The Union of Snapshot-Based Backup Technology and Data Deduplication

Chris Mavromatis

EMC Proven Professional Knowledge Sharing 2010

Chris Mavromatis
Consulting Engineer – Advanced Technical Support
EMC Global Services
Table of Contents

Scope ................................................................................................................................ 3
Tape vs. Disk .................................................................................................................... 3
Array Based Data Protection ............................................................................................. 4
  Array Backup Techniques ............................................................................................. 4
  CoW Technique ............................................................................................................ 5
Array Based Restore ........................................................................................................ 6
Data Duplication ............................................................................................................. 7
Best of Both Worlds ........................................................................................................ 8
Solution Components ..................................................................................................... 9
Backup Workflow ............................................................................................................ 9
Configuration .................................................................................................................. 12
  Avamar Grid Configuration ......................................................................................... 12
  CLARiiON Configuration ......................................................................................... 14
  NetWorker Configuration ......................................................................................... 18
Performing a Backup ....................................................................................................... 27
Performing Recovers ...................................................................................................... 30
General Configuration Tips ............................................................................................. 32
Conclusion ...................................................................................................................... 33
Glossary .......................................................................................................................... 34

Disclaimer: The views, processes or methodologies published in this compilation are those of the authors. They do not necessarily reflect EMC Corporation’s views, processes, or methodologies
Scope
This article reviews some of the current issues behind tape technology as a method of data backup, and its replacement with disk based backup solutions. The main scope of this article is to illustrate how these disk based backup technologies can merge into a single solution to provide the reliability of a seasoned backup enterprise software platform, the snapshot based backup functionality of the EMC CLARiiON disk array, and data deduplication from the Avamar data store.

Tape vs. Disk
During the past 15 years, tape backups have been the predominant tool for backup and recovery operations. Tape solutions answered the call for backup challenges surrounding backup, recovery, and the proverbial “golden copy in the vault” for disaster recovery. With the explosive data growth rates in today’s data centers, tape technology and traditional backup software solutions are starting to plateau in their ability to achieve designated backup windows and SLAs (Service Level Agreements). Granted, there have been improvements in data transfer rates in tape drive technology over the past several years, as you can see here:

- AIT-1 - 4MB/s
- DLT-8000 - 6MB/s
- LTO-2 - 35MB/s
- LTO-4 - 120MB/s

However, even with these improvements, when compared to a typical desktop hard drive of 0.5 Gbits/s, the performance characteristics become quite evident. High-end drives can achieve sustained transfer rates of up to 1 Gbit/s.

Backup-to-disk solutions not only provide fantastic transfer rates but the fact that tape drive media has a linear data access method is another significant factor making disks a superior media. During recovery scenarios, tape media commonly finds itself positioning to align to the location of the data on tape, whereas the disk drives’ intrinsic random access characteristics lend themselves nicely to restore tasks.
Array Based Data Protection

Beyond the basic backup-to-disk scenarios where backup software directs a data stream to a disk device, high-end disk arrays such as EMC CLARiiON® or EMC Symmetrix® expand the capacity plan of a disaster recovery solution, and enhance it to what is now commonly referred to as Business Continuity. Through advanced disk array capabilities such as mirroring, Business Continuance Volumes (BCV), or Copy on Write (COW), these high-end disk arrays provide near-line storage capability such that when the original data image is lost/damaged, businesses can continue to operate with zero, or minimal service interruption.

Array Backup Techniques

*Split-mirror technique (STD – BCV)*

In split-mirror, or disk mirroring, every write request to the original data is automatically duplicated to other mirrors or copies of that data. A mirror is redundant and is not a frozen image or snapshot. The mirror can be temporarily suspended, or split, to create a snapshot. The disk subsystem temporarily stops making updates to the mirrored copy; the data is frozen at that point.

You can use the split-mirror for the backup process or other purposes. Mirrors create a snapshot of the data with the split capability. Unlike the CoW technique, a full data copy is available. This requires three copies of the volume to provide continuous processing for backup. When the backup is complete, the mirror is resumed. In this setup, there are primary and secondary real-time copies, and another PIT copy of the data.

