF PSWITCH command, you would 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 would include adding a new alternate once the switch of the primary had completed successfully, thereby ensuring that you do not have a single point of failure in the CDS configuration.


```
VPCPSPN1         Sysplex Resource Management                        X2C1
Actions:   V iew     M odify

   TYPE      Primary DSN           Alternate DSN
_  SYSPLEX   SYS1.X2.XCF.CDSA0     SYS1.X2.XCF.CDSA1
_  ARM       SYS1.X2.ARM.CDSA0     SYS1.X2.ARM.CDSA1
_  BPXMCDS   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    CFXRC      -            -           -
             -          -


1 Use Pri/Spare1 CDS-es  2 Use Alt/Spare2 CDS-es  3 USE Normal CDS-es
4 Use Alt  CFs                           6 USE Normal CF-s

Selection ===>  _
  F1=Help            F3=Return                                F6=Roll
```

Figure 5-7 GDPS/XRC Sysplex Resource Management panels

Although it may not receive as much attention as recovering from a disaster, the capability of GDPS to perform Planned Actions is used far more frequently, and provides tremendous value in terms of faster turnaround and mistake avoidance.

5.5 GDPS/XRC monitoring and alerting

The GDPS SDF panel, which is discussed in 5.4.1, “NetView interface” on page 111, 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 monitoring items are found to be in a state deemed to be not normal by GDPS, an alert is posted on SDF. GDPS 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 GDPS is the BCP Internal Interface connection to all CPCs in the recovery site on which the GDPS could 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). Any alerts generated on any of these systems is 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 would have to investigate (or escalate, as appropriate) and corrective action would 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/XRC 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 impact the ability of GDPS/XRC to carry out recovery operations. This will maximize the installation’s chance of achieving its RPO and RTO commitments.

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 GDPS-related settings adhere to GDPS best practices.

The z/OS Health Checker infrastructure is intended to check a variety of settings to see whether these settings adhere to z/OS best practices values. For settings that are found to be not in line with best 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 GDPS-related parameter settings such as z/OS PARMLIB settings or NetView settings, and we document the recommendations and best practices for these settings in the GDPS publications. If these settings do not adhere to recommendations, this could hamper the ability of GDPS to perform critical functions in a timely manner.

Although GDPS monitoring would detect that GDPS was not able to perform a particular task and raise an alert, the monitor alert may be too late, at least for that particular instance of an incident. Often, if there are changes in the customer environment, this might necessitate adjustment of some parameter settings associated with z/OS, GDPS, and other products. It is possible that you could miss making these adjustments, which may 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 due to a setting that is perhaps less than ideal.

For example, there are a number of address spaces associated with GDPS/XRC and best 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. Additionally, a number of the best practices values that are checked, as well as the frequency of the checks, are customer-customizable to cater to unique customer environments and requirements.

5.6 Other GDPS-related facilities

In this section we describe miscellaneous facilities provided by GDPS/XRC that may assist in various different ways, such as reducing the window when DR capability is not available.

5.6.1 z/OS asynchronous system logger support

Beginning with z/OS V1.7, system logger provides new support for XRC that allows you to choose asynchronous writes to staging data sets for logstreams. Previously, all writes had to be synchronous. This limited the throughput for high-volume logging applications such as WebSphere, CICS, and IMS. The ability to perform asynchronous writes can allow the use of z/OS Global Mirror (XRC) for some applications for which it was not previously practical.

GDPS/XRC has extended its automation for XRC to provide the ability to configure and manage the staging data set remote copy pairs.

5.7 Flexible testing

By combining z/OS Global Mirror with FlashCopy, you can create a consistent point-in-time tertiary copy of the XRC data sets and secondary disks at your recovery site. The tertiary devices can then be used to test your disaster recovery procedures, while the GDPS/XRC sessions between production and the recovery site are running, thus ensuring disaster readiness is maintained at all times. In addition, these devices can be used for purposes other than DR testing, nondisruptive data backup, data mining, or application testing, for example.

With the addition of GDPS/XRC Zero Suspend FlashCopy, enterprises are now able to create the tertiary copy of the XRC data sets and secondary disks without having to suspend the 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 SSIDs involved and the largest number of devices in any SSID and the speed of the processor.

Zero Suspend FlashCopy will normally be executed on the GDPS controlling system in the recovery site, where there would likely 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 if this facility can be used in your environment.

If you have requirements to test the recovery capabilities and maintain the currency of the replication environment you will need to provide additional disk capacity to support FlashCopy. By providing this additional usable copy of the data, you now have the flexibility to perform on-demand DR testing and other nondisruptive activities, while maintaining current DR readiness.

5.7.1 Usage of FlashCopy Space Efficient

As discussed in “FlashCopy Space Efficient (FlashCopy SE)” on page 33, by using FlashCopy 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. Support is added to allow FlashCopy Space Efficient 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, panels, or FlashCopies automatically taken by GDPS, can use Space Efficient targets.

You should have an understanding of some of the characteristics of Space Efficient FlashCopy to determine if this method of creating a point in time copy will satisfy your business requirements. For example, will it be acceptable to your business if, due to some unexpected workload condition, the repository on the disk subsystem for the Space Efficient devices gets full and your FlashCopy is invalidated such 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 could consider using standard FlashCopy instead.

5.8 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 would be 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:

- 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.
- Peer-to-Peer VTS 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.

5.9 GDPS/XRC prerequisites

Important: For the latest GDPS/XRC prerequisite information, refer to the GDPS Web site:

`http://www-03.ibm.com/systems/z/advantages/gdps/getstarted/gdpsxrc.html`

5.10 Comparison of GDPS/XRC versus other GDPS offerings

There are so many features and functions available in the various members of the GDPS family that it is sometimes difficult to recall them all, and remember which offerings support them. 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/XRC	GDPS/GM
Continuous availability	Yes	Yes	No	No
Disaster recovery	Yes	Yes	Yes	Yes
Supported distance	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	Virtually unlimited	Virtually unlimited
FlashCopy support	Yes	Yes	Yes	Yes
Reduced impact initial copy/resynch	Yes	Yes	N/A	N/A
PtP VTS support	Yes	No	Yes	No
Production sysplex automation	Yes	No	No	No
Span of control	Both sites	Both sites (disk only)	Recovery site	Disk at both sites. Recovery Site (CBU / LPARs)
GDPS scripting	Yes	No	Yes	Yes

Feature	GDPS/PPRC	GDPS/PPRC HM	GDPS/XRC	GDPS/GM
Monitoring, alerting, and health checks	Yes	Yes	Yes	Yes
Query Services	Yes	Yes	No	No
MGM	Yes (IR or non-IR)	Yes (Non-IR only)	N/A	Yes (IR or non-IR)
MzGM	Yes	Yes	Yes	N/A
Open LUN	Yes	Yes	No	Yes
z/OS equivalent functionality for Linux for System z	Yes	No	Yes	Yes
Heterogeneous support via DCM	Yes (VCS and SA AppMan)	No	Yes (VCS only)	Yes (VCS only)
Web interface	Yes	Yes	No	Yes

5.11 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 customers 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.

6

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.
  - Management of 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 Architecture (FBA) data. Depending on your application requirements, the consistency of the FBA data can be coordinated with the CKD data.
- Controlling the GDPS-managed disk resources 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 some respects to GDPS/XRC in that it supports virtually unlimited distances. However, the underlying IBM Global Mirror (GM), remote copy technology also supports FBA and CKD devices in a consistency group so GDPS/GM includes extensions to support that.

GDPS/GM could be viewed somewhat 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, meaning that it supports the same mix of CKD and FBA 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, and, like PPRC, Global Mirror requires that the primary and secondary disk subsystems are from the same vendor.

On the other hand, GDPS/GM resembles GDPS/XRC in that it supports virtually unlimited distances between the application and recovery sites. Also, 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 sysplexes; in contrast, each GDPS/PPRC installation supports remote copy for just 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 in the recovery site. The early releases of GDPS/GM only managed z/OS disk-resident data. Based on customer requirements and enhancements to the underlying technology, this support has been expanded to cover non-z/OS data as well. This section discusses the support provided by GDPS for these various data types.

Traditional System z (CKD) data

As described in 2.4.3, “Global Mirror” on page 28, Global Mirror protects the integrity of the remote copied data by creating consistency groups at intervals specified by the installation. The whole process is managed by the master disk subsystem, using information provided by GDPS/GM. As a result, there are no restrictions relating to which operating systems are 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 is from the z/OS system running 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.

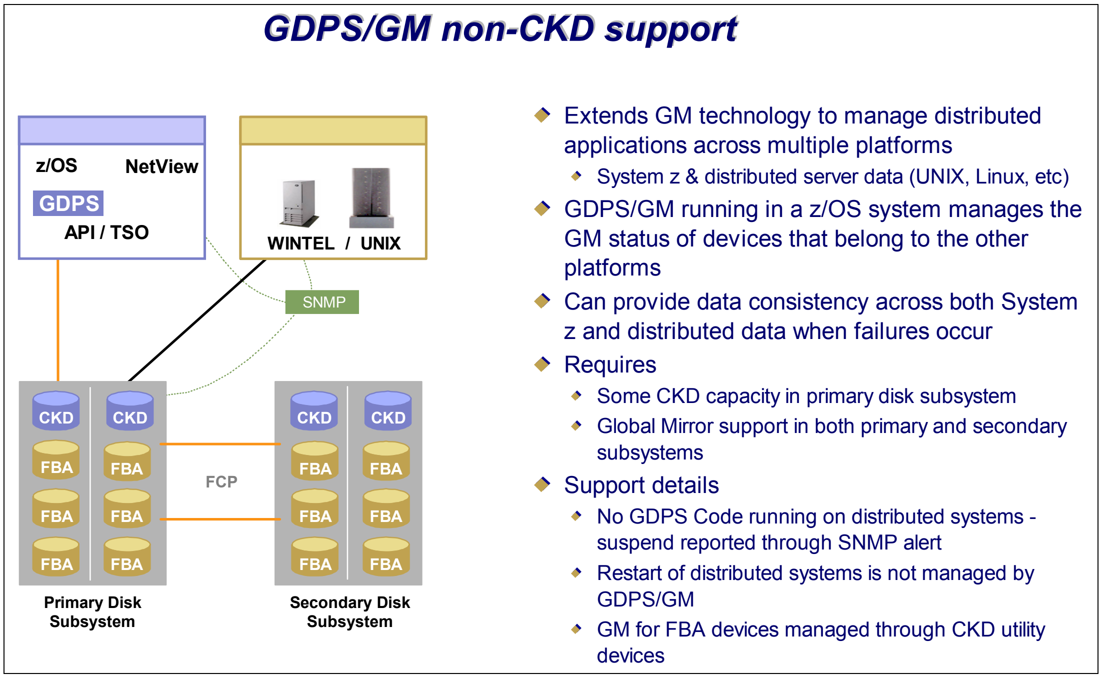
GDPS/Global Mirror uses devices in the primary and secondary disk subsystems to execute the commands to manage the environment. You must designate at least one volume in each primary LSS as a GDPS *utility device* for mirroring of CKD devices. These utility devices do not need to be dedicated devices; that is, they could be one of the devices that are being mirrored as part of your Global Mirror session. The utility devices should themselves be mirrored.

Global Mirror supports both CKD and FBA devices. If the CKD and FBA devices are in the same Global Mirror session, they will be in the same consistency group.

Distributed (FBA) data

GDPS/GM supports remote copy of FBA devices (also known as Open LUNs) written by distributed systems (including SCSI disks written by Linux for System z). If the FBA devices are in the same disk subsystem 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 (even if the consistency group interval is the same on both sessions).

Every disk subsystem cluster that contains FBA devices that are remote copied using Global Mirror must contain at least one CKD device that will be used as the utility device by GDPS/GM. If you are remote copying a large volume of FBA data, more than one utility device may be required to provide the required levels of performance. A sample configuration is shown in Figure 6-1.



GDPS/GM non-CKD support

- Extends GM technology to manage distributed applications across multiple platforms
  - System z & distributed server data (UNIX, Linux, etc)
- GDPS/GM running in a z/OS system manages the GM status of devices that belong to the other platforms
- Can provide data consistency across both System z and distributed data when failures occur
- Requires
  - Some CKD capacity in primary disk subsystem
  - Global Mirror support in both primary and secondary subsystems
- Support details
  - No GDPS Code running on distributed systems - suspend reported through SNMP alert
  - Restart of distributed systems is not managed by GDPS/GM
  - GM for FBA devices managed through CKD utility devices

Figure 6-1 GDPS/GM FBA support

Note that this support is limited to protecting the FBA *data*. There is no automation or system restart capability for the *systems* that use those devices.

6.2 GDPS/Global Mirror configuration

At its most basic, a GDPS/GM configuration consists of at least one production system, 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.

Note: Global Mirror uses three sets of disks:

- Primary disks in the application site - we refer to those as A disks in this document
- Secondary disks in the recovery site - we refer to those as B disks in this document
- FlashCopy targets in the recovery site - we refer to those as C disks in this document

The GDPS/GM control file (called GEOMPARM) also refers to the disks using this designation.

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 via the K-sys. The K-sys role is purely 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 any of the system(s) it is managing data for - in fact, it is recommended that the K-sys be placed in a monoplex on its own. It can, however, share infrastructure resources (sysres, catalogs, and so on) with members of a production sysplex. You can even remote copy the K-sys disks if you want to - it does not have the isolation requirements of the Controlling system in a GDPS/PPRC configuration.

The R-sys is primarily responsible for carrying out all recovery actions following a disaster. The R-sys is also responsible for carrying out the actions required to reconfigure the recovery site disks. These disks can then be used in the event of a real disaster or for DR testing. Refer to 6.6, “Flexible testing” on page 139 for more details. The R-sys can also be used to query the remote copy configuration.

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).

Both the K-sys and R-sys should be 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. If you wish to have a different consistency point for different sets of disks (for example for mainframe and Open disks) you need to have two (or more) Global Mirror sessions.

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 are defined in the GDPS remote copy configuration 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.

However, although the panel interface or web interface would support global operations for all disks in a session, they are primarily intended for viewing the configuration and performing operations against single disks (such as adding a number of new A/B/C devices to the configuration). 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 FBA), and your RPO.

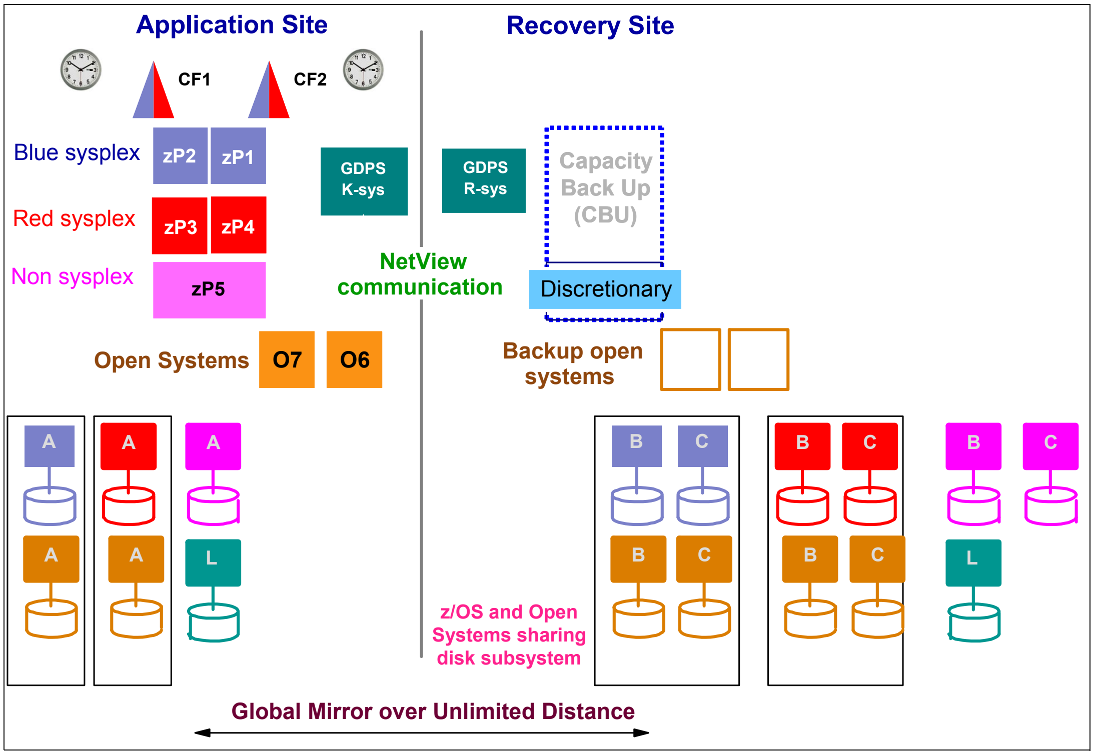


Figure 6-2 GDPS/GM configuration

Figure 6-2 shows a typical GDPS/GM configuration. The application site contains:

- z/OS systems spread across a number of sysplexes
- A non-sysplexed system
- Two distributed systems
- The K-sys
- The primary disks (identified by A)
- The K-sys disks (marked by L)

The recovery site contains:

- 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, some things are fixed:

- All the GM primary disks and the K-sys *must* be in the application site[1].
- All the GM secondary disks, the FlashCopy targets used by GM, and the GDPS R-sys *must* be in the recovery site[2].

[1] The application site is the site 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 may also see this site referred to as the “local site” or the “A-site.”

Some aspects that are different in GDPS/GM as compared to the other GDPS offerings are:

- 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, due to 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 be further apart, meaning that the RPO may 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 *only* run in the K-sys and R-sys.

6.2.1 GDPS/GM in a three-site configuration

GDPS/GM can be combined with GDPS/PPRC (or GDPS/HM) in a three-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 exploitation and GDPS HyperSwap, and GDPS/GM provides disaster recovery in a remote site.

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.

After you understand the base capabilities described in 2.4.4, “Combining remote copy technologies for CA and DR” on page 30, refer to Chapter 8, “Combining Local/Metro continuous availability with out-of-region disaster recovery” on page 173 for a more detailed discussion of 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 may not have the latest configuration information in the event of a subsequent disaster, resulting in an impact to the recovery operation.

[2] The recovery site is the site where the mirrored copy of the production disks are located, and it is the site into which production systems are failed over to in the event of a disaster. You may also see this site referred to as the “remote site” or the “R-site.”

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 would, again, impact the remote copy or recovery operation.

Because GDPS/GM is really a disaster recovery offering, rather than a continuous availability offering, it does not support the concept of site switches that GDPS/PPRC provides. It is expected that a switch to the recovery site will only be performed 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 to act as B and C disks in the application site), or you would use an alternate mechanism, like Global Copy, outside the control of GDPS.

6.3 GDPS/GM management for distributed systems and data

As previously mentioned, it is possible for GDPS/GM to manage FBA disks on behalf of distributed systems that use these disks, either in the same session as System z 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), and therefore it is not possible to have a common consistency point between the System z CKD data and the distributed data.

However, for environments where a common consistency point is not a requirement, DCM together with VCS does provide some key availability and recovery capabilities that may be of interest. DCM is discussed in further detail in 7.3.2, “DCM support for VCS” on page 154.

6.4 Managing the GDPS environment

Like GDPS/PPRC, GDPS/GM runs in both the application and recovery sites. However, GDPS/GM only runs in one system in the application site (K-sys) and it does not provide any management of the production systems. The K-sys has the following main roles:

- It is the point of GDPS/GM control for operators and system programmers in normal operations. Changes to the remote copy definitions, or to the definitions or scripts to be used at the recovery site, are defined in the K-sys and automatically propagated to the R-sys.
- 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 activated on the K-sys.

In the recovery site, GDPS/GM only runs in one system: the R-sys. However, the role and capabilities of the R-sys are very different from those of the K-sys. Even though both are GDPS Controlling systems, there are fundamental differences between them.

The R-sys is mainly used to drive recovery following a failure, or in preparation for a disaster recovery test. The R-sys can prepare the disks in the recovery site for use, and is used to perform configuration changes (such as activating CBU and backup LPARs) to bring up the production systems in the recovery site. It does not do anything with any resources in the application site.

This section describes the capabilities of both systems and how they are used to control the GDPS/GM environment.

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 131), which is also based on NetView facilities. In normal operations, the operators would interact mainly with the K-sys while there is also a similar set of interfaces for the R-sys.

The NetView interface for GDPS actually consists of two parts. The first, and potentially the most important, is the Status Display Facility (SDF). Any time there is a change of status to something that GDPS does not consider as *normal* and can impact the ability to recover, or something in GDPS that requires manual intervention, GDPS will send an alert to SDF.

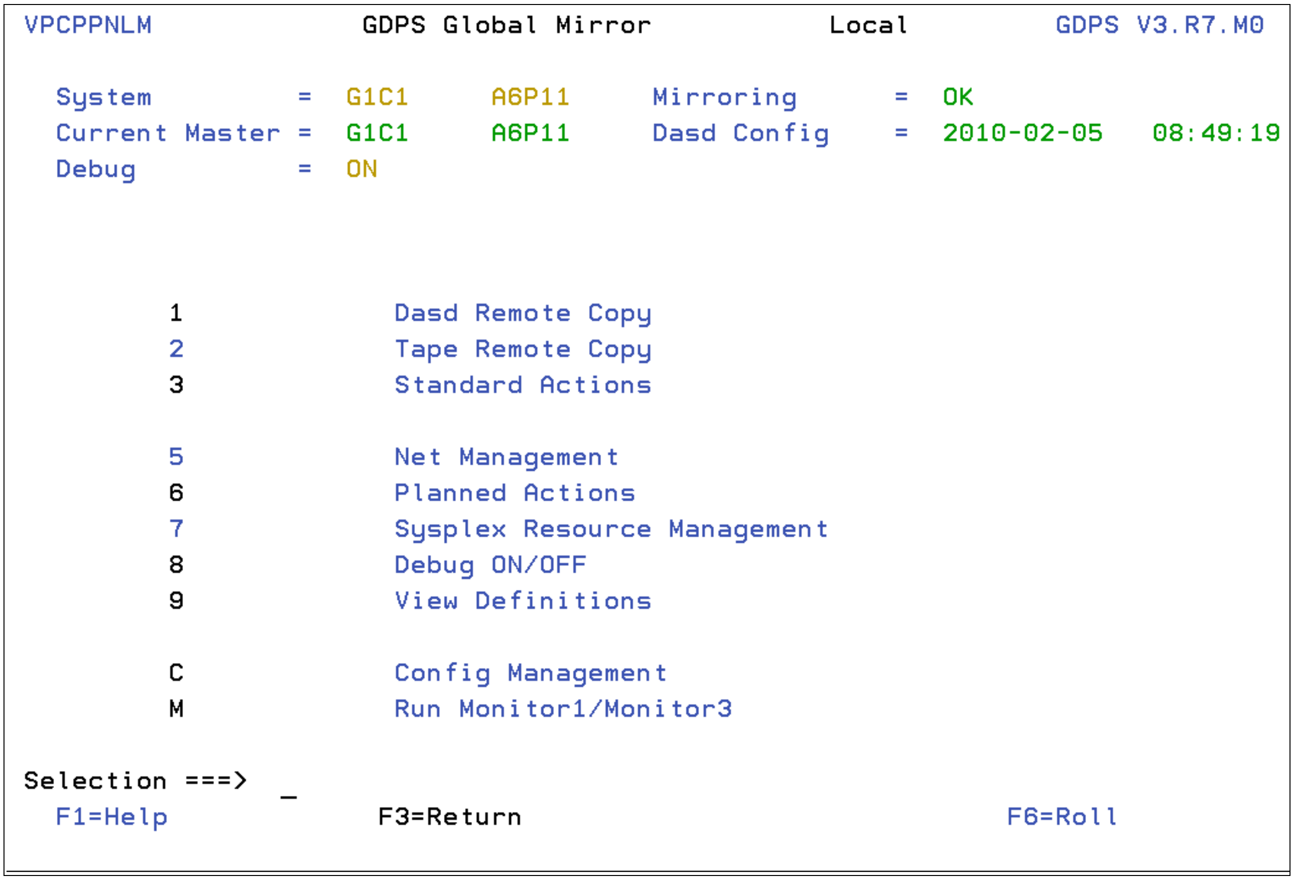
SDF provides a dynamically updated color-coded panel that provides the status of the systems and highlights any problems in the remote copy configuration. If something changes in the environment that requires attention, the color of the associated field on the panel will change. K-sys sends alerts to 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 as well.

