NETAPP TECHNICAL REPORT

Hardware Upgrade of WORM Data

Manish M Agarwal, NetApp

February 2009 | TR-3752

GUIDELINES FOR HANDLING HARDWARE UPGRADES OF WORM DATA

This document presents best practices to be followed during the process of hardware upgrades while making sure of the immutability of WORM data and preserving its retention period. Note that following these guidelines does not automatically ensure legal compliance.

TABLE OF CONTENTS

1 INTRODUCTION

Businesses today are faced with stringent regulatory and corporate governance requirements. They increasingly rely on some usage of WORM (write once, read many) data storage to meet these requirements. WORM data storage helps in making sure of the immutability and permanence of data, which is valuable from a regulatory compliance standpoint. At the same time WORM also helps businesses that place a premium on protecting certain business records or critical data files from accidental or intentional alteration or deletion.

NetApp® SnapLock® and SnapLock for SnapVault® are features of Data ONTAP® that implement a high-performance, disk-based magnetic WORM storage. These are flexible and scalable solutions that are supported on any NetApp storage platform. SnapLock is an open solution that utilizes standard protocols to enable seamless integration with ISV archival applications as well as custom applications.

SnapLock for SnapVault is a storage-efficient solution that mitigates compliance risk by helping to make backups compliant and by delivering fast data access and recovery. SnapLock for SnapVault leverages both SnapLock and SnapVault technologies to provide unaltered versions of enterprise data along with a ComplianceJournal, which tracks changes between versions. The advanced security features of SnapLock and SnapLock for SnapVault now enables customers to meet all regulatory requirements across all data classes (structured, semi-structured, and unstructured) and all levels of compliance.

This document presents best practices to be followed during the process of hardware upgrades while making sure of the immutability of WORM data and preserving its retention period. Note that following these guidelines does not automatically ensure legal compliance.

2 SCOPE OF THE DOCUMENT

This technical report covers only the hardware upgrade scenarios involving WORM data. This does *not* include the following:

1. Changing version of Data ONTAP: For software upgrades and downgrades, refer to the product documentation. The release-specific *Data ONTAP Upgrade Guide* contains the step-by-step procedure as well as the necessary checklists.
2. Compliant data migration: Migrating data from one Data ONTAP system to another in a manner that enables compliance and (as a result allows) subsequent deletion of original copy. Steps for creating replicas of the original WORM data are described in this report; following these does not automatically guarantee that the copy is legally compliant. Also, even when the copy is made as described, SnapLock does not allow deletion of the original copy of the WORM data if it is under SnapLock Compliance protection. SnapLock Enterprise allows deletion of data by system administrators.
3. Migrating WORM data from a competitor solution (for example EMC Centera) to a NetApp SnapLock solution: This document does not cover this scenario. There are Professional Services offerings from NetApp available to perform this migration (refer to section 6).

Note that this document only provides guidelines that are relevant to SnapLock WORM data. It is not a detailed and step-by-step set of instructions for hardware upgrades in general. This document is not a replacement for the *NetApp Hardware: System Upgrade Procedure*, but a supplement to it.

3 HARDWARE UPGRADE SCENARIOS

This section details the different scenarios and the guidelines for handling each.

3.1 REPLACING A DISK

When a single disk hosting WORM data needs to be replaced, it is possible to do so using the command line interface.

If a spare disk of size greater than the original disk is used, then the extra space on the new disk will be unusable: for example, if the disk that needs to be replaced is of size 250GB while the new disk that it is being replaced with is of size 500GB, then after the replacement is complete only 250GB will be usable. The following warning would appear to get the administrator’s confirmation:

```
NTAPC1> priv set advanced
```

```
NTAPC1*> disk replace start 0a.16 0a.45
*** You are about to copy and replace the following file system disk ***
  Disk /aggr0/plex0/rg0/0a.16

      RAID Disk Device  HA  SHELF BAY CHAN Pool Type  RPM  Used (MB/blks)
Phys (MB/blks)
      --------- ------  ------------- ---- ---- ---- ----- --------------
--------------
      dparity   0a.16   0a    1   0   FC:B   0  ATA   7200 211377/432901760
212718/435647712
***
Disk 0a.45 is bigger than disk 0a.16.
Only 211 GB will be used on disk 0a.45.
Really replace disk 0a.16 with 0a.45? y (ENTER RESPONSE)
disk replace: Disk 0a.16 was marked for replacing.
```

Note: After the replacement is complete, the old disk will be added to the pool of spare disks. It can then be used in any volume or aggregate (including those other than SnapLock).

