

IBM XIV Storage System



Product overview

IBM XIV Storage System



Product overview

Note:

Before using this information and the product it supports, read the information in "Notices used in this document" on page ix and "Notices" on page 155.

This edition applies to version 10, release 2.4, of IBM XIV Storage System and to all subsequent releases and modifications until otherwise indicated in new editions.

© Copyright IBM Corporation 2008, 2011.

US Government Users Restricted Rights – Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Contents

Introduction	v
Purpose and scope	v
Document version	v
Intended audience	v
Publications and related information	v
Notices used in this document	ix
Document conventions	ix
Terms and abbreviations	ix
Getting information, help, and service	x
How to send your comments	x

Chapter 1. Overview: The IBM XIV

Storage System	1
Features and functionality	1
Hardware overview	1
Hardware components	1
Supported interfaces	3
Management options	4
Reliability	4
Redundant components and no single point of failure	5
Data mirroring	5
Self-healing mechanisms	5
Protected cache	5
Redundant power	5
Performance	6
Total load balance	6
Intelligent caching for improved performance	6
Functionality	7
Upgradability	8

Chapter 2. Volumes and snapshots

overview	11
The volume life cycle	12
Support for Symantec Storage Foundation Thin Reclamation	13
Snapshots	14
Redirect on write	14
Storage utilization	15
The snapshot auto-delete priority	15
Snapshot name and association	16
The snapshot lifecycle	16
Snapshot and snapshot group format	21

Chapter 3. Host system attachment

Balanced traffic and no single point of failure	23
Dynamic rate adaptation	23
Attaching volumes to hosts	24
Advanced host attachment	24
CHAP authentication of iSCSI hosts	24
Clustering hosts into LUN maps	25
Volume mappings exceptions	27
Supporting VMWare extended operations	28
Writing zeroes	28

Hardware-assisted locking	28
Fast copy	29
QoS performance classes	29
Updating the call home contact list	30
Host system attachment commands	31

Chapter 4. Consistency groups

overview	33
Creating a consistency group	33
Taking a snapshot of a consistency group	33
The snapshot group life cycle	35
Restoring a consistency group	36

Chapter 5. Storage pools overview

Protecting snapshots on a Storage Pool level	38
--	----

Chapter 6. Thin provisioning

Chapter 7. Target connectivity

Defining a remote target object	43
Adding ports to remote target	44
Connecting between local and target ports	44
Symmetric connectivity for mirroring	45

Chapter 8. Synchronous remote mirroring

mirroring	47
Remote mirroring basic concepts	47
Remote mirroring operation	48
Configuration options	49
Volume configuration	49
Communication errors	50
Coupling activation	50
Synchronous mirroring statuses	51
Link status	52
Operational status	52
Synchronization status	52
I/O operations	53
Synchronization process	54
State diagram	54
Coupling recovery	55
Uncommitted data	55
Constraints and limitations	55
Last-consistent snapshots	56
Secondary locked error status	58
Role switchover	58
Role switchover when remote mirroring is operational	58
Role switchover when remote mirroring is nonoperational	58
Resumption of remote mirroring after role change	60
Miscellaneous	62
Remote mirroring and consistency groups	62
Using remote mirroring for media error recovery	62

Supported configurations	62
I/O performance versus synchronization speed optimization	62
Implications regarding other commands	62

Chapter 9. Asynchronous remote mirroring 65

Chapter scope	66
Features highlights	66
Terminology	67
Technological overview	68
Replication scheme	69
Snapshot-based technology	69
Mirroring special snapshots	70
Initializing the mirroring	70
The sync job	72
Mirroring schedules and intervals	72
The mirror snapshot (ad-hoc sync job)	74
Determining replication and mirror states	74
Walking through the asynchronous mirroring process	84
Peers roles	90
Activating the mirroring	90
Mirroring consistency groups	92
Setting a consistency group to be mirrored	94
Setting-up a mirrored consistency group	94
Adding a mirrored volume to a mirrored consistency group	95
Removing a volume from a mirrored consistency group	95
Administering asynchronous mirroring	95
Accommodating disaster recovery scenarios	95
Overview of select asynchronous mirroring commands	101

Chapter 10. IP and Ethernet connectivity 105

Ethernet ports	105
IP and Ethernet connectivity	105
Management connectivity	107
Field technician ports	108
Configuration guidelines summary	109

Chapter 11. Data migration. 111

Data migration overview	111
I/O handling in data migration	112
Data migration stages	112
Handling failures	114

Chapter 12. Event handling 115

Event information	115
Viewing events	116
Defining events notification rules	116
Alerting events configuration limitations	117
Defining destinations	117
Defining gateways	117

Chapter 13. Access control 119

User roles and permission levels	119
--	-----

Predefined users	120
Application administrator	121
Authentication methods	122
Native authentication	123
LDAP-authentication	124
Switching between LDAP and native authentication modes	129
Administering access control	130
Logging and event reporting	130
Access control commands	130
Glossary of access control concepts	132

Chapter 14. Tivoli Storage Productivity Center and SMI-S interoperability. 135

Chapter 15. Nondisruptive code load 137

Glossary 139

Safety and environmental notices . . . 145

Danger notices	145
Labels	147
Caution notices	147
Attention notices	148
Laser safety	148
Usage restrictions	148
Rack safety	149
Rack installation	149
Rack relocation (19" rack)	150
Product recycling and disposal	150
Battery return program	152
Fire suppression systems	153

Notices 155

Notices	156
Copyrights	157
Trademarks	157
Electronic emission notices	158
Federal Communications Commission (FCC) Class A Statement	158
Industry Canada Class A Emission Compliance Statement	158
Avis de conformité à la réglementation d'Industrie Canada	158
European Union (EU) Electromagnetic Compatibility Directive	158
Australia and New Zealand Class A statement	159
Germany Electromagnetic Compatibility Directive	159
People's Republic of China Class A Electronic Emission Statement	160
Taiwan Class A warning statement	160
Japan VCCI Class A ITE Electronic Emission Statement	160
Korean Class A Electronic Emission Statement	161

Index 163

Introduction

The IBM® XIV® Storage System is designed for secure, dependable, enterprise-grade data storage and access, straightforward installation and upgrading, and full scalability. The system contains proprietary and innovative algorithms that offset hardware malfunctions, minimize maintenance, and provide flexibility. The system uses off-the-shelf hardware components that are easy to integrate and support.

Purpose and scope

This document contains a complete hardware and software system overview of the IBM XIV Storage System. Relevant tables, charts, graphic interfaces, sample outputs, and appropriate examples are also provided.

Document version

This document supports version 10.2 of the IBM XIV Storage System code.

Intended audience

This document is a reference for administrators and IT staff that work with the IBM XIV Storage System.

Publications and related information

Product manuals, other IBM publications, and websites contain information that relates to the IBM XIV Storage System.

To view a PDF file, you need Adobe Acrobat Reader, which can be downloaded for free from the Adobe (www.adobe.com/products/acrobat/readstep.html) website.

Information centers

From the IBM XIV Storage System Information Center (publib.boulder.ibm.com/infocenter/ibmxiv/r2/index.jsp), you can browse all product documentation.

Publications

Information that is available in the information center is also available in a set of publications, in PDF format.

IBM XIV Storage System

-  *IBM XIV Storage System Product Overview* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/GA32-0791-02.pdf)

This document contains an overview of the IBM XIV Storage System hardware and software.

-  *IBM XIV Storage System Planning Guide* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/GA32-0770-05.pdf)

This document defines installation requirements for IBM XIV Storage System. It is important to ensure that you meet all requirements to guarantee a fast and reliable installation.

- *IBM XIV Storage System Host Attachment Guides*

These documents provide information about attaching host systems to the IBM XIV Storage System:

-  *Host System Attachment Guide for AIX* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/hag_1_5_x/GA32-0643-06.pdf)
-  *Host System Attachment Guide for HP-UX* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/hag_1_5_x/GA32-0645-02.pdf)
-  *Host System Attachment Guide for Linux* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/hag_1_5_x/GA32-0647-02.pdf)
-  *Host System Attachment Guide for Solaris* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/hag_1_5_x/GA32-0649-02.pdf)
-  *Host System Attachment Guide for Windows* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/hag_1_5_x/GA32-0652-03.pdf)

-  *IBM XIV Storage System XCLI Reference Guide* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/GC27-2213-10.pdf), GC27-2213

This document describes the IBM XIV command-line interface (XCLI) system and utility commands used to manage and maintain the XIV system, including the command syntax, parameter descriptions, output descriptions, and examples.

-  *IBM XIV Storage System XCLI User Manual* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/GA32-0638-03.pdf), GA32-0638

This document describes how to use the IBM XIV command-line interface (XCLI) to run XIV system and utility commands.

VSS Provider - Xprov

-  *IBM XIV Storage System VSS Provider - Xprov Release Notes* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/xProv_2_2_4.pdf)

This document describes the supported environment, new features, known issues, and installation information.

Remote Mirroring for VCS Cluster

-  *IBM XIV Storage System Remote Mirroring for VCS Installation Guide* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/IBM_XIV_Remote_Mirroring_Agent_for_VCS_2.1.0_Installation_Guide.pdf)
This guide describes how to install and configure the VERITAS Cluster Server (VCS) enterprise agent for IBM XIV Remote Mirroring.
-  *IBM XIV Storage System Remote Mirroring for VCS Release Notes for Windows* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/IBM_XIV_Remote_Mirroring_Agent_for_VCS_2.1.0_Release_Notes_for_Windows.pdf)
This document describes the supported environment, new features, fixes, and known issues.
-  *IBM XIV Storage System Remote Mirroring for VCS Release Notes for Solaris* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/IBM_XIV_Remote_Mirroring_Agent_for_VCS_2.1.0_Release_Notes_for_Solaris.pdf)
This document describes the supported environment, new features, fixes, and known issues.

MPIO Management Console

-  *IBM XIV Storage System MPIO Management Console User's Guide* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/GA32-0795.pdf), GA32-0746
This guide discusses the IBM XIV MPIO Management Console application, which provides monitoring and management capabilities to the multipath subsystems of the IBM XIV Storage System.

Remote Support Proxy

-  *IBM XIV Storage System Remote Support Proxy Installation and User's Guide* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/GA32-0795.pdf), GA32-0795
This guide describes how to install, configure, and use the IBM XIV Storage System Remote Support Proxy to connect an XIV system to the XIV Remote Support Center.

Management Console for VMware vCenter

-  *IBM XIV Storage System Management Console for VMware vCenter User's Guide* (publib.boulder.ibm.com/infocenter/ibmxiv/r2/topic/com.ibm.help.xiv.doc/docs/GA32-0820-01.pdf), GA32-0820
This guide provides installation and configuration instructions for the IBM XIV Management Console for VMware vCenter.

IBM Redbooks publications and technical papers

Various IBM Redbooks® publications, Redpapers, and white papers are available for the IBM XIV Storage System. For additional papers, see the IBM XIV Storage System (www.ibm.com/systems/storage/disk/xiv/) website.

-  *IBM XIV Storage System: Architecture, Implementation, and Usage* (www.redbooks.ibm.com/abstracts/sg247659.html?Open&cm_sp=MTE10970)

This Redbooks publication describes the concepts, architecture, and implementation of the XIV system, which is designed to be a scalable enterprise storage system based upon a grid array of hardware components.

-  *IBM XIV Storage System with the Virtual I/O Server and IBM i*
(www.redbooks.ibm.com/abstracts/redp4598.html?Open&cm_sp=MTE11087)
This Redbooks publication discusses and explains how you can connect the XIV system to the IBM i operating system through the Virtual I/O Server (VIOS). A connection through the VIOS is especially useful for IT centers that have many small IBM i partitions. When using the VIOS, the fibre-channel host adapters can be installed in the VIOS and shared by many IBM i clients using virtual connectivity to the VIOS.
-  *XIV Storage System: Host Attachment and Interoperability*
(www.redbooks.ibm.com/redpieces/abstracts/sg247904.html?Open)
This Redbooks publication describes how to attach an XIV system to various hosting operating system platforms in combination with databases and other storage-oriented application software. It also provides solutions for combining the IBM XIV Storage System with other storage platforms, host servers, or gateways.
-  *IBM XIV Storage System: Copy Services and Migration*
(www.redbooks.ibm.com/abstracts/sg247759.html?Open)
This Redbooks publication describes IBM XIV Storage System copy and migration functions for various data protection scenarios, to enhance your business continuance, data migration, and online-backup solutions. These include point-in-time copies (also known as snapshots and full volume copies) and remote-copy capabilities in synchronous or asynchronous mode. This book also discusses how to integrate the snapshot function with the IBM Tivoli FlashCopy Manager, built-in migration capability, and migration alternatives based on the IBM SAN Volume Controller (SVC).

Related websites

View these websites to get more information about the XIV system.

-  IBM XIV Storage System (www.ibm.com/systems/storage/disk/xiv/)
Use this website to learn about the XIV system, including features and hardware summary. This website also has links to white papers, Redbooks publications, and product documentation.
-  IBM Support Portal (www.ibm.com/storage/support)
Use this website to obtain downloadable files, links to submit and track problems, and support phone numbers and contacts.
-  IBM Systems Storage forum (www.ibm.com/developerworks/forums/forum.jspa?forumID=846)
Use this forum to share ideas with knowledgeable experts and discover how the latest IBM storage solutions can address your business challenges. Forum topics include storage management, storage virtualization, business continuity, infrastructure simplification, disk storage systems, and storage software products and solutions.

Notices used in this document

The caution and danger statements used in this document also appear in the multilingual *IBM XIV Storage System Safety Notices* document. Each caution and danger statement is numbered for easy reference to the corresponding statements in the safety document.

The following types of notices and statements are used in this document:

Note These notices provide important tips, guidance, or advice.

Important

These notices provide information or advice that might help you avoid inconvenient or problem situations.

Attention

These notices indicate possible damage to programs, devices, or data. An attention notice is placed just before the instruction or situation in which damage could occur.

Caution

These statements indicate situations that can be potentially hazardous to people because of some existing condition, or where a potentially dangerous situation might develop because of some unsafe practice. A caution statement is placed just before the description of a potentially hazardous procedure step or situation.

Danger

These statements indicate situations that can be potentially lethal or extremely hazardous to people. A danger statement is placed just before the description of a potentially lethal or extremely hazardous procedure step or situation.

Document conventions

The following conventions are used in this document:

boldface

Text in **boldface** represents menu items and lowercase or mixed-case command names.

italics Text in *italics* is used to emphasize a word. In command syntax, it is used for variables for which you supply actual values.

monospace

Text in monospace identifies the data or commands that you type, samples of command output, or examples of program code or messages from the system.

Terms and abbreviations

A complete list of terms and abbreviations can be found in the “Glossary” on page 139.

Getting information, help, and service

If you need help, service, technical assistance, or just want more information about IBM products, you will find a variety of sources to assist you. Table 1 provides a list of Web pages that you can view to get information about IBM products and services and to find the latest technical information and support:

Table 1. IBM Web sites for help, information and service

Web site	Description
http://www.ibm.com	Main IBM home page
http://www.ibm.com/storage/support	IBM Support home page
http://www.ibm.com/planetwide	IBM Support page with pointers to the relevant contact information for a specific country

How to send your comments

Your feedback is important in helping us provide the most accurate and high-quality information. If you have comments or suggestions for improving this document, send us your comments by e-mail to starpubs@us.ibm.com or use the Readers' Comments form at the back of this publication. Be sure to include the following:

- Exact publication title
- Form number (for example, GC26-1234-02)
- Page numbers to which you are referring

If the Reader's Comment Form in the back of this manual is missing, you can direct your mail to:

International Business Machines Corporation
Information Development
Department GZW
9000 South Rita Road
Tucson, Arizona 85744-0001 U.S.A.

When you send information to IBM, you grant IBM a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Chapter 1. Overview: The IBM XIV Storage System

This chapter covers the various features and functions of the IBM XIV Storage System, including an overview of its hardware and software components. This overview includes a brief description of its key design principles from the system's point of view, but does not cover the functionality from the user's point of view, which is covered in the subsequent chapters.

Features and functionality

The IBM XIV Storage System is characterized by the following set of features:

- iSCSI and Fibre Channel (FC) interfaces
- Multiple host access
- Unique data distribution method that eliminates "hot spots"
- Constant, predictable high performance with zero tuning
- Extremely fast rebuild time in the event of disk failure
- Large processing power and cache
- Innovative snapshot functionality, including support for practically unlimited number of snapshots, snap-of-snap and restore-from-snap
- Fault tolerance, failure analysis, and self-healing algorithms
- No single-point-of-failure
- Support for storage pools administrative units
- Support for thin provisioning
- Support for instant space reclamation
- Synchronous and asynchronous replication of a volume (as well as a consistency group) to a remote system
- Data migration
- Easy assignment and reassignment of storage capacity
- Remote configuration management
- Notifications of events through e-mail, SNMP, or SMS messages
- Non-disruptive maintenance and upgrades
- Management software, including a graphical user interface (GUI) and a command-line interface (CLI)
- Support for query of configuration information by Tivoli Productivity Center 4.1.
- Support for managing snapshot operations with Tivoli Storage Manager for Copy Services v. 6.1

Hardware overview

This section provides a general overview of the IBM XIV Storage System hardware.

Hardware components

The IBM XIV Storage System configuration includes data modules, interface modules, Ethernet switches, and uninterruptible power supply units.

The following figures show the IBM XIV Storage System major hardware components as they are seen from front and back.

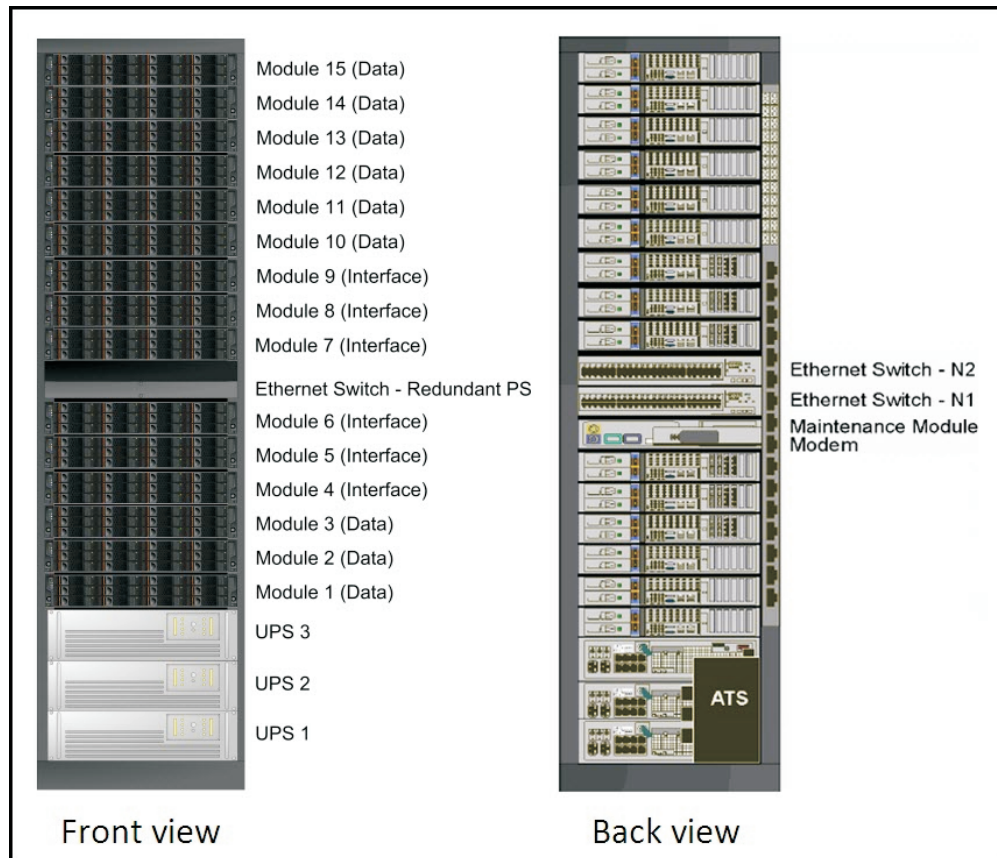


Figure 1. IBM XIV Storage System

Interface Modules

Each contains 12 disk drive modules (DDMs), CPU, Cache and Host Interface adaptors.

Host Interface adaptors

Each interface module provides Fibre Channel ports. Some host interface modules also provide iSCSI ports. Hosts may attach to the system by using the Fibre Channel and iSCSI ports. (See the most current version of the XIV Interoperability Matrix for specific supported configurations.)

SATA disk drives and Cache memory

Each Interface module contains 12 SATA disk drives and cache memory. SATA disk drives are used as the non-volatile memory for storing data in the storage grid and cache memory is used for caching data previously read, pre-fetching of data from a disk, and for delayed destaging of previously written data.

Data modules

Each Data module contains 12 SATA disk drives and cache memory. SATA disk drives are used as the nonvolatile memory for storing data in the storage grid and cache memory is used for caching data previously read, prefetching of data from a disk, and for delayed destaging of previously written data.

Uninterruptible power supply module complex

The uninterruptible power supply module complex consists of three uninterruptible supply units. It maintains an internal power supply in the event of a temporary failure of the external power supply. In the case of a continuous external power failure, the uninterruptible power supply module complex maintains power long enough for a safe and ordered shutdown of the IBM XIV Storage System. The IBM XIV Storage System can sustain the failure of one uninterruptible power supply unit while protecting against external power failures.

Ethernet switch RPS

A redundant power source for the Ethernet switches .

Maintenance module

Allows remote support access using a modem.

ATS The Automatic Transfer System (ATS) switches between line cords in order to allow redundancy of external power.

Modem

Allows the system to receive a connection for remote access by IBM support. The modem connects to the maintenance module.

Ethernet switches

Provide redundant internal GB Ethernet networks. All the modules in the system are linked through an internal redundant Gigabit Ethernet network. The network is composed of two independent Ethernet switches. Each module is directly attached to both switches, and the switches are also linked to each other. Because the switches are configured in an active-active configuration, the network topology uses maximum bandwidth while being tolerant to any individual failure in a network component (port, link, or switch).

Hardware connectivity

Data and interface modules are generically referred to as "modules". Modules communicate with each other by means of the internal Gigabit Ethernet network. Each module contains redundant Gigabit Ethernet ports used for module to module communication. The ports are all linked to the internal network through the redundant Ethernet switches. In addition, for monitoring purposes, the UPSs are directly connected by Ethernet and USB to individual modules.

Supported interfaces

The following interfaces are supported by the IBM XIV Storage System:

- Fibre Channel for host-based I/O
- Gigabit Ethernet for host-based I/O using the iSCSI protocol
- Gigabit Ethernet for management (GUI or CLI) connectivity
- Remote access interfaces:
 - Call-home connection - connecting the IBM XIV Storage System to an IBM trouble-ticketing system.
 - Broadband connection (VPN) - provides a two-way broadband access to the system for remote access by IBM support.
 - Modem - for incoming calls only. The customer has to provide telephone line and number. The modem provides secondary means for providing remote access for IBM Support.

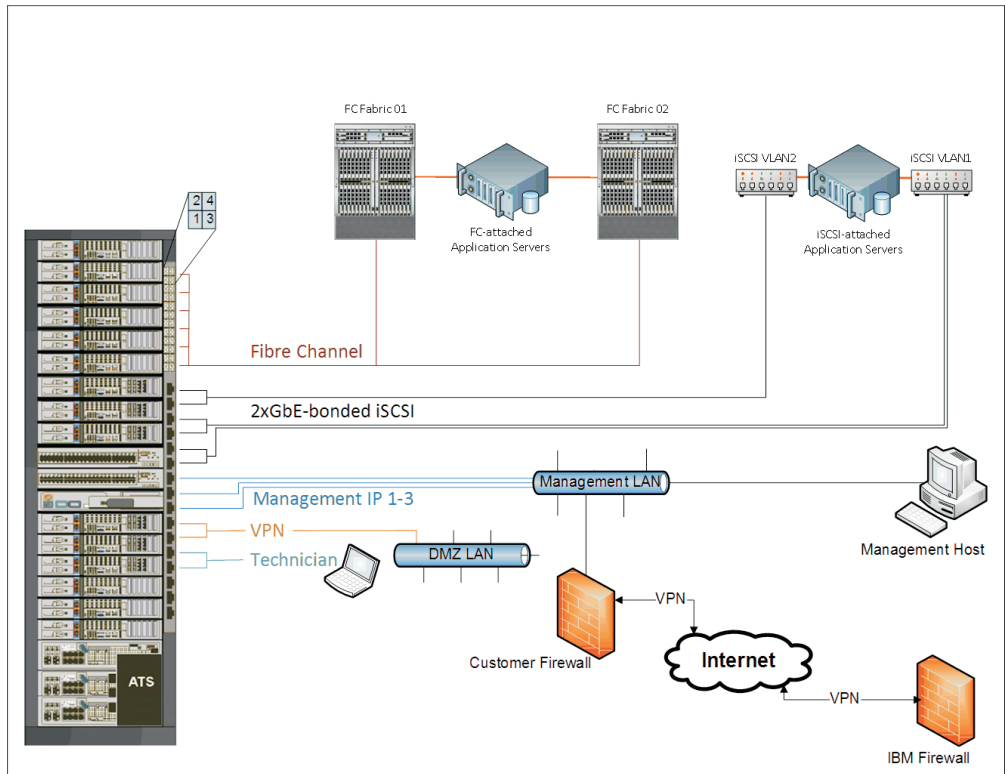


Figure 2. The IBM XIV Storage System Interfaces

Management options

The IBM XIV Storage System provides several management options.

GUI and CLI management applications

These applications must be installed on each workstation that will be used for managing and controlling the system. All configurations and monitoring aspects of the system can be controlled through the GUI or the CLI.

SNMP

Third-party SNMP-based monitoring tools are supported using the IBM XIV MIB.

E-mail notifications

The IBM XIV Storage System can notify users, applications or both through e-mail messages regarding failures, configuration changes, and other important information.

SMS notifications

Users can be notified through SMS of any system event.

Reliability

IBM XIV Storage System reliability features include data mirroring, spare storage capacity, self-healing mechanisms, and data virtualization.

Redundant components and no single point of failure

IBM XIV Storage System hardware components are fully redundant, and ensure failover protection for each other to prevent a single point of system failure.

System failover processes are transparent to the user because they are swiftly and seamlessly completed.

Data mirroring

Data arriving from the host for storage is temporarily placed in two separate caches before it is permanently written to two disk drives located in separate modules. This guarantees that the data is always protected against possible failure of individual modules, and this protection is in effect even before data has been written to the nonvolatile disk media.

Self-healing mechanisms

The IBM XIV Storage System includes built-in mechanisms for self-healing to take care of individual component malfunctions and to automatically restore full data redundancy in the system within minutes.

Self-healing mechanisms dramatically increase the level of reliability in the IBM XIV Storage System. Rather than necessitating a technician's on-site intervention in the case of an individual component malfunction to prevent a possible malfunction of a second component, the automatically restored redundancy allows a relaxed maintenance policy based on a pre-established routine schedule.

Self-healing mechanisms are not just started in a reactive fashion following an individual component malfunction, but also proactively - upon detection of conditions indicating potential imminent failure of a component. Often, potential problems are identified well before they might occur with the help of advanced algorithms of preventive self-analysis that are continually running in the background. In all cases, self-healing mechanisms implemented in the IBM XIV Storage System identify all data portions in the system for which a second copy has been corrupted or is in danger of being corrupted. The IBM XIV Storage System creates a secure second copy out of the existing copy, and it stores it in the most appropriate part of the system. Taking advantage of the full data virtualization, and based on the data distribution schemes implemented in the IBM XIV Storage System, such processes are completed with minimal data migration.

As with all other processes in the system, the self-healing mechanisms are completely transparent to the user, and the regular activity of responding to I/O data requests is thoroughly maintained with no degradation to system performance. Performance, load balance, and reliability are never compromised by this activity.

Protected cache

IBM XIV Storage System cache writes are protected. Cache memory on a module is protected with ECC (Error Correction Coding). All write requests are written to two separate cache modules before the host is acknowledged. The data is later destaged to disks.

Redundant power

Redundancy of power is maintained in the IBM XIV Storage System through the following means:

- Three uninterruptible power supply units - the system can run indefinitely on two uninterruptible power supply units. No system component will lose power if a single uninterruptible power supply unit fails.
- Redundant power supplies in each data and interface module. There are two power supplies for each module and each power supply for a module is powered by a different uninterruptible power supply unit.
- Redundant power for Ethernet switches - each Ethernet switch is powered by two uninterruptible power supply units. One is a direct connect; one is through the Ethernet switch redundant power supply.
- Redundant line cords - to protect against the loss of utility power, two line cords are supplied to the ATS. If utility power is lost on one line cord, the ATS automatically switches to the other line cord, without impacting the system.
- In the event of loss of utility power on both line cords, the uninterruptible power supply units will maintain power to the system until an emergency destage of all data in the system can be performed. Once the emergency destage has completed, the system will perform a controlled power down.

Performance

IBM XIV Storage System performance features include total load balancing, intelligent caching and 1 Gb Ethernet connections.

Total load balance

The fundamental principle of the IBM XIV Storage System architecture is to evenly distribute the handling and storage of data (associated with its logical volumes) across all hardware components in the system.

This optimum state of total load balance is preserved in all circumstances and under any kinds of configuration changes in the system, including changes due to failing hardware, so that the very high performance parameters that derive from it are fully scalable.

Moreover, the load distribution of IBM XIV Storage System is maintained when expanding the system. When modules are added, data already written to the system is redistributed in order to evenly spread the data among all modules and disk drives in the expanded system. At least two copies of the data are maintained within the system for the entire redistribution process. Reliability, availability, and performance are unaffected during and after these redistribution processes.

Data distribution in the IBM XIV Storage System is equivalent to a full, optimal virtualization of volumes across the storage resources of the system. Under this virtualization, I/O activity performed in the system takes full advantage of all the available physical resources at any point in time. Write or read requests directed at any particular volume harness the entire CPU power, internal bandwidth, and disk capacity, nearly eliminating bottlenecks.

Intelligent caching for improved performance

The IBM XIV Storage System provides intelligent caching through advanced algorithms for cache management, and through the use of a distributed cache.

The algorithms used by the IBM XIV Storage System use innovative approaches for promotion, demotion and destaging data in the cache memory. In addition, the system uses highly efficient and novel prefetch routines. These algorithms are

transparent to the host requesting the data. They result in short response times and low latencies for data requests handled by the system.

Additionally, each data or interface module caches data only for its local disks. This provides the benefit of allowing a very high bandwidth connection between the cache and the drives being cached. Also, because each data and interface module devotes its full processing power to managing an independent cache, and because this independent cache only handles data for its local disks, smaller slots can be freely used without degrading performance. The IBM XIV Storage System uses cache slots as small as 4KB and as large as 1MB, allowing highly efficient use of the cache memory. Cache slot size is automatically determined by the system, and varies based on the access patterns of the data in the region being cached.

Functionality

IBM XIV Storage System functions include point-in-time copying, automatic notifications, and ease of management through a GUI or CLI.

Snapshot management

The IBM XIV Storage System provides powerful snapshot mechanisms for creating point-in-time copies of volumes. These snapshots can be used for backup, testing, or recovery from logical errors.

The snapshot mechanisms include the following features:

- Differential snapshots, where only the data that differs between the source volume and its snapshot consumes storage space
- Instant creation of a snapshot without any interruption of the application, making the snapshot available immediately
- Practically unlimited quantity of snapshots, with no minimal performance or space overhead
- Writable snapshots, which can be used for a testing environment; storage space is only required for actual data changes
- Snapshot of a writable snapshot can be taken
- High performance that is independent of the number of snapshots or volume size
- The ability to restore from snapshot to volume or snapshot

Consistency groups for snapshots

Volumes can be put in a consistency group to facilitate the creation of consistent point-in-time snapshots of all the volumes in a single operation. This is essential for applications that use several volumes concurrently and need a consistent snapshot of all these volumes.

Storage pools

The storage space of the IBM XIV Storage System can be administratively portioned into storage pools to enable the control of storage space consumption for specific applications or departments. Storage pools are used to administer the storage resources of volumes and snapshots.

Remote monitoring and diagnostics

IBM XIV Storage System can email important system events to IBM Support. This allows IBM to immediately detect hardware failures warranting immediate attention and react swiftly (e.g., dispatch service personnel). Additionally, IBM support personnel can conduct remote support and generate diagnostics for both maintenance and support purposes. All remote support is subject to customer permission and remote support sessions are protected with a challenge response security mechanism.

SNMP

Third-party SNMP-based monitoring tools are supported for the IBM XIV Storage System MIB.

Multipathing

The parallel design underlying the activity of the Host Interface modules and the full data virtualization achieved in the system implement thorough multi-pathing access algorithms. Thus, as the host connects to the system through several independent ports, each volume can be accessed directly through any of the Host Interface modules, and no interaction has to be established across the various modules of the Host Interface array.

Automatic event notifications

The system can be set to automatically transmit appropriate alarm notification messages through SNMP traps, or e-mail messages. The user can configure various triggers for sending events and various destinations depending on the type and severity of the event. The system can also be configured to send notifications until a user acknowledges their receipt.

Management through GUI and CLI

The IBM XIV Storage System offers a user-friendly and intuitive GUI application and CLI commands to configure and monitor the system. These feature comprehensive system management functionality, encompassing hosts, volumes, consistency groups, storage pools, snapshots, mirroring relationships, data migration, events and more.

External replication mechanisms

External replication and mirroring mechanism are an extension of the *internal replication mechanisms* and of the overall functionality of the system. These features provide protection against a site disaster to ensure production continues. The mirroring can be performed over either Fibre Channel or iSCSI, and the host-to-storage protocol is independent of the mirroring protocol.

Upgradability

The IBM XIV Storage System is available in a partial rack system comprised of as few as six (6) modules, or as many as fifteen (15) modules per rack.

Partial rack systems may be upgraded by adding data and interface modules, up to the maximum of fifteen (15) modules per rack.

The system supports a non-disruptive upgrade of the system, as well as hotfix updates.

Chapter 2. Volumes and snapshots overview

The volume is the basic data-storage entity as defined by the SCSI protocol. A *volume* is a list of blocks (where the size of each block is 512 bytes), that are being presented to a SCSI host as a logical disk. Using the SCSI protocol that is implemented over FC or iSCSI, the host can read and write volume data.

Snapshots of volumes represent the data on a volume at a specific point in time. Figure 3 shows a volume with snapshots taken at two different points in time.

volumes_and_snapshots

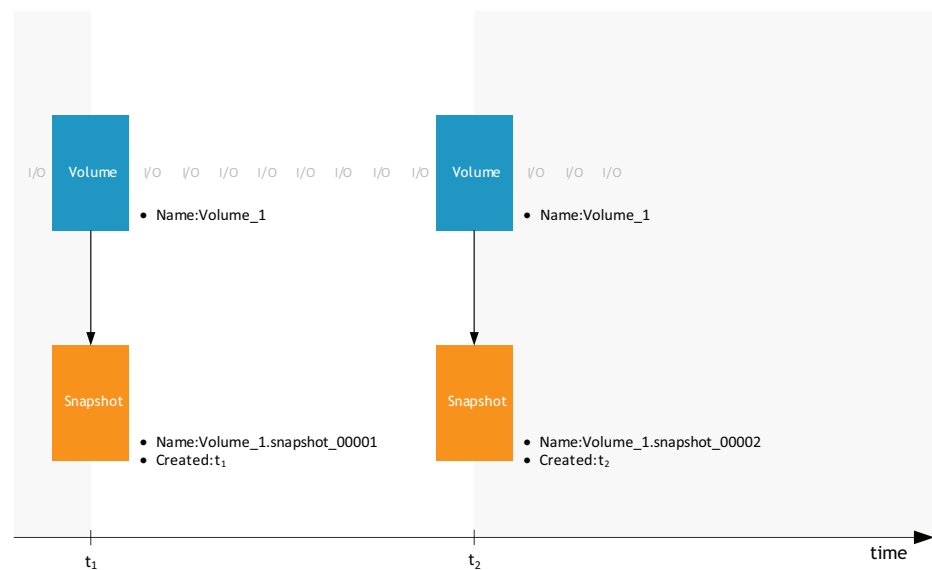


Figure 3. A volume is shown with snapshots taken at two different points in time.

Virtualization, mirroring, and thin provisioning of volumes is facilitated through:

Snapshots

An unlimited number of snapshots can be taken with minimal impact on performance.

Consistency groups

Volumes can be grouped into consistency groups to take simultaneous snapshots of a large number of volumes and easily manage snapshots for these volumes.

Storage pools

Each volume must be associated with a storage pool to manage its storage capacity.

Note: In the storage system industry literature, volumes are sometimes referred to as *disks*, *LUNs* or *devices*.

The volume life cycle

The volume is the basic data container that is presented to the hosts as a logical disk.

The term *volume* is sometimes used for an entity that is either a volume or a snapshot. Hosts view volumes and snapshots through the same protocol. Whenever required, the term *master volume* is used for a volume to clearly distinguish volumes from snapshots.

Each volume has two configuration attributes: a name and a size. The volume name is an alphanumeric string that is internal to the IBM XIV Storage System and is used to identify the volume to both the GUI and CLI commands. The volume name is not related to the SCSI protocol. The volume size represents the number of blocks in the volume that the host sees.

The volume can be managed by the following commands:

Create Defines the volume using the attributes you specify

Resize Changes the virtual capacity of the volume. For more information, see Chapter 6, “Thin provisioning,” on page 39.

Copy Copies the volume to an existing volume or to a new volume

Format
Clears the volume

Lock Prevents hosts from writing to the volume

Unlock
Allows hosts to write to the volume

Rename
Changes the name of the volume, while maintaining all of the volumes previously defined attributes

Delete Deletes the volume. See Instant Space Reclamation below.

The following query commands list volumes:

Listing Volumes
This command lists details of all volumes, or a specific volume according to a given volume or pool.

Finding a Volume Based on a SCSI Serial Number
This command prints the volume name according to its SCSI serial number.

These commands are available when you use both the IBM XIV Storage Management GUI and the IBM XIV command-line interface (XCLI). See the *IBM XIV Storage System XCLI User Manual* for the commands that you can issue in the XCLI.

Figure 4 on page 13 shows the commands you can issue for volumes.

volume_life_cycle

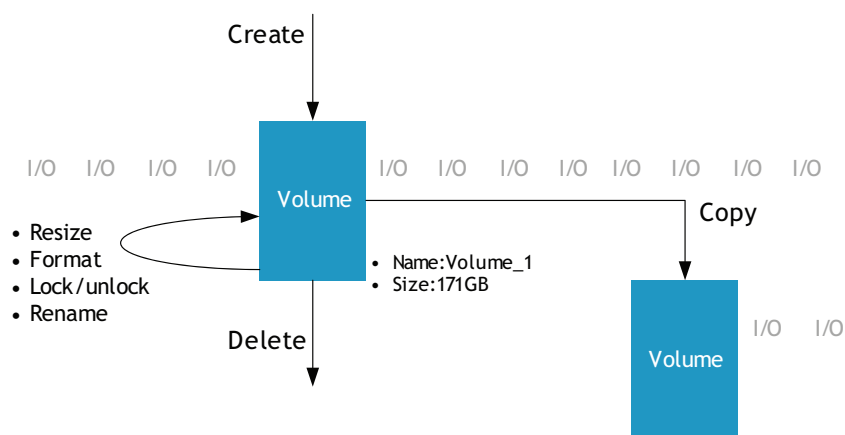


Figure 4. Volume operations

Support for Symantec Storage Foundation Thin Reclamation

The IBM XIV Storage System supports Symantec's Storage Foundation Thin Reclamation API.

The IBM XIV Storage System features instant space reclamation functionality, enhancing the existing IBM XIV Thin Provisioning capability. The new instant space reclamation function allows IBM XIV users to optimize capacity utilization, thus saving costs, by allowing supporting applications, to instantly regain unused file system space in thin-provisioned XIV volumes instantly.

The IBM XIV Storage System is one of the first high-end storage systems to offer instant space reclamation. The new, instant capability enables third party products vendors, such as Symantec Thin Reclamation, to interlock with the The IBM XIV Storage System such that any unused space is detected instantly and automatically, and immediately reassigned to the general storage pool for reuse.

This enables integration with thin-provisioning-aware Veritas File System (VxFS) by Symantec, which ultimately enables to leverage the IBM XIV Storage System thin-provisioning-awareness to attain higher savings in storage utilization.

For example, when data is deleted by the user, the system administrator can initiate a reclamation process in which the IBM XIV Storage System frees the non-utilized blocks and where these blocks are reclaimed by the available pool of storage.

Instant space reclamation doesn't support space reclamation for the following objects:

- Mirrored volumes

- Volumes that have snapshots
- Snapshots

Snapshots

A *snapshot* is a logical volume reflecting the contents of a given source volume at a specific point-in-time.

