HITACHI
Inspire the Next

Best Practice Library

Guidelines for Hitachi TrueCopy™ Remote Replication Software and Hitachi Universal Replicator Software

By Hitachi Data Systems Technical Marketing

February 2007

Hitachi Data Systems

Executive Summary

Whether your business IT operations span the campus, continent or globe, delays in resuming IT operations following an unexpected disruption can have a dramatic impact on bottom-line financial performance and competitive advantage. Hitachi Data Systems provides innovative storage-based technology solutions for enabling continuous business operations for all companies. For Hitachi Data Systems, business continuity means partnering with our customers to understand key business problems and requirements to provide robust application-focused storage solutions that enhance operational efficiency and resilience.

This document provides best practices for tailoring remote data replication solutions based on Hitachi TrueCopy™ Heterogeneous Remote Replication software bundle with TrueCopy Synchronous and TrueCopy Asynchronous software, and Hitachi Universal Replicator software for Universal Storage Platform and Network Storage Controller, including the qualification, design, implementation, operation, and tuning in remote replication environments based on these offerings.

Although the recommendations documented here may generally represent good practices, configurations may vary. Please contact your Hitachi Data Systems representative or authorized dealer, or visit Hitachi Data Systems online at http://www.hds.com for further information on Business Continuity and Disaster Recovery Solutions from Hitachi Data Systems.

Document Revision Level

Revision	Date	Description
1.0	February 2007	Initial Release

Source Documents for this Revision

Hitachi Data Systems Architect – Business Continuity Certification Prep Course

Hitachi Universal Replicator Three Data Center Multi-Target Planning and Proof of Concept *by Roy Strouse*

Audience

This document is intended for storage and system administrators responsible for designing and implementing business continuity solutions based on Hitachi Data Systems storage systems.

Contributors

The information included in this document represents the expertise of a number of skilled practitioners within Hitachi Data Systems in the successful delivery of remote replication solutions to our customers. In addition, a broad range of additional skills and talents working across many disciplines within Hitachi Data Systems provided many hours of effort on quality assurance of this material. Thanks go to:

Douglas E. Babcock	Managing Principal, Business Continuity Mainframe, Global Solutions Services
Christina Gallagher	Account Consultant, Sales
Tom Goulding	Solution Architect, Global Solution Services
John Hickman	Field Product Manager, Enterprise Solutions
Vladimir Igolnikov	Principal Technical Consultant, Global Solution Services
Carl Isenburg	Principal Technical Consultant, Global Solution Services
Gian Jagai	Technical Consultant, Global Solution Services
Bill Martin	Product Manager, Disaster Recovery Program Products, Global Solutions Strategy and Development
Chris Myers	Technical Consultant, Global Solution Services
Keith O’Toole	Technical Consultant, Global Solution Services
Roselinda R Schulman, CBCP	Director, Data Protection, Global Solution Business Line
Leland Sindt	Technical Consultant, Global Solution Services
Roy Strouse	Advanced Technical Consultant, Solutions Sales & Services
Alex Tse	Solutions Service Practice Manager, Global Solutions Services

Delivery of this guide was managed by Technical Marketing, Global Solutions Strategy and Development.

Contact Information

All documents in this series can be found on www.hds.com.

Please send comments on this document. Make sure to include the document title, and revision. Please refer to specific section(s) and paragraph(s) whenever possible.

E-mail: TM@hds.com

Fax: 408-562-5477

Mail:
Technical Marketing, M/S 34-88
Hitachi Data Systems
750 Central Expressway
Santa Clara, CA 95050-2627

Thank you! (All comments become the property of Hitachi Data Systems Corporation.)

Contents

Best Practices Library

Guidelines for Hitachi TrueCopy™ Remote Replication Software and Hitachi Universal Replicator Software

By Hitachi Data Systems Technical Marketing

Introduction

The world has changed significantly in the past few years. Devastating terrorist acts and threats, the seemingly increased frequency of widespread power-grid disruptions, and the emergence of regulatory requirements for infrastructure protection are all placing stringent, yet necessary, data protection requirements on many organizations. Regardless of the industry, as more and more businesses operate in a 24/7 environment—especially large enterprises where global operations are the norm—they need an increasingly competitive edge to maintain profitability and stay in business.

In the complex and challenging global environment, well-planned business continuity or proven disaster recovery practices for nonstop data availability have become critical to organizations if they are to survive any type of outage.

Problem Description

Most information technology related disruptions are actually locally contained events, such as data corruption, viruses and human error, as opposed to physical disasters like fire, earthquakes, hurricanes, etc.These events occur frequently and pose a more common threat to businesses than physical disasters. However, because they are less visible to the general public, these disruptions may be taken less seriously.

The real challenge facing IT management lies in getting the organization to think proactively and to deploy best practices and technologies that can be leveraged to maximize business operations instead of adopting a reactive “fix-it” posture. The true test of sound IT infrastructure is the ability to prevent outages from occurring in the first place and minimizing the effects of those incidents when they do occur. Companies today must follow the continuous business paradigm which combines high-availability solutions with advanced disaster recovery techniques. The ultimate goal is to be able to manage both planned and unplanned situations with minimal or zero disruption.

When an unplanned event does occur, the ideal scenario is:

: : Recovery happens almost automatically with no loss of data

: : Costs of the solution and resources are minimal

: : Impact to the production environment is zero

While technology is moving forward at a rapid pace to reach this ideal scenario, many other business and technology concerns exist, including some significant trade-offs dictated by technology, budgets and personnel resources.

Technical Issues and Challenges

Options for replicating data for business continuity generally fall into one of two categories: synchronous (real-time) or asynchronous (near real-time). Business managers must consider distance at which replication takes place versus the possible impacts to application performance. They also must evaluate the critical replication issues posed by single event or rolling disasters, including latency, sequence—or write order—fidelity, and data consistency. Using the results of this analysis, business managers can select and implement the right business continuity and recovery solution for their organization's needs.

Synchronous and Asynchronous Replication

Synchronous replication ensures that a remote copy of the data is identical to the primary copy at the time the primary copy is created or updated. In synchronous replication, an I/O update operation is not considered done until completion is confirmed at both the primary and mirrored sites. If the operation fails to complete at the remote site, actions taken will be determined by replication and other software settings i.e. clustering, path failover.

One benefit of synchronous replication is that data can be recovered quickly. After a disruption at the primary site, business operations at the remote site can begin immediately with a consistent copy of the data. Only I/Os in-flight at the instant of disruption may be lost. Because neither the primary nor remote site will have a record of those transactions, the business processing rolls back to the last commonly confirmed state.

The drawback to synchronous replication is its distance limitation. Fibre Channel, the primary enterprise storage transport protocol, can theoretically extend over several hundred kilometers. However, latency quickly becomes an application problem as propagation delays lengthen with increased distance. Propagation delays can significantly slow down an application by forcing it to wait for confirmation of each storage write operation.This means the practical distance for synchronous replication on a busy system depends on the application response time tolerance and other factors, which typically ranges from 20 to 100 miles (about 35 to 160 kilometers) —not far enough to be clear of a wide-area or regional disaster.

To mitigate the distance limitations of synchronous methods, asynchronous technologies have been developed, implementing a buffering mechanism to accumulate write operations for subsequent transmission after I/O completion has been acknowledged to the host. By eliminating the wait for a response from the remote site for each I/O, this approach eliminates the propogation delay that hinders synchronous copy techniques.

The main benefit of asynchronous replication is the ability to have the secondary storage system at long distances from the primary storage system without impacting the application at the primary site. Implementations of this replication strategy can extend to thousands of kilometers.

The downside to asynchronous replication is the potential for data loss between the primary and remote sites. Because of the slight time lag between data being stored at the primary and remote sites, updates lost in-flight during an outage can mean the remote center cannot pick up operations instantly at the point the primary site failed. In such a situation, asynchronous replication caching, sequence numbering, time stamps, and other techniques used to automatically preserve write-order fidelity and data integrity at the remote site are essential.

The Rolling Disaster Challenge

A rolling disaster occurs when an unplanned outage event takes place over a span of time—anywhere from a few minutes to several hours. During a rolling disaster, not all systems, storage and network connections fail at precisely the same moment. In this situation, a system may still be able to process transactions and issue updates to primary storage devices, but due to earlier failures, updates may not replicate successfully to the secondary site. Rolling disasters pose a challenge because they may result in corrupted and unusable data at the remote site, requiring difficult and very lengthy recovery processes.

To protect against rolling disasters, a data replication technology must be able to freeze remote replicas at a point in time prior to or during the onset of the outage. This ability to create point-in-time images of data is what differentiates remote copy technology from simple mirroring.

Because the remote and local I/O of a synchronous replication succeed or fail together, this replication approach does not introduce data inconsistencies following a disaster. Rolling disasters are primarily a challenge for remote asynchronous replication, and one of the principle areas of concern is write order fidelity.

Write Order Fidelity

Database and file managers maintain very complex internal data structures, including indexes, structured data tables, directories, logs and so forth. Database applications should have atomicity (whereby database transactions follow the atomic rule: if one part fails, the whole transaction fails), being asset compliant in writing data to disk. Thus, careful write sequencing and strict adherence to write dependencies allow file systems and databases to preserve the integrity of these internal structures no matter what I/O activity is in progress when failure occurs. Each write is carefully sequenced so that, at any point in time, a correct file system or database state can be recreated.

Resynchronization refers to the process of updating the remote data copy following a planned or unplanned suspension. Traditional remote replication technologies track changes in storage system cache during normal paired operation, building and maintaining a record of changed data in the event of cache overflow during unexpectedly high change rate or a replication link failure. During resynchronisation, data in the remote storage system may not be updated in the same sequence as it was written in the primary storage system; it will therefore not be consistent. If the primary storage system were to fail or if access to the primary storage system was lost during this process, there would not be any consistent data available for the application to continue running. Taking a local copy of the data as part of the recovery process before starting the resynchronisation could alleviate the application data consistency.

