HITACHI
Inspire the Next

Hitachi Data Systems

HNAS Replication Best Practices Guide

Hitachi Data Systems

MK-92HNAS070-00

Hitachi, Ltd., reserves the right to make changes to this document at any time without notice and assumes no responsibility for its use. This document contains the most current information available at the time of publication. When new or revised information becomes available, this entire document will be updated and distributed to all registered users.

Some of the features described in this document might not be currently available. Refer to the most recent product announcement for information about feature and product availability, or contact Hitachi Data Systems Corporation at https://portal.hds.com.

Notice: Hitachi, Ltd., products and services can be ordered only under the terms and conditions of the applicable Hitachi Data Systems Corporation agreements. The use of Hitachi, Ltd., products is governed by the terms of your agreements with Hitachi Data Systems Corporation.

Notice of Export Controls

Export of technical data contained in this document may require an export license from the United States government and/or the government of Japan. Contact the Hitachi Data Systems Legal Department for any export compliance questions.

Document Revision Level

Revision	*Date*	*Description*
Version 1.0	6/15/15	Document created by GSS FCS Services Engineering

Contributors

The information included in this document represents the expertise, feedback, and suggestions of a number of skilled practitioners. This document is the result of the contribution of the following subject matter experts

- Rich McClelland
- Francisco Salinas
- Gary Mirfield
- David Restor
- Bent Knudsen
- Nathan King
- Phil Wood
- Troy Pillon
- Victor Abyad

Contact

Hitachi Data Systems
2845 Lafayette Street
Santa Clara, California 95050-2627
https://portal.hds.com

North America: 1-800-446-0744

Table of Contents

Introduction

The Hitachi NAS (HNAS) platform is a versatile, intelligent, and scalable multi-protocol solution. HNAS can be configured in many ways to meet a variety of requirements. This document details the best practices for configuring and using HNAS replication and related features. It is assumed that you are familiar with networking concepts and the HNAS platform.

HNAS replication concepts

Object replication

HNAS object replication provides a high-speed means of asynchronously replicating file systems and related configurations, such as CIFS shares and NFS exports from a source server to a destination, or target server, using object-level replication.

Files and directories are made up of objects. Detecting and copying objects from a source to a target requires fewer system resources than detecting files and directories (which include directory structures and metadata, while objects do not). An object-level replication detects and replicates only those objects that have changed on the source file system, thereby using minimal system resources. Object replication is the fastest HNAS method available of performing replications using the system.

In an object replication, a snapshot of a file system is replicated to another server, typically remote, to provide backup and recovery of the source data. The replicated file system may be promoted for use in a disaster recovery situation to allow client access. Additionally, the roles of the source and target servers can be reversed in a disaster, allowing the target server to quickly take over the responsibilities of the source server.

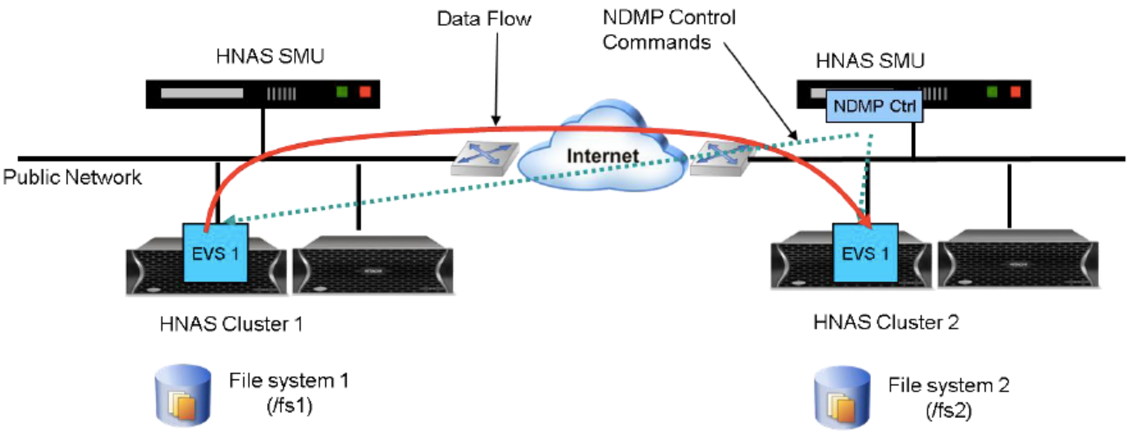
When an object replication first runs, it takes an initial snapshot of the source file system and replicates it on the target server. Subsequent, incremental replications take a new snapshot of the source file system which are compared to the target file system snapshot. Only the differences between the newest snapshot and the last snapshot are copied to the target.

File replication

HNAS file replication provides asynchronous replication between two HNAS servers for disaster recovery (DR), backup, and other purposes.

HNAS uses a built-in NDMP engine to perform replication. This allows file replication to copy both file data and file system metadata which includes virtual volumes and quotas. HNAS file replication is not specific to an HNAS model or version, and supports replication from a source and target within one major release e.g. from 11.x to 12.x.

The following figure illustrates a standard replication topology.



HNAS file replication is not limited to replication between two clusters. It can also provide replication within the same HNAS Enterprise Virtual Server (EVS), within a node between two EVSs, and between two nodes in the same cluster or different clusters remote from each other. This flexibility makes it a versatile replication utility. It copies both file data and file system metadata which includes virtual volumes and quotas.

Note: HNAS does not support overlapping paths between replication policies.

Comparing file and object replication

The HNAS server supports two types of asynchronous replication: File and object. The following table shows a side-by-side comparison between the two different methods of replication.

Type	*File replication*	*Object replication*
Controlled by	SMU	HNAS server
Method	NDMP	Object base
File system support	WFS 1 & 2	WFS 2
Converting FS Block size	Yes	No
Target assessable	Yes (syslock)	Not without promoting
Select IP address	Yes	Yes
Restart fail replication	Yes, checkpoint every 5 minutes	Restarts at beginning of failed session
Schedule frequency	1 Hour	1 minute
Script replication	None	SSC
Cascading	Yes (Post/Pre script)	No
CVL1 or 2 aware	Yes	No (inflates files, one-way)

Type	*File replication*	*Object replication*
File clones	Yes (inflates files)	11.1 (one-way)
Dedupe support	No (inflates files)	No (inflates files, one-way)
V2I clones/backup snapshots	Exclude /.jetapi	11.1 (one-way)
Disaster Recovery (DR)		
DR functions	GUI (Manual) CLI (Scripted)	GUI or CLI
TPA (Transfer Primary Access)	Yes	No
Transfer access point during replication	No*	Yes
Syslock support	Yes	No
Non-disrupted NFS failover	No	Yes
Promote multiple file systems	No	Yes

Note: *Unlike object replication, file replication does not automatically transfer share and export configuration from the source to target filesystems. It is strongly recommended to manage provisioning processes to manually create share and exports on both source and target systems at the time of creation. This ensures that the share and export configuration is present on the target to facilitate disaster recovery.

Object replication

Object replication architecture

Object replication operates at the file system level by copying the objects that make up the files, directories, and metadata for the files and directories in the file system. Files and directories are made up of objects, such as files, directories, security descriptors, snapshot lists, root directory, and many others.

Data Replication has the following characteristics:

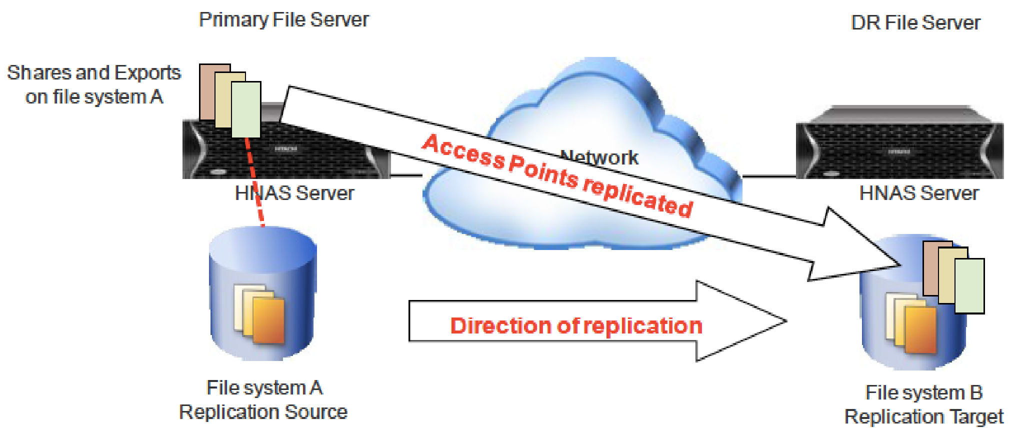
- Object-based
- Hardware accelerated
- Performs a checksum on transferred data
- Tracks changed objects for quick incremental
- Replication state stored on the file systems

Access Point Recovery has the following characteristics:

- NFS exports
- Access configuration list

- All export settings
- CIFS share
- Shares
- SAA
- All CIFS settings

The following figure illustrates the direction of an object replication from a source file server to a target server. Objects on the source server, such as file systems, are replicated to the target server, where they can be instantly accessed during disaster recovery.