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 6-3 on page 131. You will see that some of the options are not enabled (options 2, 5, and 7 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-provided 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).
- View and refresh the definitions of the remote copy configuration.



```
VPCPPNLM                GDPS Global Mirror           Local         GDPS V3.R7.M0

  System         =  G1C1     A6P11     Mirroring     =  OK
  Current Master =  G1C1     A6P11     Dasd Config   =  2010-02-05   08:49:19
  Debug          =  ON



         1             Dasd Remote Copy
         2             Tape Remote Copy
         3             Standard Actions

         5             Net Management
         6             Planned Actions
         7             Sysplex Resource Management
         8             Debug ON/OFF
         9             View Definitions

         C             Config Management
         M             Run Monitor1/Monitor3

Selection ===>
  F1=Help           _     F3=Return                          F6=Roll
```

Figure 6-3 Main GDPS/GM panel

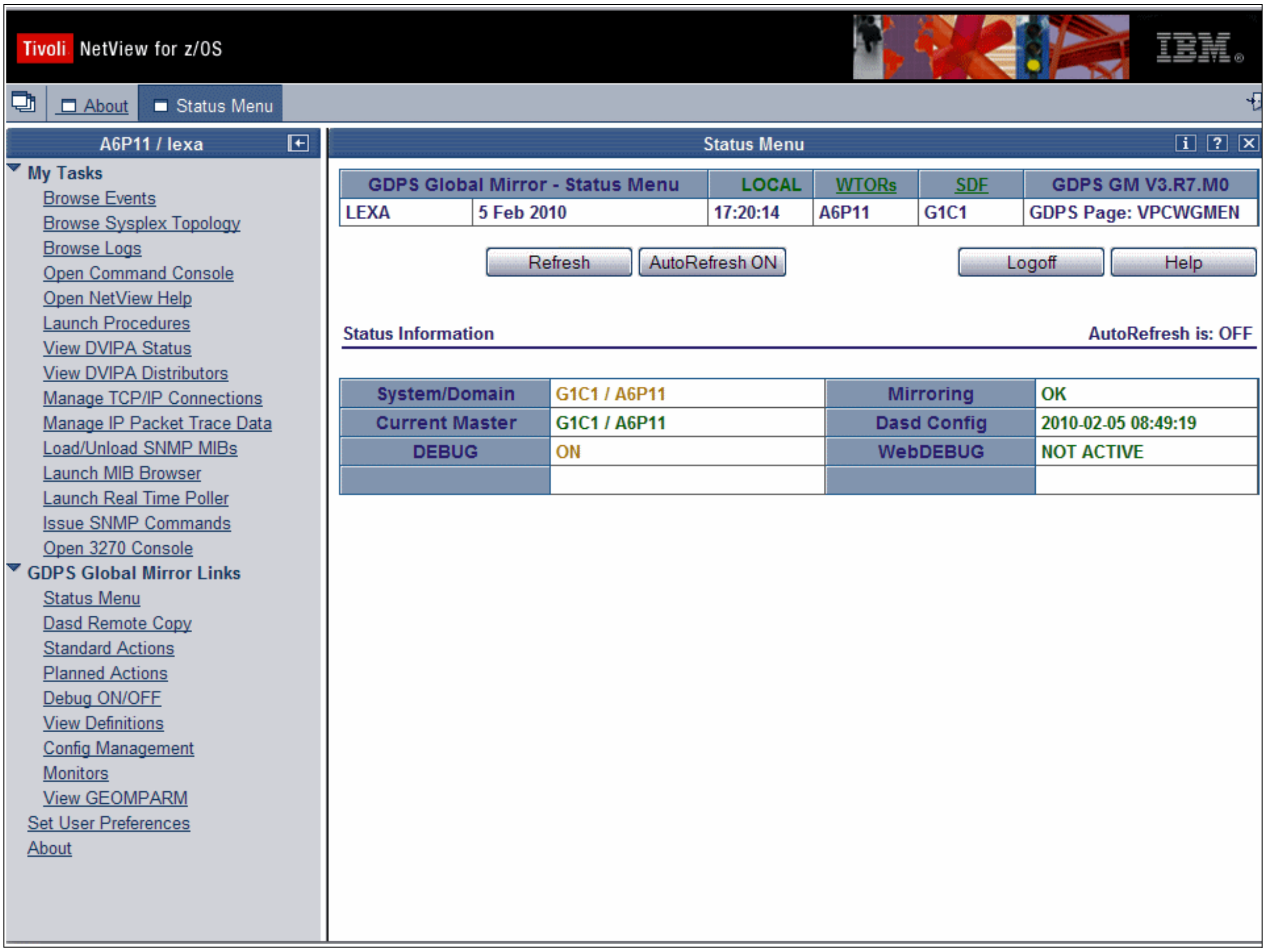
Web graphical user interface

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 is split into 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 screen of the GDPS/GM web interface is shown in Figure 6-4 on page 132. The left frame, shown below `GDPS Global Mirror links`, allows you to select the menu options. These options can be displayed at all times, or you can optionally collapse the frame.



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

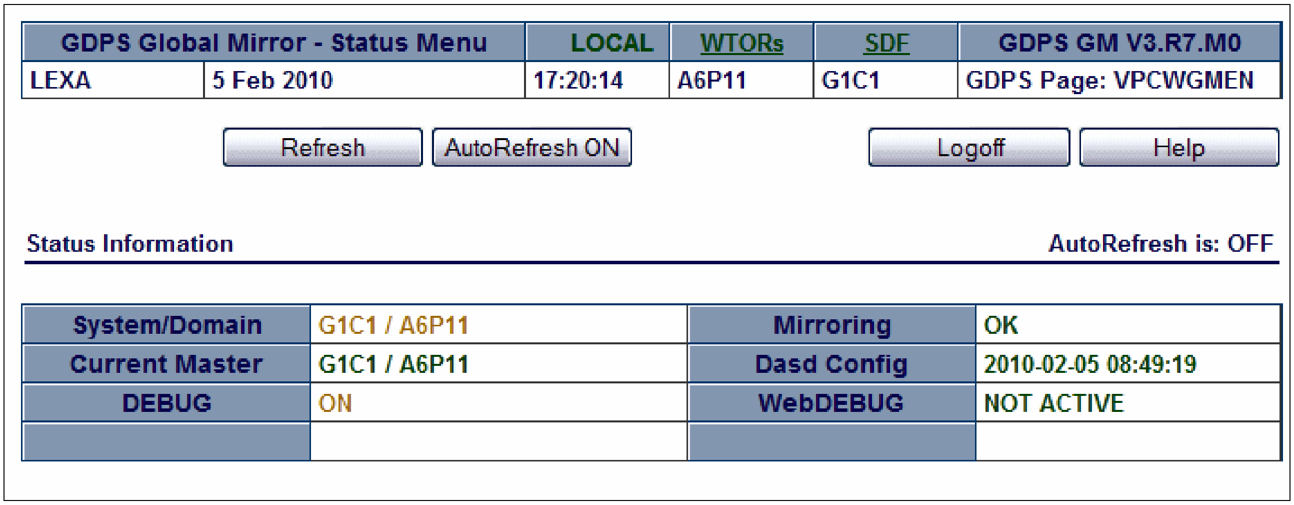
Status Information AutoRefresh is: OFF

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 Full view of GDPS main panel with task bar and status information

Main Status panel

The GDPS Web interface status frame shown in Figure 6-5 on page 133 is the equivalent to the main GDPS panel. The information 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 6-5 GDPS Web interface - Main Status panel

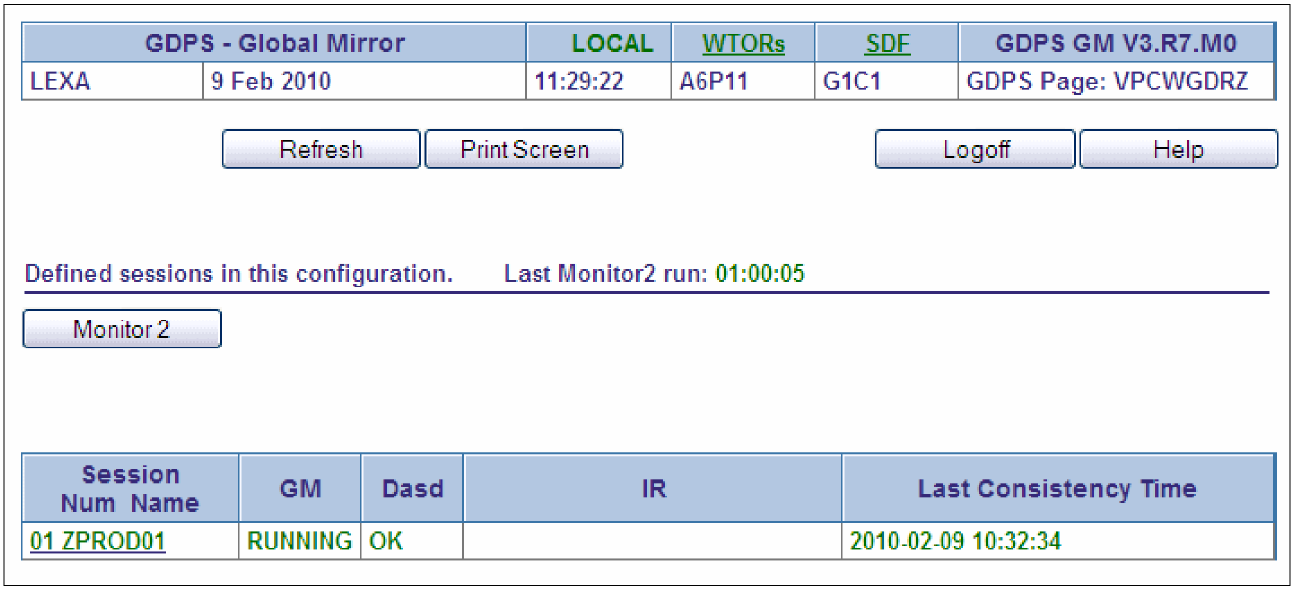
Remote copy panels

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, you can (and should) use the GDPS-provided Disk Remote Copy panels.

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 gets the configuration information from the K-sys; no `GEOMPARM` file is specified on the R-sys.

After the configuration is known to GDPS, you can use the panels to check that the current configuration matches the desired one. You can start, stop, pause, and resynch mirroring. These actions can be carried out at the device, LSS, or session level, as appropriate. However, we recommend that GDPS control scripts are used for actions at the session level.

Figure 6-6 on page 134 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 only option provided is to View the configuration. Control of the end-to-end remote copy environment can only be carried out from the K-sys; the R-sys can only control the devices in the recovery site.



GDPS - Global Mirror		LOCAL	WTORs	SDF	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

Session Num Name	GM	Dasd	IR	Last Consistency Time
01 ZPROD01	RUNNING	OK		2010-02-09 10:32:34

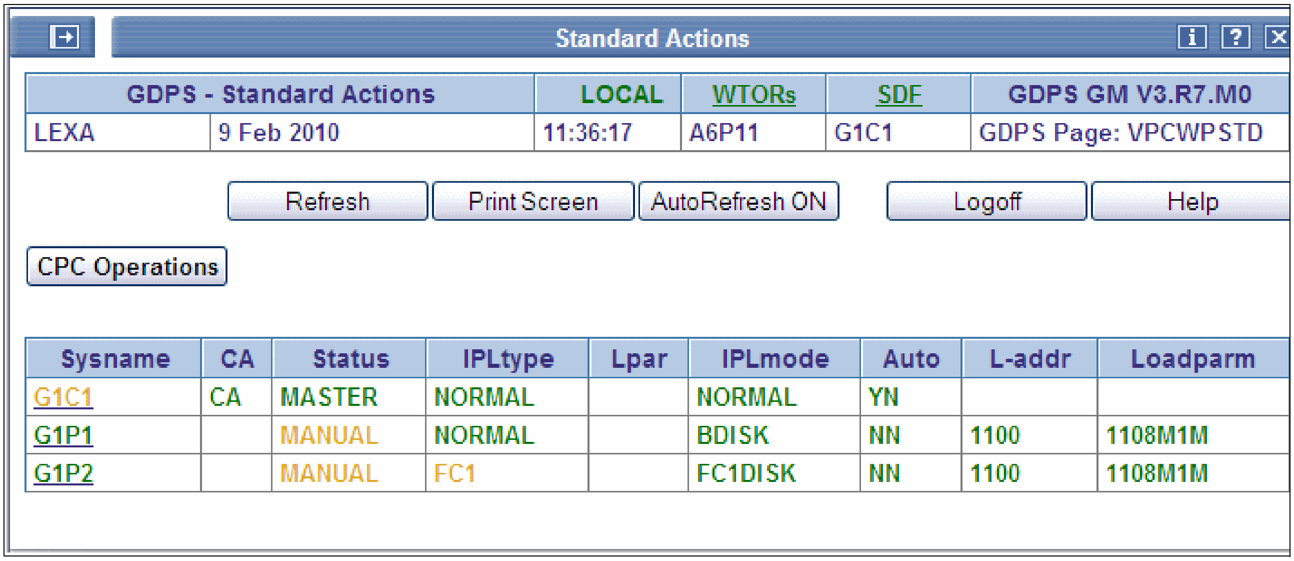
Figure 6-6 Disk Mirroring panel for GDPS/GM K-sys

Remember that these GDPS-provided panels are *not* intended to be a remote copy monitoring tool. Because of the overhead involved in gathering the information to populate the NetView panels, GDPS only gathers this information 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 any problem is detected.

Standard Actions

As previously explained, the K-sys does not provide any management functions for any systems, and 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.

On the K-sys, the only Standard Action available is the one to define or change the IPL information for recovery systems (production systems when they are recovered in the recovery site). Changes made on this panel are propagated automatically to the R-sys. The K-sys Standard Actions panel is shown in Figure 6-7 on page 135.

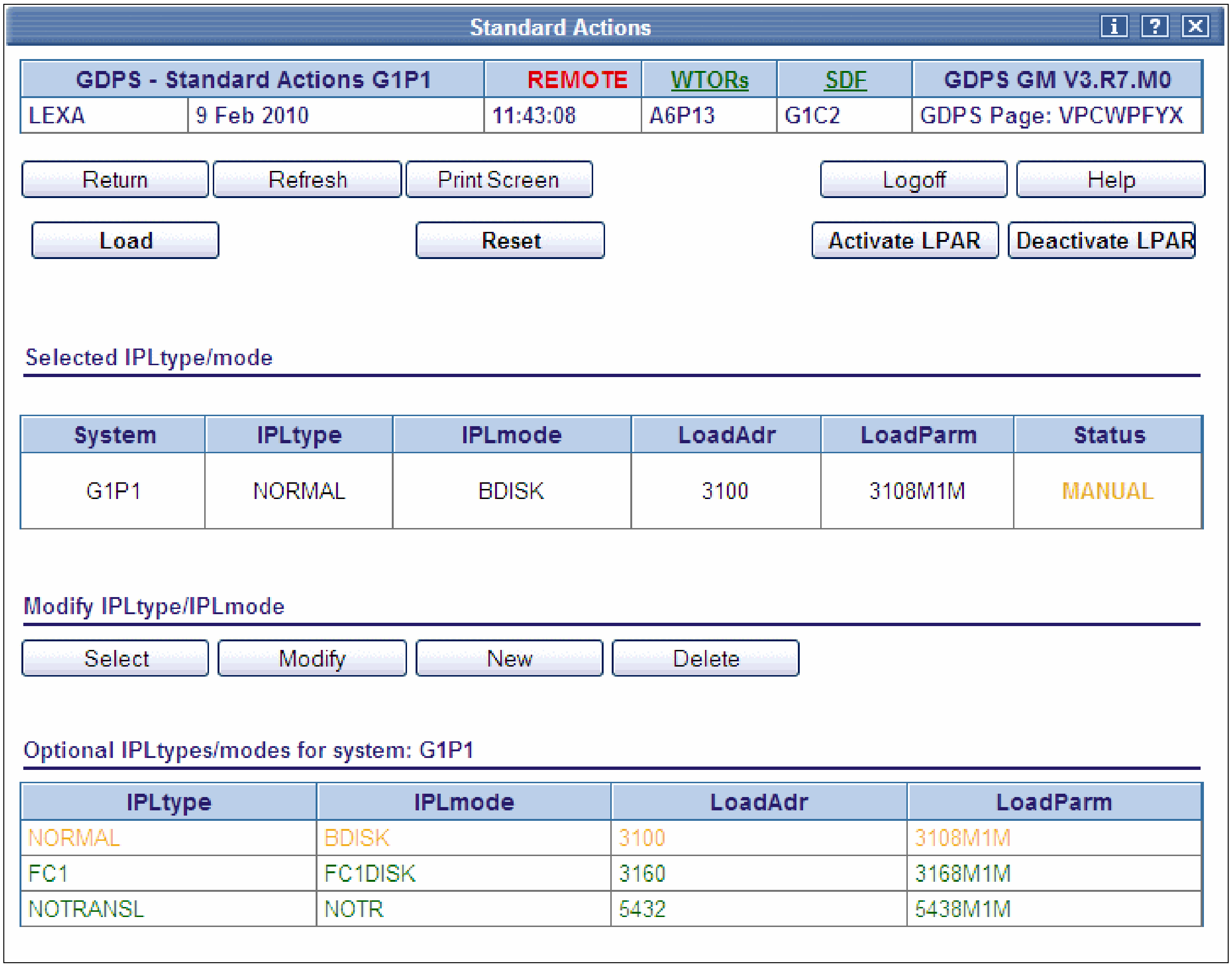


GDPS - Standard Actions		LOCAL	WTORs	SDF	GDPS GM V3.R7.M0
LEXA	9 Feb 2010	11:36:17	A6P11	G1C1	GDPS Page: VPCWPSTD

Sysname	CA	Status	IPLtype	Lpar	IPLmode	Auto	L-addr	Loadparm
G1C1	CA	MASTER	NORMAL		NORMAL	YN		
G1P1		MANUAL	NORMAL		BDISK	NN	1100	1108M1M
G1P2		MANUAL	FC1		FC1DISK	NN	1100	1108M1M

Figure 6-7 GDPS/GM K-sys Standard Actions panel

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 6-8 on page 136. Functions are provided to activate and deactivate LPARs, to IPL and reset systems, and to update the IPL information for each system.



GDPS - Standard Actions G1P1		REMOTE	WTORs	SDF	GDPS GM V3.R7.M0
LEXA	9 Feb 2010	11:43:08	A6P13	G1C2	GDPS Page: VPCWPFYX

Selected IPLtype/mode

System	IPLtype	IPLmode	LoadAdr	LoadParm	Status
G1P1	NORMAL	BDISK	3100	3108M1M	MANUAL

Modify IPLtype/IPLmode

Optional IPLtypes/modes for system: G1P1

IPLtype	IPLmode	LoadAdr	LoadParm
NORMAL	BDISK	3100	3108M1M
FC1	FC1DISK	3160	3168M1M
NOTRANSL	NOTR	5432	5438M1M

Figure 6-8 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-step, or are intended to impact just one resource. Examples are IPLing a system, updating the IPL address or the Loadparm to be used the next time a system is IPLed, or activating an LPAR.

So, 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 would consist of four separate Standard Actions that you would initiate sequentially from the panel.

GDPS scripts

Nearly all the functions that can be initiated via the panels (and more) are also available from GDPS scripts. A script is a "program" consisting of one or more GDPS functions. In addition to the low-level functions available through the panels, scripts can invoke functions with a single command that might require 6 or 7 separate steps if performed through the panels.

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 via a GDPS script; 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 very powerful because they can access the full capability of GDPS. The ability to invoke all the GDPS functions through a script provides:

- 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 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 System 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!

- Automatic checking of results from commands

  Because the results of many GDPS commands can be very 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 could have been six GM commands executed against thousands of devices) completed successfully before proceeding with the next command in the script.

- 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 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. An example of a planned action in GDPS/GM would be a script that would prepare the secondary disks and LPARs for a disaster recovery test. Such a script would carry out 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 while 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 would run another script in the R-sys to:

- 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 shutdown 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 would 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 job. This is especially suited to processes that are run regularly, and have some interaction with the GDPS environment.

6.5 GDPS/GM monitoring and alerting

We discuss the GDPS SDF panel in 6.4.1, “NetView panel interface” on page 130. 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 could 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 additionally 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 as long as the K-sys is up and running.

If an alert is posted, the operator would have to investigate (or escalate, as appropriate) and corrective action would 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/GM 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 impact the ability of GDPS/GM to carry out recovery operations. This will maximize the installation's chance of achieving its 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 GDPS-related settings adhere to GDPS best practices.

The z/OS Health Checker infrastructure is intended to check a variety of settings to determine whether these settings adhere to z/OS *best practices* values. For those settings that are found to be not in line with best 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 GDPS-related parameter settings, such as z/OS PARMLIB settings or NetView settings and the recommendations and best practices for these settings have been documented in the GDPS publications. If these settings do not adhere to recommendations, this could hamper the ability of GDPS to perform critical functions in a timely manner.

Although GDPS monitoring would detect that GDPS was not able to perform a particular task and raise an alert, the monitor alert may be too late, at least for that particular instance of an incident. Often, if there are changes in the customer environment, this might necessitate adjustment of some parameter settings associated with z/OS, GDPS and other products. It is possible that you could miss making these adjustments, which may 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, due to a setting that is perhaps less than ideal.

For example, there are a number of address spaces associated with GDPS/GM and best 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 GDPS-recommended characteristics.

Similar to z/OS and other products that provide health checks, GDPS health checks are optional. Additionally, a number of the best practices values that are checked, as well as the frequency of the checks, are customer-customizable to cater to unique customer environments and requirements.

6.6 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, note that 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 would be used to give you a consistent restart point. All updates made to the production disks after the start of the test would need to be recreated, 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*. F disks are additional 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 not impact your RPO and DR capability.

Currently, GDPS/GM supports the definition and management of a single F disk for each Metro-Global Mirror triplet (A, B, and C disk combos) in the configuration. To reduce management and operational complexity, support was added to 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 reducing complexity, this capability provides constraint relief for customers with many mirrored devices. Without this capability, a maximum of approximately 21 K devices could be mirrored, bringing the number of UCBs in the R-sys to the 63 K limit. With this function, up to almost 32 K devices could be mirrored before hitting the 63 K UCB limit in the R-sys.

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 you should consider installing the additional disk capacity to support F disks in your Global Mirror environment.

6.6.1 Usage of FlashCopy Space Efficient

As discussed in "FlashCopy Space Efficient (FlashCopy SE)" on page 33, by using FlashCopy Space Efficient (SE) volumes, you may be able to lower the amount of physical storage needed, and thereby reduce the cost associated with providing a tertiary copy of the data. Support is added to allow FlashCopy Space Efficient 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 recommend 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 would reclaim 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 due to 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 could be used for short-term, less-expensive testing, but would not be suitable for actual recovery due to their non-temporary nature.

6.7 Services component

As you have seen, 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 would be extremely rare to find a team that possessed 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.