Using Symmetrix device mirroring for backups offers this key advantage: the backup of a static mirror-copy of the data rather than the production data itself. For example, databases need special handling to prepare for backup, either shutdown or online backup mode. But, with mirroring, they need this special handling for only the brief time that is needed to split the mirrors, not for the duration of the backup. During the backup, the database server can function normally using the production disks for the database.
**CoW Technique**

When a backup is requested using the CoW technique, the disk subsystem creates a second pointer, or a snapshot index, and represents it as a new, and full, copy. It is the same to the user. The snapshot is created by saving the original data to a snapshot index whenever data in the archetypal volume is updated. The snapshot process creates an empty snapshot index, maintaining the original values that later changes in the archetypal volume after creation. The snapshot is a combination of the archetypal volume data with the snapshot index containing original data changed in the archetypal volume.

The CoW technique requires a fraction of the archetypal volume disk space. Average snapshot disk space requirements are 10 to 20 percent of the archetypal volume space. The requirement depends on how long the snapshot is active and how many writes are made to the archetypal volume (that is, snapshot index). CoW is efficient except in a heavy write environment or when the copy is required to be active for a long time.

CoW technique-based snapshots are good for impact-less backup operations because:

- Snapshot creation is quick when compared with split-mirror technique-based snapshots.

- Snapshots require less storage space than split-mirror technique snapshots, which require the same amount of storage space as the source volume.
CoW technology-based snapshots depend on the underlying storage space for repository and cache and the archetypal volume combined, to present a particular PIT view of the source volume. Repository and cache storage space is allocated at the time of snapshot creation and, depending on the vendor's implementation, is fixed or can be expanded. Load on repository and cache increases as more writes occur on the source volume and may overrun the repository and cache storage space. The snapshot view is broken and the snapshot is rendered useless. Furthermore, when multiple snapshots exist for a given archetypal volume, large amounts of write activity can break all or some of the snapshots.

**Array Based Restore**

During restore requests from backups taken from mirror snapshots, or Copy on Write snapshots, the restore operation is internal to the disk array. In the case of STD – BCV mirror restores, users can sync the BCV back on the STD which results in data on the production STD to revert to the image in which the STD and BCV were split.

In the case of CoW, you can perform restores so that the data on the source LUN (Logical Unit Number) can revert to its original image as of the time the snap session was started. This is achieved by taking all of the write operations written to the snap reserve LUN pool, and applied back to the source LUN.

Both of the restore operations provide restore capabilities based on the internal array capability to transfer data within itself. The restores are vastly faster than traditional tape restore methods where the data is read from sequential tape media and sent to the application server via an IP, SCSI, or SAN transfer pipe.

Finally, another important advantage of disk array based backup and restore is that they are both server-less and LAN free operations. Because the data movement is internal to the disk array, these operations do not impose a load on the production data server. This is an important differentiating factor in a data centers’ backup strategy.
Data Duplication
The game changer for disk based backups

Cost is often the deciding factor for continuing to use tapes over disk based backup solutions. Traditionally, disk based backup solutions have been more expensive than tape cartridges. As companies need to perform daily backups, the sheer amount of data needing protection is very large, thus making the mass storage capacity of tape cartridges less expensive than disk. Imagine for a moment that you have 10 users with an identical 100MB Word document on their PCs. During backup, all of these 10 files will be backed up regardless of the fact that they are common to one another, and thus 1 GB of storage would be consumed on tape media.

If any one of these users modifies their file, even one word within the document, traditional backup software to tape will result in making a backup of the entire 100MB file all over again. Basically – regardless if a single byte or the entire file is modified, the entire file has to be backed up again. You can easily see how backups would require large amounts of space on a weekly basis, thus making tape the most cost effective backup solution.

Data deduplication is the game changing technology that allows businesses to successfully deploy full disk based backup solutions.

As the name suggests, the goal of data deduplication is to eliminate redundant data from backups. With the appropriate client agents installed on the servers that are to be backed up, data deduplication technology analyzes the 100MB file that is targeted for backup at the source machine, and determines if the data is already known to the backup server. If it is, the file does not need to be transmitted to the backup server, and thus the backup is now completed very quickly since the entire file does not need to be sent over again. This workflow is true even if the machine that is doing the backup has not previously backed up this file. If another machine on the network has already backed up this 100MB Word file, then all machines on the data zone benefit from the deduplication.
In the above example, if a very small portion of the file was modified, the deduplication software would once again analyze the 100MB file, and determine that only a small fraction of the file needs to be sent over to the backup server, thus making the backup process faster due to the time savings of not retransmitting the complete 100MB file.

The deduplication backup software keeps track of all of the machines and users that performed the backups, and manages the pointers that reference the original data, and all additional chunks from all backups taken. Data deduplication technology now makes disk based backup solutions a very cost effective business proposition.