Planning considerations

The following information is required to successfully plan the implementation of object replication.

- Recovery Point Objective (RPO) describes the amount of data lost as measured in time.
- Recovery Time Objective (RTO) is the duration of time and a service level within which a business process must be restored after a disaster to avoid unacceptable consequences associated with a break in continuity.
- Are the system(s) using the same HNAS software release?
- Will replication be used for disaster recovery?
- File system capacity available on source and destination systems
- Network details including:
    - Distance to destination system
    - WAN type
    - Sharing bandwidth
    - Bandwidth throttling
    - RTT (`ping –s 1518 –c 50`)
- Are there any firewalls in the path?

- Is the replication network isolated?

Object replication considerations

Object-level replication has the following limitations:

- Object replication is only available for use with WFS-2 file systems. WFS-1 formatted file systems cannot be configured for object replication.
- Object replication works at the file system level only. Entire file systems may be replicated using object replication, but individual files or directories cannot.
- During disaster recovery failover, target file systems are not accessible until promoted to primary. Because the file system is being replicated as its constituent objects, the file system may appear to be corrupted if a user attempts to access it during a replication before all file system objects have been replicated. Target file systems are not accessible until the file system is promoted to primary during failover.
- CNS tree structures are not replicated; they must be manually created on the target file system if CNS is used with object replication.

Object replication best practices

Follow these recommended best practices to ensure the correct use of object replication:

Capacity

Ensure that the replication target is at least as big as the source file system, to always be able to replicate everything from the source. This is especially important if you have more snapshots, and so more space used by snapshots, on the replication target.

If the source file system contains cross-volume links (CVLs), it is necessary to provide additional space for them on the target file system as these are replicated by reverse migrating the data (although they remain migrated on the source file system).

Snapshots

For snapshot rule-based replications, the schedule for the snapshot rule should ensure a snapshot is created before the replication runs, ensuring that a new snapshot is available for the replication.

For example, sometimes an administrator may want to keep hourly snapshots for the last day and daily snapshots for the last month on the replication target. This can be achieved with two policies between the same source and target file systems. The first policy would use a destination snapshot rule with a queue size of 24 and be scheduled hourly. The second policy would use a destination snapshot rule with a queue size of 30 and be scheduled daily. Care should be taken with the scheduling to ensure the daily policy does not start while the hourly policy is running, as this would prevent it from running.

Cross-volume links

If the source file system contains cross-volume links (CVLs), it is necessary to provide additional space for them on the source if they have been modified on the replication target during a role reversal and have to be reverse-replicated.

iSCSI

If the source file system contains iSCSI logical units that need to be in a consistent state when the replication occurs, a script should be used to take the snapshot used by the replication, after flushing the logical unit's data to the server.

Configuring object replication

Initial system setup

When first setting up the system to use object replication, note the following considerations regarding the source and target file systems:

- File systems at the source must have access points enabled.
- File systems at the target must be formatted as a replication target.

Object replication configuration

The basic steps necessary to configure the replication policy, the schedule, and the first replication run are:

1. Create the target WFS-2 file system as a replication target.
2. Enable global file system access transfer on the source.
3. Configure share/export access transfer if required.
4. Create a policy using previously gathered information for the following:
    - Policy name
    - EVS/File system
    - IP address
    - Snapshot (auto/rule)
5. Create a schedule for the newly create policy:
    - Select policy
    - Initial run date/time
    - Run until
    - Every (minute, hour, day, week, month, or continuous)
6. Run the policy.
7. Verify that data replication was successful.

Disaster recovery

Disaster recovery offers the ability to copy your primary access of a file system to your destination file system. Disaster recovery allows you to promote the file system and recover your access. The recovery process will create all your shares, SAA and exports, but you are still required to have your environment configured (DNS, computer object, routes) in order to allow the client access to the new location of your disaster recovery site data.

After you restore services to the primary site system, you will need to verify the integrity of the file system, and then demote the file system to replication target. Demoting the file system prevents any users from accessing the data on the primary site.

Disaster recovery and data replication

To better understand the disaster recovery process, it is useful to study and understand the basic concepts behind the replication process.

At the beginning of the replication, a list of entries from the persona objects is generated. The persona object is used in replication to verify the current state of the replication. Each file system contains a persona object with the following information:

File system identifier (FSID)

- Already generated for each file system
- At the destination, this is the FSID of the source file system used to generate the replica

Replication snapshot identifier

- Only used where the file system is the target file system of a replication
- Used to identify the snapshot on the source file system

Replication restart information

- Only used on the target file system of a replication
- Used to restart interrupted replications

Replication status

- Only used on the target file system of a replication
- Indicates the current status of the replication, as either complete or incomplete

At the beginning of every replication, the persona entries are generated and compared with the replication target, providing the current state of the replication's source and target file system.

During the initial stages of the replication, the file system IDs (FSID), the old/new snapshot ID and the synchronization status of the last replication are all validated before attempting to transfer any data. It is very important to understand the current state of the snapshots on the source and destination before promoting and demoting either of the file systems.

When a replication is started, follow these steps to validate the status of the replication:

1. Take a snapshot at the source and generate the list of the persona entries.
2. Send a start replication message to the destination, with the list of persona entries.
3. The destination receives the message and generates the list of the persona entries on the target file system.
4. Find the latest entry in both lists that match.
5. If no match is found:
    - If the target file system has an unset FSID, then a full replication is performed.

- Otherwise, the replication fails and the destination replies with an old snapshot ID.

6. If the entries for the target file system do not match an FSID and snapshot ID on the entries for the source, the target has to be rolled back to the latest snapshot where a match is found. The replication fails, and the system administrator needs to take appropriate action to remedy the problem.
7. If the entry in the target file system matches an entry in the source, but is marked as incomplete:
   - The destination requests that the replication resume by sending the old snapshot ID and the new snapshot ID corresponding to the interrupted replication, with the object number where the previous replication was interrupted.
   - If the new snapshot ID no longer exists on the source file system, the destination is rolled back to the old snapshot, and the target requests an incremental replication instead.
8. If the entry in the target file system matches an entry in the source, and is marked as complete, the destination requests an incremental replication by sending the old snapshot ID and the new snapshot ID.

Important: When promoting and demoting file systems to minimize data loss, roll back to the latest successful replication snapshots. If you roll back your disaster recovery site file system to a previous replication snapshot, you will lose all data and snapshots taken after that snapshot. When demoting the primary site’s file system, roll back the file system to the same point in time as when you promoted the disaster recovery site file system.

Replication target limitations

Currently, there are some limitations on the target file system:

- There is no way to access the data on a replication target.
- No access points can exist on a replication target.
- You cannot do an NDMP backup to your tape devices from the replication target.

Failover considerations

You should be aware of the customer RTO before failing over to a disaster recovery system. Determine how much time is necessary to cut over all the equipment to the disaster recovery system. If the fail over time is going to be longer than the outage, it is recommended to wait out the failure and restore services on the primary site instead.

Also consider the following factors:

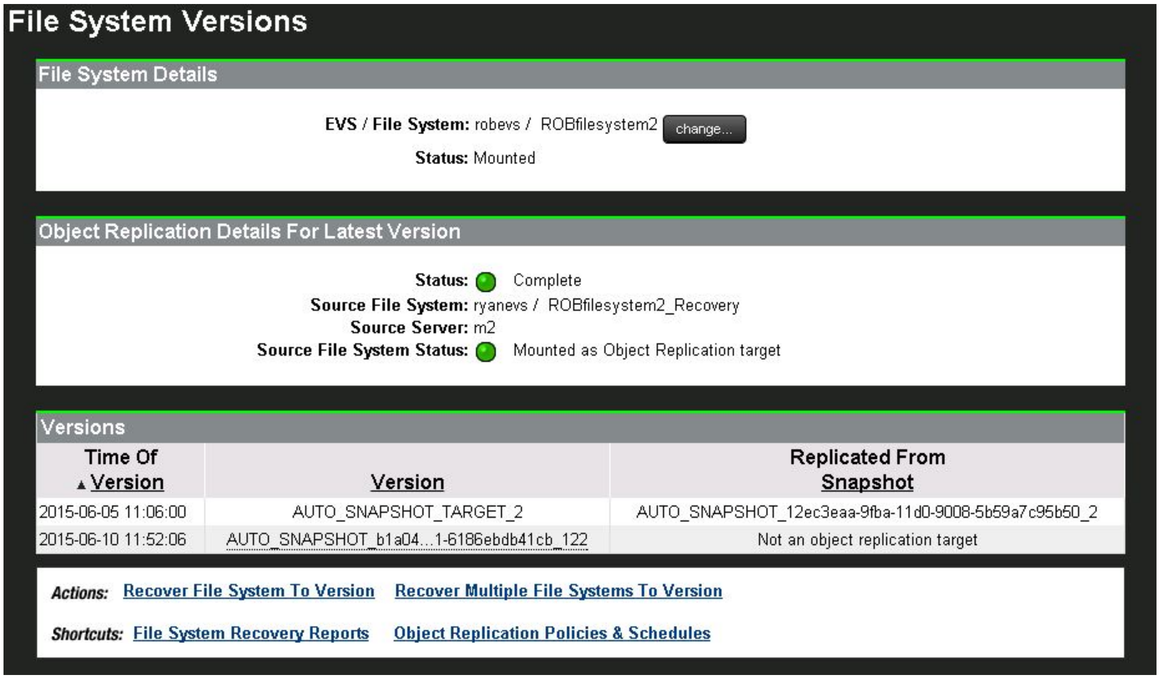
- The duration of the outage
- The duration of the customer RTO
- Whether a snapshot occurred during the replication
- The current state of the file systems

Snapshots

To promote or demote a file system, you must roll back the file system to the last successful replication snapshot. You should take note of what snapshots are available to reduce the amount of lost data during the promoting of a file system. To note the time and version of the snapshots, you can go to the file system version page on the SMU.