6.8 GDPS/GM prerequisites

Important: For the latest GDPS/GM prerequisite information, refer to the GDPS product Web site:

http://www-03.ibm.com/systems/z/advantages/gdps/getstarted/gdps_gm.html

6.9 Comparison of GDPS/GM versus other GDPS offerings

There are so many features and functions available in the various members of the GDPS family that it is sometimes difficult to remember them all, and to recall which offerings support them. To position the offerings Table 6-1 on page 142 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/XRC	GDPS/GM
Continuous availability	Yes	Yes	No	No
Disaster recovery	Yes	Yes	Yes	Yes
Supported distance	200 km 300 km (BRS configuration)	200 km 300 km (BRS configuration)	Virtually unlimited	Virtually unlimited
FlashCopy support	Yes	Yes	Yes	Yes
Reduced impact initial copy/resynch	Yes	Yes	N/A	N/A
PtP VTS support	Yes	No	Yes	No
Production sysplex automation	Yes	No	No	No
Span of control	Both sites	Both sites (disk only)	Recovery site	Disk at both sites. Recovery Site (CBU / LPARs)
GDPS scripting	Yes	No	Yes	Yes
Monitoring, alerting, and Health Checks	Yes	Yes	Yes	Yes
Query Services	Yes	Yes	No	No
MGM	Yes (IR or non-IR)	Yes (Non-IR only)	N/A	Yes (IR or non-IR)
MzGM	Yes	Yes	Yes	N/A
Open LUN	Yes	Yes	No	Yes
z/OS equivalent functionality for Linux for System z	Yes	No	Yes	Yes
Heterogeneous support via DCM	Yes (VCS and SA AppMan)	No	Yes (VCS only)	Yes (VCS only)
Web interface	Yes	Yes	No	Yes

6.10 Summary

GDPS/GM provides automated disaster recovery capability over virtually unlimited distances for both CKD and FBA devices. It does not have a requirement for a z/OS System Data Mover system like XRC does, but 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 bring up the production systems. The powerful script 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 in order 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 much more user-friendly interface for monitoring and managing the remote copy configuration. This includes the initialization and monitoring of the GM volume pairs based upon policy, and performing routine operations on installed storage subsystems.

7

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 System z, 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 customers to manage data and coordinate disaster recovery across multiple platforms. The various extensions are available in one or more of the GDPS offerings. The extensions described in this chapter are:

- Open LUN Management function
  - Available for GDPS/PPRC, GDPS/PPRC HyperSwap Manager, and GDPS/Global Mirror
- GDPS/PPRC Multiplatform Resiliency for System z (also known as xDR)
  - Available for GDPS/PPRC
- 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

7.1 Open LUN Management function

As discussed in 3.1.3, “Protecting distributed (FBA) data” on page 51, many enterprises today run applications that update data across multiple platforms. For these enterprises there is a need to manage and protect not only the System z data, but also the data residing on the non-System z servers. GDPS/PPRC, GDPS/PPRC HyperSwap Manager, and GDPS/Global Mirror have been extended to manage a heterogeneous environment of System z 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 System z and distributed systems platforms, GDPS/PPRC and GDPS/Global Mirror can manage the Metro Mirror or Global Mirror remote copy configurations, as well as the FlashCopy for distributed systems storage.

Open LUN support extends the GDPS/PPRC Freeze capability to Open LUN (FBA) devices residing in supported disk subsystems to provide data consistency for not only the System z data, but also the data on Open LUNs. If you are using the GDPS/PPRC function known as xDR (described in 7.2, “GDPS/PPRC Multiplatform Resiliency for System z” on page 147), V3.8 now supports native Linux for System z systems running on SCSI attached Fixed Block Architecture (FBA or FB) disks. In addition to providing the Freeze capability for FBA disks, the new support provides 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 should be handled for Open LUN (FB disks) and System z (CKD disks), when mirroring or primary disk problems are detected:

> **Note:** Irrespective 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.

- You can select to Freeze all GDPS-managed devices.

  If this option is used, both the CKD and FBA devices are in a single consistency group. Any Freeze trigger, either for the System z or Open LUN devices, will result in both the Open LUN and the System z 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 System z and distributed servers that update the same data, for example using DB2 DRDA®, the Distributed Relational Database Architecture™.

- You can select to Freeze devices by group.

  If this option is selected, the CKD and xDR controlled FBA devices are in a separate consistency group from the non-xDR FBA devices. The Freeze will only be performed on the group for which the Freeze trigger was received. If the Freeze trigger occurs for a System z disk device, only the CKD and xDR controlled FBA devices will be frozen. If the trigger occurs for an non-xDR controlled FBA disks, only they will be frozen.

The Global Mirror remote copy technology, described in 2.4.3, “Global Mirror” on page 28, inherently provides data consistency for both System z and distributed Systems data.

A sample Open LUN GDPS/PPRC configuration is shown in Figure 7-1 on page 147.

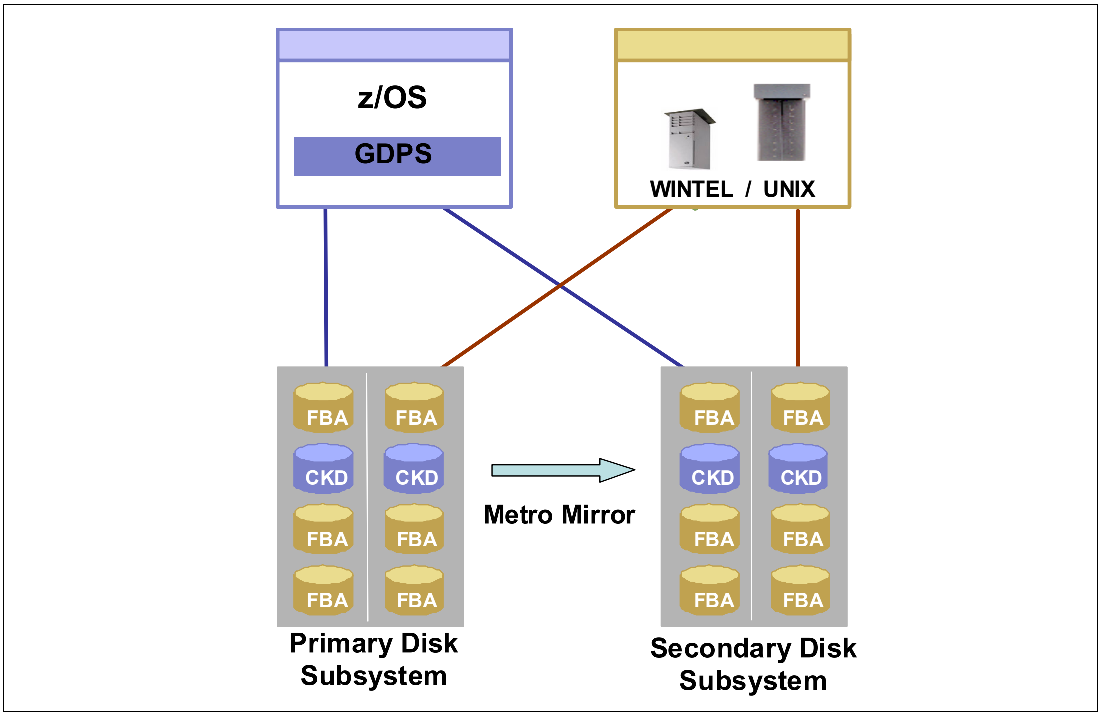


Figure 7-1 Open LUN support

Note also that this support is limited to managing the remote copy of the non-System z *data,* and providing data consistency of all data. There is no automation or system restart capability for the non-System z *systems*, unless you have Distributed Cluster Management (DCM) support as described in 7.3, "Distributed Cluster Management" on page 153. Also, there is no HyperSwap support for the Open LUN devices, because HyperSwap involves changing information at the control block level within the operating system.

7.2 GDPS/PPRC Multiplatform Resiliency for System z

Note: For the remainder of this section, Linux on System z may also be referred to as Linux. The terms are interchangeable.

As discussed in 3.3.1, "Multiplatform Resiliency for System z (also known as xDR)" on page 58, when customers implement a multitiered architecture, with application servers that run on Linux on System z and database servers that run on z/OS, there is a need to provide a coordinated near Continuous Availability/Disaster Recovery solution for both z/OS and Linux. GDPS/PPRC provides this capability with a function called **"Multiplatform Resiliency for System z (also known as xDR)"**. To provide these capabilities, GDPS/PPRC communicates with System Automation for MultiPlatforms (SA MP).

Linux can run either:

- As a guest under z/VM

  For Linux guests, SA MP is running in a dedicated Linux guest (known as the *proxy*) and in each Linux server guest. In Linux terms, the proxy is a cluster made up of only one node.

  The proxy guest is used by GDPS to communicate commands to z/VM and to monitor for disk errors that are reported to GDPS. Status information, Freeze and HyperSwap triggers are passed to GDPS/PPRC by SA MP. GDPS/PPRC uses SA MP to pass commands to z/VM and Linux guests.

The disks being used by z/VM and Linux in this configuration must be CKD disks.

> **Note:** The xDR function can be used even if the guests under z/VM are other than Linux. The only requirement is that you have the proxy as a dedicated Linux guest.

- Run natively in its own partition

  In this configuration, no specific Linux node is designated as the proxy. SA MP is running in each Linux partition, and each system receives and executes commands and reports disk errors.

  Status information and Freeze and HyperSwap triggers are passed to GDPS/PPRC by SA MP. GDPS/PPRC uses SA MP to pass commands to native Linux systems.

  The disks being used by native Linux in this configuration can be either CKD or FBA disks. GDPS/PPRC V3.8 xDR adds support for native Linux for System z systems running on SCSI attached Fixed Block Architecture (FBA or FB) disks.

  This support builds on the existing Open LUN management capability which provides PPRC and Freeze for FB disks to additionally provide planned and unplanned HyperSwap for the FB disks used by xDR-managed native Linux systems.

GDPS/PPRC will provide the reconfiguration capabilities for the Linux servers and data in the same manner as for z/OS systems and data. To support planned and unplanned outages, GDPS/PPRC provides the following types of recovery actions:

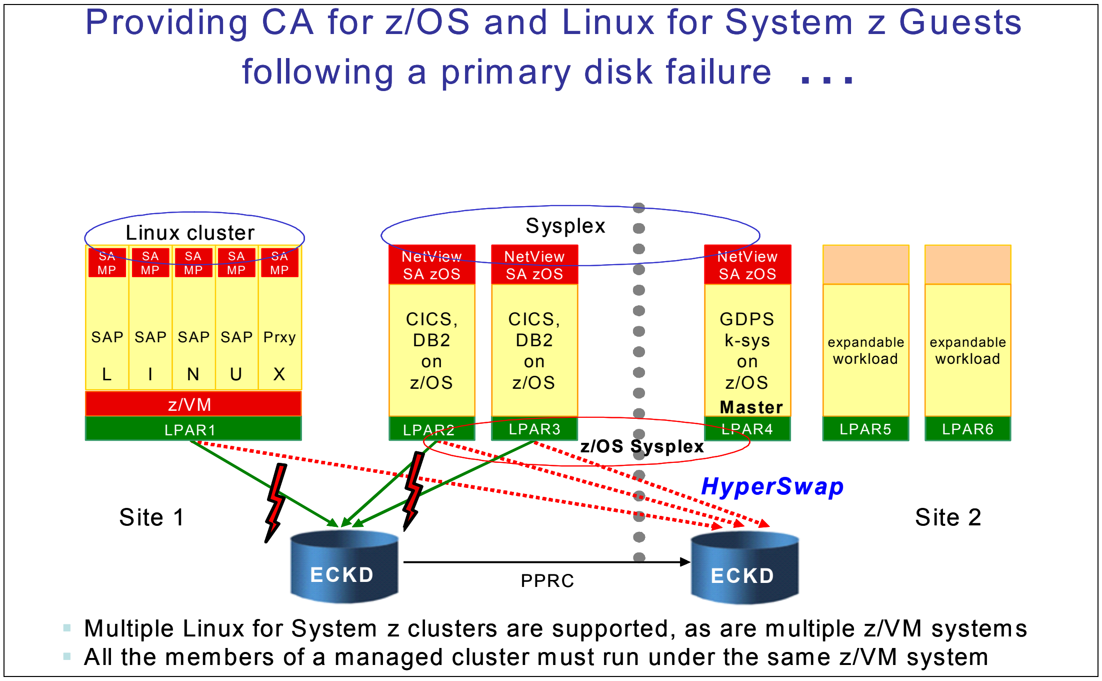
- Re-IPL failing operating system images
- Site takeover or failover of a complete production site
- Coordinated planned and unplanned HyperSwap of disk subsystems, transparent to the operating system images and applications using the disks.
- Graceful shutdown/startup of the Linux on System z cluster, nodes in the cluster and the z/VM host.
- Data consistency with freeze functions across z/OS and Linux

7.2.1 Guest Linux under z/VM

z/VM provides a HyperSwap function so that the virtual device associated with one real disk can be swapped transparently to another disk. HyperSwap can be used to switch to secondary disk storage subsystems mirrored by Metro Mirror. If there is a hard failure of a storage device, GDPS/PPRC coordinates the HyperSwap with z/OS for continuous availability spanning the multitiered application. For site failures, GDPS/PPRC invokes the Freeze function for data consistency and rapid application restart, without the need for data recovery.

From a Freeze and HyperSwap perspective, the disks used by Linux and z/VM are in the same disk group as the z/OS disks, so a Freeze-inducing error on one of those disks will result in a Freeze on all the z/OS disks. It also works the other way around: a Freeze trigger introduced in one of the z/OS systems will result in a Freeze on all disks (z/OS, z/VM and Linux). Similarly, an unplanned HyperSwap trigger can be detected on z/OS, z/VM, or guest disks, and will result in a swap of all disks. A planned HyperSwap can transparently swap all disks (z/OS, z/VM, guest), as well.

Figure 7-2 on page 149 shows an example of a configuration where several Linux nodes are running as guests under z/VM. Note that one of the Linux guests is the proxy. The figure also illustrates the actions taken if a disk failure is detected and HyperSwap is invoked by GDPS/PPRC.



Providing CA for z/OS and Linux for System z Guests following a primary disk failure . . .

- Multiple Linux for System z clusters are supported, as are multiple z/VM systems
- All the members of a managed cluster must run under the same z/VM system

Figure 7-2 HyperSwap example - Linux running as a guest under z/VM

7.2.2 Native Linux on System z

Figure 7-3 on page 150 shows an example of a configuration where several Linux nodes are running natively in their own partitions, and all of them are under GDPS control.

Note that:

- None of the Linux systems have been designated as the proxy.
- Linux LPARs 1 and 2 are using CKD disks, whereas Linux LPARs 8 and 9 are using SCSI attached FBA disks

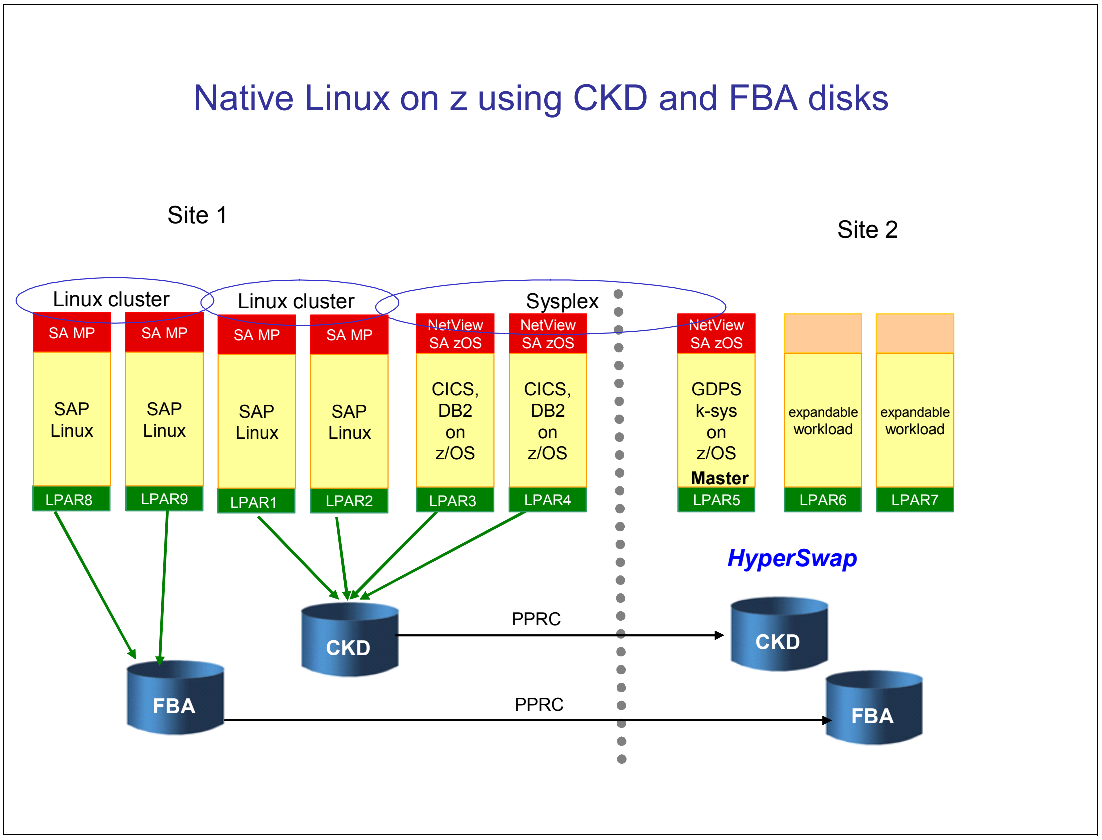
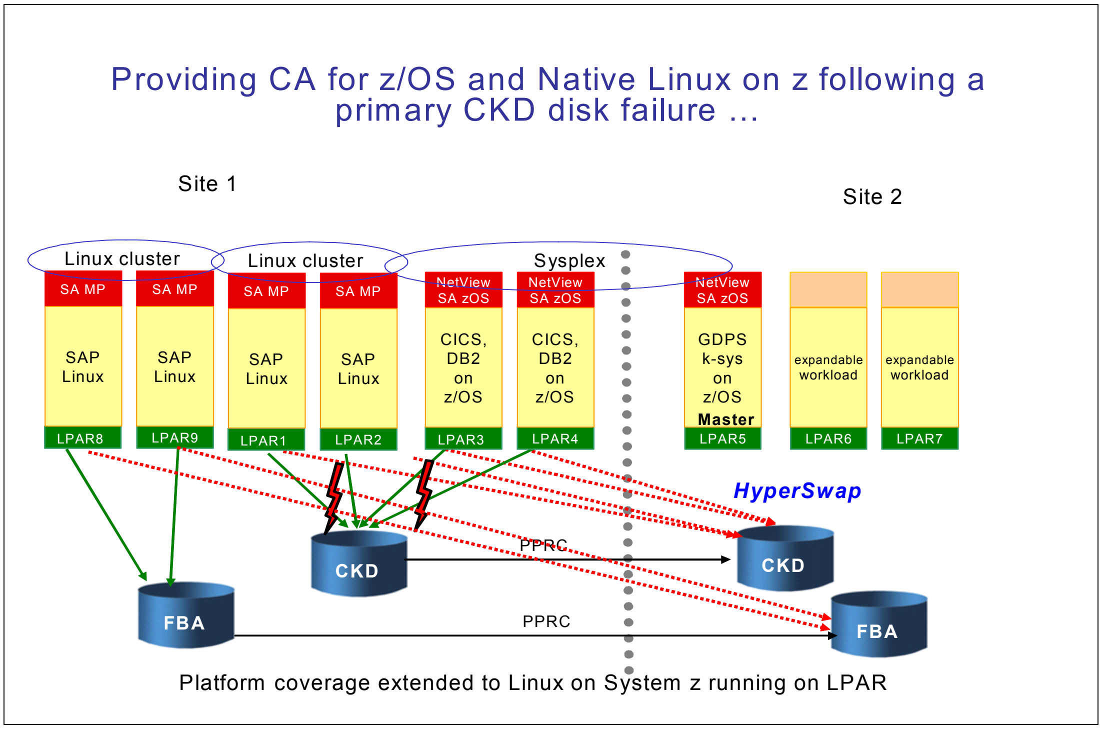


Figure 7-3 Native Linux on System z - LPARs using CKD and FBA disks

In this configuration, when a primary disk problem is detected either for a CKD or xDR FB disk, and the environment is enabled for HyperSwap at the time the trigger occurs, a HyperSwap is performed for the CKD and xDR FB disks.

Figure 7-4 on page 151 illustrates the actions taken if a disk failure is detected on the CKD disks and HyperSwap is invoked by GDPS/PPRC. Even though the disk failure was associated with the CKD disks, both the CKD and FBA disks are swapped to the secondary copy of the disks.



Providing CA for z/OS and Native Linux on z following a primary CKD disk failure ...

Platform coverage extended to Linux on System z running on LPAR

Figure 7-4 HyperSwap example following a CKD disk failure

7.2.3 Support for two GDPS Controlling systems

Prior to GDPS V3.5, xDR supported only one GDPS controlling system (also referred to as the GDPS Master K-sys), as illustrated in Figure 7-2 on page 149 and Figure 7-4. xDR functions could only be processed 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 7-5 on page 152 shows an xDR configuration with two GDPS controlling systems following a HyperSwap of the primary disks from Site1 to Site2. Note that 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 only 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 recommend that you run GDPS with two Controlling systems and enable xDR to support two Controlling systems.

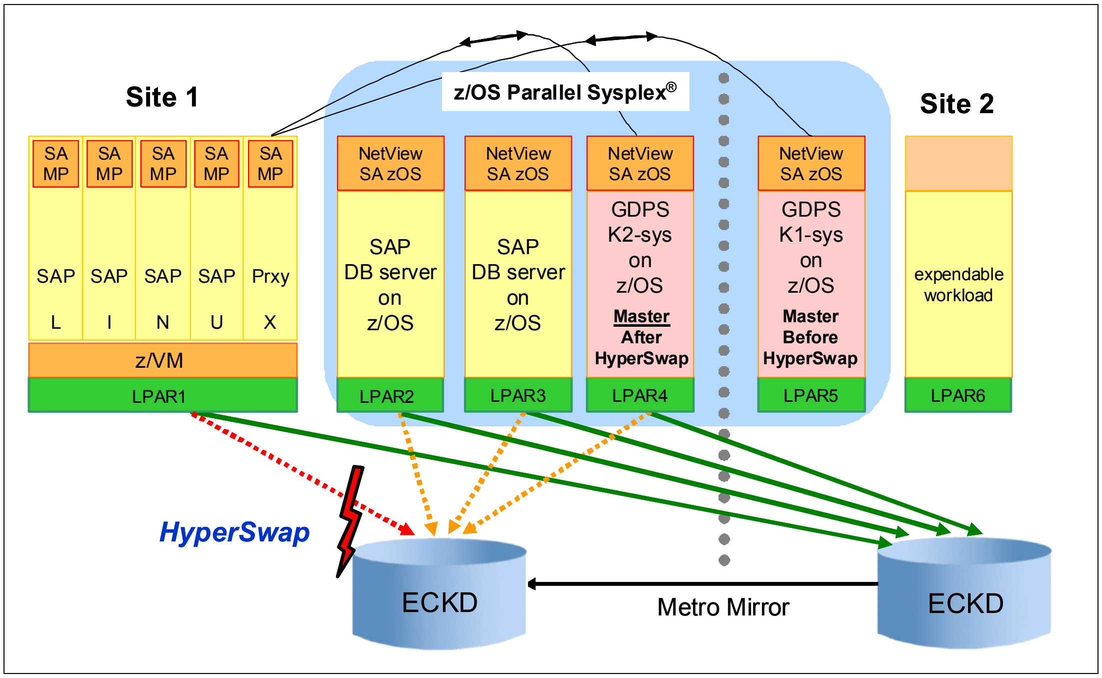


Figure 7-5 xDR configuration with two controlling systems after a HyperSwap

7.2.4 Disk and LSS sharing

When Linux is running as a guest under z/VM, xDR supports the sharing of LSSs by multiple z/VM systems. This facilitates the efficient sharing of resources, provides configuration flexibility, and simplifies the setup required to keep the LSSs separate. Additionally, it enables xDR environments to exploit the z/VM Cross System Extension (CSE) capability.

For example, suppose that you have more than one z/VM system and you want to be able to perform the following tasks:

- Share the RACF® database across your systems
- Manage one VM Directory for all the systems
- Ensure that a minidisk may only be linked RW on one guest on one system, and have all the systems enforce that
- Share the z/VM System Residence volumes

z/VM Cross System Extension 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. Refer to *z/VM CP Planning and Administration*, SC24-6083, for details about CSE.