**Best of Both Worlds**

It is now possible to merge these distinct backup methods of array based backups and deduplication backups into a single solution that gives you the ability to benefit from both disk array features, and data deduplication technology. You can achieve this merge of technologies through the use of:

- NetWorker™ backup server
- PowerSnap application module
- Avamar data store
- CLARiiON disk array

This article assumes a minimal knowledge of each of the above technologies. We will focus on the methods to merge these technologies. We will use the NetWorker backup server as the main backup platform to maintain the backup configuration, schedule the backups, and drive daily backup operations. Although NetWorker independently is not able to perform data deduplication, it is able to incorporate with an Avamar data grid which in turn has the main function of performing data deduplication.
EMC CLARiiON is the high-end disk array that will perform the snapshots of the production data using the previously mentioned CoW method. The bridge between the NetWorker backup software and the CLARiiON disk array is achieved through PowerSnap. PowerSnap is an add-on module for the NetWorker backup server that allows the backup server to integrate with the disk array, and thus drives the snapshot based backup operations.

Solution Components
Please assume the following:

Production server “as2” is a Windows 2003 server. It has access to a LUN which has been provisioned from the CLARiiON disk array, and is mounted on drive letter W:\

Proxy server “tito” is also a Windows 2003 server. Its primary purpose is to be the proxy host that is going to mount the backup snapshot of W:\ onto itself, and validate the snapshot operation. It is also responsible for sending a copy of the contents of W:\ to the backup server for backup purposes.

NetWorker backup server “ecc1” is a Windows 2003 server. Its role is to provide the main backup functionality on the data zone, integrate with both the CLARiiON disk array through the usage of the add-on plug-in PowerSnap, and also integrate with the Avamar grid that is installed on the network.

Avamar grid “avamar01-01” is a single node Avamar server. It will be the target for the deduplication data we are going to generate.

Backup Workflow

During backups of the W:\drive of client as2, we configure NetWorker such that we will take a PowerSnap backup of the data residing on this drive. Recall that W:\ is actually a drive which is a CLARiiON LUN, and for which we have configured Snapview™ (within the CLARiiON array) to perform a CoW snapshot.
At the start of the backup, the server as2 that contains the PowerSnap client software communicates with the CLARiiON array to start a Snapview session against the LUN W:\. This operation begins tracking the I/O change requests against the drive W:\ internally within the CLARiiON array.

Next, the proxy server “tito” tries to import the snapshot LUN (not the original) of W:\ onto itself so that it can validate that the snapshot of the data was successfully performed. After the proxy server has imported the snapshot and validated the snapshot, it sends the contents to the Avamar grid in a deduplicated fashion.

This is achieved by analyzing the data that is read off of the snapshot drive of W:\ and comparing the data against the local cache database. This will determine if the file is already saved (and not changed) from a previous backup attempt. If it needs to be backed up again, this file will be opened, chunked into smaller pieces, and a hash id calculated for that chunk of data.

Next, the hash id is compared against a second local cache database to determine if the data is already known to the Avamar grid. If a cache hit occurs, this portion of the file data has already been backed up previously and does not need to be transmitted to the Avamar grid. If a cache miss occurs, a final lookup is performed again at the Avamar grid to determine if this data chunk has potentially been backed up previously from another client. If this lookup succeeds, the data does not need to be transmitted to the Avamar grid since the whole concept of data deduplication is to attain but a single copy of common data regardless of its origin. Avamar grid pointers will be updated to reflect that this data chunk was saved for our client even though no data was actually sent, so that during recovery the data will be sent back to the as2 server.

This chunking, lookup, and data transmission process repeats until all of the data on the snapshot drive has been analyzed and saved. Next, the proxy host deports the snapshot LUN, passes back to the application host as2, and sends a signal to the NetWorker backup server that the backup was completed.
Upon completion, we are left with a snapshot on the CLARiiON array that can be used in one of two ways.

1. We can use this snapshot to once again mount it on the proxy host and perform a file-by-file restore of select data.

2. We can perform a “roll back” of the snapshot within the disk array itself (not having to mount any snapshots on any host), and quickly roll back the data at the point in time of the snapshot. This is the fastest way of restoring large amounts of data.

Furthermore, we have the contents of W:\ protected off of the array, and onto the Avamar grid (through the use of NetWorker backup server) in a deduplicated fashion. This gives us the ability to restore all the data in the event that the array experiences a catastrophic failure. For an additional level of protection, the Avamar grid itself is commonly replicated to a secondary Avamar server. It is imported to keep in mind that the backup process stated above is “server-free”, as all of the intense processing is performed onto the proxy server tito and NOT on the application host as2. This makes this backup solution a very lucrative proposition for production hosts that need minimal service performance impact during the backup window.

