HITACHI
Inspire the Next

Best Practices Guide for Disaster Recovery Replication

HDS CONFIDENTIAL

March 2010

Hitachi Storage Solutions

Hitachi Data Systems

Contents

Introduction

This document has been written to assist in the planning, design, and implementation of the replication and recovery process for the Hitachi High-performance NAS platform. The Hitachi High-performance NAS platform offers a policy−based file system−level incremental replication process to establish asynchronous copies of file systems, virtual volumes, directories, and under limited circumstances snapshots. When a primary source system undergoes failure and is no longer available, the replicated copy can be brought online at the target destination to continue service and reestablish business operations.

Note: There are multiple ways to copy data using various commands and utilities such as the accelerated data copy (ADC) client utility, manual use of data protection options in the System Management Unit (SMU), and the Transfer Primary Access (TPA) feature. Although these processes replicate data to make copies, they are not considered part of a scheduled policy−based replication process that establishes ongoing incremental copies of data between two sites with the appropriate disaster recovery steps and a state−based recovery process in place.

Note: There is a unique block−based replication capability called remote volume mirroring (RVM) that is specific to the implementation of LSI storage. The use of this option will be covered in a separate paper. This option provides block−based replication between storage arrays across a backend SAN infrastructure.

Overview of the process

The process includes:

- Identify and organize important information assets that need to be protected and replicated as part of a disaster recovery plan.
- Determine a rough approximation of data set size and rate of change associated with each individual policy setting to evaluate a sustainable operation.

  Note: This step might require onsite testing prior to production to validate distance, latency, bandwidth, scripts, and performance capabilities of the system.
- Determine recovery point objectives (RPOs) to establish frequency of replication and associated potential data loss up to that recovery point.
- Establish recovery time objectives (RTOs) and procedures depending on the mode of failure preceding the recovery operations.

Document the recovery process for future reference, including all process steps, commands, account permissions, events, and notification settings.

Recovery Point Objective

The recovery point objective describes a point in time to which data must be restored in order to be acceptable. This is often thought of as the time between the last available backup and the time a replication will potentially occur. The RPO is established based on tolerance for loss of data or reentering of data.

It is a function of the replication software and the bandwidth between sites. The RPO in conjunction with the RTO is the basis on which data protection strategy is developed.

In some cases, hot backup mode could be needed for some applications and databases.

This information helps to determine the replication method: synchronous or asynchronous.

Recovery Time Objective

The *recovery time objective* is the boundary of time and service level within which the data replication process must be accomplished to avoid unacceptable consequences associated with a break in continuity.

The RTO is established during the business impact analysis by the customer (usually in conjunction with Hitachi Data Systems personnel). The RTOs are then presented to senior management for acceptance.

It should be noted that the RTO attaches to the business process and not the resources required to support the process; for instance:

- Data has to be available
- Servers and applications have to be up
- The network has to be up and ports open for application access

The RTO and the results of the impact analysis in its entirety provide the basis for identifying and analyzing viable strategies for inclusion in the DR plan. Viable strategy options would include any which would enable resumption of a business process in a time frame at or near the RTO. This helps determine the professional services required and includes alternate workarounds.

States of Data

The states of data include:

- Runable: synchronous
- Restartable: asynchronous, file system consistent images
- Recoverable: backups
- Data corruption (scrambled eggs)

Bandwidth

When evaluating the site's bandwidth:

- Never assume the customer has as much WAN bandwidth as they believe to have
- Verify with network reports that display total bandwidth and utilization
- Use FTP to test for bandwidth, ensuing you run the test throughout the day to measure bandwidth utilization changes
- On most networks, the WAN is busiest first thing in the morning, just before lunch, and just before quitting time
- Check latency, as it is a large factor of bandwidth: the more latency, the less total available bandwidth

Replication Overview

In general, replication is the copying of information to multiple systems in such a way that the information is consistent across those systems. There are two basic types of replication:

- Synchronous
- Asynchronous

Synchronous Replication

Synchronous replication is commonly hardware and storage specific. It uses the two−phase commit process available with most relational database management system (RDBMS) products. In a two−phase commit environment, when an update to the master database is processed, the master system connects to all other systems (slave databases) that require the update, locks those databases at the record level, and then updates them simultaneously. If any connection to another system is not available, the update is rejected.

Synchronous replication has a distance limit, which varies with each vendor. Most customers using synchronous replication use storage vendors such as EMC, HDS, IBM, and so forth.

The cost is generally twice the cost of the hardware, plus months of professional services to analyze and deploy.