In summary, if you want to share LSSs and additionally share disks:

- In one LSS you may place disks for as many xDR-managed z/VM systems as you wish.
- If desired, any GDPS-managed z/VM disk can be shared by multiple xDR-managed z/VM systems. This requires that you also implement z/VM CSE.

7.2.5 Customer Verification Program

An xDR Customization Verification Program (CVP) is introduced with V3.8 to verify that installation and customization activities have been carried out correctly for both xDR native and guest Linux on System z environments. This helps identify any issues with the customization of the environment where many components exist with very specific setup and customization requirements.

It is an operator initiated program that can be used after initial setup, as well as 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.

7.2.6 xDR Extended Monitoring

The GDPS HyperSwap Monitor provides checking for z/OS systems to ascertain whether the GDPS managed z/OS systems meet required conditions. Any system not meeting 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.

Prior to GDPS 3.8, similar monitoring was not available for xDR systems. An xDR system (either native or z/VM) could be reset during a planned or unplanned HyperSwap, if it was found to be not HyperSwap ready, because it did not meet one or more of the various environmental conditions for HyperSwap.

GDPS HyperSwap monitoring, previously available for the z/OS systems is being extended for xDR systems (native and z/VM guest environments). A number of environmental conditions required for HyperSwap for xDR systems are now 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.

Additionally, any such xDR system that does not meet all environmental conditions that are monitored will be marked as being '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 in order to avoid having the system reset if an unplanned swap is triggered.

7.3 Distributed Cluster Management

Distributed Cluster Management (DCM) allows the management and coordination of planned and unplanned outages across non-System z distributed clusters in coordination with the System z workloads that GDPS is responsible for.

As discussed in 7.1, “Open LUN Management function” on page 146 and 7.2, “GDPS/PPRC Multiplatform Resiliency for System z” on page 147, many enterprises have requirements to provide automated failover and rapid recovery of business-critical applications residing not only on System z, but also residing on other platforms such as UNIX, AIX®, Windows and Linux. In addition, when you have a multitiered architecture, there is a need to provide a coordinated near Continuous Availability/Disaster Recovery solution for applications that may be residing on non-System z servers and those residing on System z.

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. Refer to “DCM functions (VCS)” on page 161 and “DCM functions (SA AppMan)” on page 168 for more details.

7.3.1 Distributed Cluster Management terminology

This section presents new terminology and provides 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, refer to “VCS terminology” on page 155. For terminology applicable only to the DCM support for SA AppMan, refer to “SA AppMan terminology” on page 164.

Distributed Cluster Management (DCM)
The GDPS capability to manage and coordinate disaster recovery across distributed servers that are clustered using high availability clustering solutions alongside the System z workloads that GDPS is responsible for.

Application site A site in which the applications (both distributed applications and System z applications) normally reside. This site is also referred to as *Site1* by GDPS with DCM.

Recovery site 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 A group of servers and other resources that act like a single system and enable high availability as well as, in some cases, load balancing and parallel processing.

K-sys The GDPS Controlling System.

R-sys The GDPS Remote Controlling System in a GDPS/GM configuration. It is located in the Recovery site.

Geographically Dispersed Open Clusters (GDOC)
An IBM services offering to help customers 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, you could 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.

7.3.2 DCM support for VCS

This section describes how the functions available with GDPS/PPRC (see Chapter 3, “GDPS/PPRC” on page 43), GDPS/XRC (see Chapter 5, “GDPS/XRC” on page 107), and GDPS/GM (see Chapter 6, “GDPS/Global Mirror” on page 123) 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 designed to be extensible to other high availability clustering solutions.

VCS terminology

This section presents new 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 term 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 term refers to functionality 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 via VCS GCO. (In this section, you may 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 GDPS-managed VCS resources. 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.

Veritas Cluster Server overview

Veritas Cluster Server (VCS) from Symantec 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 could be to a local server in the same site or, for disaster recovery, the failover could be to a remote cluster located several thousand miles away. VCS supports multiple operating systems, such as IBM AIX, Sun Solaris, HP-UX, and Linux. It also supports multiple hardware, software, and database replication technologies. For more information about Veritas Cluster Server, go to:

http://www.symantec.com/business/products/overview.jsp?pcid=2247&pvid=20_1

Figure 7-6 on page 156 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, the GCO option is required.

The configurations shown in the red and blue circles 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).

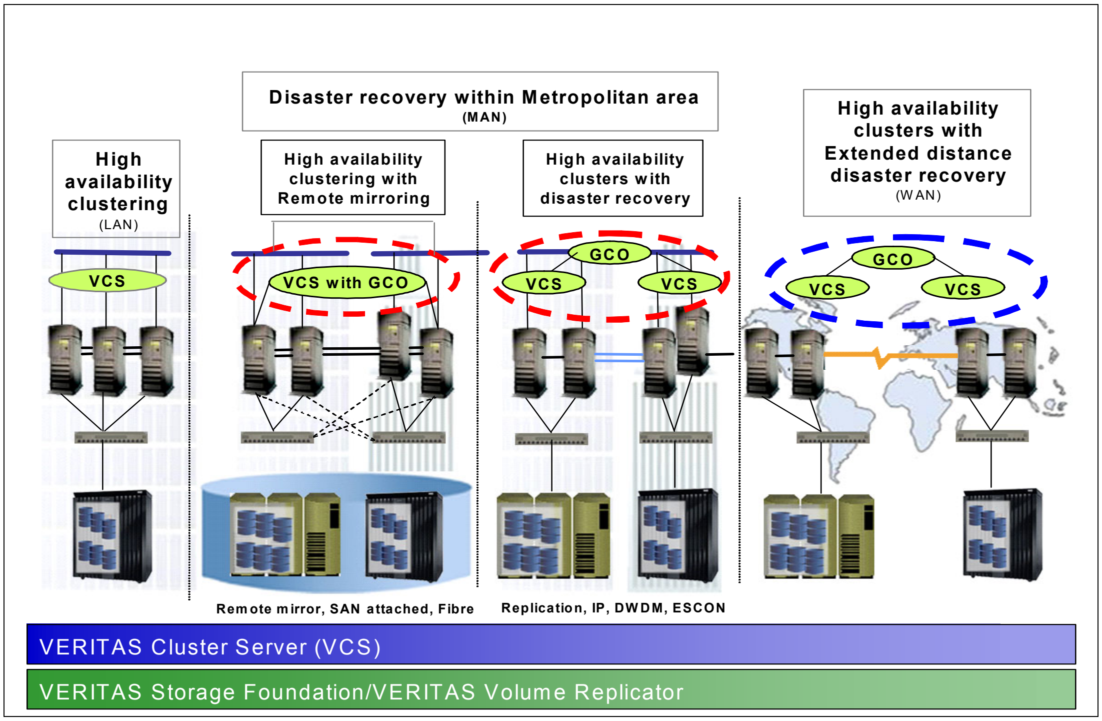

Disaster recovery within Metropolitan area
(MAN)
High availability clustering
(LAN)
High availability clustering with Remote mirroring
High availability clusters with disaster recovery
High availability clusters with Extended distance disaster recovery
(WAN)
VCS
VCS with GCO
GCO
VCS
VCS
GCO
VCS
VCS
Remote mirror, SAN attached, Fibre
Replication, IP, DWDM, ESCON
VERITAS Cluster Server (VCS)
VERITAS Storage Foundation/VERITAS Volume Replicator

Figure 7-6 Veritas Cluster Server configurations

Integrated configuration of GDPS/PPRC and VCS clusters

Figure 7-7 on page 157 illustrates the various components of a GDPS/PPRC configuration integrated with a VCS configuration that has the GCO option. The GDPS/PPRC configuration comprises of a multi-site Parallel Sysplex with a set of primary disks in Site1 being mirrored to a set of secondary disks in Site2 using Metro Mirror. The disk mirroring is managed by GDPS/PPRC as described in Chapter 3, “GDPS/PPRC” on page 43. There is a minimum of one GDPS K-sys in Site2, and optionally there can be a second K-sys in Site1.

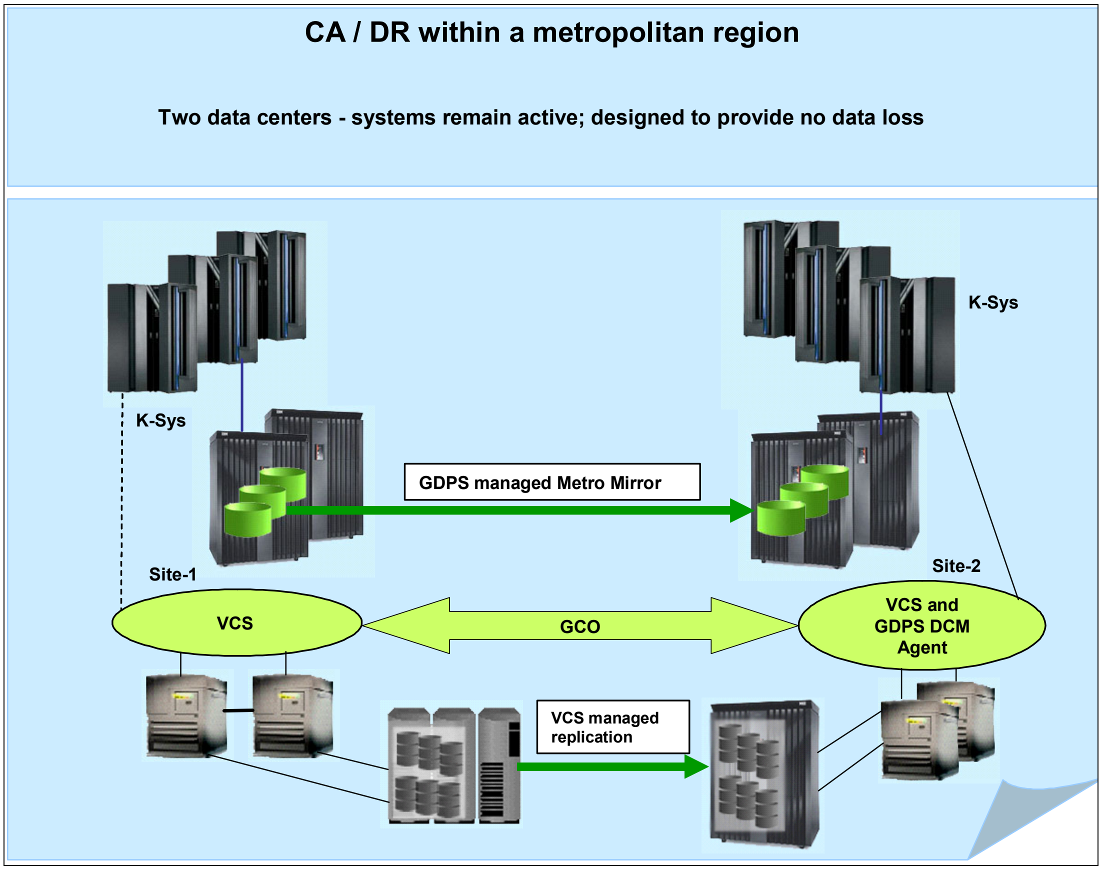

CA / DR within a metropolitan region
Two data centers - systems remain active; designed to provide no data loss
K-Sys
K-Sys
GDPS managed Metro Mirror
Site-1
Site-2
VCS
GCO
VCS and GDPS DCM Agent
VCS managed replication

Figure 7-7 GDPS/PPRC integration with VCS using GCO - metropolitan distance

The VCS configuration in this example is comprised 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 in the event of 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 customers that have sufficient resources in the recovery site to support this function.

Integrated configuration of GDPS/XRC and VCS clusters

Figure 7-8 illustrates the various components of a GDPS/XRC configuration integrated with a VCS configuration that has the GCO option. The GDPS/XRC configuration comprises of one or more System Data Movers (SDMs) and a GDPS K-sys in a sysplex or Parallel Sysplex in Site2, the recovery site.

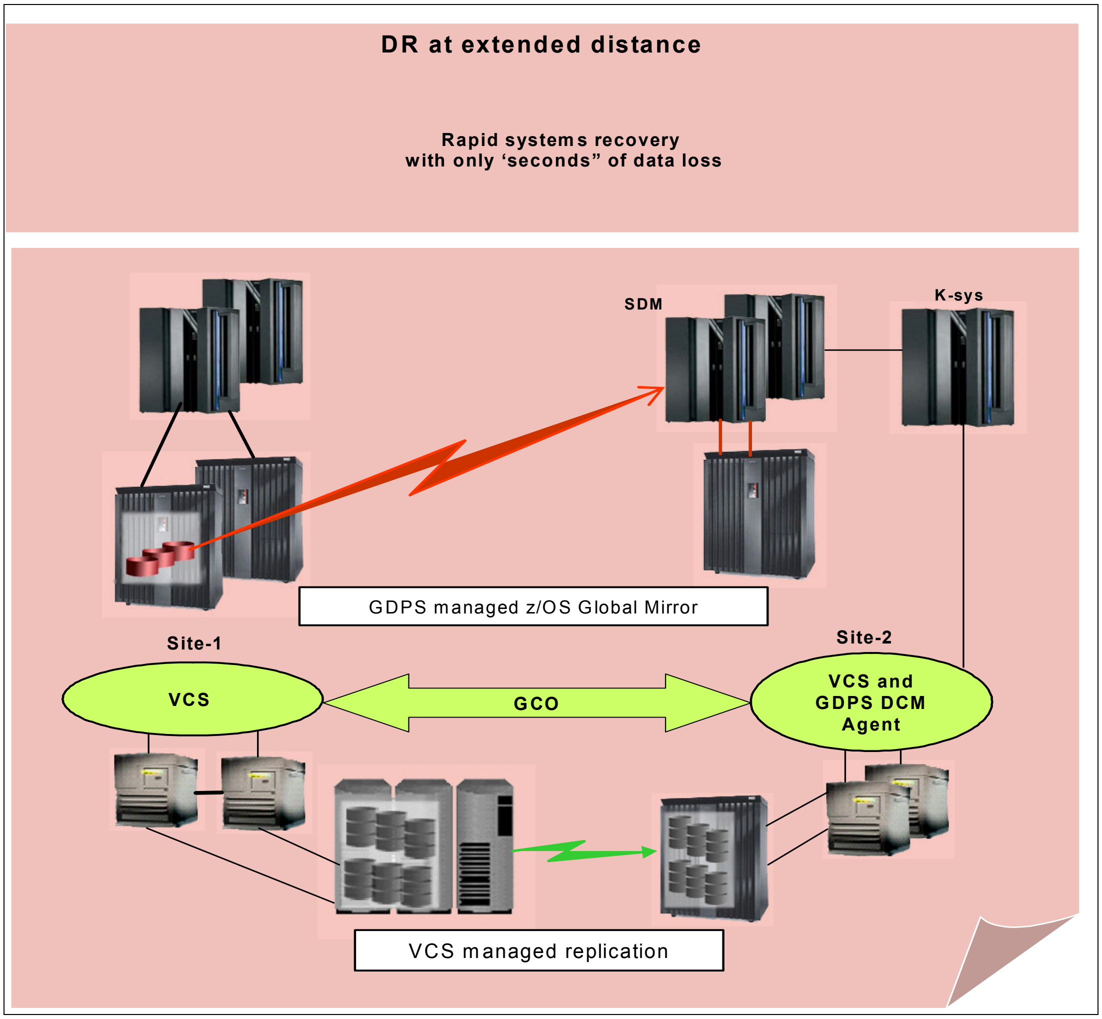


Figure 7-8 GDPS/XRC integration with VCS using GCO - unlimited distance

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 107.

The VCS configuration in this example is comprised 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 in the event of 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 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. 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 customers that have sufficient resources in the recovery site to support this function.

Integrated configuration of GDPS/GM and VCS clusters

Figure 7-9 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 123.

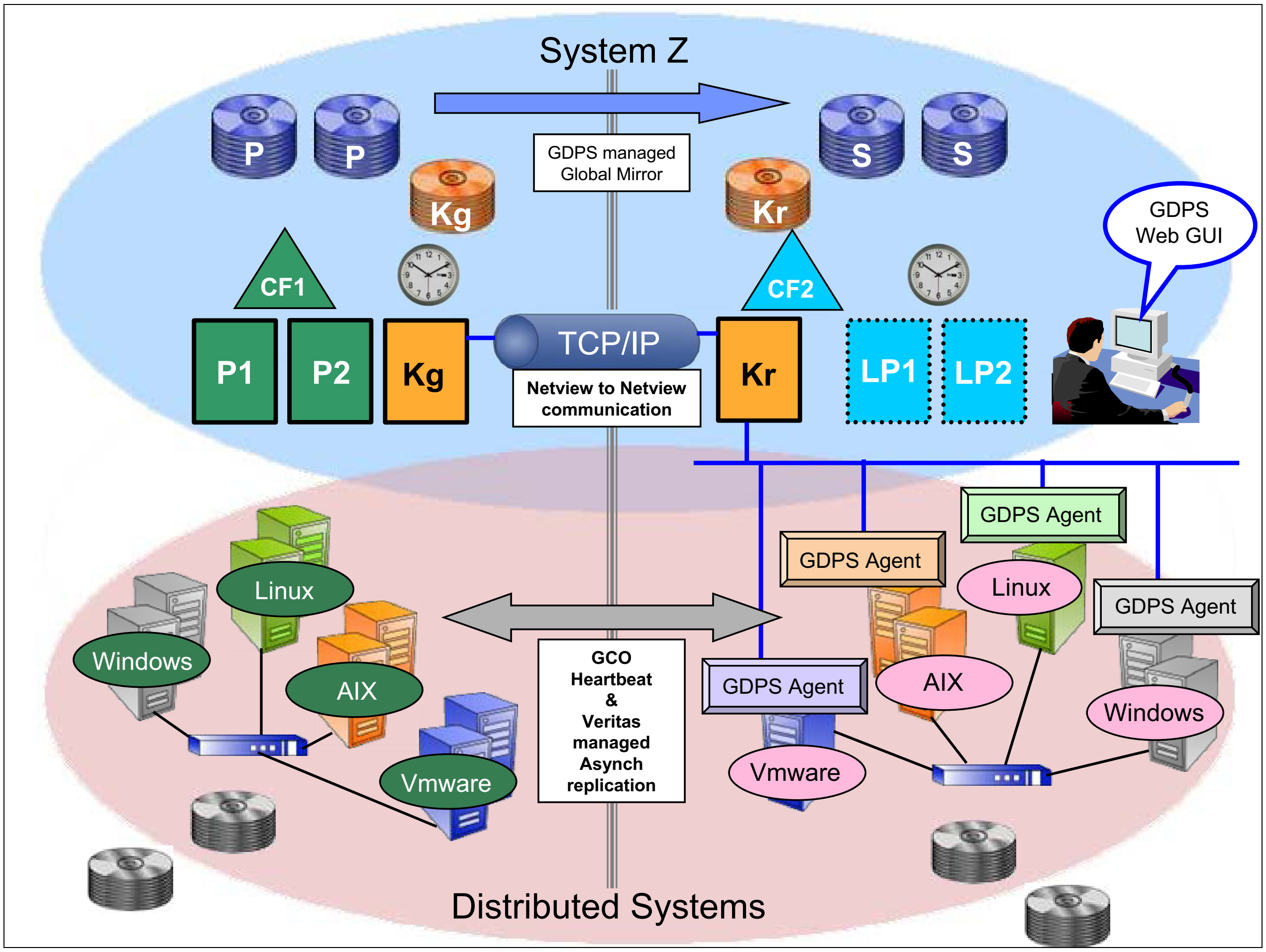


Figure 7-9 GDPS/GM integration with VCS using GCO - unlimited distance

The VCS configuration in this example is comprised 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 in the event of 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 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 only active 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 customers 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: System x®, p, and z. 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 System z workload. In this case:

- 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, Sun, and so on).

  There could be multiple global clusters of different server types; for example:

  - AIX VCS cluster 1 in Site1, AIX cluster 2 in Site2 - comprising 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 7-10 on page 161 depicts 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.

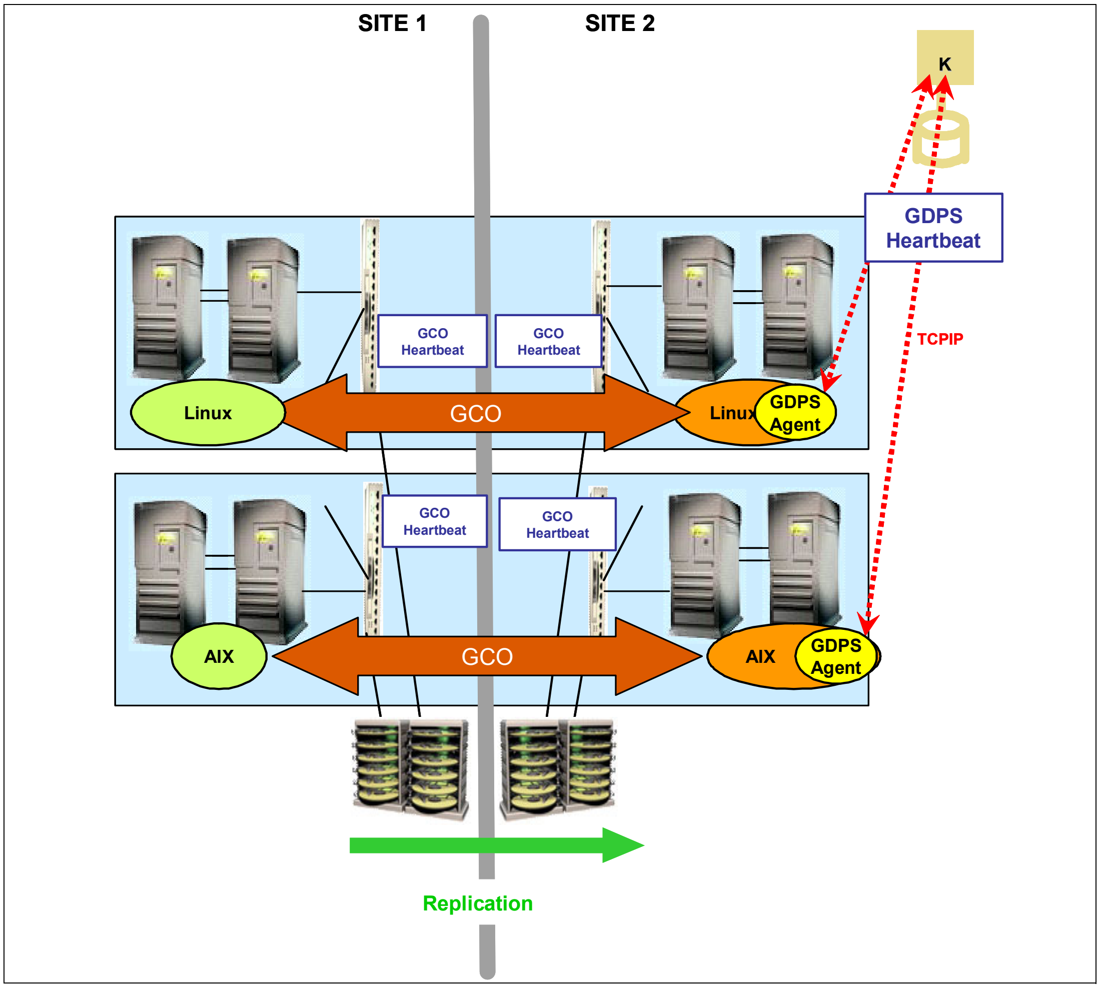


Figure 7-10 Multiple VCS cluster support

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 may 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 System z servers in the event of 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 7-11 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 via communication links, as shown in the figure.