The IBM XIV Storage System uses advanced snapshot mechanisms to create a virtually unlimited number of volume copies without impacting performance. Snapshot taking and management are based on a mechanism of internal pointers that allow the master volume and its snapshots to use a single copy of data for all portions that have not been modified.

This approach, also known as Redirect-on-Write (ROW) is an improvement of the more common Copy-on-Write (COW), which translates into a reduction of I/O actions, and therefore storage usage.

With the IBM XIV snapshots, no storage capacity is consumed by the snapshot until the source volume (or the snapshot) is changed.

Redirect on write

The IBM XIV Storage System uses the Redirect-on-Write (ROW) mechanism.

The following items are characteristics of using ROW when a write request is directed to the master volume:

1. The data originally associated with the master volume remains in place.
2. The new data is written to a different location on the disk.
3. After the write request is completed and acknowledged, the original data is associated with the snapshot and the newly written data is associated with the master volume.

In contrast with the traditional copy-on-write method, with redirect-on-write the actual data activity involved in taking the snapshot is drastically reduced. Moreover, if the size of the data involved in the write request is equal to the system's slot size, there is no need to copy any data at all. If the write request is smaller than the system's slot size, there is still much less copying than with the standard approach of Copy-on-Write. Figure 5 on page 15 provides an example of ROW.

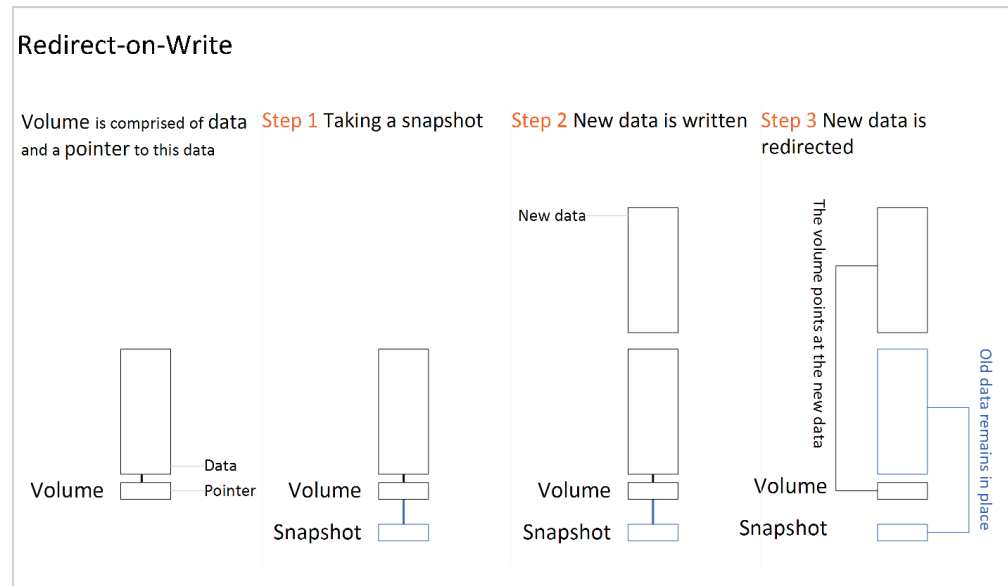


Figure 5. Example of the Redirect-on-Write process

The metadata established at the beginning of the snapshot mechanism is independent of the size of the volume to be copied. This approach allows the user to achieve the following important goals:

Continuous backup

As snapshots are taken, backup copies of volumes are produced at frequencies that resemble those of *Continuous Data Protection* (CDP). Instant restoration of volumes to virtually any point in time is easily achieved in case of logical data corruption at both the volume level and the file level.

Productivity

The snapshot mechanism offers an instant and simple method for creating short or long-term copies of a volume for data mining, testing, and external backups.

Storage utilization

The IBM XIV Storage System allocates space for volumes and their Snapshots in a way that whenever a Snapshot is taken, additional space is actually needed only when the volume is written into.

As long as there is no actual writing into the volume, the Snapshot does not need actual space. However, some applications write into the volume whenever a Snapshot is taken. This writing into the volume mandates immediate space allocation for this new Snapshot. Hence, these applications use space less efficiently than other applications.

The snapshot auto-delete priority

Snapshots are associated with an *auto-delete priority* to control the order in which snapshots are automatically deleted.

Taking volume snapshots gradually fills up storage space according to the amount of data that is modified in either the volume or its snapshots. To free up space when the maximum storage capacity is reached, the system can refer to the

auto-delete priority to determine the order in which snapshots are deleted. If snapshots have the same priority, the snapshot that was created first is deleted first.

Snapshot name and association

A snapshot can either be taken of a source volume, or from a source snapshot. The name of a snapshot is either automatically assigned by the system at creation time or given as a parameter of the XCLI command that creates it. The snapshot's auto-generated name is derived from its volume's name and a serial number. The following are examples of snapshot names:

MASTERVOL.snapshot_XXXXX
NewDB-server2.snapshot_00597

Parameter	Description	Example
MASTERVOL	The name of the volume.	NewDB-server2
XXXXXX	A five-digit, zero filled snapshot number.	00597

The snapshot lifecycle

The roles of the snapshot determine its life cycle.

Figure 6 shows the life cycle of a snapshot.

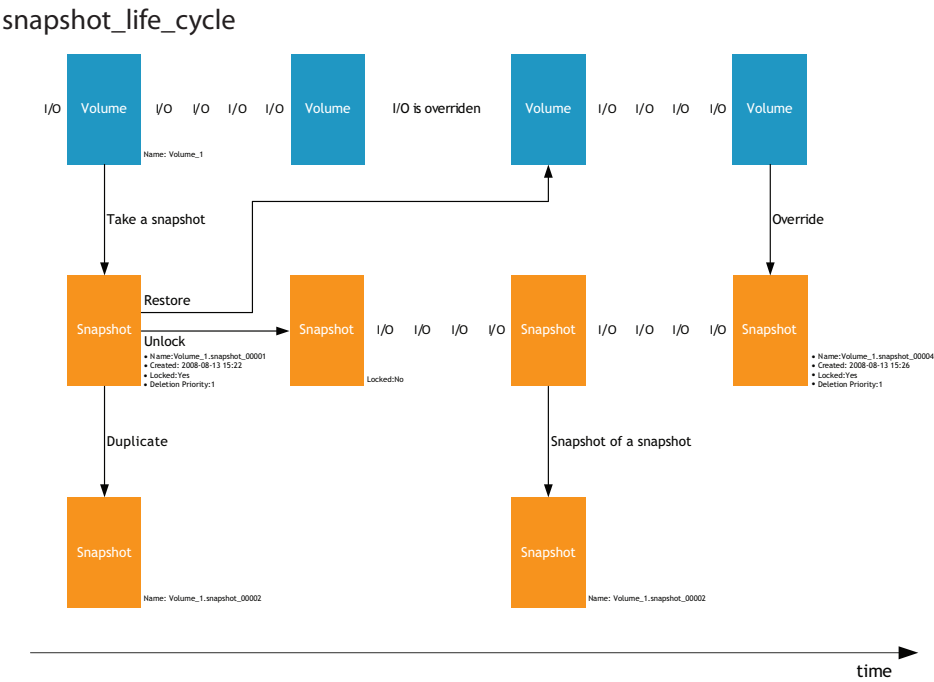


Figure 6. The snapshot life cycle

The following list describes the life cycle:

Create Creates the snapshot.

Restore

Copies the snapshot back onto the volume. The main snapshot functionality is the capability to restore the volume.

Unlocking

Unlocks the snapshot to make it writable and sets the status to Modified. Re-locking the unlocked snapshot disables further writing, but does not change the status from Modified.

Duplicate

Duplicates the snapshot. Similar to the volume, which can be snapshotted infinitely, the snapshot itself can be duplicated.

A snapshot of a snapshot

Creates a backup of a snapshot that was written into. Taking a snapshot of a writable snapshot is similar to taking a snapshot of a volume.

Overwriting a snapshot

Overwrites a specific snapshot with the content of the volume.

Delete Deletes the snapshot.

Creating a snapshot

First, a snapshot of the volume is taken. The system creates a pointer to the volume, hence the snapshot is considered to have been immediately created. This is an atomic procedure that is completed in a negligible amount of time. At this point, all data portions that are associated with the volume are also associated with the snapshot.

Later, when a request arrives to read a certain data portion from either the volume or the snapshot, it reads from the same single, physical copy of that data.

Throughout the volume life cycle, the data associated with the volume is continuously modified as part of the ongoing operation of the system. Whenever a request to modify a data portion on the master volume arrives, a copy of the original data is created and associated with the snapshot. Only then the volume is modified. This way, the data originally associated with the volume at the time the snapshot is taken is associated with the snapshot, effectively reflecting the way the data was before the modification.

Locking and unlocking snapshots

Initially, a snapshot is created in a locked state, which prevents it from being changed in any way related to data or size, and only enables the reading of its contents. This is called an *image* or *image snapshot* and represents an exact replica of the master volume when the snapshot was created.

A snapshot can be unlocked after it is created. The first time a snapshot is unlocked, the system initiates an irreversible procedure that puts the snapshot in a state where it acts like a regular volume with respect to all changing operations. Specifically, it allows write requests to the snapshot. This state is immediately set by the system and brands the snapshot with a permanent modified status, even if no modifications were performed. A *modified snapshot* is no longer an image snapshot.

An unlocked snapshot is recognized by the hosts as any other writable volume. It is possible to change the content of unlocked snapshots, however, physical storage space is consumed only for the changes. It is also possible to resize an unlocked snapshot.

Master volumes can also be locked and unlocked. A locked master volume cannot accept write commands from hosts. The size of locked volumes cannot be modified.

Duplicating image snapshots

A user can create a new snapshot by duplicating an existing snapshot. The duplicate is identical to the source snapshot. The new snapshot is associated with the master volume of the existing snapshot, and appears as if it were taken at the exact moment the source snapshot was taken. For image snapshots that have never been unlocked, the duplicate is given the exact same creation date as the original snapshot, rather than the duplication creation date.

With this feature, a user can create two or more identical copies of a snapshot for backup purposes, and perform modification operations on one of them without sacrificing the usage of the snapshot as an untouched backup of the master volume, or the ability to restore from the snapshot.

A snapshot of a snapshot

When duplicating a snapshot that has been changed using the unlock feature, the generated snapshot is actually a snapshot of a snapshot. The creation time of the newly created snapshot is when the command was issued, and its content reflects the contents of the source snapshot at the moment of creation.

After it is created, the new snapshot is viewed as another snapshot of the master volume.

Restoring volumes and snapshots

The restoration operation provides the user with the ability to instantly recover the data of a master volume from any of its locked snapshots.

Restoring volumes

A volume can be restored from any of its snapshots, locked and unlocked. Performing the restoration replicates the selected snapshot onto the volume. As a result of this operation, the master volume is an exact replica of the snapshot that restored it. All other snapshots, old and new, are left unchanged and can be used for further restore operations. A volume can even be restored from a snapshot that has been written to. Figure 7 on page 19 shows a volume being restored from three different snapshots.

restoring_a_volume

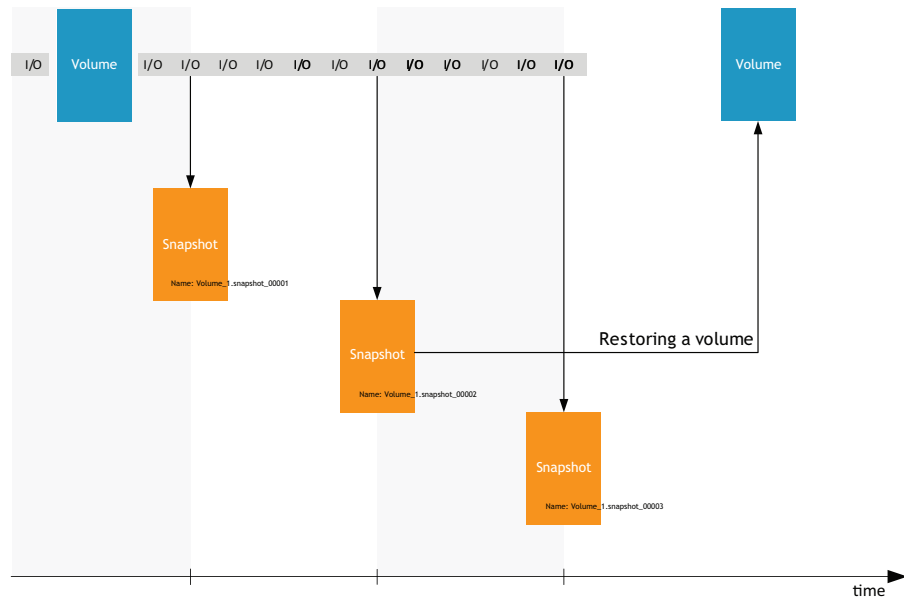


Figure 7. Restoring volumes

Restoring snapshots

The snapshot itself can also be restored from another snapshot. The restored snapshot retains its name and other attributes. From the host perspective, this restored snapshot is considered an instant replacement of all the snapshot content with other content. Figure 8 on page 20 shows a snapshot being restored from two different snapshots.

restoring_a_snapshot

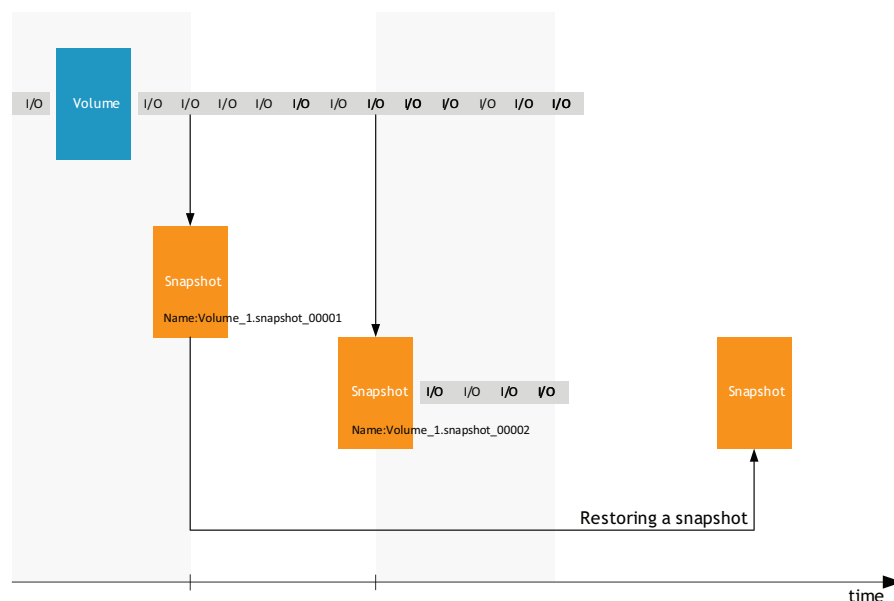


Figure 8. Restoring snapshots

Full Volume Copy

Full Volume Copy overwrites an existing volume, and at the time of its creation it is logically equivalent to the source volume.

After the copy is made, both volumes are independent of each other. Hosts can write to either one of them without affecting the other. This is somewhat similar to creating a writable (unlocked) snapshot, with the following differences and similarities:

Creation time and availability

Both Full Volume Copy and creating a snapshot happen almost instantly. Both the new snapshot and volume are immediately available to the host. This is because at the time of creation, both the source and the destination of the copy operation contain the exact same data and share the same physical storage.

Singularity of the copy operation

Full Volume Copy is implemented as a single copy operation into an existing volume, overriding its content and potentially its size. The existing target of a volume copy can be mapped to a host. From the host perspective, the content of the volume is changed within a single transaction. In contrast, creating a new writable snapshot creates a new object that has to be mapped to the host.

Space allocation

With Full Volume Copy, all the required space for the target volume is reserved at the time of the copy. If the storage pool that contains the target volume cannot allocate the required capacity, the operation fails and has no effect. This is unlike writable snapshots, which are different in nature.

Taking snapshots and mirroring the copied volume

The target of the Full Volume Copy is a master volume. This master volume can later be used as a source for taking a snapshot or creating a mirror. However, at the time of the copy, neither snapshots nor remote mirrors of the target volume are allowed.

Redirect-on-write implementation

With both Full Volume Copy and writable snapshots, while one volume is being changed, a redirect-on-write operation will ensure a split so that the other volume maintains the original data.

Performance

Unlike writable snapshots, with Full Volume Copy, the copying process is performed in the background even if no I/O operations are performed. Within a certain amount of time, the two volumes will use different copies of the data, even though they contain the same logical content. This means that the redirect-on-write overhead of writes occur only before the initial copy is complete. After this initial copy, there is no additional overhead.

Availability

Full Volume Copy can be performed with source and target volumes in different storage pools.

Snapshot and snapshot group format

This operation deletes the content of a snapshot - or a snapshot group - while maintaining its mapping to the host.

The purpose of the formatting is to allow customers to backup their volumes via snapshots, while maintaining the snapshot ID and the LUN ID. More than a single snapshot can be formatted per volume.

Required reading

Some of the concepts this topic refers to are introduced in this chapter as well as in a later chapter on this document. Consult the following reading list to get a grasp regarding these topics.

Snapshots

“Snapshots” on page 14

Snapshot groups

“The snapshot group life cycle” on page 35

Attaching a host

Chapter 3, “Host system attachment,” on page 23

The format operation results with the following

- The formatted snapshot is read-only
- The format operation has no impact on performance
- The formatted snapshot does not consume space
- Reading from the formatted snapshot always returns zeroes
- It can be overridden
- It can be deleted
- Its deletion priority can be changed

Restrictions

No unlock

The formatted snapshot is read-only and can't be unlocked.

No volume restore

The volume that the formatted snapshot belongs to can't be restored from it.

No restore from another snapshot

The formatted snapshot can't be restored from another snapshot.

No duplicating

The formatted snapshot can't be duplicated.

No re-format

The formatted snapshot can't be formatted again.

No volume copy

The formatted snapshot can't serve as a basis for volume copy.

No resize

The formatted snapshot can't be resized.

Use case

1. Create a snapshot for each LUN you would like to backup to, and mount it to the host.
2. Configure the host to backup this LUN.
3. **Format the snapshot.**
4. Re-snap. The LUN ID, Snapshot ID and mapping are maintained.

Restrictions in relation to other XIV operations

Snapshots of the following types can't be formatted:

Internal snapshot

Formatting an internal snapshot hampers the process it is part of, therefore is forbidden.

Part of a sync job

Formatting a snapshot that is part of a sync job renders the sync job meaningless, therefore is forbidden.

Part of a snapshot group

A snapshot that is part of a snapshot group can't be treated as an individual snapshot.

Snapshot group restrictions

All snapshot format restrictions apply to the snapshot group format operation.

Chapter 3. Host system attachment

The IBM XIV Storage System attaches to hosts of various operating systems.

The IBM XIV Storage System attaches to hosts through a set of Host Attachment Kits and complementary utilities.

Note: The term *host system attachment* was previously known as *host connectivity* or *mapping*. These terms are obsolete.

Balanced traffic and no single point of failure

All host traffic (I/O) is served through up to six interface modules (modules 4-9). Although the IBM XIV Storage System distributes the traffic across all system modules, the storage administrator is responsible for ensuring that host I/O operations are equally distributed among the different interface modules.

The workload balance should be watched and reviewed when host traffic patterns change. The IBM XIV Storage System does not automatically balance incoming host traffic. The storage administrator is responsible for ensuring that host connections are made redundantly in such a way that a single failure, such as in a module or HBA, will not cause all paths to the machine to fail. In addition, the storage administrator is responsible for making sure the host workload is adequately spread across the different connections and interface modules.

Dynamic rate adaptation

The IBM XIV Storage System provides a mechanism for handling insufficient bandwidth and external connections for the mirroring process.

The mirroring process replicates a local site on a remote site (See the Chapter 8, "Synchronous remote mirroring," on page 47 and Chapter 9, "Asynchronous remote mirroring," on page 65 chapters later on this document). To accomplish this, the process depends on the availability of bandwidth between the local and remote storage systems.

The mirroring process' sync rate attribute determines the bandwidth that is required for a successful mirroring. Manually configuring this attribute, the user takes into account the availability of bandwidth for the mirroring process, where the IBM XIV Storage System adjusts itself to the available bandwidth. Moreover, in some cases the bandwidth is sufficient, but external IOs latency causes the mirroring process to fall behind incoming IOs, thus to repeat replication jobs that were already carried out, and eventually to under-utilize the available bandwidth even if it was adequately allocated.

The IBM XIV Storage System prevents IO time-outs through continuously measuring the IO latency. Excess incoming IOs are pre-queued until they can be submitted. The mirroring rate dynamically adapts to the number of pre-queued incoming IOs, allowing for a smooth operation of the mirroring process.

Attaching volumes to hosts

While the IBM XIV Storage System identifies volumes and snapshots by name, hosts identify volumes and snapshots according to their logical unit number (LUN).

A *LUN* is an integer that is used when attaching a system's volume to a registered host. Each host can access some or all of the volumes and snapshots on the storage system, up to a set maximum. Each accessed volume or snapshot is identified by the host through a LUN.

For each host, a LUN identifies a single volume or snapshot. However, different hosts can use the same LUN to access different volumes or snapshots.

Advanced host attachment

The IBM XIV Storage System provides flexible host attachment options.

The following host attachment options are available:

- Definition of different volume mappings for different ports on the same host
- Support for hosts that have both Fibre Channel and iSCSI ports. Although it is not advisable to use these two protocols together to access the same volume, a dual configuration can be useful in the following cases:
 - As a way to smoothly migrate a host from Fibre Channel to iSCSI
 - As a way to access different volumes from the same host, but through different protocols

CHAP authentication of iSCSI hosts

The MS-CHAP extension enables authentication of initiators (hosts) to the IBM XIV Storage System and vice versa in unsecured environments.

When CHAP support is enabled, hosts are securely authenticated by the IBM XIV Storage System. This increases overall system security by verifying that only authenticated parties are involved in host-storage interactions.

Definitions

The following definitions apply to authentication procedures:

CHAP Challenge Handshake Authentication Protocol

CHAP authentication

An authentication process of an iSCSI initiator by a target through comparing a secret hash that the initiator submits with a computed hash of that initiator's secret which is stored on the target.

Initiator

The host.

Oneway (unidirectional CHAP)

CHAP authentication where initiators are authenticated by the target, but not vice versa.

Supported configurations

CHAP authentication type

Oneway (unidirectional) authentication mode, meaning that the Initiator (host) has to be authenticated by the IBM XIV Storage System.

MDS CHAP authentication utilizes the MDS hashing algorithm.

Access scope

CHAP-authenticated Initiators are granted access to the IBM XIV Storage System via mapping that may restrict access to some volumes.

Authentication modes

The IBM XIV Storage System supports the following authentication modes:

None (default)

In this mode, an initiator is not authenticated by the IBM XIV Storage System.

CHAP (oneway)

In this mode, an initiator is authenticated by the IBM XIV Storage System based on the pertinent initiator's submitted hash, which is compared to the hash computed from the initiator's secret stored on the IBM XIV Storage System.

Changing the authentication mode from None to CHAP requires an authentication of the host. Changing the mode from CHAP to None doesn't require an authentication.

Complying with RFC 3720

The IBM XIV Storage System CHAP authentication complies with the CHAP requirements on the following Web site:<http://tools.ietf.org/html/rfc3720>

Secret length

The secret has to be between 96 bits and 128 bits; otherwise, the system fails the command, responding that the requirements are not fulfilled.

Initiator secret uniqueness

Upon defining or updating an initiator (host) secret, the system compares the entered secret's hash with existing secrets stored by the system and determines whether the secret is unique. If it is not unique, the system presents a warning to the user, but does not prevent the command from completing successfully.

Clustering hosts into LUN maps

To enhance the management of hosts, the IBM XIV Storage System allows clustering them together, where the clustered hosts are provided with identical mappings. The mapping of volumes to LUN identifiers is defined per cluster and applies to all of the hosts in the cluster.

Adding and removing hosts to a cluster are done as follows:

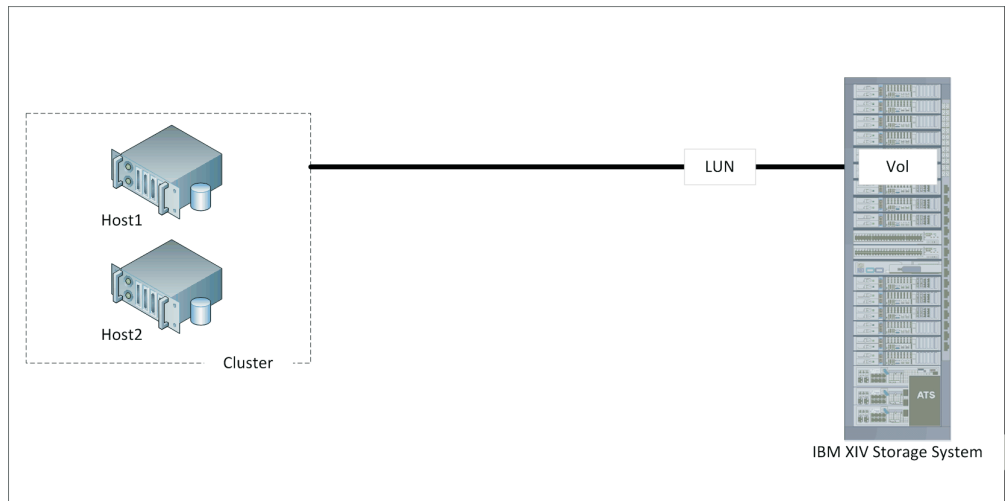


Figure 9. A Volume, a LUN and clustered Hosts

Adding a host to a cluster

Adding a host to a cluster is a straightforward action in which a host is added to a cluster and is connected to a LUN:

- Changing the host's mapping to the cluster's mapping.
- Changing the cluster's mapping to be identical to the mapping of the newly added host.

Removing a host from a cluster

The host is disbanded from the cluster, maintaining its connection to the LUN:

- The host's mapping remains identical to the mapping of the cluster.
- The mapping definitions do not revert to the host's original mapping (the mapping that was in effect before the host was added to the cluster).
- The host's mapping can be changed.

Notes:

- The IBM XIV Storage System defines the same mapping to all of the hosts of the same cluster. No hierarchy of clusters is maintained.
- Mapping a volume to a LUN that is already mapped to a volume.

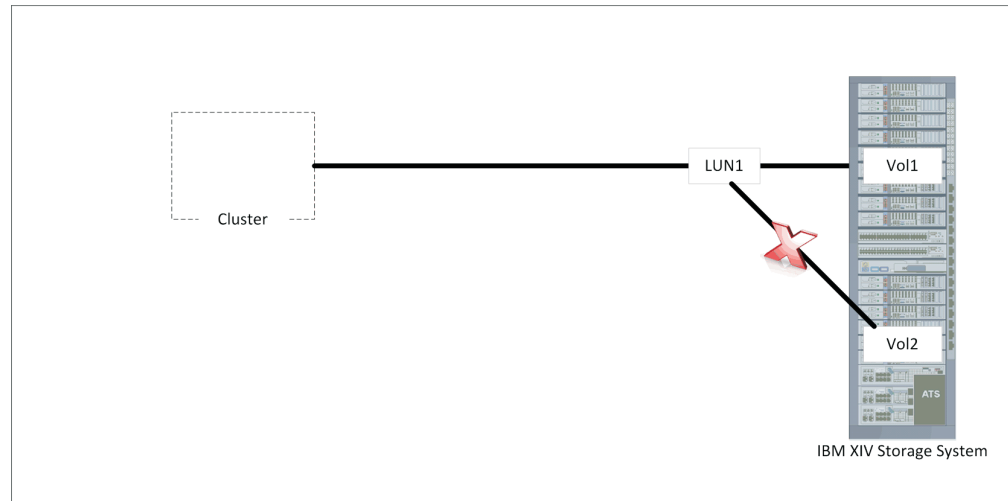


Figure 10. You can't map a volume to a LUN that is already mapped

- Mapping an already mapped volume to another LUN.

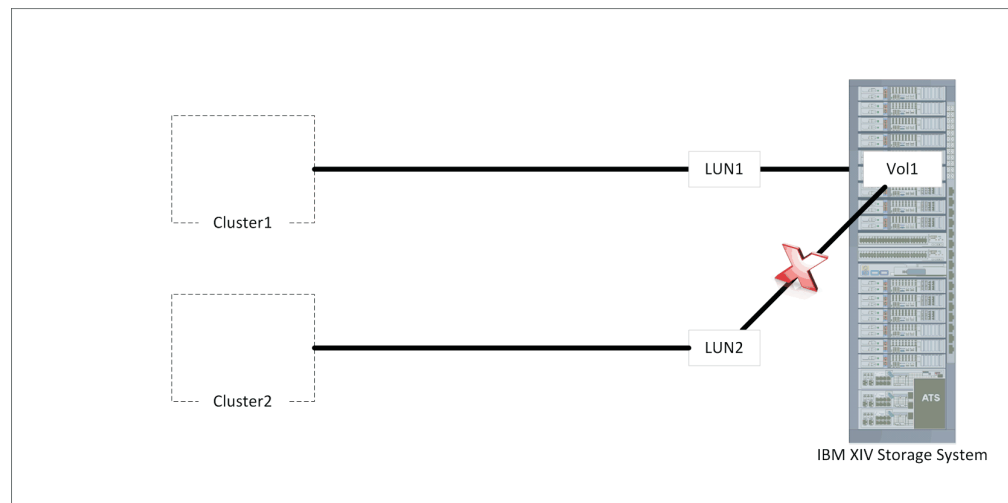


Figure 11. You can't map a volume to a LUN, if the volume is already mapped.

Volume mappings exceptions

The IBM XIV Storage System facilitates association of cluster mappings to a host that is added to a cluster. The system also facilitates easy specification of mapping exceptions for such host; such exceptions are warranted to accommodate cases where a host must have a mapping that is not defined for the cluster (e.g., Boot from SAN).

Mapping a volume to a host within a cluster

It is impossible to map a volume or a LUN that are already mapped.

For example, the host *host1* belongs to the cluster *cluster1* which has a mapping for the volume *vol1* to *lun1*:

1. Mapping *host1* to *vol1* and *lun1* **fails** as both volume and LUN are already mapped.
2. Mapping *host1* to *vol2* and *lun1* **fails** as the LUN is already mapped.
3. Mapping *host1* to *vol1* and *lun2* **fails** as the volume is already mapped.

4. Mapping *host1* to *vol2* and *lun2* **succeeds** with a warning that the mapping is host-specific.

Listing volumes that are mapped to a host/cluster

Mapped Hosts that are part of a Cluster are listed (that is, the list is at a Host-level rather than Cluster-level).

Listing mappings

For each Host, the list indicates whether it belongs to a Cluster.

Adding a host to a cluster

Previous mappings of the Host are removed, reflecting the fact that the only relevant mapping to the Host is the Cluster's.

Removing a host from a cluster

The Host regains its previous mappings.

Supporting VMWare extended operations

The IBM XIV Storage System supports VMWare extended operations that were introduced in VMWare ESX Server 4 (VMWare vStorage API).

The purpose of the VMWare extended operations is to offload operations from the VMWare Server onto the storage system. The IBM XIV Storage System supports the following operations:

Full copy

The ability to copy data from one storage array to another without writing to the ESX Server.

Block zeroing

Zeroing-out a block as a means for freeing it and make it available for provisioning.

Hardware-assisted locking

Allowing for locking volumes within an atomic command.

Writing zeroes

The Write Zeroes command allows for zeroing large storage areas without sending the zeroes themselves.

Whenever a new VM is created, the ESX server creates a huge file full of zeroes and sends it to the storage system. The Write Zeroes command is a way to tell a storage controller to zero large storage areas without sending the zeroes. To meet this goal, both VMWare's generic driver and our own plug-in utilizes the WRITE SAME 16 command.

This method differs from the former method where the host used to write and send a huge file full of zeroes.

Note: The write zeroes operation is not a thin provisioning operation, as its purpose is not to allocate storage space.

Hardware-assisted locking

The hardware-assisted locking feature utilizes VMWare new Compare and Write command for reading and writing the volume's metadata within a single operation.

Upon the replacement of SCSI2 reservations mechanism with Compare and Write by VMWare, the IBM XIV Storage System provides a faster way to change the metadata specific file, along with eliminating the necessity to lock all of the files during the metadata change.

The legacy VMWare SCSI2 reservations mechanism is utilized whenever the VM server performs a management operation, that is to handle the volume's metadata. This method has several disadvantages, among them the mandatory overall lock of access to all volumes, which implies that all other servers are refrained from accessing their own files. In addition, the SCSI2 reservations mechanism entails performing at least four SCSI operations (reserve, read, write, release) in order to get the lock.

The introduction of the new SCSI command, called Compare and Write (SBC-3, revision 22), results with a faster mechanism that is displayed to the volume as an atomic action that doesn't not require to lock any other volume.

Note: The IBM XIV Storage System supports single-block Compare and Write commands only. This restriction is carried out in accordance with VMWare.

Backwards compatibility

The IBM XIV Storage System maintains its compatibility with older ESX versions as follows:

- Each volume is capable of connecting legacy hosts, as it still supports SCSI reservations.
- Whenever a volume is blocked by the legacy SCSI reservations mechanism, it is not available for an arriving COMPARE AND WRITE command.
- The Admin is expected to phase out legacy VM servers to fully benefit from the performance improvement rendered by the hardware-assisted locking feature.

Fast copy

The Fast Copy functionality allows for VN cloning on the storage system without going through the ESX server.

The Fast copy functionality speeds up the VM cloning operation by copying data inside the storage system, rather than issuing READ and WRITE requests from the host. This implementation provide a great improvement in performance, since it saves host to storage system intra-storage system communication. Instead, the functionality utilizes the huge bandwidth within the storage system.

QoS performance classes

The IBM XIV Storage System allows the user to allocate more I/O rates for important applications.

The QoS Performance Classes feature allows the user to allocate more I/O to applications that are considered to be more important, through prioritizing their hosts. Each of the hosts that are connected to the storage system is associated with a group. This group is attributed with a rate limitation.

This limitation attribute and the association of host with the group limit the I/O rates of a specified host in the following way:

- Host rate limitation groups are independent of other forms of host grouping (i.e. Clusters)
- The group can be associated with an unlimited number of hosts
- By default, the host is not associated with any host rate limiting group

Max bandwidth limit attribute

The host rate limitation group has a max bandwidth limit attribute, which is the number of blocks per second. This number could be either:

- A value between *min_rate_limit_bandwidth_blocks_per_sec* and *max_rate_limit_bandwidth_blocks_per_sec* (both are available from the storage system's configuration).
- Zero (0) for unlimited bandwidth.

Commands

Host rate limiting is carried out through the following commands:

Performance class level

- perf_class_create**
Create a Performance Class.
- perf_class_delete**
Delete a Performance Class.
- perf_class_rename**
Renaming a Performance Class.
- perf_class_list**
Listing Performance Classes and their hosts

Host level

- perf_class_add_host**
Adding a host to a group.
- perf_class_remove_host**
Removing a host from a group.
- perf_class_list_hosts**
Listing the group's hosts.
- perf_class_set_rate**
Setting a rate limitation for a group.
- perf_class_get_rate**
Displaying the rate limitation value per group.

Updating the call home contact list

It is important that the contact list remain current to ensure the appropriate personnel is notified in the event of a service issue that requires IBM to contact the customer.

Updating the contact list is done through the XIV GUI. Please refer to the XIV GUI for the procedures for making the appropriate changes.

Host system attachment commands

The IBM XIV Storage System manages host attachment through host commands, cluster commands, and mapping commands.

Table 2 and Figure 12 on page 32 list the command sets that are available.

Table 2. Commands used for host attachment

Command set	Commands
Host commands	For adding and managing hosts and providing them with ports, the following commands are available: • host_define • host_add_port • host_remove_port • host_update • host_list • host_rename • host_delete
Cluster commands	For creating and managing clusters, and populating them with hosts, the following commands are available: • cluster_create • cluster_add_host • cluster_remove_host • cluster_list • cluster_rename • cluster_delete
Mapping commands	For creating and managing maps and populating them with hosts, the following commands are available: • map_vol • unmap_vol • mapping_list • vol_mapping_list
Other	As of today, HP-UX hosts requires a special attachment method, provided by the following command: • special_type_set

Host System Attachment Commands

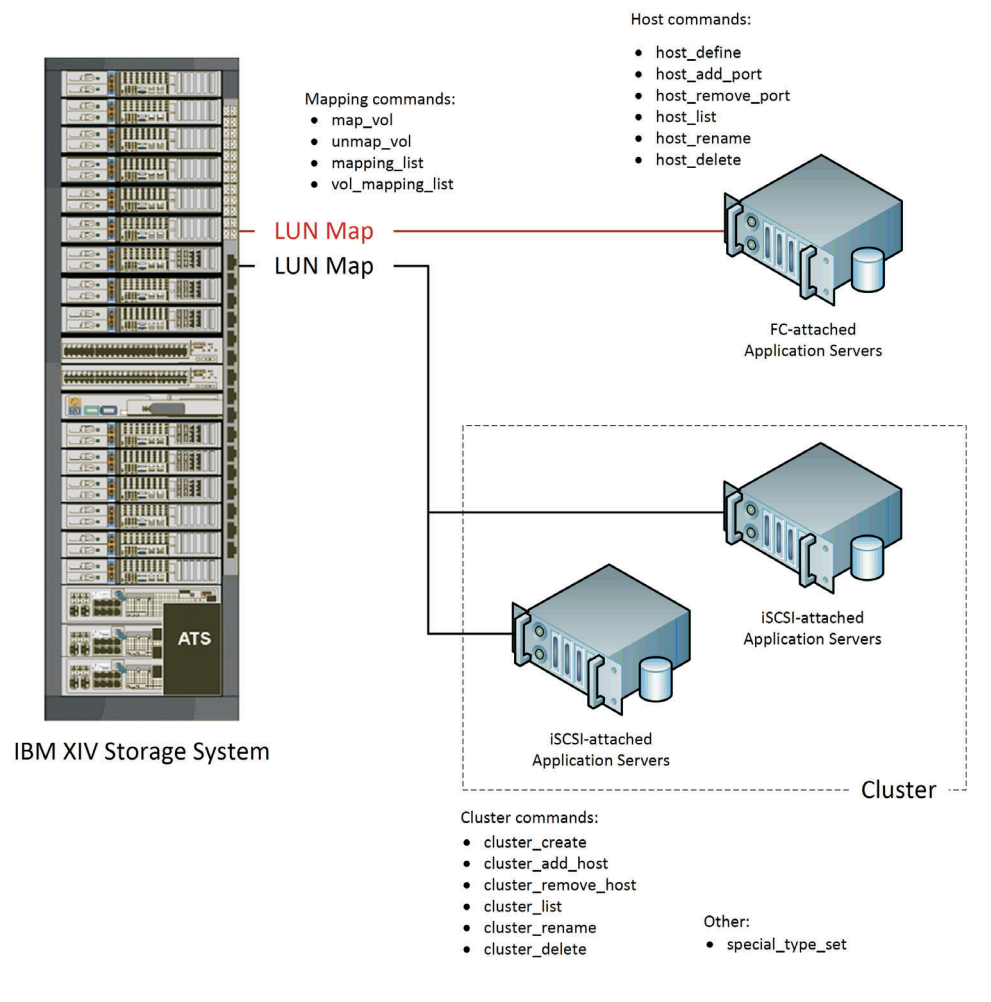


Figure 12. Host system attachment commands

Chapter 4. Consistency groups overview

Consistency groups can be used to take simultaneous snapshots of multiple volumes, thus ensuring consistent copies of a group of volumes.

Creating a synchronized snapshot set is especially important for applications that use multiple volumes concurrently. A typical example is a database application, where the database and the transaction logs reside on different storage volumes, but all of their snapshots must be taken at the same point in time.

This chapter contains the following sections:

- “Creating a consistency group”
- “Taking a snapshot of a consistency group”
- “The snapshot group life cycle” on page 35
- “Restoring a consistency group” on page 36

Creating a consistency group

Consistency groups are created empty and volumes are added to them later on.

The consistency groups is an administrative unit of multiple volumes facilitates simultaneous snapshotting of multiple volumes, mirroring of volume groups, and administration of volume sets.