When using synchronous replication, the most important requirements are bandwidth, latency, and application response time. The pro using synchronous replication is zero RPO with runable data; cons include:

- Hardware−specific
- Expensive
- Difficult to deploy
- Hard to maintain
- High bandwidth requirements

Asynchronous Replication

Asynchronous replication can be deployed by either storage or software vendors. Most storage vendors have the feature built in. Software vendor products−−such as Veritas Volume Replicator (VVR), NSI Double−Take, EMC RepliStor, and Network Appliance ReplicatorX (formerly a Topio product), to name a few−−are integrated with applications to perform replication.

Generally, it uses less bandwidth and there are no application performance problems due to latency.

The Hitachi Data Systems replication solution uses network data management protocol (NDMP) to replicate data. By default, it uses file system snapshots for replication of data sets. Hitachi NAS Patforms can replicate entire file systems. In addition, it can select the folder to replicate, including virtual volumes.

Incremental data replication (IDR) allows system administrators to setup a scheduled, incremental backup of a volume or a directory on the volume. Multiple schedules can be defined on a per server basis with support for pre and post scripting, enabling automated functions to occur prior and post the IDR schedule.

IDR uses snapshots as a basis for replication, maintaining the last snapshot as the reference point for the next replication to occur. Using snapshots as a means for replication also ensures files that would otherwise be skipped, as they are in use, are replicated and protected.

Using snapshots for replication also ensures that only changed files are replicated, drastically reducing the amount of replication traffic. The Hitachi High-performance NAS platform system can support inter−Hitachi High-performance NAS platform server replication (local or remote) as well as intra− Hitachi High-performance NAS platform server replication, using Hitachi NAS Platforms' unique tiered storage.

Incremental block replication (IBR) replicates only the changed block of a file. By default, it only replicates files greater than 32 MB in size. The best use case scenario is with databases.

IBR uses a schedule for replication. It can be called a *continuous data replicator*. By setting the schedule to replicate the data blocks, it moves to the next block as each replication schedule finishes.

The con using asynchronous replication is high RPO (more than a minute); pros include:

- Inexpensive
- No distance limitation
- Easy to deploy (no drive speed matching)
- Lower bandwidth requirements

Note: Filtering NDMP events is recommended when resolving any problems, as all start and stop events are recorded.

See the following section for Hitachi Data Systems−specific services.

Hitachi Data Systems Replication Services

Hitachi High-performance NAS Platform Replication Process Overview

There are several file−level replication capabilities offered on the Hitachi High-performance NAS platform, each of which is briefly described in the following sections. All of these replication processes leverage the inherent snapshot and NDMP (backup and recovery) capabilities. The replication services can be used to transfer entire file systems, directories, or virtual volumes. It cannot be used to transfer storage pools, storage volumes, or RAID sets. The replication process preserves the required user and group file system permissions and quota information. The process does not preserve any share or export information. This part of the configuration must be manually re−created after a replication process is complete; the process is explained in the recovery section.

Although the Hitachi High-performance NAS platform server supports single full copies and a variety of replication−based backup options, the emphasis here is on using replication to incrementally transfer data between two systems over time between a primary source system and a secondary target destination. The target system is usually in standby waiting for recovery should the primary system become unavailable for any reason. The concatenated scheduling of a backup job for data protection after a replication process on the target system is possible.

Accelerated Data Copy Client Service

A standard Hitachi Data Systems Storage System feature, the accelerated data copy client is a stand−alone network data management protocol (NDMP) client utility that can be used to manage operations on NDMP servers running on the Hitachi High-performance NAS platform server. The ADC capability can be used for data protection services associated with replication and tape backup. File system copies can be made between file systems on different Hitachi High-performance NAS platform servers and between file systems within a single Hitachi High-performance NAS platform server (inter and intra respectively). ADC uses a simple client piece of software with a very simple command line interface that produces a simple log of events and data transferred.

Important: The ADC client is not to be used to set up replication and recovery process on Hitachi High-performance NAS platform servers and is usually reserved only for Hitachi Data Systems support, service and engineering personnel. Originally, it was designed as a command line interface NDMP client to test NDMP request for backup to tape resources. For most general purposes the following section on Policy Based Intelligent Data Replication will be the primary interface for customers to configure and operate replication.

Incremental Data Replication Service

Incremental data replication is the preferred method of replication for disaster recovery configurations. Typically, a disaster recovery configuration represents two physically separate nodes that are remotely located several miles (km) from each other for disaster recovery purposes. IDC extends the capabilities of ADC and includes a number of additional software capabilities that helps to ensure replication between two nodes is coordinated and understood relative to various node failures, failover, and fail−back scenarios. The Hitachi High-performance NAS platform server supports policy−based asynchronous full and incremental file system−based data replication processes. This allows administrators to setup and configure scheduled replication jobs independent from or coordinated with other backup strategies. The configuration of retention of these policies is only through the use of a single SMU. The IDR policies can be configured to perform continuous incremental, periodic incremental, and individual single full complete data replications.