In a disaster scenario, the primary system probably will be unavailable, so you need to access the file system version through the SMU on the disaster recovery system. Note the time of the replication snapshot and the versions of the source and destination, and use the latest version when promoting the file system to a normal (`rw`) file system.

The following illustration shows the time of the replication snapshot and the versions of the source and destination.



File System Versions

File System Details

EVS / File System: robevs / ROBfilesystem2

Status: Mounted

Object Replication Details For Latest Version

Status: ● Complete

Source File System: ryanevs / ROBfilesystem2_Recovery

Source Server: m2

Source File System Status: ● Mounted as Object Replication target

Versions

Time Of ▲Version	Version	Replicated From Snapshot
2015-06-05 11:06:00	AUTO_SNAPSHOT_TARGET_2	AUTO_SNAPSHOT_12ec3eaa-9fba-11d0-9008-5b59a7c95b50_2
2015-06-10 11:52:06	AUTO_SNAPSHOT_b1a04...1-6186ebdb41cb_122	Not an object replication target

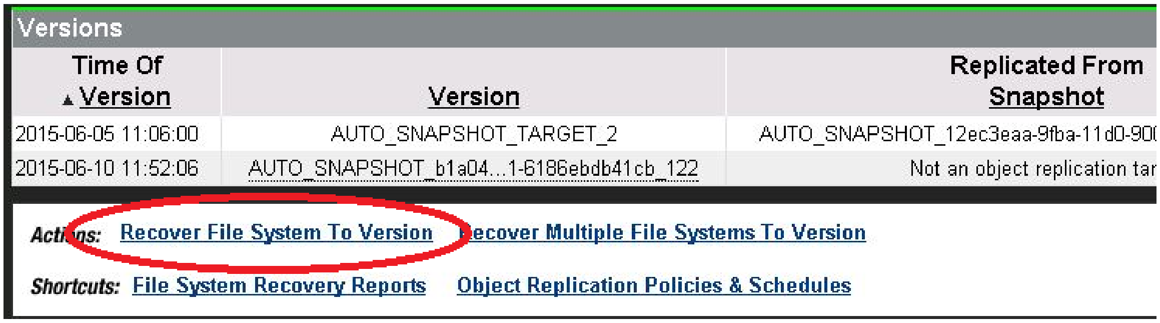
Actions: Recover File System To Version   Recover Multiple File Systems To Version

Shortcuts: File System Recovery Reports   Object Replication Policies & Schedules

Promoting the file system of the disaster recovery system

To promote a file system to a normal (`rw`) file system, in the SMU, under **Actions**, select **Recover File System to Version**. If you need to recover multiple file systems you, select **Recover Multiple File System to Version**. When promoting the file system to a normal (`rw`) file system, verify that the snapshots are the latest before proceeding.

The following illustration shows the relevant SMU section, with the **Recover File System to Version** command link highlighted.



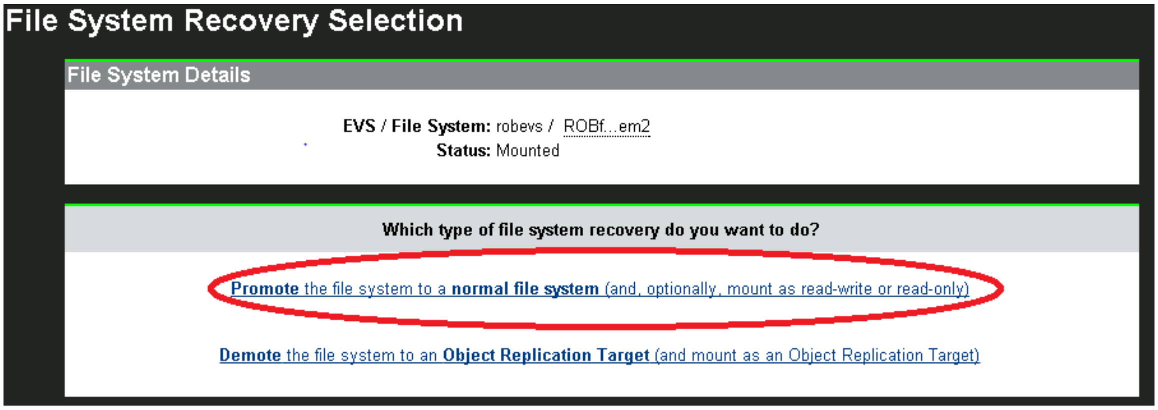
Versions

Time Of ▲Version	Version	Replicated From Snapshot
2015-06-05 11:06:00	AUTO_SNAPSHOT_TARGET_2	AUTO_SNAPSHOT_12ec3eaa-9fba-11d0-90
2015-06-10 11:52:06	AUTO_SNAPSHOT_b1a04...1-6186ebdb41cb_122	Not an object replication tar

Acti ns: Recover File System To Version   ecover Multiple File Systems To Version

Shortcuts: File System Recovery Reports   Object Replication Policies & Schedules

Next, select the option to promote the file system to a normal file system, as shown here.



File System Recovery Selection

File System Details

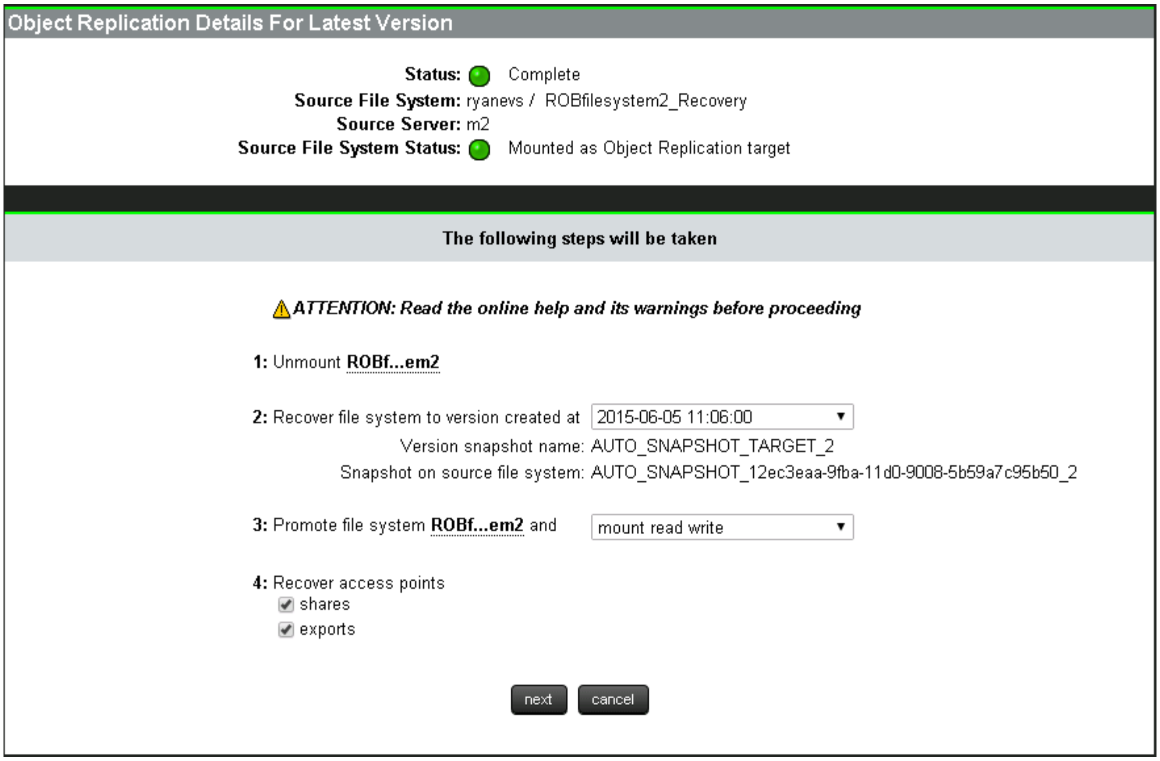
EVS / File System: robevs / ROBf...em2
Status: Mounted

Which type of file system recovery do you want to do?