A replication solution must address write order fidelity, ensuring that remote writes are made in the same order as those at the primary site. To ensure the integrity of asynchronously replicated data in a rolling disaster, replication technology must employ techniques to automatically preserve write order fidelity at the remote site.

Data Consistency

In the context of data replication, data consistency represents the ability to recover from a failure or disruptive event. A fundamental concept of data consistency that enables quick recovery is the “dependent write”, that pervasive logic among complex data structures comprising databases, file systems, etc. that determines the sequence in which writes are issued. A dependent write is a data update that cannot be executed until a previous write—on which it is dependent—has been executed. It is this logic that preserves the integrity, the consistency, of the data and allows systems and applications to restart after a sudden failure.

There are three types of data consistency that have different implications at different levels within the application and data architecture where the meaning of the data can have different logical dependencies. These are I/O, transaction, and application consistency.

I/O consistency, or crash recovery consistency, refers to data that is not necessarily transaction consistent, but is still in a restartable state. If the writing of data is interrupted in the middle of a transaction and fails to complete, this leaves the resulting data in an I/O consistent state if the sequence of dependent writes haas been maintained—the data is recoverable. When the application is restarted, the data will be rolled back or rolled forward to a transaction consistent state.

A transaction is a logical unit of work that may include hundreds or thousands of updates. Transaction consistency is achieved when an application is shut down (quiesced), or when the application/database or other system component rolls back or rolls forward after a restart. Restarts could result from a sudden power failure, system crash or other disaster.

An application may be made up of many different types of data, such as multiple database componenets as well as flat files. Application consistency is the state in which individual components have each been recovered to a transaction consistent state. Collectively, the components need to be synchronized based on the application requirements.

The difference between the primary site failure and resumed operations at the remote site represents the Recovery Point Objective (RPO – explained later in this document) which will always be kept to a minimum in Hitachi Data Systems asynchronous replication implementations.

Evaluation Criteria

Evaluation of any data replication solution should consider the following:

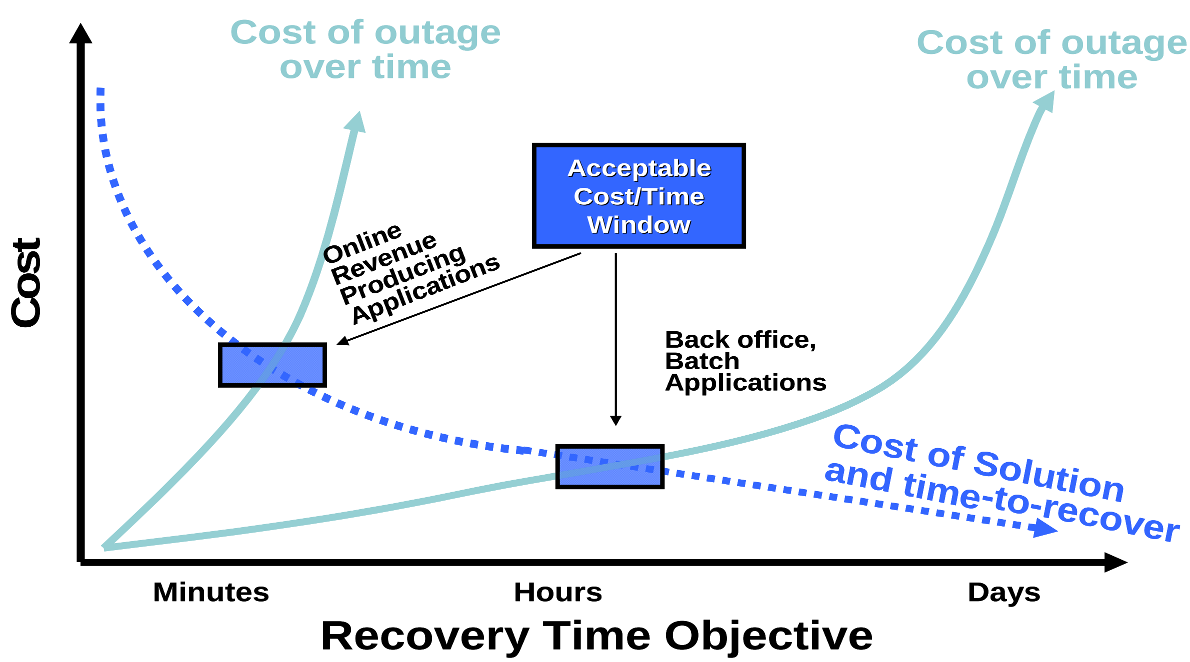
Data Consistency

Does the remote copy technology provide I/O consistency, which is a bedrock requirement to successfully use the replica copies?

Cost versus Benefit

When evaluating the cost of business continuity solutions, the greatest cost component is usually the bandwidth needed to support remote replication. The greatest benefit is maintaining a minimal RPO. This balance is illustrated in Figure 1.

Figure 1: Recovery Time Versus Cost



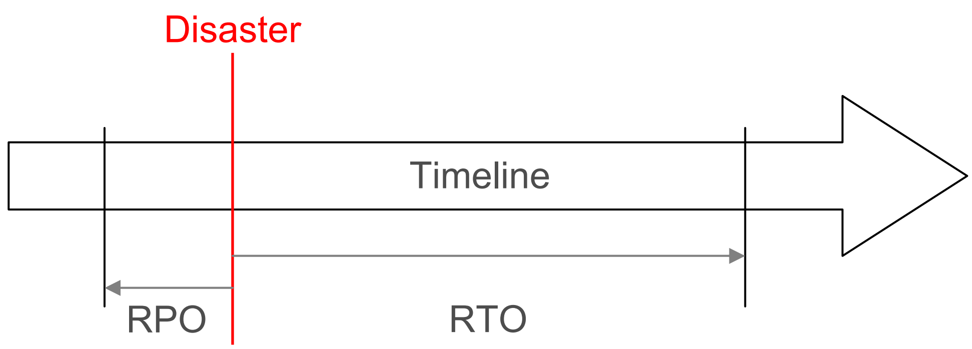
Each application must be evaluated separately to identify the costs versus benefits.

As more is spent on bandwidth, the replica copy can be more up to date. The ideal solution allows IT management to balance costs against benefits and maximize return on investment.

RPO and RTO

Fundamental to developing an effective disaster recovery solution is identifying a business's risk tolerance as expressed in terms of Recovery Point Objectives and Recovery Time Objectives (RPO and RTO, respectively). RPO represents the worst case time between the interruption in operations and the last recoverable backup, where potentially lost data weighs against cost. RTO represents the time to resume operations after the interruption.

Figure 2: Recovery Point and Recovery Time Objectives



Evaluation of risk tolerance is stated in terms of how much data must be recovered to resume operations, the RPO, and the outage duration, the RTO.

Does the remote copy technology allow for minimal RPO? Batch copy techniques usually cost less and often promise lower bandwidth requirements by extending RPO. Is this a false economy? Hitachi technology balances RPO against bandwidth costs while also providing mechanisms for bandwidth management.

Making the Choice

Clearly remote storage replication for recovery and business continuity requires more than just shipping data over a network. The selection process starts with an assessment of the potential risks and their probability.

Two-data-center (2DC) replication strategies are viable for most in-region recovery—for example, serving as a hot site for campus-level or metro-level server cluster—and for out-of-region recovery sites where propagation delays are not an issue.

Synchronous replication provides very fast recovery time (low RTO) and good data currency (low RPO). If your organization cannot tolerate any data loss and operations must be resumed quickly following an outage, synchronous replication is likely to be the best choice. Of course, the decision must also factor in how far the data has to be replicated to clear any likely disaster zone, balanced against how much degradation of application performance can be tolerated.

Asynchronous replication provides better protection against regional disasters, albeit with less favorable RPO. If your organization can tolerate being down while the last few transactions are reconstructed—or cannot tolerate the performance impact of synchronous propagation delays—asynchronous replication may prove to be a less costly option.

When synchronous and asynchronous replication is combined into "three data center" (3DC) replication solutions, maximum protection and flexibility in recovery is achieved. A three data center strategy offers the best of both worlds: fast recovery and excellent data currency for local site failures, combined with advanced protection from regional disasters.

Solution Overview

Consistent with a long history of delivering technologically advanced storage-based solutions for the enterprise, Hitachi Data Systems offers a diverse portfolio of Application Optimized Storage™ solutions for business continuity. While providing the right type of access and availability of data to applications is essential, just as important is the strategy used to protect the data itself. This strategy should be based on business requirements, including mitigating risk, regulatory compliance, and employment of best practices.

For organizations with demanding heterogeneous data replication needs for business continuity or improved IT operations, Hitachi Universal Replicator and Hitachi TrueCopy™ Heterogenous Remote Replication software with synchronous and asynchronous capabilities provide the enterprise-class performance associated with storage-system-based replication while delivering truly resilient business continuity without the need for redundant servers or replication appliances. Both Hitachi Universal Replicator software and Hitachi TrueCopy™ Remote Replication software deliver simplified remote data replication across Hitachi TagmaStore® Universal Storage Platform and Hitachi TagmaStore Network Storage Controller internal and externally-attached storage. Both Universal Replicator and TrueCopy support Geographically Dispersed Parallel Sysplex (GDPS), an IBM service offering for system failover, workload balancing, and data mirroring.

Synchronous Remote Replication

For distances within the same metropolitan area, Hitachi TrueCopy™ Heterogenous Remote Replication software with synchronous capabilities provides a no-data-loss, rapid restart solution. TrueCopy Remote Replication Synchronous software yields the highest degree of data integrity because its real-time copies are the same as the originals.

Asynchronous Remote Replication

Hitachi Universal Replicator software and Hitachi TrueCopy™ Asynchronous software can be deployed for wide-area disaster protection across virtually any distance.