Also note that the disk in question can only be replaced by a disk of equal or larger size. Any attempt to replace the disk with one of lower size will fail with the following error:

```
disk replace: Disk <disk-name> is not of appropriate size.
```

For a detailed set of instructions for the relevant Data ONTAP release, refer to the *Storage Management Guide*.

3.2 REPLACING THE CONTROLLER

This section deals with the case where the storage controller on the system is being replaced by another one. There are several different possibilities. The instructions for handling WORM data differ according to the specific case.

I. USING EXISTING DISKS

The old disks hosting the WORM data are going to be used with the new controller. In these cases there is no need to create a copy of the WORM data. The upgrade is done “in place.” To be able to do this the following must be true: the new controller should support the old disks: that is, physical disks hosting the SnapLock volumes should be supported with the new controller. In this case the physical disk shelves hosting the SnapLock volumes can be attached to the new controller. Follow the procedure listed in *NetApp Hardware: System Upgrade Procedure*.

In case the new controller has ComplianceClock™ initialized, refer to NetApp TR-3618 for possible effects this could have on ComplianceClock on the new controller.

II. USING NEW DISKS

The old disks hosting the WORM data are going to be replaced with newer disks. In this case the WORM data needs to be copied, except for a special case described at the end of this section. The disk replacement could be due to one or more of the following conditions:

1. Existing disks are not supported on the new controller head.
2. The disks are being replaced with faster and/or larger capacity disks.
3. More expensive disks are being replaced with cheaper disks.

In these cases the WORM data cannot be upgraded in place, and a copy of the WORM data needs to be created on the new system. Section 4 explains the steps involved in this.

Note that creating a copy of the WORM data will not result in the disks housing the original copy of the WORM data to be now available for reuse. That is the advantage of using the disk replace command (when applicable). The `disk replace` command can be used when all the following conditions hold true:

a. The disks are also being replaced (along with the controller).

b. The new disks are supported on the old controller.

c. The new disks are equal in capacity to the old disks that are being replaced OR the loss of additional capacity is acceptable (refer to section 3.1).

In this case the new disks can be connected to the old controller and the procedure described in section 3.1 repeated for each of the disks in the aggregate or traditional volume hosting the WORM data. Once this is complete, the new disks can be moved over to the new controller by following the procedure listed in the documentation titled *NetApp Hardware: System Upgrade Procedure*.

For possible effects this could have on ComplianceClock, refer to TR-3618.

4 CHAIN OF CUSTODY FOR COPIES OF WORM DATA

Note: Following these steps does not automatically ensure legal compliance. Rather it is intended as a starting point for planning so that the important steps are included in the overall migration plan. Consultation with your legal department is highly recommended to make sure that your unique specific compliance requirements are met.

This section describes the steps involved in copying WORM data to a new volume (one hosted on the new system and/or disks). Once these steps are completed, the data can be made available online on the new system, and the clients accessing the data can be pointed to the new copy. The procedure involves the following stages:

4.1 DATA COPY

The WORM data from the old system can be copied over to the new system using a `vol copy` or `aggr copy`. Refer to *Commands: Manual Page Reference, Volume 1* and *Commands: Manual Page Reference, Volume 2* for further information.

Care should be taken to create the volume of same variant of SnapLock on the new system.

Note: It is possible to use technologies such as SnapMirror® to create a replica of the WORM data residing on a SnapLock Compliance volume onto a SnapLock Enterprise or a regular volume.

Here is an example using `vol copy` to copy the WORM data from volume `fvol_compliance` on NTAPSRC to `fvol_comp_copy` on NTAPDEST. While the example only shows the actual command line options for `vol copy`, the logical steps would be similar for `aggr copy` as well.