Promote the file system to a **normal file system** (and, optionally, mount as read-write or read-only)

Demote the file system to an **Object Replication Target** (and mount as an Object Replication Target)

The following illustration shows the file system recovery progress page.



Object Replication Details For Latest Version

Status: ● Complete
Source File System: ryanevs / ROBfilesystem2_Recovery
Source Server: m2
Source File System Status: ● Mounted as Object Replication target

The following steps will be taken

⚠ ***ATTENTION: Read the online help and its warnings before proceeding***

1: Unmount **ROBf...em2**

2: Recover file system to version created at 2015-06-05 11:06:00 ▾
Version snapshot name: AUTO_SNAPSHOT_TARGET_2
Snapshot on source file system: AUTO_SNAPSHOT_12ec3eaa-9fba-11d0-9008-5b59a7c95b50_2

3: Promote file system **ROBf...em2** and mount read write ▾

4: Recover access points
☑ shares
☑ exports

Once you have selected all the options, it will take a few minutes for the server to roll back the file system and restore all the access points. You can monitor the server's progress on the "File System Recovery Progress" page on the SMU, as shown above.

Restoring Client Services

Verify that the disaster recovery site's environment resources are available, and that they include the new IP and computer objects in a cutover run boot. You may need scripts to update the client to the new location of its data.

Verify that the following resources are available on the disaster recovery environment:

- The disaster recovery system's EVS IP Address

- The disaster recovery system's EVS Computer Object
- Routes in place for clients
- Firewall allowing access for the file services

Once you have promoted the file system at the disaster recovery site, redirect the clients to the new location of their data:

- **UNIX clients**: Change the CNAME's A record in the DNS to the new IP address.
- **CIFS**: DFS – A namespace offers a simple way to redirect users to other servers that contain required data when their primary server is down. Advantages include simple setup, no outages, and no client-visible changes.
- **DNS aliasing**: Create CNAME records pointing Primary file server names to the server.
- All shares will be visible when the server is accessed.
- Use the window's startup scripts to point to the new location.
- Have the user logoff and on to regain access to the data on the disaster recovery system.

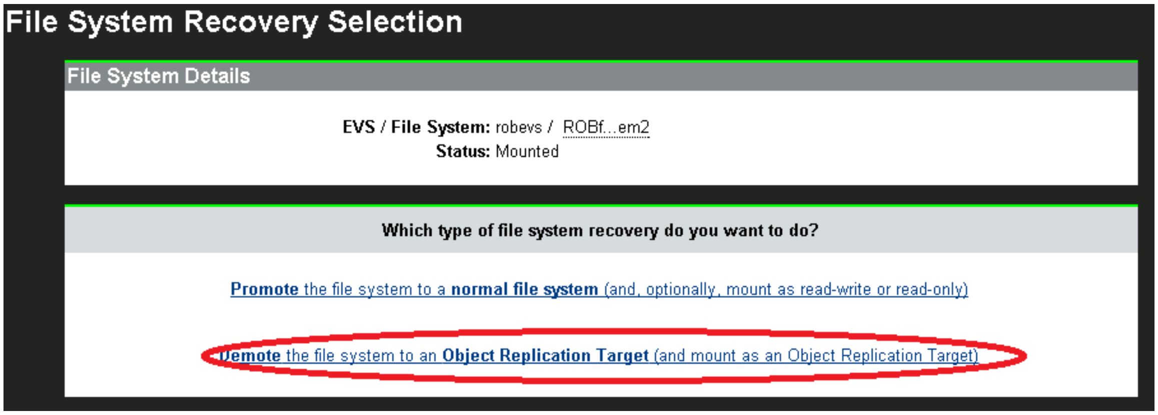
Whichever method is used to redirect your users to the disaster recovery system, ensure that when the primary site services are restored, as DDNS does not redirect the clients to the primary site. Once you demote the file system, the access points are not available for viewing.

Demote the primary site file system

When demoting the primary site file system, this removes all the access points from the file system and prevents users from accessing the file system.

On the SMU, from the Demote File System to Object Replication Target page, verify that the demote process uses the same point in time as the disaster recovery snapshot (the one used to promote the file system). Before you demote the file system to a replication target, you can access the file system to recover any critical file. Be sure to `syslock` the file system as soon as you recover the file system.

The following illustration shows the command link for demoting the file system.



File System Recovery Selection

File System Details

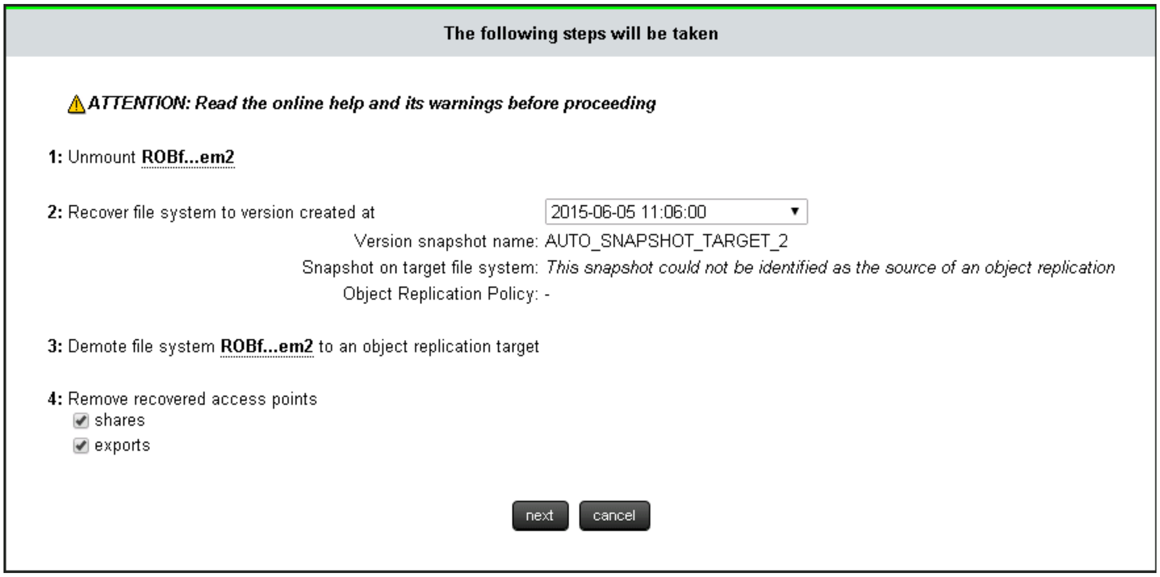
EVS / File System: robevs / ROBf...em2
Status: Mounted

Which type of file system recovery do you want to do?

Promote the file system to a **normal file system** (and, optionally, mount as read-write or read-only)

Demote the file system to an **Object Replication Target** (and mount as an Object Replication Target)

The next illustration shows the Demote File System to Replication Target page:



The following steps will be taken

⚠ *ATTENTION: Read the online help and its warnings before proceeding*

1: Unmount **ROBf...em2**

2: Recover file system to version created at 2015-06-05 11:06:00 ▾

Version snapshot name: AUTO_SNAPSHOT_TARGET_2
Snapshot on target file system: *This snapshot could not be identified as the source of an object replication*
Object Replication Policy: -

3: Demote file system **ROBf...em2** to an object replication target

4: Remove recovered access points
☑ shares
☑ exports

Once you have selected all the options, it will take a few minutes for the server to roll back the file system and remove all the access points. You can monitor the progress on the File System Recovery Progress page in the SMU.

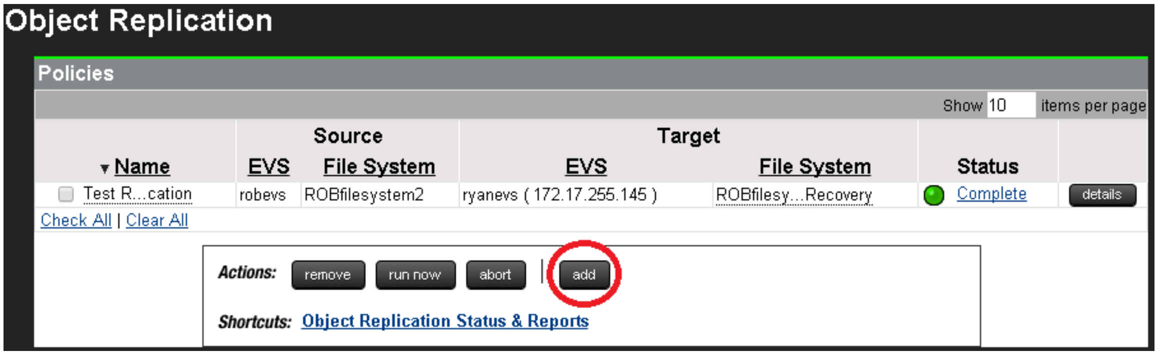
Primary site recovery

Synchronizing the file systems and doing a cutback of the access points to the primary site requires a maintenance window. Depending on the amount of data, it may take a few incremental backups to fully synchronize the file system. Once the incremental is up to date, you can monitor the completion time to pick the day with the least amount of changed data. This will shorten the outage needed to cut back to the primary site.