TrueCopy Asynchronous software delivers premier data integrity with minimal performance impact on the primary system. Able to operate at any distance, TrueCopy software supports fast restarts and recovery by ensuring proper database update sequences for each transaction by using a unique method of sequence numbers and timestamps in each data record to ensure proper sequencing and data integrity during transmission and recovery.

Universal Replicator software for the Universal Storage Platform and the Network Storage Controller provides advanced replication among all of the storage systems certified for external attachment to these two storage platforms, permitting data to be copied from any supported device to any other supported device, regardless of operating system or protocol differences. Using industry-leading controller-based virtualization, the Hitachi storage platforms enable this single replication tool to operate against all heterogeneous storage resources in a tiered infrastructure. This significantly reduces the complexity and cost of replilcating data, both locally and long distance.

A unique asynchronous implementation, Universal Replicator software at the primary site writes designated records to cache and a specific set of disk journal volumes. The remote Universal Storage Platform then reads the records from the journal cache or volumes, offloading the primary system by pulling them across the communication link, instead of making the primary system push them as in most other approaches. By writing records to journals instead of keeping them solely in storage system cache, Universal Replicator software does not consume available cache, freeing resources for production transactions, eliminating the most common cause of asynchronous replication failure, and permitting the replication bandwidth to be sized towards average utilization instead of peak demand.

Three Data Center Configurations

For enterprise environments, Universal Replicator ensures availability of up-to-date copies of data in dispersed locations by leveraging synchronous capabilities of Hitachi TrueCopy Remote Replication software. Three data center configurations, illustrated under "Reference Architecture" below, include the following:

Three Data Center Cascade replicates data from the primary site to an intermediate site via TrueCopy Synchronous software and then to a third remote location with Universal Replicator software.

Three Data Center Multi-Target simultaneously copies data from a central location to a hot stand-by site via TrueCopy Synchronous software and to a third site via Universal Replicator.

Three Data Center Multi-Target with Delta Resync supports recovery of the remote site from the synchronous copies of journal data at the hot stand-by site if the primary site has failed.

Large Data Center Configurations

Universal Replicator 4x4 in the mainframe environment supports a single consistency group spanning up to four storage systems at either or both primary and remote sites in any "NxN" combination up to 4x4: 3x3, 2x1, etc. In this configuration, the Universal Replicator Extended Consistency Group feature supports up to 16,000 volumes in a single consistency group for each storage system in the complex. Universal Replicator 1x1, 2x2, 3x3 and 4x4 configurations would support up to 16,000, 32,000, 48,000 and 64,000 volumes in one consistency group.

Local Data Copies

Hitachi ShadowImage™ In-System Replication software plays a key role in many of the recommended remote replication architectures by creating disk-based data copies within a single Hitachi storage system. ShadowImage software provides a safeguard formission critical application consistency, near instant recovery from data corruption and "point-in-time" (PIT) data copies for immediate and nondisruptive access and sharing of information for decision support, test and development, or to optimize tape backup operations.

Replication Management

Hitachi Business Continuity Manager software for IBM® z/OS® offers centralized, enterprise-wide replication management for IBM z/OS mainframe environments. Through a single, consistent interface based on familiar TSO/ISPF full-screen panels, Business Continuity Manager software automates Hitachi Universal Replicator, Hitachi ShadowImage™ In-System Replication, and Hitachi TrueCopy™ Remote Replication software operations, accessing key replication metrics with built-in performance monitoring. Business Continuity Manager presents views to the status of all enterprise-wide replication objects in real time and provides automatic notification of key events completion, such as pair state transitions, timeout thresholds, and other system events.

Hitachi HiCommand® Replication Monitor software simplifies administration of the entire suite of Hitachi replication products for open systems and mainframe environments with a single, easy-to-use display for monitoring and visualizing volume replication configurations and status information. It streamlines storage administration and replication management functions by interfacing with Hitachi Device Manager software for Hitachi storage systems and replication software. This allows storage administrators to get a visual reference for data under replication management, as well as a point-in-time status indicator of replicated pairs including recovery point.

Business Impact and Benefits

Deployment of Hitachi Data Systems' remote replication technologies deliver business benefits through a variety of common applications, among them:

Disaster Recovery and Business Continuity

Remote copy technology can greatly reduce the potential losses incurred during a large scale disaster. Because a nearly simulataneous duplicate of production disk resources can be located at sites far enough removed from production facilities as to be unaffected by disasters such as floods, hurricanes, terrorism, earthquakes, and similar events. By failing over to the remote facility, processing can be quickly resumed at the recovery facility with minimal loss of information.

Data Migration

Moving production data from one storage system to another can be enormously disruptive and require significant service outages as well as planning and logistic effort. Remote copy can be used to substantially minimize the complexity of moving data to target devices, and reduce outage durations that are needed when production applications cut over to the new environment

Data Center Relocation

Long distance data center relocation can require long duration outages if storage devices are off-line while de-installed, transported to their new home, and then finally brought on-line. With remote copy, outage durations are significantly smaller since an up-to-date can be created at the new facility while production applications are still on-line.

Further, the suite of Hitachi Data Systems remote replication technologies deliver asynchronous and synchronous remote replication of data from one storage system to another for everyday uptime improvement and rapid recovery in the event of an outage. Specifically:

Hitachi TrueCopy™ Heterogenous Remote Replication Software

- Supports fast restarts and recovery by ensuring proper database update sequences for each transaction during transmission between enterprise storage systems
- Improves service levels by reducing planned and unplanned downtime of customer-facing applications

Hitachi Universal Replicator Software

- Reduces cache utilization and maximizes the use of transmission-line bandwidth by leveraging performance-optimized disk-based journals
- Reduces costs, requiring only one product to provide asynchronous copy services for use across all attached storage systems
- Can significantly reduces RPO through advanced point-in-time recovery capabilities afforded by the use of journaling technology
- Maintains protection of data in the event of total network outages – depending on journal size and workload write rate – for swift recovery upon resumption of network connectivity by eliminating the need for a full volume copy to recover the remote site, therefore reducing the exposure to lost data
- Ensures availability of up-to-date copies of data in dispersed locations by leveraging synchronous capabilities of Hitachi TrueCopy Remote Replication software, including replication to multiple data centers as well as to both remote and hot standby data centers
- Eliminates the need for any difficult to understand and implement recovery procedures as with other products/solutions
- Provides single pane-of-glass heterogeneous replication through the virtualization capabilities of the Universal Storage Platform, allowing the replication of any volume hosted on any supported externally attached storage system

In addition, in the event of a site-wide failure at the primary site, the journal at the intermediate **Three Data Center Cascade** site can continue to propagate the reaming I/O to the remote site.

Three Data Center Multi-Target minimizes failure points because each replication leg is independent of the other. And similar to Universal Replicator Three Data Center Cascade, in the event of a sitewide failure at the production site, the journal at the synchronous site in the **Three Data Center Multi-Target with Delta Resync** environment can continue to propagate the remaining I/O to the remote site.

Hitachi ShadowImage™ In-System Replication Software

- Shortens restart and recovery times with the consistency-group function, which provides multivolume, point-in-time copies for applications and databases that share or span multiple volumes.
- Reduces recovery from data corruption time dramatically through the ShadowImage QuickRestore feature, which allows an immediate restore to a disk-resident, point-in-time data copy.
- Replicates large data volumes without impacting service levels, timing out, or affecting performance levels.
- Enables normal backup operations on a copy of up-to-date production data while critical applications continue to run unaffected

Hitachi Business Continuity Manager Software

- Dramatically reduces recovery times by automating complex disaster recovery and planned outage functions
- Allows proactive problem avoidance and optimum performance to ensure that service-level objectives are met or exceeded by providing access to critical system performance metrics and thresholds
- Eliminates hours of tedious input and costly human error when configuring and protecting complex, mission-critical applications and data through its auto-discovery capability
- Universal Replicator and ShadowImage In-System Replication software support Business Continuity Manager’s mainframe ATTIME Split functionality, allowing real-time ShadowImage software replication of Universal Replicator software’s remote volumes without suspending the Universal Replicator pairs.

Hitachi HiCommand® Replication Monitor Software

- Monitors data currency and recovery points for both open and mainframe environments
- Offers advanced replication system status reporting and built-in capabilities for monitoring and managing replicated volumes for active problem avoidance
- Provides an enhanced topological –like view of user-selected copy groups for simplified management

Reference Architecture

The following tables (Table 1 through Table 8) detail the spectrum of Hitachi Data Systems’ remote replication architectures from simplest to most advanced.