Note: `vol copy` works at a block level and thus requires that the destination of the copy be at the same or higher version of Data ONTAP. Also, the source and the destination volumes should both be either flexible volumes (FlexVol® volumes) or traditional volumes (tradvol).

1. Find the size of the source volume (`fvol_compliance` on NTAPSRC):

   ```
   NTAPSRC> vol size fvol_compliance
   ```

2. Create a corresponding volume on the destination (`fvol_comp_copy` on NTAPDEST). The size of the destination should be greater than or equal to the source volume.

   Note: This will require the creation of a SnapLock aggregate of the same variant (that is, Enterprise or Compliance) of SnapLock as the source aggregate. The variant of SnapLock is selected at an aggregate level, which is inherited by all volumes on the aggregate. The exception is a traditional volume (or tradvol), which does not reside on an aggregate abstraction and can have its own variant of SnapLock set.

   ```
   NTAPDEST> vol create fvol_comp_copy <aggr name> <size>
   ```

3. The new destination volume needs to be first put into a “restricted” mode before the copy can be initiated.

   ```
   NTAPDEST> vol restrict fvol_comp_copy
   ```

4. Use the command line on the source system to initiate the copy:

   ```
   NTAPSRC> vol copy start –S fvol_compliance NTAPDEST:fvol_comp_copy
   ```

   Note: It is possible to copy just a single Snapshot™ copy of the source volume (`fvol_compliance` on NTAPSRC). In the example above all the Snapshot copies will get copied over.

If ComplianceClock on the destination system is ahead of ComplianceClock on the source, bringing the restricted volume online will cause ComplianceClock on the destination system to shift backward to match the value of the source. For more information on ComplianceClock, refer to TR-3618.

Note that it is possible to bring the new copy online at this point and then complete the remaining two steps in parallel with the client accesses to the new copy of the data. This is important for large data sets where the verification phase can take a long time. Also note that allowing data access after this step will make the task of verification a little harder. This is explained in more detail in the next section.

4.2 VERIFICATION

Data copies created using `vol copy` are identical. However, this step is required to generate a persistent record of all the files and their contents in both the source and the destination copy of the SnapLock data. This record will be useful to verify at a later that the contents and other properties such as the retention period of the WORM data did not change in process of the copy. The verification results will be especially useful if the source is discarded or destroyed (so it is not around to verify the contents at a later date) after the copy.

Once the WORM data has been copied over to the destination a check can be run to test the following conditions:

1. All the files from the source were copied over.
2. The contents of the files that were copied over are the same as the source.
3. The relevant metadata for each of the file is the same between the copy and the source.
4. The retention period in effect on the source and the destination are similar.
5. SnapLock related options (both volume level and system wide) are the same on both sides.

To avoid dealing with constantly changing data it is advisable to do the comparison based on some recent Snapshot copy of the data.

TESTS 1, 2, and 3

This can be done by generating and comparing “fingerprints” of the files on the source and the destination or by doing a byte by byte comparison of the contents and the metadata.

> **Note**: Data ONTAP release 7.3.1 onward will include a command to generate the fingerprints of files. This will also be accessible via an ONTAPI™ call `file-get-fingerprint`. For earlier versions of Data ONTAP the verification script will have to compute the fingerprints by reading the files and the attributes.
>
> **Note**: Unless the ONTAPI call `file-get-fingerprint` is being used, the source and the destination will need to be exported via NFS or CIFS.

```
NTAPC1> file fingerprint

usage: file fingerprint [-a {md5 | sha-256}] [-m] [-d] [-x] <path>

- Calculates the fingerprint of the file using md5 or sha-256 digest
algorithm.

NTAPC1> file fingerprint /vol/fvol/dir1/file1

/vol/fvol/dir1/file1

Data Fingerprint: JudpdnTfzrrJz/3uWJCmatKcqWmIoIXmVIk3qUuilzg=

Metadata Fingerprint: umG3Q0GxS6i2HB9TE98B3KNRKiPq/pFUd5XkrLe9hgI=
```