When a replication policy is first set up, the SMU performs an initial full copy of the source file system (or directory) to a target destination. After the initial copy is successful, incremental copies are performed at the scheduled intervals. IDR replicates all files modified since the last replication transfer. IDR allows a number of filtering capabilities to limit which files are targeted for the replication process.

Incremental Block Replication Service

Incremental block replication is an optional licensed feature that is used in conjunction with IDR. IBR allows partial file transfers instead of full file transfers to reduce the overall amount of data that needs to be replicated between sites. This feature is run under the same policies and schedules as IDR, but allows for specific settings related to large file sizes. The primary setting is associated with a file size threshold.

Incremental block−level replication can optionally be used to replicate large files more efficiently. During a standard incremental data replication without IBR, files that have been changed since the last scheduled replication are replicated to the target in full. When using the optional IBR capability, the incremental block−level replication only transfers the block−level changes within individual files and not the whole file, greatly reducing the amount of data replicated. This option was designed to reduce the overall replication time and can increase performance.

IBR also is good for large databases applications. Specific file sets with regions that have a high rate of change can be targeted. This also applies to iSCSI LUNs, which look like a large file to the system. The iSCSI LUNs can be large, so if the LUN is new or mostly empty the IBR function will help to dramatically reduce the transfer load and time.

The Underlying Snapshot and NDMP Background

The Hitachi High-performance NAS platform server uses snapshot and NDMP technologies to transfer copies of data within servers and between servers. The *snapshot* is a time consistent image of a file system that is backed up using a proprietary NDMP stream, which is transferred and restored on another file system in the remote node, resulting in an exact replica on the destination target. The NDMP transfer is a file level data transfer process. The transfer uses the same Ethernet ports as the data servicing NFS exports and CIFS shares to clients on the network. For this reason the NDMP stream will govern its workflow based on overall system performance to insure it does not have a detrimental impact on the primary NAS data services.

After successful completion of a replication, the NDMP engine automatically takes a snapshot on the target. This supports various features such as:

- A replication policy (optionally) that deletes files on the target which no longer appear on the source
- Rolling back an incomplete replication

Replication Implementation Process

Reference Materials and Documentation

The "Policy−Based Data Replication" section in Chapter 7, "Data Protection," of the *Hitachi High-performance NAS platform server Systems Administration Guide* is an excellent source of information associated with the rules, policies, and setup options for managing the replication services. There are also excellent screen shots within the guide.

Replication Configuration

Replication policies and schedules are configured and stored on the systems management unit, which provides a web user interface to simplify administration. Configuring a policy−based replication process requires the administrator to set up the following processes:

- The *Replication policy* identifies the data source, the replication target, and optionally the replication rule. Pre−replication and post−replication scripts can also be set up on the policy page.
- *Replication rules* are optional configuration parameters that allow specific functions to be enabled (or disabled) or achieve optimal performance. Most of these are filtering functions that impact the transfer payload to optimize the duration of replication processes.
- *Replication schedules* define the schedule, timing, and policy based on the scheduled data and time.

To locate the IDR function on the SMU, choose Data Protection. Replication management provides for the creation of policies, schedules, and rules combined with the ability to monitor status and view reports.

Replication policies are individually defined as a one−way transfer from point−to−point between a source to a target destination pair. A target for one replication policy could very well be a source for another replication policy.

The available sources for replication are defined as follows:

- A file system
- A directory
- A virtual volume*
- A snapshot

The available targets for replication are defined as follows:

- A file system
- A directory
- A virtual volume*

* Virtual volumes are not supported for unmanaged replication nodes, which are typically the case when two sites in excess of 500 meters from each other are used and are not recommended for disaster recovery scenarios and design.

Although a SMU schedules and starts all replications, data being replicated flows directly from the source to target via the NDMP data stream without passing through the SMU.

Important Note: Multiple replication policies can be established that run simultaneously. For example, a single Hitachi High-performance NAS platform server can be the destination target for several independent source

Hitachi High-performance NAS platform servers. In general, it is best to avoid overlap of replication policies that use the same file system, virtual volume, directory, or snapshot because changes to one of these systems could impact the other. This also includes the impact of backup operations because of the use of NDMP in the backup process could interfere with the NDMP used in the replication process if they are sharing access to resources. So timing and schedules that use the same resources should be design to not interfere from a scheduling standpoint.

Distances and Performance

The distance can be both short (within a single data center) and long (between remote data centers). A key design element is that typically two source and target systems configured for disaster recovery replication have independent backend SAN storage pools and independent management SMUs. The SMU for the source typically is configured for managing transfers that originate from the source to a destination target. The destination target typically has its own management SMU, but the source SMU must be able to see the Ethernet data ports on the target SMU. Conversely, the SMU at the target is used to schedule jobs originating on the target (which now becomes a source). This is called a *managed* or *unmanaged server* (detailed in the following section).