Table 1: Real-time Copy

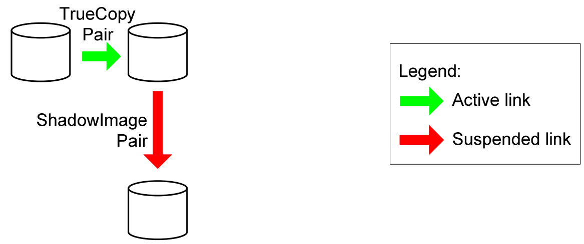
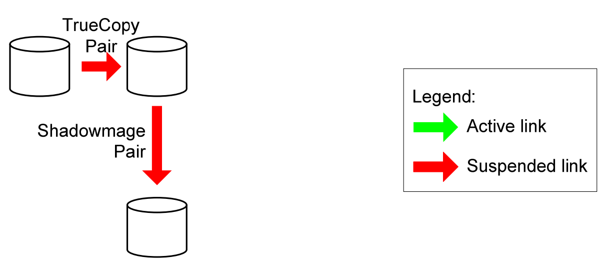
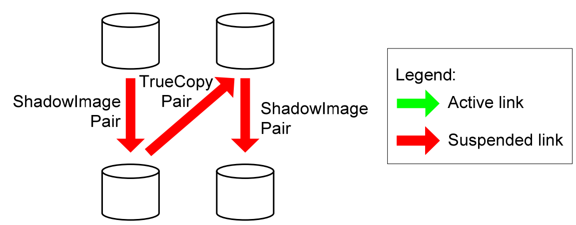
Real-time Copy	
RPO with Site Wide Instant Disaster	Low RPO (0 for TrueCopy Sync, minutes for TCA)
Recovery from Logical Corruption	Planned local or remote Point in Time recovery images
Recovery from Logical Corruption during Disaster Recovery Testing	No recovery during DR testing unless additional point in time copies are added
Host response time	Sensitive to bandwidth, cache, remote storage system health, and distance if using synchronous copy
Application Interaction Complexity	Application quiesce required to automate point in time split
Bandwidth	High Size link to peak workload
Storage Capacity Requirements (Primary to Secondary)	1:2
Front End Director processor requirements per storage system	> 2 dependent on workload
Manageability Ranking	Simplest
Implementation Complexity Ranking **(1=lowest, 5=highest)**	1

Table 2: Point-In-Time Three-Copy Model

Point-In-Time Three-Copy Model	
RPO with Site Wide Instant Disaster	High RPO (hours)
Recovery from Logical Corruption	Planned local or remote Point in Time recovery images

Recovery from Logical Corruption during Disaster Recovery Testing	RPO = Test duration + resynch duration
Host response time	Sensitive to resynch timing (with initial copy as an upper bound), copy pace
Application Interaction Complexity	Application quiesce required before suspending TrueCopy pair(s)
Bandwidth	Lower Size link to peak RPO Average + safety margin for production activity during resynch
Storage Capacity Requirements (Primary to Secondary)	1:2
Front End Director processor requirements per storage system	>2 dependent on workload
Manageability Ranking	Simple
Implementation Complexity Ranking **(1=lowest, 5=highest)**	3

Table 3: Point-In-Time Four-Copy Model

Point-In-Time Four-Copy Model	
RPO with Site Wide Instant Disaster	High RPO (hours)
Recovery from Logical Corruption	Planned local or remote Point in Time recovery images
Recovery from Logical Corruption during Disaster Recovery Testing	RPO = Test duration + resynch duration
Host response time	Sensitive to ShadowImage placement and ShadowImage resynch duration
Application Interaction Complexity	Application quiesce required to automate point in time split
Bandwidth	Lowest Size to peak RPO Average Compared to Batch DR, TrueCopy pair does not need to catch up to production updates)
Storage Capacity Requirements (Primary to Secondary)	2:2

Front End Director processor requirements per storage system	>2 dependent on workload
Manageability Ranking	More detailed
Implementation Complexity Ranking **(1=lowest, 5=highest)**	4

Table 4: Universal Replicator Asynchronous Replication

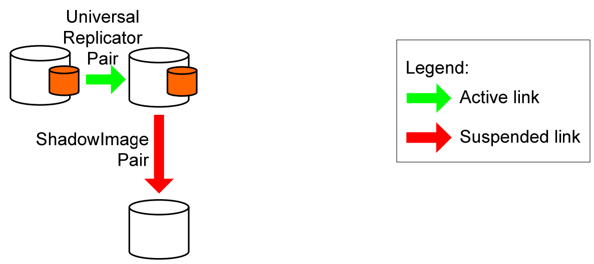
Universal Replicator Asynchronous Replication	
RPO with Site Wide Instant Disaster	Flexible RPO (minutes to hours)
Recovery from Logical Corruption	Planned local or remote Point in Time recovery images
Recovery from Logical Corruption during Disaster Recovery Testing	No recovery during DR testing unless additional point in time copies are added
Host response time	Sensitive to journal throughput and journal placement
Application Interaction Complexity	Application quiesce required to automate point in time split
Bandwidth	Flexible Size to peak workload or to peak RPO average Journals can be used to decrease bandwidth requirements by increasing RPO
Storage Capacity Requirements (Primary to Secondary)	~1.3:2.3 Roughly, when using Universal Replicator, every three parity groups of production volumes will require one parity group dedicated to journals. Actual requirements are dependent on specific workload.
Front End Director processor requirements per storage system	>4 dependent on workload
Manageability Ranking	Moderate Adding workload or volumes requires re-evaluation of journal configuration

Table 5: Universal Replicator Three Data Center Cascade

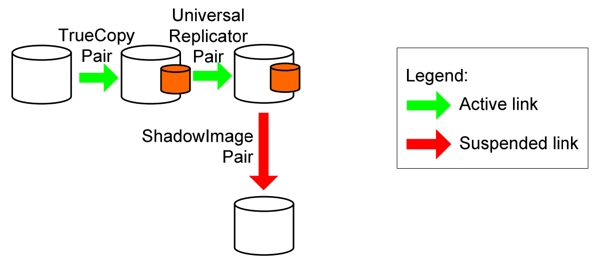
Universal Replicator Three Data Center Cascade	
RPO with Site Wide Instant Disaster	RPO as low as 0 Can provide for no data loss
Recovery from Logical Corruption	Planned local or remote Point in Time recovery images
Recovery from Logical Corruption during Disaster Recovery Testing	No recovery during DR testing unless additional point in time copies are added
Host response time	Sensitive to bandwidth, cache, remote storage system health, journal throughput, journal placement and distance
Application Interaction Complexity	Application quiesce required to automate point in time split
Bandwidth	Highest Size to peak for TrueCopy links
Storage Capacity Requirements (Primary to Secondary)	1:~1.3:~2.3 Roughly, when using Universal Replicator, every three parity groups of production volumes will require one parity group dedicated to journals. Actual requirements are dependent on specific workload.
Front End Director processor requirements per storage system	>2 for TrueCopy link dependent on workload >4 for Universal Replicator link dependent on workload
Manageability Ranking	Most detailed Adding workload or volumes requires re-evaluation of journal configuration
Implementation Complexity Ranking **(1=lowest, 5=highest)**	5

Table 6: Universal Replicator Three Data Center Multi-Target

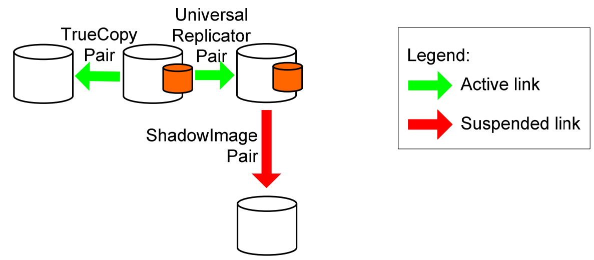
Universal Replicator Three Data Center Multi-Target	
RPO with Site Wide Instant Disaster	RPO as low as 0 Can provide for no data loss
Recovery from Logical Corruption	Planned local or remote Point in Time recovery images
Recovery from Logical Corruption during Disaster Recovery Testing	No recovery during DR testing unless additional point in time copies are added
Host response time	Sensitive to bandwidth, cache, remote storage system health, journal throughput, journal placement and distance
Application Interaction Complexity	Application quiesce required to automate point in time split
Bandwidth	Highest Size to peak for TrueCopy links
Storage Capacity Requirements (Primary to Secondary)	1:~1.3:~1.3 Roughly, when using Universal Replicator, every three parity groups of production volumes will require one parity group dedicated to journals. Actual requirements are dependent on specific workload.
Front End Director processor requirements per storage system	>2 for TrueCopy link dependent on workload >4 for Universal Replicator link dependent on workload
Manageability Ranking	Most detailed Adding workload or volumes requires re-evaluation of journal configuration
Implementation Complexity Ranking **(1=lowest, 5=highest)**	5

Table 7: Universal Replicator Three Data Center Multi-Target with Delta Resync

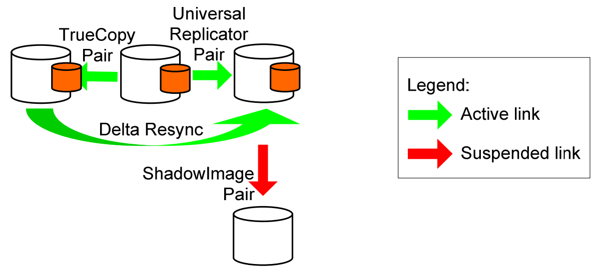
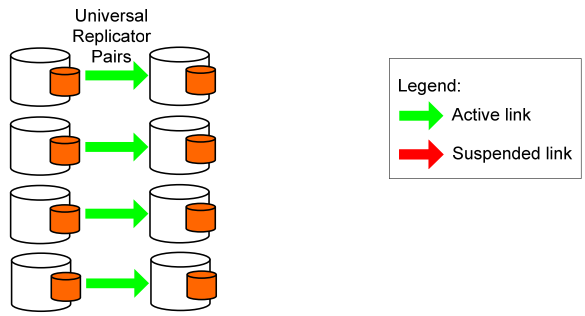
Universal Replicator Three Data Center Multi-Target with Delta Resynch	
RPO with Site Wide Instant Disaster	RPO as low as0 Can provide for data loss
Recovery from Logical Corruption	Planned local or remote Point in Time recovery images
Recovery from Logical Corruption during Disaster Recovery Testing	No recovery during DR testing unless additional point in time copies are added
Host response time	Sensitive to bandwidth, cache, remote storage system health, journal throughput, journal placement and distance
Application Interaction Complexity	Application quiesce required to automate point in time split
Bandwidth	Highest Size to peak for TrueCopy all links
Storage Capacity Requirements (Primary to Secondary)	1:1:1 Roughly, when using Universal Replicator, every three parity groups of production volumes will require one parity group dedicated to journals. Actual requirements are dependent on specific workload.
Front End Director processor requirements per storage system	>2 for TrueCopy link dependent on workload >4 for Universal Replicator link dependent on workload
Manageability Ranking	Most detailed Adding workload or volumes requires re-evaluation of journal configuration
Implementation Complexity Ranking **(1=lowest, 5=highest)**	5

Table 8: Universal Replicator Three Data Center Four-by-Four

Universal Replicator Three Data Center 4X4	
RPO with Site Wide Instant Disaster	Flexible RPO (minutes to hours)
Recovery from Logical Corruption	Planned local or remote Point in Time recovery images
Recovery from Logical Corruption during Disaster Recovery Testing	No recovery during DR testing
Host response time	Sensitive to bandwidth, cache, remote storage system health, journal throughput, journal placement and distance
Application Interaction Complexity	Application quiesce required to automate point in time split
Bandwidth	Flexible Size to peak workload or to peak RPO average Journals can be used to decrease bandwidth requirements by increasing RPO
Storage Capacity Requirements (Primary to Secondary)	1:1 Roughly, when using Universal Replicator, every three parity groups of production volumes will require one parity group dedicated to journals. Actual requirements are dependent on specific workload.
Front End Director processor requirements per storage system	Consistency groups can span directors at either or both locations, unmatched 1 to 4 for Universal Replicator link(s) dependent on workload
Manageability Ranking	Most detailed Adding workload or volumes requires re-evaluation of journal configuration
Implementation Complexity Ranking **(1=lowest, 5=highest)**	3

Architectural Alternatives

Table 9 below presents the costs and benefits of several common options to the reference architectures defined in Tables 1 through 8 above.