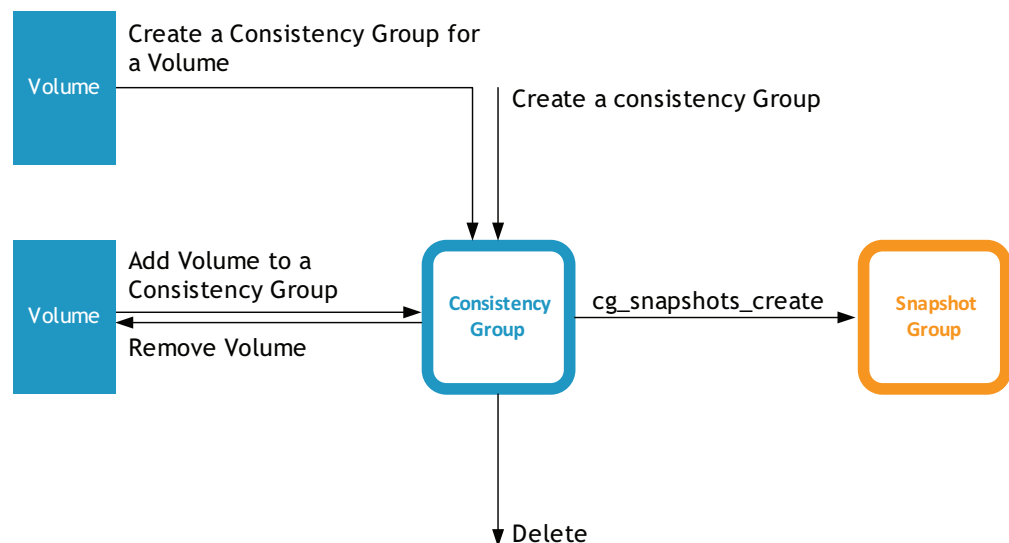


Figure 13. The consistency group's life-cycle

Taking a snapshot of a consistency group

Taking a snapshot for the entire consistency group means that a snapshot is taken for each volume of the consistency group at the same point-in-time. These snapshots are grouped together to represent the volumes of the consistency group at a specific point in time.

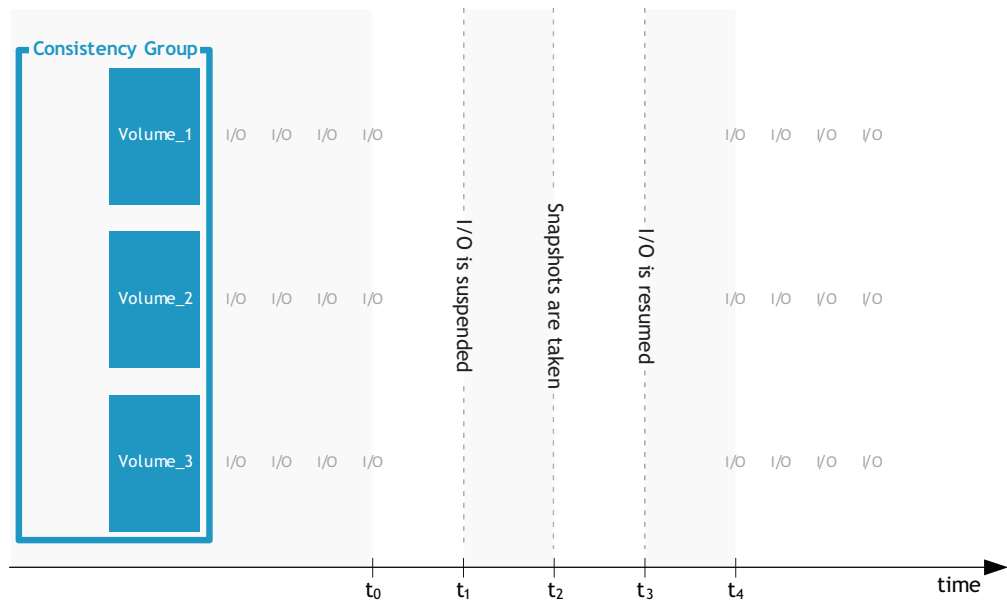


Figure 14. A snapshot is taken for each volume of the consistency group

In Figure 14, a snapshot is taken for each of the consistency group's volumes in the following order:

Time = t_0

Prior to taking the snapshots, all volumes in the consistency group are active and being read from and written to.

Time = t_1

When the command to snapshot the consistency group is issued, I/O is suspended .

Time = t_2

Snapshots are taken at the same point in time.

Time = t_3

I/O is resumed and the volumes continue their normal work.

Time = t_4

After the snapshots are taken, the volumes resume active state and continue to be read from and written to.

Most snapshot operations can be applied to each snapshot in a grouping, known as a *snapshot set*. The following items are characteristics of a snapshot set:

- A snapshot set can be locked or unlocked. When you lock or unlock a snapshot set, all snapshots in the set are locked or unlocked.
- A snapshot set can be duplicated.
- A snapshot set can be deleted. When a snapshot set is deleted, all snapshots in the set are also deleted.

A snapshot set can be disbanded which makes all the snapshots in the set independent snapshots that can be handled individually. The snapshot set itself is deleted, but the individual snapshots are not.

The snapshot group life cycle

Most snapshot operations can be applied to snapshot groups, where the operation affects every snapshot in the group.

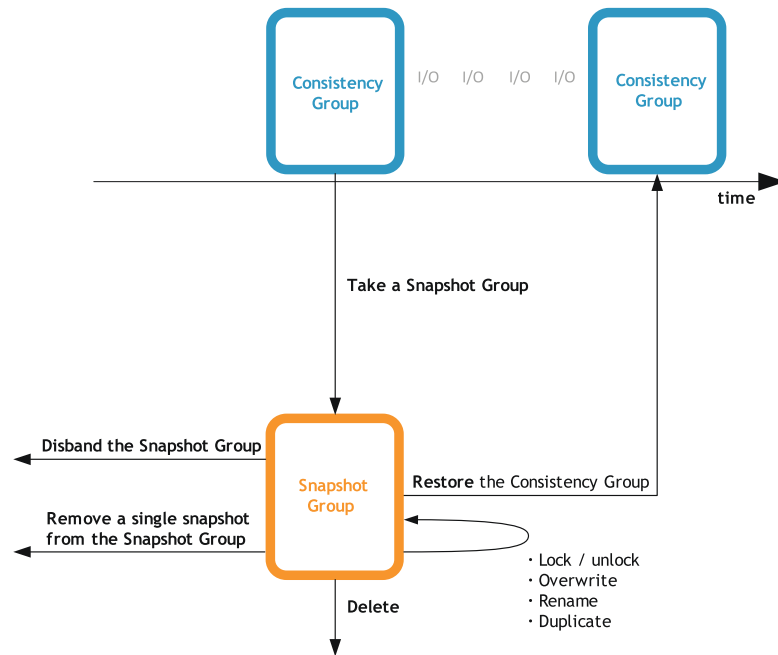


Figure 15. Most snapshot operations can be applied to snapshot groups

Taking a snapshot group

Creates a snapshot group. See “Taking a snapshot of a consistency group” on page 33.

Restoring consistency group from a snapshot group

The main purpose of the snapshot group is the ability to restore the entire consistency group at once, ensuring that all volumes are synchronized to the same point in time. See “Restoring a consistency group” on page 36.

Listing a snapshot group

This command lists snapshot groups with their consistency groups and the time the snapshots were taken.

Note: All snapshots within a snapshot group are taken at the same time.

Lock and unlock

Similar to unlocking and locking an individual snapshot, the snapshot group can be rendered writable, and then be written to. A snapshot group that is unlocked cannot be further used for restoring the consistency group, even if it is locked again.

The snapshot group can be locked again. At this stage, it cannot be used to restore the master consistency group. In this situation, the snapshot group functions like a consistency group of its own.

Overwrite

The snapshot group can be overwritten by another snapshot group.

Rename

The snapshot group can be renamed.

Restricted names

Do not prefix the snapshot group's name with any of the following:

1. most_recent
2. last_replicated

Duplicate

The snapshot group can be duplicated, thus creating another snapshot group for the same consistency group with the time stamp of the first snapshot group.

Disbanding a snapshot group

The snapshots that comprise the snapshot group are each related to its volume. Although the snapshot group can be rendered inappropriate for restoring the consistency group, the snapshots that comprise it are still attached to their volumes. Disbanding the snapshot group detaches all snapshots from this snapshot group but maintains their individual connections to their volumes. These individual snapshots cannot restore the consistency group, but they can restore its volumes individually.

Changing the snapshot group deletion priority

Manually sets the deletion priority of the snapshot group.

Deleting the snapshot group

Deletes the snapshot group along with its snapshots.

Restoring a consistency group

Restoring a consistency group is a single action in which every volume that belongs to the consistency group is restored from a corresponding snapshot that belongs to an associated snapshot group.

Not only does the snapshot group have a matching snapshot for each of the volumes, all of the snapshots have the same time stamp. This implies that the restored consistency group contains a consistent picture of its volumes as they were at a specific point in time.

Note: A consistency group can only be restored from a snapshot group that has a snapshot for each of the volumes. If either the consistency group or the snapshot group has changed after the snapshot group is taken, the restore action does not work.

Chapter 5. Storage pools overview

The storage space of the IBM XIV Storage System is portioned into *storage pools*, where each volume belongs to a specific storage pool.

Storage pools provide the following benefits:

Improved management of storage space

Specific volumes can be grouped together in a storage pool. This enables you to control the allocation of a specific storage space to a specific group of volumes. This storage pool can serve a specific group of applications, or the needs of a specific department.

Improved regulation of storage space

Snapshots can be automatically deleted when the storage capacity that is allocated for snapshots is fully consumed. This automatic deletion is performed independently on each storage pool. Therefore, when the size limit of the storage pool is reached, only the snapshots that reside in the affected storage pool are deleted. For more information, see “The snapshot auto-delete priority” on page 15.

Facilitating thin provisioning

Thin provisioning is enabled by Storage Pools.

Storage pools as logical entities

A storage pool is a logical entity and is not associated with a specific disk or module. All storage pools are equally spread over all disks and all modules in the system.

As a result, there are no limitations on the size of storage pools or on the associations between volumes and storage pools. For example:

- The size of a storage pool can be decreased, limited only by the space consumed by the volumes and snapshots in that storage pool.
- Volumes can be moved between storage pools without any limitations, as long as there is enough free space in the target storage pool.

Note: For the size of the storage pool, please refer to the IBM XIV Storage System data sheet.

All of the above transactions are accounting transactions, and do not impose any data copying from one disk drive to another. These transactions are completed instantly.

Moving volumes between storage pools

For a volume to be moved to a specific storage pool, there must be enough room for it to reside there. If a storage pool is not large enough, the storage pool must be resized, or other volumes must be moved out to make room for the new volume.

A volume and all its snapshots always belong to the same storage pool. Moving a volume between storage pools automatically moves all its snapshots together with the volume.

Protecting snapshots on a Storage Pool level

Snapshots that participates in the mirroring process can be protected in case of Pool space depletion.

This is done by attributing both snapshots (or snapshot groups) and the storage pool with a deletion priority. The snapshots are attributed with a deletion priority between 0 - 4 and the storage pool is configured to disregard snapshots whose priority is above a specific value. Snapshots with a lower delete priority (i.e., higher number) than the configured value might be deleted by the system whenever the Pool space depletion mechanism implies so, thus protecting snapshots with a priority equal or higher to this value.

Follow the link to a full description of the “Pool space depletion” on page 79 mechanism.

Chapter 6. Thin provisioning

The IBM XIV Storage System supports thin provisioning, which provides the ability to define logical volume sizes that are much larger than the physical capacity installed on the system. Physical capacity needs only to accommodate written data, while parts of the volume that have never been written to do not consume physical space.

This chapter discusses:

- Volume hard and soft sizes
- System hard and soft sizes
- Pool hard and soft sizes
- Depletion of hard capacity

Volume hard and soft sizes

Without thin provisioning, the size of each volume is both seen by the hosts and reserved on physical disks. Using thin provisioning, each volume is associated with the following two sizes:

Hard volume size

This reflects the total size of volume areas that were written by hosts. The hard volume size is not controlled directly by the user and depends only on application behavior. It starts from zero at volume creation or formatting and can reach the volume soft size when the entire volume has been written. Resizing of the volume does not affect the hard volume size.

Soft volume size

This is the logical volume size that is defined during volume creation or resizing operations. This is the size recognized by the hosts and is fully configurable by the user. The soft volume size is the traditional volume size used without thin provisioning.

System hard and soft size

Using thin provisioning, each IBM XIV Storage System is associated with a *hard system size* and *soft system size*. Without thin provisioning, these two are equal to the system's capacity. With thin provisioning, these concepts have the following meaning:

Hard system size

This is the physical disk capacity that was installed. Obviously, the system's hard capacity is an upper limit on the total hard capacity of all the volumes. The system's hard capacity can only change by installing new hardware components (disks and modules).

Soft system size

This is the total limit on the soft size of all volumes in the system. It can be set to be larger than the hard system size, up to 79TB. The soft system size is a purely logical limit, but should not be set to an arbitrary value. It must be possible to upgrade the system's hard size to be equal to the soft size, otherwise applications can run out of space. This requirement means that enough floor space should be reserved for future system hardware upgrades, and that the cooling and power infrastructure should be able to

support these upgrades. Because of the complexity of these issues, the setting of the system's soft size can only be performed by IBM XIV support.

Pool hard and soft sizes

The concept of storage pool is also extended to thin provisioning. When thin provisioning is not used, storage pools are used to define capacity allocation for volumes. The storage pools control if and which snapshots are deleted when there is not enough space.

When thin provisioning is used, each storage pool has a soft pool size and a hard pool size, which are defined and used as follows:

Hard pool size

This is the physical storage capacity allocated to volumes and snapshots in the storage pool. The hard size of the storage pool limits the total of the hard volume sizes of all volumes in the storage pool and the total of all storage consumed by snapshots. Unlike volumes, the hard pool size is fully configured by the user.

Soft pool size

This is the limit on the total soft sizes of all the volumes in the storage pool. The soft pool size has no effect on snapshots.

Thin provisioning is managed for each storage pool independently. Each storage pool has its own soft size and hard size. Resources are allocated to volumes within this storage pool without any limitations imposed by other storage pools. This is a natural extension of the snapshot deletion mechanism, which is applied even without thin provisioning. Each storage pool has its own space, and snapshots within each storage pool are deleted when the storage pool runs out of space regardless of the situation in other storage pools.

The sum of all the soft sizes of all the storage pools is always the same as the system's soft size and the same applies to the hard size.

Storage pools provide a logical way to allocate storage resources per application or per groups of applications. With thin provisioning, this feature can be used to manage both the soft capacity and the hard capacity.

Depletion of hard capacity

Thin provisioning creates the potential risk of depleting the physical capacity. If a specific system has a hard size that is smaller than the soft size, the system will run out of capacity when applications write to all the storage space that is mapped to hosts. In such situations, the system behaves as follows:

Snapshot deletion

Snapshots are deleted to provide more physical space for volumes. The snapshot deletion is based on the deletion priority and creation time.

Volume locking

If all snapshots have been deleted and more physical capacity is still required, **all the volumes in the storage pool are locked and no write commands are allowed**. This halts any additional consumption of hard capacity.

Note: Space that is allocated to volumes that is unused (that is, the difference between the volume's soft and hard size) can be used by snapshots in the same storage pool.

The thin provisioning implementation in the IBM XIV Storage System manages space allocation per storage pool. Therefore, one storage pool cannot affect another storage pool. This scheme has the following advantages and disadvantages:

Storage pools are independent

Storage pools are independent in respect to the aspect of thin provisioning. **Thin provisioning volume locking on one storage pool does not create a problem in another storage pool.**

Space cannot be reused across storage pools

Even if a storage pool has free space, this free space is never reused for another storage pool. This creates a situation where volumes are locked due to the depletion of hard capacity in one storage pool, while there is available capacity in another storage pool.

Important: If a storage pool runs out of hard capacity, all of its volumes are locked to all write commands. Although write commands that overwrite existing data can be technically serviced, they are blocked to ensure consistency.

Instant space reclamation

The IBM XIV Storage System supports Symantec's Storage Foundation and Thin Reclamation API.

Chapter 7. Target connectivity

The target connectivity feature controls the way the IBM XIV Storage System utilizes the bandwidth for connected remote systems.

This feature defines the communication topology between a local storage system and a remote storage system, enabling remote mirroring and data migration between the two storage systems. The connectivity can be defined so that both local and target storage systems can write to and read from each other.

This chapter explains the way the bandwidth is prioritized and concepts behind remote target definitions. It also describes the process of manually defining remote targets through CLI commands.

Defining a remote target object

Define a remote target object in order to allow for data migration or mirroring (from other IBM XIV Storage Systems).

The remote target object - a storage system other than the local - is defined through the `target_define` command. The `target_define` command specifies the following connectivity attributes:

target The local name of the target storage system.

protocol

The protocol used to access the remote target (either Fibre Channel or iSCSI).

iscsi_name

The iSCSI name of the target.

system_id

The ID of the target as viewed by the target.

xiv_features

Declares whether the target is an IBM XIV Storage System. If so, mirroring is applicable. If not, only data migration is applicable.

Remote target objects are referenced in the `mirror_create` (the command that creates a remote mirror) and `dm_create` (creating a data migration process) commands.

Target-related commands

Remote target objects can be manipulated through the following commands:

target_delete

Deletes a target. The deleted target is no longer viewed by the local storage system.

target_rename

Renames the target.

Note: The attribute that is renamed is the name of the target as viewed by the local storage system.

target_list

Lists the target's ports along with their activity status and mirroring-related values (see further on this chapter).

target_update

Updates mirroring-related attributes.

Adding ports to remote target

After a remote target object is defined, you can add the remote target port addresses to the local system.

Defining a port is done by simply providing the Fibre Channel port or iSCSI address. This is done according to the following guidelines:

- The port type must conform to the target type.
- Both iSCSI target and the iSCSI initiator ports on the remote system are required for remote mirroring.

The ports are added to the system through the `target_port_add` command. This command provides the local system with the following:

target The name of the target system.

ipaddress

The IP address of the port on the target system (for iSCSI type targets only).

fcaddress

The FC address of the port on the target system (for FC type targets only).

The added port is set to **active** state, making it available for use.

Port-related commands

After the ports are added, they can be managed using the following commands:

target_port_delete

Deletes the address of a specific target port.

target_port_deactivate

Sets the port to be unavailable for use without deleting it.

target_port_activate

Reactivate an inactive port.

target_port_list

Listing ports of a given target along with their addresses and active status.

Connecting between local and target ports

Once added, the target system's ports are connected to the ports of the local system.

The connectivity between ports is an optional path that is available for use. It is defined using the `target_connectivity_define` command:

target The target system.

ipaddress/fcaddress

The port address (ipaddress for iSCSI; fcaddress for FC).

local_ipinterface

The interface on the local system.

local_port

The FC port (applicable for FC connectivity only).

Table 3. Port connectivity commands

Command	Description
target_connectivity_define	This command is used to define connectivity between a local system port and a remote port on a remote target. This command defines the connectivity as an optional path that can be used. The user must ensure that the two ports are actually connected.
fc_connectivity_list ipinterface_run_traceroute	These commands are used to list which remote target ports are visible. Local target ports are shown using the fc_connectivity_list command and remote target IP interfaces are listed using the ipinterface_run_traceroute command.
target_connectivity_list	This command is used to view the current definitions and the current connection status.
target_connectivity_delete target_connectivity_activate target_connectivity_deactivate	These commands are used to delete, activate, and deactivate target connectivity.

Port connectivity-related commands

List commands

fc_connectivity_list

Lists all available targets throughout the FC network.

ipinterface_run_traceroute

Lists all available connectivity to remote IP nodes using the ICMP trace-route mechanism.

target_connectivity_list

Lists the local system's ports definitions and status.

Management commands

target_connectivity_delete

Deletes a port connectivity.

target_connectivity_deactivate

Setting a connection between ports to be unavailable.

target_connectivity_activate

Activates a deactivated connectivity.

Symmetric connectivity for mirroring

Remote mirroring features role switchover between the local and remote system. This requires the target system to have a symmetric connection to the local system. That is, just as the local system have a definition of the remote system, the remote system has to view the local system as its own remote.

Symmetric connectivity is achieved through the target_mirroring_allow command.

Chapter 8. Synchronous remote mirroring

IBM XIV Storage System features synchronous and asynchronous remote mirroring for disaster recovery. Remote mirroring can be used to replicate the data between two geographically remote sites. The replication ensures uninterrupted business operation if there is a total site failure.

Remote mirroring provides data protection for the following types of site disasters:

Site failure

When a disaster happens to a site that is remotely connected to another site, the second site takes over and maintains full service to the hosts connected to the first site. The mirror is resumed after the failing site recovers.

Split brain

After a communication loss between the two sites, each site maintains full service to the hosts. After the connection is resumed, the sites complement each other's data to regain mirroring.

Synchronous and asynchronous remote mirroring

The two distinct methods of remote mirroring - synchronous and asynchronous - are described on this chapter and on the following chapter. Throughout this chapter, the term *remote mirroring* refers to *synchronous* remote mirroring, unless clearly stated otherwise.

Remote mirroring basic concepts

Synchronous remote mirroring provides continuous availability of critical information in the case of a disaster scenario.

A typical remote mirroring configuration involves the following two sites:

Primary site

The location of the master storage system.

A local site that contains both the data and the active servers.

Secondary site

The location of the secondary storage system.

A remote site that contains a copy of the data and standby servers.

Following a disaster at the master site, the servers at the secondary site become active and start using the copy of the data.

Master

The volume or storage system which is mirrored. The master volume or storage system is usually at the master site.

Slave The volume or storage system to which the master is mirrored. The slave volume or storage system is usually at the Secondary site.

One of the main goals of remote mirroring is to ensure that the secondary site contains the same (consistent) data as the master site. With remote mirroring, services are provided seamlessly by the hosts and storage systems at the secondary site.

The process of ensuring that both storage systems contain identical data at all times is called *remote mirroring*. Synchronous remote mirroring is performed during each write operation. The write operation issued by a host is sent to both the master and the slave storage systems.

To ensure that the data is also written to the secondary system, acknowledgement of the write operation is only issued after the data has been written to both storage systems. This ensures the consistency of the secondary storage system to the master storage system except for the contents of any last, unacknowledged write operations. This form of mirroring is called synchronous mirroring.

In a remote mirroring system, reading is performed from the master storage system, while writing is performed on both the master and the slave storage systems, as previously described.

The IBM XIV Storage System supports configurations where server pairs perform alternate master or secondary roles with respect to their hosts. As a result, a server at one site might serve as the master storage system for a specific application, while simultaneously serving as the secondary storage system for another application.

Remote mirroring operation

Remote mirroring operations involve *configuration, initialization, ongoing operation, handling of communication failures, and role switching* activities.

The following list defines the remote mirroring operation activities:

Configuration

Local and remote replication peers are defined by an administrator who specifies the primary and secondary volumes. For each coupling, several configuration options can be defined.

Initialization

Remote mirroring operations begin with a master volume that contains data and a formatted slave volume. The first step is to copy the data from the master volume to the slave volume. This process is called *initialization*. Initialization is performed once in the lifetime of a remote mirroring coupling. After it is performed, both volumes are considered synchronized.

Ongoing Operation

After the initialization process is complete, remote mirroring is activated. During this activity, all data is written to the master volume and then to the slave volume. The write operation is complete after an acknowledgement from the slave volume. At any point, the master and slave volumes are identical except for any unacknowledged (pending) writes.

Handling of Communication Failures

From time to time the communication between the sites might break down, and it is usually preferable for the primary site to continue its function and to update the secondary site when communication resumes. This process is called *synchronization*.

Role Switching

When needed, a replication peer can change its role from master to slave

or vice versa, either as a result of a disaster at the primary site, maintenance operations, or because of a drill that tests the disaster recovery procedures.

Configuration options

The remote mirroring configuration process involves configuring volumes and volume pair options.

When a pair of volumes point to each other, it is referred to as a *coupling*. In a *coupling relationship*, two volumes participate in a remote mirroring system with the slave peer serving as the backup for the master peer. The coupling configuration is identical for both master volumes and slave volumes.

Table 4. Configuration options for a volume

Name	Values	Definition
Role	None, Master, Slave	Role of a volume. (Primary and Secondary are designations.)
Peer	Remote target identification and the name of the volume on the remote target.	Identifies the peer volume.

For more information about configuring a volume, see “Volume configuration.”

Table 5. Configuration options for a coupling

Name	Values	Definition
Activation	Active, Stand-by.	Activates or deactivates remote mirroring.

For more information about configuring a coupling, see “Communication errors” on page 50.

Volume configuration

The role of each volume and its peer volumes on the IBM XIV Storage System must be defined for it to function within the remote mirroring process.

The following concepts are to be configured for volumes and the relations between them:

- Volume role
- Peer

The *volume role* is the current function of the volume. The following volume roles are available:

None The volume is created using normal volume creation procedures and is not configured as part of any remote mirroring configuration.

Master volume

The volume is part of a mirroring coupling and serves as the master volume.

All write operations are made to this master volume. It ensures that write operations are made to the slave volume before acknowledging their success.

Slave volume

This volume is part of a mirroring coupling and serves as a backup to the master volume.

Data is read from the slave volume, but cannot be written to it.

A *peer* is a volume that is part of a coupling. A volume with a role other than none has to have a peer designation, and a corresponding master or slave volume assigned to it.

Configuration errors

In some cases, configuration on both sides might be changed in a non-compatible way. This is defined as a *configuration error*. For example, switching the role of only one side when communication is down causes a configuration error when connection resumes.

Mixed configuration

The volumes on a single storage system can be defined in a mixture of configurations.

For example, a storage system can contain volumes whose role is defined as master, as well as volumes whose roles are defined as slave. In addition, some volumes might not be involved in a remote mirroring coupling at all.

The roles assigned to volumes are *transient*. This means a volume that is currently a master volume can be defined as a slave volume and vice versa. The term *local* refers to the master volume, and *remote* refers to the slave volume for processes that switch the master and slave assignments.

Communication errors

When the communication link to the secondary volume fails or the secondary volume itself is not usable, processing to the volume continues as usual. The following occurs:

- The system is set to an unsynchronized state.
- All changes to the master volume are recorded and then applied to the slave volume after communication is restored.

Coupling activation

Remote mirroring can be manually activated and deactivated per coupling. When it is activated, the coupling is in *Active* mode. When it is deactivated, the coupling is in *Standby* mode.

These modes have the following functions:

Active Remote mirroring is functioning and the data is being written to both the master and the slave volumes.

Standby

Remote mirroring is deactivated. The data is not being written to the slave volume, but it is being recorded in the master volumes which will later synchronize the slave volume.

Standby mode is used mainly when maintenance is performed on the secondary site or during communication failures between the sites. In this mode, the master volumes do not generate alerts that the mirroring has failed.

The coupling lifecycle has the following characteristics:

- When a coupling is created, it is always initially in Standby mode.
- Only a coupling in Standby mode can be deleted.
- Transitions between the two states can only be performed from the UI and on the volume.

Synchronous mirroring statuses

The status of the synchronous remote mirroring volume represents the state of the storage volume in regard to its remote mirroring operation.

The state of the volume is a function of the status of the communication link and the status of the coupling between the master volume and the slave volume.

Table 6 describes the various statuses of a synchronous remote mirroring volume during remote mirroring operations.

Table 6. Synchronous mirroring statuses

Entity	Name	Values	Definition
Link	Status	• Up • Down	Specifies if the communications link is up or down.
Coupling	Operational status	• Operational • Non-operational	Specifies if remote mirroring is working.
	Synchronization status	• Initialization • Synchronized • Unsynchronized • Consistent • Inconsistent	Specifies if the master and slave volumes are consistent.
	Last-secondary-timestamp	Point-in-time date	Time stamp for when the secondary volume was last synchronized.
	Synchronization process progress	Synchronization status	Amount of data remaining to be synchronized between the master and slave volumes due to non-operational coupling.
	Secondary locked	Boolean	True, if secondary was locked for writing due to lack of space; otherwise false. This can happen during the synchronization process when there is not enough space for the last-consistent snapshot.
	Configuration error	Boolean	True, if the configuration of the master and secondary slave is inconsistent.

Link status

The status of the communication link can be either *up* or *down*. The link status of the master volume is, of course, also the link status of the slave volume.

Operational status

The coupling between the master and slave volumes is either *operational* or *non-operational*. To be operational, the link status must be up and the coupling must be activated. If the link is down or if the remote mirroring feature is in Standby mode, the operational status is non-operational.

Synchronization status

The synchronization status reflects the consistency of the data between the master and slave volumes. Because the purpose of the remote mirroring feature is to ensure that the slave volume is an identical copy of the master volume, this status indicates whether this objective is currently attained.

The possible synchronization statuses for the master volume are:

Initialization

The first step in remote mirroring is to create a copy of the data from the master volume to the slave volume, at the time when the mirroring was set to place. During this step, the coupling status remains initialization.

Synchronized (master volume only)

This status indicates that all data that was written to the primary volume and acknowledged has also been written to the secondary volume. Ideally, the primary and secondary volumes should always be synchronized. This does not imply that the two volumes are identical because at any time, there might be a limited amount of data that was written to one volume, but was not yet written to its peer volume. This means that their write operations have not yet been acknowledged. These are also known as pending writes.

Unsynchronized (primary volume only)

After a volume has completed the initialization stage and achieved the synchronized status, a volume can become unsynchronized.

This occurs when it is not known whether all the data that was written to the primary volume was also written to the secondary volume. This status occurs in the following cases:

- **Communications link is down** - As a result of the communication link going down, some data might have been written to the primary volume, but was not yet written to the secondary volume.
- **Secondary system is down** - This is similar to communication link errors because in this state, the primary system is updated while the secondary system is not.
- **Remote mirroring is deactivated** - As a result of the remote mirroring deactivation, some data might have been written to the primary volume and not to the secondary volume.

It is always possible to reestablish the synchronized status when the link is reestablished or the remote mirroring feature is reactivated, no matter what was the reason for the unsynchronized status.

Because all updates to the primary volume that are not written to the secondary volume are recorded, these updates are written to the secondary volume. The

synchronization status remains unsynchronized from the time that the coupling is not operational until the synchronization process is completed successfully.

Synchronization progress status

During the synchronization process while the secondary volumes are being updated with previously written data, the volumes have a dynamic synchronization process status.

This status is comprised of the following sub-statuses:

Size to complete

The size of data that requires synchronization.

Part to synchronize

The size to synchronize divided by the maximum size-to-synchronize since the last time the synchronization process started. For coupling initialization, the size-to-synchronize is divided by the volume size.

Time to synchronize

Estimate of the time, which is required to complete the Synchronization process and achieve synchronization, based on past rate.

Last secondary time stamp

A time stamp is taken when the coupling between the primary and secondary volumes becomes non-operational.

This time stamp specifies the last time that the secondary volume was consistent with the primary volume. This status has no meaning if the coupling's synchronization state is still *initialization*. For synchronized coupling, this timestamp specifies the current time. Most importantly, for an unsynchronized coupling, this timestamp denotes the time when the coupling became non-operational.

The timestamp is returned to current only after the coupling is operational and the primary and secondary volumes are synchronized.

I/O operations

I/O operations are performed on the primary and secondary volumes across various configuration options.

I/O on the primary

Read All data is read from the primary (local) site regardless of whether the system is synchronized.

Write

- If the coupling is operational, data is written to both the primary and secondary volumes.
- If the coupling is non-operational, an error is returned.

The error reflects the type of problem that was encountered. For example, remote mirroring has been deactivated, there is a locked secondary error, or there is a link error.

I/O on the secondary

A secondary volume can have LUN maps and hosts associated with it, but it is only accessible as a read-only volume. These maps are used by the backup hosts

when a switchover is performed. When the secondary volume becomes the primary volume, hosts can write to it on the remote site. When the primary volume becomes a secondary volume, it becomes read-only and can be updated only by the new primary volume.

Read Data is read from the secondary volume like from any other volume.

Write An attempt to write on the secondary volume results in a volume read-only SCSI error.

Synchronization process

When a failure condition has been resolved, remote mirroring begins the process of synchronizing the coupling. This process updates the secondary volume with all the changes that occurred while the coupling was not operational.

This section describes the process of synchronization.

State diagram

Couplings can be in the *Initialization*, *Synchronized*, *Timestamp*, or *Unsynchronized* state.

The following diagram shows the various coupling states that the IBM XIV Storage System assumes during its lifetime, along with the actions that are performed in each state.

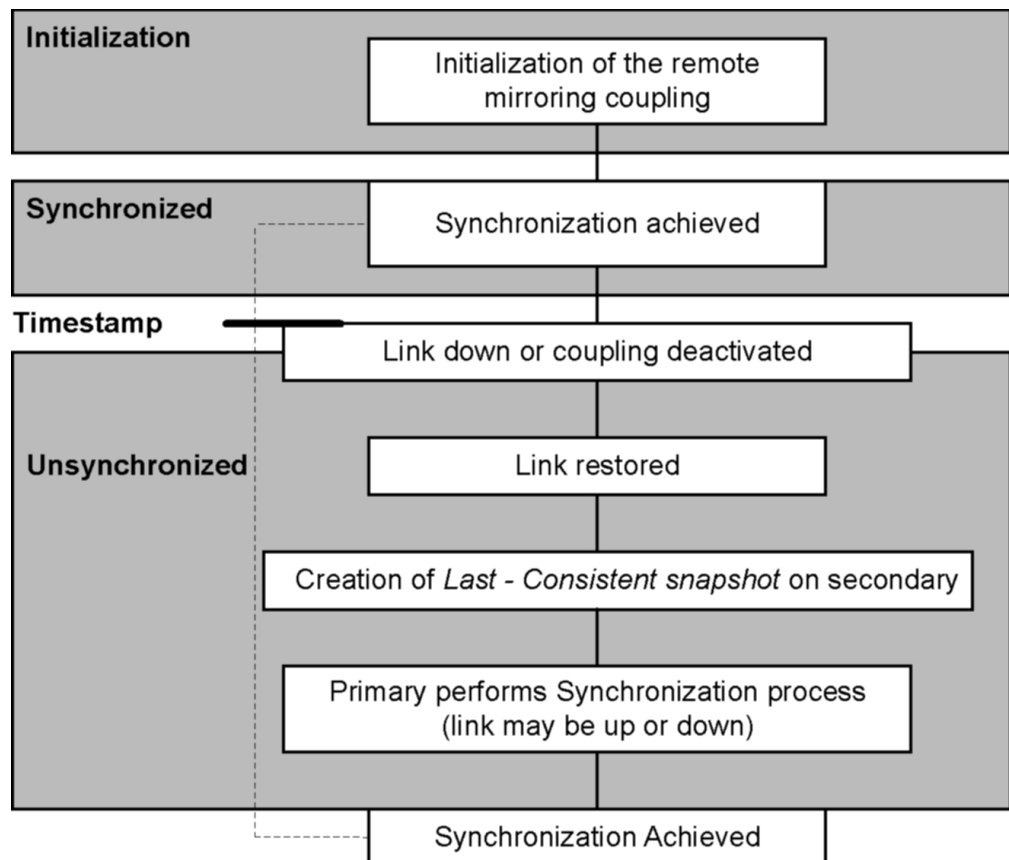


Figure 16. Coupling states and actions

The following list describes each coupling state:

Initialization

The secondary volume has a Synchronization status of Initialization. During this state, data from the primary volume is copied to the secondary volume.

Synchronized

This is the working state of the coupling, where both the primary and secondary volumes are consistent.

Timestamp

Remote mirroring has become non-operational so a time stamp is recorded. During this status, the following actions take place:

1. Coupling deactivation, or the link is down
2. Coupling is reactivated, or the link is restored.

Unsynchronized

Remote mirroring is non-operational because of a communications failure or because remote mirroring was deactivated. Therefore, the primary and secondary volumes are not synchronized. When remote mirroring resumes, steps are taken to return to the Synchronized state.

Coupling recovery

Remote mirroring recovers from non-operational coupling.

When remote mirroring recovers from a non-operational coupling, the following actions take place:

- If the secondary volume is in the Synchronized state, a last-consistent snapshot of the secondary volume is created and named with the string `secondary-volume-time-date-consistent-state`.
- The primary volume updates the secondary volume until it reaches the Synchronized state.
- The primary volume deletes the special snapshot after all couplings that mirror volumes between the same pair of systems are synchronized.

Uncommitted data

When the coupling is in an Unsynchronized state, for Best-effort coupling, the system must track which data in the primary volume has been changed, so that these changes can be committed to the secondary when the coupling becomes operational again.

The parts of the primary volume that must be committed to the secondary volume and must be marked are called *uncommitted data*.

Note: There is only metadata that marks the parts of the primary volume that must be written to the secondary volume when the coupling becomes operational.

Constraints and limitations

Coupling has constraints and limitations.

The following constraints and limitations exist:

- The Size, Part, or Time-to-synchronize are relevant only if the Synchronization status is Unsynchronized.

- The last-secondary-time stamp is only relevant if the coupling is Unsynchronized.

Last-consistent snapshots

Before the synchronization process is initiated, a snapshot of the secondary volume is created. This snapshot is created to ensure the usability of the secondary volume in case of a primary site disaster during the synchronization process.

If the primary volume is destroyed before the synchronization is completed, the secondary volume might be inconsistent because it may have been only partially updated with the changes that were made to the primary volume. The reason for this possible inconsistency is the fact that the updates were not necessarily performed in the same order in which they were written by the hosts.

To handle this situation, the primary volume always creates a snapshot of the last-consistent secondary volume after re-connecting to the secondary machine, and before starting the synchronization process.

The last consistent snapshot

The Last Consistent snapshot (LCS) is created by the system on the Slave peer in synchronous mirroring just before mirroring resynchronization needs to take place. Mirroring resynchronization takes place after link disruption, or a manual mirroring deactivation. In both cases the Master will continue to accept host writes, yet will not replicate them onto the Slave as long as the link is down, or the mirroring is deactivated.

Once the mirroring is restored and activated, the system takes a snapshot of the Slave (LCS), which represents the data that is known to be mirrored, and only then the non yet mirrored data written to the Master is replicated onto the Slave through a resynchronization process.

The LCS is deleted automatically by the system once the resynchronization is complete for all mirrors on the same target , but if the Slave peer role is changed during resynchronization $\triangle$ this snapshot will not be deleted.

The external last consistent snapshot

Prior to the introduction of the external last consistent snapshot, whenever the peer's role was changed back to Slave and sometime whenever a new resynchronization process had started, the system would detect an LCS on the peer and would not create a new one. If, during such an event, the peer was not part of a mirrored consistency group (mirrored CG) it would mean that not all volumes have the same LCS timestamp. If the peer was part of a mirrored consistency group, we would have a consistent LCS but not as current as possibly expected. This situation is avoided thanks to the introduction of the external last consistent snapshot.

Whenever the role of a Slave with an LCS is changed to Master while mirroring resynchronization is in progress (in the system/target not specific to this volume), the LCS is renamed external last consistent (ELCS). The ELCS retains the LCS deletion priority of 0. If the peer's role is later changed back to Slave and sometime afterwards a new resynchronization process starts, a new LCS will be created.

Subsequently changing the Slave role again will rename the existing external last consistent snapshot to external last consistent x (x being the first available number starting from 1) and will rename the LCS to external last consistent. The deletion priority of external last consistent will be 0, but the deletion priority of the new external last consistent x will be the system default (1), and can thus be deleted automatically by the system upon pool space depletion.

It is crucial to validate whether the LCS or an ELCS (or even ELC x) should serve as a restore point for the Slave peer if resynchronization cannot be completed. While snapshots with deletion priority 0 are not automatically deleted by the system to free space, the external last consistent and external last consistent x snapshots can be manually deleted by the administrator if so required. As the deletion of such snapshots might leave an inconsistent peer without a consistent snapshot to be restored from (in case the resynchronization cannot complete due to Master unavailability) $\triangle$ it should generally be avoided even when pool space is depleted, unless the peer is guaranteed to be consistent.

Manually deleting the last consistent snapshot

- Only the XIV support team can delete the last consistent snapshot. The XIV support team uses the `delete_mirror_snapshots` CLI command.
- The XIV support team can also configure a mirroring so that it does not create the last consistent snapshot. This is required when the system that contains the secondary volume is fully utilized and an additional snapshot cannot be created.

Last consistent snapshot timestamp

A timestamp is taken when the coupling between the primary and secondary volumes becomes non-operational. This timestamp specifies the last time that the secondary volume was consistent with the primary volume.

This status has no meaning if the coupling's synchronization state is still *Initialization*. For synchronized couplings, this timestamp specifies the current time. Most importantly, for unsynchronized couplings, this timestamp denotes the time when the coupling became non-operational.

Table 7 provides an example of a failure situation and describes the time specified by the timestamp.

Table 7. Example of the last consistent snapshot time stamp process

Time	Status of coupling	Action	Last consistent timestamp
Up to 12:00	Operational and synchronized		Current
12:00 - 1:00	Failure caused the coupling to become non-operational. The coupling is Unsynchronized.	Writing continues to the primary volume. Changes are marked so that they can be committed later.	12:00
13:00	Connectivity resumes and remote mirroring is operational. Synchronization begins. The coupling is still Unsynchronized.	All changes since the connection was broken are committed to the secondary volume, as well as current write operations.	12:00
13:15	Synchronized		Current

Secondary locked error status

While the synchronization process is in progress, there is a period in which the secondary volume is not consistent with the primary volume, and a last-consistent snapshot is maintained. While in this state, I/O operations to the secondary volume can fail because there is not enough space. There is not enough space because every I/O operation potentially requires a copy-on-write of a partition.