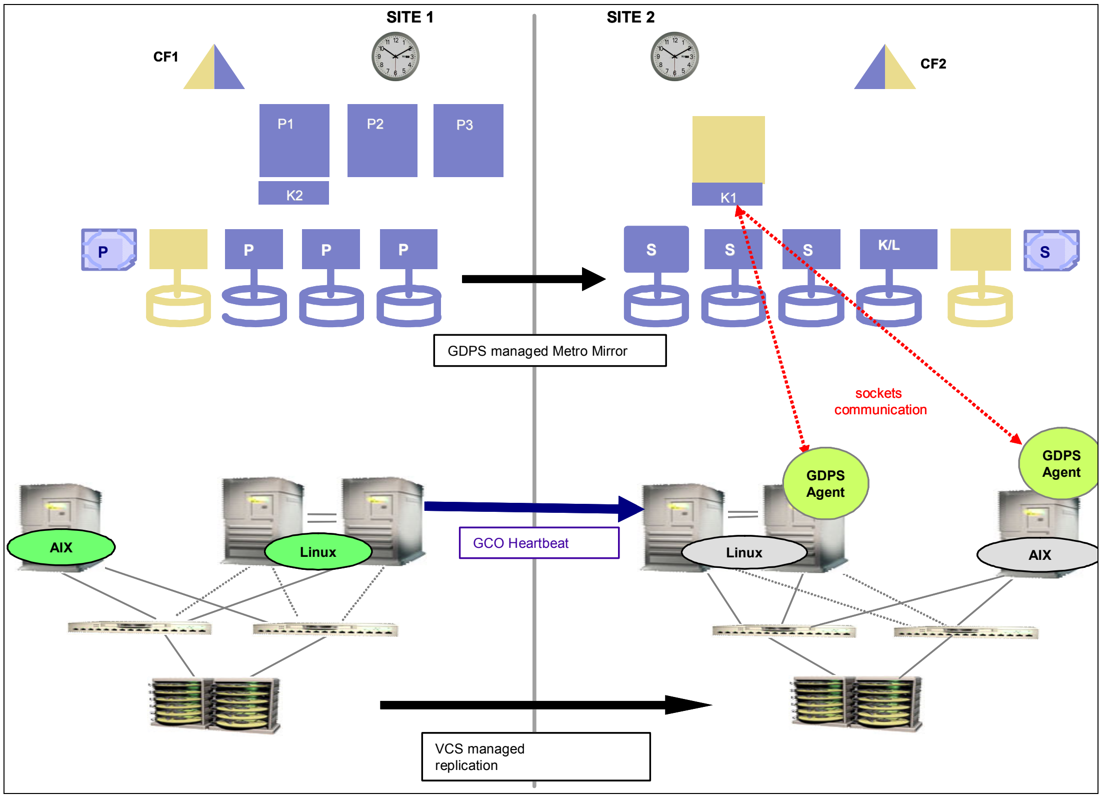


Figure 7-11 Example of GDPS/PPRC configuration with VCS clusters

The failure scenario shown in Figure 7-12 on page 163 is a Site1 failure when one or more failures occurred in Site1, which could 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.

Now that GDPS DCM support provides 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.

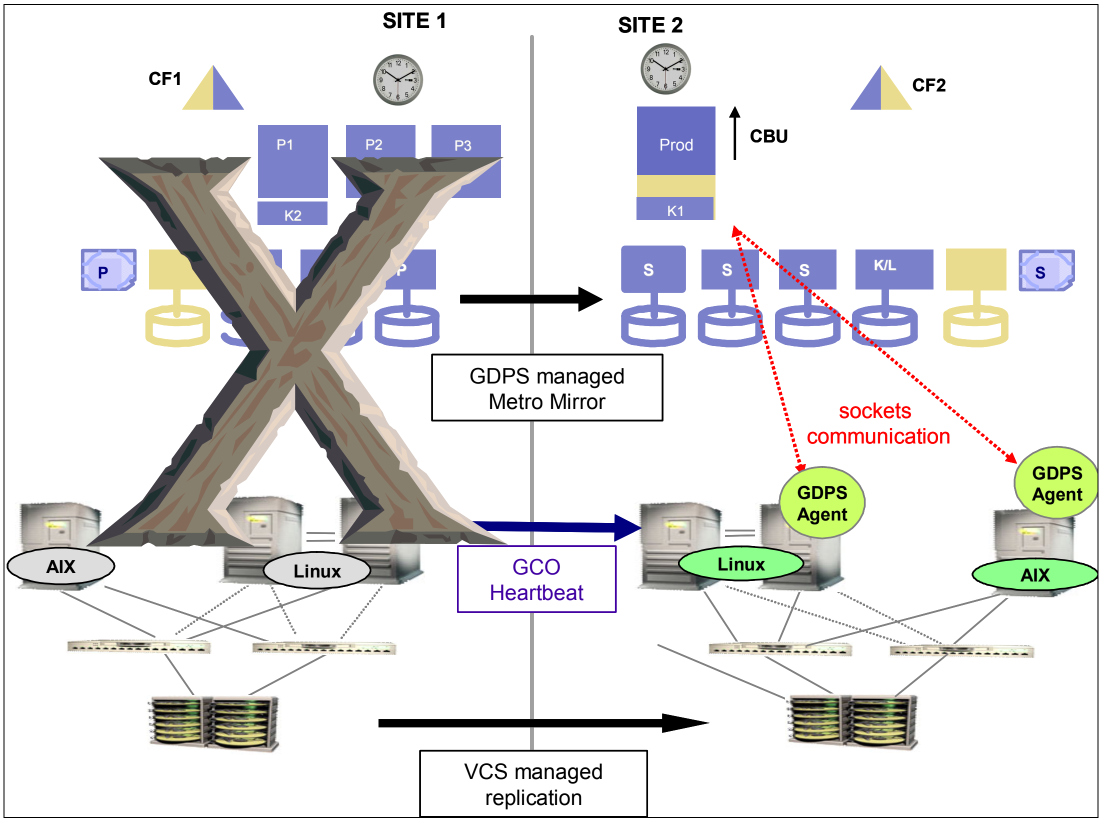


Figure 7-12 Example of GDPS/PPRC configuration with VCS clusters - Site1 failure

An existing GDPS script that previously performed failover and restart of System z 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 performs the following actions to move System z resources from Site1 to Site2:

- It resets production systems in Site1.
- It reconfigures the secondary disks.
- It activates CBU for servers in Site2.
- It switches couple data sets to those in Site2.
- It activates partitions in Site2.
- It reIPLs P1, P2, and P3 in Site2.

GDPS performs the following actions to move VCS clusters from Site1 to Site2:

- It forces a switch of the service group for the AIX cluster from Site1 to Site2.
- It 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 the System z resources and the VCS clusters when there is an unplanned outage that affects Site1.

7.3.3 DCM support for SA AppMan

This section describes how the functions available with GDPS/PPRC (see Chapter 3, “GDPS/PPRC” on page 43) have been integrated with functions provided by IBM Tivoli System Automation Application Manager (SA AppMan).

SA AppMan terminology

This section presents new terminology and provides a brief description for each term that is applicable to the DCM support for SA AppMan.

GDPS agent This term 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 GDPS-managed cluster resources. There is one GDPS agent for all SA AppMan cluster sets, running in Site2.

The GDPS agent is only available if you have enabled the Distributed Disaster Recovery (DDR) functionality 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 term refers to the SA AppMan feature that provides the interaction with GDPS. As mentioned, the GDPS agent is only available if you have enabled DDR.

First Level Automation (FLA) domain

This term 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.

Domain This term refers to the automation scope of an automation product instance such as SA MP, IBM High Availability Cluster Multiprocessing (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 term 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 term refers to a cluster that is dispersed across two sites.

Cluster set This term 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 term refers to applications that are critical to the business (such as databases and Web servers).

Discretionary workload

This term 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 to easily coordinate and manage across cluster technologies, so you can better control your enterprise business services.

In the example shown in Figure 7-13 on page 166, 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 among 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.

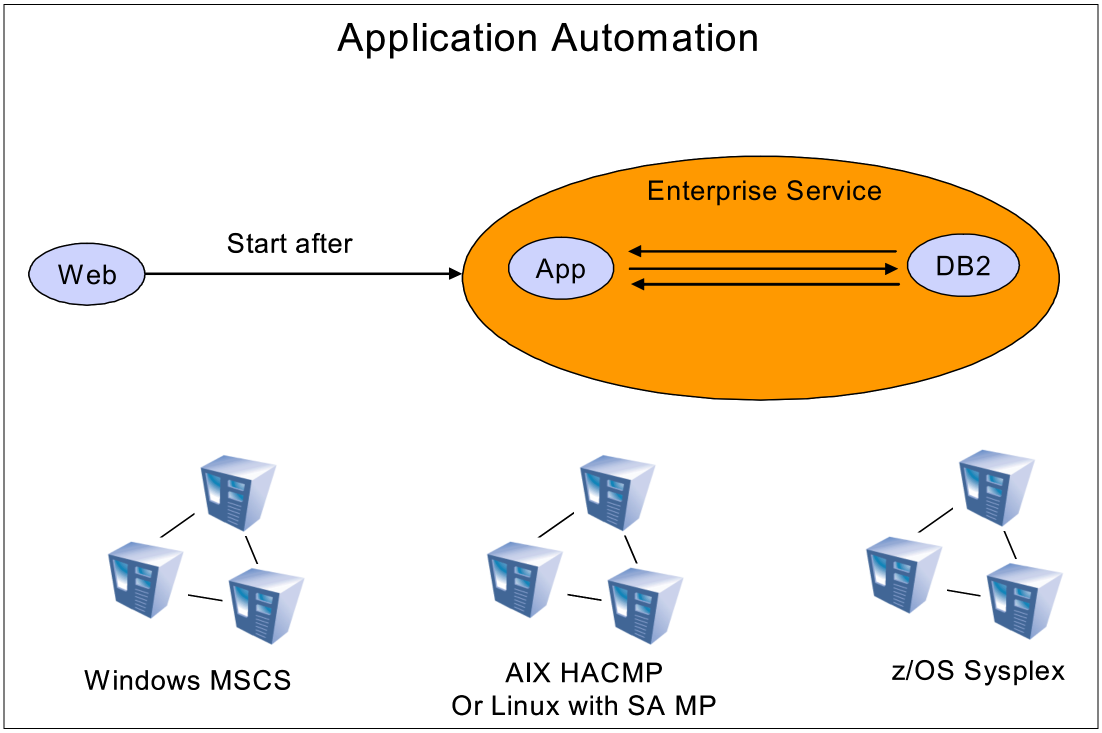


Figure 7-13 Sample end-to-end automation

SA AppMan provides adapters to help manage any combination of the following major clustering technologies:

- Veritas Cluster Server (VCS) on Solaris
- High Availability Cluster Multiprocessing (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

Integrated configuration of GDPS/PPRC and SA AppMan

Figure 7-14 on page 167 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 multi-site 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.

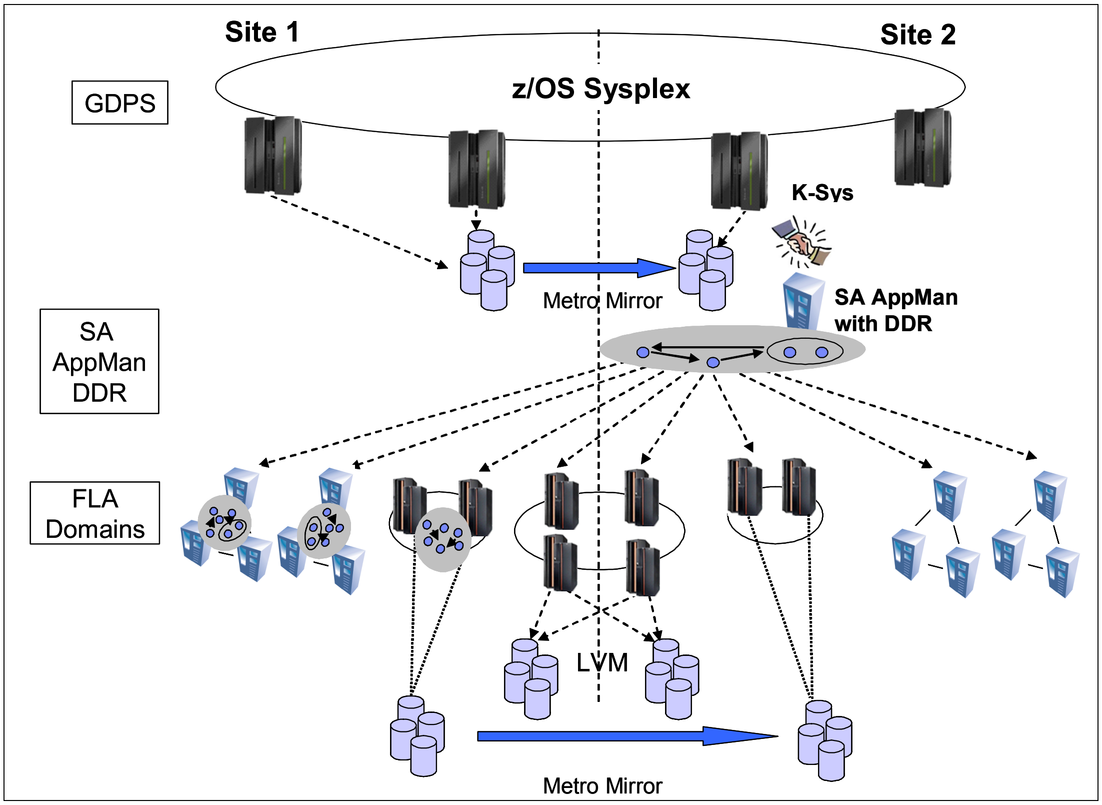


Figure 7-14 SA Application Manager - Distributed Disaster Recovery configuration

In the middle of the figure, one of the distributed servers has the SA AppMan feature called Distributed Disaster Recovery (DDR). Because the DDR feature includes the GDPS agent functionality 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.

SA AppMan code that communicates with GDPS is implemented in its own server 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 (System p, System x, System i®, and so on), and multiple applications can be running. As defined in “SA AppMan terminology” on page 164, 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 7-14 on page 167.

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 very much like the non-stretched cluster, because all applications runs 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 7.1, “Open LUN Management function” on page 146 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 nor by 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 via LVM such 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 assist 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 may 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.

- Scripting

  The scripting capability in GDPS provides workflow integration for actions taken on distributed servers and System z servers in the event of 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 7-15 on page 170 shows a sample configuration consisting 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 via LVM.

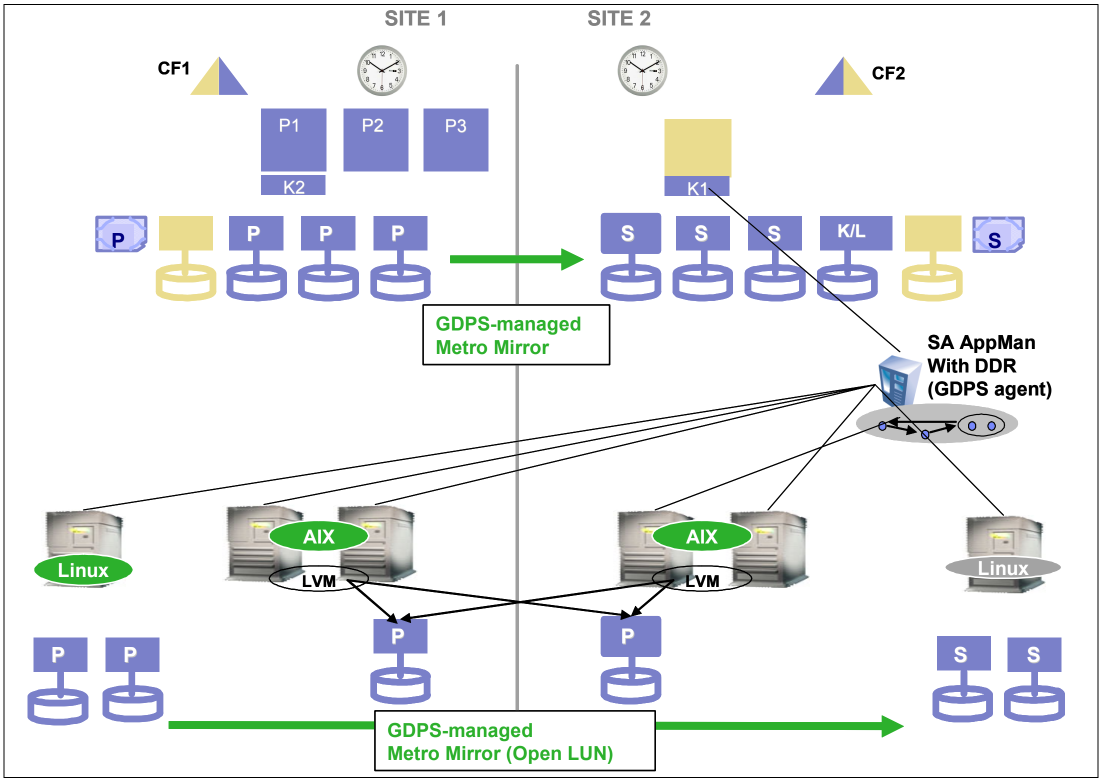


Figure 7-15 Sample GDPS/PPRC DCM configuration

The DDR feature that includes the GDPS agent communicates with the GDPS K-sys via communication links as shown in Figure 7-15.

Figure 7-16 on page 171 represents a failure scenario in which one or more failures occur in Site1, which could 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.

Now that GDPS DCM support provides 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.

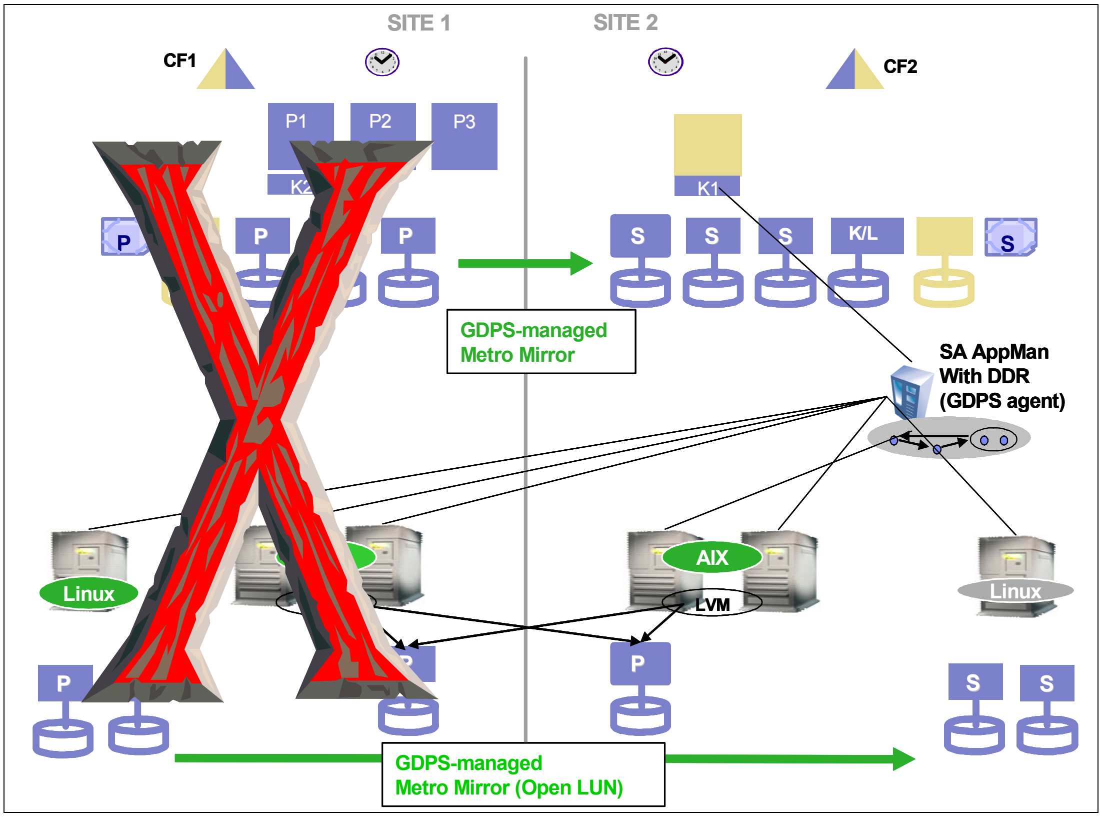


Figure 7-16 Sample GDPS/PPRC DCM configuration - Site1 failure

An existing GDPS script that previously performed failover and restart of System z resources can be extended to also include statements to automate the failover of SA AppMan clusters to Site2. When the additional script statements are executed as part of the site takeover script, GDPS automation performs the following actions to move System z resources from Site1 to Site2:

- It resets production systems in Site1.
- It reconfigures the secondary disks.
- It activates CBU for servers in Site2.
- It switches couple data sets to those in Site2.
- It activates partitions in Site2.
- It reIPLs P1, P2, and P3 in Site2.

GDPS performs the following actions to move the non-stretched SA AppMan Linux cluster from Site1 to Site2:

- It makes available those Metro Mirror secondary devices in Site2 (managed by Open LUN) that are associated with the Linux cluster.
- It sends a Reset Site1 command to SA AppMan to power off all distributed production systems in Site1.
- It 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) may 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 System z resources and the AppMan clusters, when there is an unplanned outage that affects Site1.

7.3.4 Summary

The Distributed Cluster Management function of GDPS can provide a single point of control to monitor and manage both System z resources and distributed server resources. DCM can also provide coordinated failover for planned and unplanned events that can affect either the System z resources or the distributed server resources or both. In short, you can attain business resiliency across your entire enterprise.

8

Combining Local/Metro continuous availability with out-of-region disaster recovery

In this chapter, for customers that have requirements for both continuous availability locally and regional disaster recovery protection, we discuss the capabilities and considerations for implementing GDPS/Metro Global Mirror (GDPS/MGM) and GDPS/Metro z/OS Global Mirror (GDPS/MzGM).

GDPS/MGM and GDPS/MzGM combine the continuous availability attributes of GDPS/PPRC with the out-of-region disaster recovery capabilities of GDPS/GM or GDPS/XRC to protect critical business data in the event of a wide-scale disruption, while providing for fast automated recovery in the event of any failure.

> **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 facilitate readability, only GDPS/PPRC will be used in the text for most of the discussions. If a particular function is not supported by GDPS/PPRC HyperSwap Manager, it will be mentioned.

Functions provided by these technologies include:

- Three-copy disk mirroring using GDPS/PPRC 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 in the event of a wide-scale disruption.
- Multi-site management of the remote copy environment to maintain data integrity and data consistency across all three disk copies.
- Support transparent switching to secondary disks in the event of a primary disk storage subsystem failure using GDPS/PPRC with 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 System z using GDPS/PPRC and GDPS/GM, assuming that only one site is lost in the event of 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.

8.1 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 utilize 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.

8.2 Design considerations

In the following sections we describe design considerations to keep in mind, including three-copy solutions versus three-site solutions; multi-target and cascading topologies; and cost considerations.

8.2.1 Three-copy solutions versus three-site solutions

It is not always the case that customers implementing a three-copy mirroring solution will have three independent data centers (shown in Figure 8-1), each with the capability to run production workloads.

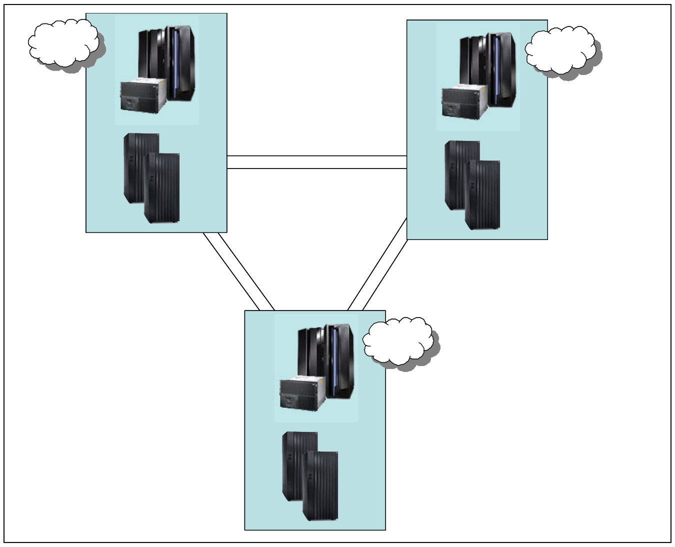


Figure 8-1 Three-site solution

Having three distinct locations with both the connectivity required for the replication and connectivity for user access is expensive and may not provide sufficient cost justification. Additionally, 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 8-2), along with a third copy of the data at a remote recovery location, provides you with many of the benefits of a full three-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.