Table 9: Design Variations for Replication Architectures

Option	Benefits	Costs
Additional ShadowImage In-System Replication software copies at recovery site	• DR testing without impact to RPO	• Additional storage capacity for additional ShadowImage copies • Incremental increase to ShadowImage software licenses • Additional scripting to accommodate additional pairs
Additional ShadowImage In-System Replication software copies at production site	• Local Recovery from logical corruption	• Additional storage capacity for additional ShadowImage copies • Incremental increase to ShadowImage software licenses • Additional scripting to accommodate additional pairs
Symmetric/bi-directional remote copy links	• Provides for reverse replication to support post-outage failback	• Sufficient ports on storage system and channel extenders must be available to support bi-directional replication
Fully symmetric/bi-directional configuration (remote copy links plus recovery volumes)	• Ability to move applications to the recovery facility while still maintaining all recovery options • Reduces administrative complexity because operations can be performed from either site with minimal change in procedures • Simplifies and accelerates the failback process	• Additional storage necessary to provide symmetric ShadowImage configurations at both the primary and secondary sites • Sufficient ports on storage system and channel extender must be available to support bi-directional replication
Hitachi Copy on Write Snapshot software	• Up to 64 Point-In-Time Copies • Ability to reduce storage requirements to support additional copies (dependent on locality of write activity)	• Host I/O Performance Impact • Pool overflow conditions will prevent recovery from secondary volumes • May be inappropriate for use as testing volumes due to pool capacity considerations

Implementation Planning

Data Collection Tools and Processes

Successful implementation planning involves collection of data for a comprehensive view of the replication environment. In addition to business statistical data, Hitachi Data Systems representatives may employ or recommend the following tools in the solution planning process:

Risk Analysis and Remote Copy Planning and Design Services

Hitachi Data Systems Global Solution Services provides a number of thorough planning, implementation and integration services for the data replication environment. Key among them for objectively evaluating business requirements and risks for distance replication are the Risk Analysis and Remote Copy Planning and Design services.

Starting with the **Risk Analysis Service** will help identify and quantify the probability of loss occurrence and expected losses for the company. Hitachi Data Systems Global Solutions Services help determine which risks are acceptable and which require mitigation; they will develop a risk model to evaluate critical exposures, examine regional and local risks to calculate expected frequency and probable losses, and prioritize risks based on likelihood of occurrence. The following deliverables are included in the scope of this service:

- An IT Infrastructure and Facility Risk/Exposure Analysis
- A Business Unit and Facility Risk/Exposure Analysis
- A Risk Analysis Report covering local and regional risks and exposures, a prioritized matrix of exposures and expected occurrences, and a suggested approach for mitigating and addressing the most pressing risks and exposures, and
- An Engagement Findings Executive Presentation summarizing the Risk Analysis Report

With the most critical risks and exposures identified, the **Remote Copy Planning and Design Service** applies data replication best practices to produce a detailed study of the existing environment and a documented high-level strategy for implementing the most appropriate and cost effective distance replication solution. Deliverables from the Remote Copy Planning and Design Service include:

- An audit of host and storage environment hardware and software to be included in the replication environment
- A report of workload and performance characteristics of the volumes in the replication environment
- Documented objectives for the replication environment
- Documented mechanisms and techniques to support achievement of those objectives
- A strategic recommendation report providing feedback and recommendations for an overall approach, schedule and key success factors, and
- A recommended configuration identifying volumes, copy groups, update frequency and other management criteria

Remote Copy Expert Assistant

Remote Copy Expert Assistant (RCEA) is a tool used by Hitachi Data Systems representatives to automate and streamline tasks needed to deliver remote copy services such as the Remote Copy Planning and Design Service. RCEA encompasses the following steps:

Data Collection

Workload, performance and configuration metrics are collected at the host and storage levels. Ideally, data should be collected spanning a full four- to six-week business cycle to include standard data processing peaks like month or quarter end; collection timeframes for multiple servers should overlap.

- For the Mainframe
  - SAS can be used for data analysis
  - RMF Magic is a third-party analysis tool
- For Open Systems
  - RCEA from the Tools Competency Center uses common data collection scripts
  - Excel can be used to compile data manually

Data Processing and Analysis

The data is then transported to Hitachi Data Systems where it secured and workloads are modeled to identify an optimal replication solution. RCEA also allows Hitachi Data Systems representatives to do combined storage and host data analysis

Resource Requirements

Bandwidth and Replication Paths

In many respects, replication traffic is processed much like any other workload on a storage system. Write I/O on the primary device is transferred across a wire to the secondary device. For Hitachi remote copy products, this traffic uses SAN Fibre Channel connections between SCSI initiator ports at the production facility to SCSI target ports on the secondary storage system. Those Fibre Channel paths have a specific bandwidth capacity, and multiple connections may be necessary to ensure sufficient capacity.

It is an industry best practice to dedicate separate bandwidth offering the highest Quality of Service to data replication.

TrueCopy Synchronous software requires high bandwidth and low latency. TrueCopy Asynchronous and Universal Replicator software require less bandwidth and will tolerate some latency. However, variation in latency over time or “jitter” should be kept to minimum.

To maintain a continous replica copy, bandwidth must exceed the average write workload that occurs during any given RPO interval subject to the capacity limitations of the buffering mechanism. This means that if an organization wants to maintain an RPO of twenty minutes, the twenty minute interval with greatest write activity must be identified. With that, the bandwidth and buffer capacity required to keep up with this traffic can be calculated. In practice, an absolute peak interval cannot be identified; data on hand can be used to revise resource requirements up to accomodate gaps in data and uncertainty due to assumptions made during the assessment process. Further, in calculating bandwidth requirements for synchronous replication, network latency and protocol conversion are added twice to initial storage response time since data transmission via Fibre Channel protocol requires two round trips—one for command and one for data.

Network bandwidth recommendations are offered through the Remote Copy Planning and Design Service, but the final network choice is the customer’s responsibility. Table 10 below weighs the pros, cons and applications for many available bandwidth options.

Table 10: Network Connectivity Options for Remote Replication

Connectivity	Pros	Cons	Used for
Dark Fiber	• Bandwidth • Highest Quality of Service	• Cost • Availability • Complexity	• SAN extension and Synchronous replication
DWDM	• Bandwidth • High Quality of Service	• Cost	• SAN extension and Synchronous replication
Optical Carrier Networks	• Quality (packet loss and latency) • Availability • Cost efficiency		• Could be shared or dedicated for different network services
Ethernet (IP) Networks	• Lowest cost • Shared with other data services • Highest availability	• Require Fibre protocol conversion • Highest protocol overhead • Latency jitter due to routing	• Most widely used network services

Redundancy Requirements

In addition, it's important to identify redundancy requirements. As a best practice, replication over distance should occur over redundant independent wide area network (WAN) circuits. In the event of a telco outage, a seperate independent circuit must be available that will accomodate the production write workload.

Processing Capacity on the Storage Systems

Remote Copy operations incur overhead on the storage system; additional processor cycles are consumed to service the additional steps necessary. On Hitachi storage systems, this additional workload occurs within the processors for the front-end director ports, which are also used to service host I/O. Sufficient front-end director ports should be allocated to accomodate the production write workload alongside additional requirements for handling remote copy operations.

Channel Extension

As mentioned above, Hitachi replication traffic is processed using SAN Fibre Channel connections. When dealing with SAN connections over long distances, channel extension devices are typically deployed to overcome distance limitations of the Fibre Channel protocol.

Buffer Capacity

As RPO increases, sufficient buffer capacity and performance are required to contain and service all write operations that occur within that interval. For aggressive RPO targets, buffer resources can be small. But if RPO is balanced against bandwidth requirements to save on recurring costs, those buffer requirements – in terms of both capacity and performance – will go up.

Storage for Replica Copies

There are two ways to determine storage requirements for replica copies when using Hitachi remote copy, depending on product:

TrueCopy Synchronous, Asynchronous and Universal Replicator software: Primary and secondary storage devices must have equivalent storage capacity.

Hitachi remote copy products impose no inherent limitations according to RAID levels, drive speed, and other characteristics.

Software Requirements

Hitachi HiCommand® Business Continuity Manager software

- Business Continuity Manager 5.0 is required to take advantage of Universal Replicator Multi-Target with Delta Resynch and the Extended Consistency Group feature available with Universal Replicator 4x4

Resource Measurement Facility

- Performance Monitor with Export Tool software installed at each site

Command Device

A command device is used to provide in-band management and control of the replication process using native storage protocols. This command device is a disk volume allocated on the storage system which is dedicated to buffering and communicating commands to control and monitor pair operations. The command device cannot be used for production data.