Whenever I/O operations fail because there is not enough space, all couplings in the system are set to the secondary-locked status and the coupling becomes non-operational. The administrator is notified of a critical event, and can free space on the system containing the secondary volume.

If this situation occurs, contact an IBM XIV field engineer. After there is enough space, I/O operations resume and remote mirroring can be reactivated.

Role switchover

Role switchover when remote mirroring is operational

Role switching between primary and secondary volumes can be performed from the IBM XIV Storage Management GUI or the XCLI after the remote mirroring function is operational. After role switchover occurs, the primary volume becomes the secondary volume and vice versa.

There are two typical reasons for performing a switchover when communications between the volumes exist. These are:

Drills Drills can be performed on a regular basis to test the functioning of the secondary site. In a drill, an administrator simulates a disaster and tests that all procedures are operating smoothly.

Scheduled maintenance

To perform maintenance at the primary site, switch operations to the secondary site on the day before the maintenance. This can be done as a preemptive measure when a primary site problem is known to occur.

This switchover is performed as an automatic operation acting on the primary volume. This switchover cannot be performed if the primary and secondary volumes are not synchronized.

Role switchover when remote mirroring is nonoperational

A more complex situation for role switching is when there is no communication between the two sites, either because of a network malfunction, or because the primary site is no longer operational. The XCLI command for this scenario is `reverse_role`. Because there is no communication between the two sites, the command must be issued on both sites concurrently, or at least before communication resumes.

Switchover procedures differ depending on whether the primary and secondary volumes are connected or not. As a general rule, the following is true:

- When the coupling is deactivated, it is possible to change the role on one side only, assuming that the other side will be changed as well before communication resumes.

- If the coupling is activated, but is either unsynchronized, or nonoperational due to a link error, an administrator must either wait for the coupling to be synchronized, or deactivate the coupling.
- On the secondary volume, an administrator can change the role even if coupling is active. It is assumed that the coupling will be deactivated on the primary volume and the role switch will be performed there as well in parallel. If not, a configuration error occurs.

Switch secondary to primary

The role of the secondary volume can be switched to primary using the IBM XIV Storage Management GUI or the XCLI. After this switchover, the following is true:

- The secondary volume is now the primary.
- The coupling has the status of unsynchronized.
- The coupling remains in standby mode, meaning that the remote mirroring is deactivated. This ensures an orderly activation when the role of the other site is switched.

The new primary volume starts to accept write commands from local hosts. Because coupling is not active, in the same way as any primary volume, it maintains a log of which write operations should be sent to the secondary when communication resumes.

Typically, after switching the secondary to the primary volume, an administrator also switches the primary to the secondary volume, at least before communication resumes. If both volumes are left with the same role, a configuration error occurs.

Secondary consistency

If the user is switching the secondary to a primary volume, and a snapshot of the last-consistent state exists, then the link was broken during the process of synchronizing. In this case, the user has a choice between using the most-updated version, which might be inconsistent, or reverting to the last_consistent snapshot. Table 8 outlines this process.

Table 8. Disaster scenario leading to a secondary consistency decision

Time	Status of remote mirroring	Action
Up to 12:00	Operational	Volume A is the primary volume and volume B is the secondary volume.
12:00	Nonoperational because of communications failure	Writing continues to volume A and volume A maintains the log of changes to be committed to volume B.
13:00	Connectivity resumes and remote mirroring is operational.	A last_consistent snapshot is generated on volume B. After that, volume A starts to update volume B with the write operations that occurred since communication was broken.
13:05	Primary site is destroyed and all information is lost.	
13:10		Volume B is becoming the primary. The operators can choose between using the most-updated version of volume B, which contains only parts of the write operations committed to volume A between 12:00 and 13:00, or use the last-consistent snapshot, which reflects the state of volume B at 12:00.

If a last-consistent snapshot exists and the role is changed from secondary to primary, the system automatically generates a snapshot of the volume. This

snapshot is named *most_updated snapshot*. It is generated to enable a safe reversion to the latest version of the volume, when recovering from user errors. This snapshot can only be deleted by the IBM XIV Storage System support team.

If the coupling is still in the initialization state, switching cannot be performed. In the extreme case where the data is needed even though the initial copy was not completed, a volume copy can be used on the primary volume.

Switch primary to a secondary

When coupling is inactive, the primary machine can switch roles. After such a switch, the primary volume becomes the secondary one.

Unsynchronized primary becoming a secondary

Because the primary volume is inactive, it is also in the unsynchronized state, and it might have an uncommitted data list. All these changes will be lost. When the volume becomes secondary, this data must be reverted to match the data on the peer volume, which is now the new primary volume. In this case, an event is created, summarizing the size of the changes that were lost.

The uncommitted data list has now switched its semantics, and instead of being a list of updates that the local volume (old primary, new secondary) needs to update on the remote volume (old secondary, new primary), the list now represents the updates that need to be taken from the remote to the local volume.

Upon reestablishing the connection, the local volume (current secondary, which was the primary) will update the remote volume (new primary) with this uncommitted data list to update, and it is the responsibility of the new primary volume to synchronize these lists to the local volume (new secondary).

Resumption of remote mirroring after role change

When the communication link is resumed after a switchover of roles in which both sides were switched, the coupling now contains one secondary and one primary volume. See “Reconnection when both sides have the same role” on page 61 for the description of the behavior of the system if there is a user error and both sides have the same role.

Note: After a role switchover, the coupling is in standby. The coupling can be activated before or after the link resumes.

Table 9 on page 61 describes the system when the coupling becomes operational, meaning after the communications link has been resumed and the coupling has been reactivated. When communications is resumed, the new primary volume (old secondary) might be in the unsynchronized state, and have an uncommitted data list to synchronize.

The new secondary volume (old primary), might have an uncommitted data list to synchronize from the new primary volume. These are write operations that were written after the link was broken and before the role of the volume was switched from primary to secondary. These changes must be reverted to the content of the new primary volume. Both lists must be used for synchronization by the new primary volume.

Table 9. Resolution of uncommitted data for synchronization of the new primary volume

Time	Status of remote mirroring	Action
Up to 12:00	Operational and synchronized	Volume A is the primary volume and volume B is the secondary volume.
12:00 to 12:30	Communication failure, coupling becomes nonoperational	Volume A still accepts write operations from the hosts and maintains an uncommitted data list marking these write operations. For example, volume A accepted a write operation to blocks 1000 through 2000, and marks blocks 1000 through 2000 as ones that need to be copied to volume B after reconnection.
12:30	Roles changed on both sides	Volume A is now secondary and volume B is primary. Volume A should now revert the changes done between 12:00 and 12:30 to their original values. This data reversion is only performed after the two systems reconnect. For now, volume A reverts the semantics of the uncommitted data list to be data that needs to be copied from volume B. For example, blocks 1000 through 2000 need to be copied now from volume B.
12:30 to 13:00	Volume B is primary, volume A is secondary, coupling is nonoperational	Volume A does not accept changes because it is a secondary in a nonoperational coupling. volume B is now a primary in a nonoperational coupling, and maintains its own uncommitted data list of the write operations that were performed since it was defined as the primary. For example, if the hosts wrote blocks 1500 through 2500, volume B marks these blocks to be copied to volume A.
13:00	Connectivity resumes	Volume B and volume A communicate and volume B merges the lists of uncommitted data. Volume B copies to volume A both the blocks that changed in volume B between 12:30 and 13:00, as well as the blocks that changed in volume A between 12:00 and 12:30. For example, volume B could copy to volume A blocks 1000 through 2500, where blocks 1000 through 1500 would revert to their original values at 12:00 and blocks 1500 through 2500 would have the values written to volume B between 12:30 and 13:00. Changes written to volume A between 12:00 and 12:30 are lost.

Reconnection when both sides have the same role

Situations where both sides are configured to the same role can only occur when one side was switched while the link was down. This is a user error, and the user must follow these guidelines to prevent such a situation:

- Both sides need to change roles before the link is resumed.
- As a safety measure, it is recommended to first switch the primary to secondary.

If the link is resumed and both sides have the same role, the coupling will not become operational.

To solve the problem, the user must use the role switching mechanism on one of the volumes and then activate the coupling.

In this situation, the system behaves as follows:

- If both sides are configured as secondary volumes, a minor error is issued.
- If both sides are configured as primary volumes, a critical error is issued. Both volumes will be updated locally with remote mirroring being nonoperational until the condition is solved.

Miscellaneous

Remote mirroring and consistency groups

The following assumptions ensure that consistency group procedures are compatible with the remote mirroring function:

- All volumes in a consistency group are mirrored on the same system (as all primaries on a system are mirrored on the same system).
- All volumes in a consistency group have the same role.
- The last_consistent snapshot is created and deleted system-wide, and therefore it is consistent across the consistency group.

Note: An administrator can incorrectly switch the roles of some of the volumes in a consistency group, while keeping others in their original role. This is not prevented by the system and is detected at the application level.

Using remote mirroring for media error recovery

If a media error is discovered on one of the volumes of the coupling, the peer volume is then used for a recovery.

Supported configurations

- Either fibre-channel or iSCSI can serve as the protocol between the primary and secondary volumes. A volume accessed through one protocol can be synchronized using another protocol.
- The remote system must be defined as an XIV in the remote-target connectivity definitions.
- All the peers of volumes that belong to the same consistency group on a system must reside on a single remote system.
- Primary and secondary volumes must contain the same number of blocks.

I/O performance versus synchronization speed optimization

The IBM XIV Storage System has two global parameters, controlling the maximum rate used for initial synchronization and for synchronization after nonoperational coupling.

These rates are used to prevent a situation where synchronization uses too many of the system or communication line resources.

This configuration parameter can be changed at any time. These parameters are set by the IBM XIV Storage System technical support representative.

Implications regarding other commands

- Renaming a volume changes the name of the last_consistent and most_updated snapshots.
- Deleting all snapshots does not delete the last_consistent and most_updated snapshot.
- Resizing a primary volume resizes its secondary volume.
- A primary volume cannot be resized when the link is down.
- Resizing, deleting, and formatting are not permitted on a secondary volume.

- A primary volume cannot be formatted. If a primary volume must be formatted, an administrator must first deactivate the mirroring, delete the mirroring, format both the secondary and primary volumes, and then define the mirroring again.
- Secondary or primary volumes cannot be the target of a copy operation.
- Locking and unlocking are not permitted on a secondary volume.
- Last_consistent and most_updated snapshots cannot be unlocked.
- Deleting is not permitted on a primary volume.
- Restoring from a snapshot is not permitted on a primary volume.
- Restoring from a snapshot is not permitted on a secondary volume.
- A snapshot cannot be created with the same name as the last_consistent or most_updated snapshot.

Chapter 9. Asynchronous remote mirroring

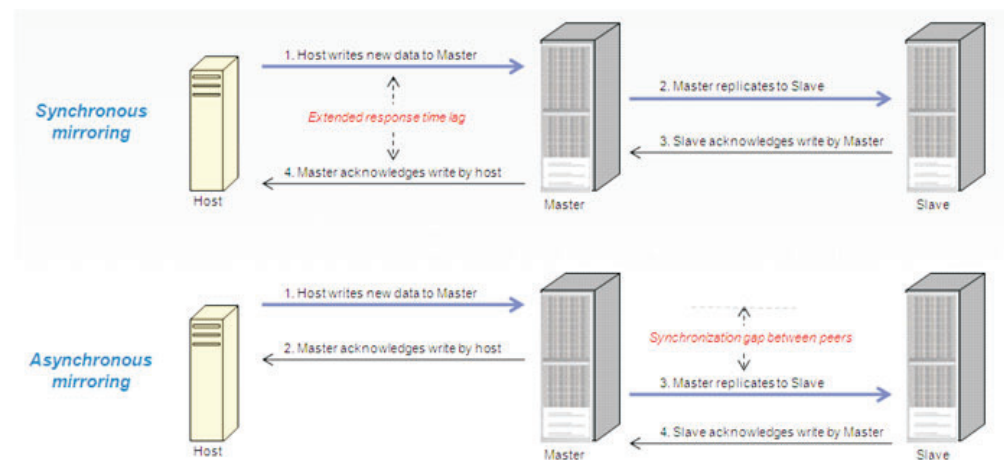
Asynchronous mirroring enables you to attain high availability of critical data through a process that asynchronously replicates data updates that are recorded on a primary storage peer to a remote, secondary peer.

The relative merits of asynchronous and synchronous mirroring are best illustrated by examining them in the context of two critical objectives:

- Responsiveness of the storage system
- Currency of mirrored data

With synchronous mirroring, host writes are acknowledged by the storage system only after being recorded on both peers in a mirroring relationship. This yields high currency of mirrored data (both mirroring peers have the same data), yet results in less than optimal system responsiveness because the local peer cannot acknowledge the host write until the remote peer acknowledges it. This type of process incurs latency that increases as the distance between peers increases.

XIV features both asynchronous mirroring and synchronous mirroring. Asynchronous mirroring is advantageous in various use cases. It represents a compelling mirroring solution in situations that warrant replication between distant sites because it eliminates the latency inherent to synchronous mirroring, and might lower implementation costs. Careful planning of asynchronous mirroring can minimize the currency gap between mirroring peers, and can help to realize better data availability and cost savings.



With synchronous mirroring (top), response time (latency) increases as the distance between peers increases, but both peers are synchronized. With asynchronous mirroring (bottom), response time is not sensitive to distance between peers, but the synchronization gap between peers is sensitive to the distance.

Figure 17. Synchronous mirroring compared to asynchronous mirroring

Note: Synchronous mirroring is covered in chapter 8.

Chapter scope

This chapter presents the following key aspects of the XIV asynchronous mirroring functionality:

- Feature highlights
- Terminology
- A technology overview of XIV asynchronous mirroring
- Initializing XIV asynchronous mirroring
- Walking through an XIV asynchronous mirroring process
- Administering XIV asynchronous mirroring
- Asynchronous mirroring of consistency groups
- Accommodating disaster recovery scenarios with XIV asynchronous mirroring
- Analyzing XIV asynchronous mirroring

Features highlights

The IBM XIV Storage System blends existing and new XIV technologies to produce an advanced mirroring solution with unique strengths.

The following are highlights of IBM XIV Storage System asynchronous mirroring:

Advanced Snapshot-based Technology

IBM XIV asynchronous mirroring is based on XIV snapshot technology, which streamlines implementation while minimizing impact on general system performance. The technology leverages functionality that has previously been effectively employed with synchronous mirroring and is designed to support mirroring of complete systems – translating to hundreds or thousands of mirrors.

Mirroring of Consistency Groups

IBM XIV supports definition of mirrored consistency groups, which is highly advantageous to enterprises, facilitating easy management of replication for all volumes that belong to a single consistency group. This enables streamlined restoration of consistent volume groups from a remote site upon unavailability of the primary site.

Automatic and Manual Replication

Asynchronous mirrors can be assigned a user-configurable schedule for automatic, interval-based replication of changes, or can be configured to replicate changes upon issuance of a manual (or scripted) user command. Automatic replication allows you to establish crash-consistent replicas, whereas manual replication allows you to establish application-consistent replicas, if required. The XIV implementation allows you to combine both approaches because you can define mirrors with a scheduled replication and you can issue manual replication jobs for these mirrors as needed.

Multiple RPOs and Multiple Schedules

IBM XIV asynchronous mirroring enables each mirror to be specified a different RPO, rather than forcing a single RPO for all mirrors. This can be used to prioritize replication of some mirrors over others, potentially making it easier to accommodate application RPO requirements, as well as bandwidth constraints.

Flexible and Independent Mirroring Intervals

IBM XIV asynchronous mirroring supports schedules with intervals ranging between 20 seconds and 12 hours. Moreover, intervals are

independent from the mirroring RPO. This enhances the ability to fine tune replication to accommodate bandwidth constraints and different RPOs.

Flexible pool management

IBM XIV asynchronous mirroring enables the mirroring of volumes and consistency groups that are stores in thin provisioned pools. This applies to both mirroring peers.

Bi-Directional Mirroring

IBM XIV systems can host multiple mirror sources and targets concurrently, supporting over a thousand mirrors per system. Furthermore, any given IBM XIV Storage System can have mirroring relationships with several other IBM XIV systems. This enables enormous flexibility when setting mirroring configurations.

The number of systems with which the Storage System can have mirroring relationships is specified out side this document (see the IBM XIV Storage System Data Sheet).

Concurrent Synchronous and Asynchronous Mirroring

The IBM XIV Storage System can concurrently run synchronous and asynchronous mirrors.

Easy Transition between Peer Roles

IBM XIV mirror peers can be easily changed between master and slave.

Easy Transition From Independent Volume Mirrors into Consistency Group Mirror

The IBM XIV Storage System allows for easy configuration of consistency group mirrors, easy addition of mirrored volumes into a mirrored consistency group, and easy removal of a volume from a mirrored consistency group while preserving mirroring for such volume.

Control over Synchronization Rates per Target

The asynchronous mirroring implementation enables administrators to configure different system mirroring rates with each target system.

Comprehensive Monitoring and Events

IBM XIV systems generate events and monitor critical asynchronous mirroring-related processes to produce important data that can be used to assess the mirroring performance.

Easy Automation via Scripts

All asynchronous mirroring commands can be automated through scripts.

Terminology

Mirror coupling (sometimes referred to as *coupling*)

A pairing of storage peers (either volumes or consistency groups) that are engaged in a mirroring relationship.

Master and Slave

The roles that correspond with the source and target storage peers for data replication in a mirror coupling. These roles can be changed by a system administrator after a mirror is created to accommodate customer needs and support failover and failback scenarios. A valid mirror can have only one master peer and only one slave peer.

Peer Designation

A user-configurable mirroring attribute that describes the designation associated with a coupling peer. The master is designated by default as primary and the slave is designated by default as secondary. These values

serve as a reference for the original peer's designation regardless of any role change issued after the mirror is created, but should not be mistaken for the peer's role (which is either master or slave).

Last replicated snapshot

A snapshot that represents the latest state of the master that is confirmed to be replicated to the slave.

Most recent snapshot

A snapshot that represents the latest synchronized state of the master that the coupling can revert to in case of disaster.

Sync Job

The mirroring process responsible for replicating any data updates recorded on the master since the Last Replicated Snapshot was taken. These updates are replicated to the slave.

Schedule

An administrative object that specifies how often a Sync Job is created for an associated mirror coupling.

Interval

A schedule parameter that indicates the duration between successive Sync Jobs.

RPO Recovery Point Objective – an objective for the maximal data synchronization gap acceptable between the master and the slave. An indicator for the tolerable data loss (expressed in time units) in the event of failure or unavailability of the master.

RTO Recovery Time Objective - an objective for the maximal time to restore service after failure or unavailability of the master.

Technological overview

The IBM XIV Storage System asynchronous mirroring blends existing and new technologies.

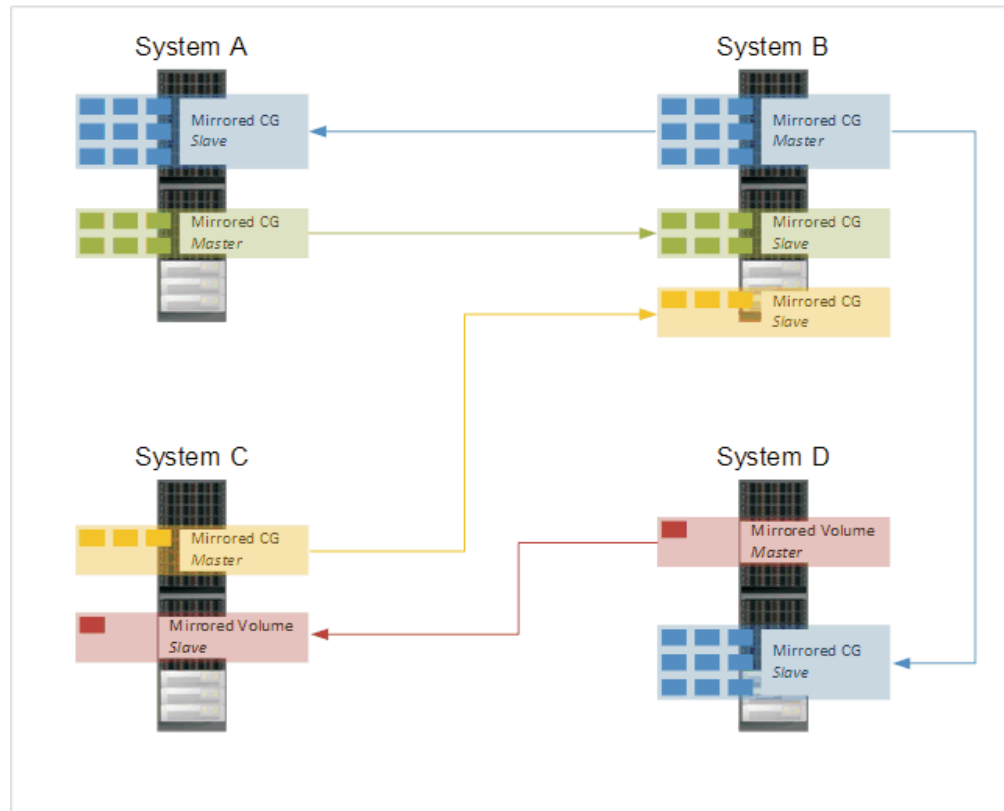
The asynchronous mirroring implementation is based on snapshots and features the ability to establish automatic and manual mirroring with the added flexibility to assign each mirror coupling with a different RPO. The ability to specify a different schedule for each mirror independently from the RPO helps accommodate special mirroring prioritization requirements without subjecting all mirrors to the same mirroring parameters. The paragraphs below detail the following IBM XIV asynchronous mirroring aspects, technologies, and concepts:

- The replication scheme
- The snapshot-based technology
- IBM XIV asynchronous mirroring special snapshots
- Initializing IBM XIV asynchronous mirroring
- The mirroring replication unit: the Sync Job
- Mirroring schedules and intervals
- The manual (ad-hoc) Sync Job
- Determining mirror state through the RPO
- Mirrored consistency groups
- IBM XIV asynchronous mirroring and pool space depletion

Replication scheme

IBM XIV asynchronous mirroring supports establishing mirroring relationships between an IBM XIV Storage System and other XIV systems.

Each of these relationships can be either synchronous or asynchronous and a system can concurrently act as a master in one relationship and act as the slave in another relationship. There are also no practical distance limitations for asynchronous mirroring – mirroring peers can be located in the same metropolitan area or in separate continents.



Each IBM XIV Storage System can have mirroring relationships with other IBM XIV Storage Systems. Multiple concurrent mirroring relationships are supported with each target.

Figure 18. The replication scheme

Snapshot-based technology

IBM XIV features an innovative snapshot-based technology for asynchronous mirroring that facilitates concurrent mirrors with different recovery objectives.

With IBM XIV asynchronous mirroring, write order on the master is not preserved on the slave. As a result, a snapshot taken of the slave at any moment is most likely inconsistent and therefore not valid. To ensure high availability of data in the event of a failure or unavailability of the master, it is imperative to maintain a consistent replica of the master that can ensure service continuity.

This is achieved through XIV snapshots. XIV asynchronous mirroring employs snapshots to record the state of the master, and calculates the difference between successive snapshots to determine the data that needs be copied from the master to

the slave as part of a corresponding replication process. Upon completion of the replication process, a snapshot is taken of the slave and reflects a valid replica of the master.

Below are select technological properties that explain how the snapshot-based technology helps realize effective asynchronous mirroring:

- XIV's support for a practically unlimited number of snapshots facilitates mirroring of complete systems with practically no limitation on the number of mirrored volumes supported
- XIV implements memory optimization techniques that further maximize the performance attainable by minimizing disk access.

Mirroring special snapshots

The *status* and scope of the synchronization process is determined through the use of snapshots.

The following two special snapshots are used:

most_recent snapshot

This snapshot is the most recent snapshot taken of the master (being either a volume or consistency group), prior to the creation of a new replication process (Sync Job – see below). This snapshot exists only on the master.

last_replicated snapshot

This is the most recent snapshot that is confirmed to have been fully replicated to the slave. Both the master and the slave have this snapshot. On the slave, the snapshot is taken upon completion of a replication process, and replaces any previous snapshot with that name. On the master, the *most_recent* snapshot is renamed *last_replicated* after the slave is confirmed to have a corresponding replica of the master's *most_recent* snapshot.



XIV maintains three snapshots per mirror coupling: two on the master and one on the slave. A valid (recoverable) state of the master is captured in the *last_replicated* snapshot, and on an identical snapshot on the slave. The *most_recent* snapshot represents a recent state of the master that needs to be replicated next to the slave. The system determines the data to replicate by comparing the master's snapshots.

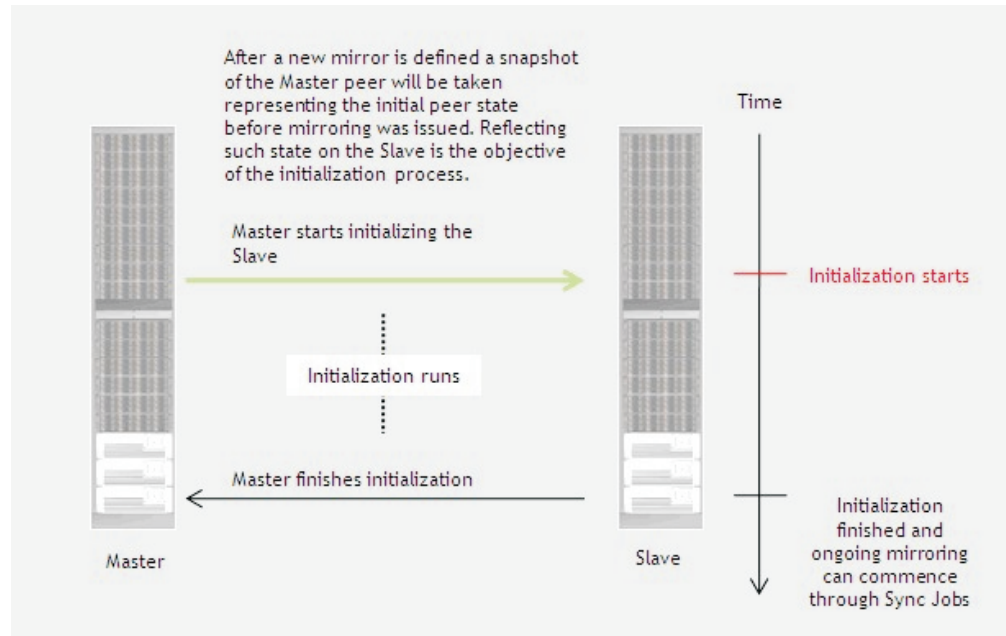
Figure 19. Location of special snapshots

Initializing the mirroring

An XIV mirror is easily created using the CLI or GUI. First, the mirror is created and activated, then an initialization phase starts.

XIV mirrors are created in standby state and must be explicitly activated. During the Initialization phase, the system generates a valid replica of the state of the master on the slave. Until the Initialization is over, there is no valid replica on the

slave to help recover from a disaster on the master. Once the Initialization phase ends, a snapshot of the slave is taken. This snapshot represents a valid replica of the master and can be used to restore a consistent state of the master in disaster recovery scenarios.



Mirroring starts with an initialization phase, during which the data on the master (at the beginning of the phase) is being replicated to the slave. Before the actual replication starts, a snapshot of the master is taken. This snapshot determines the scope of the replication for the initialization phase.

Figure 20. Asynchronous mirroring over-the-wire initialization

Off-line replicating the master onto the slave

The IBM XIV Storage System allows for this volume replica to be transferred off-line to the slave.

At the beginning of the Initialization of the mirror, the user states which volume will be replicated from the master to the slave. This replica of the master is typically much larger than the schedule-based replicas that accumulate differences that are made during a small amount of time. The IBM XIV Storage System allows for this volume replica to be transferred off-line to the slave. This method of transfer is sometimes called "Truck Mode" and is accessible through the `mirror_create` command.

Off-line initialization of the mirror replicates the master onto the slave without being required to utilize the inter-site link. The off-line replication requires:

- Specifying the volume to be mirrored.
- Specifying the initialization type to the mirror creation command.
- Activating the mirroring.

Meeting the above requirements, the system takes a snapshot of the master, and compares it with the slave volume. Only areas where differences are found are replicated as part of the initialization. Then, the slave peer's content is checked against the master and not just automatically considered a valid replica. This check

optimizes the initialization time through taking into consideration the available bandwidth between the master and slave and whether the replica is identical to the master volume (that is, there where no writes to the master during the initialization).

The sync job

Data synchronization between the master and slave is achieved through a process run by the master called a Sync Job.

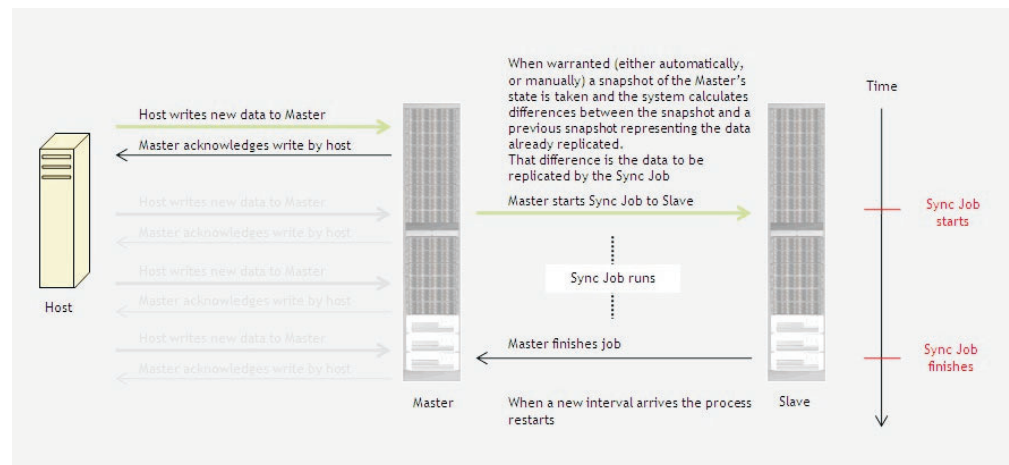
The Sync Job updates the slave with any data that was recorded on the master since the latest Sync Job was created. The process can either be started automatically based on a user-configurable schedule, or manually based on a user-issued command.

Calculating the data to be replicated as part of the Sync Job

When the Sync Job is started, a snapshot of the master's state at that time is taken (the `most_recent` snapshot).

After any outstanding Sync Job are completed, the system calculates the data differences between the above snapshot and the most recent master snapshot that corresponds with a consistent replica on the slave (the `last_replicated` snapshot). That difference constitutes the data to be replicated next by the Sync Job.

Replication is very efficient because it only copies data differences between mirroring peers. For example, if only a single block was changed on the master, only a single block will be replicated to the slave.



Data synchronization between the master and slave is achieved through a process run by the master called a Sync Job. The process replicates data changes between the master's snapshots to the slave.

Figure 21. : The Asynchronous mirroring Sync Job

Mirroring schedules and intervals

The IBM XIV Storage System implements a scheduling mechanism that is used to drive a recurring asynchronous mirroring process.

Each asynchronous mirror has a specified schedule, and the schedule's interval indicates how often a Sync Job is created for that mirror.

Asynchronous mirroring has the following features:

- The schedule concept. A *schedule* specifies an interval for automatic creation of Sync Jobs; a new Sync Job is normally created at the arrival of a new interval.
- A Sync Job is not created if another scheduled Sync Job is running when a new interval arrives
- Custom schedules can be created by users
- Schedule intervals can be set to any of the following values: 30 seconds, 1 min, 2 min, 5 min, 10 min, 15 min, 30 min, 1 hour, 2 hours, 3 hours, 6 hours, 8 hours, 12 hours. The schedule start hour is 00:00.

Note: The IBM XIV Storage System offers a built-in, non-configurable schedule called *min_interval* with a 20s interval. It is only possible to specify a 20s schedule using this predefined schedule.

- When creating a mirror, two schedules are specified - one per peer. The slave's schedule can help streamline failover scenarios - controlled by either XIV or a 3rd party process.
- A single schedule can be referenced by multiple couplings on the same system.
- Sync Job creation for mirrors with the same schedule takes place at exactly the same time. This is in contrast with mirrors having different schedules with the same interval. Despite having the same interval, Sync Jobs for these types of mirrors are not guaranteed to take place at the same time.
- A unique schedule called *Never* is provided to indicate that no Sync Jobs are automatically created for the pertinent mirror (see below).

Schedules are local to a single XIV system

Schedules are local to the XIV system where they are defined and are set independently of system-to-system relationships. A given source-to-target replication schedule does not mandate an identical schedule defined on the target for reversed replication. To maintain an identical schedule for reverse replication (if the master and slave roles need be changed), independent identical schedules must be defined on both peers.

Schedule sensitivity to timezone difference

The schedules of the peers of a mirroring couple have to be defined in a way that won't be impacted from timezone differences. For example, if the timezone difference between the master and slave sites is two hours, the interval is 3 hours and the schedule on one peer is (12PM, 3AM, 6AM,...), the schedule on the other peer needs to be (2AM, 5AM, 8AM). Although some cases do not call for shifted schedules (for example, a timezone difference of 2 hours and an interval of one hour), this issue can't be overlooked.

The master and the slave also have to have their clocks synchronized (for example using NTP). Avoiding such a synchronization could hamper schedule-related measurements, mainly RPO.

The Never schedule

The system features a special, non-configurable schedule called *Never* that denotes a schedule with no interval. This schedule indicates that no Sync Jobs are automatically created for the mirror so it is only possible to issue replication for the mirror through a designated manual command.

Note: A manual snapshot mirror can be issued to every mirror that is assigned a user-defined schedule.

The mirror snapshot (ad-hoc sync job)

You can manually issue a dedicated command to run a mirror snapshot, in addition to using the schedule-based option.

This type of mirror snapshot can be issued for a coupling regardless of whether it has a schedule. The command creates a new snapshot on the master and manually initiates a Sync Job that is queued behind outstanding Sync Jobs.

The mirror snapshot:

- Accommodates a need for adding manual replication points to a scheduled replication process.
- Creates application-consistent replicas (in cases where consistency is not achieved via the scheduled replication).

The following characteristics apply to the manual initiation of the asynchronous mirroring process:

- Multiple mirror snapshot commands can be issued – there is no maximum limit on the number of mirror snapshots that can be issued manually.
- A mirror snapshot running when a new interval arrives delays the start of the next interval-based mirror scheduled to run, but does not cancel the creation of this Sync Job.
 - The interval-based mirror snapshot will be cancelled only if the running snapshot mirror (ad-hoc) has never finished.

Other than these differences, the manually initiated Sync Job is identical to a regular interval-based Sync Job.

Determining replication and mirror states

The mirror state indicates whether the master is mirrored according to objectives that are specified by the user.

As asynchronous mirroring endures a gap that may exist between the master and slave states, the user must specify a restriction objective for the mirror – the RPO – or Recovery Point Objective. The system determines the mirror state by examining if the master's replica on the slave. The mirror state is considered to be OK only if the master replica on the slave is newer than the objective that is specified by the RPO.

RPO and RTO

The evaluation of the synchronization status is done based on the mirror's RPO value. Note the difference between RPO and RTO.

RPO Stands for Recovery Point Objective and represents a measure of the maximum data loss that is acceptable in the event of a failure or unavailability of the master.

RPO units

Each mirror must be set an RPO by the administrator, expressed in time units. Valid RPO values range between 30 seconds and 24 hours. An RPO of 60 seconds indicates that the slave's state should not be older than the master's state by more than 60 seconds. The

system can be instructed to alert the user if the RPO is missed, and the system's internal prioritization process for mirroring is also adjusted.

RTO Stands for Recovery Target Objective and represents the amount of time it takes the system to recover from a failure or unavailability of the master.

The mirror's RTO is not administered in XIV.

Mirror status values

The mirror status is determined based on the mirror state and the mirroring status.

During the progress of a Sync Job and until it completes, the slave replica is inconsistent because write order on the master is not preserved during replication. Instead of reporting this state as inconsistent, the mirror state is reported based on the timestamp of the slave's last_replicated snapshot as one of the following:

RPO_OK

Synchronization exists and meets its RPO objective.

RPO_Lagging

Synchronization exists but lags behind its RPO objective.

Initializing

Mirror is initializing.

Definitions of mirror state and status:

The mirror status is determined based on the mirror state and the mirroring status.

Mirror state

During the progress of a Sync Job and until it completes, the slave replica is inconsistent because write order on the master is not preserved during replication. Instead of reporting this state as inconsistent, the mirror state is reported based on the timestamp of the slave's last_replicated snapshot as one of the following:

Definition of RPO_OK

Synchronization exists and meets its RPO objective.

Definition of RPO_Lagging

Synchronization exists but lags behind its RPO objective.

Initializing

Mirror is initializing.

Mirroring status

The mirroring status denotes the status of the replication process and reflects the activation state and the link state.

Effective recovery currency

Measures as the delta between the current time and the timestamp of the last_replicated snapshot

Declaring on RPO_OK

The Effective recovery currency is positive.

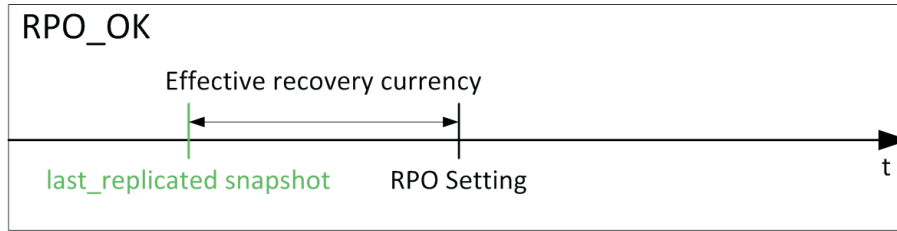


Figure 22. The way RPO_OK is determined

Declaring on RPO_Lagging

The Effective recovery currency is negative.

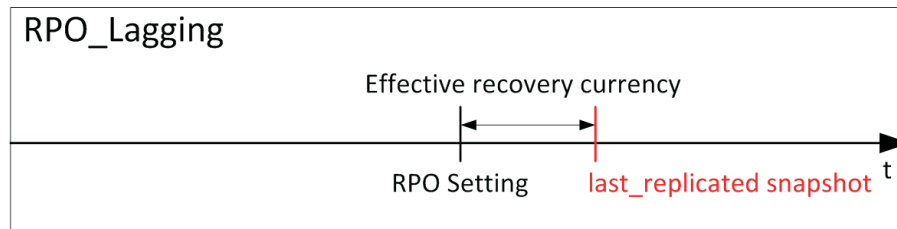


Figure 23. The way RPO_Lagging is determined

Examples of the way mirror status is determined:

The mirroring status denotes the status of the replication process and reflects the activation state and the link state.

The following example portrays the way the mirroring status is determined. The time-axis denotes time and the schedule of the sync jobs (t_0 - t_5). Both RPO states are displayed in red and green at the upper section of the image.

First sync job - RPO is OK

Time: t_0

A sync job starts. RPO_OK is maintained as long as the sync job ends before t_1 .

Time: t_a

As the sync job ends at t_a , prior to t_1 , the status is RPO_OK.

Effective recovery currency

During the sync job run the value of Effective recovery currency (the black graph on the upper section of the image) changes. This value goes up as we are getting farther from t_0 , goes down - to the RPO setting - once the sync job complete, and does not resume climbing as long as the next schedule has arrived.

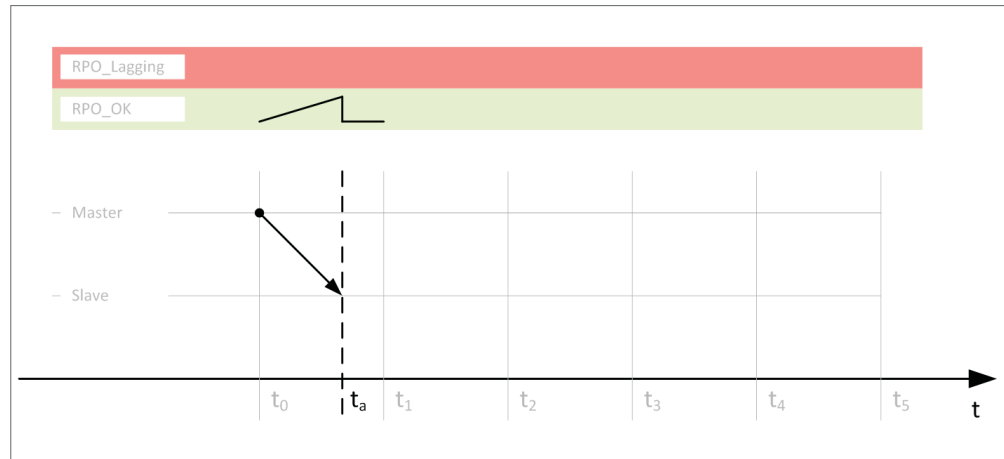


Figure 24. Determining the asynchronous mirroring status – example part 1

Second sync job - RPO is lagging

Time: t_1

A sync job starts. RPO_OK is maintained as long as the sync job ends before t_2 .