When comparing the file metadata it is suggested that the following file attributes be considered:

- File type (only regular files can be WORM protected by SnapLock)
- File size
- User ID of the file owner
- Group ID of the file owner
- Security ID (SID) for the owner (visible only from CIFS)
- Time of last modification (mtime)
- Time of last access (atime): with SnapLock this represents the file retention period for the file

- File creation time (this is only visible from a CIFS client, not visible to an NFS client)
- Time of last status change (ctime: only visible from NFS clients, not visible from CIFS clients)
- File permissions and other security attributes

Other attributes that may or may not be important from a compliance standpoint depending on the application are:

- Access control lists (ACLs) on the file
- Alternate or named data streams on the file (Note: SnapLock currently does not protect these)
- Link count on the file (for files with hard links)

Confirming that every file on the source has a matching file on the destination (that is, the file path and name should match) and the contents and the metadata for each of the files takes care of the first three of the five requirements mentioned at the start of this subsection.

Note: If instead of doing the verification immediately after the copy, the verification is done after the new system has been in use for a while, then the file metadata may not match completely: for example, the retention times on the WORM files may have been extended (it cannot be shortened). Moreover, new WORM or non-WORM content might have been created. Expired WORM files may also have been deleted. In these cases, depending on the situation, the verification could be relaxed to take this into account.

TEST 4

It should be ascertained that the retention periods in effect on either end are the same. Matching the time of last access (done above) will make sure that the retention time stamp is the same on either end. However, for the absolute time stamps to make sense the value of ComplianceClock also needs to be compared on either ends. ComplianceClock on the new system should either be in sync or behind the source. If ComplianceClock on the new system is behind, it will result in the retention period being that much longer.

TEST 5

Finally, the volume and system options related to SnapLock should also be compared to make sure that the SnapLock variant, the default retention period, the maximum and the minimum retention periods, and the SnapLock options in effect are the same. The relevant SnapLock options are:

a) Volume level options:

   Use the following command to find the options:

```
NTAPC1> vol options <volume-name>
```

   The following are the SnapLock related options:

   - One of the following will be set: `snaplock_compliance or snaplock_enterprise`
   - `snaplock_default_period`
   - `snaplock_minimum_period`
   - `snaplock_maximum_period`

b) System options:

   Use the following command to find the system options related to SnapLock:

```
NTAPC1> options
```

   Look for `snaplock.autocommit_period` and `snaplock.compliance.write_verify`

4.3 REPORT

This is a summary of actions taken with details that are pertinent from an audit perspective. The report should include information from the verification phase. The report should be stored as a WORM record with retention period equal to the maximum retention period of any record in the volume.

The following is a checklist of items that might be relevant for a report for the migration:

CHECKLIST

1. Why the copy is necessary

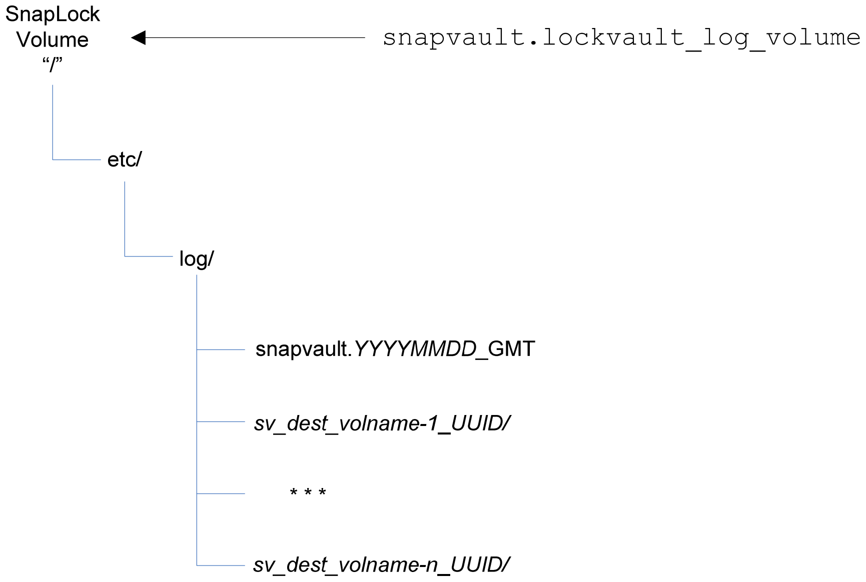
2. Who is doing it, and their position in the company security/compliance hierarchy
3. System information (output of `sysconfig` command) including the NVRAM ID and the hostname of the system for the source
4. The same system information for the destination system
5. Volume information for the volume being copied (use command line interface or the ONTAPI call `volume-list-info`):
    a. Volume name
    b. Volume UUID
    c. Volume type (FlexVol or tradvol)
    d. Name of the aggregate hosting the volume
    e. Aggregate UUID (use the ONTAPI call `aggr-list-info`)
    f. Volume expiry date (Data ONTAP 7.3.1 also has a command `vol status -w`)
    g. Whether the volume is a SnapLock Compliance or a SnapLock Enterprise volume
6. The same volume information for the copy of the volume
7. Information described in section 4.2 for every file on both the source and the destination
8. ComplianceClock value on the source at both the start of the operation and at the end of it (use the `date -c` command or ONTAPI call `clock-get-clock`)
9. ComplianceClock value on the destination at the end of the data copy and verification

5 SNAPLOCK FOR SNAPVAULT

SnapLock for SnapVault needs special handling due to the presence of ComplianceJournal on the secondary. ComplianceJournal resides on a SnapLock volume (enterprise or compliance) configured by a system level option `snapvault.lockvault_log_volume`.

5.1 FORMAT OF COMPLIANCEJOURNAL

It contains the journal of SnapLock for SnapVault related activities. It has the following directory hierarcy:



- Operations Log file: `/etc/log/snapvault.`*`YYYYMMDD`*`_GMT` contains a log of system (SnapVault secondary) wide activity. There are multiple such files, typically a separate file being created on Sunday morning at midnight. Each file contains start and finish times for each SnapVault update. For example:

```
tgt Tue Jan 20 19:05:00 GMT sv_dest_vol xfer:sv_hourly Target_start
dst Tue Jan 20 19:05:00 GMT NTAPSRC:/vol/src_vol0
NTAPDEST:/vol/sv_dest_vol/qtree_src_vol0 Request (Update)
dst Tue Jan 20 19:05:06 GMT NTAPSRC:/vol/src_vol0
NTAPDEST:/vol/sv_dest_vol/qtree_src_vol0 Start
dst Tue Jan 20 19:05:20 GMT NTAPSRC:/vol/src_vol0
NTAPDEST:/vol/sv_dest_vol/qtree_src_vol0 End (2928 KB)
tgt Tue Jan 20 19:05:25 GMT sv_dest_vol xfer:sv_hourly
Target_create_snapshot (sv_hourly.20090106_080423_GMT)
tgt Tue Jan 20 19:05:27 GMT sv_dest_vol xfer:sv_hourly Target_end
```

- Files-transferred log file: Each /etc/log/*sv_dest_volname_UUID* directory contains a subdirectory for each source volume (named as the volume or QTree name followed by its UUID). This subdirectory contains the log of activity between the source and destination volume pairs. Here is an example of the log contents:

```
# Transfer type: Incremental Start
# From: NTAPSRC:/vol/src_vol0
# To: NTAPDEST:/vol/sv_dest_vol/qtree_src_vol0
# Start time: Tue Jan  6 08:04:03 GMT 2009
# Source snapshot timestamp: Tue Jan 20 19:00:05 GMT 2009

Mon Jan 19 23:21:31 GMT 2009 Delete File    14       ./dir1/testfile
Mon Jan 19 23:22:36 GMT 2009 Delete File    42
./dir1/sm/ckpts/NTAPDEST/sv_dest_vol_qtree_src_vol0
Tue Jan 20 18:42:40 GMT 2009 Modify Attr     5       ./dir1
Tue Jan 20 18:42:40 GMT 2009 Create File    14       ./dir1/testfile
Tue Jan 20 19:00:04 GMT 2009 Modify File    20
./dir1/log/snapmirror
Tue Jan 20 00:18:58 GMT 2009 Modify Attr     6       ./dir2
Tue Jan 20 00:18:52 GMT 2009 Create File    14       ./dir2/filexyz
Tue Jan 20 00:18:56 GMT 2009 Create File    14       ./dir2/file111
Tue Jan 20 00:18:58 GMT 2009 Create File    14       ./dir2/file123
# End of Log File
```