The first backup of the data which is deduplicated and sent to the Avamar grid does take longer then the subsequent backups. This is because the first backup “seeds” the local cache files that would contain the references of the hash ids from the data which is to be backed up. All subsequent backups would greatly benefit from a populated cache file as the data would commonly be found to be a duplicate of previous backups, and thus would not need to be re-transmitted to the Avamar grid.

This is but just a single workflow scenario. It is possible to create snapshot backups that run several hours per day, leave the snapshot available for the restore onto the CLARiiON, and then send the data from the first/last/or all of these snapshot backups to the Avamar grid in a deduplicated fashion.
Configuration

Avamar Grid Configuration

The first step is to prepare the Avamar server for NetWorker backups. This can be broken down into several small tasks.

A - Download the latest NetWorker Linux client release from powerlink.emc.com, and copy it onto the Avamar Utility node.

B – Unpack Networker Linux client package and install

For example:

```
root@avamar01-01:/chris/linux_x86_64/#: rpm -ivh --nodeps lgtoclnt-7.5.2-1.x86_64.rpm
Preparing...  ########################################### [100%]
1:lgtoclnt  ########################################### [100%]
```

To install EMC HomeBase Agent run the below script as 'root' user:

```
/opt/homebase-agent/setup-homebase.sh
```

C- Start the NetWorker Client background process.

```
root@avamar01-01:/chris/linux_x86_64/#: /etc/init.d/networker start
root@avamar01-01:/chris/linux_x86_64/#: ps -ef | grep nsr
root  25226  1  20:04 ? 00:00:00 /usr/sbin/hasrexcud
root  25833 6236 0 20:04 pts/0 00:00:00 grep nsr
```

NetWorker Client background process
Modify the Avamar mcserver.xml file.

Change the Avamar server mcserver.xml "allow_duplicate_client_names" setting from false to true.

To modify the Avamar server mcserver.xml preferences file:

1. Open a command shell.
2. Log on to the Avamar server utility node as the user admin.
3. Ensure that you are not still logged in as root.
4. Load the admin OpenSSH key. For example:
   
   ```bash
   ssh-agent bash
   ssh-add ~admin/.ssh/admin_key
   ```
   
   You are prompted to enter a passphrase.

5. Enter the admin user account passphrase.
6. Stop the administrator server (mcs):
   
   ```bash
   dpnctl stop mcs
   ```
   
   Wait for this command to complete.

7. Open the following file in a Unix text editor:
   
   `/usr/local/avamar/var/mc/server_data/prefs/mcserver.xml`

8. Change the "allow_duplicate_client_names" setting to true.
9. Save your changes.
10. Restart the administrator server (mcs):
    
    ```bash
    dpnctl start mcs
    ```
    
    Wait for this command to complete.
D - Add a NetWorker Domain in the Avamar Server.

Connect to the Avamar utility node using the Avamar Administrator GUI, and add a new domain called “NetWorker.”

**CLARiiON Configuration**

The first step to configure the CLARiiON disk array is to ensure that snapshots outside of the NetWorker framework are possible.

A - Allocate LUNs in the CLARiiON snap reserve LUN pool through the Navisphere GUI

For example (please see the following page):
Select LUNs for each of the SPs from the left column, and add them to the LUN pool.
B – Create a snapshot for the LUN.
In our scenario, drive W:\ on our production host and as2 reside on CLARiiON LUN 0. Using Navisphere, locate LUN 0 and create a snapshot

For example:

Give the snapshot a name, and ensure that the snapshot belongs to a storage group for which the proxy host (in our case tito) is a member.
C – start a Snapview session from the application host as2
For example: open a command line window, and issue the following command.

C:\Program Files\EMC\Navisphere CLI>navicli.exe -h 10.5.160.24 snapview -
startsession test -LUN 0

You must activate a snapshot on this session before you can access it.

D – Move onto the proxy host tito, and activate the previously started session.
For example:

C:\Program Files\EMC\Navisphere Admsnap>admsnap activate -s test
Scanning for new devices.
Activated session test on device E:

E – Once the snapshot is activated, we can be ensured that later steps of performing
backups using our automated workflow will also succeed. At this point, we can clear up
the previous steps from C and D. Issue the deactivate statement from the proxy host tito

C:\Program Files\EMC\Navisphere Admsnap>admsnap deactivate -s test
Deactivated session test on device E:

From the application host as2 issue the following command
C:\Program Files\EMC\Navisphere CLI>navicli.exe -h 10.5.160.24 snapview -
stopsession test

Do you really want to stop the session: test (y/n)? y

At this point we have successfully created a snapshot, imported it into the proxy host,
and then deported the snap and cleaned up all created sessions.
**NetWorker Configuration**

A - Create a deduplication storage node within NetWorker.