Time: t_2

The sync job should have ended at this point, but it is still running.

The sync job the was scheduled to run on this point in time is cancelled.

Time: t_b

As the sync job ends at t_b , which is after t_2 , the status is RPO_Lagging.

Effective recovery currency

The value of Effective recovery currency keeps climbing as long as the next sync job hasn't finished.

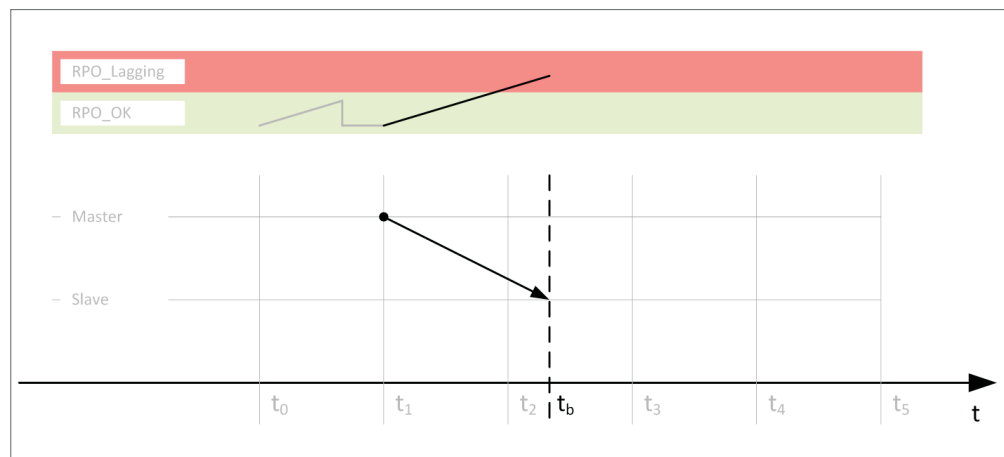


Figure 25. Determining the asynchronous mirroring status – example part 2

Third sync job - RPO is OK

Time: t_3

A new sync job starts. At this point the status is RPO_Lagging.

Time: t_c

As the sync job ends prior to t_4 , the status is RPO_OK.

Effective recovery currency

The value of Effective recovery currency keeps climbing until the next sync job has finished (this happens at t_c). This value immediately returns to the RPO setting until the time of the next schedule.

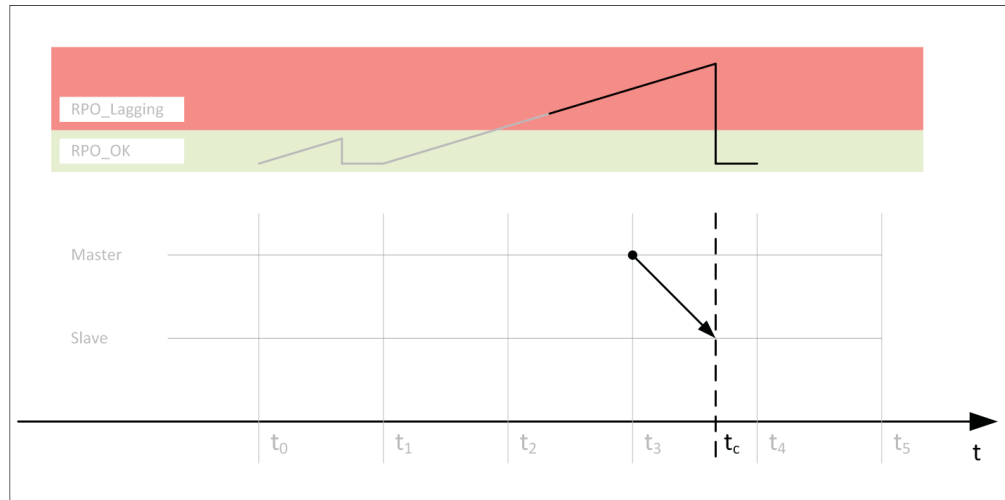


Figure 26. Determining Asynchronous mirroring status – example part 3

The added-value of multiple RPOs

The system bases its internal replication prioritization on the mirror RPO; hence – support for multiple RPOs corresponding with true recovery objectives helps optimize the available bandwidth for replication.

The added-value of multiple Schedules

You can attain a target RPO using multiple schedule interval options. A variable schedule that is decoupled from the RPO helps optimize the replication process to accommodate RPO requirements without necessarily modifying the RPO.

Mirrored consistency groups

IBM XIV enables mirrored consistency groups and mirroring of volumes to facilitate the management of mirror groups.

Asynchronous mirroring of consistency groups is accomplished by taking a snapshot group of the master consistency group in the same manner employed for volumes, either based on schedules, or manually through a dedicated command option.

The peer synchronization and status are managed on a consistency group-level, rather than on a volume-level. This means that administrative operations are carried out on the whole consistency group, rather than on a specific volume within the consistency group. This includes operations such as activation, and mirror-wide settings such as a schedule.

The synchronization status of the consistency group reflects the combined status of all mirrored volumes pertaining to the consistency group. This status is determined

by examining the (system-internally-kept) status of each volume in the consistency group. Whenever a replication is complete for all volumes and their state is RPO_OK, the consistency group mirror status is also RPO_OK. On the other hand, if the replication is incomplete or any of the volumes in a mirrored consistency group has the status of RPO_Lagging, the consistency group mirror state is also RPO_Lagging.

Storage space required for the mirroring

IBM XIV enables to manage the storage required for the mirroring on thin-provisioned pools on both the master and the slave.

Throughout the course of the mirroring, the `last_replicated` and `most_recent` snapshots may exceed the space allocated to the volume and its snapshots. The lack of sufficient space on the master can prevent host writes, where the lack of space on the slave can disrupt the mirroring process itself.

IBM XIV enables to manage the storage required for the mirroring on thin-provisioned pools. This way, The IBM XIV Storage System manages and allocates space according to the schemes described in the Chapter 6, “Thin provisioning,” on page 39 chapter.

Upon depletion of space on each of the peers, the “Pool space depletion” mechanism takes effect.

Pool space depletion

Pool space depletion is a mechanism that takes place whenever the mirroring can no longer be maintained due to lack of space for incoming write requests issued by the host.

Whenever a pool does not have enough free space to accommodate the storage requirements warranted by a new host write, the system runs a multi-step procedure that progressively deletes snapshots within that pool until enough space is made available for a successful completion of the write request.

This multi-step procedure is progressive, meaning that the system proceeds to the next step only if following the execution of the current step, there is still insufficient space to support the write request.

The following processes and concepts are introduced upon the depletion of pool space:

- The Pool space deletion mechanism utilizes the concept of “Protecting snapshots through setting their deletion priority”
- “Pool space depletion on the master” on page 82
- “Pool space depletion on the slave” on page 84

These processes apply to thin-provisioned pools as well.

Protecting snapshots through setting their deletion priority:

Protected snapshots have precedence over other snapshots during the Pool space deletion process.

The concept of protected snapshots assigns the Storage Pool with an attribute that is compared with the snapshots' auto-deletion priority attribute. Whenever a snapshot has a deletion priority that is higher than the pool's attribute, it is considered protected.

For example, if the deletion priority of the depleting storage is set to 3, the system will delete snapshots with the deletion priority of 4. Snapshots with priority level 1, 2 and 3 will not be deleted.

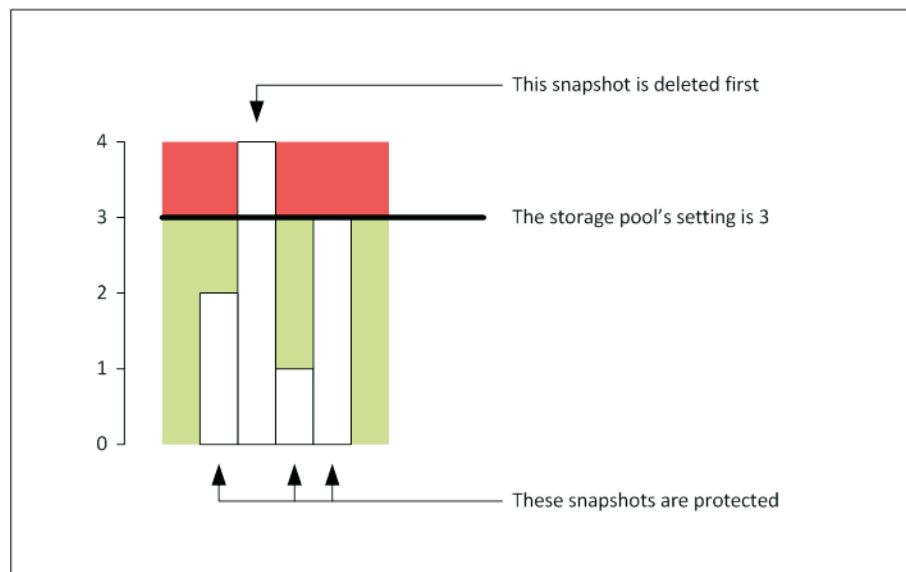


Figure 27. The deletion priority of the depleting storage is set to 3

If the deletion priority of the depleting storage is set to 4, the system will deactivate mirroring before deleting any snapshots.

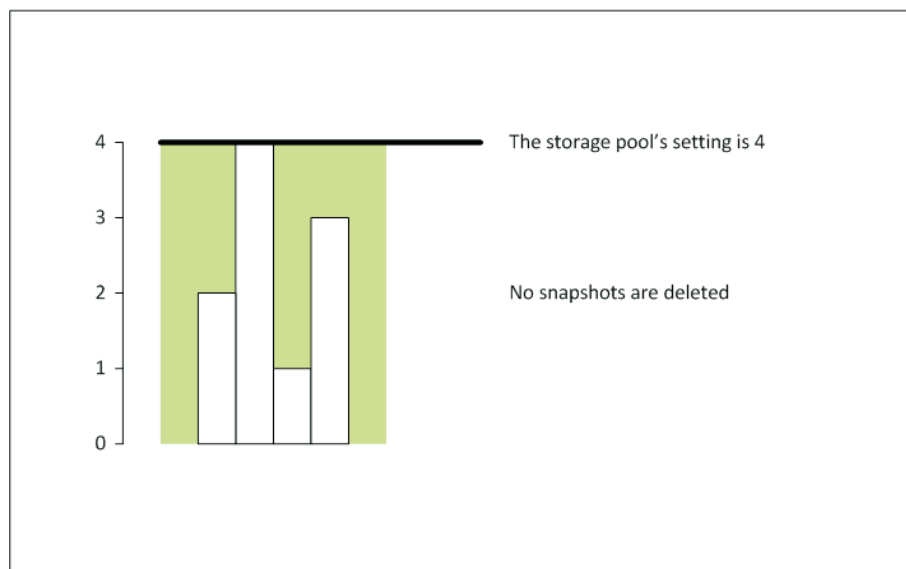


Figure 28. The deletion priority of the depleting storage is set to 4

If the deletion priority of the depleting storage is set to 0, the system can delete any snapshot regardless of deletion priority.

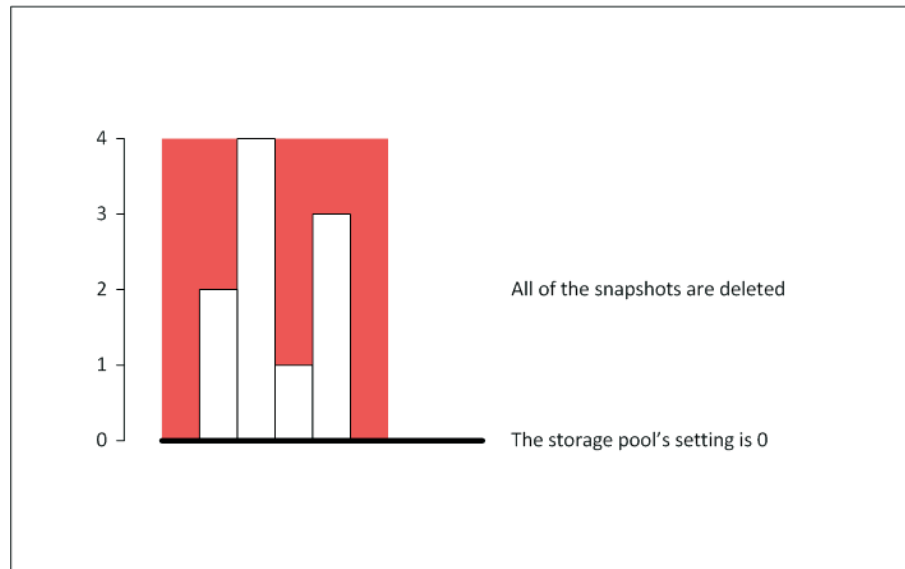


Figure 29. The deletion priority of the depleting storage is set to 0

See here for specifics regarding the deletion priority:

- “The snapshot auto-delete priority” on page 15
- The snapshot group auto-delete priority
- The Storage Pool's deletion settings

Deletion priority conventions

Protecting the deletion priority of the last replicated snapshot

The deletion priority of mirroring-related snapshots is set implicitly by the system and cannot be customized by the user (see below).

Last replicated and most recent snapshots

The deletion priority of the asynchronous mirroring `last_replicated` and `most_recent` snapshots on the master is set to 1.

Last replicated snapshot on the slave

The deletion priority of the `last_replicated` snapshot on the slave and the is set to 0 (see below).

Default value of the snapshot protecting CLI

By default, the value of the `protected_snapshot_priority` parameter of the `pool_config_snapshots` command is 0.

Changing this value

If the `protected_snapshot_priority` parameter is changed, the system and user-created snapshots with a deletion priority nominally equal or lower than the protected setting will be deleted only after the internal mirroring snapshots are .

For example, if the `protected_snapshot_priority` is changed to 1, all system and user-created snapshots with deletion priority 1 (which includes ALL snapshots created by the user, assuming that their deletion priority was not changed) will be protected and will be deleted only after internal mirroring snapshots are.

Other snapshots

Non mirroring-related snapshots are created by default with a deletion priority 1.

Protecting the last replicated snapshots

The last replicated snapshots represent a consistent replica of the Master in asynchronous mirroring. Both the Master and the Slave have a last replicated snapshot, however, these two snapshots are protected differently.

LRS_{slave}

The slave must have an available consistent copy of the master at all times, where the master does not have to have such availability (as the LRS_{master} itself is regarded consistent). As a result, this snapshot is never deleted. Upon pool space depletion on the Slave, whenever there is no space for the mirroring process, the pool will be locked.

The deletion priority of the LRS on the slave is 0.

LRS_{master}

The last replicated snapshot on the master is available for deletion during pool space depletion.

The deletion priority of the LRS on the master is 1.

Pool space depletion on the master:

The depletion procedure on the master takes the following steps.

Step 1 - deletion of unprotected snapshots

The following snapshots are deleted:

- Regular (not related to mirroring) snapshots
- Snapshots of the mirroring processes that are no longer active
- The snapshot of any Snapshot Mirror (a.k.a. ad hoc sync job) that has not started yet

The deletion is subject to the deletion priority of the individual snapshot. In the case of deletion priority clash, older snapshots are deleted first.

Success criteria:

The user reattempts operation, re-enables mirroring and resumes replication. If this fails, the system proceeds to step 2 (below).

Step 2 - deletion of the snapshot of any outstanding (pending) scheduled sync job

If replication still does not resume after the actions taken on step 1:

The following snapshots are deleted:

- All snapshots that were not deleted in step 1.

Success criteria:

The system reattempts operation, re-enables mirroring and resumes replication.

Step 3 - automatic deactivation of the mirroring and deletion of the snapshot designated as the mirror most_recent snapshot

If the replication still does not resume:

The following takes place:

- An automatic deactivation of the mirroring
- Deletion of the most_recent snapshot
- An event is generated.

Ongoing ad-hoc sync job

The snapshot created during the ad-hoc sync job is considered as a `most_recent` snapshot, although it is not named as such and not suplicated with a snapshot in that name. Following the completion of the ad-hoc sync job, and only after this completion, the snapshot is duplicated and the duplicate is named `last_replicated`.

Upon a manual reactivation of the mirroring process:

1. The mirroring activation state changes to Active
2. A `most_recent` snapshot is created
3. A new sync job starts

Step 4 - deletion of the `last_replicated` snapshot

If more space is still required:

The following takes place:

- Deletion of the `last_replicated` snapshot (on the Master)
- An event is generated.

Following the deletion:

1. The mirroring remains deactivated, and must be manually reactivated.
2. The mirroring changes to *change tracking* state. Host I/O to the Master are tracked but not replicated
3. The system marks storage areas that were written into since the `last_replicated` snapshot was created

Step 5 - deletion of the `most_recent` snapshot that is created when activating the mirroring in *Change Tracking* state

If more space is still required:

The following takes place:

- Deletion of the `most_recent` snapshot (on the Master).
- An event is generated.

Following the deletion:

Deletion of this `most_recent` snapshot in this state leaves the master with neither a snapshot nor a bitmap, mandating full initialization. To minimize the likelihood for such deletion, this snapshot is automatically assigned a special (new) deletion priority level. This deletion priority implies that the system should delete the snapshot only after all other `last_replicated` snapshots in the pool were deleted. Note that the new priority level will only be assigned to a mirror with a consistent Slave replica and not to a mirror that was just created (whose first state is also initialization).

Step 6 - deletion of protected snapshots

If more space is still required:

The following takes place:

- An event is generated.
- Deletion of all protected snapshots, regardless of the mirroring. These snapshots are deleted according to their deletion priority and age.

Following the deletion:

- The Master's state changes to **Init** (distinguished from the Initialization phase mirrors start with)
- The system stops marking new writes
- A `most_recent` snapshot is created
- The system creates and runs a sync job encompassing all of the tracked changes tracked
- following the completion of this sync job, a `last_replicated` snapshot is created on the Master, and the mirror state changes to `rpo_ok` or `rpo_lagging`, as warranted by the Effective RPO

If pool space depletes during the Init:

- The Master's state remains Initialization
- An event is generated
- The mirroring is deactivated
- The `most_recent` snapshot is deleted (mandating a Full Initialization)

Upon manual mirroring activation during the Init:

- The Master's state remains Initialization
- A `most_recent` snapshot is created
- The system starts a Full Initialization based on the `most_recent` snapshot

Pool space depletion on the slave:

Pool space depletion on the Slave means that there is no room available for the `last_replicated` snapshot. In this case, the mirroring is deactivated.

Snapshots with a Deletion Priority of 0 are special snapshots that are created by the system on the Slave peer and are not automatically deleted to free space, regardless of the pool space depletion process. The asynchronous mirroring slave peer has one such snapshot: the `last_replicated` snapshot.

Walking through the asynchronous mirroring process

This section walks you through creating an asynchronous mirroring relationship, starting from the initialization all the way through completing the first scheduled Sync Job.

Step 1

Time is 01:00 when the command to create a new mirror is issued. In this example, an RPO of 120 minutes and a schedule of 60 minutes are specified for the mirror.

The mirroring process must first establish a baseline for ensuing replication. This warrants an Initialization process during which the current state of the master is replicated to the slave peer. This begins with the host writes being briefly blocked (1). The state of the master peer can then be captured by taking a snapshot of the master state: the `most_recent` snapshot (2), which serves as a baseline for ensuing schedule-based mirroring. After this snapshot is created, host writes are no longer blocked and continue to update the storage system (3). At this time, no snapshot exists on the slave yet.

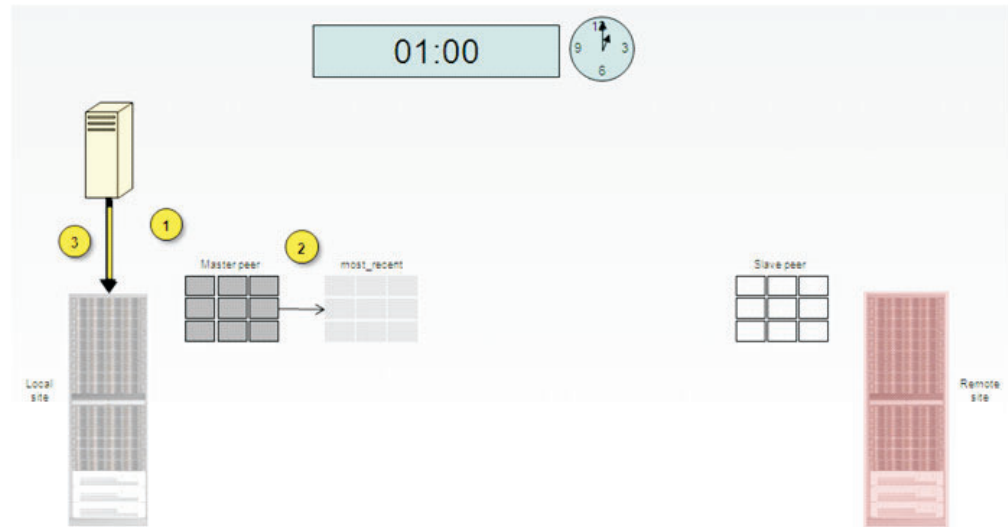


Figure 30. Asynchronous mirroring walkthrough – part 1

Step 2

After the state of the master is captured, the data that needs to be replicated as part of the Initialization process is calculated. In this example, the master's `most_recent` snapshot represents the data to be replicated through the first Sync Job (4).

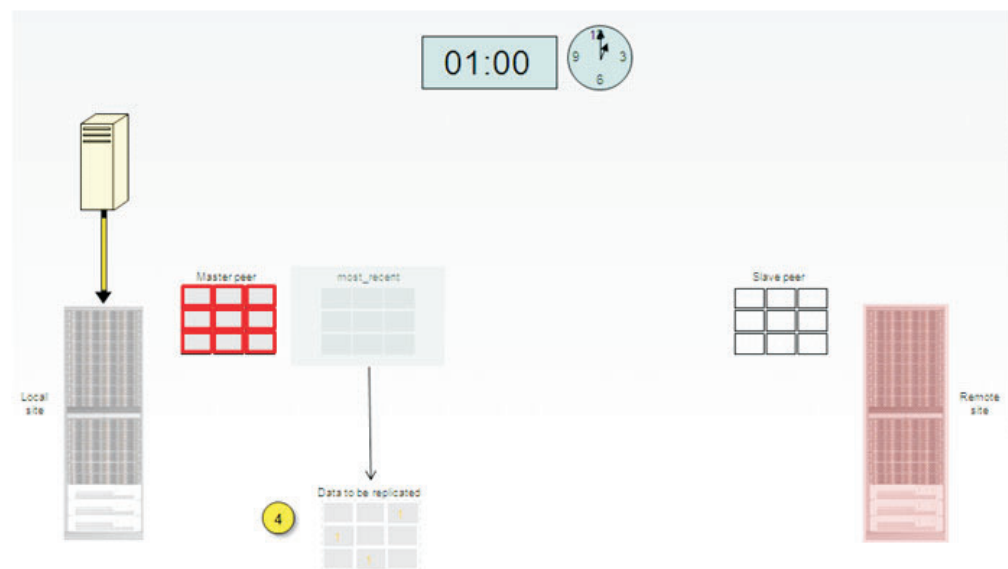


Figure 31. Asynchronous mirroring walkthrough – part 2

Step 3

During this step, the Initialization Sync Job is well in progress. The master continues to be updated with host writes – the updates are noted in the order they are written – first 1, then 2 and finally 3. The initialization Sync Job replicates the initial master peer's state to the slave peer (5).

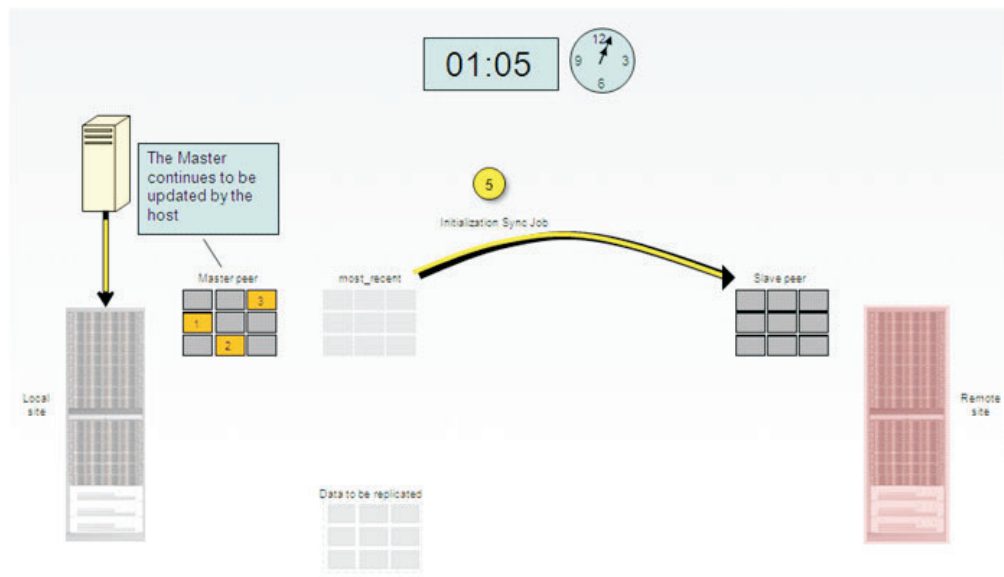


Figure 32. Asynchronous mirroring walkthrough – part 3

Step 4

Moments later, the Initialization Sync Job completes. After it completes, the slave's state is captured by taking a snapshot: the last_replicated snapshot (6). This snapshot reflects the state of the master as captured in the most_recent snapshot. In this example, it is the state just before the initialization phase started.

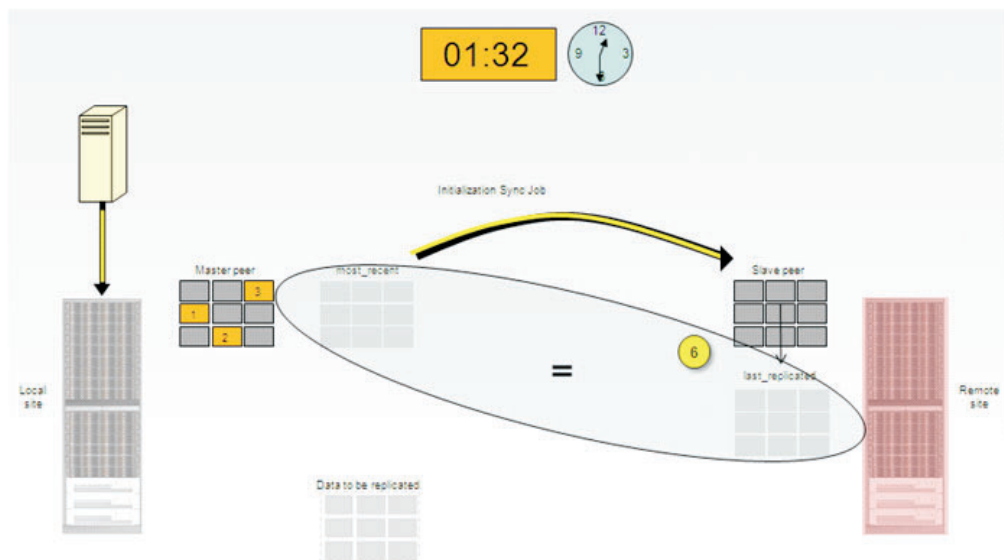


Figure 33. Asynchronous mirroring walkthrough – part 4

Step 5

During this step, the master's last_replicated snapshot is created. The most_recent snapshot on the master is renamed last_replicated (7) and represents the most recent point-in-time that the master can be restored if needed (because this state is captured in the slave's corresponding snapshot).

Step 7

A new Sync Job is set. The data to be replicated is calculated based on the difference between the master's most_recent snapshot and the last_replicated snapshot (4).



Figure 36. Asynchronous mirroring walkthrough – part 7

Step 8

The Sync Job is in process. During the Sync Job, the master continues to be updated with host writes (update 5).

The sync job data is not replicated to the slave in the order by which it was recorded at the master – the order of updates on the slave is different.

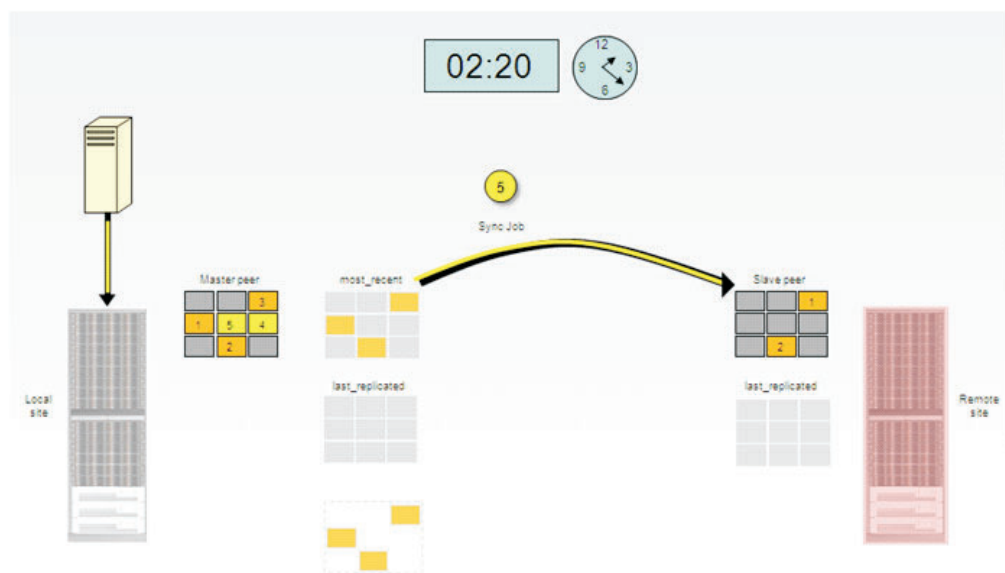


Figure 37. Asynchronous mirroring walkthrough – part 8

Step 9

The Sync Job is completed. The Slave's last_replicated snapshot is deleted (6) and replaced (in one atomic operation) by a new last_replicated snapshot.

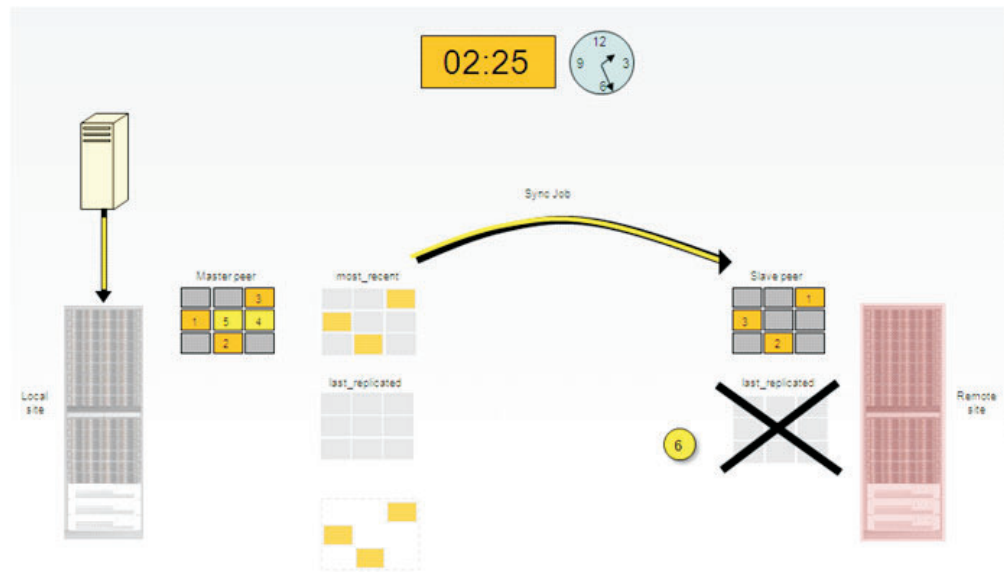


Figure 38. Asynchronous mirroring walkthrough – part 9

Step 10

The Sync Job is completed with a new last_replicated snapshot representing the updated slave's state (7).

The slave's last_replicated snapshot reflects the master's state as captured in the most_recent snapshot. In this example, it is the state at the beginning of the mirror schedule's interval.

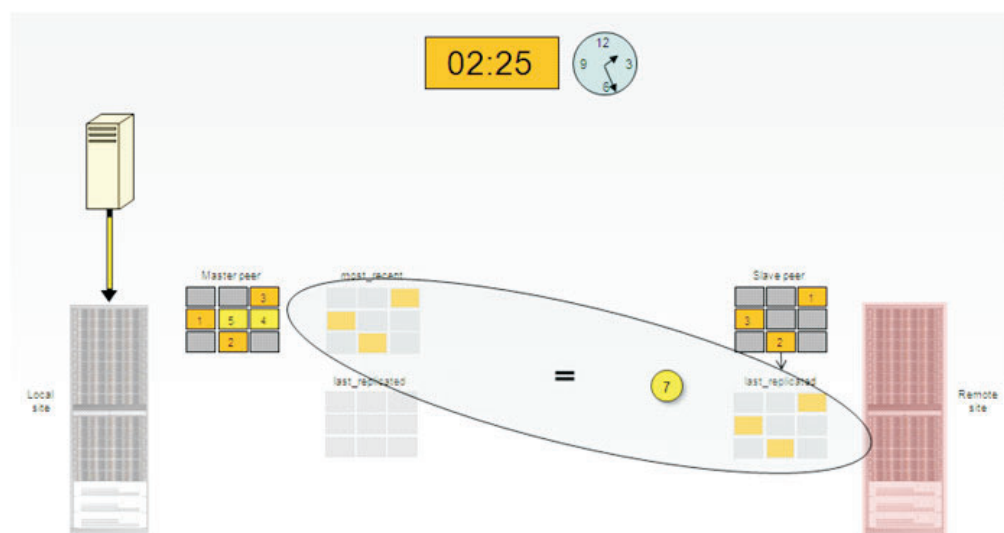


Figure 39. Asynchronous mirroring walkthrough – part 10

Step 11

A new master last_replicated snapshot created. In one transaction - the current last_replicated snapshot on the master is deleted (8) and the most_recent snapshot is renamed the last_replicated (9).

The interval sync process is now complete - the master and slave both have an identical restore time point to which they can be reverted if needed.

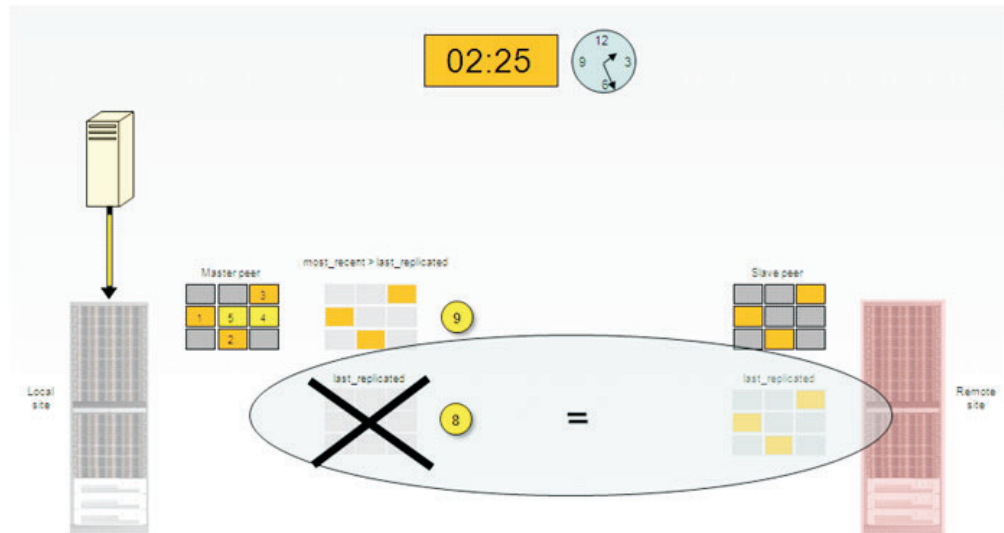


Figure 40. Asynchronous mirroring walkthrough – part 11

Peers roles

Peers' statuses denote their roles within the coupling definition.

After creation, a coupling has exactly one peer that is set to be the master peer, and exactly one peer that is set to be the slave peer. Each of the peers can have the following available statuses:

None The peer is not part of a coupling definition.

Master

The actual source peer in a replication coupling. This type of peer serves host requests, and is the source for synchronization updates to the slave. A master peer can be changed to a slave directly while in asynchronous mirroring.

Slave The actual target peer in a replication coupling. This type of peer does not serve host requests, and accepts synchronization updates from a corresponding master. A slave can be changed to a master directly while in asynchronous mirroring.

Activating the mirroring

The state of the mirroring is derived from the state of its components.

The remote mirroring process hierarchically manages the states of the entities that participate in the process. It manages the states for the mirroring based on the states of the following components:

- Link
- Activation

The following mirroring states are possible:

Non-operational

The coupling state is defined as non-operational if at least one of the following conditions is met:

- The activation state is standby.
- The link state is error.
- The slave peer is locked.

Operational

All of the following conditions must be met for the coupling to be defined as operational:

- The activation state is active.
- The link is OK.
- The peers have different roles.
- The slave peer is not locked.

Link states

The link state is one of the factors determining the coupling operational status.

The *link state* reflects the connection from the master to the slave. A failed link or a failed slave system both manifest as a link error. The link state is one of the factors determining the coupling operational status.

The available link states are:

OK The link is up and functioning.

Error The link is down.

Activation states

When the coupling is created, its activation is in standby state. When the coupling is enabled, its activation is in active state.

Standby

When the coupling is created, its activation is in standby state.

The synchronization is disabled:

- Sync jobs do not run.
- No data is copied.
- The coupling can be deleted.

Active The synchronization is enabled:

- Sync jobs can be run.
- Data can be copied between peers.

Regardless of the activation state:

- The mirroring type can be changed to synchronous.
- Peer roles can change.

Deactivating the coupling

Deactivating the coupling stops the mirroring process.

The mirroring is terminated by deactivating the coupling, causing the system to:

- Terminate, or delete the mirroring
- Stop the mirroring process as a result of:

- A planned network outage
- An application to reduce network bandwidth
- A planned recovery test

The deactivation pauses a running sync job and no new sync jobs will be created as long as the active state of the mirroring is not restored. However, the deactivation does not cancel the interval-based status check by the master and the slave. The synchronization status of the deactivated coupling is calculated on the start of each interval, as if the coupling was active.

Deactivating a coupling while a sync job is running, and not changing that state before the next interval begins, leads to the synchronization status becoming RPO_Lagging, as described in the following outline. Upon the deactivation:

On the master

The activation state changes to standby; replication pauses (and records where it paused); replication resumes upon activation.

Note: An ongoing sync job resumes upon activation, no new sync job will be created until the next interval.

On the slave

Not available.

Regardless of the state of the coupling:

- Peer roles can be changed

Note: For consistency group mirroring: deactivation pauses all running sync jobs pertaining to the consistency group. It is impossible to deactivate a single volume sync job within a consistency group.

Mirroring consistency groups

Grouping volumes into a consistency group provides a means to maintain a consistent snapshot of the group of volumes at the secondary site.

The following assumptions make sure that consistency group semantics work with remote mirroring:

Consistency Group-level management

Mirroring of consistency groups is managed on a consistency group-level, rather than on a volume-level. For example, the synchronization status of the consistency group is determined after examining all mirrored volumes that pertain to the consistency group.

Starting with an empty consistency group

Only an empty consistency group can be defined as a mirrored consistency group. If you want to define an existing non-empty consistency group as mirrored, the volumes within the consistency group must first be removed from the consistency group and added back only after the consistency group is defined as mirrored.

Adding a volume to an already consistency group

Only mirrored volumes can be added into a mirrored consistency group. This operations requires the following:

- Volume peer is on the same system as the peers of the consistency group

- Volume replication type is identical to the type used by the consistency group. For example, `async_interval`.
- Volume belongs to the same storage pool of the consistency group
- Volume has the same schedule as the consistency group
- Volume has the same RPO as the consistency group
- Volume and consistency group are in the same synchronization status (SYNC_BEST_EFFORT for synchronous mirroring, RPO OK for asynchronous mirroring)

If the mirrored consistency group is configured with a user-defined schedule, meaning not using the Never schedule:

Mirrored consistency group or volume should not have non-started snapshot mirrors, non-finished snapshot mirrors (ad hoc Sync Jobs), or both.