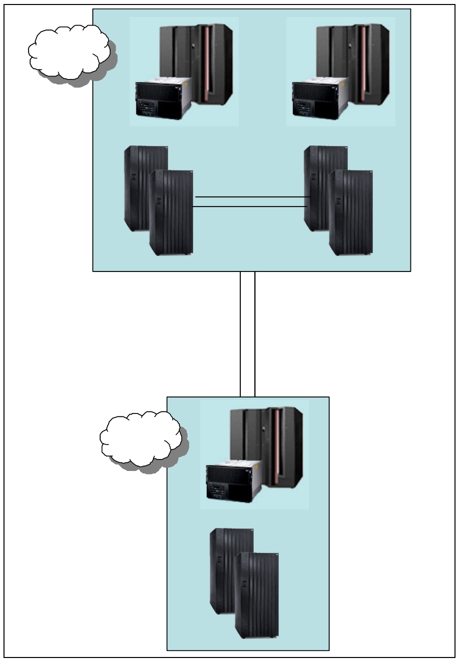


Figure 8-2 Two-site solution

Another benefit of the two-data center solution, especially in a System z environment, is that you can take full advantage 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 8-3 on page 177 illustrates another variation of this scenario, in which the primary data center is some type of campus location with separate machine rooms or buildings, each with the ability to run production workloads.

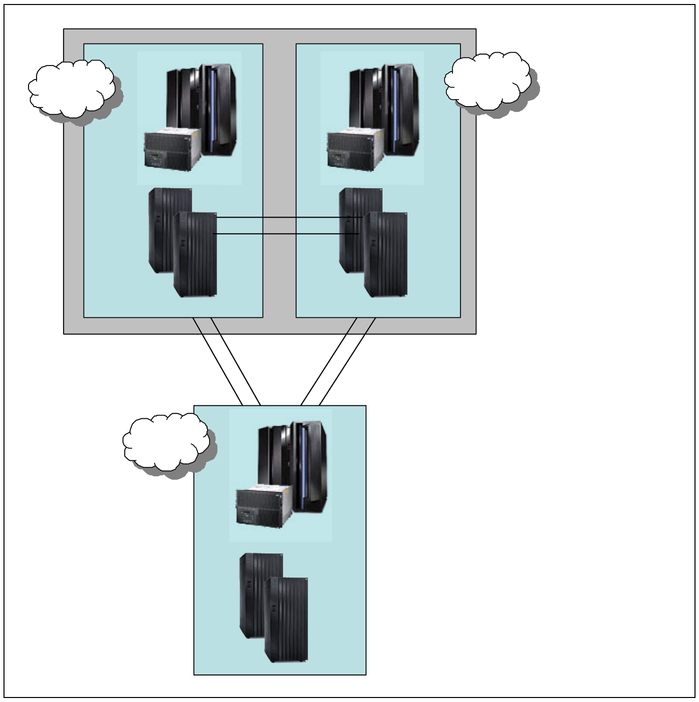


Figure 8-3 Two-site solution - Campus and Recovery site

In the past customers often used the bunker topology (shown in Figure 8-4) to create a solution that could provide mirroring at extended distances but still handle a primary site failure without data loss.

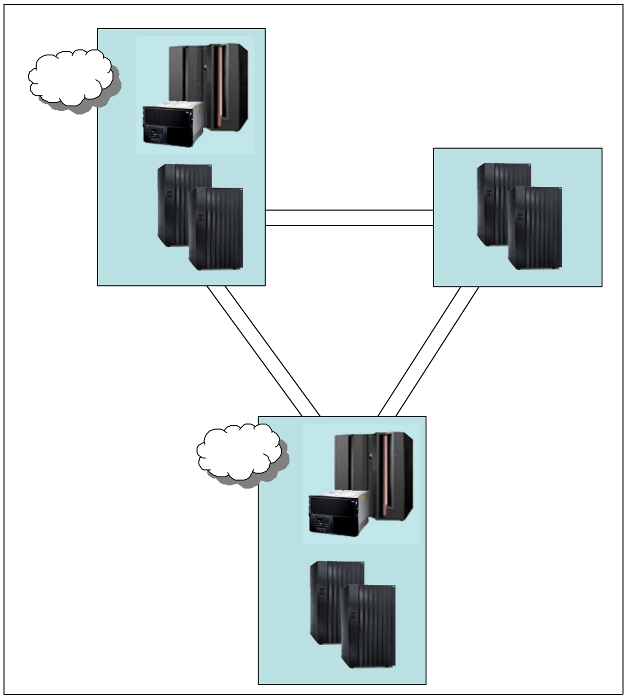


Figure 8-4 Two sites and an intermediate bunker

There are a number of arguments against this approach:

1. For guaranteed zero data loss you need a policy in which, if the mirroring stops, the production applications are also stopped. There are customers 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.
2. 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.
3. 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.

8.2.2 Multitarget and cascading topologies

Multitarget 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:

1. A viable recovery copy/capability is available at all times in a location other than where production is running. It is possible that there will be regulatory requirements that demand this.
2. Any single site failure will only result in 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.

The first requirement implies that there are no situations where both offsite copies would be compromised.

The second requirement makes it extremely desirable to have a capability to perform incremental resynchronization between any two copies of the data. Not having this would result in an extended period of exposure to additional data loss in case of a second failure.

8.2.3 Cost considerations

The third location is, in many situations, regarded as an insurance copy and as mainly providing regulatory compliance. This may imply that costs for this location are kept to an absolute minimum.

Reducing the network bandwidth to the remote location can provide significant cost savings for the overall cost of the solution. Given that a synchronous copy is already available, trading off the RPO versus the cost of the network may be a good 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. Shared hardware assets and the removal of the requirement to invest in an additional physical location 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.

8.3 GDPS Metro/Global Mirror solution

GDPS provides two “three-site” solutions:

- GDPS Metro/Global Mirror (GDPS/MGM) is a cascading data replication solution for both System z and distributed systems data.
- GDPS Metro/z/OS Global Mirror (GDPS/MzGM) is a multitarget data replication solution for System z data.

This section describes the capabilities and requirements of the GDPS Metro/Global Mirror (GDPS/MGM) solution.

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

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 with GDPS/PPRC or GDPS/PPRC HyperSwap Manager. GDPS/GM is used to asynchronously replicate data from the secondary disks to a third disk subsystem located 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 FBA devices can be mirrored to the recovery site, thereby protecting both System z and open system data.

For enterprises that require consistency across both distributed systems and System z 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 in the event of a wide-scale disruption.

8.3.1 GDPS/MGM overview

The GDPS/MGM configuration shown in Figure 8-5 on page 180 is a three-site continuous availability and DR solution. In this example, Site1 and Site2 are running an active/active workload (refer to 3.2.3, “Active/active configuration” on page 55) and are located within metropolitan distances to ensure optimal application performance. All data required to recover critical workloads is resident on disk and mirrored. Each site is configured with sufficient spare capacity to handle failed-over workloads in the event of a site outage.

The third site, or recovery site, can be located at virtually unlimited distance from Site1 and 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 for continued disaster recovery protection in the event of a Site2 disaster, or a failure of the disk subsystems in Site2. See “Incremental resynchronization for GDPS/MGM” on page 180 for more details. There is sufficient CPU capacity installed to support the R-sys. CBU is installed and GDPS will invoke CBU on System z to provide the additional capacity needed to support production workloads in the event disaster recovery is invoked.

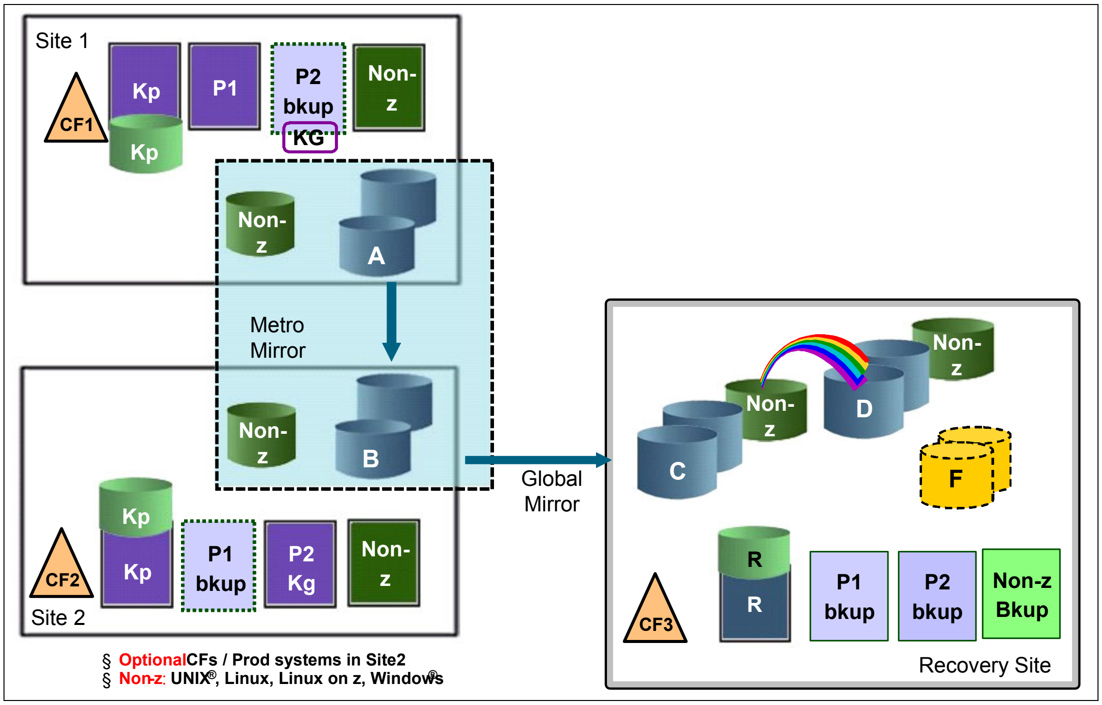


Figure 8-5 GDPS Metro Global Mirror configuration

The *A disks* are synchronously mirrored to the *B disks* located in Site2 using Metro Mirror. The *B disks* are then asynchronously mirrored to a third (C) set of disks located in the recovery site using Global Mirror. A fourth (D) set of disks, also located in the recovery site, are the FlashCopy targets used to provide the consistent data 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 “gold” or insurance copy of the data For more detail information about Global Mirror, refer to Chapter 6, “GDPS/Global Mirror” on page 123.

Incremental resynchronization for GDPS/MGM

The incremental resynchronization functionality of Metro Global Mirror aims to allow for 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 supports a configuration where only incremental changes must be copied from Site 1 to the recovery site.

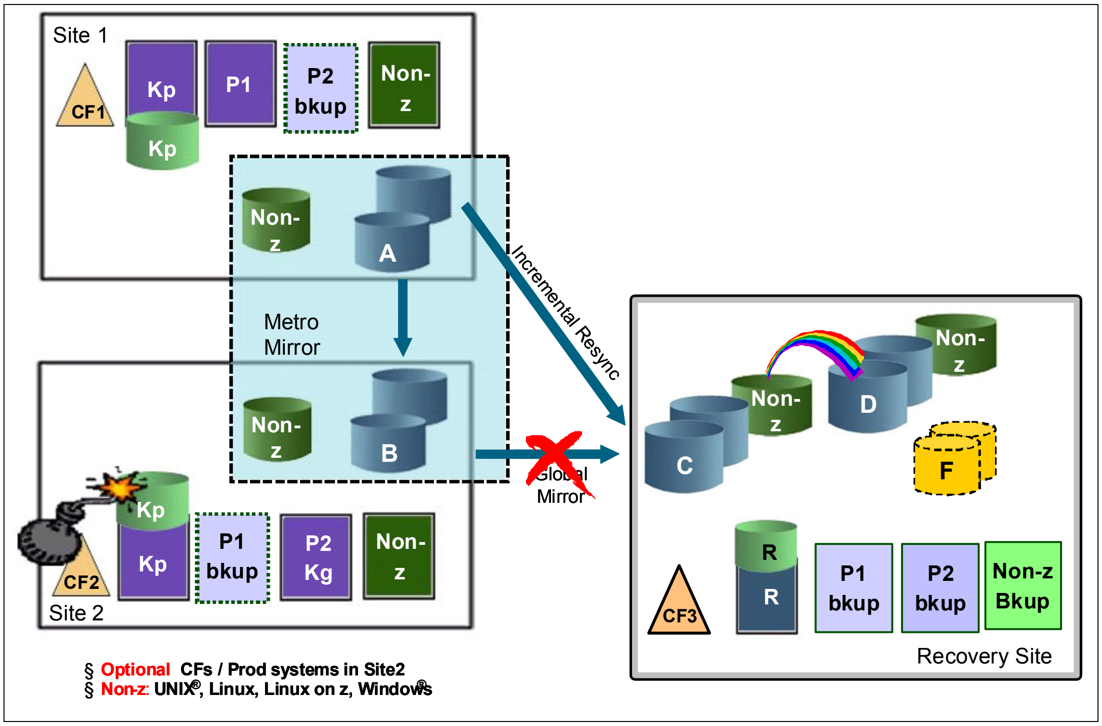


Figure 8-6 GDPS Metro Global Mirror configuration after Site2 outage

Figure 8-6 shows how GDPS/MGM can establish a Global Mirror session between the production site, Site1, and the recovery site when it detects that the intermediate site is unavailable. 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.

8.3.2 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 will freeze all *B disks*, both CKD and FBA, to prevent logical contamination of data residing on the B devices. For a more detailed description of GDPS/PPRC processing, refer to Chapter 3, “GDPS/PPRC” on page 43.

At this point, the GDPS/GM session between Site2 and the recovery site is still running and it is most likely that both locations will have the same set of data after a very short period 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.

8.3.3 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” on page 180 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 would depend on the specific environment.

8.3.4 Other considerations in a GDPS/MGM environment

With Global Mirror, it is possible to deliberately underconfigure the bandwidth provided to reduce the total cost of the solution. If there are significant peaks, then this cost saving could be considerable because the network costs are often a significant portion of ongoing costs.

8.3.5 Managing the GDPS/MGM environment

GDPS provides a range of solutions for disaster recovery and continuous availability in a System z-centric environment. GDPS/MGM provides support for Metro Global Mirror within a GDPS environment. GDPS builds on facilities provided by System Automation and NetView, and utilizes inband connectivity to manage the Metro Global Mirror relationships.

GDPS/MGM runs two different services to manage Metro Global Mirror, both of which run on z/OS systems, as explained here:

- GDPS/PPRC services run on every z/OS image in the production sysplex and the controlling systems, K1 and K2, located 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 and will take over the master function if the Master controlling system becomes unavailable.

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 in the event of a wide-scale disruption that impacts both Site1 and Site2.

As well as managing the operational aspects of Global Mirror, GDPS/GM also provides facilities to restart System z production systems in the recovery site. By providing scripting facilities, it provides a complete solution for the restart of a System z environment, in a disaster situation, without requiring expert manual intervention to manage the recovery process.

GDPS supports both System z and distributed systems’ devices in a Metro Global Mirror environment. However, GDPS requires that the disk subsystems be shared between the System z and distributed systems environments, because it requires CKD device addresses to issue the commands to manage the environment.

8.3.6 Flexible testing in a GDPS/MGM environment

To facilitate testing of site failover and failback processing, you may 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 supports an additional FlashCopy disk device, referred to as *F disks*. These are additional FlashCopy target devices that may optionally be created in the recovery site. The F disks may be used to facilitate standalone 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 supports the definition and management of a single F device for each Metro-Global Mirror *triplet* (B, C, and D disk combinations) in the configuration. To reduce management and operational complexity, support was added to 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.

8.3.7 Prerequisites for GDPS/MGM

GDPS/MGM has the following prerequisites:

- GDPS/PPRC or GDPS/PPRC HM is required. If GDPS/PPRC HM is used, the incremental Resynchronization function is not available.
- 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, refer to the GDPS product Web site, available at:

http://www.ibm.com/systems/z/advantages/gdps/getstarted

8.4 GDPS Metro z/OS Global Mirror solution

GDPS provides two “three-site” solutions:

- GDPS Metro/Global Mirror (GDPS/MGM) is a cascading data replication solution for both System z and distributed systems data.
- GDPS Metro/z/OS Global Mirror (GDPS/MzGM) is a multitarget data replication solution for System z data.

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 multitarget 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 located in a recovery site, typically out of the local metropolitan region. Because z/OS Global Mirror (XRC) only supports CKD devices, only System z data can be mirrored to the recovery site.

For enterprises looking to protect System z 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 in the event of a wide-scale disruption.

8.4.1 GDPS/MzGM overview

The solution depicted in Figure 8-7 is an example of a three-site GDPS/MzGM continuous availability and DR implementation. In this example, Site1 and Site2 are running an active/active workload (refer to 3.2.3, “Active/active configuration” on page 55) and located within metropolitan distances to ensure optimal application performance. All data required to recover critical workloads is resident on disk and mirrored. Each site is configured with sufficient spare capacity to handle failed-over workloads in the event of a site outage.

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 for continued data protection and DR capabilities in the event of a Site1 disaster, or a failure of the disk subsystems in Site1. See “Incremental resynchronization for GDPS/MzGM” on page 184 for more details. Sufficient mainframe resources are allocated to support the SDMs and GDPS/XRC Controlling system. In the event of a disaster situation, GDPS will invoke CBU to provide the additional capacity needed to support production workloads.

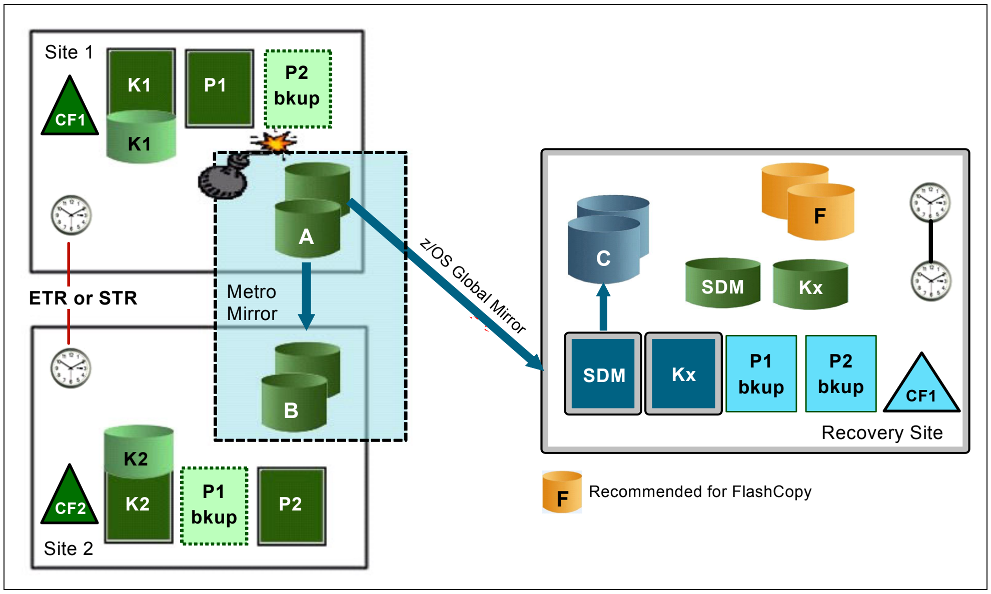


Figure 8-7 GDPS z/OS Metro Global Mirror

The A disks are synchronously mirrored to the B disks located in Site2 using Metro Mirror. In addition, A disks are asynchronously mirrored to a third (C) set of disks located in the recovery site using z/OS Global Mirror (XRC). An optional, and highly recommended, fourth (F) set of disks located in the recovery site are used to create FlashCopy of the C disks. These disks can then be used for stand-alone disaster recovery testing, or in the event of a real disaster, to create a “gold” or insurance copy of the data. For more detailed information about z/OS Global Mirror, refer to Chapter 5, “GDPS/XRC” on page 107.

Incremental resynchronization for GDPS/MzGM

The incremental resynchronization functionality of Metro z/OS Global Mirror aims to allow for incremental resynchronization between Site2 and the recovery site, when Site1, or the disk subsystems in Site1, are not available. Without incremental resynchronization, if Site1 becomes unavailable the data at the recovery site starts to age because data could no longer be replicated. The disaster recovery capability could be restored by establishing a new z/OS Global Mirror 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 very large configurations.

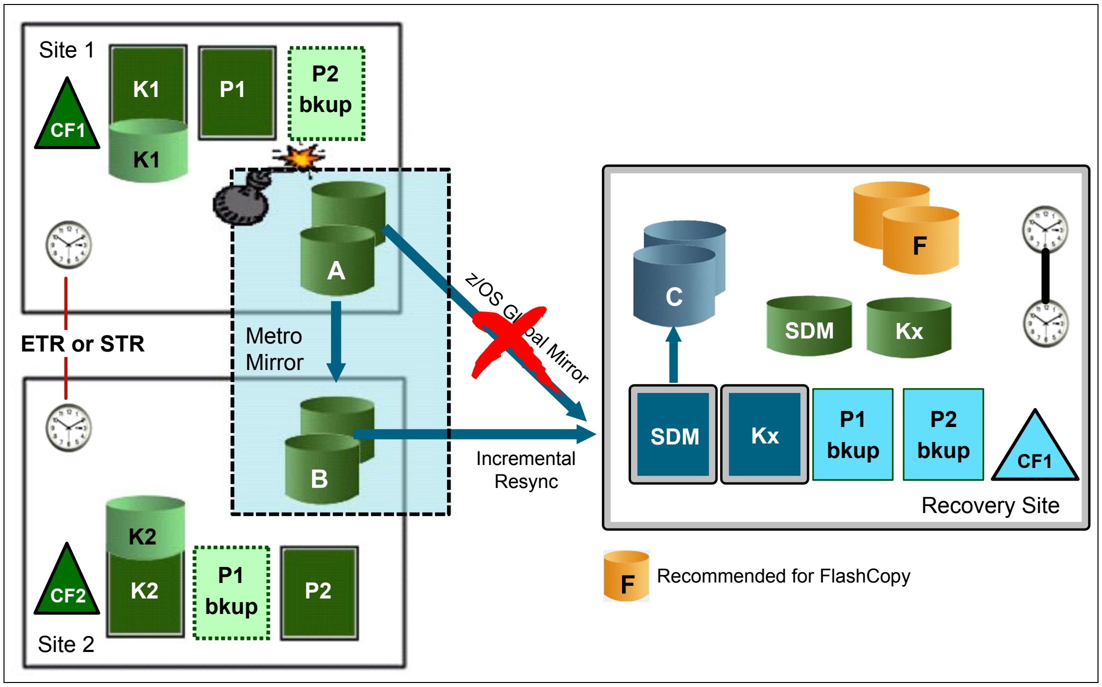


Figure 8-8 GDPS Metro z/OS Global Mirror configuration after Site1 outage

Figure 8-8 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. 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.

8.4.2 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, refer to Chapter 3, “GDPS/PPRC” on page 43.

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.

Note that 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.

8.4.3 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.

8.4.4 Other considerations in a GDPS/MzGM environment

In the event of a regional disaster where both Site1 and Site2 are lost, production will be moved to the recovery location. At some point it will become necessary to move production workloads back to either Site1 or Site2. To move data back to either site, the z/OS Global Mirror (XRC) remote copy environment must be designed to allow the mirroring session to be reversed. This means ensuring that the proper connectivity and resources are configured in the alternate sites to support the SDMs.

8.4.5 Management of 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 8.3.5, “Managing the GDPS/MGM environment” on page 182, are also provided by GDPS/MzGM.

GDPS/XRC services run on the Kx Controlling system located 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 has no awareness of 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 will activate the production LPARs and Coupling Facilities, and is able to respond to certain z/OS initialization messages. However, it cannot automate the complete startup of the production systems. For this, the K1 or K2 systems could be used to automate the application startup and recovery process in the production sysplex.