Disaster recovery steps

Following are the typical steps used for disaster recovery, to return to using the primary site:

1. Create a new policy to synchronize the file systems. The navigation path in the SMU is **Home > Data Protection> Object Replication**. On the Object Replication page, click **add**.



Object Replication

Policies

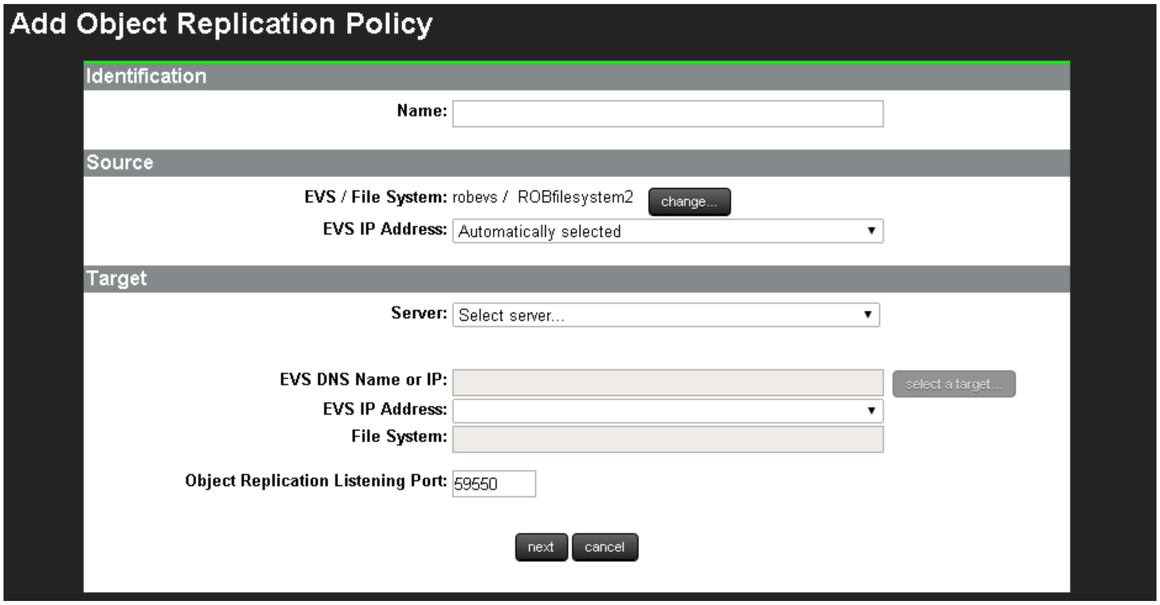
Show 10 items per page

	Source		Target		Status	
▾ Name	EVS	File System	EVS	File System		
□ Test R...cation	robevs	ROBfilesystem2	ryanevs (172.17.255.145)	ROBfilesy...Recovery	● Complete	details

Check All | Clear All

Actions: remove run now abort add
Shortcuts: Object Replication Status & Reports

2. The Add Object Replication Policy page displays. Enter the required information to add the policy.



Add Object Replication Policy

Identification

Name:

Source

EVS / File System: robevs / ROBfilesystem2

EVS IP Address: Automatically selected

Target

Server: Select server...

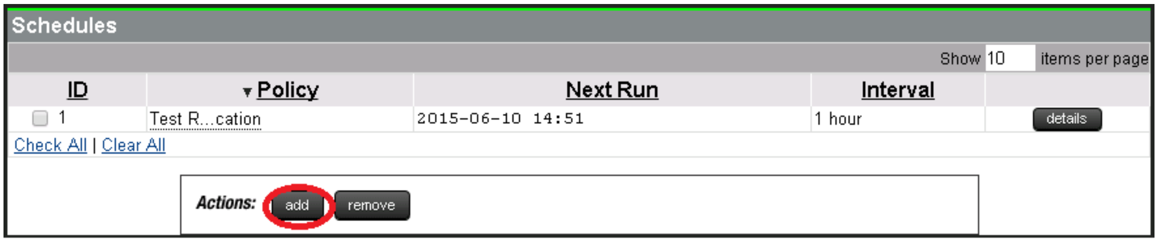
EVS DNS Name or IP:

EVS IP Address:

File System:

Object Replication Listening Port: 59550

3. Create a schedule with the time you will perform the cutback. On the Schedules display, click **add**.

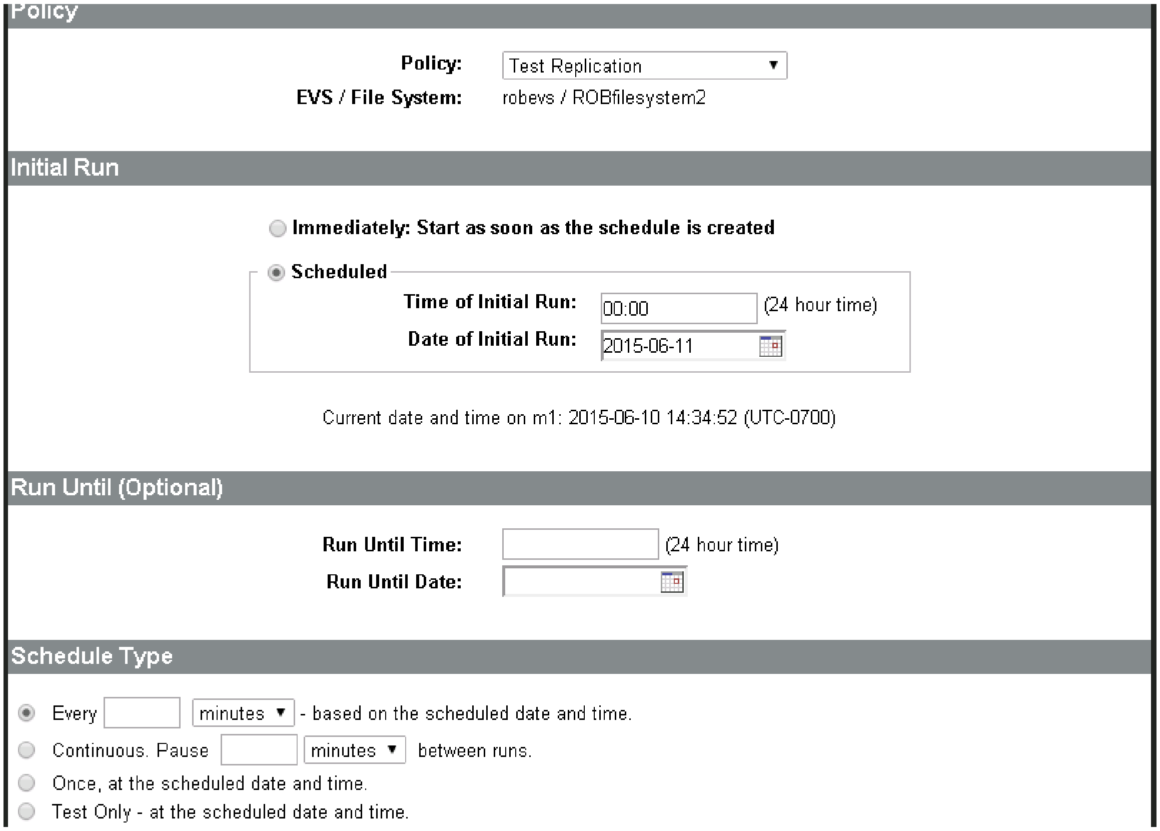


Schedules

Show 10 items per page

	ID	Policy	Next Run	Interval	
□	1	Test R...cation	2015-06-10 14:51	1 hour	details

4. On the next page displayed, enter the schedule specifications.



Policy

Policy: Test Replication

EVS / File System: robevs / ROBfilesystem2

Initial Run

○ Immediately: Start as soon as the schedule is created

● Scheduled

Time of Initial Run: 00:00 (24 hour time)

Date of Initial Run: 2015-06-11

Current date and time on m1: 2015-06-10 14:34:52 (UTC-0700)

Run Until (Optional)

Run Until Time: (24 hour time)

Run Until Date:

Schedule Type

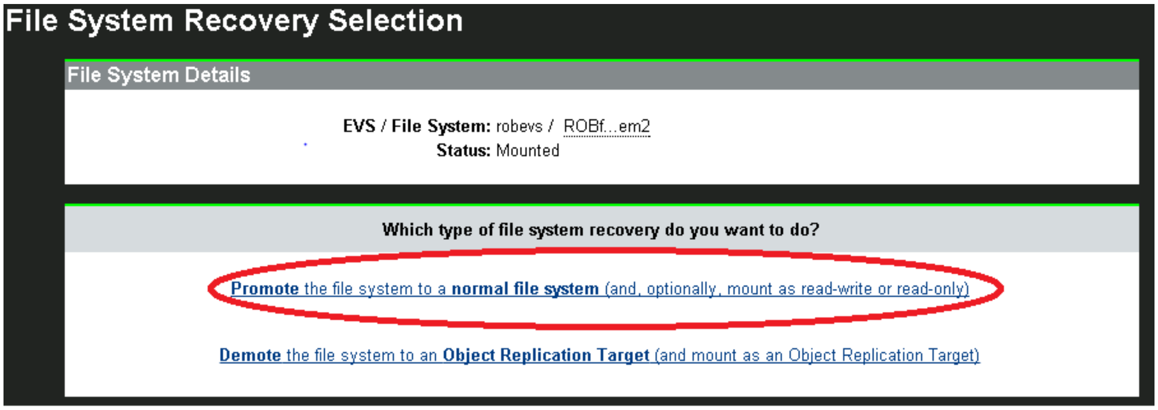
● Every minutes - based on the scheduled date and time.