If the mirrored consistency group is configured with a Never schedule:

Mirrored consistency group or volume should not have non-started, non-finished snapshot mirrors, non-finished snapshot mirrors (ad hoc Sync Jobs), or both. The status of the mirrored consistency group shall be Initialization until the next Sync Job is completed.

Adding a mirrored volume to a non-mirrored consistency group

It is possible to add a mirrored volume to a non-mirrored consistency group, and it will retain its mirroring settings.

A single sync job for the entire consistency group

The mirrored consistency group has a single Sync Job for all pertinent mirrored volumes within the consistency group.

Location of the mirrored consistency group

All mirrored volumes in a consistency group are mirrored on the same system.

Retaining mirroring attributes of a volume upon removing it from a mirrored consistency group

When removing a volume from a mirrored consistency group, the corresponding peer volume is removed from the peer consistency group. Mirroring is retained (same configuration as the consistency group from which it was removed). Peer volume is also removed from peer consistency group. Ongoing consistency group Sync Jobs will continue.

Mirroring activation of a consistency group

Activation and deactivation of a consistency group affects all consistency group volumes.

Role updates

Role updates concerning a consistency group affects all consistency group volumes.

Dependency of the volume on its consistency group

- It is not possible to directly activate, deactivate, or update role of a given volume within a consistency group from the UI.
- It is not possible to directly change the interval of a given volume within a consistency group.
- It is not possible to set independent mirroring of a given volume within a consistency group.

Protecting the mirrored consistency group

Consistency group-related commands, such as moving a consistency group, deleting a consistency group and so on, are not allowed as long as the consistency group is mirrored. You must remove mirroring before you can delete a consistency group, even if it is empty.

Setting a consistency group to be mirrored

Volumes added to a mirrored consistency group have to meet some prerequisites.

Volumes that are mirrored together as part of the same consistency group share the same attributes (see here).

- Target
- Pool
- Sync type
- Mirror role
- Schedule
- Mirror state
- Last_replicated snapshot timestamp

In addition, their snapshots are all part of the same last_replicated snapshot group.

To obtain the consistency of these attributes, setting the consistency group to be mirrored is done by first creating a consistency group, then setting it to be mirrored and only then populating it with volumes. These settings mean that adding a new volume to a mirrored consistency group requires having the volume set to be mirrored exactly as the other volumes within this consistency group, including the last_replicated snapshot timestamp (which entails an RPO_OK status for this volume).

Note: A non-mirrored volume cannot be added to a mirrored consistency group. It is possible, however, to add a mirrored volume to a non-mirrored consistency group, and have this volume retain its mirroring settings.

Setting-up a mirrored consistency group

The process of creating a mirrored consistency group comprises the following steps.

Step 1 Define a consistency group as mirrored (the consistency group must be empty).

Step 2 Activate the mirror.

Step 3 Add a corresponding mirrored volume into the mirrored consistency group. The mirrored consistency group and the mirrored volume must have the following identical parameters:

- Source and target
- Pools
- Mirroring type
- RPO
- Schedule names (both local and remote)
- Mirror state is RPO_OK
- Mirroring status is Activated

Note: It is possible to add a mirrored volume to a non-mirrored consistency group. In this case, the volume retains its mirroring settings.

Adding a mirrored volume to a mirrored consistency group

After the volume is mirrored and shares the same attributes as the consistency group, you can add the volume to the consistency group after certain conditions are met.

The following conditions must be met:

- The volume is on the same system as the consistency group
- The volume belongs to the same storage pool as the consistency group
- Both the volume and the consistency group do not have outstanding sync jobs, either scheduled or manual (ad hoc)
- The volume and consistency group have the same synchronization status (synchronized="best_effort" and async_interval="rpo_ok")
- The volume's and consistency group's special snapshots, most_recent and last_replicated, have identical timestamps (this is achieved by assigning the volume to the schedule that is utilized by the consistency group)
- In the case that the consistency group is assigned with schedule="never", the status of the consistency group is initialization as long as no sync job has run.

Removing a volume from a mirrored consistency group

Removal of a volume from a mirrored consistency group is easy and preserves volume mirroring.

When you remove a volume from a mirrored consistency group, the corresponding peer volume is removed from the peer consistency group; mirroring is retained with the same configuration as the consistency group from which it was removed. All ongoing consistency group's sync jobs keep running.

Administering asynchronous mirroring

Accommodating disaster recovery scenarios

A disaster is a situation where one of the sites (either the master or the slave) fails, or the communication between the master site and the slave site is lost.

The IBM XIV asynchronous mirroring attains synchronization between Master and Slave peers through a recurring data replication process called a Sync Job. Running at user-configurable schedules, the Sync Job takes a snapshot (called the most_recent snapshot) of the Master and compares this snapshot with another snapshot representing the latest Master state already replicated to the Slave (and a valid recovery point for DR scenarios - called last_replicated snapshot). The Sync Job then synchronizes the Master data corresponding to these differences with the Slave. At the completion of a sync job, a new last_replicated snapshot is created both on the Slave and on the Master.

Disaster recovery scenarios handle cases in which one of the snapshots mentioned above becomes unavailable. These cases are:

Unplanned service disruption

1 Failover

Unplanned service disruption starts with a failover to the Slave.

The Slave is promoted and becomes the new Master, serving host requests

2 Recovery

Next, whenever the Master and link are restored, the replication is set from the promoted Slave (the new Master) onto the demoted Master (the new Slave).

Alternately: No recovery

If recovery is not possible, a new mirroring is established on the Slave. The original mirroring is deleted and a new mirroring relationship is defined.

3 Failback

Following the recovery, the original mirroring configuration is reestablished. The Master maintains its role and replicates to the Slave.

Planned service disruption

1 Failover

Planned service disruption starts with a failover to the Slave. The Slave is promoted to become the new Master, and the Master is demoted to become the new Slave. The promoted Slave serves host requests, and replicates to the demoted Master.

2 Failback

Following the recovery, the original mirroring configuration is reestablished. The Master maintains its role and replicates to the Slave.

Testing

Testing the slave replica.

Unintentional/erroneous role change

1 Recovery

Recovery from unintentional/erroneous implementation of the XCLI/GUI change role command on the Slave.

These cases are described in the following topics.

Note: Please contact XIV Support in case of disaster or for any testing of Disaster Recovery, in order to get clear guidelines and to secure a successful test.

Unplanned service disruption

Following a failure, the Master is no longer servicing, nor available.

The failure disrupts the replication of all outstanding sync jobs and results in the loss of the data that was recorded since the timestamp of the last_replicated snapshot.

Failover

The Slave must be promoted to Master because of the disruption.

Procedure

1. On the host systems, configure the host connectivity so that the hosts can communicate with the Slave.

2. On the Slave system, promote the Slave to become the Master by issuing the `mirror_change_role` command. This will implicitly revert the Slave to the `last_replicated` snapshot.
3. Map the mirror-related volumes on the Slave to the relevant host by issuing the `map_vol` command on the Slave system.

Recovery

Whenever the Master becomes available after the Failover and the original replication configuration is desired (i.e., failback), the replication has to be established from the new Master (the promoted Slave) to the old Master, and only afterwards a proper Failback procedure could be carried out

Before you begin

Note: Following a Failover, do not change the role of the new Master back to Slave prior to a proper establishing of an operational mirroring replication from the new Master to the old Master. Failing to do so will result in loss of the data written to the Slave after its promotion to Master (since the old Master will be reverted to its `last_replicated` snapshot, dated to before the Failover).

Prior to the establishing of the mirroring between the promoted Slave and the old Master, verify the following:

1. The role of the old Slave is now Master
2. Hosts are configured to write directly to the new Master
3. The mirroring is properly configured
4. Connectivity is up between the old master and the new Master (such connectivity is required before the Master can be changed to Slave, to ensure that the Master will be reverted to a point-in-time corresponding with the `last_replicated` snapshot on the other new Master)

Procedure

1. On the Master system, demote the Master to Slave by issuing the `mirror_change_role` command.

Note: In case of pool space depletion and the deletion of the new Master's snapshot, the demotion will not be possible.

2. On the slave system, activate the mirroring by issuing the `mirror_activate` command. Mirroring upon failover is in standby state and must be explicitly set to active state. Upon activation, the mirroring will run from the promoted Slave (the new Master) to the demoted Master (the new Slave).

No recovery

Whenever the Master cannot be recovered, yet mirroring needs be established for the data on the Slave, the mirroring definition is deleted and a new mirroring definition is set.

Procedure

1. Delete the mirroring between Master and Slave by issuing the `mirror_delete` command on the Slave system.
2. Perform the following actions to define a new mirroring on the Slave system:
 - a. Connect to the host by setting configuration with relevant target issuing the relevant commands.
 - b. Create new mirror by issuing the `mirror_create` command.

- c. Activate the mirror by issuing the `mirror_activate` command.

Failback

Failback restores the original mirroring configuration to what it was prior to the failover.

Before you begin

Prior to carrying out the steps listed below, verify the following:

1. The role of the old Slave is now Master
2. The role on the old Master is now Slave
3. Hosts are configured to write directly to the new Master
4. The mirroring is properly configured

Procedure

1. On the host systems, configure the host connectivity so that hosts can communicate with the old Master.
2. Perform the following actions on the old Slave system:
 - a. Stop I/O from the hosts.
 - b. Wait for the next scheduled sync job to start. Issue the `schedule_list` command to check the mirror's schedule interval. Determine if it is worthwhile to change the mirror's schedule to one with a shorter interval (through issuing the `mirror_change_schedule` command) to expedite the creation of the next sync job.
 - c. Change the mirror schedule to 'never'. This schedule setting guarantees that no new sync jobs will be automatically created by the system.
 - d. Wait for the active sync job to complete. You can verify that no sync job is running, by issuing the `sync_job_list` command and ensuring that no sync jobs are listed for the mirror.
 - e. Switch the roles between the peers by issuing the `mirror_switch_roles` command. This will demote the Master to Slave and will promote the Slave to Master.

Once the link is up – the mirroring will resume (there is no need to explicitly activate it).
3. Perform the following actions on the old Master system:
 - a. Verify that mirroring schedule parameters are configured as needed.
 - b. Map all mirror-related volumes on the Master to the relevant host by issuing the `map_vol` command.

Planned service disruption

The amount of data loss can be predetermined when the service disruption is a planned process.

Failover

The service disruption is planned. The Master is unable to serve hosts requests for a planned period of time during which the Slave needs to function as Master. Being a planned process, the failover can be set to occur immediately following the completion of a sync job, resulting in minimal data loss, or no data loss (if hosts are quiesced (paused) before the creation of the last sync job).

Procedure

1. On the host systems, configure the host connectivity so that hosts can communicate with the Slave.
2. Perform the following steps on the Master system:
 - a. Stop I/O for the relevant hosts.
 - b. Wait for the next scheduled sync job to start. Issue the `schedule_list` command to check the mirror's schedule interval. Determine if it is worthwhile to change the mirror's schedule to one with a shorter interval (by issuing the `mirror_change_schedule` command) to expedite the creation of the next sync job.
 - c. Change the mirror schedule to 'never'. This schedule setting guarantees that no new sync jobs will be automatically created by the system.
 - d. Wait for the active sync job to complete. You can verify that no sync job is running by issuing the `sync_job_list` command and ensuring that no sync jobs are listed for the mirror.
 - e. Switch the roles between the peers by issuing the `mirror_switch_roles` command. This will demote the Master to Slave and will promote the Slave to Master.

Once the link is up, the mirroring will resume (there is no need to explicitly activate it).
3. Perform the following steps on the Slave system:
 - a. Verify that mirroring schedule parameters are configured as needed.
 - b. Map all mirror-related volumes on the Master to the relevant host by issuing the `map_vol` command.
 - c. Deactivate the mirroring by issuing the `mirror_deactivate` command for the period during which the Master system will be unavailable.

What to do next

Since the `mirror_switch_roles` command yields an active mirroring relationship, the mirroring does not need to be explicitly recovered following the planned Failover procedure. However, the mirroring must be explicitly deactivated (using the `mirror_deactivate` command) as long as the demoted Master is unavailable. After the Master becomes available again, the mirroring must be activated using the `mirror_activate` command.

Failback

Failback restores the original mirroring configuration to what it was prior to the failover.

Before you begin

Prior to carrying out the steps listed below, verify the following:

1. The role of the old Slave is now Master
2. The role on the old Master is now Slave
3. Hosts are configured to write directly to the new Master
4. The mirroring is properly configured

Procedure

1. On the host systems, configure the host connectivity so that hosts can communicate with the old Master.
2. Perform the following actions on the old Slave system:

- a. Stop I/O from the hosts.
 - b. Wait for the next scheduled sync job to start. Issue the `schedule_list` command to check the mirror's schedule interval. Determine if it is worthwhile to change the mirror's schedule to one with a shorter interval (through issuing the `mirror_change_schedule` command) to expedite the creation of the next sync job.
 - c. Change the mirror schedule to 'never'. This schedule setting guarantees that no new sync jobs will be automatically created by the system.
 - d. Wait for the active sync job to complete. You can verify that no sync job is running, by issuing the `sync_job_list` command and ensuring that no sync jobs are listed for the mirror.
 - e. Switch the roles between the peers by issuing the `mirror_switch_roles` command. This will demote the Master to Slave and will promote the Slave to Master.
Once the link is up – the mirroring will resume (there is no need to explicitly activate it).
3. Perform the following actions on the old Master system:
 - a. Verify that mirroring schedule parameters are configured as needed.
 - b. Map all mirror-related volumes on the Master to the relevant host by issuing the `map_vol` command.

Testing for service disruption

The IBM XIV Storage System allows for validating the mirror replica on the Slave without disrupting the mirroring.

The validation is achieved through mapping the Slave to the host.

Note: The Failover and Failback processes cannot be tested without disrupting the mirroring.

Failover test Procedure

1. Promote the Slave to Master by issuing the `mirror_change_role` command on the Slave system. This will implicitly revert the Slave to the last_replicated snapshot.
2. Map all mirror-related volumes on the Master to the relevant host by issuing the `map_vol` command on the Slave system.
3. Verify that the new Master is functional and hosts can be read from and write into it.

Following the test Procedure

1. Perform the following actions on the Slave system:
 - a. Remove the mappings between the Slave and the hosts.
 - b. Demote the new Master back to Slave by issuing the `mirror_change_role` command. The peer will be reverted to the last_replicated snapshot.

Note: Ensure that the connectivity is up between the old Master and new Master, in order to ensure that the Master will be reverted to a point-in-time corresponding with the last_replicated snapshot on the new Master.

2. Issue the `mirror_activate` command on the Master system.

Nondisruptive testing Procedure

1. Perform the following actions to duplicate the last_replicated snapshot:
 - a. Duplicate the last_replicated snapshot by issuing the `snapshot_duplicate` (or `snap_group_duplicate`) command.
 - b. Map all mirror-related volumes on the Master to the relevant host by issuing the `map_vol` command.
2. Perform the following actions to copy the last_replicated snapshot:
 - a. Copy the last_replicated snapshot to a new volume by issuing the `vol_copy` command.
 - b. Map all mirror-related volumes on the Master to the relevant host by issuing the `map_vol` command.

Note: The duplicated/copied replica is an independent copy that is created solely for testing purposes, and cannot become a mirroring peer in itself. Therefore, if the duplicated/copied replica is written into by hosts as part of the test, the new data cannot be updated to the real mirror replica.

Unintentional/erroneous application of role change

If the XCLI/GUI `mirror_change_role` command was unintentionally issued on the Slave, the Slave can simply be changed back to its original role.

Overview of select asynchronous mirroring commands

General

mirror_create

This command creates a new mirror definition.

This command includes parameters for specifying mirror type, target, RPO, schedule, the ability to initialize from referenced snapshot, and more.

mirror_delete

This command deletes a mirror.

The command deletes the mirroring relationship – not the corresponding objects (volumes or consistency groups). The command deletes the mirroring definition from both master and slave. In case there is no communication between the master and slave, this command could be operated from the slave.

mirror_activate

This command activates the mirror.

The command sets the mirror to Active state. In this state, the slave is updated together with the master.

mirror_deactivate

This command deactivates the mirror.

The command sets the mirror to Standby state. In this state, only the master is updated. This command can deactivate a batch of mirrors as well.

mirror_change_role

This command changes the role of the local mirror peer between master and slave.

The command needs be run on each mirror peer separately.

For synchronous mirroring, another command, `mirror_switch_roles`, can be used.

mirror_switch_role

This command switches between the roles of the master and the slave volumes. In order to switch roles, the following conditions have to be met:

- The coupling is operational and synchronized
- The master volume is identical to the last_replicated snapshot (in other words - all pertinent volume updates need be replicated)
- In asynchronous mirroring - the mirrored volume has to be identical to the last_replicated snapshot.

The command can be performed only on the master.

Prior to the switch, the system stops accepting new writes to the local volume and performs all pending writes. Only after all pending writes have been committed, the roles are switched. After the command is executed, the coupling is deactivated and the user has to activate it in order to restart synchronization. It is advised to create a snapshot before deactivating the coupling in order to enable recovery from logical errors due to incorrect server configurations.

After the command is executed, the volume that was previously the master volume becomes the slave volume and the volume that was previously the slave volume becomes the master volume.

mirror_change_rpo^(*)

This command changes the mirror RPO.

The RPO has to be larger than the mirror's corresponding schedule interval.

mirror_change_schedule^(*)

This command changes the mirror's assigned schedule.

The schedule's interval has to be shorter than the mirror's corresponding RPO.

mirror_change_remote_schedule^(*)

This command changes the schedule of a remote peer.

The schedule's interval doesn't have to be shorter than the mirror's corresponding RPO.

mirror_change_designation

This command changes the peers' designation, so the primary becomes secondary and the secondary becomes a primary.

Mirroring has to be operational for this command to take effect.

Special

mirror_create_snapshot

This command creates a snapshot on both mirror peers; the snapshot sets the scope for an ad-hoc sync job.

In asynchronous replication, the command establishes a process that takes a point-in-time snapshot (at the moment the command was issued) of the source peer (Master) and synchronizes that point-in-time with the Slave.

The process sets a new sync job to copy the differences between that snapshot and the most recent snapshot that is guaranteed to be synchronized with the target peer.

In contrast, in synchronous replication, the command takes a snapshot of the source peer (Master) and the target peer (Slave) at exactly the same time.

mirror_cancel_snapshot

This command cancels all ad-hoc sync jobs for a master.

The command cancels all snapshot mirrors (ad-hoc sync jobs) of a specified master volume or a master consistency group, that have not started to run yet.

Statistics

mirror_list

This command lists details of defined mirrors

The output details include schedule, designation, role, sync type, sync state, active state, operational state, sync progress, size to synchronize, rpo, last_replicated_snapshot timestamp, and more.

mirror_statistics_get^(*)

This command lists information on completed sync jobs for a specified mirror.

The command presents statistics that are automatically gathered by the system on past sync jobs corresponding to a specified mirrored volume or consistency group. The displayed information includes the following: mirrored object type (volume/cg), Job type (scheduled/ad hoc), Job size (MB), Date and time created, Date and time started to run, Date and time completed, and more.

sync_job_list^(*)

This command lists information on pending or running sync jobs.

The output details the following; Mirroring coupling (volume/CG), type, Mirror Schedule, creation time, and more.

rpo_thresholds_set^(*)

This command sets a system-wide RPO threshold for all mirrors that once crossed will trigger the creation of an event.

Absolute and relative thresholds can be specified.

rpo_thresholds_get^(*)

This command displays the RPO threshold set by rpo_threshold_set.

Schedules

schedule_create

This command defines a schedule.

Valid schedule intervals are: 30s, 1m, 2m, 5m, 10m, 15m, 30m, 1h, 2h, 3h, 4h, 6h, 8h, 12h (there is also a min_interval schedule with a non-configurable 20s interval).

schedule_delete

This command deletes a schedule definition.

schedule_list

This command lists all schedule definitions.

schedule_rename

This command renames a schedule.

schedule_change

This command changes the interval of an existing schedule.

(*) The command is only applicable to asynchronous mirroring. It is irrelevant for synchronous mirroring.

Chapter 10. IP and Ethernet connectivity

The following topics provide a basic explanation of the various Ethernet ports and IP interfaces that can be defined and various configurations that are possible within the IBM XIV Storage System.

The IBM XIV Storage System IP connectivity provides:

- iSCSI services over IP or Ethernet networks
- Management communication

Ethernet ports

The following three types of Ethernet ports are available:

iSCSI service ports

These ports are used for iSCSI over IP or Ethernet services. A fully equipped rack is configured with six Ethernet ports for iSCSI service. These ports should connect to the user's IP network and provide connectivity to the iSCSI hosts. The iSCSI ports can also accept management connections.

Management ports

These ports are dedicated for IBM XIV command-line interface (XCLI) and IBM XIV Storage Management GUI communications, as well as being used for outgoing SNMP and SMTP connections. A fully equipped rack contains three management ports.

Field technician ports

These ports are used for incoming management traffic only (usage is both XCLI and IBM XIV Storage Management GUI access). The ports are utilized only for the field technician's laptop computer and must not be connected to the user's IP network.

IP and Ethernet connectivity

The order of the ports included in the patch panel is outlined here.

External system communication is performed through the patch panel. The order of the ports included in the patch panel are shown in "IP and Ethernet connectivity," and outlined in Table 10 on page 106.

	Function, label, or both	From (See Note 1.)	Position (port number)	Comment
Fibre Channel Module 9 2 4 1 3 Module 8 2 4 1 3 Module 7 2 4 1 3 Module 6 2 4 1 3 Module 5 2 4 1 3 Module 4 2 4 1 3 iSCSI Module 9 1 2 Module 8 1 2 Module 7 1 2 Management Module 6 1 Module 5 1 Module 4 1 VPN Module 6 1 Module 4 1 Laptop - Module 5 1 Module 4 1 Maintenance Module 1 Reserved 1 Modem 1	Fibre channel (See Note 2.)	Module 9 Module 8 Module 7 Module 6 Module 5 Module 4	1 - 3 2 - 4	Ports providing connectivity for fibre-channel communications. The fibre-channel connectivity arrangement is the same for all modules (module 9 through module 4). In a partial rack configuration of six modules, only the fibre-channel ports for modules 5 and 6 are active.
	iSCSI	Module 9	1	Ports providing iSCSI connectivity.
		Module 9	2	In a partial rack configuration of six modules, none of the iSCSI ports is active.
		Module 8	1	
		Module 8	2	
		Module 7	1	
		Module 7	2	
	Management	Module 6	1	Ports providing both IBM XIV command-line interface (XCLI) and IBM XIV Storage Management GUI based interface access, as well as both SNMP and SMTP connectivity.
		Module 5	1	
		Module 4	1	
VPN	Module 6	1	Ports providing communication to a virtual private network (VPN).	
	Module 4	1		
Laptop	Module 5	1	Ports for the field service technician to connect a laptop. This port is used for XCLI and IBM XIV Storage Management GUI communication only.	
	Module 4	1		
Maintenance	Module	1	Ports that provide connections to the maintenance module.	
Reserved	N/A	1	Reserved for future use.	
	N/A	1	Reserved for future use.	

Notes:

- Not applicable (N/A)
- The *fibre channel* label does not actually appear on the patch panel. It is only used in this table to indicate the position of the fibre-channel ports.

iSCSI connections

- As an iSCSI target serves hosts through the iSCSI protocol

- As an iSCSI initiator for remote mirroring, when connected to another
- As an iSCSI initiator for data migration when connected to a third party iSCSI storage system
- As XCLI and IBM XIV Storage Management GUI access over the iSCSI ports

iSCSI definitions

iSCSI ports can be defined according to the following criteria:

- Each iSCSI port can be defined as an IP interface.
- The following configuration options are listed for each iSCSI IP interface:
 - IP address (mandatory)
 - Network mask (mandatory)
 - Default gateway (optional)
 - MTU

Default
1,536

Maximum
4,500

Management connectivity

Management connectivity is used for the following functions:

- Executing XCLI commands through the IBM XIV command-line interface (XCLI)
- Controlling the IBM XIV Storage System through the IBM XIV Storage Management GUI
- Sending e-mail notification messages and SNMP traps about event alerts

To ensure management redundancy in case of module failure, the IBM XIV Storage System management function is accessible from three different IP addresses. Each of the three IP addresses is handled by a different hardware module. The various IP addresses are transparent to the user and management functions can be performed through any of the IP addresses. These addresses can be accessed simultaneously by multiple clients. Users only need to configure the IBM XIV Storage Management GUI or XCLI for the set of IP addresses that are defined for the specific system.

Note: All management IP interfaces must be connected to the same subnet and use the same network mask, gateway, and MTU.

XCLI and IBM XIV Storage Management GUI management

The IBM XIV Storage System management connectivity system allows users to manage the system from both the XCLI and IBM XIV Storage Management GUI. Accordingly, the XCLI and IBM XIV Storage Management GUI can be configured to manage the system through iSCSI IP interfaces. Both XCLI and IBM XIV Storage Management GUI management is run over TCP port 7778. With all traffic encrypted through the Secure Sockets Layer (SSL) protocol.

System-initiated IP communication

The IBM XIV Storage System can also initiate IP communications to send event alerts as necessary. Two types of system-initiated IP communications exist:

Sending e-mail notifications through the SMTP protocol

E-mails are used for both e-mail notifications and for SMS notifications through the SMTP to SMS gateways.

Sending SNMP traps

Note: SMPT and SNMP communications can be initiated from any of the three IP addresses. This is different from XCLI and IBM XIV Storage Management GUI, which are user initiated. Accordingly, it is important to configure all three IP interfaces and to verify that they have network connectivity.

Field technician ports

The IBM XIV Storage System supports two Ethernet ports. These ports are dedicated for the following reasons:

- Field technician use
- Initial system configuration
- Direct connection for service staff when they can not connect to customer network
- Directly manage the IBM XIV Storage System through a laptop computer

Laptop connectivity - configuring using DHCP

Two field technician ports are provided for redundancy purposes. This ensures that field technicians will always be able to connect a laptop to the IBM XIV Storage System. These two ports use a Dynamic Host Configuration Protocol (DHCP) server. The DHCP server will automatically assign IP addresses to the user's laptop and connects the laptop to the IBM XIV Storage System network. A laptop connected to any of the field technician ports is assigned an IP address and the IBM XIV Storage Management GUI or IBM XIV command-line interface (XCLI) will typically use the predefined configuration *direct-technician-port*.

Note: The two field technician laptop ports are used only to connect directly to the IBM XIV Storage System and should never be connected to the customer's network.

Laptop connectivity - configuring without DHCP

If the technician's laptop is not setup to receive automatic IP configuration information through DHCP, the laptop should be defined using these parameters:

IP address:

14.10.202.1

Netmask:

255.255.255.0

Gateway:

none

MTU:

1536

The field technician ports accept both XCLI and IBM XIV Storage Management GUI communications. SNMP and SMTP alerts are not sent through these ports.

Note: Each of the field technician ports is connected to a different module. Therefore, if a module fails, the port will become inoperative. When this happens, the laptop should be connected to the second port. See “IP and Ethernet connectivity” on page 105 for descriptions of the ports on each module.

Configuration guidelines summary

When shipped, the IBM XIV Storage System does not have any IP management configurations. Accordingly, the following procedures should be performed when first setting up the system:

- Connecting a laptop to one of the field technician laptop ports on the patch panel
- Configuring at least one management IP interface
- Continuing the configuration process from either the technician port or from the configured IP interface

Note: It is important to define all three management IP interfaces and to test outgoing SNMP and SMTP connections from all three interfaces. The `dest_test` command can be used to test outgoing notifications on a specific module.

Chapter 11. Data migration

The use of any new storage system frequently requires the transfer of large amounts of data from the previous storage system to the new IBM XIV Storage System. This can require many hours or even days; usually an amount of time that most enterprises cannot afford to be without a working system. The data migration feature enables production to be maintained while data transfer is in progress.

Given the nature of the data migration process, it is recommended that you consult and rely on the IBM XIV Storage System support team when planning a data migration.

Data migration overview

The data migration feature enables the smooth transition of a host working with the previous storage system to an IBM XIV Storage System by:

- Immediately connecting the host to the XIV and providing the host with direct access to the most up-to-date data even before data has been copied from the previous storage system.
- Synchronizing the XIV from the previous storage system by transparently copying the contents of the previous storage system to the XIV as a background process.

During data migration, the host is connected directly to the XIV and is disconnected from the previous storage system. The XIV is connected to the previous storage system, as shown in Figure 41. The XIV and the previous storage system must remain connected, until both storage systems are synchronized and data migration is completed. The previous storage system perceives the XIV as a host, reading from and optionally writing to the volume that is being migrated. The host reads and writes data to the XIV, while the XIV might need to read or write the data to the previous storage system to serve the command of the host.

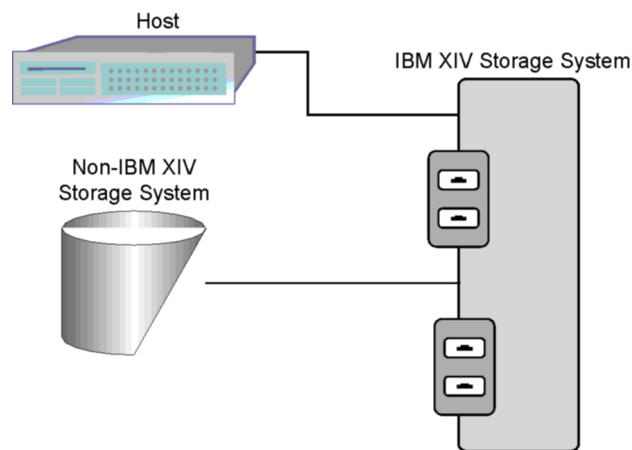


Figure 41. The Data migration process is performed per volume.

The communication between the host and the XIV and the communication between the XIV and the previous storage system is fibre-channel.

I/O handling in data migration

Serving read requests

During this stage the IBM XIV Storage System will serve all the host's data read requests. This will be performed in a transparent manner without requiring any action by the host, as follows:

- If the requested data has already been copied to the XIV, it is served from the XIV.
- If the requested data has not yet been copied to the XIV, the XIV will retrieve it from the previous storage system and then serve it to the host.

Serving write requests

During this stage the XIV will serve all host's data write requests. This will be performed in a transparent manner without requiring any action by the host.

Data migration provides the following two alternative XIV configurations for handling write requests from a host:

Source updating:

A host's write requests are written by the XIV to the XIV, as well as to the previous storage system. In this case, the previous storage system remains completely updated during the background copying process. Throughout the process, the volume of the previous storage system and the volume of the XIV are identical.

Write commands are performed synchronously, so the XIV only acknowledges the write operation after writing to itself, writing to the previous storage system, and receiving an acknowledgement from the previous storage system. Furthermore, if, due to a communication error or any other error, the writing to the previous storage system fails, the XIV will report to the host that the write operation has failed.

No source updating:

A host's write requests are only written by the XIV to the XIV and are not written to the previous storage system. In this case, the previous storage system is not updated during the background copying process, and therefore the two storage systems will never be synchronized. The volume of the previous storage system will remain intact and will not be changed throughout the data migration process.

Data migration stages

Figure 42 on page 113 describes the process of migrating a volume from a previous storage system to the IBM XIV Storage System. It also shows how the XIV synchronizes its data with the previous storage system, and how it handles the data requests of a host throughout all these stages of synchronization.

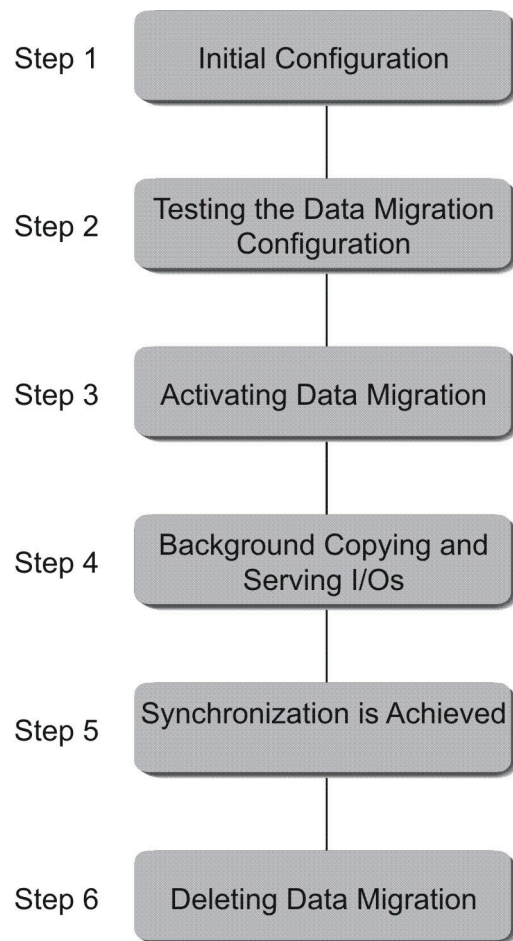


Figure 42. Data migration steps

Initial configuration

The XIV volume must be formatted before data migration can begin. The XIV must be connected as a host to the previous storage system whose data it will be serving.

The volume on the previous storage system and the volume on the XIV must have an equal number of blocks. This is verified upon activation of the data migration process.

You can then initiate data migration and configure all hosts to work directly and solely with the XIV.

Data migration is defined through the `dm_define` command.

Testing the data migration configuration

Before connecting the host to the XIV, use the `dm_test` XCLI command to test the data migration definitions to verify that the XIV can access the previous storage system.

Activating data migration

After you have tested the connection between the XIV and the previous storage system, activate data migration using the `dm_activate` XCLI command and connect the host to the XIV. From this point forward, the host reads and writes data to the XIV, and the XIV will read and optionally write to the previous storage system.

Data migration can be deactivated using the `dm_deactivate` command. It can then be activated again. While the data migration is deactivated, the volume cannot be accessed by hosts (neither read nor write access).

Background copying and serving I/O operations

Once data migration is initiated, it will start a background process of sequentially copying all the data from the previous storage system to the XIV.

Synchronization is achieved

After all of a volume's data has been copied, the data migration achieves synchronization. After synchronization is achieved, all read requests are served from the XIV.

If source updating was set to Yes, the XIV will continue to write data to both itself and the previous storage system until data migration settings are deleted.

Deleting data migration

Data migration is stopped by using a delete command. It cannot be restarted.

Handling failures

Upon a communication error or the failure of the previous storage system, the IBM XIV Storage System will stop serving I/O operations to hosts, including both read and write requests.

If the XIV encounters a media error on the previous storage system (meaning that the XIV cannot read a block on the previous storage system), then the XIV will reflect this state on its own storage system (meaning that it will mark this same block and an error on its own storage system). The state of this block will indicate a media error even though the disk in the XIV has not failed.

Chapter 12. Event handling

The IBM XIV Storage System monitors the health, the configuration changes, and the activity of your storage systems, and generates system events. These events are accumulated by the system and can help the user in the following two ways:

- Users can view past events using various filters. This is useful for troubleshooting and problem isolation.
- Users can configure the system to send one or more notifications, which are triggered upon the occurrence of specific events. These notifications can be filtered according to the events, severity and code. Notifications can be sent through e-mail, SMS messages, or SNMP traps.

Event information

Events are created by various processes, including the following:

- Object creation or deletion, including volume, snapshot, map, host, and storage pool
- Physical component events
- Network events

Each event contains the following information:

- A system-wide unique numeric identifier
- A code that identifies the type of the event
- Creation timestamp
- Severity
- Related system objects and components, such as volumes, disks, and modules
- Textual description
- Alert flag, where an event is classified as alerting by the event notification rules, as described in “Defining events notification rules” on page 116
- Cleared flag, where alerting events can be either uncleared or cleared. This is only relevant for alerting events.

Event information can be classified with one of the following severity levels:

Critical

Requires immediate attention

Major Requires attention soon

Minor Requires attention within the normal business working hours

Warning

Nonurgent attention is required to verify that there is no problem

Informational

Normal working procedure event

Viewing events

The IBM XIV Storage System provides the following variety of criteria for displaying a list of events:

- Before timestamp
- After timestamp
- Code
- Severity from a certain value and up
- Alerting events, meaning events that are sent repeatedly according to a snooze timer
- Uncleared alerts

The number of displayed filtered events can be restricted.

Defining events notification rules

The IBM XIV Storage System monitors the health, configuration changes, and activity of your storage systems and sends notifications of system events as they occur. Event notifications are sent according to the following rules:

Which events

The severity, event code, or both, of the events for which notification is sent.

Where The destinations or destination groups to which notification is sent, such as cellular phone numbers (SMS), e-mail addresses, and SNMP addresses.

Notifications are sent according to the following rules:

Destination

The destinations or destination groups to which a notification of an event is sent. Destinations are described in “Defining destinations” on page 117.

Filter A filter that specifies which events will trigger the sending of an event notification. Notification can be filtered by event code, minimum severity (from a certain severity and up), or both.

Alerting

To ensure that an event was indeed received, an event notification can be sent repeatedly until it is cleared by an XCLI command or the IBM XIV Storage Management GUI. Such events are called alerting events. Alerting events are events for which a snooze time period is defined in minutes. This means that an alerting event is resent repeatedly each snooze time interval until it is cleared. An alerting event is uncleared when it is first triggered, and can be cleared by the user. The cleared state does not imply that the problem has been solved. It only implies that the event has been noted by the relevant person who takes the responsibility for fixing the problem. There are two schemes for repeating the notifications until the event is clear: snooze and escalation.

Snooze

Events that match this rule send repeated notifications to the same destinations at intervals specified by the snooze timer until they are cleared.

Escalation

You can define an escalation rule and escalation timer, so that if events are not cleared by the time that the timer expires, notifications are sent to the

predetermined destination. This enables the automatic sending of notifications to a wider distribution list if the event has not been cleared.

Alerting events configuration limitations

The following limitations apply to the configuration of alerting rules:

- Rules cannot escalate to nonalerting rules, meaning to rules without escalation, snooze, or both.
- Escalation time should not be defined as shorter than snooze time.
- Escalation rules must not create a loop (cycle escalation) by escalating to itself or to another rule that escalates to it.
- The configuration of alerting rules cannot be changed while there are still uncleared alerting events.

Defining destinations

Event notifications can be sent to one or more destinations, meaning to a specific SMS cell number, e-mail address, or SNMP address, or to a destination group comprised of multiple destinations. Each of the following destinations must be defined as described:

SMS destination

An SMS destination is defined by specifying a phone number. When defining a destination, the prefix and phone number should be separated because some SMS gateways require special handling of the prefix.

By default, all SMS gateways can be used. A specific SMS destination can be limited to be sent through only a subset of the SMS gateways.

E-mail destination

An e-mail destination is defined by an e-mail address. By default, all SMTP gateways are used. A specific destination can be limited to be sent through only a subset of the SMTP gateways.

SNMP managers

An SNMP manager destination is specified by the IP address of the SNMP manager that is available to receive SNMP messages.

Destination groups

A destination group is simply a list of destinations to which event notifications can be sent. A destination group can be comprised of SMS cell numbers, e-mail addresses, SNMP addresses, or any combination of the three. A destination group is useful when the same list of notifications is used for multiple rules.

Defining gateways

Event notifications can be sent by SMS, e-mail, or SNMP manager. This step defines the gateways that will be used to send e-mail or SMS.

E-mail (SMTP) gateways

Several e-mail gateways can be defined to enable notification of events by e-mail. By default, the IBM XIV Storage System attempts to send each e-mail notification through the first available gateway according to the order that you specify. Subsequent gateways are only attempted if the first attempted gateway returns an error. A specific e-mail destination can also be defined to use only specific gateways.

All event notifications sent by e-mail specify a sender whose address can be configured. This sender address must be a valid address for the following two reasons:

- Many SMTP gateways require a valid sender address or they will not forward the e-mail.
- The sender address is used as the destination for error messages generated by the SMTP gateways, such as an incorrect e-mail address or full e-mail mailbox.

E-mail-to-SMS gateways

SMS messages can be sent to cell phones through one of a list of e-mail-to-SMS gateways. One or more gateways can be defined for each SMS destination.

Each such e-mail-to-SMS gateway can have its own SMTP server, use the global SMTP server list, or both.

When an event notification is sent, one of the SMS gateways is used according to the defined order. The first gateway is used, and subsequent gateways are only tried if the first attempted gateway returns an error.

Each SMS gateway has its own definitions of how to encode the SMS message in the e-mail message.

Chapter 13. Access control

The IBM XIV Storage System features role-based authentication either natively or by using LDAP-based authentication. The system provides:

Role-based access control

Built-in roles for access flexibility and a high level of security according to predefined roles and associated tasks.

Two methods of access authentication

The IBM XIV Storage System supports the following methods of authenticating users:

Native authentication

This is the default mode for authentication of users and groups on the IBM XIV Storage System. In this mode, users and groups are authenticated against a database on the system.