For mainframe environments, a minimum of two command devices is required per pair of storage systems; for bi-directional replication (reverse synch), add an additional pair of command devices.

A route is a path connecting storage systems by command devices so that commands can traverse to remote storage systems. In a two datacenter configuration, a route will start on the production storage system and end at the recovery storage system. In a cascaded three datacenter configuration, the route will start on the production storage system command device, proceed to the intermediate storage system's command device, and terminate on the last storage system's command device.

Sizing & Design

Remote Copy Planning and Design Service

As described above,. Design of any disaster recovery solution begins with delivery of the Remote Copy Planning and Design Service by Hitachi Data Systems Global Solution Services which examines the current configuration of the environment considered for replication, establishes the business and technical requirements for the implementation, and proposes a high-level solution design to meet established requirements. Whereas the Remote Copy Planning and Design Service will cover many of the following details, specific attention should be given to understand the existing and near-term business, technical and environmental requirements covered below.

Data Gathering – Business Requirements

The most effective best practice for planning and implementing a data replication solution is to understand the goals up front: business and technical requirements, objectives, timelines, what's possible. Goals for the project might include:

- Disaster recovery
- Business continuity
- Migration
- Relocation

- Along with the business goals for replication, it is important to identify the timeframe for implementation.

Understanding Data Protection Classifications

With an understanding of recovery point and time objectives, the type of disasters against which a business may protect its data are divided in two categories.

Disasters of **scale** would include:

– Point disaster representing a single event at a single point of time. This could be caused by human error, or isolated hardware failure.

– Site-wide disaster impacting operations at an entire facility, caused by fire, earthquake or hurricane.

Disasters of **time** include:

– Immediate disaster where a single, distinct event impacts all components at the same time, such as a meteor.

– Rolling disaster where several components fail at different points in time, for example an air conditioning or power failure takes down a server, then storage, network, eventually the entire site.

Regulatory Requirements

Data protection strategies address and support regulatory requirements such as

- Sarbanes-Oxley–corporate and financial reporting
- Basel II–back office systems and risk management tools for banks
- Email archiving–litigation defense under business records laws

A risk analysis can assist in determining recovery objectives and the appropriate level of data protection.

Data Gathering – Environmental Requirements

Environmental requirements, like business requirements, must consider the existing components. Hitachi Data Systems' Remote Copy Planning and Design service will help to uncover all applicable elements in the existing environment, including:

- Current infrastructure – hardware and software

☐ Servers and storage

☐ Operating systems

☐ System monitoring software

☐ Software licenses

☐ Volume Manager

☐ Multi-pathing

☐ Enterprise scheduler

☐ File system to LUN map

☐ Hardware connections

In order to support the replication environment, connectivity must be in place between the storage systems at each site.

It is Hitachi Data Systems' best practice to utilize processors on different processor clusters distributed across power boundary regions for TrueCopy operations. On the Universal Storage Platform, ports 1X/5X, 3X/7X, 2X/6X, and 4X/8X share port processors and must be configured with similar functionality. It is therefore important to note that ports that share processors with designated replication initiator or target ports cannot be used for LUN target ports.

- [ ] Routers

Routers may also be customer-selected, new or existing. Again, Hitachi Data Systems recommends a second circuit be provisioned for growth and redundancy.

- [ ] Network and channel extension

Use SAN extension if using Dense Wave Division Multiplexing (DWDM) or similar, or plan to merge fabrics over sites and links. DWDM is useful for short distance SAN environments, offering very high bandwidth but limited distance. DO NOT use SAN extension if extending SAN over telecom or over medium to long distance or cannot merge fabrics over sites and links.

- [ ] WAN dedicated to replication?
- [ ] Telecommunications

The telecommunications between routers can be customer-selected; the overall architecture of the replication solution is not dependent on the telecommunications protocol. Bandwidth should be monitored constantly for usage. Also, Hitachi Data Systems recommends a second circuit be provisioned for growth and redundancy.

Distance between replication sites must be taken into account in managing the delay in the transmission of data. Hitachi Data Systems assumes a millisecond of delay for each 125 miles or 200 kilometers.

- [ ] Third party site
- [ ] Backup Schedule, tape rotation
- [ ] Existing disaster recovery servers, storage
- [ ] Change control procedures related to implementation, project plans, timelines
- [ ] Current recovery techniques
- [ ] Existing disaster recovery test schedule and written plan
- [ ] Data and workload, growth rates
- [ ] Replication infrastructure
- [ ] Servers, host environment for CCI

Command Control Interface (CCI) servers should have Fibre Channel connectivity to the storage system(s) and uninhibited network access to all other CCI servers in the configuration.

- [ ] Preferred scripting language
- [ ] Distance between replication sites
- [ ] Replication direction – bi-directional? Failback after failover
- [ ] WAN Quality of Service

- Workload

Without a clearly stated and agreed upon workload it is impossible to accurately design a configuration that will meet the success criteria for the implementation. Workload and environment characteristics required to size any installation include:

- ☐ Maximum I/O rate
- ☐ Read/write ratio
- ☐ Predominant write block size
- ☐ Distance between sites
- ☐ FICON or ESCON host attachment
- ☐ Number of devices and their use
- ☐ I/O driver requirement and selection
- ☐ I/O consistency goals
- ☐ Data replication goals
- ☐ Configuration Software products/components required
- ☐ WAN workload

Data Gathering – Technical Requirements

Infrastructure

- New environment
- Conversion of previous environment

Application Specific Questions

- ☐ Version
- ☐ Function
- ☐ RPO and RTO values
- ☐ Backup requirements
- ☐ Recovery requirements
- ☐ Other data copies for backup, reporting, etc.
- ☐ Storage system

Vendor Preferences

Recovery Techniques

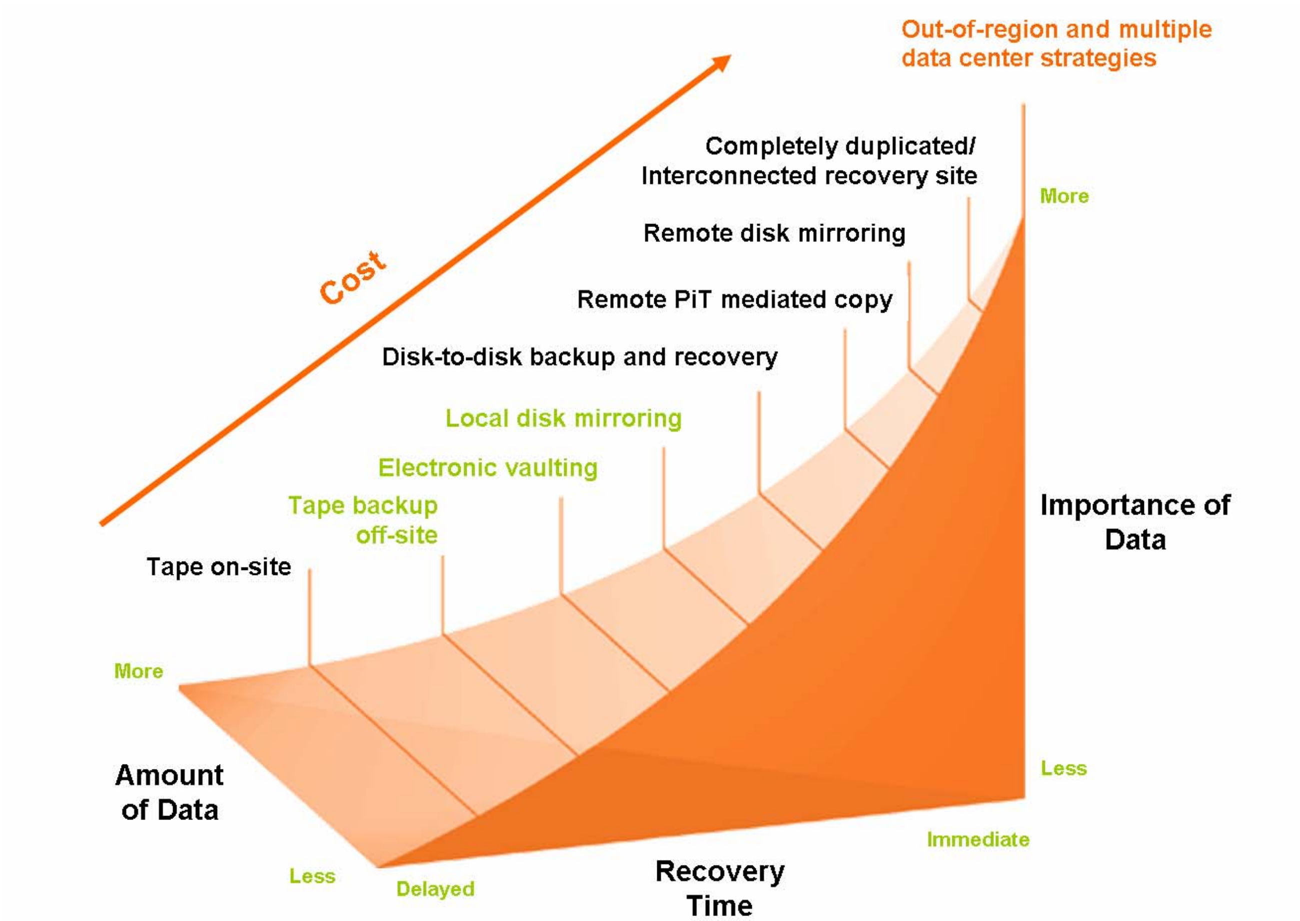
Recovery may be accomplished through a variety of means, such as:

- Manual data re-entry
- File-level recovery
- Tape recovery
- Rebuild hardware environment from equipment/resources on hand
- Rapid recovery from point-in-time copy

- Cluster failover

Each technique should be weighed against recovery time, cost and value of the data as illustrated in Figure 3.