5.2 HANDLING THE LOG VOLUME

Both the system level log and the log for each pair of source and destination volumes contain volume names and volume UUIDs. The volume names and UUIDs may change when the SnapLock for SnapVault destination is copied. To preserve the meaning of these logs it is suggested that the following steps be taken:

1. The volume hosting ComplianceJournal also be copied (according to steps mentioned in section 4) to the new system. The report corresponding to this copy should be added to the new copy of the volume and made WORM with a retention period greater than or equal to the maximum retention period for all files in the volume.

2. Each of the SnapLock for SnapVault destination volumes should also be copied to the new system. The reports corresponding to each of these copies should be added to the volume hosting ComplianceJournal on the new system: under the directory corresponding to the destination volume. Since the data copy report contains the old and the new volume names and their corresponding UUIDs, it will help in translating old ComplianceJournal entries on the new system.

3. The `snapvault.lockvault_log_volume` option should be pointed to the volume hosting ComplianceJournal on the new system.

4. The SnapLock for SnapVault relationships can now be reestablished between the source volumes and the new system containing the copies of the old destination volume.

6 SERVICE OFFERINGS

NetApp Professional Services have service offerings for both the initial implementation of SnapLock environment and for copying WORM data due hardware upgrades or migrating from a third-party platform to a NetApp platform. For an updated list of services offered, contact NetApp Professional Services. Here are some of the services currently offered:

1. SnapLock and SnapLock for SnapVault Implementation Service

   These services will help customers with the implementation of SnapLock and SnapLock for SnapVault.

2. Enterprise Vault on NetApp Implementation Service

   Veritas™ Enterprise Vault™ software, NetApp storage systems, and NetApp software provide highly robust storage, archival, records-retention, and information life cycle management solutions for Microsoft® Exchange Server, Windows®, and UNIX® file servers, and Microsoft SharePoint® Portal Server. When deployed with NetApp SnapLock Compliance or SnapLock Enterprise software, these solutions help protect companies' business and legal interests by preventing the modification or deletion of archived messages and important documents through robust WORM (write once, read many) storage capabilities. The joint NetApp and Veritas solution provides enterprises with an easily managed, highly scalable, and available data management infrastructure for Exchange, file systems, and SharePoint at a low total cost of ownership. The combined solution, including joint support and professional services, simplifies content archiving and enables compliance with strict records-retention regulations.

3. Data Archival and Compliance Solution: EMC Centera Replacement Program

   NetApp Professional Services has teamed up with Procedo to deliver the industry’s leading archive solution composed of the best technology and professional services specifically targeted for EMC Centera customers hosting e-mail archive data in EmailXtender or Symantec™ Enterprise Vault application environments.

   The solution is ideally suited to customers with ever-expanding e-mail data archives who are discovering EMC Centera and EmailXtender environments can't keep pace with their needs. Services professionals with deep and proven experience and extensive training make the customer’s transition from EMC Centera to more flexible NetApp platforms smooth and efficient.

7 REFERENCES

Archive and Compliance Management Guide

Data Protection Online Backup and Recovery Guide

Commands: Manual Page Reference, Volume 1

Commands: Manual Page Reference, Volume 2

NetApp Hardware: System Upgrade Procedure

Storage Management Guide

TR-3263: WORM Storage on Magnetic Disks Using SnapLock Compliance and SnapLock Enterprise

TR-3738: SnapLock Record Retention Date and Implementation Strategy

TR-3618: Understanding SnapLock ComplianceClock