The maximum distance is regulated by the overall delay of the network system, which is a maximum of 15 minutes. This component of the configuration will need to be test. The first transfer is typically the largest and requires a full complete transfer. If the initial full transfer is too large, alternative methods could be tape or just making a one−time copy. The performance of the transfer depends on the network connectivity between the two locations.

Note: Replication within a cluster is uncommon because clusters must share access to the same backend SAN storage subsystem and they recovery file systems and data in a method independent of replication processes. A better example of replication between two clusters would be the use that allows one cluster to protect the assets of another.

Managed and Unmanaged Hitachi High-performance NAS platform servers

The *system management server* manages multiple Hitachi High-performance NAS platform servers. Replications can be configured between Hitachi High-performance NAS platform servers that are both managed and not managed by the same SMU. In this case, use the source NAS platform and SMU to set up the replication policy for data originating on the source. When an unmanaged system is used, additional information is required. This includes the IP address, user name, and password for the backup operator group.

Restriction: The SMU must have all the data ports visible on the same network at both the source and target systems. Essentially, partitioned networks are not allowed in replication configurations. If the SMU cannot see both ends of the replication system then the replication will not start and might fail to complete.

Best Practice: Two SMUs should be implemented, one at each site. They can be configured to provide standby backup services to each other should one fail. Each SMU can manage the local Hitachi High-performance NAS platform server in the respective sites. This requires the unmanaged Hitachi High-performance NAS platform server option to be implemented.

Best Practice: In general, use the SMU attached to the source Hitachi High-performance NAS platform server to configure the replication policy related to the source of the data. A source originates data streams, a target receives them. If two sites are replicating to each other, set up the corresponding policies on each side with independent file systems or directories dedicated and designed to avoid potential overlap.

Replication Policies

To add a replication policy:

1. From the SMU Home page, choose **Data Protection > Replication**; in the Policy area, select **add.**
2. Check **Not a managed server**.

   You need both the NDMP user name and password on the remote destination server.
3. Ensure you are on the correct Hitachi High-performance NAS platform server or cluster (source), and on the next page, enter the following:
   - Enter a name for the replication policy.
   - Select the source EVS or file system.

     If needed, determine the path to a specific directory.
   - Select the target EVS or file system.

     If needed determine the path to a specific directory.

     Note: In the case of an unmanaged target, virtual volumes cannot be specified and the NDMP user name, password, and IP address are needed.

     Note: For the processing options:

     - Keep the snapshot rules set to the default: None; the replication engine creates and deletes snapshots automatically on this setting.
     - Replication scripts are only for application or database replications that are used to quince the I/O, take a snapshot, and then restart the I/O. This is beyond the scope of this document. Contact Professional Services for more information.
   - Assign replication rules as needed.

     This assumes previously existing replication rules have been defined.
   - Ensure the destination location is capacious enough to hold the source data set being replicated and the incremental changes.

     As a best practice, the two file systems should be equal in size.

Replication Rules

Replication rules are optional configuration parameters that allow replications to be tuned to enable or disable specific functions or to achieve optimal performance.

To add a rule:

1. From the SMU Home page, choose **Data Protection > Replication**.
2. Click **add new rule.**
3. Enter a name and description for the new rule.
4. When setting rule definition:

- Use the defaults for the simplest approach.
- Enter the exclusion of files or directories as needed.
- Use Block Replication, if licensed. This option cuts down on data transferred to increase performance. It requires separately licensed software. The default is enabled, which chooses a 32 MB file size.
- The changed directory list default is disabled. This helps to track changes and speed replication activities. It is an internal function to the file system. Call Hitachi Data Systems Support if you are considering this option.
- Keep the Number of Read Ahead Processes on default: 1. The range is 0 to 32..
- Keep the Pause option on default: Yes.
- Keep the Take a Snapshot option on default: Yes.
- Set the Delete the Snapshot option to Last for incremental replications. The default is Last, which is typical for disaster recovery.
- Keep Migrated File Exclusion set to Disable, unless you have a spread of data across a migration that needs to be replicated. The default is disable.
- Keep the Migrated File Remigration set to Disable, unless you establish a migration apability on the destination target.The default is disable.

.

Replication Schedules

Replications can be scheduled and rescheduled at any time and with any of the following scheduling options:

- *Periodic replication* occurs at preset times, which can be set to run daily, weekly, monthly, or at intervals specified in number of hours or days.
- *Continuous replication* starts a new replication job after the previous one has ended. The new replication job can start immediately or after a specified number of hours.
- *One−time replication* runs once at a specific time. This is a single full complete replication.