Figure 3: Recovery Techniques on the Data Protection Continuum



RTO, based on the value of specific data, maps to a range of technical approaches, costs, and degrees of data protection.

Table 11 further details specific RPO and RTO ranges and resource planning recommendations for each of technology tiers represented by these recovery techniques.

Table 11: Disaster Recovery Tiers

Technology Tier	RPO Range	RTO Range	Minimum # Disk Copies	Distance
Tape Backup[1]	24-168 hours	2-168 hours	N/A	Any
Virtual Tape[1]	12-48 hours	1-24 hours	Disk Pool	Any if replicated
Disk Point-in-Time[1]	2-36 hours	15 minutes- 12 hours	3[2]	Any
Synchronous Remote Copy	0-2 minutes	1-8 hours	2[2]	Limited
Synchronous Remote Copy w/Failover	0-2 minutes	5-60 minutes	2[2]	Limited
Asynchronous Remote Copy	0-5 minutes[3]	30 minutes- 8 hours	2[2]	Any
Asynchronous Remote Copy w/Failover	0-5 minutes[3]	30-90 minutes	2[2]	Any
Three Data Center	0-2 minutes	1-8 hours	3-7[2,4]	Any

Each technology tier supports a clear range in RPO, RTO and distance.

Note 1 Depends on how much data is being recovered and how often backups or point-in-time copies are taken

Note 2 Best practice is one additional copy for doing DR testing without impacting the ongoing replication session

Note 3 Network problems will extend the RPO

Note 4 Depends on vendor and method deployed

Journal Groups

Hitachi Universal Replicator journal groups function similarly to Hitachi TrueCopy asynchronous software consistency groups but with an expandable sidefile in the form of array group storage.

There are a number of points to consider when planning the journal groups:

- Number of journal groups
- Number of array groups per journal group
- Number of journal volumes per journal group
- Journal volume size
- Journal volume emulation
- Extended consistency groups
- Delta Resynch with Three Datacenter Multi-Target
- Universal Replicator 4x4

Number of Journal Groups

For planning purposes and as a general Rule of Thumb (ROT), the Universal Storage Platform can conservatively drive around 15,000 write I/Os per Hitachi Universal Replicator journal group, regardless of write block size.

In a production environment determine how many journal groups (think consistency groups) are required based upon information such as:

- The need to maintain I/O consistency between multiple/all applications
- The IOPS rate, R/W ratio and predominant write block size
- The single journal group write I/O rate planning limit of ~15,000 write I/Os per second (IOPS)

This requires an idea of which volumes/applications will run in each journal group, how many write I/Os in each journal group (15,000 maximum write I/Os per journal group) and the predominant write block size for each journal group.

The journal group array group requirement for **EACH** journal group must be calculated.

√ This task is best left to the GSS practitioner during the Remote Copy Planning and Design Services engagement.

NOTE: If choosing to use the EXTCG (Extended Consistency Group, aka 4x4) feature therefore tieing together multiple journal groups to a single I/O consistency point, the workload must still be intelligently divided into multiple journal groups so as not to exceed the single journal group maximum write rate of ~15,000 write IOPS as mentioned above, and the journal group array requirements for each group must be calculated.

Number of Array Groups per Journal Group

Hitachi Data Systems representatives and authorized dealers have tools to determine how many array groups will be required to support the expected maximum R/W I/O rate, the number of array groups depending on their RAID type, HDD size/configuration, their respective throughput specifications, and also how many Hitachi Universal Replicator data and command paths are required to support this environment.

Regarding journal group arrays (simply put), current Hitachi guidelines for Hitachi Universal Replicator journal group configuration are as follows:

- Each journal group must have its own array groups, not shared by any host volumes or any other journal group activity.
- Care and consideration must be given to the physical placement of each journal array group to allow for the best performance possible.

NOTE: The journal group array type used to construct the Local "A" site journal volumes should be of the same size and type used to construct the corresponding journal group at the Remote "B" site.

Number of Journal Volumes per Journal Group

A journal group can have up to 16 journal volumes.

One of the main considerations when planning how many journal volumes should be in a given journal group is whether or not to plan for dynamic expansion of a journal group.

Journal volumes can be dynamically added into a given journal group, non-disruptively, up to the maximum of 16, while the customer is running.

Two points to note here:

- When dynamically adding journal volumes to a group, only the data area – not the metadata area – gets dynamically expanded, until a suspension of the journal group takes place.
- Journal volumes cannot be dynamically removed from a journal group without first suspending the group.

This means that when planning on the number of journal volumes to put in a journal group, starting off with the maximum number of 16 will not allow for dynamically adding volumes should the need arise.

Journal volumes do not have to be of equal size in a given group but as a Rule Of Thumb always try to make all journal volumes in a given group the same size for best efficiency.

Determining When to Dynamically Add Journal Volumes

There are two factors that affect this decision.

If the proper amount of cache is configured in the primary and the secondary storage systems (+50% or 25% of the 'normal' requirement for TrueCopy and Universal Replicator respectively), the journal volumes should consume minimum storage area during normal customer operation.

The real consideration for journal group storage capacity is to allow the Hitachi Universal Replicator pair environment to remain in "Duplex" status and continue to log write updates without going into a suspended state during times of network outage, network shortages, peak write rate greater than planned for, write pending situations for any number of reasons, or any time that the write update in cache cannot be sent to the remote site in a timely fashion and must be destaged down to the journal volume disks. The current maximum time for Path Blockade Watch is currently 60 minutes.

Determining Journal Volume Size

This is probably the simplest part. After deciding how many array groups are required for a given journal group and their specific type and size (for example RAID5 7+1 144GB), add up the total amount of storage and divide by the number of journal volumes desired in that given journal group.

An example:

- (2) Journal groups, supporting 10,000 4K write IOPS each.
- (2) RAID5 7+1 parity groups per journal group @ 1150 usable GB/parity group
- (10) OPEN-V journal volumes per journal group @ 230 usable GB/journal volume
- Each journal group would be 2.3TB useable capacity
- Total journal capacity for this configuration: 4.6TB

The by-product of the required array group configuration for each of these journal groups is enough journal storage to allow for close to 98 minutes of complete network outage while sustaining the 10,000 write IOPS rate @ 4K workload without suspending the pairs.

If the network returns, even if only partially, before the 98 minutes are up, the journaled write updates will again be transmitted to the remote site immediately moving the RPO forward without the need for a destructive write resync that is required for ALL other cache based solutions. **No other solution can do this today.**

Remember to use good array placement Rules of Thumb for installing the journal array groups across the BEDs to avoid contention. The journal arrays should be the best performing array groups in the configuration.

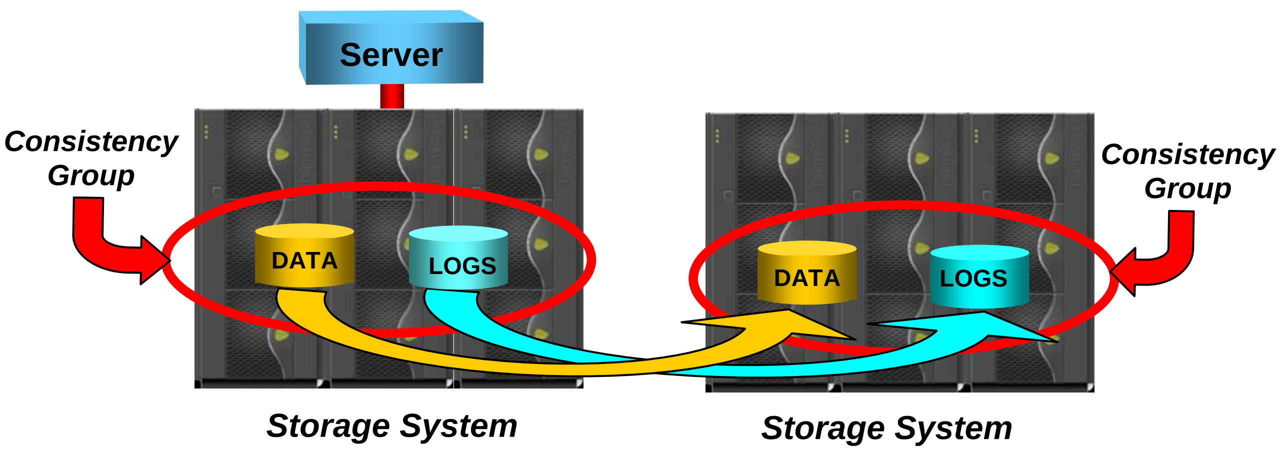
Journal Volume Emulation

For both Open Systems and mainframe environments, Hitachi Data Systems best practice would be to use OPEN-V emulation. Although other emulation modes are available in the mainframe environment, OPEN-V emulation provides the most desireable throughput characteristics.

Consistency Groups

I/O consistency, the idea of having all data changes written at the remote site in the same order in which they were written at the host site by the applications, maintaining changed data order integrity, is called "guaranteed in order writes". This is illustrated below in Figure 4. With Hitachi Universal Replicator software, this is realized at the individual journal group level following the same concept as the consistency group with TrueCopy Asynchronous software. All volumes that are within the same journal group will have their write I/O consistency maintained at the remote site.

Figure 4. Consistency Group



Recoverability with replication is dependent on in-order delivery. This is provided by consistency groups.

With the Universal Replicator software feature called EXTCG or Extended Consistency Groups, or '4X4', a single replication environment can have up to 16 individual journal groups, each containing up to 4096 pairs, I/O consistent with each other, yielding a single I/O consistency scheme for as many as 65,536 volumes.

The long term goal, as is true with almost every data center, is to have a single I/O consistency point across their entire shop, which will require the EXTCG feature to be implemented at some point in the future.

A consistency group (CTG) is a TrueCopy feature that ensures in-order replication and consequently I/O consistency across multiple LUNs.