LDAP When enabled, the system authenticates the users against an LDAP repository.

Following the description of access control methods, this chapter provides auxiliary information:

- Reporting of logging and events
- Reference to access control cli commands
- Glossary for access control terms

User roles and permission levels

User roles allow specifying which roles are applied and the various applicable limits. Table 11 describes the currently available roles, their level of access and typical use.

Note: None of these system-defined users have access to data.

Table 11. Available user roles

User role	Permissions and limits	Typical usage
Read only	Read only users can only list and view system information.	The system operator, typically, but not exclusively, is responsible for monitoring system status and reporting and logging all messages.

Table 11. Available user roles (continued)

User role	Permissions and limits	Typical usage
Application administrator	Only application administrators can perform the following tasks: • Creating snapshots of specifically assigned volumes • Mapping their own snapshot to a specifically assigned host • Deleting their own snapshot	Application administrators typically manage applications that run on a particular server. Application managers can be defined as limited to specific volumes on the server. Typical application administrator functions are the following: • Managing backup environments: – Creating a snapshot for backups – Mapping a snapshot to backup server – Deleting a snapshot after backup is complete – Updating a snapshot for new content within a volume • Managing software testing environment: – Creating a new application instance – Testing the new application instance
Storage administrator	The storage administrator has permission to perform all functions, except: • Maintenance of physical components or changing the status of physical components • Only the predefined administrator, named <i>admin</i>, can change the passwords of other users	Storage administrators are responsible for all administration functions.
Technician	The technician is limited to the following tasks: • Physical system maintenance • Phasing components in or out of service	Technicians maintain the physical components of the system. Only one predefined technician is specified per system. Technicians are IBM XIV Storage System technical support team members.

Notes:

1. All users can view the status of physical components; however, only technicians can modify the status of components.
2. User names are case sensitive.
3. Passwords are case sensitive.

Predefined users

There are several predefined users configured on the IBM XIV Storage System. These users cannot be deleted.

Storage administrator

This user id provides the highest level of customer access to the system.

Predefined user name: admin

Default password: adminadmin

Technician

This user id is used only by IBM XIV Storage System service personnel.

Predefined user name: technician

Default password: Password is predefined and is used only by the IBM XIV Storage System technicians.

Note: Predefined users are always authenticated by the IBM XIV Storage System, even if LDAP authentication has been activated for them.

Application administrator

The primary task of the application administrator is to create and manage snapshots. Application administrators manage snapshots of a specific set of volumes. The user group to which an application administrator belongs determines the set of volumes which the application administrator is allowed to manage.

User groups

A user group is a group of application administrators who share the same set of snapshot creation permissions. This enables a simple update of the permissions of all the users in the user group by a single command. The permissions are enforced by associating the user groups with hosts or clusters. User groups have the following characteristics:

- Only users who are defined as application administrators can be assigned to a group.
- A user can belong to only a single user group.
- A user group can contain up to eight users.
- If a user group is defined with `access_all="yes"`, application administrators who are members of that group can manage all volumes on the system.

Storage administrators create the user groups and control the various permissions of the application administrators.

User group and host associations

Hosts and clusters can be associated with only a single user group. When a user belongs to a user group that is associated with a host, it is possible to manage snapshots of the volumes mapped to that host. See “Command conditions” on page 122 for additional information about commands for application administrators. User and host associations have the following properties:

- User groups can be associated with both hosts and clusters. This enables limiting application administrator access to specific volumes.
- A host that is part of a cluster cannot also be associated with a user group.
- When a host is added to a cluster, the associations of that host are broken. Limitations on the management of volumes mapped to the host is controlled by the association of the cluster.
- When a host is removed from a cluster, the associations of that host become the associations of the cluster. This enables continued mapping of operations so that all scripts will continue to work.

Listing hosts

The command `host_list` lists all groups associated with the specified host, showing information about the following fields:

Range All hosts, specific host

Default

All hosts

Listing clusters

The command `cluster_list` lists all clusters that are associated with a user group, showing information about the following fields:

Range All clusters, specific cluster

Default

All clusters

Command conditions

The application administrator can perform specific operations through a set of commands. Many of the commands are condition-dependent. Table 12 lists the various commands that application administrators can execute according to association definitions and applicable conditions. If the application administrator is a member of a group defined with `access_all="yes"`, then it is possible to perform the command on all volumes.

Table 12. Application administrator commands

Relevant command	Conditions
<code>cg_snapshot_create</code>	At least one volume in the consistency group is mapped to a host or cluster that is associated with a user group containing the user executing the command.
<code>map_vol</code> <code>unmap_vol</code>	Application administrators can use these commands to map snapshots of volumes that meet both of the following conditions: 1. The volumes were created by an application administrator. 2. The master volume is mapped to a host or cluster associated with a user group containing the user.
<code>vol_lock</code> <code>snapshot_duplicate</code> <code>snapshot_delete</code> <code>snapshot_change_priority</code>	The snapshot master volume is mapped to a host or cluster that is associated with the user group containing the user and the snapshot that was created by an application administrator.
<code>snap_group_lock</code> <code>snap_group_duplicate</code> <code>snap_group_delete</code> <code>snap_group_change_priority</code>	At least one of the master volumes of the snapshots in this group is mapped to a host or cluster that is associated with the user group containing the user, and the specific snapshot group was created by an application administrator.
<code>snapshot_create</code>	The volume is mapped to a host or cluster that is associated with the user group containing the user executing the command. If the command overwrites a snapshot, then the overwritten snapshot must be one previously created by an application administrator.

Authentication methods

The following authentication methods are available:

Native (default)

The user is authenticated by the IBM XIV Storage System based on the submitted username and password, which are compared to user credentials defined and stored on the IBM XIV Storage System.

The user must be associated with an IBM XIV Storage System user role that specifies pertinent system access rights.

This mode is set by default.

LDAP

The user is authenticated by an LDAP directory based on the submitted username and password, which are used to connect with the LDAP server.

Note: See the LDAP section for details.

Predefined users authentication

The administrator and technician roles are always authenticated by the IBM XIV Storage System, regardless of the authentication mode. They are never authenticated by LDAP.

Native authentication

This is the default mode for authentication of users and groups on the IBM XIV Storage System. In this mode, users and groups are authenticated against a database on the system. User management is performed locally on the IBM XIV Storage System.

User configuration

Configuring users requires defining the following options:

Role Specifies the role category that each user has when operating the system. The role category is mandatory. See Table 11 on page 119 in “User roles and permission levels” on page 119 for explanations of each role.

Name Specifies the name of each user allowed to access the system.

Password

All user-definable passwords are case sensitive.

Passwords are mandatory, can be 6 to 12 characters long, use uppercase or lowercase letters as well as the following characters: ~!@#\$\$%^&*()_+-= {} | ; < > ? , . / \ [] .

E-mail E-mail is used to notify specific users about events through e-mail messages. E-mail addresses must follow standard addressing procedures. E-mail is optional. Range: Any legal e-mail address.

Phone and area code

Phone numbers are used to send SMS messages to notify specific users about events. Phone numbers and area codes can be a maximum of 63 digits, hyphens (-) and periods (.) Range: Any legal telephone number; The default is N/A

Password management

Password management is very straightforward in the IBM XIV Software System. The IBM XIV Software System provides a high level of security. The following restrictions apply when working with passwords:

- For security purposes, passwords are not shown in user lists.
- Passwords are user changeable. Users can change only their own passwords.

- A user with the predefined password of admin can change the passwords of other users.
- Passwords are changeable from both the XCLI and the IBM XIV Storage Management GUI.
- Passwords are case-sensitive.
Range: 6 - 12 characters (a - z, A - Z, 0 - 9)

LDAP-authentication

Lightweight Directory Access Protocol (LDAP) support enables the IBM XIV Storage System to authenticate users through an LDAP repository.

When LDAP authentication is enabled, the username and password of a user accessing the IBM XIV Storage System (through CLI or GUI) are used by the IBM XIV system to login into a specified LDAP directory. Upon a successful login, the IBM XIV Storage System retrieves the user's IBM XIV group membership data stored in the LDAP directory, and uses that information to associate the user with an IBM XIV administrative role.

The IBM XIV group membership data is stored in a customer defined, pre-configured attribute on the LDAP directory. This attribute contains string values which are associated with IBM XIV administrative roles. These values might be LDAP Group Names, but this is not required by the IBM XIV Storage System. The values the attribute contains, and their association with IBM XIV administrative roles, are also defined by the customer.

Supported domains

The IBM XIV Storage System supports LDAP authentication of the following directories:

- Microsoft Active Directory
- SUN directory
- Open LDAP

LDAP multiple-domain implementation

In order to support multiple LDAP servers that span over different domains, and in order to use the **memberOf** property, the IBM XIV Storage System allows for more than one role for the Storage Administrator and the Read-Only roles.

The predefined XIV administrative IDs “admin” and “technician” are always authenticated by the IBM XIV Storage System, whether or not LDAP authentication is enabled.

Responsibilities division between the LDAP directory and the storage system

Following are responsibilities and data maintained by the IBM XIV system and the LDAP directory:

LDAP directory

- Responsibilities - user authentication for IBM XIV users, and assignment of IBM XIV related group in LDAP.
- Maintains - Users, username, password, designated IBM XIV related LDAP groups associated with IBM XIV Storage System.

IBM XIV Storage System

- Responsibilities - Determination of appropriate user role by mapping LDAP group to an IBM XIV role, and enforcement of IBM XIV user system access.
- Maintains - mapping of LDAP group to IBM XIV role.

LDAP authentication process

In order to use LDAP authentication, you must perform the following major steps:

1. Define an LDAP server and system parameters
2. Define an XIV user on this LDAP server. The storage system uses this user when searching for authenticated users. This user is later on referred to as system's configured service account.
3. Identify an LDAP attribute in which to store values associated with IBM XIV user roles
4. Define a mapping between values stored in the LDAP attribute and IBM XIV user roles
5. Enable LDAP authentication

Once LDAP has been configured and enabled, the predefined user is granted with login credentials authenticated by the LDAP server, rather than the IBM XIV Storage System itself.

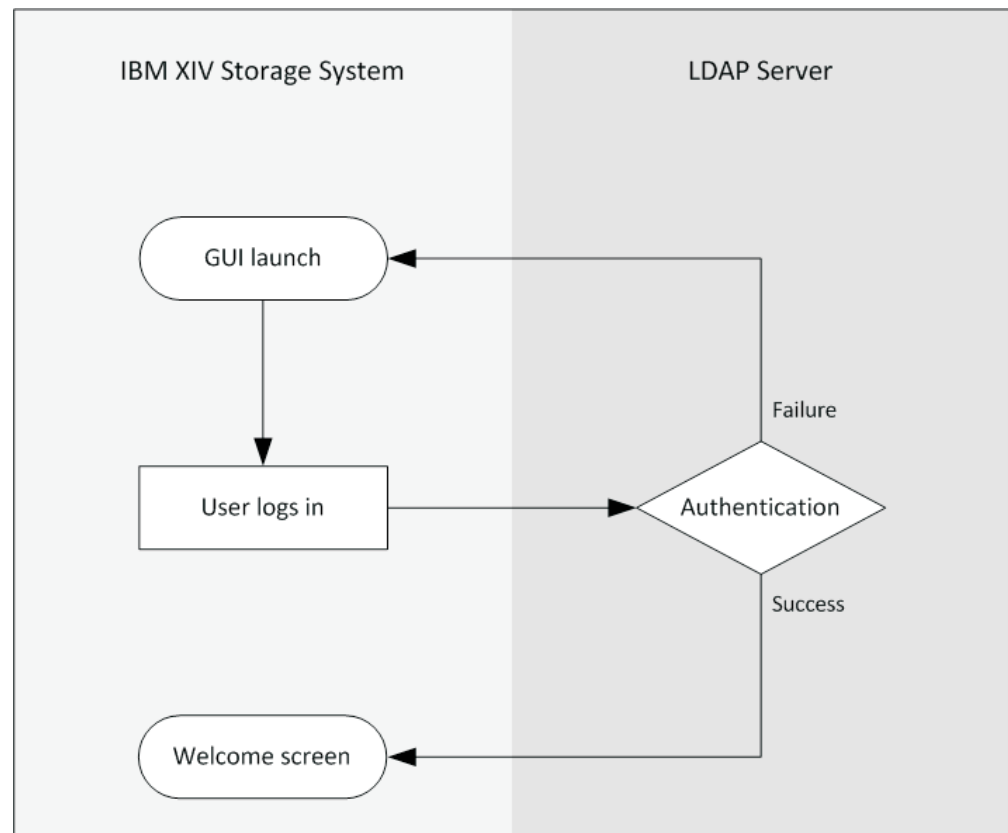


Figure 43. The XIV logs on to a specified LDAP directory

Testing the authentication

The storage administrator can test the LDAP configuration prior to its activation. Look for the `ldap_test` command later on this chapter.

LDAP configuration scenario

Following is an overview of an LDAP configuration scenario:

1. Storage administrator defines the LDAP server(s) to the IBM XIV storage system.
2. Storage administrator defines the LDAP base DN, communication, and timeout parameters to the IBM XIV storage system.
3. Storage administrator defines the LDAP XIV group attribute to be used for storing associations between LDAP groups and XIV storage administrator roles. These are the *storage administrator* and *readonly* roles using the `ldap_config_set` command.
4. Storage administrator defines the mapping between LDAP group name and IBM XIV application administrator roles using the `user_group_create` command.
5. Storage administrator enables LDAP authentication.

LDAP login scenario

LDAP-authenticated login scenario takes the following course:

Initiation

If initiated from the GUI

1. User launches the IBM XIV Storage System GUI.
2. IBM XIV Storage System presents the user with a login screen.
3. User logs in submitting the required user credentials (e.g., username and password).

If initiated from the CLI

1. User logs into the CLI with user credentials (username and password).

Authentication

1. IBM XIV Storage System attempts to log into LDAP directory using the user-submitted credentials.
2. **If login fails:**
 - The IBM XIV Storage System attempts to log into the next LDAP server.
 - **If login fails again** on all servers, a corresponding error message is returned to the user.
3. **If login succeeds**, the IBM XIV Storage System will determine the IBM XIV role corresponding to the logged-in user, by retrieving the user-related attributes from the LDAP directory. These attributes were previously specified by the IBM XIV-to-LDAP mapping.
 - The IBM XIV Storage System will inspect whether the user role is allowed to issue the CLI.
 - If the CLI is permitted for the user's role, it will be issued against the system, and any pertinent response will be presented to the user.
 - If the CLI is not permitted for the user's role, the IBM XIV Storage System will send an error message to the user.

Supported user name characters

The login mechanism supports all characters, including `@`, `*` and `\`, to allow names of the following format:

- UPN: name@domain
- NT domain: domain\name

Searching within indirectly-associated groups:

In addition to the users search, the IBM XIV Storage System allows for searching indirectly-associated Active Directory groups.

Searching for indirectly-associated Active Directory groups is done separately from the user search that was described above, in the “Users validation” section. This search of indirectly-associated group utilizes the group attribute *memberof* and it conveys the following flow.

Note: This search does not apply to SUN directory, as you get all the indirectly-associated groups on the users validation query.

The IBM XIV Storage System search for the group membership starts with the groups the user is directly associated with and spans to other groups. The *memberof* attribute is searched for within each of these groups. The search goes on until one of the following stop criteria is met:

Stop when found

- A group membership that matches one of the configured LDAP rules is found
- The search command is set to stop searching upon finding a group.

Don't stop when found

- A group membership that matches one of the configured LDAP rules is found
- The search command does not stop once a group membership is found. It is set to continue onto the next group.
- The search command is set to stop upon reaching a search limit (see Reaching a limit below).

Multiple findings

- More than a single group membership that matches one of the configured LDAP rules were found
 - Every match will be counted once even if it was found several times (arrived at it from several branches).
 - The search doesn't avoid checking groups that were previously checked from other branches.

Reaching a limit

One of the following limits is met (the limits are set as part of the search command):

- The search reached the search depth limit.
This search attribute limits the span of the search operation within the groups tree.
- The search reached the maximum number of queries limit.

Users validation

During the login, the system validates the user as follows:

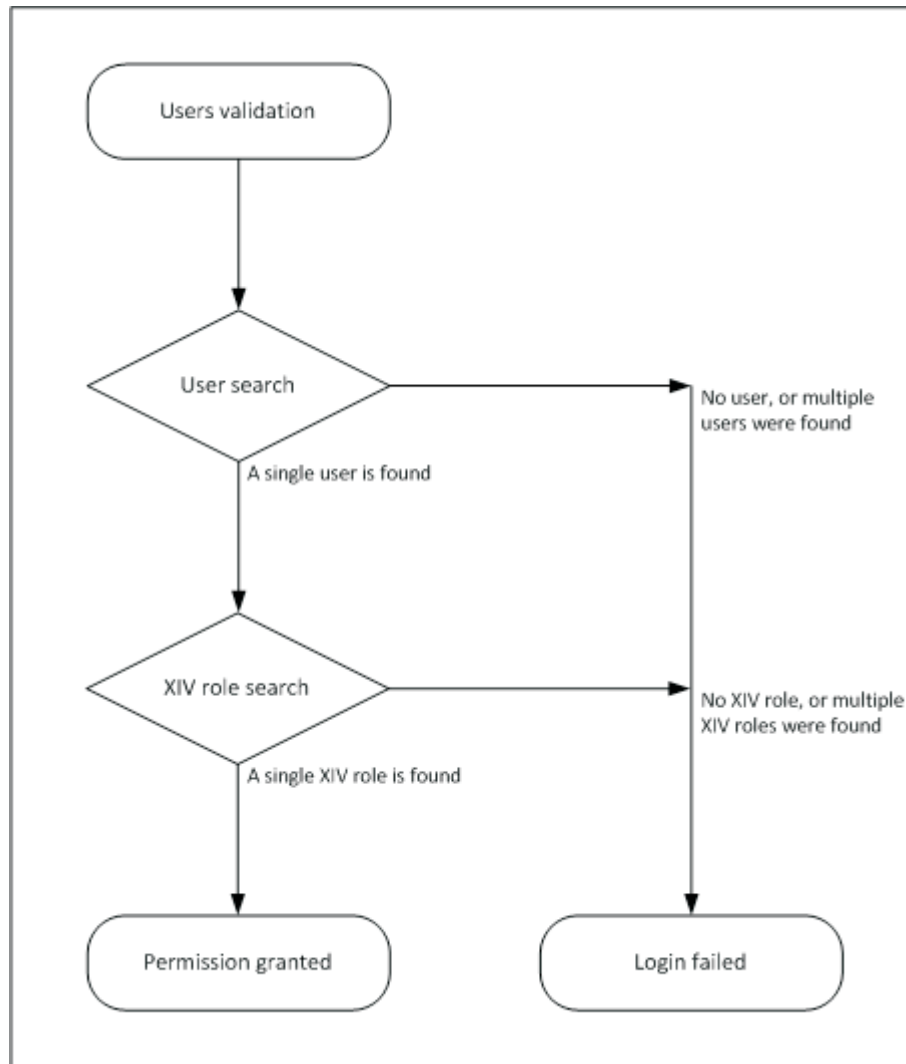


Figure 44. The way the system validates users through issuing LDAP searches

Step 1 - Issuing a user search

The system issues an LDAP search for the user's entered username.

The request is submitted on behalf of the system's configured service account and the search is conducted for the LDAP server, base DN and reference attribute as specified in the XIV LDAP configuration (see above).

The base DN specified in the XIV LDAP configuration serves as a reference starting point for the search – instructing LDAP to locate the value submitted (the username) in the attribute specified (whose value is specified in `user_name_attr`).

Step 2 - If a single user is found - issuing an XIV role search

The system issues a second search request, this time submitted on behalf of the user (with the user's credentials), and will search for XIV roles associated with the user, based on XIV LDAP configuration settings (as specified in parameter `xiv_group_attr`).

If a single XIV role is found - permission is granted

The system inspects the rights associated with that role and

grant login to the user. The user's permissions are in correspondence with the role associated by XIV, base on XIV LDAP configuration.

If no XIV role is found for the user, or more than one role was found

If the response by LDAP indicates that the user is either not associated with an XIV role (no user role name is found in the referenced LDAP attribute for the user), or is actually associated with more than a single role (multiple roles names are found) – login will fail and a corresponding message will be returned to the user.

If no such user was found, or more than one user were found

If LDAP returns no records (indicating no user with the username was found) or more than a single record (indicating that the username submitted is not unique), the login request fails and a corresponding message is returned to the user.

Establishing an IBM XIV service account for LDAP queries

The IBM XIV Storage System carries out the LDAP search through a service account. This service account is established via the `ldap_config_set` command (see in the Access Control commands section below).

Switching between LDAP and native authentication modes

This section describes system behavior when switching between LDAP authentication and native authentication.

After changing authentication modes from native to LDAP

The system will start authenticating users other than "admin" or "technician" against the LDAP server, rather than the local IBM XIV storage system user database. However, the local user account data will not be deleted:

- Users without an account on the LDAP server will not be granted access to the XIV system.
- Users with an LDAP account who are not associated with an XIV role on the LDAP directory will not be granted access to the XIV system.
- Users with an LDAP account who are associated with an XIV role on the LDAP directory will be granted access to the XIV system if the following conditions are met:
 - The XIV role on LDAP is mapped to a valid XIV role on the XIV system
 - The user is associated only to one XIV role on LDAP

The following commands related to user account management will be disabled. These operations must be performed on the LDAP directory.

- `user_define`
- `user_rename`
- `user_update`
- `user_group_add_user`
- `user_group_remove_user`

After changing authentication modes from LDAP to native

The system will start authenticating users against the locally defined IBM XIV user database. Users and groups which were defined prior to switching from native to

LDAP authentication will be reenabled. The IBM XIV system will allow local management of users and groups. The following commands related to user account management will be enabled:

- `user_define`
- `user_rename`
- `user_update`
- `user_group_add_user`
- `user_group_remove_user`

Users must be defined locally on XIV and be associated with local IBM XIV user groups in order to gain access to the system.

Administering access control

Logging and event reporting

Command execution log

The IBM XIV Software System provides logs for tracking command execution and who has performed the command. This log is implemented through the event log.

Object creation tracking

For each system object such as a volume, host, snapshot, consistency group, or pool, the system keeps track of the user who executed the command that created the object. This is part of the object's attributes shown in the relevant list command.

Event report destinations

The system provides mechanisms for sending reports to variety of destinations, including:

- Telephone numbers
- E-mail addresses

The configured e-mail or phone number per user can serve as the destinations for these notifications. See “User configuration” on page 123 for an explanation of destination parameters.

For a detailed explanation about handling events and setting up notification, see Chapter 12, “Event handling,” on page 115.

Access control commands

The following IBM XIV command-line interface (XCLI) commands are available for managing role-based access control (RBAC). For a detailed explanation of these commands, see the chapter detailing “Access control” in the relevant (for the release you are using) *IBM XIV Storage System XCLI User Manual*.

User-related commands

You can use the following user-related commands to manage role-based access control:

user_define

Defines a new user.

user_update

Updates the attributes of the user.

user_list

Lists all users, or a specific user.

user_rename

Renames the user.

user_delete

Deletes the user.

User groups-related commands

You can also use the following user group-related commands to manage role-based access control:

user_group_create

Creates a user group.

user_group_update

- Assigns the user group with a Lightweight Directory Access Protocol (LDAP) role.
- Updates the user group name.

user_group_add_user

Adds a user to a user group.

user_group_remove_user

Removes a user from a user group.

user_group_list

Lists all user groups along with their users.

user_group_rename

Renames a user group.

user_group_delete

Deletes a user group.

Role-based access control commands

The following list of access-related commands can be used to manage role-based access control:

access_define

Associates a user group with a host and a cluster.

access_delete

Dissociates a user group from the host and cluster with which it is associated.

access_list

Lists access associations.

Configuration-related commands

You can also use the following LDAP server configuration-related commands:

ldap_config_set

Sets up the LDAP configuration parameters.

ldap_config_get

Lists the configuration attributes of an LDAP server that works with the storage system.

ldap_mode_set

Enables/disables LDAP authentication to the storage system.

ldap_mode_get

Returns the authentication mode of the storage system (active/inactive).

ldap_user_test

This command authenticates the user's credentials on the LDAP machine.

ldap_test

Validates the LDAP settings prior to the activation.

Non-LDAP commands

The following commands are available in non-LDAP mode and are not available in LDAP mode:

user_define

Defining a new user on the XIV system.

user_update

Modifying the XIV user's details.

user_rename

Renaming an XIV user.

user_group_add_user

Adding a user to an XIV Application Administrator user group.

user_group_remove_user

Removing a user from an XIV application administration user group.

Glossary of access control concepts

The following key definitions apply throughout the discussion on role-based access control:

LDAP Lightweight Directory Access Protocol.

LDAP attribute

A property of an LDAP object, with a single or multiple values. A special object attribute is designated by an LDAP administrator to hold user group memberships values corresponding to XIV roles.

LDAP authentication

A method for authenticating users by validating the user's submitted credentials against data stored on an LDAP directory.

LDAP Directory

A hierarchical database stored on an LDAP server and accessed through LDAP calls.

LDAP Server

A server that provides directory services through LDAP.

LDAP status

The status of an LDAP server.

Microsoft Active Directory

Microsoft Active Directory provides a directory (lookup), DNS and authentication services.

XIV Mapping

An association of data on the LDAP server (a specific LDAP attribute) and data on the XIV. This is required to determine which access rights to grant to an authenticated LDAP user.

Chapter 14. Tivoli Storage Productivity Center and SMI-S interoperability

The IBM XIV Storage System integrates with Tivoli Storage Productivity Center (TPC) by providing system configuration information.

Tivoli Storage Productivity Center then displays this information to TPC end users. In addition to query support, TPC provides a means to launch the XIV GUI directly from within the TotalStorage Productivity Center GUI.

Queries of system configuration are supported starting with the Tivoli Storage Productivity Center version 4.1 and IBM XIV Storage System microcode version 10.1 and up. Tivoli Storage Productivity Center will be allowed to query configuration data including:

- Storage pools
- Volumes
- Disks
- Hosts and host mappings
- Fibre Channel ports

These queries, when combined with SCSI inquiry data that Tivoli Storage Productivity Center collects from the hosts, allows it to correlate LUNs reported by the IBM XIV Storage System to LUNs as seen by the host systems. Also, when the IBM XIV Storage System is a back end to the IBM System Storage SAN Volume Controller (SVC), the Tivoli Storage Productivity Center can correlate LUNs reported by the IBM XIV Storage System to SAN Volume Controller-managed disks (MDisks).

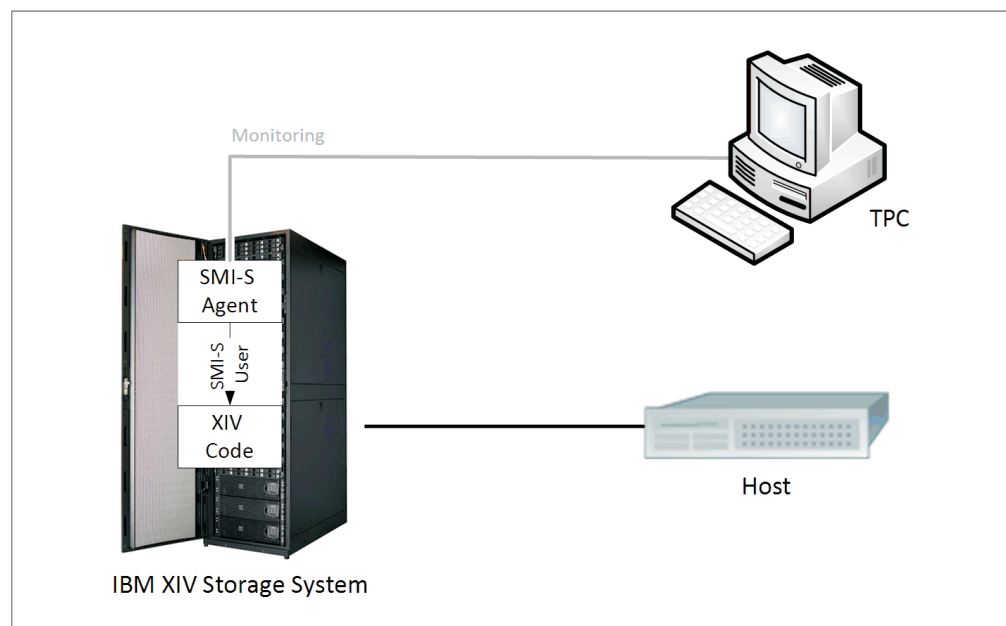


Figure 45. TPC interoperability with IBM XIV Storage System

TPC utilizes a CIM agent which is embedded on the IBM XIV Storage System, where SMI-S 1.2 is supported. This agent queries the IBM XIV Storage System using the `smis_user` ID. The `smis_user` ID is allowed read only access to the IBM XIV Storage System.

The IBM XIV CIM provider will publish itself as a service using the Service Location Protocol (SLP) as a Service Agent (SA). This CIMOM (Common Information Model Object Manager) provider broadcasts its address to allow a directory look-up by a CIM agent, in this case, TotalStorage Productivity Center. This then allows Tivoli Storage Productivity Center to query for the IP address, namespace, and supported profiles for the IBM XIV CIM provider, thus discovering it.

CIM and IBM XIV user groups

An IBM XIV CIM user is permitted to run commands like `vol_list` and `pool_list`, regardless of the IBM XIV user group it belongs to.

Related TPC documentation

Relevant documentation includes:

- Adding a user
- Collecting the log
- Application Programming Interface Reference, available on the IBM XIV Information Center

Chapter 15. Nondisruptive code load

Non-disruptive-code-load (a.k.a. Hot Upgrade) enables the IBM XIV Storage System to upgrade its software from a current version to a newer version without disrupting application service.

The upgrade process is run on all modules in parallel and is designed to be quick enough so that the applications' service on the hosts will not be damaged. The upgrade requires that neither data migration nor rebuild processes are run, and that all internal network paths are active.

During the Non-Disruptive Code Load process there is a point in time dubbed the 'upgrade-point-of-no-return', before which the process can still be aborted (either automatically by the system - or manually through a dedicated CLI). Once that point is crossed - the Non-Disruptive Code Load process is not reversible.

Following are notable characteristics of the Non-disruptive code load:

Duration of the upgrade process

The overall process of downloading new code to storage system and moving to the new code is done online to the application/Host.

The duration of the upgrade process is affected by the following factors:

- The upgrade process requires to stop all IOs. If there are a lot of IOs in the system, or there are slow disks, the system might not be able to stop the IOs fast enough, so it will restart them and try again after a short while, taking into consideration some retries.
- The upgrade process installs a valid version of the software and then retains its local configuration. This process might take a considerable amount of time, depending on the future changes in the structure of the configuration.

Prerequisites and constraints

- The process cannot run if a data migration process or a rebuild process is active. An attempt to start the upgrade process when either a data migration or a rebuild process is active will fail.
- Generally, everything that happens after the point-of-no-return is treated as if it happened after the upgrade is over.
- As long as the overall hot upgrade is in progress (up to several minutes) no management operations are allowed (save for status querying), and no events are processed.
- Prior to the point-of-no-return, a manual abort of the upgrade is available.

Effect on mirroring

Mirrors are automatically deactivated before the upgrade, and reactivated after it is over.

Effect on management operations

During the Non-Disruptive Code Load process it is possible to query the system about the upgrade status, and the process can also be aborted manually before the 'point-of-no-return'. If a failure occurs before this point - the process will be aborted automatically.

Handling module or disk failure during the upgrade

If the failure occurs before the point-of-no-return, it will abort the upgrade.
If it happens after that point, the failure is treated as if it happened after the upgrade is over.

Handling power failure during the upgrade

As for power failure before the point-of-no-return - power is being monitored during the time the system prepares for the upgrade (before the point-of-no-return). If a power failure is detected, the upgrade will be aborted and the power failure will be taken care of by the old version.

Glossary

The following is an alphabetical list of terms and abbreviations that are used throughout this document.

Active directory

Microsoft Active Directory (AD) provides directory (lookup), DNS and authentication services.

Alerting event

An event that triggers recurring event notifications until it is cleared.

API See *Application program interface (API)*.

Application program interface (API)

The interface through which the application accesses the operating system and the other services.

Authorization level

The authorization level determines the permitted access level to the various functions of the IBM XIV Storage Management GUI:

Read only

Only viewing is allowed.

Full Enables access to all the configuration and control functions, including shutdown of the system. This level requires a password.

Auto delete priority

As the storage capacity reaches its limits, snapshots are automatically deleted to make more space. The deletion takes place according to the value set for each snapshot, as follows:

1 last to be deleted

4 first to be deleted

Each snapshot is given a default auto delete priority of 1 at creation.

Clearing events

The process of stopping the recurring event notification of alerting events.

CLI The IBM XIV command-line interface (XCLI). See *Command line interface (CLI)*

Command line interface (CLI)

The nongraphical user interface used to interact with the system through set commands and functions. The IBM XIV command-line interface (XCLI) for the IBM XIV Storage System.

Completion code

The returned message sent as a result of running CLI commands.

Consistency group

A cluster of specific volumes that can all be snapshotted, mirrored and administered simultaneously as a group. A volume can only be associated with a single consistency group.

The volumes within a consistency group are grouped into a single volume set. The volume set can be snapshotted into multiple snapshot sets under the specific consistency group. See also *Snapshot set, Volume set*.

Coupling

The two peers (volumes or consistency groups) between which a mirroring relationship was set.

Data module

A module dedicated to data storage. A fully-configured rack contains 9 dedicated data modules, each with 12 disks.

Destination

See *Event destination*.

Escalation

A process in which event notifications are sent to a wider list of event destinations because the event was not cleared within a certain time.

Event destination

An address for sending event notifications.

Event notification rule

A rule that determines which users are to be notified, for which events and by what means.

Event notification

The process of notifying a user about an event.

Event A user or system activity that is logged (with an appropriate message).

Fabric The hardware that connects workstations and servers to storage devices in a SAN. The SAN fabric enables any-server-to-any-storage device connectivity through the use of fibre-channel switching technology.

FC-AL Also known as arbitrated loop. A fibre-channel topology that requires no fibre-channel switches. Devices are connected in a one-way loop fashion.

FC-HBA

Fibre-channel host bus adapter.

FC See *Fibre channel*.

Fibre channel

Serial data transfer architecture developed by a consortium of computer and mass storage device manufacturers and now being standardized by ANSI.

Functional area

One of the high level groupings of icons (functional modules) of the left-hand pane of the IBM XIV Storage Management GUI screen. For example: Monitor, Configuration or Volume management. See *Functional module*.

Functional module

One of the icons of a functional area, on the left-hand pane of the IBM XIV Storage Management GUI screen. For example, System (under Monitor) or Hosts and LUNs (under Configuration). See *Functional area*.

Graphical user interface (GUI)

On-screen user interface supported by a mouse and a keyboard.

H/W Hardware.

HBA Host bus adapter.

Host interface module

The interface data module serves external host requests with the ability to store data. A fully-configured rack has 6 interface data modules.

- Host** A host is a port name of a host that can connect to the system. The system supports fibre channel and iSCSI hosts.
- I/O** Input/output.
- Image snapshot**
A snapshot that has never been unlocked. It is the exact image of the master volume it was copied from, at the time of its creation. See also *snapshot*.
- Internet Protocol**
Specifies the format of packets (also called datagrams), and their addressing schemes. See also *Transmission Control Protocol (TCP)*.
- IOPs** Input/output (I/O) per second.
- IP** See *Internet Protocol*.
- iSCSI** Internet SCSI. An IP-based standard for linking data storage devices over a network and transferring data by carrying SCSI commands over IP networks.
- Latency**
Amount of time delay between the moment an operation is issued, and the moment it is committed.
- LDAP** Lightweight Directory Access Protocol.
- LDAP attribute**
An attribute defined in an LDAP directory data model.
- LDAP authentication**
A method for authenticating users by validating the user's submitted credentials against data stored on an LDAP directory.
- LDAP directory**
A hierarchical database stored on an LDAP server and accessed through LDAP calls.
- LDAP server**
A server that provides directory services through LDAP.
- LDAP status**
The status of an LDAP server.
- Load balancing**
Even distribution of load across all components of the system.
- Locking**
Setting a volume (or snapshot) as unwritable (read-only).
- LUN map**
A table showing the mappings of the volumes to the LUNs.
- LUN** Logical unit number. Exports a systems volume into a registered host.
- Master volume**
A volume that has snapshots is called the master volume of its snapshots.
- MIB** Management information base. A database of objects that can be monitored by a network management system. SNMP managers use standardized MIB formats to monitor SNMP agents.
- Microsoft Active directory**
See *Active Directory*

Mirror peer

A peer (volume or consistency group) that is designated to be a replica of a specified source peer data.

Mirroring

See *Remote mirroring*.

Modified State

A snapshot state. A snapshot in modified state can never be used for restoring its master volume.

Multipathing

Enables host interface modules direct access to any volume.

Peer Denotes a constituent side of a coupling. Whenever a coupling is defined, a designation is specified for each peer - one peer is designated primary and the other is designated secondary.

Pool See *Storage pool*.

Primary peer

A peer whose data is mirrored for backup on a remote storage system.

Rack The cabinet that stores all of the hardware components of the system.

Remote mirroring

The process of replicating the content of a source peer (volume or consistency group) to a designated mirror peer.

Remote target connectivity

A definition of connectivity between a port set of a remote target and a module on the local storage system.

Remote target

An storage system on a remote site, used for mirroring, data migration, and so on.

Role Denotes the actual role that the peer is fulfilling as a result of a specific condition, either a master or a slave.

Rule See *Event notification rule*.

SAN Storage area network.

SCSI Small computer system interface.

Secondary peer

A peer that serves as a backup of a primary peer.

SMS gateway

An external server that is used to send SMSs.

SMTP gateway

An external host that is used to relay e-mail messages through the SMTP protocol.

Snapshot set

The resulting set of synchronized snapshots of a volume set in a consistency group. See also *Consistency group*, *Volume set*.

Snapshot

A point-in-time snapshot or copy of a volume. See also *Image snapshot*.

SNMP agent

A device that reports information through the SNMP protocol to SNMP managers.

SNMP manager

A host that collects information from SNMP agents through the SNMP protocol.

SNMP trap

An SNMP message sent from the SNMP agent to the SNMP manager, where the sending is initiated by the SNMP agent and not as a response to a message sent from the SNMP manager.

SNMP

Simple Network Monitor Protocol. A protocol for monitoring network devices. See also *MIB*, *SNMP agent*, *SNMP manager*, *SNMP trap*.

Snooze

The process of sending recurring event notifications until the events are cleared.

Storage pool

A reserved area of virtual disk space serving the storage requirements of the volumes.

Sync best effort mode

A mode of remote mirroring in which I/O operations are not suspended when communication between a primary and secondary volume is broken.

Synchronization

The process of making the primary volume and secondary volume identical after a communication down time or upon the initialization of the mirroring.

Target See *Remote target*.

TCP/IP

See *Transmission Control Protocol*, *Internet Protocol*.

Thin provisioning

Thin provisioning provides the ability to define logical volume sizes that are much larger than the physical capacity installed on the system.

Transmission Control Protocol

Transmission Control Protocol (TCP) on top of the Internet Protocol (IP) establishes a virtual connection between a destination and a source over which streams of data can be exchanged. See also *IP*.

Trap See *SNMP trap*.

Unassociated volume

A volume that is not associated with a consistency group. See *Consistency group*.

Uninterruptible power supply

The uninterruptible power supply provides battery backup power for a determined period of time, particularly to enable the system to power down in a controlled manner, on the occurrence of a lengthy power outage.

Volume cloning

Creating a snapshot from a volume.

Volume set

A cluster of specific volumes in a consistency group, which can all be snapshotted simultaneously, thus, creating a synchronized snapshot of all of them. The volume set can be snapshotted into multiple snapshot sets of the specific consistency group. See also *Snapshot set*, *Volume set*.

Volume

A discrete unit of storage on disk, tape or other data recording medium that supports some form of identifier and parameter list, such as a volume label or input/output control.

A volume is a logical address space, having its data content stored on the systems disk drives. A volume can be virtually any size as long as the total allocated storage space of all volumes does not exceed the net capacity of the system. A volume can be exported to an attached host through a LUN. A volume can be exported to multiple hosts simultaneously. See also *Storage pool, Unassociated volume*.

WWPN

World Wide Port Name

XCLI IBM XIV command-line interface (XCLI) command set. See *Command line interface*.

XDRP The disaster recovery program for the XIV – The remote mirror feature of the XIV.

XIV-LDAP mapping

An association of data on the LDAP server (a specific LDAP attribute) and data on the XIV. This is required to determine the access rights that should be granted to an authenticated LDAP user.