Best Practice: Do a continuous replication that starts the next replication process immediately after the prior one to reduce the recovery point to objective, which minimizes any potential data loss.

Note: Continuous replication is the best practice for replication for disaster recovery.

Scheduled Periodic and Continuous Incremental Replications

The replication operation consists of an initial backup snapshot of the source file system, which produces an NDMP backup data stream that is routed directly to the destination system, in which it is continuously recovered. This is similar to an ordinary backup and recovery operation, except the data never makes it to tape.

Note: Replication require NDMP version 3 or 4 to run. Setting the protocol version to 2 prevents these replications from running. Version 2 is for backup operations only.

Other Performance Considerations

The performance of the replication process is variable and depends primarily on the amount of data being transferred. The amount of data being transferred depends on rate of change of the number of files in the file

system being transferred. It is also important to consider the number of simultaneous replication processes running on the same system. The replication process is file based and not designed for large data transfers or multi−terabyte data set transfers. Transfer rates can be achieved as high as 50 MB/sec. All transfers are completed across the TCP/IP connections on the Ethernet data ports.

Recommended Server Settings

The speed of recovery is important, depending the business or application access. In order to prepare a system for the fastest level of recovery, use the following recommendations.

1. Identify share names, share paths, and file systems on both the source and destination Hitachi High-performance NAS platform servers.
2. Ensure enough space or same size file systems are used on both the source and destination Hitachi High-performance NAS platform servers.
3. If multiple sources replicate to a common target, dedicate a file system for each source on the target to allow separate file system rollback and recovery processes, associated only with the sources that fail.

Disaster Recovery Process

This recovery process is related to establishing access to data from either a source or target node, depending on failure conditions.

Recovering from a Replication Failure

The system administrator can deal with replication failures as follows:

- Restart the replication from where it left off
- Allow the next replication to continue
- Rollback the replication to the snapshot taken after the last successful replication

The appropriate action depends on the situation. Restarting the replication is the preferred option when replications do not occur that often. However, if replications are performed continuously (or frequently enough), a new replication may start before the user has a chance to restart the old one, in which case the second option occurs automatically. In disaster recovery (DR) situations, rollback is the only option available, as a failure in the source server may make continuing the replication impossible.

Replication rollback is a licensed feature.

Primary failure scenarios include:

- Replication fails; source node available
- Replication fails; source node not available

Important: Shares and export data is not replicated and needs to be copied periodically or recreated, as needed. This is a manual process and takes time.

Replication Recovery Mode Definitions

During a recovery that involves rollback, the replication source and destination operate in a number of different modes. These modes control possible actions. Destination modes include:

- NORMAL−−the data on the destination must not be updated except by the replication process itself (for example, the destination should be treated as read−only as far as network users are concerned). Normal replications can take place as long as the source mode also is NORMAL.
- ROLLBACK−−the destination is in the process of being rolled back to the last good copy. After the rollback has started, it must complete before any further action can be taken. If a rollback action fails, it must be restarted. The destination should still be treated as read−only, and cannot be enabled as the working store until rollback has completed.
- DECOUPLED−−a rollback has completed on the destination. The destination can be brought online in place of the source and used as the working store. Start recovery replications in this mode.
- COPYBACK−−enter this mode when the replication is being restored by copying changes back to the source file system. The destination should be treated as read−only.

Source modes include:

- NORMAL−−the source data may be updated. Normal replications can take place as long as the destination mode also is NORMAL.
- SUSPENDED−−the source enters this mode when a destination rollback starts. The source data should not be updated in this state.
- ROLLBACK−−the source is in the process of being rolled back to the last good copy. Once the rollback has started, it must complete before any further action can be taken. If a rollback action fails, it must be restarted. The source should still be treated as read−only (it must not be changed except byte replication software until recovery process has completed).
- SYNCHRONIZED−−a rollback has completed on the source. The source is now ready for restoring data from the destination.
- COPYBACK−−enter this mode when the source is being restored by copying changes back to the source system. Both the source and the destination should be treated as read−only.

Resume Source

In most circumstances, if a replication fails, the first and best practice approach is to resume the replication process. This assumes that the replication process was interrupted, but that the primary node providing file services is still up and running or was only temporarily down. Essentially, it is still the primary source of the data under protection and remains the source for that data while the replication process is resuming from a failure.

Note: If the source system is close to being recovered, then you may want to delay activation of the target node, as it is easier to recover the replication if the destination has not been modified.

If a replication fails and the source node recovers quickly, the following best practices are recommended:

- Resume is the default best practice.
- Wait until the next increment occurs in the schedule. It will automatically adjust to the difference in snapshots and capture all the correct changes.
- Resume the replication, which will automatically try to run the replication again using the same snapshot. The node is brought back online, assuming a temporary outage.