Each server can be contained in a separate consistency group. A drawback of using a single consistency group for all servers is that no single server at a time can be tested at the DR site without minimal disruption to the entire consistency group.

As the TrueCopy Asynchronous software consistency group grows, utilization increases for the remote storage system's channel adapter ports and sidefile, as data is sequenced before being committed to the secondary volumes. Hitachi Data Systems recommends monitoring remote sidefile utilization and channel adapter utilization to ensure usage thresholds are not exceeded whereby TrueCopy pairs are suspended.

If the consistency group grows in a Universal Replicator software environment, the number and size of the journal groups need to be reevaluated.

Hitachi Data Systems also recommends mapping application business processes to determine which servers actually need grouped data consistency. This will enable greater scalability with more and smaller consistency groups in the environment and less chance for TrueCopy pair suspensions as the environment grows.

Additional Cache Requirements

For the distance replication environment, it is recommended that cache be increased by +50% in excess of the “normal” cache requirement for TrueCopy Synchronous and Asynchronous software. Universal Replicator has an additional cache requirement of +25% over the “normal” cache requirement.

As stated earlier, if the proper amount of cache is configured in the primary and the secondary storage systems (+25% of “normal”), the journal volumes should have minimum storage area consumed during normal operations.

Installation & Configuration

Copy Paths

All Universal Replicator and TrueCopy Synchronous paths used in the Three Data Center configuration are configured and confirmed functional using Storage Navigator.

Following the physical copy path setup using Storage Navigator, the Business Continuity Manager PATHSET xml file is created. This method works well to verify that the correct path connectivity exists before the attempt is made to create the path structure with Business Continuity Manager, thereby helping to determine whether it is a hardware or software problem should an error occur when subsequently creating the Business Continuity Manager PATHSET file.

Journal Groups

All journal groups are constructed using Storage Navigator.

Beginning with Hitachi Business Continuity Manager for Mainframe Replication

Configuring Business Continuity Manager

If configuration of Business Continuity Manager is begun with one “Configuration File Prefix” (in the ‘set defaults’ panel) and then left unfinished, be sure to use the same prefix when resuming or else the XML files will not share the same naming convention and Business Continuity Manager will not function.

If Business Continuity Manager will be used to manage pairs, avoid managing them with Storage Navigator as well. Business Continuity Manager can’t always see pairs created with Storage Navigator.

Device Address Domain

Business Continuity Manager uses Device Address Domains (DAD) to identify/specify volumes that are local to a given host. Each host in the replication environment should have its own DAD ID.

Beginning with Hitachi Command Control Interface for Open Systems Replication

Use the latest version of CCI for the microcode on the storage system.

Review appropriate ECNs for the version of RAID Manager/CCI being used.

Follow implementation steps from the appropriate program product installation materials.

HORCM

- HORCM instances should be unique
- In general, use HORCM0 for P-VOLs and HORCM1 for S-VOLs. This avoids any confusion when creating pairs and avoids the likelihood of creating a pair in the wrong direction.
- Don’t reuse HORCM instance numbers on multiple hosts. For multiple CCI hosts in an environment, don’t use multiple HORCM0’s and HORCM1’s. Extend the instance naming convention to accommodate the configuration. For example:
- Host A could have HORCMA0; Host B could have HORCMB0, etc.
- Or HORCMTC0 could be used for TrueCopy S-VOLs

- Determine what is appropriate for the project.
- Add a comment to the HORCM files to document any extended naming convention

CCI Poll Values

For Thunder, set the poll(10ms) value greater than or equal to 6000 using the formula of 6000*number-of-instances using the command device.

Naming Conventions for DEV_GROUPs and DEV_NAMEs

DEV_GROUP naming should be related to the application that those volumes support

- Example: EXCHANGE1 or SQL_PAYROLL or ORA_PROD

DEV_NAMES should tie to the physical device.

- P-VOL LDEV to S-VOL LDEV, e.g., 001D_012E or 0013_0122
- It's better to not use LUN numbers in the naming convention, since the context for that LUN number may be unclear (absolute LUN? An HSD LUN? A host-specific LUN?).
- Remember that DEV_NAMEs must be unique. One DEV_NAME cannot be used in two different DEV_GROUPs.

Automation

The daily operations of the replication environment should be automated to a high degree and should notify the Operations Team when manual intervention is required, with extensive logging.

Scripting Interface

Hitachi Data Systems' replication products provide a Command Line Interface (CLI) facility Command Control Interface (CCI). Hitachi Data Systems recommends that CCI be installed on one host at each site for the entire environment - this allows for centralized scheduling and administration. The servers will communicate with each other using TCP/IP (user defined UDP ports - one per CCI server for TrueCopy).

The CCI servers will be connected to their respective storage arrays via a Fibre Channel connection and mapping to a command device. CCI communicates with the storage system through the command device. It is a user-selected, dedicated logical volume on the Universal Storage Platform that functions as the interface to the CCI software on the Windows/UNIX host. The command device is dedicated to CCI communications and cannot be used by any other applications. It accepts replication software read and write commands that are executed by the storage system. The command device also returns read requests to the Windows/UNIX host. The volume designated as the command device is used only by the storage system and is blocked from the user. The command device uses 16MB, and the remaining volume space is reserved for CCI and its utilities. It can be any OPEN-V device that is accessible by the host. A LU Size Expansion (LUSE) volume cannot be used as a command device. A volume as small as 36MB can be used as a command device. It will appear with -CM appended to its label. (Example: OPEN-V-CVS-CM)

CCI commands provide functionality for managing both TrueCopy and Universal Replicator operations. All TrueCopy and Universal Replicator software recovery functions can be controlled by scripts utilizing these commands. The scripts can be automated using '*cron*' or any desired system scheduler. It is also recommended that the scripts be written in a standard, open, and common scripting language, such as Korn Shell. It is recommended that a script be put in place to periodically verify that the replication paths are still operational.

Configuration Testing

RAID Manager/CCI testing should be performed on the completed installation using test pairs

- to verify proper solution configuration,
- to verify expected behavior from RAID Manager/CCI, and
- to protect production volumes.

Daily Operations

Application Interaction

When creating a point-in-time copy of any disk-resident data, it is valuable to ensure that as much data as possible has left the application and server’s I/O buffers and been written to disk. It is Hitachi Data Systems best practice that all I/O buffers be flushed immediately prior to suspending TrueCopy pairs. See “Application Interaction Complexity” for each reference architecture detailed in tables 1 through 8.

Disaster Recovery Testing

Recovery procedures that use replicated volumes are less complicated and involve fewer human resources than recovering from tapes, dramatically reducing the time required for recovery tests.

NOTE: During testing of the failover server recovery, all scheduled replication processes for the tested applications should be suspended; they can be rescheduled after testing is complete.

Scheduling Disaster Recovery Testing

Full Disaster Recovery tests should be scheduled once per year.

Scaling for Growth and Change

Optimization & Tuning

Initial Copy

- In a Universal Replicator Three Data Center Multi Target configuration there are TrueCopy synchronous pairs and Universal Replicator asynchronous pairs, BOTH having the same Primary volume in common.
- An initial copy of the TrueCopy Sync pairs AND the Hitachi Universal Replicator Asynchronous pairs cannot be done at the same time. One or the other must be in Duplex status before initiating the other's initial copy operation.
- The Rule Of Thumb is that the TrueCopy Synchronous software initial copy should be done first.
- Using default initial copy parameters, Hitachi Universal Replicator initial copy performance varies from ~80MB/sec to ~165MB/sec during the course of the initial copy operation regardless of whether a single LCU or multiple LCUs are being copied.

Hitachi Universal Replicator Reverse Resync

NOTE: V7+1 (50-07-69) or greater ucode is required for correct Reverse Resync operation. For Three Data Center, the minimum level is V8 (50-08-06).

- This function is performed between the SiteA Universal Storage Platform and the SiteC Universal Storage Platform using a single LCU.
- All LCU volumes are reverse suspended/reverse resynchronized using Business Continuity Manager with the copy direction being changed to SiteC being the primary and SiteA becoming the secondary site.
- The reversed pairs are then suspended normally.
- The "primary" volumes are varied online to the SiteC host.
- Data may be created on any of the Primary volumes.
- The pairs are then resumed and the data is automatically resynchronized (copied) to the "secondary" volumes in SiteA. This is how it is supposed to work.
- The pairs are then suspended forward/resynchronized forward, returning the copy direction to its original "normal" direction of primary in SiteA to secondary in SiteC.

Appendix A References

Hitachi Data Systems documentation:

- *Hitachi TrueCopy™ Remote Replication Users Guide*
- *Universal Storage Platform user documentation*
- *Command Control Interface (CCI) User and Reference Guide*

IBM® documentation:

- *Planning for IBM Remote Copy*, SG24-2595
- *DFSMS MVS V1 Remote Copy Guide and Reference*, SC35-0169
- *OS/390 Advanced Copy Services*, SC35-0395 (replaces *Advanced Copy Services*, SC35-0355)

Appendix B Glossary

RCP&D: Remote Copy Planning and Design – a service delivered by Hitachi Data Systems Global Solutions Services to assess the current environment considered for replication, establish the business and technical requirements for the implementation, and propose a high-level solution design to meet established requirements

RPO: Recovery Point Objective – the point to which data must be recovered after an interruption or disaster

RTO: Recovery Time Objective – the length of time for which data and services can remain unavailable

Hitachi Data Systems

Corporate Headquarters
750 Central Expressway
Santa Clara, California 95050-2627
U.S.A.
Phone: 1 408 970 1000
www.hds.com
info@hds.com

Asia Pacific and Americas
750 Central Expressway
Santa Clara, California 95050-2627
U.S.A.
Phone: 1 408 970 1000
info@hds.com

Europe Headquarters
Sefton Park
Stoke Poges
Buckinghamshire SL2 4HD
United Kingdom
Phone: + 44 (0)1753 618000
info.eu@hds.com

TM-002-00 February 2007