○ Continuous. Pause minutes between runs.

○ Once, at the scheduled date and time.

○ Test Only - at the scheduled date and time.

5. Synchronize the file systems.
6. Use `syslock` on the file system for the disaster recovery site and perform the final synchronization.
7. Promote the file system of the primary site.



File System Recovery Selection

File System Details

EVS / File System: robevs / ROBf...em2

Status: Mounted

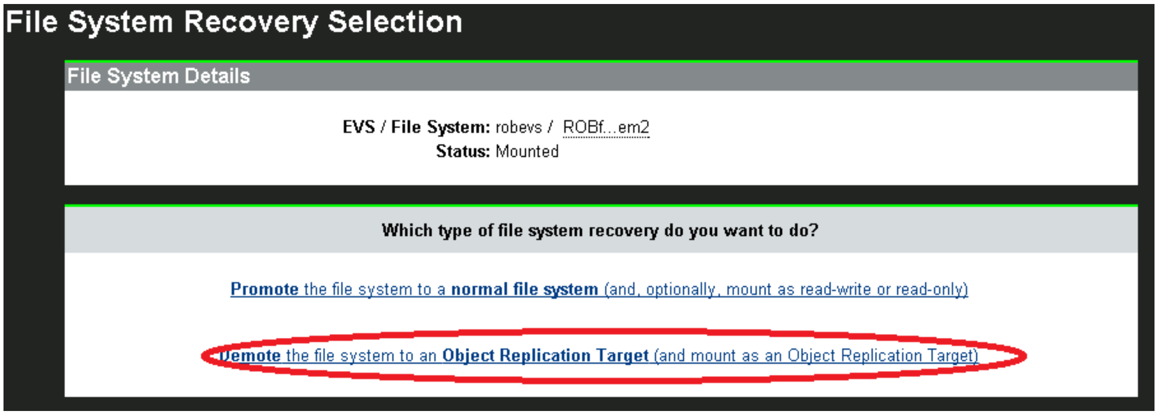
Which type of file system recovery do you want to do?

Promote the file system to a **normal file system** (and, optionally, mount as read-write or read-only)

Demote the file system to an **Object Replication Target** (and mount as an Object Replication Target)

8. Verify that all access points have been created.
9. Redirect the client to the primary site file system.
10. Verify that you can access the data from a client.

11. Demote the file system for the disaster recovery site to a replication target.



File System Recovery Selection

File System Details

EVS / File System: robevs / ROBf...em2
Status: Mounted

Which type of file system recovery do you want to do?

Promote the file system to a **normal file system** (and, optionally, mount as read-write or read-only)

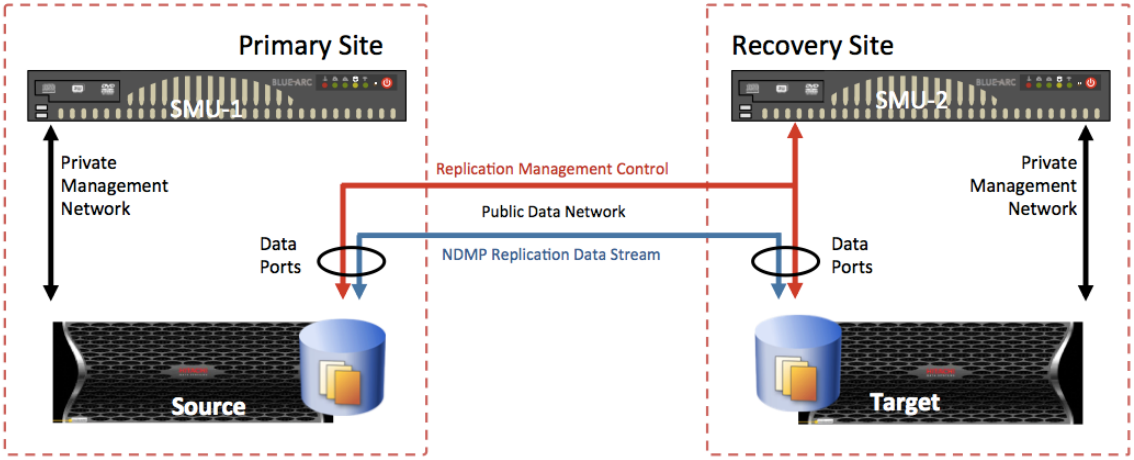
Demote the file system to an **Object Replication Target** (and mount as an Object Replication Target)

12. Reactivate the policy schedule if it was disabled.
13. Verify that the replication is running successfully.
14. Allow user access to the file system.

File replication

File replication architecture

The following illustration shows the connections used in setting up file replication.


Primary Site
SMU-1
Private Management Network
Data Ports
Source
Replication Management Control
Public Data Network
NDMP Replication Data Stream
Recovery Site
SMU-2
Private Management Network
Data Ports
Target

The SMU uses the management connection to the Admin Virtual node (EVS) to discover the parameters needed to set up the replication. The SMU then uses a control connection to the EVS to set up the replication.

Note: The SMU communicates with the EVS because the replication software must run on the node where the file system is hosted, which may not be the same as the node where the Admin Virtual node (EVS) is running.

The SMU asks the receiving EVS to issue a TCP listen. The SMU passes the addresses on which the EVS is listening to the sending EVS, which then issues a connect request to set up the data connection.

Once the session between EVSs is established, the data is sent over the data connection between the NAS servers.

Planning considerations

File replication can be configured only by using the SMU Web Manager. To do this, you will need to be familiar with the customer environment to be successful. To plan the implementation, collect the following information:

- Recovery Point Objective (RPO), which describes the amount of data lost, as measured in time.
- Recovery Time Objective (RTO), which is the duration of time and a service level within which a business process must be restored after a disaster in order to avoid unacceptable consequences associated with a break in continuity.
- System managed by a common SMU
- Is the file replication going to be used for disaster recovery?
- File system layout:
  - Average file sizes
  - Files per directory
  - Wide data structure or narrow
- Physical distance to destination system
- WAN type
- Sharing bandwidth
- Bandwidth throttling
- RTT (`ping –s 1518 –c 50`)
- Any firewalls in the replication path
- Timeout value
- Isolated replication network

File replication network communication considerations

In certain environments, the EVS IP addresses or networks may not accessible by the SMU. There are different configurations to take into consideration in these situations.

- If SMU is unable to communicate with both the source and target EVS.
- If this is a managed replication, the SMU is managing both systems.
- When doing a managed replication, but SMU is unable to manage the target system through the management network.
- If the management network is not being used.
- If a segregated network is being used to isolate management from data replication.
- If SMU can communicate with the EVS, but there is a requirement to isolate the data transfer.
- If SMU cannot communicate between either the EVS due to a secured network.

- If a specific network for management and data transfer is required.

File replication IP selection process

HNAS needs to determine the addresses used to set up the initial control connections from the SMU to the EVS and also the data connection between the EVSs.

- When setting up the *control* connection, the IP address used by default is the first IP address associated with the EVS. HNAS has mechanisms to deal with problem issues.
- There may be situations where the SMU cannot connect to the EVSs at all. The management connection can be configured through the private management network. In this case, the CLI command `ndmp-management-ports-set` is used to configure a relay from the management network to the EVS.

In scenarios where specific IP addresses are required, use the policy addresses file (`<policy_name>_addresses`) to specify the addresses for the data and control connections.

When setting up the *data* connection, the SMU asks the receiving EVS to set up a TCP listener. It listens on all IP addresses associated with the EVS and sends a list of these IP addresses back to the SMU.

The SMU sends this list to the source EVS and asks it to connect. At this point, the source HNAS needs to choose a local IP address associated with the local EVS and a remote IP address from the list passed from the receiver in order to create the connection. The selection is as follows:

- HNAS first searches for a pair of IP addresses within the same subnet. If found, HNAS chooses this pair. Note that HNAS will perceive a conflict if there are addresses at both ends which appear to be on the same subnet but in fact are on two different private subnets.
- If the SMU connected to the EVS directly, HNAS assumes the connected address must be on a public network and chooses that IP address. (This is true for both of the IP address pairs.)
- If the SMU was connected through a management network relay (using the `ndmp-management-ports-set` CLI command), the first IP address configured for the EVS is used.

Using the policy addresses file

Certain situations require the ability to specify both the IP address used to set up the NDMP connection control protocol and the data transfer IP address.