Assuming a partial transfer occurred; the latter two options will roll back the data on the target node prior to resuming or starting the next replication. See the "Restarting Replications" section for more information.

Activate Target

If there is a problem with the system used as a replication source, such that it becomes unavailable for some time, disaster recovery (DR) actions are needed. When this occurs, the first action is to bring the destination online as the working store.

Case A: Replication Immediately Interrupted, Partial Transfer, and Using Rollback for Disaster Recovery

Replications set up for disaster recovery purposes typically occurs frequently, each starting soon after the last one completed. As a result, if the source file system is to fail, the failure is likely to occur in the middle of a replication. Though the target file system will be in a consistent state, the data might be inconsistent because only part of it has been copied to the target. As this can be unacceptable for some customer applications, the Hitachi High-performance NAS platform servers provides a way to *rollback* the target file system to its state prior to the start of the replication.

Replication rollback should only be used when the destination is going to take over (at least temporarily) as the primary file system. The NDMP engine performs file system rollbacks using snapshots taken automatically on the target file system after the successful completion of the previous replication. The file system rollback is started from the SMU. The actions can be described as *undoing a replication*, with files being copied from a prereplication snapshot to the live file system. The rollback only affects directories that are part of the replication (for instance, if the file system is used as the target for multiple replication sources, replications from other sources are not affected by the rollback operation). Any additional steps required for making the target file system accessible to users must be performed by the system administrator.

As a best practice, the intent is to keep the consistency of two copies between the two locations. This way the customers can always rollback to this consistent state.

If the SMU in the source site is down, which controlled the replication policies, the SMU in the remote site for the target systems will need to be used to roll back the file system using the last known good snapshot. Rolling back the snapshot should put the status of the replication into a DECOUPLED state.

Rolling Back an Incomplete Replication

Replication rollback is performed from the SMU. If the replication fails, the Modify Schedule page displays as follows:

Click the **restart** button to restart the replication from its last checkpoint. Click the **rollback** button to restore the replication target contents to the snapshot taken after the last replication successfully completed.

Replication rollback is the first step in recovering from a replication failure in a DR situation. Additional steps are typically required to make the target file system accessible to network users and to restart replications should the source come back online.

Case B: Replication Completed and Source Failed

If a consistent state exists between the source and target systems, and the source failed in between replications, then there is no need to rollback the file system. Proceed to the next section, "Bringing the Replication Target Online."

Because the file system is incrementally recovered during the replication process, it is immediately available after a replication process has completed.

Bringing the Replication Target Online

After the replication source fails and the target file system rolls back, it might be necessary to bring the replication target online to replace the source. To bring the target online (performed manually by the system administrator):

1. Bring the target file system online to replace the source file system. It is necessary to create equivalent shares, exports, and other file system−related configuration settings on the target file system to mimic those of the source file system.
2. Instruct network clients to connect to the replication target. Typically, this requires different EVS names or IP address.
   - DNS changes: Change the IP address of the source Hitachi High-performance NAS platform servers resource record to the IP address of the recovery target Hitachi High-performance NAS platform servers destination.
   - On a single client, at the command prompt, enter arp −d *.
   - Do an NSLOOKUP using the primary Hitachi High-performance NAS platform servers host name to test that the name resolves to the disaster recovery Hitachi High-performance NAS platform server's IP address.
   - Do an NSLOOKUP using the disaster recovery Hitachi High-performance NAS platform servers IP address to test that the IP address resolves to the primary Hitachi High-performance NAS platform servers host name.
   - If the name and host resolution work correctly, all clients will need to have their ARP cache cleared. This can be accomplished by rebooting all clients are running the arp -d* command locally on each system. It is recommended to set up a batch file to script this for allhost using remote execution scripts on the clients.

Restarting Replications

If the source system or the source data are not recoverable, then replication of the working copy of the data (now located on the destination) must be set up from scratch. However, if the source system comes back online and the source data is still intact, then it might save time to restart replications as described in this section.

A script is provided to help in restarting replications. It is run from a command line prompt on the SMU. To run the script, the user must be logged in as root and in `/usr/local/adc_replic` directory. It is invoked as follows:

`sh replication_recovery.sh` *`command policy_name`* `[`*`schedule_id`*`]`

As the recovery takes place, the source and destination take on different modes of operation, which affect what actions can be taken. The commands issued from the **replication_recovery** script are used to switch from one mode to another, with the ultimate goal of restarting replications (and possibly restoring access to the original source). The following replication recovery commands are supported:

- STATUS finds the recovery status
- REENABLE restores replications to the source (before the destination is activated)
- SYNCHRONIZE performs a rollback operation on the source (to resynchronize)
- REVERSE switches roles (so that the original source now becomes the destination)

- COPYBACK copies data from the destination back to the source

The usual recovery sequence and use of these commands are as follows:

1. When the source system goes offline and appears that it might not be recoverable in a short period of time, then a destination rollback should be started to prepare the destination to take over in place of the source. This rollback action can take a few moments. The source mode is changed to SUSPENDED and destination to ROLLBACK. If the rollback fails, it must be rerun.