8.4.6 Flexible testing of the GDPS/MzGM environment

To facilitate testing of site failover and failback processing, you may 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 now 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 SSIDs 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 may 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 now have the flexibility to perform on-demand DR testing and other nondisruptive activities, while maintaining up-to-date DR readiness.

8.4.7 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 the latest GDPS prerequisite information, refer to the GDPS product Web site, available at:

http://www.ibm.com/systems/z/advantages/gdps/getstarted

9

Sample continuous availability and disaster recovery scenarios

In this chapter we describe a number of common customer scenarios and requirements, along with what we believe to be the most suitable solution for each case.

The scenarios we discuss are:

- A customer with a single data center that has already implemented Parallel Sysplex with data sharing and workload balancing wishes to move to the next level of availability.
- A customer with two centers needs a disaster recovery capability that will permit application restart in the remote site following a disaster.
- A customer 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 customer with two sites at continental distance needs to provide a disaster recovery capability.

9.1 Introduction

In the following sections we describe how the various GDPS service offerings can address different continuous availability (CA) and disaster recovery (DR) requirements. As every business is unique, the following sections are not a complete list of all the ways the offerings can address your specific business' needs, but they do serve to illustrate key capabilities.

Please note that in the figures accompanying the text we show minimal configurations for clarity. Many customer configurations would be more complex than this, but both configurations are obviously supported.

9.2 Continuous availability in a single data center

In the first scenario the customer has just one data center, but wishes to have higher availability. The customer has already implemented data sharing for their critical applications, and exploits 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 could take down all their applications. The configuration is shown in Figure 9-1.

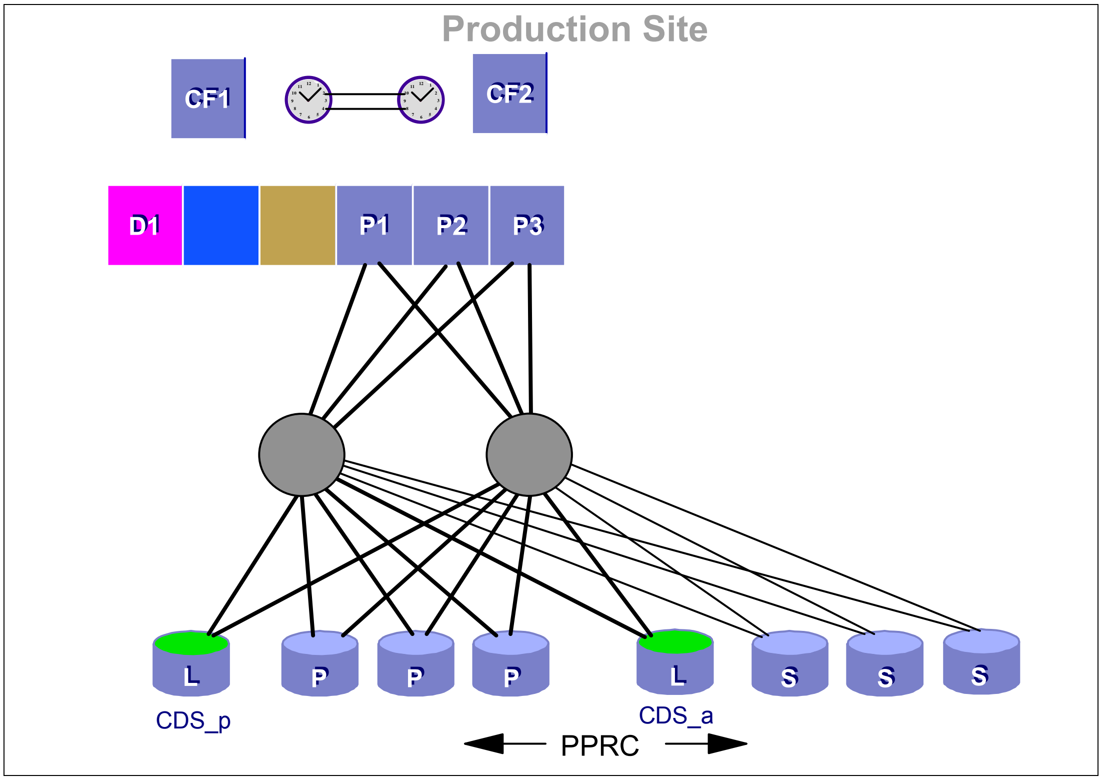


Figure 9-1 Data sharing, workload balancing, mirroring - single site

From a disaster recovery perspective, the customer 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 would like 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 9-1 summarizes the customer's situation and requirements, and shows which of those requirements can be addressed by the most suitable GDPS offering for this customer's requirements—GDPS/PPRC HyperSwap Manager.

Table 9-1 Mapping customer 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(1)
Ability to easily move to GDPS/PPRC in the future	Y
Note 1: 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 customer 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 of the need for HyperSwap, RCMF/PPRC is not suitable for them. And, because they only have a single site, do not currently have a VTS, and do not currently have the time to properly implement GDPS system and resource management, the full GDPS/PPRC offering is more than they need right now.

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 9-2 on page 192 shows the customer 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 the systems in the GDPS sysplex over to what were previously the secondary volumes. The darker lines connecting the secondary volumes in the picture indicate that the processor-to-control unit channel capacity is now similar to that used for the primary volumes.

.

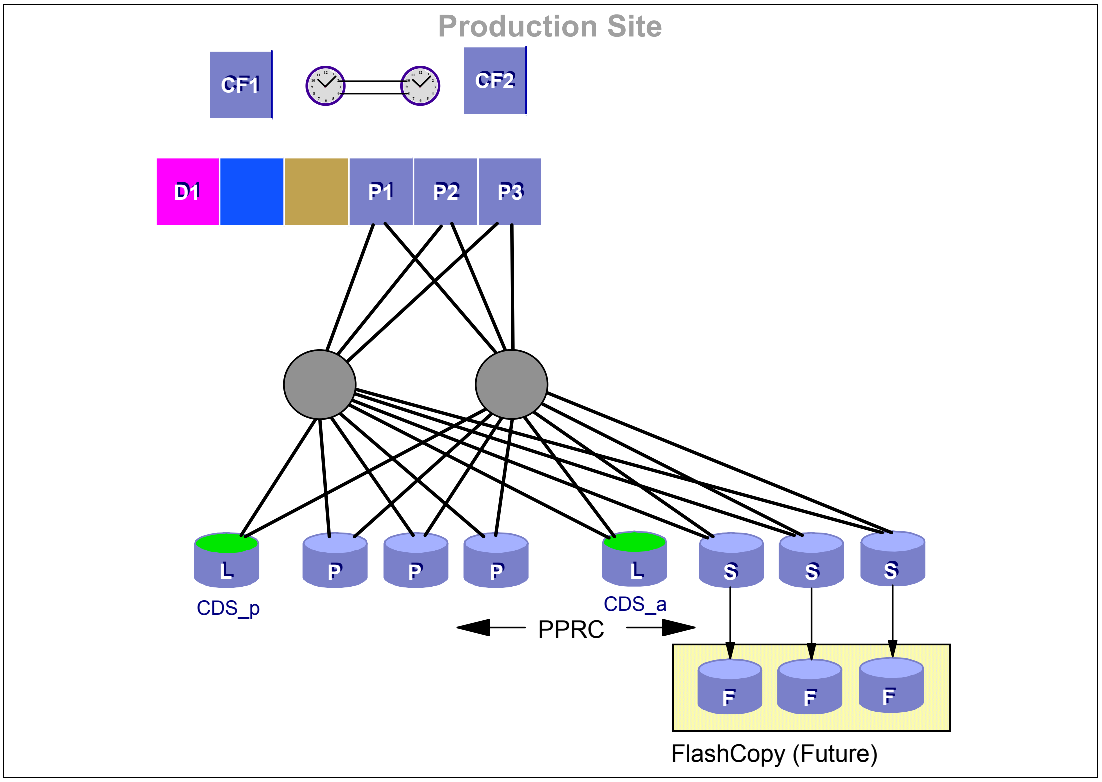


Figure 9-2 Continuous Availability within a single data center

After the customer has implemented GDPS and enabled the HyperSwap function, the customer's next move will be to install the additional disk capacity so it can use FlashCopy. The customer will then be able to use the Freeze function to create a consistent view that can be FlashCopied to create a set of volumes that can then be full-volume dumped for disaster recovery. This will not only create a more consistent set of backup tapes than the customer has today (because today it is backing up a running system), but also the backup window will now only be a few seconds rather than the hours it is currently taking. This will allow the customer to take more frequent backups should it want to do so.

9.3 DR across two data centers at metro distance

The next scenario relates to a customer that is under pressure to provide a disaster recovery capability in a very short time frame, perhaps for regulatory reasons. The customer 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 customer 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 customer 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. RCMF/PPRC is not an option in this case because it does not provide the Freeze capability. Because they will not be utilizing the additional capabilities delivered by GDPS/PPRC in the immediate future, the customer decides to implement the lower cost GDPS/PPRC HyperSwap Manager option. Table 9-2 on page 193 summarizes the customers situation and requirements, and shows how those requirements can be addressed by GDPS/PPRC HyperSwap Manager.

Table 9-2 Mapping customer 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(1) (FlashCopy)
Ability to move to GDPS/PPRC in the future	Y
Note 1: FlashCopy is used to create a consistent set of secondary volumes prior to a resynchronization, following a suspension of remote copy sessions.	

This customer 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 customer also has a requirement to have a consistent restart point at *all* times (even during DR testing).

This customer will implement GDPS/PPRC HyperSwap Manager, with the Control system in the primary site, and the secondary disks in the remote site. The secondary storage subsystems are configured with sufficient capacity to be able to use FlashCopy for the secondary devices; this will allow the customer 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 customer will install the additional intersite channel bandwidth required to be able to HyperSwap between the sites. This configuration is shown in Figure 9-3. Later, in preparation for a move to full GDPS/PPRC, the customer will move the Control system (and its disks) to the remote site.

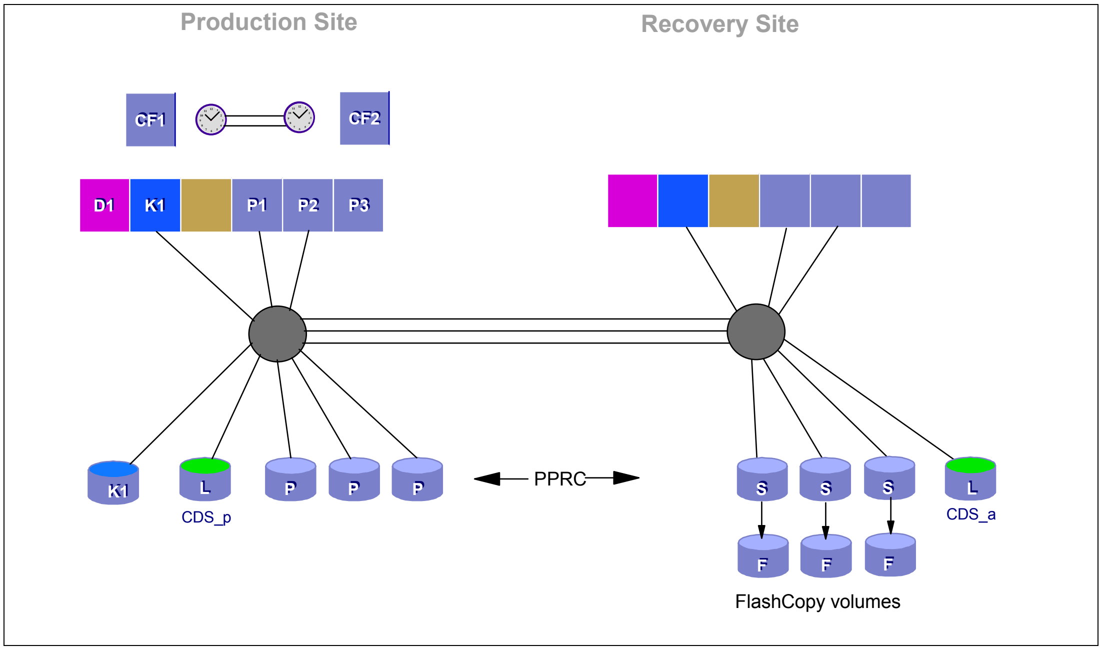


Figure 9-3 GDPS/PPRC two-site HM configuration

9.4 DR and CA across two data centers at metro distance

The customer in the next scenario has two centers within metro distance of each other. The customer already uses PPRC to remote copy the primary disks (both CKD and Open LUNs) to the second site. It also has the infrastructure in place for a cross-site sysplex; however, all the production work still runs in the systems in the primary site.

The customer 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 not only disaster recovery, but also 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 customer is already doing remote mirroring, its 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 customer 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 customer’s next area of focus will be continuous availability, across both planned and unplanned outages of applications, systems, and complete sites.

The customer is also investigating the use of z/VM and Linux for zSeries to consolidate some of its thousands of PC servers onto the mainframe. However, this is currently a lower priority than its other tasks.

Because of the disaster recovery and continuous availability requirements of this customer, together with the work it has already done and the infrastructure it has in place, the obvious GDPS offering for it is GDPS/PPRC. Table 9-3 shows how this offering addresses this customer’s needs.

Table 9-3 Mapping customer 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(1) (FlashCopy)
Remote copy and remote consistency support for Open LUN 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
Note 1: FlashCopy is used to create a consistent set of secondary volumes prior to a resynchronization, following a suspension of remote copy sessions.	

Although this customer has performed a significant amount of useful work already, fully exploiting the capabilities of GDPS/PPRC is going to take a significant amount of time, so the project has been broken up as follows:

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 remote copy management facilities of GDPS/PPRC becomes familiar, enable the Freeze capability, initially as Freeze=Go, then moving to Freeze=Stop as soon as the customer is confident with the stability of the remote copy infrastructure.

   While the customer has PPRC today, it does not have the consistency on the remote disks that is required to do a restart rather than a recovery following a disaster. The GDPS Freeze capability will add this consistency, and enhance this with the ability to ensure zero (0) data loss following a disaster when a Freeze=Stop policy is implemented.

3. Add the Open LUN 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 Open LUN disks, a second Control system will be set up[1].

   Although the customer does not currently have transactions that update both the z/OS *and* Open disks, the ability to Freeze all disks at the same point in time makes cross-platform recovery significantly simpler.

   In the future, if the customer 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 "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.

   The GDPS system and sysplex management capabilities are a very 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 customer sometimes has outages as a result of planned maintenance to its primary site. Using the scripts, the customer can exploit HyperSwap to keep its applications available as it moves its systems

[1] Only the GDPS Control system can see the Open LUN disks. So a second Control system is recommended to ensure the Open LUNs can always be managed, even if a Control system is down for some reason.

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. IBM recommends that a Control system be set up in each site when using HyperSwap, to ensure there is always a system available to initiate a HyperSwap, regardless of where the primary disks may be at that time. In the case of this customer, it has already set up the second Control system when it added the Open LUN devices to the GDPS configuration.

   The customer will exploit both Planned HyperSwap (to move its primary disks prior to planned maintenance on the primary subsystems) and Unplanned HyperSwap (allowing the customer to continue processing across a primary subsystem failure).

7. Finally, and assuming the consolidation onto Linux for zSeries has proceeded, the Heterogeneous Disaster Recovery capability will be implemented to add HyperSwap support for Linux guests, and to enable the use of errors on Linux disks to act as triggers for a Freeze or HyperSwap.

   Although HyperSwap support is not available for users of Open LUN devices (because HyperSwap requires changes to the operating system using the device), it *is* available to Linux guests running under z/VM Version 5.1 or later. This means that not only do you get Freeze support for those Linux guests, but you also have the ability to swap the Linux disks from the primary to secondary disk subsystem, with only a temporary pause to the processing of those guests. And because it is all managed by the same GDPS, the swap could 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 Linux disk could result in a HyperSwap of the Linux disks and the z/OS disks.

The projected final configuration is shown in Figure 9-4 on page 197 (in the interests of clarity, we have not included the Linux components in the picture).

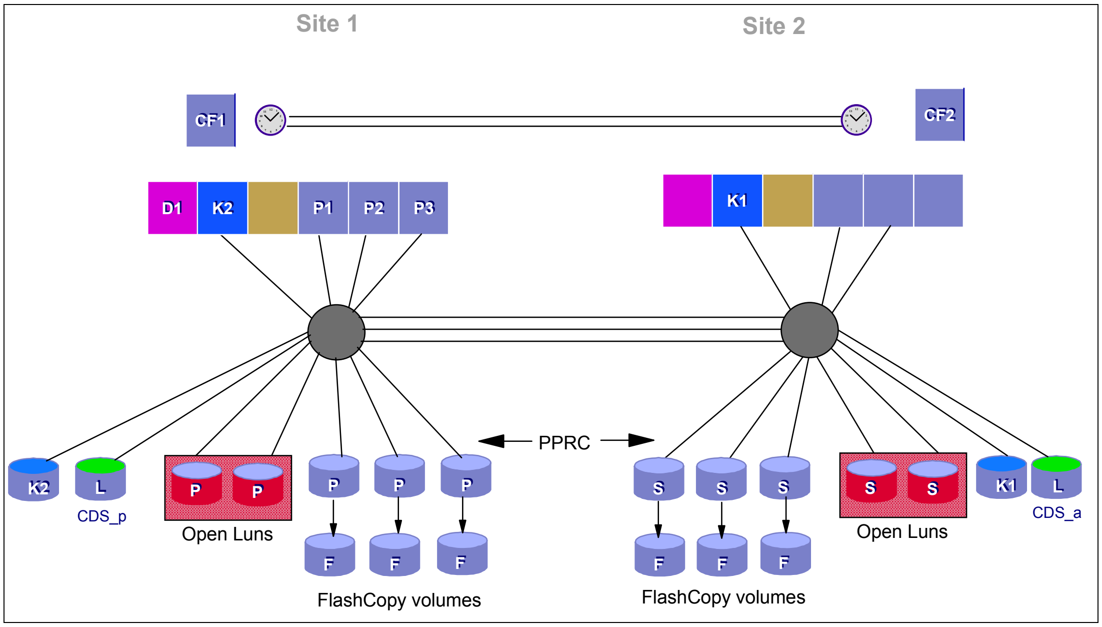


Figure 9-4 Active/standby workload GDPS/PPRC configuration

9.4.1 Active/active workload

As said previously, this customer 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 customer plans to move to an active/active workload configuration, with some production systems in the primary site, and others in the “recovery” site.

To derive the maximum benefit from this configuration, it should be 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 multi-site data sharing and HyperSwap means that the customer’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 could potentially result in a temporary application outage is an instantaneous outage of all resources in the primary site; this could result in the database managers in the recovery site having to be restarted.

The move to an active/active workload may require creating some minor changes to the GDPS definitions, some 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. However, apart from that, there is no fundamental change in the way GDPS is set up or operated.

9.5 DR in two data centers, global distance

The customer in this scenario has a data center in Asia, and another in Europe. Following the tsunami disaster in 2004, the customer decides to remote copy its production sysplex data to its data center in Europe. The customer 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 customer has about 10,000 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 customer is GDPS/XRC. Due to the large distance between the two sites (approaching 15000 km), a synchronous remote copy method would be out of the question. Because of the large number of volumes, multiple SDMs will be required to manage the remote copy, and coupled SDMs are not supported by RCMF/XRC. Table 9-4 shows how the customer's configuration and requirements map to the capabilities of GDS/XRC.

Table 9-4 Mapping customer 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 a number of factors, most notably the available bandwidth)
Maintain consistency of secondary volumes	Y
Maintain consistency of secondary volumes during resynch	Y(1) (FlashCopy)
Over 10,000 volumes	Y (exploit coupled SDM support)
Requirement for data replication for and between multiple storage vendors products	Y
Only z/OS disks need to be mirrored	Y
Automated recovery of disks and systems following a disaster	Y(GDPS script support)
Note 1: FlashCopy is used to create a consistent set of secondary volumes prior to a resynchronization, following a suspension of remote copy sessions.	

The first step for the customer 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 customer 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 Control system LPAR). This will allow the system 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 9-5 on page 199. The GDPS systems are in the same sysplex and reside on the same processor as 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.

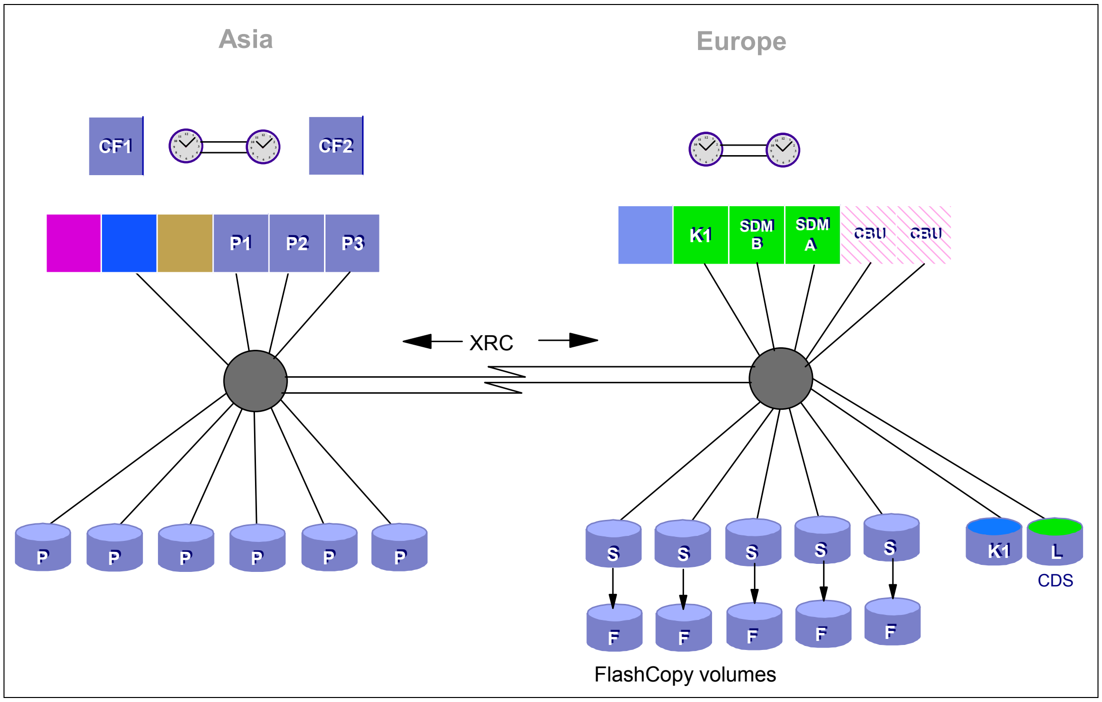


Figure 9-5   Final GDPS/XRC configuration

9.6 Other configurations

There are, of course, 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.

A

RCMF/PPRC

In this chapter we discuss the capabilities and prerequisites of the RCMF/PPRC offering.

The RCMF/PPRC offering is a small subset of the function provided by GDPS/PPRC and GDPS/PPRC HM, providing just the ability to manage your remote copy configuration using a set of NetView panels. It is not a disaster recovery offering because it does not include the Freeze function and therefore cannot guarantee consistency across multiple LSSs. It also does not provide any of the system management capabilities delivered by GDPS/PPRC. It is intended purely as a tool to manage your remote copy environment, while providing a similar user interface should you decide to implement full GDPS/PPRC or GDPS/PPRC HM in the future.

Introduction

RCMF/PPRC is a facility that provides the user a single point of control to manage a PPRC remote copy configuration. It simplifies the storage administrator's remote copy management functions by managing the remote copy configuration, rather than individual remote copy pairs. A high-level user interface is provided in the form of a NetView interface that allows the user to manage the remote copy configuration without having to use the complex syntax of TSO commands. By providing a high level interface, the potential for making mistakes is reduced.

The RCMF/PPRC offering has been functionally stabilized such that no significant features are expected to be added in future releases.