To support this functionality, it is possible to set up a policy addresses file in:

```
/opt/smu/adc_replic/conf/replication_policies
```

With the name `<policy_name>_addresses`, where `<policy_name>` is the replication policy name.

Therefore, if you created a policy named `home-dir`, the policy addresses file would be `home-dir_addresses` and this can have the following contents:

```
SRC_CTRL_ADDRESS=x.x.x.x SRC_DATA_ADDRESS=y.y.y.y DEST_CTRL_ADDRESS=z.z.z.z
DEST_DATA_ADDRESS=w.w.w.w
```

The replication scripts take the CTRL addresses as fixed addresses to use when accessing the source and destination NDMP servers. The DATA addresses are used for inter-server connections.

If the SMU containing the replication policies cannot connect to the local serving EVS by any route, the storage administrator may still need to use a management connection (`ndmp-management-ports-set`) to the local HNAS server.

In this case, the CTRL address should be omitted from the policy addresses file. For instance, if the SMU is local to the source server and can only connect through the management interface, but it can connect to the destination server EVS, the addresses file might resemble the following:

```
SRC_DATA_ADDRESS x.x.x.x DEST_CTRL_ADDRESS y.y.y.y DEST_DATA_ADDRESS z.z.z.z
```

In addition, you would configure the `ndmp-management-ports-set` command on the local HNAS server to the SMU containing the configured replication policies.

The policy addresses file may be used for an existing policy on HNAS system versions 6.1 and higher. Any IP address changes are picked up on the next replication run. If the IP address changes, you can change the IP address in the policy addresses file between replication and the next run will pick up the new IP address. When reversing the replication direction, it may be necessary to update the policy addresses file.

NDMP port ranges

Certain situations require the ability to specify the TCP port range used by NDMP replication. This is particularly useful when firewalls exist between source and target systems. Beginning with HNAS release 12.4, the NDMP option `data_port_range` is available through the CLI and allow you to control the port(s) available to listen for data connections during TCP NDMP operations.

With the feature enabled, NDMP replication or ADC copy operations assign a listening port within the specified range.

- NDMP currently holds a listening port for the duration of an NDMP session, which can result in the port being blocked for extended periods of time.
- If multiple NDMP sessions are in progress, they are each assigned a port within the port range. If there are no available ports within the defined range an ‘Address already in use’ message is returned.
- Ensure the port range is sufficient for the maximum number of concurrent NDMP sessions.

The setting is implemented from the CLI using the command:

```
ndmp-option data_port_range [{ any | <numeric_port_value> |
<numeric_port_range> }]
```

The default setting is "any," meaning any available port.

A valid port number range consists of a minimum and a maximum port number, separated by a dash (-), with the maximum port number greater than or equal to the minimum port number. An example range specification would be “2000-2010.”

Note: A range with 0 as one port number and an otherwise valid non-zero port number as the other is not a valid range.

There should be enough ports in the specified range to support the required number of NDMP operations. One port will be used for each NDMP session for the duration of its connection. If one or more ports in the specified range are already in use for some other purpose, those ports will not be available for NDMP. A wider port range may be required to compensate.

Tuning replication performance

There are many factors to consider in calculating the expected replication performance between HNAS systems. Understanding the network topology is key to a successful implementation. These include:

- Topology:
  - Bandwidth
  - Latency
  - Network throttling
  - CoS (Class of Service)
  - Firewalls
  - Switches
- File system layout:
  - Average file size
  - Directory structure
  - Average files per directory
- System utilization:
  - Peak system usage
  - Replication scheduling: Continuous, hourly, or daily
- Conflicting NDMP operations (backups)

The distance between source and destination HNAS systems can be both short (within a single data center) and long (between remote data centers).

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 needs to be tested during implementation. 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 used, such as restoring from tape. The performance of the transfer depends on the network connectivity between the two locations.

To optimize replication for specific network latency conditions, it is possible to customize the TCP windows scaling factor on the HNAS. A larger window allows HNAS to transmit more data and improves overall throughput over WAN links.

The RTT (Round Trip Delay Time) latency can be discovered by using the CLI command `ping–s 1518 –c 50`. If this is not feasible, it is possible to estimate the RTT by calculating the distance in kilometers and then dividing it by the speed of light per second (200,000 kilometers per second) and adding an overhead factor of 50%:

```
(<kilometers>/200,000)+50% = RTT
```

To calculate the expected TCP throughput vs. window size factor for a specific RTT use the formula:

```
(<windows size factor>/<RTT(ms)>)/2
```

The following example illustrates the potential throughput for a specific scaling factor based on an RTT latency of ~44ms:

- Scaling factor:1=128 kilobytes/44.2730475 milliseconds/2 = 1.4116941 MBps
- Scaling factor:2=256 kilobytes/44.2730475 milliseconds/2 = actual 2.8233882 MBps
- Scaling factor:3=512 kilobytes/44.2730475 milliseconds/2 = actual 5.6467764 MBps
- Scaling factor:4=1024 kilobytes/44.2730475 milliseconds/2 = actual 11.2935528 MBps
- Scaling factor:5=2048 kilobytes/44.2730475 milliseconds/2 = actual 22.5871056 MBps
- Scaling factor:6=4096 kilobytes/44.2730475 milliseconds/2 = actual 45.1742112 MBps
- Scaling factor:7=8192 kilobytes/44.2730475 milliseconds/2 = actual 90.3484225 MBps
- Scaling factor:8=16 384 kilobytes/44.2730475 milliseconds/2 = actual 180.696845 MBps

Note: A scaling factor of 2 is set on the HNAS by default.

To set the windows scaling factor on HNAS, use the command `ipeng –F <window scale factor>` on the source HNAS system. Choose the appropriate value according to maximum available bandwidth.

Managed and unmanaged Hitachi NAS platform servers

The system management server manages multiple Hitachi NAS platform servers.

Replications can be configured between Hitachi 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.

Important: 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, the replication will not start and might fail to complete.

File replication considerations

File replication has several considerations to note. First, NDMP replication will run on WFS-1 or WFS-2 file systems, meaning it is supported on all HNAS platforms.

Additionally:

- The SMU should be able to communicate to the first IP address in each EVS.
- The SMU at the remote disaster recovery site should manage the replication jobs.
- File replication is asynchronous.
- The maximum number of concurrent sessions per HNAS head is 50. File replication uses one session on source and destination. Two sessions are used when replication occurs within the same head.
- NDMP backup also uses one session. Consider the number of backups and replications running concurrently and the potential affect a node failover can have.

- Overlapping paths in replication policies are not supported.

File replication best practices

The following guidelines should be considered when setting up file replication.

SMU

In general, use the SMU attached to the target Hitachi 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.

File system

Ensure the replication target file system is at least as large as the source file system.

Share and export configuration

Unlike object replication, file replication does not automatically transfer share and export configuration from the source to target filesystems. It is strongly recommended to manage provisioning processes to manually create share and exports on both source and target systems at the time of creation. This ensures that the share and export configuration is present on the target to facilitate disaster recovery.

Snapshot retention

The default automatic snapshot retention time is seven days. Therefore, if the initial full and first incremental replication is going to take longer than seven days, increase the automatic snapshot retention time to accommodate.

Note: The maximum retention time is 40 days.

If the snapshot retention time is extended, consider the file system change rate and allocate additional space required by snapshot.

Target snapshots

Business needs may require additional snapshots at the destination. For example, an organization might need hourly snapshots for the last day and daily snapshots for the last month on the replication target. This can be achieved with snapshot rules on the destination system.

Snapshot rules

Ensure the queue depth for a snapshot rule is large enough to prevent the baseline snapshot from being removed before the next replication starts. (Once the next replication has started, the baseline snapshot is protected from deletion.)

For snapshot rule-based replications, the schedule for the snapshot rule should ensure a snapshot is created before the replication runs, ensuring that a new snapshot is available for the replication.

Clone files

If the source file system contains clone files, it is necessary to provide additional space for their replication on the target file system, as they will be inflated on the target file system. Also, exclude the `/.jetapi` directory in the replication rule.

Multiple connections

By default, each replication policy uses a default of four processes and twelve read-ahead threads. Increasing these provides best improvement when replicating random files and directories (non-sequential disk access). Read-ahead processes should generally be larger if typical size of files is small. The processes and threads should be increased in the same ratio and can be adjusted in the GUI.

Note: The use of additional connections increases the load on the system and could negatively impact performance.

Cross-volume links

If the source file system contains cross-volume links (CVLs) or external cross-volumes links (XVL):

- There is an option to recreate the data migration paths on the target system and select the re-migrate data in a replication rule

  OR

- Provision additional capacity on the destination file system for the fully inflated data.

Note: This could be an issue in a disaster recovery scenario as the primary site file system may not be able to support the addition space for a reverse replication.

ISCSI

If the source file system contains iSCSI logical units that need to be in a consistent state when the replication occurs, a script should be used to take the snapshot used by the replication after flushing the logical unit data to the server

Change Directory List

The Change Directory List (CDL) is a journaling function which maintains a list of file system changes between snapshot operations. This feature can be used to improve replication performance by reducing the need to traverse the file system structure to determine changed each time a replication policy is run by avoiding directories which do not contain changed files.