Safety and environmental notices

This section contains information about:

- Safety notices and labels
- Laser safety
- Rack safety
- Product recycling and disposal
- Battery return program
- Fire suppression systems

Danger notices

A danger notice calls attention to a situation that is potentially lethal or extremely hazardous to people. A lightning bolt symbol accompanies a danger notice to represent a dangerous electrical condition. A sample danger notice follows.



DANGER

An electrical outlet that is not correctly wired could place hazardous voltage on metal parts of the system or the devices that attach to the system. It is the responsibility of the customer to ensure that the outlet is correctly wired and grounded to prevent an electrical shock.

A comprehensive danger notice provides instructions on how to avoid shock hazards when servicing equipment. Unless instructed otherwise, follow the procedures in the following danger notice.



DANGER

Electrical voltage and current from power, telephone, and communication cables are hazardous.

To avoid a shock hazard:

- Do not connect or disconnect any cables or perform installation, maintenance, or reconfiguration of this product during an electrical storm.
- Connect all power cords to a properly wired and grounded electrical outlet. Ensure outlet supplies proper voltage and phase rotation according to the system rating plate.
- Connect any equipment that will be attached to this product to properly wired outlets.
- When possible, use one hand only to connect or disconnect signal cables.
- Never turn on any equipment when there is evidence of fire, water, or structural damage.
- Disconnect the attached power cords, telecommunications systems, networks, and modems before you open the device covers, unless instructed otherwise in the installation and configuration procedures.
- Connect and disconnect cables as described below when installing, moving, or opening covers on this product or attached devices.

To Disconnect:

1. Turn everything OFF (unless instructed otherwise).
2. Remove power cords from the outlet.
3. Remove signal cables from connectors.
4. Remove all cables from devices.

To Connect:

1. Turn everything OFF (unless instructed otherwise).
2. Attach all cables to devices.
3. Attach signal cables to connectors.
4. Attach power cords to outlet.
5. Turn device ON.

Labels

As an added precaution, safety labels are often installed directly on products or product components to warn of potential hazards.

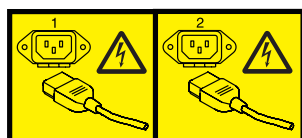
The actual product safety labels might differ from these sample safety labels:



DANGER

Hazardous voltage, current, or energy levels are present inside any component that has this label attached.

Do not service, there are no serviceable parts.



DANGER

Multiple power cords

To remove all power to the device, disconnect all power cords.

Caution notices

A caution notice calls attention to a situation that is potentially hazardous to people because of some existing condition. A caution notice can be accompanied by different symbols, as shown in the examples in Table 13.

Table 13. Caution notice symbols

If the symbol is...	It means...
	A hazardous electrical condition with less severity than electrical danger.
	A generally hazardous condition not represented by other safety symbols.
 Class I	A hazardous condition due to the use of a laser in the product. Laser symbols are always accompanied by the classification of the laser as defined by the U. S. Department of Health and Human Services (for example, Class I, Class II, and so forth).

Sample caution notices:



CAUTION:

This product is equipped with a 3-wire (two conductors and ground) power cable and plug. Use this power cable with a properly grounded electrical outlet to avoid electrical shock.



Class I

CAUTION:

Data processing environments can contain equipment transmitting on system links with laser modules that operate at greater than Class 1 power levels. For this reason, never look into the end of an optical fiber cable or open receptacle.

Attention notices

An attention notice indicates the possibility of damage to a program, device, or system, or to data. An exclamation point symbol may accompany an attention notice, but is not required. A sample attention notice follows:



Attention: Do not bend a fibre cable to a radius less than 5 cm (2 inches); you can damage the cable. Tie wraps are not recommended for optical cables because they can be easily overtightened, causing damage to the cable.

Laser safety

When using an NVRAM5 or NVRAM6 cluster media converter, the storage system must be installed in a restricted access location.

CAUTION:

This product contains a Class 1M laser. Do not view directly with optical instruments. (C028)

This equipment contains Class 1 laser products, and complies with FDA radiation Performance Standards, 21 CFR Subchapter J and the international laser safety standard IEC 825-2.



Class I

CAUTION:

Data processing environments can contain equipment transmitting on system links with laser modules that operate at greater than Class 1 power levels. For this reason, never look into the end of an optical fiber cable or open receptacle.

Attention: In the United States, use only SFP or GBIC optical transceivers that comply with the FDA radiation performance standards, 21 CFR Subchapter J. Internationally, use only SFP or GBIC optical transceivers that comply with IEC standard 825-1. Optical products that do not comply with these standards may produce light that is hazardous to the eyes.

Usage restrictions

The optical ports of the modules must be terminated with an optical connector or with a dust plug.

Rack safety

Rack installation



DANGER

- Always lower the leveling pads on the rack cabinet.
- Always install stabilizer brackets on the rack cabinet.
- To avoid hazardous conditions due to uneven mechanical loading, always install the heaviest devices in the bottom of the rack cabinet. Always install servers and optional devices starting from the bottom of the rack cabinets.
- Rack-mounted devices are not to be used as a shelf or work space. Do not place any object on top of rack-mounted devices.
- Each rack cabinet might have more than one power cord. Be sure to disconnect all power cords in the rack cabinet before servicing any device in the rack cabinet.
- Connect all devices installed in a rack cabinet to power devices installed in the same rack cabinet. Do not plug a power cord from a device installed in one rack cabinet into a power device installed in a different rack cabinet.

CAUTION:

- Do not install a unit in a rack where the internal rack ambient temperatures will exceed the manufacturer's recommended ambient temperature for all your rack-mounted devices.
- Do not install a unit in a rack where the air flow is compromised. Ensure that air flow is not blocked or reduced on any side, front, or back of a unit used for air flow through the unit.
- Consideration should be given to the connection of the equipment to the supply circuit so that overloading of the circuits does not compromise the supply wiring or overcurrent protection.
- To provide the correct power connection to a rack, refer to the rating labels located on the equipment in the rack to determine the total power requirement of the supply circuit.
- This drawer is a fixed drawer and should not be moved for servicing unless specified by manufacturer. Attempting to move the drawer partially or completely out of the rack may cause the rack to become unstable or cause the drawer to fall out of the rack.

Rack relocation (19" rack)

CAUTION:

Removing components from the upper positions in the rack cabinet improves rack stability during relocation. Follow these general guidelines whenever you relocate a populated rack cabinet within a room or building:

- Reduce the weight of the rack cabinet by removing equipment starting at the top of the rack cabinet. When possible, restore the rack cabinet to the configuration of the rack cabinet as you received it. If this configuration is not known, you must do the following:
 - Remove all devices in the 32U position and above.
 - Ensure that the heaviest devices are installed in the bottom of the rack cabinet.
 - Ensure that there are no empty U-levels between devices installed in the rack cabinet below the 32U level.
 - If the rack cabinet you are relocating is part of a suite of rack cabinets, detach the rack cabinet from the suite.
 - Inspect the route that you plan to take when moving the rack to eliminate potential hazards.
 - Verify that the route that you choose can support the weight of the loaded rack cabinet. Refer to the documentation that came with your rack cabinet for the weight of a loaded rack cabinet.
 - Verify that all door openings are at least 760 x 2030 mm (30 x 80 in.).
 - Ensure that all devices, shelves, drawers, doors, and cables are secure.
 - Ensure that the four leveling pads are raised to their highest position.
 - Ensure that there is no stabilizer bracket installed on the rack cabinet during movement.
 - Do not use a ramp inclined at more than ten degrees.
 - Once the rack cabinet is in the new location, do the following:
 - Lower the four leveling pads.
 - Install stabilizer brackets on the rack cabinet.
 - If you removed any devices from the rack cabinet, repopulate the rack cabinet from the lowest position to the highest position.
 - If a long distance relocation is required, restore the rack cabinet to the configuration of the rack cabinet as you received it. Pack the rack cabinet in the original packaging material, or equivalent. Also, lower the leveling pads to raise the casters off of the pallet and bolt the rack cabinet to the pallet.

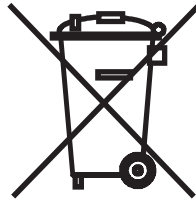
For additional information, refer to the documentation for your rack cabinet.

Product recycling and disposal

This unit must be recycled or discarded according to applicable local and national regulations. IBM encourages owners of information technology (IT) equipment to responsibly recycle their equipment when it is no longer needed. IBM offers a variety of product return programs and services in several countries to assist equipment owners in recycling their IT products. Information on IBM product recycling offerings can be found on IBM's Internet site at: www.ibm.com/ibm/environment/products/index.shtml

Esta unidad debe reciclarse o desecharse de acuerdo con lo establecido en la normativa nacional o local aplicable. IBM recomienda a los propietarios de equipos de tecnología de la información (TI) que reciclen responsablemente sus equipos cuando éstos ya no les sean útiles. IBM dispone de una serie de programas y servicios de devolución de productos en varios países, a fin de ayudar a los propietarios de equipos a reciclar sus productos de TI. Se puede encontrar información sobre las ofertas de reciclado de productos de IBM en el sitio web de IBM

www.ibm.com/ibm/environment/products/index.shtml.



Notice: This mark applies only to countries within the European Union (EU) and Norway.

Appliances are labelled in accordance with European Directive 2002/96/EC concerning waste electrical and electronic equipment (WEEE). The Directive determines the framework for the return and recycling of used appliances as applicable throughout the European Union. This label is applied to various products to indicate that the product is not to be thrown away, but rather reclaimed upon end of life per this Directive.

Remarque: Cette marque s'applique uniquement aux pays de l'Union Européenne et à la Norvège.

L'étiquette du système respecte la Directive européenne 2002/96/EC en matière de Déchets des Equipements Electriques et Electroniques (DEEE), qui détermine les dispositions de retour et de recyclage applicables aux systèmes utilisés à travers l'Union européenne. Conformément à la directive, ladite étiquette précise que le produit sur lequel elle est apposée ne doit pas être jeté mais être récupéré en fin de vie.

注意: このマークは EU 諸国およびノルウェーにおいてのみ適用されます。

この機器には、EU 諸国に対する廃電気電子機器指令 2002/96/EC(WEEE) のラベルが貼られています。この指令は、EU 諸国に適用する使用済み機器の回収とリサイクルの骨子を定めています。このラベルは、使用済みになった時に指令に従って適正な処理をする必要があることを知らせるために種々の製品に貼られています。

In accordance with the European WEEE Directive, electrical and electronic equipment (EEE) is to be collected separately and to be reused, recycled, or recovered at end of life. Users of EEE with the WEEE marking per Annex IV of the WEEE Directive, as shown above, must not dispose of end of life EEE as unsorted municipal waste, but use the collection framework available to customers for the return, recycling and recovery of WEEE. Customer participation is important to minimize any potential effects of EEE on the environment and human health due to the potential presence of hazardous substances in EEE. For proper collection and treatment, contact your local IBM representative.

Battery return program

This product may contain a sealed lead acid, nickel cadmium, nickel metal hydride, lithium, or lithium ion battery. Consult your user manual or service manual for specific battery information. The battery must be recycled or disposed of properly. Recycling facilities may not be available in your area. For information on disposal of batteries outside the United States, go to www.ibm.com/ibm/environment/products/batteryrecycle.shtml or contact your local waste disposal facility.

In the United States, IBM has established a return process for reuse, recycling, or proper disposal of used IBM sealed lead acid, nickel cadmium, nickel metal hydride, and other battery packs from IBM Equipment. For information on proper disposal of these batteries, contact IBM at 1-800-426-4333. Please have the IBM part number listed on the battery available prior to your call.

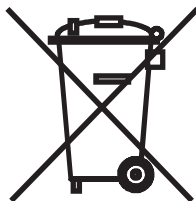
For Taiwan:



Please recycle batteries.

廢電池請回收

For the European Union:



Notice: This mark applies only to countries within the European Union (EU).

Batteries or packaging for batteries are labeled in accordance with European Directive 2006/66/EC concerning batteries and accumulators and waste batteries and accumulators. The Directive determines the framework for the return and recycling of used batteries and accumulators as applicable throughout the European Union. This label is applied to various batteries to indicate that the battery is not to be thrown away, but rather reclaimed upon end of life per this Directive.

Les batteries ou emballages pour batteries sont étiquetés conformément aux directives européennes 2006/66/EC, norme relative aux batteries et accumulateurs en usage et aux batteries et accumulateurs usés. Les directives déterminent la marche à suivre en vigueur dans l'Union Européenne pour le retour et le recyclage

des batteries et accumulateurs usés. Cette étiquette est appliquée sur diverses batteries pour indiquer que la batterie ne doit pas être mise au rebut mais plutôt récupérée en fin de cycle de vie selon cette norme.

注意: このマークは EU 諸国およびノルウェーにおいてのみ適用されます。

この機器には、EU 諸国に対する廃電気電子機器指令 2002/96/EC(WEEE) のラベルが貼られています。この指令は、EU 諸国に適用する使用済み機器の回収とリサイクルの骨子を定めています。このラベルは、使用済みになった時に指令に従って適正な処理をする必要があることを知らせるために種々の製品に貼られています。

In accordance with the European Directive 2006/66/EC, batteries and accumulators are labeled to indicate that they are to be collected separately and recycled at end of life. The label on the battery may also include a chemical symbol for the metal concerned in the battery (Pb for lead, Hg for mercury and Cd for cadmium). Users of batteries and accumulators must not dispose of batteries and accumulators as unsorted municipal waste, but use the collection framework available to customers for the return, recycling and treatment of batteries and accumulators. Customer participation is important to minimize any potential effects of batteries and accumulators on the environment and human health due to the potential presence of hazardous substances. For proper collection and treatment, contact your local IBM representative.

This notice is provided in accordance with Royal Decree 106/2008 of Spain: The retail price of batteries, accumulators and power cells includes the cost of the environmental management of their waste.

For California:

Perchlorate Material - special handling may apply. See www.dtsc.ca.gov/hazardouswaste/perchlorate.

The foregoing notice is provided in accordance with California Code of Regulations Title 22, Division 4.5 Chapter 33. Best Management Practices for Perchlorate Materials. This product, part or both may include a lithium manganese dioxide battery which contains a perchlorate substance.

Fire suppression systems

A fire suppression system is the responsibility of the customer. The customer's own insurance underwriter, local fire marshal, or a local building inspector, or both, should be consulted in selecting a fire suppression system that provides the correct level of coverage and protection. IBM designs and manufactures equipment to internal and external standards that require certain environments for reliable operation. Because IBM does not test any equipment for compatibility with fire suppression systems, IBM does not make compatibility claims of any kind nor does IBM provide recommendations on fire suppression systems.

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.*

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

IBM World Trade Asia Corporation
Licensing
2-31 Roppongi 3-chome, Minato-ku
Tokyo 106-0032, Japan

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 PUBLICATIONS "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 publications. IBM may make improvements and/or changes in the product(s) and/or 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.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
Almaden Research
650 Harry Road
Bldg 80, D3-304, Department 277
San Jose, CA 95120-6099
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

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 is for planning purposes only. The information herein is subject to change before the products described become available.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Notices

This document is confidential and proprietary to IBM XIV.

Contents of this and any related documentation may not be copied or transmitted in any printed or electronic manner or format, or in any other way disclosed to others, for any purpose other than that for which it is furnished, without the prior written consent of IBM Ltd.

Brand names and product names mentioned in this manual may be trademarks or registered trademarks of their respective companies and are hereby acknowledged.

2009 IBM. All rights reserved. IBM XIV, the XIV logo, and Storage Reinvented are trademarks of IBM XIV. All other trademarks are the property of their respective owners.

June, 2009 Document Revision 10.1.0

IBM Information Systems Corporate Headquarters:

USA

XIV Inc.

175 Crossing Blvd., Suite 400

Framingham, MA 01702

Tel: +1-800-4700-XIV (+1-800-470-0948)

Fax: +972-3-6959749

Israel

XIV Ltd.

1 Azrieli Center

Tel Aviv 67021

Tel: +972-3-6074672

info@xivstorage.com

www.xivstorage.com

Copyrights

© Copyright IBM Corporation 2010. US Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

References in this documentation to IBM products, programs, or services do not imply that IBM intends to make these available in all countries in which IBM operates. Any reference to an IBM product, program or service is not intended to state or imply that only IBM's product, program or service may be used. Any functionally equivalent product, program or service that does not infringe any of IBM's intellectual property rights may be used instead of the IBM product, program or service. Evaluation and verification of operation in conjunction with other products, except those expressly designated by IBM, are the user's responsibility.

Trademarks

IBM, the IBM logo, and ibm.com® are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at Copyright and trademark information (www.ibm.com/legal/copytrade.shtml).

Microsoft is a trademark of Microsoft Corporation in the United States, other countries, or both.

Electronic emission notices

The following statements apply to this product. The statements for other products intended for use with this product will appear in their accompanying manuals.

Federal Communications Commission (FCC) Class A Statement

Note: This equipment has been tested and found to comply with the limits for a Class A digital device, pursuant to Part 15 of the FCC Rules. These limits are designed to provide reasonable protection against harmful interference when the equipment is operated in a commercial environment. This equipment generates, uses, and can radiate radio frequency energy and, if not installed and used in accordance with the instruction manual, may cause harmful interference to radio communications. Operation of this equipment in a residential area is likely to cause harmful interference, in which case the user will be required to correct the interference at his own expense.

Properly shielded and grounded cables and connectors must be used in order to meet FCC emission limits. IBM is not responsible for any radio or television interference caused by using other than recommended cables and connectors or by unauthorized changes or modifications to this equipment. Unauthorized changes or modifications could void the user's authority to operate the equipment.

This device complies with Part 15 of the FCC Rules. Operation is subject to the following two conditions: (1) this device may not cause harmful interference, and (2) this device must accept any interference received, including interference that may cause undesired operation.

Industry Canada Class A Emission Compliance Statement

This Class A digital apparatus complies with Canadian ICES-003.

Avis de conformité à la réglementation d'Industrie Canada

Cet appareil numérique de la classe A est conforme à la norme NMB-003 du Canada.

European Union (EU) Electromagnetic Compatibility Directive

This product is in conformity with the protection requirements of EU Council Directive 89/336/EEC on the approximation of the laws of the Member States relating to electromagnetic compatibility. IBM cannot accept responsibility for any failure to satisfy the protection requirements resulting from a non-recommended modification of the product, including the fitting of non-IBM option cards.

This product has been tested and found to comply with the limits for Class A Information Technology Equipment according to European Standard EN 55022. The limits for Class A equipment were derived for commercial and industrial environments to provide reasonable protection against interference with licensed communication equipment.

Attention: This is a Class A product. In a domestic environment this product may cause radio interference in which case the user may be required to take adequate measures.

Properly shielded and grounded cables and connectors must be used in order to reduce the potential for causing interference to radio and TV communications and to other electrical or electronic equipment. Such cables and connectors are available from IBM authorized dealers. IBM cannot accept responsibility for any interference caused by using other than recommended cables and connectors.

European Community contact:

IBM Technical Regulations
Pascalstr. 100, Stuttgart, Germany 70569
Tele: 0049 (0)711 7851176
Fax: 0049 (0)711 785 1283
e-mail: tjahn@de.ibm.com

Australia and New Zealand Class A statement

Attention: This is a Class A product. In a domestic environment this product may cause radio interference in which case the user may be required to take adequate measures.

Germany Electromagnetic Compatibility Directive

Deutschsprachiger EU Hinweis:

Hinweis für Geräte der Klasse A EU-Richtlinie zur Elektromagnetischen Verträglichkeit

Dieses Produkt entspricht den Schutzanforderungen der EU-Richtlinie 2004/108/EG zur Angleichung der Rechtsvorschriften über die elektromagnetische Verträglichkeit in den EU-Mitgliedsstaaten und hält die Grenzwerte der EN 55022 Klasse A ein.

Um dieses sicherzustellen, sind die Geräte wie in den Handbüchern beschrieben zu installieren und zu betreiben. Des Weiteren dürfen auch nur von der IBM empfohlene Kabel angeschlossen werden. IBM übernimmt keine Verantwortung für die Einhaltung der Schutzanforderungen, wenn das Produkt ohne Zustimmung der IBM verändert bzw. wenn Erweiterungskomponenten von Fremdherstellern ohne Empfehlung der IBM gesteckt/eingebaut werden.

EN 55022 Klasse A Geräte müssen mit folgendem Warnhinweis versehen werden:

"Warnung: Dieses ist eine Einrichtung der Klasse A. Diese Einrichtung kann im Wohnbereich Funk-Störungen verursachen; in diesem Fall kann vom Betreiber verlangt werden, angemessene Maßnahmen zu ergreifen und dafür aufzukommen."

Deutschland: Einhaltung des Gesetzes über die elektromagnetische Verträglichkeit von Geräten

Dieses Produkt entspricht dem "Gesetz über die elektromagnetische Verträglichkeit von Geräten (EMVG)." Dies ist die Umsetzung der EU-Richtlinie 2004/108/EG in der Bundesrepublik Deutschland.

Zulassungsbescheinigung laut dem Deutschen Gesetz über die elektromagnetische Verträglichkeit von Geräten (EMVG) (bzw. der EMC EG Richtlinie 2004/108/EG) für Geräte der Klasse A

Dieses Gerät ist berechtigt, in Übereinstimmung mit dem Deutschen EMVG das EG-Konformitätszeichen - CE - zu führen.

Verantwortlich für die Konformitätserklärung des EMVG ist die IBM Deutschland GmbH, 70548 Stuttgart.

Generelle Informationen:

Das Gerät erfüllt die Schutzanforderungen nach EN 55024 und EN 55022 Klasse A.

People's Republic of China Class A Electronic Emission Statement

中华人民共和国“A类”警告声明

声明

此为A级产品，在生活环境中，该产品可能会造成无线电干扰。在这种情况下，可能需要用户对其干扰采取切实可行的措施。

Taiwan Class A warning statement

警告使用者：
這是甲類的資訊產品，在居住的環境中使用時，可能會造成射頻干擾，在這種情況下，使用者會被要求採取某些適當的對策。

Japan VCCI Class A ITE Electronic Emission Statement

この装置は、情報処理装置等電波障害自主規制協議会（VCCI）の基準に基づくクラスA情報技術装置です。この装置を家庭環境で使用すると電波妨害を引き起こすことがあります。この場合には使用者が適切な対策を講ずるよう要求されることがあります。

Korean Class A Electronic Emission Statement

이기는 업무용으로 전자파 적합등록을 받은 기기
이오니, 판매자 또는 사용자는 이점을 주의하시기
바라며, 만약 잘못 구입하셨을 때에는 구입한 곳에
서 비업무용으로 교환하시기 바랍니다.

Index

A

- about this document
 - how to send your comments x
- access control 119
 - command execution log 130
 - commands 130
 - event report destinations 130
 - logging and event reporting 130
 - object creation tracking 130
- access_all 122, 123
- acknowledgement 70
- active
 - activation state 90, 91
- Active Directory 127
- Active mode 50
- adding ports to remote target 44
- address, IBM x
- administering
 - access control 119
- advanced host attachment 24
- advanced snapshot mechanism 14
- alarm notification 7
- algorithms 5, 7
- application administrator 119, 121
 - access_all 122
 - command conditions 122
- associations
 - user groups and hosts 121
- asynchronous
 - remote mirroring 84, 85, 86, 87, 88, 89, 90
- asynchronous mirroring 47
 - scheduling 72
 - scope of the chapter 66
- asynchronous mirroring process
 - walkthrough 84
- asynchronous remote mirroring
 - peer roles 90
 - role transmission 74, 90, 91, 95
 - state machine 90
 - XCLI 101
- Atomic test & set 29
- ATS 29
- attention notice 148
- authentication 122
 - xiv 123
- authentication modes
 - switching 129
- auto delete priority 15
- automatic
 - recovery from failure 5
- automatic event notifications 7
- available consistent copy
 - during mirroring 79

B

- backup
 - continuous 14, 16
- backups 49

- bandwidth 2
- bandwidth utilization 23
- battery return 152
- Block zeroing 28
- Broadband connection 3

C

- cache
 - protection 5
- cache memory 2
- cache module 2
- caching 6
- call home
 - contact list 30
- Call-home connection 3
- canceled
 - sync job 69, 79, 82
- caution notices 147
 - definition 147
 - examples 147
- CDP (continuous data protection) 16
- change tracking 82
- CHAP 24
- Class A electronic emission notice 158
- CLI
 - management options 4
- CLI (command line interface) 7
- CLI management 107
- clustering
 - hosts 25
- command execution log 130
- commands
 - fc_connectivity_list 44
 - host attachment 27, 31
 - ipinterface_run_traceroute 44
 - target_connectivity_activate 44
 - target_connectivity_deactivate 44
 - target_connectivity_define 44
 - target_connectivity_delete 44
 - target_connectivity_list 44
- comments, how to send x
- component
 - redundancy 5
- components, hardware 2
- configuration 126
 - multi-rack 7
- configuration guidelines summary 109
- configured sync rate 23
- connectivity 23, 24
 - hardware 2
- consistency group 11, 36
 - creating 33
- consistency groups 7
 - and remote mirroring 62
 - mirroring 78
 - mirroring-related tasks 78
 - overview 33
 - restore 36
 - restoring 33
 - snapshots 34, 35

- contact list
 - call home 30
- continuous backup 14, 16
- continuous data protection 14
- Copy-on-Write (COW) 14, 16
- coupling 67
 - activation 50
 - constraints and limitations 55
 - definition 49
 - last consistent snapshot
 - timestamp 57
 - last-consistent snapshots 56
 - modes 50
 - states 54, 90, 91
 - types 50
 - uncommitted data 55
- COW (copy-on-write) 14
- COW (Copy-on-Write) 14, 16
- creating
 - consistency group 33
- creating a vm 28

D

- danger notices 145
 - definition 145
 - example 145
- Data (Cache) modules 2
- data differences 72
- data migration 43
 - deleting 112
 - failures 114
 - I/O handling 112
 - overview 111
 - read requests 112
 - stages
 - activating 112
 - initial configuration 112
 - synchronization 112
 - testing 112
 - write requests 112
- Data mirroring 5
- Data module 2
- data replication 72
- data virtualization 5, 7
- deactivating the mirroring 91
- defining gateways 117
- deleting snapshots
 - as part of the pool depletion
 - process 69, 79, 82
- deleting the mirroring 91
- destination groups 117
- destinations
 - defining 117
 - e-mail 117
 - SMS 117
- diagnostics 7
- diff
 - the data that is sent by the sync job 72
- differences calculation 70

- disaster recovery 47, 65, 95
- disaster recovery types 47, 48
- disconnects prevention 23
- disposal 150
- Don't stop when found
 - a stop criteria 127
- dr
 - disaster recovery 95
- Dynamic rate adaptation 23

E

- e-mail (SMTP) gateways 117
- e-mail destination 117
- e-mail notifications 4
- e-mail-to-SMS gateways 117
- ECLS 56
- ECLS x 56
- electronic emission notices
 - Australia and New Zealand Class A statement 159
 - European Union EMC Directive conformance statement 158
 - Federal Communications Commission (FCC) statement 158
 - Industry Canada Class A emission compliance statement 158
 - Japanese Voluntary Control Council for Interference (VCCI) statement 160
 - Korean Class A warning statement 161
 - People's Republic of China Class A warning statement 160
 - Taiwanese Class A warning statement 160
- environmental
 - notices 145
- error
 - link state 90, 91
- error code protection 5
- ESX
 - COMPARE AND WRITE 29
 - fast copy 29
 - SCSI2 reservations mechanism 29
 - write zeroes 28
- Ethernet connectivity 105
- Ethernet ports 105
 - field technician ports 105
 - iSCSI service ports 105
 - management ports 105
- event
 - handling 115
 - information 115
 - notification rules 116
 - viewing 116
- event notifications 7
- event report destinations 130
- external connection congestion 23
- external last consistent snapshot 56
- external replication mechanisms 7

F

- failback 95, 96, 98
- failover 95, 96, 98

- failover process 5
- failover test 100
- fast copy 29
- fc_connectivity_list 44
- FCC Class A notice 158
- features and functionality 1
- fibre-channel connectivity 105
- field technician ports
 - Ethernet ports 108
 - laptop connectivity
 - configuring using DHCP 108
 - configuring without DHCP 108
- fire suppression 153
- form, reader comment x
- format
 - snapshot and snapshot group 21
- forums v
- Full copy 28
- Full Volume Copy 20

G

- gateways
 - defining 117
 - e-mail (SMTP) 117
 - e-mail-to-SMS 117
- Gigabit Ethernet 3
- global spare storage 5
- glossary 139
- group rate limitation 29
- groups, destination 117
- GUI
 - management options 4
 - updating the call home contact list 30
- GUI (graphic user interface) 7
- GUI management 107
- gui/cli initiated LDAP login 126

H

- hard capacity, depletion 39
- hardware components 2
- hardware connectivity 2
- Hardware-assisted locking 28, 29
- HBA 23
- host
 - clustering 25
 - communication with the slave volume 96
 - rate 29
- host connectivity 23, 24
- Host Interface module 2
- host system attachment 23, 24
- host-based I/O 3
- hosts
 - associations 121
- hot upgrade 137
- how to send your comments x

I

- I/O 23
 - rate limitation 29
- I/O operations 53

IBM

- address x
- IBM XIV role
 - in relation to access control 124
- image snapshots
 - duplicating 18
- implementation
 - of LDAP 124
- indirectly associated groups
 - of LDAP users 127
- init
 - to be distinguished from initialization 82
- initialization
 - synchronization status 52
 - the first phase of asynchronous mirroring 84, 85, 86, 87, 88, 89, 90
- Initialization 70
- initialization type 70, 71
- initiator
 - iSCSI 24
- installation
 - rack 149
- instant space reclamation 13
- intelligent caching 6
- interfaces
 - call home 30
 - supported 3
- internal snapshots 21
- IP communication, system-initiated 107
- IP connectivity 105
- IP connectivity, guidelines summary 109
- ipinterface_run_traceroute 44
- iSCSI CHAP authentication 24
- iSCSI connectivity 105
- iSCSI ports
 - definition criteria 105
 - functionalities 105

L

- labels, safety 147
- laser safety 148
- last consistent snapshot 56
- last replicated snapshot
 - deletion priority 79
- last secondary timestamp 53
- last_replicated
 - deletion 69, 79, 82
 - snapshot 70
 - timestamp 75
- last_replicated snapshot 96
- latency
 - overcoming latency that is inherent to synchronous mirroring 65, 66
- LCS 56
- LDAP 132
 - authentication 122, 124
 - authentication mode
 - switch to and from 129
 - authentication scenarios 127
 - directory 124
 - group mapping 124
 - service account 129
 - use cases 125, 126
 - user validation 127
- LDAP authentication scenarios 126

- LDAP server
 - definition 125
- ldap_test 125
- life-cycle
 - of a volume 12
- Lightweight Directory Access Protocol 124
- limiting
 - host rates 29
- link
 - states 90, 91
- link status
 - synchronous mirroring statuses 52
- load balancing 6
- Locking
 - of the pool during pool space depletion 79
- logical unit numbers 23
- low sync rate 23
- LUN ID 21

M

- management connectivity 107
- management options 4
- managers, SNMP 117
- manual deletion
 - of last consistent snapshot 56
- manual reactivation
 - of the mirror process 82
- map_vol 96
- mapping
 - mirror related volumes 96
- mapping, LUN, 23
- master 67
 - role type 90
- master LRS 79
- master site 47
- master volumes 12
- max sync rate 23
- maximum number of queries
 - as a search limit 127
- mechanisms
 - self-healing 5
- memberof
 - group attribute 127
- methods
 - of access control 119
- Microsoft Active Directory 124
- mirror state 74
- mirror_activate 96
- mirror_change_role 96
- mirror_change_schedule 96
- mirror_create 70, 71, 96
- mirror_delete 96
- mirror_switch_roles 96
- mirroring
 - data 5
 - remote 47, 65, 84, 85, 86, 87, 88, 89, 90
 - scheduling asynchronous mirroring 72
 - termination, deletion, deactivation 91
- mirroring and snapshots 68, 70
- mirroring consistency groups 78, 92, 94, 95
- mirroring relationships 69

- mirroring resynchronization 56
- Modem 3
- modes
 - Active 50
 - Standby 50
- module
 - data 2
 - Host Interface 2
 - UPS 2
- modules
 - cache 5
- most_recent 70, 71
 - deletion 69, 79, 82
 - snapshot 70
- multi-rack configuration 7
- multipathing 7
- Multiple concurrent mirroring relationships 69
- Multiple findings
 - during a search within indirectly associated groups 127
- multiple targets 69

N

- native
 - authentication 122
- never
 - a default schedule 72
- Never
 - schedule 96
- Non-disruptive code load 137
- Nondisruptive testing 100
- none
 - role type 90
- nonoperational
 - coupling state 90, 91
- nonvolatile disk media 5
- notices 155
 - attention 148
 - caution 147
 - danger 145
 - electronic emission 158
 - FCC, Class A 158
 - safety and environmental 145
- notifications
 - e-mail 4
 - SMS 4
 - SNMP 4

O

- object creation tracking 130
- of an LDAP server 126
- of the initialization 70
- off-line initialization 70, 71
- OK
 - link state 90, 91
- Open LDAP 124
- operational
 - coupling state 90, 91
- operational status
 - synchronous mirroring statuses 52
- optical port terminators 148
- options
 - management 4

- outstanding sync jobs
 - disruption 96

P

- partial rack
 - patch panel layout 105
- password management 123
- patch panel 105
- PDFs v
- peers 67
- pending writes 52
- Performance classes
 - (QoS) 29
- planned service disruption 98
- Planned service disruption 95
- point of failure 5
- pool space depletion 79, 82
 - on the slave 84
- protected snapshots 79
- pool_config_snapshots 79
- predefined users 120
 - authentication 122
- protected snapshots 79
- protected_snapshot_priority 79
- provisioning
 - thin 7
- provisioning, thin 39
- publications v

Q

- QoS
 - performance classes 29

R

- rack installation
 - safety 149
- rack relocation 150
 - safety 150
- rack safety 149
- reader comment form processing x
- recovery 96, 98
 - from a disaster 95
- recovery point objective 74
- recovery target objective 74
- recycling 150
- redirect-on-write
 - in the context of mirroring 68
- Redirect-on-Write (ROW) 14, 16
- redundancy 5
- redundant components 5
- redundant power 5
- related information v
- reliability 5
- remote mirroring 43, 47
 - and consistency groups 62
 - basic concepts 47, 48
 - best-effort coupling recovery 55
 - configuration options 49
 - coupling 50
 - disaster recovery types 47, 48
 - for media error recovery 62
 - operation 47, 48
 - resuming after role change 60

- remote mirroring (*continued*)
 - role switchover 58
 - synchronization 52, 54
 - synchronous mirroring statuses 51
 - volume configuration 49
- remote monitoring 7
- replication 65, 66
- replication mechanisms 7
- replication scheme 69
- replication state 74
- restoring 36
 - snapshots 18
 - volumes 18
- restricted prefix
 - to a snapshot group 35
- restrictions, usage 148
- resynchronization 56
- role switchover 58
 - when remote mirroring is nonoperational 58
 - when remote mirroring is operational 58
- role transmission
 - within the asynchronous mirroring process 74, 90, 91, 95
- role-based
 - access control 119
- role-based access control
 - application administrator 121
 - configuring users 123
- roles
 - within the asynchronous mirroring process 90
- ROW (redirect-on-write) 14
- ROW (Redirect-on-Write) 14, 16
- rpo 74
- RPO 74
- RPO_Lagging 75, 76, 91
- RPO_OK 75, 76
- RTO 74

S

- safety
 - environmental notices 145
 - laser 148
 - rack 149
 - rack installation 149
 - rack relocation 150
- safety labels 147
- scheduling
 - asynchronous mirroring 72
- scrubbing 5
- SCSI error
 - while writing to a secondary volume 53
- search flow
 - of indirectly-associated groups 127
- secondary site 47
- secondary volume
 - last consistent 56
- self-healing
 - mechanism 5
- self-healing mechanisms 5
- semantics
 - of mirrored consistency groups 78

- service disruption
 - testing 100
- single physical copy of data 17
- single point of failure 5
- slave 67
 - promoting during failover 96
 - role type 90
- slave LRS 79
- slave replica
 - inconsistent 75
- smis_user 120
- SMS destination 117
- SMS notifications 4
- snapshot 17, 18
 - atomic procedure of creating a 17
 - format 21
 - storage utilization 15
- Snapshot
 - association 16
 - name 16
 - serial number 16
- snapshot architecture
 - in the context of mirroring 68
- snapshot group
 - format 21
- snapshot groups 34, 35, 36
- snapshot ID 21
- snapshot management 7
- snapshots 11, 14, 16
 - auto delete priority 15
 - depletion of hard capacity 39
 - duplicating 18
 - locking and unlocking 17
 - protected 79
 - restoring 18
 - taken in a point-in-time 11
 - unprotected (from deletion) 82
- Snapshots 14
- snapshotting 14, 16
 - consistency groups 7
 - snapshot management 7
- Snapshotting 14
- SNMP 7
- SNMP managers 117
- SNMP notifications 4
- spare storage 5
- standby
 - activation state 90, 91
- Standby mode 50
- states
 - coupling 54
- stop criteria
 - for searching indirectly associated groups 127
- Stop when found
 - a stop criteria 127
- storage
 - global spare 5
- storage administrator 23, 120
- storage pool 11
 - depletion of hard capacity 39
 - hard and soft sizes 39
- storage pools 7
 - mirroring 79
- moving volumes 37, 38
- overview 37, 38
- SUN directory 124

- switchover 58
- Symantec Storage Foundation Thin Reclamation 13
- symmetric connectivity for mirroring 45
- sync job 72
 - canceling 69, 79, 82
 - pausing 91
 - pending 82
 - progress 75
 - snapshot that is part of a 21
- sync rate
 - low 23
- synchronization 53
- synchronization status
 - how is it determined 75, 76
 - synchronous mirroring statuses 52
- synchronized
 - remote mirroring 47, 65, 66
- synchronous mirroring 47
 - statuses 51
- synchronous mirroring statuses 53
- synchronous remote mirroring
 - I/O operations 53
- system
 - hard and soft sizes 39
- system attachment
 - see: host system attachment 23, 24

T

- target connectivity 43, 44
- Target connectivity 45
- target object
 - defining 43
- target_connectivity_activate 44
- target_connectivity_deactivate 44
- target_connectivity_define 44
- target_connectivity_delete 44
- target_connectivity_list 44
- technician 120
- terminating the mirroring 91
- terminators
 - optical ports 148
- the unmap bit 28
- thin provisioning 7, 39
 - mirroring 79
- time stamp 53
- timestamp of last_replicated snapshot 96
- transmission
 - of roles 74, 90, 91, 95
- truck mode 70, 71
- type
 - of initialization 70, 71

U

- uncommitted data 55
- Unintentional/erroneous role change 95
- United States electronic emission Class A notice 158
- United States FCC Class A notice 158
- unplanned service disruption 96
- Unplanned service disruption 95
- unprotected snapshots
 - deletion 82
- upgradability 8

- UPS module 2
- usage restrictions 148
- use cases
 - LDAP 125, 126
- user groups 121
 - associations 121
- user roles
 - application administrator 119
 - permission levels 119
 - read only 119
 - storage administrator 119
 - technician 119
- user search
 - LDAP 127
- users validation
 - using LDAP 127

V

- vm cloning 29
- volume
 - configuration 49
- volume configuration
 - mixed configuration 50
- volumes 11, 12
 - Full Volume Copy 20
 - hard and soft sizes 39
 - restoring 18

W

- write zeroes 28

X

- XCLI
 - asynchronous mirroring 101
- xiv authentication 123
- XIV GUI
 - updating the call home contact list 30
- XIV mapping 132
- XIV role search
 - LDAP 127
- XIV-to-LDAP mapping 126

Readers' Comments — We'd Like to Hear from You

IBM XIV Storage System
Product overview

Publication No. GA32-0791-03

We appreciate your comments about this publication. Please comment on specific errors or omissions, accuracy, organization, subject matter, or completeness of this book. The comments you send should pertain to only the information in this manual or product and the way in which the information is presented.

For technical questions and information about products and prices, please contact your IBM branch office, your IBM business partner, or your authorized remarketer.

When you send comments to IBM, you grant IBM a nonexclusive right to use or distribute your comments in any way it believes appropriate without incurring any obligation to you. IBM or any other organizations will only use the personal information that you supply to contact you about the issues that you state on this form.

Comments:

Thank you for your support.

Send your comments to the address on the reverse side of this form.

If you would like a response from IBM, please fill in the following information:

Name

Address

Company or Organization

Phone No.

Email address



Cut or Fold
Along Line

Fold and Tape

Please do not staple

Fold and Tape

PLACE
POSTAGE
STAMP
HERE

International Business Machines

Fold and Tape

Please do not staple

Fold and Tape

Cut or Fold
Along Line



Printed in USA

GA32-0791-03