2. If the source system becomes available while the destination rollback is happening or before the destination has been enabled as the primary store, then the replication can be restarted by using the REENABLE replication recovery command. However, the destination rollback must complete before the replication can be restarted. The source mode is switched to NORMAL.

3. If the source system is not available by the time the destination rollback has completed, then the destination can take over as the primary working store. The destination mode is set to DECOUPLED when the rollback completes..

4. If the source system and data become available after the destination has been used as working store, then the source data must be synchronized before replications can be restarted. This synchronization process is invoked using the SYNCHRONIZE recovery command. This command initiates a rollback of the source data to the last snapshot copied by the replication. The source mode is set to ROLLBACK until the synchronization action completes. If the rollback fails, it must be rerun by reissuing the SYNCHRONIZE recovery command.

5. If the destination has taken over as the working store and the source is in SYNCHRONIZED mode, then a replication recovery can be initiated. The shortest and easiest way to restore replications is to reverse the direction of the replication with the original destination taking over the source role. If the original roles must be retained, then the changes made on the destination must be copied back to the source:

   – When the destination mode is DECOUPLED and source mode is SYNCHRONIZED, the REVERSE command swaps the roles in the replication and the replication scheduling is reenabled.

   – If the destination mode is DECOUPLED and source mode is SYNCHRONIZED then a COPYBACK command can be issued to copy changes on the destination back to the source.

     (Note that both source and destination data must be treated as read−only during this operation.)

Note: A file system rollback can take a considerable amount of time to finish.

Addendum

This addendum includes:

- Using the NDMP Engine for File System Rollback
- Frequently Asked Questions
- Replication Checklist

Using the NDMP Engine for File System Rollback

In order to perform a file system rollback, the NDMP engine *replicates* the contents of a snapshot back to the live file system. The following considerations apply:

- File system rollback is performed by copying files one at a time. As a result, it might take a few moments.
- Block level operations are not used during a file system rollback.
- If a rollback operation is interrupted prior to completion, it cannot be restarted. However, it is possible to perform a second rollback operation without losing any information.
- Quotas are not rolled back by the NDMP engine. If changes are made to the quota limit configuration as part of the failed replication, then these changes are not reversed by the rollback operation. (Because presumably the changes were to be applied anyway, it is usually expected that this is the desired result.) If the quota limit configuration needs to be changed back to its prereplication state, then the changes must be reversed manually.

Configuring target snapshots

By default, this is a default NDMP snapshot. However, if a destination snapshot rule has been defined in the replication policy, the snapshot rule is used instead. The destination snapshot rule is specified in the Add Replication Policy or Modify Replication Policy SMU pages, as follows:

Creation of the destination snapshot is suppressed if the user selects snapshot parameters that preclude incremental working.

Forcing Full Replications

When forcing a full replication:

- You may want to wipe out the target directory; otherwise, you risk keeping old and deleted files on the target.
- Delete the last source snapshot used, or delete and recreate the policy

To force a full replication every time, create a replication rule in which "Take a Snapshot" is set to No. Use this rule in the replication policy.

During a replication, if the Hitachi High-performance NAS platform server has an EVS with more than one IP address defined, it might be necessary to manually override the replication IP addresses.

To override the replication IP addresses:

1. Ensure that the SMU is running release 4.0.417j, or later.
2. Edit */usr/local/adc_replic/address_overrides*.

   This should create a new file.
3. Add the override definition in the format *override*_*this_address=with_this_address*.

   For example, the file should appear similar to the following:

   ```
   $ cat /usr/local/adc_replic/address_overrides

   10.128.1.202=192.168.7.206
   ```

Frequently Asked Questions

Are virtual volumes replicated (for instance, to the destination)?

Yes, but only if we are creating a directory that is the root of the virtual volume or if the directory is empty at the destination. Also, the virtual volume name must not exist on the destination volume.

Does SMU unconfig or SMU restore affect replications?

Not in release 4.2 or later; for instance, after a restore, replications continue to be incremental and a full replication is not necessary. Only the policies on the SMU are affected. The replication snapshots and status remain on the Hitachi High-performance NAS platform server.

How are snapshots used on the source?