The maximum number of changed objects a file system can track is one million objects (file or directory). A recommendation for enabling CDL depends on the file system structure.

- If the file system structure is narrow, enabling CDL will not result in any improvements and will probably add to the replication time while it processes the objects to path.
- Enabling CDL is beneficial for large file systems with a wide directory structure that has a few (not millions) files per directory and where the change happens in a relatively small number of directories.

When processing the change directory object list:

1. The changes are recorded as list of objects that have changed between two snapshots.
2. The replication process uses the list of objects to create a tree of all the directories with changed data, and to do this HNAS must go back through all the relevant snapshots.
    - For all the changed objects HNAS retrieves the file metadata to see if the change is relevant.
    - At this point, HNAS can filter on the directory path or virtual volume path.
3. Finally, the replication starts replicating data.

As is evident by the CDL stages, it can take some time to process data to create the relevant directory tree used for the replication.

Note: If replicating at the virtual volume level, this process runs on the entire file system for every policy on the file system.

CDL is more effective with a small number of changes, so the policy may need to run every few hours rather than every few days.

Number of additional server connections

The source and the destination of a data replication can create multiple connections as opposed to the current single connection. The different connections are used to copy different subdirectories, which will be processed in parallel by separate threads on the source and destination machines.

The option controls the number of additional server connections that will be established during a replication operation (from 0 to 30).

Increasing the number of additional server connections may improve performance by allowing multiple transfers in parallel.

Note: Each additional server connection consumes system resources and best practices indicate limiting the number of additional server connections to situations where they improve performance. Also, as the number of additional server connections is increased, more read-ahead processes are required.

Number of read-ahead processes

This setting controls the number of read-ahead processes used when reading directory entries during a replication.

Each additional read-ahead process takes up system resources, so it is best to limit the number of additional processes unless a significant performance improvement can be achieved.

While the default number of read-ahead processes is suitable for most replications, file systems made up of many small files increase the amount of time spent reading directory entries proportionately. In such cases, adding additional read-ahead processes may speed up the replication operation provided the storage can handle the additional I/O.

File replication limitations

File replication (as opposed to object replication) is a good choice in most cases where replication is needed. However, there are some limitations to be aware of:

- File replication requires the SMU to be able to communicate to the first IP address in each EVS
- File replication is asynchronous and can result in a higher RPO than other methods, including array-based replication.

Configuring file replication

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 following steps:

- **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**: 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**: Defines the schedule, timing, and policy based on the scheduled data and time.

To locate the file replication function on the SMU, select **Data Protection**. File 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 and 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*

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

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

Important: Multiple replication policies can be established that run simultaneously. For example, a single Hitachi NAS platform server can be the destination target for several independent source Hitachi 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 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. The timing and schedules that use the same resources should be designed to not interfere with each other.

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, select **Data Protection > File Replication** or **File Replication Rules**.
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.
   - Block-based Replication Minimum File Size. This option cuts down on data transferred to increase performance. The default is enabled, which chooses a 32 MB file size.
   - The changed directory list disabled by default, but is recommended to be enabled. This helps to track changes and speed replication activities.
   - The Number of Additional Server Connections allows the connection threads to be modified.
   - Keep the Number of Read Ahead Processes set to the default (1). (The range is 0 to 32.)
   - Keep the Pause While Other File Replication(s) Finish Writing option set to the default (Yes).
   - Keep the Take a Snapshot option set to the default (Yes).
   - Keep the Delete the Snapshot option set to the default (Last) for incremental replications. The default value is typical for disaster recovery.
   - Keep the Migrated File Exclusion option set to the default (Disable), unless you have a spread of data across a migration that needs to be replicated.
   - Keep the Migrated File Remigration option set to the default (Disable), unless you establish a migration capability on the destination target.
   - Keep the External Migration Links set to the default (Re-migrate).
   - Keep the Ignore File Attribute Changes option set to the default (Disable).

Replication policies

First, the NDMP user name and password on the remote destination server are needed. Then, to add a replication policy:

1. From the SMU home page, select **Data Protection > File Replication**, and in the Policy section, select **add**.
2. Enable the setting for **Not a managed server**.
3. First, make certain you are on the correct HNAS platform server or cluster (source), and on the next SMU page:
   - 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 of the remote system 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 quiesce file system I/O, take a snapshot, and then restart the I/O.

- Assign replication rules as needed. (This assumes previously existing replication rules have been defined.)
- Make certain the destination location is large 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 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.

Note: When optimal disaster recovery is the goal, it is recommended to perform a continuous replication that starts the next replication process immediately after the previous one if finished, minimizing any potential data loss. Continuous replication is the best practice for replication for disaster recovery.

Recommended server settings

The speed of recovery is important, depending the business or application access. To prepare a system for the fastest level of recovery:

- Identify share names, share paths, and file systems on both the source and destination Hitachi NAS platform servers.
- Ensure there is enough space or that the same size file systems are used on both the source and destination Hitachi NAS platform servers.
- 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

HNAS file replication can be used for effective disaster recovery. A script on the SMU named `replication_recovery.sh` is used to manage recovery, failover, failback, and incremental reverse replication.

To use this script, the replication must be managed, as opposed to un-managed. Also, the SMU must have both the source and destination HNAS clusters under management.

The best practice is use the SMU at the destination site to manage the replication. The source cluster can be managed by the local SMU, and the destination SMU is only needed to manage the replication policies and DR functions.

It is highly recommended not to perform other management functions against the local cluster using the remote SMU.

Note: A public admin EVS IP address should be configured on the source HNAS cluster.

Note: If using the policy addresses file with the `replication_recovery.sh` script, remember to change the SRC and DEST entries in the policy addresses file when failing over and failing back.

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

You can address replication failures in one of the following ways:

- Restart the replication from where it stopped.
- Allow the next scheduled replication to continue.
- Roll back 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 there is an opportunity to restart the old one, in which case the second option occurs automatically. In disaster recovery situations, rollback is the only option available, as a failure in the source server may make continuing the replication impossible.

Primary failure scenarios include:

- Replication fails, and the source node is available
- Replication fails, but the source node is not available

Important: Export data and shares are not replicated and need to be copied periodically or recreated, as needed. This is a manual process and can require significant 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 is 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 is 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**: This mode is used 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 by the replication system until recovery process has completed).
- **Synchronized**: A rollback has completed on the source. The source is ready for restoring data from the destination.
- **Copyback**: This mode is used 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 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, you may wish 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 section Restarting Replications 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 actions are needed. When this occurs, the first action is to bring the destination online as the working store.

Replication rollback and disaster recovery considerations

Replications set up for disaster recovery purposes are typically configured for continuous replication, 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. Although 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, HNAS provides a way to roll back 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 pre-replication 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.

As a best practice, the intent is to keep the consistency of two copies between the two locations. This way it is always possible to roll back to this consistent state.

If the SMU in the source site—the one which controlled the replication policies—is down, the SMU in the remote site for the target systems need sto 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.

- Click **restart** to restart the replication from its last checkpoint.
- Click **rollback** 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 disaster recovery 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.

Replication completed and source failed

If a consistent state exists between the source and target systems, and the source failed between replications, there is no need to roll back the file system. You can bring 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. This needs to be done manually, however. To bring the target online:

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 addresses.

3. DNS changes: Change the IP address of the source HNAS servers resource record to the IP address of the recovery target HNAS destination.
4. On a single client, at the command prompt, enter the command `arp -d *`
5. Do an NSLOOKUP using the primary HNAS host name to test that the name resolves to the disaster recovery HNAS IP address.
6. Do an NSLOOKUP using the disaster recovery HNAS IP address to test that the IP address resolves to the primary HNAS host name.
7. If the name and host resolution work correctly, all clients will need to have their ARP cache cleared. This can be done 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 all hosts 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. However, if the source system comes back online and the source data is still intact, it might save time to restart replications as described here.

A script is provided to help in restarting replications. It is run from a command line prompt on the SMU. To run the script, you must be in the `/opt/smu/adc_replic/` folder.

```
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.sh` script are used to switch from one mode to another, with the 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 typical recovery sequence and uses 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, 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 run again.
2. If the source system becomes available while the destination rollback is happening or before the destination has been enabled as the primary store, 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, 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 re-run by reissuing the **Synchronize** recovery command.

5. If the destination has taken over as the working store and the source is in Synchronized mode, 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 set to Decoupled and source mode is set to Synchronized, the **Reverse** command swaps the roles in the replication and the replication scheduling is re-enabled.

If the destination mode is set to Decoupled and source mode is set to Synchronized, 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 complete.

Hitachi Data Systems

Corporate Headquarters
2845 Lafayette Street
Santa Clara, California 95050-2639
U.S.A.
www.hds.com

Regional Contact Information

Americas
+1 408 970 1000
info@hds.com

Europe, Middle East, and Africa
+44 (0) 1753 618000
info.emea@hds.com

Asia Pacific
+852 3189 7900
hds.marketing.apac@hds.com