How an RCMF/PPRC configuration looks

The RCMF/PPRC offering differs from the GDPS offerings in that there is no hard requirement for a Control system with RCMF/PPRC. Also, RCMF/PPRC does not have to run in every system in the sysplex, and in fact does not even have to be in the same sysplex as the systems using the primary disk. As a result, there is no need, from an RCMF/PPRC perspective, for a cross-site sysplex. However, because you will use RCMF to recover the secondary disk in case of a disaster, you would typically run RCMF in a system in the second site, using the disk in that site. If you need to switch to using the secondary disk then, the RCMF system will be available to drive the process of recovering the secondary disk back to simplex mode.

RCMF/PPRC functions

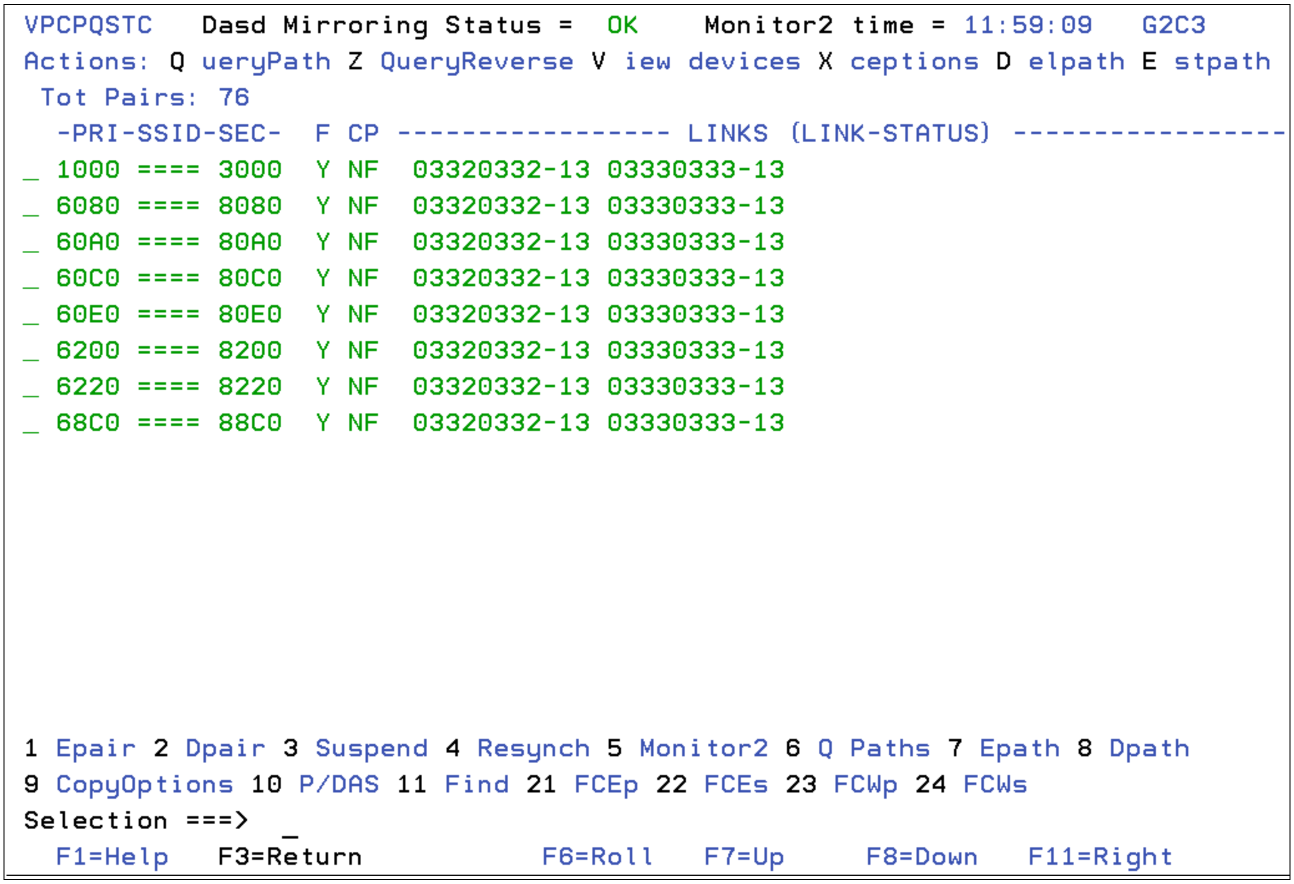
It is important to point out that RCMF is a productivity tool whose functions are executed only when initiated by the user. For instance, RCMF will provide user-initiated status and exception monitoring, but does not perform any kind of interval-driven automatic status monitoring. RCMF does not provide secondary data consistency because it does not deliver the Freeze function. RCMF provides the following functionality:

- Invoke the functions with a single key stroke
- Initiate functions per pair, subsystem, or all disks
- Automatically establish target configuration at system startup
- Add, move, remove pairs, subsystems, and links

To prevent accidental configuration changes, RCMF asks the user for a confirmation when invoking this type of function. A status change will be confirmed to the user by displaying the results of a query that is run after the configuration change is completed.

Figure A-1 on page 203 shows a sample panel, which displays the subsystem identifier (SSID) pairs that have initially been defined to be part of the RCMF/PPRC configuration, and the PPRC links between them. The panel is color-coded: when all SSID pairs are green, all the pairs are in duplex state. Behind each SSID pair, the PPRC link definitions for that pair are displayed.

The commands at the top of the panel can be executed for a selected SSID pair. The set of commands at the bottom of the panel, which are each identified by a number, when entered at the command prompt will be executed against all SSID pairs on the panel. Note that the same panel is shared between RCMF/PPRC and GDPS/PPRC, so some of the options shown (like FlashCopy Establish) are not enabled in the RCMF/PPRC offering.



```
VPCPQSTC   Dasd Mirroring Status =  OK    Monitor2 time = 11:59:09   G2C3
Actions: Q ueryPath Z QueryReverse V iew devices X ceptions D elpath E stpath
 Tot Pairs: 76
  -PRI-SSID-SEC-  F CP ----------------- LINKS (LINK-STATUS) -----------------
_ 1000 ==== 3000  Y NF  03320332-13 03330333-13
_ 6080 ==== 8080  Y NF  03320332-13 03330333-13
_ 60A0 ==== 80A0  Y NF  03320332-13 03330333-13
_ 60C0 ==== 80C0  Y NF  03320332-13 03330333-13
_ 60E0 ==== 80E0  Y NF  03320332-13 03330333-13
_ 6200 ==== 8200  Y NF  03320332-13 03330333-13
_ 6220 ==== 8220  Y NF  03320332-13 03330333-13
_ 68C0 ==== 88C0  Y NF  03320332-13 03330333-13









1 Epair 2 Dpair 3 Suspend 4 Resynch 5 Monitor2 6 Q Paths 7 Epath 8 Dpath
9 CopyOptions 10 P/DAS 11 Find 21 FCEp 22 FCEs 23 FCWp 24 FCWs
Selection ===> _
  F1=Help   F3=Return           F6=Roll   F7=Up     F8=Down   F11=Right
```

Figure A-1   RCMF/PPRC Mirroring information panel

Services component

Because RCMF/PPRC is a much more limited offering than GDPS/PPRC, the services included in the offering are similarly more limited.

An IBM services specialist will help you install RCMF/PPRC in one of your systems. The specialist will also work with you to set up some PPRC pairs, and provide knowledge transfer to a member of your staff as part of the process.

Additional services such as configuration planning, bandwidth analysis, and so on are available under a separate services agreement.

RCMF/PPRC prerequisites

RCMF/PPRC has the following prerequisites for the RCMF management system:

- Any generally supported level of z/OS.
- Any generally supported level of Tivoli NetView for z/OS.
- REXX Runtime Libraries. A set of REXX Runtime Libraries is shipped with RCMF for customers who do not already have the libraries installed on their system.
- There are no additional hardware requirements for RCMF beyond those required for PPRC.

Comparison of supported features

There are so many features and functions in the various members of the GDPS family, it is sometimes difficult to remember them all, and which offerings support them. To make it a little easier to position the offerings, Table A-1 lists the features and functions and which ones are delivered by the various GDPS and RCMF PPRC-based offerings.

Table A-1 Supported features matrix

	RCMF/PPRC	**GDPS/PPRC HyperSwap Manager**	**GDPS/PPRC**
Continuous availability	**No**	Yes (with HyperSwap)	Yes (with HyperSwap)
Disaster recovery	**No**	Yes (with Freeze)	Yes (with Freeze)
FlashCopy support	**No**	Yes	Yes
Production sysplex automation	**No**	No	Yes
Supported distance	**Up to 300 km, depending on location of RCMF system**	100 km	100 km
PtP VTS support	**No**	No	Yes
Span of control	**None**	Both sites (disk only)	Both sites
Open LUN	**No**	Yes	Yes
Multiplatform heterogeneous support	**No**	No	Yes
Monitoring and alerting	**Manually-initiated**	Yes	Yes
GDPS scripts capability	**No**	No	Yes
Web interface	**No**	Yes	Yes

B

RCMF/XRC

In this chapter we discuss the capabilities and prerequisites of the RCMF/XRC offering.

The RCMF/XRC offering is a subset of the function provided by GDPS/XRC, providing the ability to manage your remote copy configuration using a set of NetView panels and to create a consistent set of secondary volumes for disaster recovery purposes. It does not provide any of the system management or scripting capabilities delivered by GDPS/XRC. It is intended as a tool to manage your remote copy environment, while providing a similar user interface should you decide to implement full GDPS/XRC in the future.

RCMF/XRC

RCMF/XRC is a facility that provides the user a single point of control to manage an XRC remote copy configuration. It simplifies the storage administrator's remote copy management functions by managing the remote copy configuration, rather than individual remote copy pairs. Since in a XRC configuration, the consistency group processing performed by the SDM assures data consistency, RCMF/XRC can be considered as a valid DR option, even though it does not provide the full automation capabilities provided by GDPS/XRC.

A high-level user interface is provided in the form of a full panel interface that allows the user to manage the remote copy configuration without having to use the complex syntax of TSO commands. By providing a high-level interface, the potential for making mistakes is almost eliminated.

RCMF/XRC is a productivity tool with functions that are executed *only* when initiated by the user. For instance, RCMF will provide user-initiated status and exception monitoring, but does not perform any kind of interval-driven automatic status monitoring.

The RCMF/XRC offering has been functionally stabilized such that no significant features are expected to be added in future releases.

How an RCMF/XRC configuration looks

The RCMF/XRC offering differs from the GDPS/XRC offering mainly in that there is no requirement for a Control system with RCMF/XRC. RCMF typically runs in the SDM image and does not require any connectivity to the application site devices above and beyond that required for the SDM. The SDM system must be connected to each of the primary and secondary disk subsystems that it manages.

RCMF/XRC does not provide support for the Coupled SDM or Multi-SDM functions. The full GDPS/XRC offering is required for Coupled SDM support.

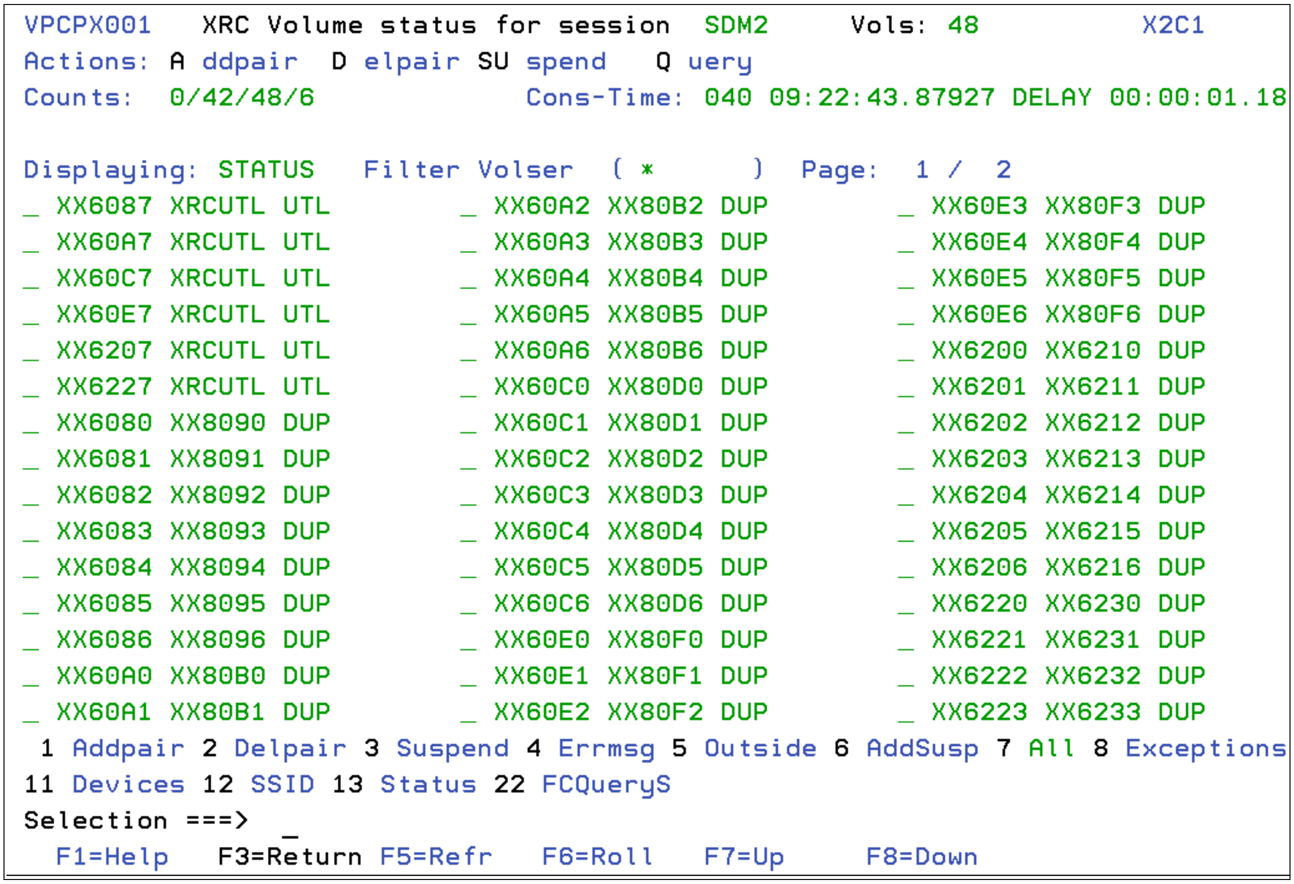
RCMF/XRC functions

RCMF/XRC provides the following functionality:

- Invoke functions with a single key stroke
- Initiate functions per pair, per subsystem, or for all devices
- Automatically establish the target configuration at system startup
- Add, move, remove pairs, subsystems, and links

To prevent accidental configuration changes, RCMF asks the user for a confirmation when invoking this type of function. A status change will be confirmed to the user by displaying the results of a query that is run after the configuration change is completed.

Figure B-1 on page 207 shows a sample RCMF/XRC NetView panel displaying all device pairs in the XRC session. The user has the option to expand the view to include device numbers (option 11) or SSIDs (option 12). The panel is color-coded: when all SSID pairs are green, all the pairs are in duplex state.



```
VPCPX001   XRC Volume status for session  SDM2     Vols: 48          X2C1
Actions: A ddpair  D elpair SU spend   Q uery
Counts:  0/42/48/6             Cons-Time: 040 09:22:43.87927 DELAY 00:00:01.18

Displaying: STATUS   Filter Volser  ( *       )  Page:  1 /  2
_ XX6087 XRCUTL UTL         _ XX60A2 XX80B2 DUP         _ XX60E3 XX80F3 DUP
_ XX60A7 XRCUTL UTL         _ XX60A3 XX80B3 DUP         _ XX60E4 XX80F4 DUP
_ XX60C7 XRCUTL UTL         _ XX60A4 XX80B4 DUP         _ XX60E5 XX80F5 DUP
_ XX60E7 XRCUTL UTL         _ XX60A5 XX80B5 DUP         _ XX60E6 XX80F6 DUP
_ XX6207 XRCUTL UTL         _ XX60A6 XX80B6 DUP         _ XX6200 XX6210 DUP
_ XX6227 XRCUTL UTL         _ XX60C0 XX80D0 DUP         _ XX6201 XX6211 DUP
_ XX6080 XX8090 DUP         _ XX60C1 XX80D1 DUP         _ XX6202 XX6212 DUP
_ XX6081 XX8091 DUP         _ XX60C2 XX80D2 DUP         _ XX6203 XX6213 DUP
_ XX6082 XX8092 DUP         _ XX60C3 XX80D3 DUP         _ XX6204 XX6214 DUP
_ XX6083 XX8093 DUP         _ XX60C4 XX80D4 DUP         _ XX6205 XX6215 DUP
_ XX6084 XX8094 DUP         _ XX60C5 XX80D5 DUP         _ XX6206 XX6216 DUP
_ XX6085 XX8095 DUP         _ XX60C6 XX80D6 DUP         _ XX6220 XX6230 DUP
_ XX6086 XX8096 DUP         _ XX60E0 XX80F0 DUP         _ XX6221 XX6231 DUP
_ XX60A0 XX80B0 DUP         _ XX60E1 XX80F1 DUP         _ XX6222 XX6232 DUP
_ XX60A1 XX80B1 DUP         _ XX60E2 XX80F2 DUP         _ XX6223 XX6233 DUP
 1 Addpair 2 Delpair 3 Suspend 4 Errmsg 5 Outside 6 AddSusp 7 All 8 Exceptions
11 Devices 12 SSID 13 Status 22 FCQueryS
Selection ===>  _
  F1=Help   F3=Return F5=Refr   F6=Roll   F7=Up     F8=Down
```

Figure B-1 RCMF/XRC Mirroring status panel

The commands at the top of the panel can be executed for a selected pair. The set of commands at the bottom of the panel, which are each identified by a number, when entered at the command prompt will be executed against all pairs on the panel.

A filtering capability exists that allows users to find a subgroup of the volumes in the session. Filtering applies to the current view. For example, if you select option 11, filtering operates on device numbers. With option 12 active, filtering operates on SSIDs. With option 13, it operates on volume serial numbers. Note that RCMF/XRC shares some panels with the GDPS/XRC offering, so some of the options on the panel are not supported by RCMF/XRC. For example, the option to query FlashCopy status is not enabled in RCMF/XRC.

Services component

Because RCMF/XRC is a more limited offering than GDPS/PPRC, the services included in the offering are similarly more limited.

An IBM services specialist will help you install RCMF/XRC in your SDM system. The specialist will also work with you to set up some XRC pairs, and provide knowledge transfer to a member of your staff as part of the process.

Additional services such as configuration planning, bandwidth analysis, and so on are available under a separate services agreement.

RCMF/XRC prerequisites

RCMF/XRC has the following prerequisites for the RCMF management system:

- Any generally supported level of z/OS.
- Any generally supported level of Tivoli NetView for z/OS.
- REXX Runtime Libraries. A set of REXX Runtime Libraries is shipped with RCMF for customers who do not already have the libraries installed on their system.
- There are no additional hardware requirements for RCMF beyond those required for XRC.

Comparison of supported features

There are so many features and functions in the various members of the GDPS family, it is sometimes difficult to remember them all, and which offerings support them. To make it a little easier to position the offerings, Table B-1 compares the most important features in RCMF/XRC and GDPS/XRC.

Table B-1 Comparison of RCMF/XRC and GDPS/XRC

	RCMF/XRC	**GDPS/XRC**
Continuous availability	**No**	No
Disaster recovery	**Yes**	Yes (with SDM)
FlashCopy support	**No**	Yes Zero Suspend FC
Coupled SDM support	**No**	Yes
Supported distance	**Virtually unlimited**	Virtually unlimited
Span of control	**N/A**	Recovery site
Monitoring and alerting	**Manually initiated**	Yes
GDPS scripts capability	**No**	Yes

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.

DFSMSdss. A functional component of DFSMS/MVS used to copy, dump, move, and restore data sets and volumes.

Disaster Recovery. 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. See also 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 comprised 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- 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

JCL. Job Control Language.

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. 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.

LSS. See 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. 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. 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 only accept 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 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.

Related publications

The publications listed in this section are considered particularly suitable for a more detailed discussion of the topics covered in this document.

IBM Redbooks publications

The following IBM Redbooks documents are referenced in this document:

- *IBM TotalStorage Virtual Tape Server: Planning, Implementing, and Monitoring,* SG24-2229
- *IBM zSeries Connectivity Handbook,* SG24-5444
- *Implementing ESS Copy Services on S/390,* SG24-5680
- *IBM TotalStorage Solutions for Disaster Recovery,* SG24-6547
- *IBM System Storage DS8000: Copy Services with IBM System z,* SG24-6787
- *Server Time Protocol Planning Guide*, SG24-7280
- *Server Time Protocol Implementation Guide*, SG24-7281

The following IBM Redpaper publications contain information about the DWDMs that are qualified for use with GDPS:

- *A Disaster Recovery Solution Selection Methodology,* REDP-3847
- *Planning for Disaster Recovery in a Heterogeneous Environment,* REDP-3848
- *zSeries Qualified WDM Vendor: Adva Optical Networking,* REDP-3903
- *zSeries Qualified WDM Vendor: Nortel Networks,* REDP-3904
- *zSeries Qualified WDM Vendor: Cisco Systems,* REDP-3905
- *zSeries Qualified WDM Vendor: Lucent Techologies,* REDP-3906

Other publications

These publications are also relevant as further information sources:

- *z/VM CP Planning and Administration*, SC24-6083
- *System z Capacity on Demand User's Guide,* SC28-6846
- *Tivoli NetView for OS/390 Installation and Administration Guide,* SC31-8236
- *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

Online resources

These Web sites are also relevant as further information sources:

The following page, on the IBM ResourceLink Web site contains a list of the qualified DWDM vendors:

https://www.ibm.com/servers/resourcelink/lib03020.nsf/pages/zseriesQualifiedExtendersAndWdmProductsForGdpsSolutions?OpenDocument&pathID=

Maintenance information for Advanced Copy Services is available at:

http://www.ibm.com/servers/storage/support/solutions/bc/index.html

Information white paper about system-managed CF structure duplexing:

http://www.ibm.com/servers/eserver/zseries/library/techpapers/gm130103.html

There are also a number of related Web sites that discuss various aspects of Disaster Recovery and Business Resilience:

http://www.ffiec.gov/ffiecinfobase/booklets/bcp/bus_continuity_plan.pdf
http://www.sec.gov/news/speech/spch050103mag.htm
http://www.dmreview.com/article_sub.cfm?articleId=1011020
http://www.omm.com/webdata/content/publications/client_alert_financial_services_2004_05_06.htm
http://www.sec.gov/news/studies/34-47638.htm
http://www.export.gov/safeharbor/
http://www.continuitycentral.com/news01588.htm
http://www.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP100825

How to get IBM Redbooks

You can search for, view, or download Redbooks, Redpapers, Hints and Tips, draft publications and Additional materials, as well as order hardcopy Redbooks or CD-ROMs, at this Web site:

ibm.com/redbooks

Help from IBM

IBM Support and downloads

ibm.com/support

IBM Global Services

ibm.com/services

Index

IBM Redbooks

GDPS Family - An Introduction to Concepts and Capabilities

(0.2”spine)
0.17”<->0.473”
90<->249 pages

GDPS Family - An Introduction to Concepts and Capabilities

Overview of business resilience requirements

Technologies that support IT resilience

Introduction to the GDPS family of offerings

This IBM Redbooks publication presents an overview of the GDPS family of offerings and the role they play in delivering a business IT resilience solution.

It begins with a discussion of general concepts of business IT resilience and disaster recovery along with some 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 each offering can achieve your defined goals for disaster recover and high availability. Also covered are the features that simplify and enhance data replication activities, the prerequisites for implementing each offering, and some hints for planning for the future as well as immediate business requirements. Tables provide an easy-to-use summary and comparison of the offerings, and the additional planning and implementation services available from IBM are explained.

Finally, a number of practical customer scenarios and requirements are described, along with the most suitable GDPS solution for each case.

The intended audience for this book includes Systems Programmers, Technical Support Managers, Operations Managers, Availability Managers, and Disaster Recovery Planners.

SG24-6374-06 ISBN 0738434426