If the policy does not have a snapshot rule name selected, then when the replication begins, an NDMP_AUTO_SNAPSHOT is taken. This will be purged according to the system's backup snapshot options (specifically, the maximum auto−snapshot retention time value). The default is 7 days; the maximum value is 40 days.

If the policy has a specific snapshot rule name selected, then the most recent snapshot with that name is used. See the *Hitachi High-performance NAS Platform System Administration Guide* for more information.

Will the post−replication script be run if the replication fails?

No, the replication must complete successfully before the post−replication will run.

Does replication copy all blocks of a sparse file (for instance, iSCSI LUNs)?

No, if you create a new, empty 50 GB iSCSI LUN and replicate it, only a couple of bytes need to be transferred.

What is the difference between run now and restart buttons?

Restart attempts to pick up where it left off prior to the previous replication failure or where it was aborted.The following also applies:

- Does not run the prereplication script
- Does not take an automatic source snapshot
- Appends to the previous replication report

Run now starts the replication as though it has been scheduled. The following also applies:

- Runs the prereplication script
- Takes a source snapshot, if appropriate
- Creates a new replication report

When can I restart a replication?

The restart button is available when a replication fails or is aborted (for instance, when the replication status is "failed"). The previous replication must have actually started. If the target was unmounted, it never started.

The source file system snapshot must be enabled (this is the default, but it is possible to disable snapshot usage); for example, replication must be copying from a snapshot.

Does the restart feature work with ADC as well?

There are parameters that can be added to the ADC parameter file to cause it to take checkpoints. If the transfer fails, a *restart* parameter file is left, which can be sued to restart the operation.

How can I replicate from a specific snapshot on the source (for example, a snapshot taken 2 weeks ago)?

Create a policy in which the source path directory points to the exact snapshot directory (for example, */.snapshot/specificSnap002/root/dir1/dir2/*).

Note: when adding a policy in releases before 4.2, snapshot directories are not displayed by default when browsing the available paths. To display the snapshot directories, enter the **/.snapshot/** path and click **browse**.

Replication does not attempt to take an automatic snapshot because it is already replicating from a snapshot.

In nonclustered, two Hitachi High-performance NAS platform server configurations, how does the primary Hitachi High-performance NAS platform server handle replication and failover (for instance, can the NFS and CIFS failover to the data replication facility)?

Failover is not automatic in this scenario. Hitachi High-performance NAS Platform does not provide automatic failover in the case of synchronous or asynchronous replication. Any data replication scenario involving replication is going to require administrator action to bring the data replication site online.

How can I take a data backup that needs replication, restore it on a remote site, and then set up replication to use that data and continue from that point?

There is no clean method of performing this task at this time; however, the workaround involves the following:

1. Set up two Hitachi High-performance NAS platform servers locally.
2. Create an EVS on each Hitachi High-performance NAS platform server.
3. Place the source file system on the first server and the target on the second.
   These should be in a server farm; that is, they should see the same storage.
4. Create a replication policy from the source to the *unmanaged* target.
   The unmanaged target is the second server.
5. Allow the initial replication to occur.
6. Relocate the target Hitachi High-performance NAS platform server and its storage offsite.
7. Once the target server reaches the destination, change the replication policy to update the target IP of the replication.
   The next replication will be an incremental replication.

Replication Checklist

- The checklist includes:
- What is the RPO requirement?
- What is the RTO requirement?

- What Hitachi Data Systems replication product are they going to use?
- How much total data do they want to protect?
- What are the file systems they want to protect?
- What is the daily data change rate?
- What is the expected amount of data (during each replication cycle)?
- Do you require rolling site protection and for how long at the data replication site?
- What is the WAN link speed?
- How much of the WAN is available for replication?
- Use Iperf (Google search: iperf)
- What does the WAN latency like?
- Do they have a WAN accelerator? If so, which one?
- What kinds of data are to be replicated? (Open files (databases), user shares (common home directories), vertical data sets (add for each vertical), and so forth.)
- How much data is there?
- Where is the data located: file system or LUN?
- What is the bandwidth between sites?
- What is the distance between sites?
- What is the latency between sites?
- Can you use Iperf, which is supported on most open−system operating systems?

Hitachi Data Systems Corporation

Corporate Headquarters 750 Central Expressway, Santa Clara, California 95050-2627 USA
Contact Information: + 1 408 970 1000 www.hds.com / info@hds.com

Asia Pacific and Americas 750 Central Expressway, Santa Clara, California 95050-2627 USA
Contact Information: + 1 408 970 1000 www.hds.com / info@hds.com

Europe Headquarters Sefton Park, Stoke Poges, Buckinghamshire SL2 4HD United Kingdom
Contact Information: + 44 (0) 1753 618000 www.hds.com / info.emea@hds.com

March 2010