The deduplication node is the name of the Avamar utility node, and is the target from which we are going to send our deduplicated data. Since NetWorker is driving the backup process, we need to define this node, so that the PowerSnap proxy host can send the data onto it. We define the deduplication node through the NetWorker console, “Devices” button.

For example:
B - Create NetWorker Group, Snapshot polices, and Pool resources.

- Snapshot Policy: defines the frequency of the snaps (Number of Snapshots), along with how many of these snapshots will be retained by the disk array (Retain Snapshots), and which snapshots will be sent to tape or disk media for storage (Backup Snapshots).

In our example, we are going to perform one snap per day, keep this snapshot on the disk array until the next backup, and send this snapshot also to the backup server for storage (which will actually be directed to the Avamar grid). With the above criteria we will therefore set the following option Number of Snapshot = 1, Retain Snapshots = 1, Backup Snapshots = All.

Continues on the following page.
Create a NetWorker Group resource and place the application host as2 in the created group.
In the Setup tab give the group a name (in our case we are calling it SnapDedupe1) and ensure that the Client retries tab is set to 0.

Create a Pool resource to capture the metadata generated for the backup. It is important to remember that we are going to be conducting a snapshot backup, and also sending the data to the Avamar grid in a deduplicated fashion so no data will actually be stored on the NetWorker backup server.

Meta-data will be stored on the NetWorker backup server. This metadata is necessary to track the backup taken, and required for restore purposes. Select the Media icon on the NetWorker server to create this media pool, and select new. Enter the name and the Group.
Modify your Group resource to tie in all of the previous steps.
Create a disk type device on the NetWorker backup server to capture the metadata (not the actual backup data) that will be generated during the backup.
C – Create the client resource

You must create a client resource for the as2 application host that contains the data we are going to backup AND another client resource for the proxy host machine tito that will be performing the actual data movement. In our example, the Save Set field will be W:\, and the Group will be SnapDedupe1.
Finally, select the Apps & Modules tab. Select the Deduplication backup option (check box), the Deduplication node that was configured from the previous step, along with the PowerSnap specific application information. In this example, you can see that we have to specify the IP address of the CLARiiON SPs, and the name of the proxy host (NSR_DATA_MOVER).
Performing a Backup

Backups are initiated via the NetWorker backup server scheduler, or via an ad-hoc method through the NetWorker GUI Group control.

This backup will create a snapshot on the CLARiiON disk array and be kept until the next 24 hour period when the next backup will occur. At that time, the snapshot will be released, and a new snapshot created. The snapshot sessions can be observed from the CLARiiON Navisphere™ GUI.
You can use the NetWorker GUI on the backup server and a command line tool called “mminfo” to display the contents of the media data after the backup has completed.

D:\Software\powersnap>mminfo -av

<table>
<thead>
<tr>
<th>volume</th>
<th>client</th>
<th>date</th>
<th>time</th>
<th>size</th>
<th>ssid</th>
<th>fl</th>
<th>lvl</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>snap_00</td>
<td>ecc1</td>
<td>2/28/2010</td>
<td>11:31:35 AM</td>
<td>7 KB</td>
<td>2978737127</td>
<td>cr</td>
<td>full</td>
<td>index:as2</td>
</tr>
<tr>
<td>snap_00</td>
<td>as2</td>
<td>2/28/2010</td>
<td>11:28:36 AM</td>
<td>5 KB</td>
<td>2995514320</td>
<td>cb</td>
<td>full</td>
<td></td>
</tr>
<tr>
<td>BRC_CONNDIR_SAV</td>
<td>snap_00</td>
<td>as2</td>
<td>2/28/2010</td>
<td>11:04:28 AM</td>
<td>7 KB</td>
<td>3012291514</td>
<td>cb</td>
<td>full</td>
</tr>
<tr>
<td>snap_00</td>
<td>as2</td>
<td>2/28/2010</td>
<td>11:04:32 AM</td>
<td>2 KB</td>
<td>3029068655</td>
<td>cbP</td>
<td>full</td>
<td>W:\</td>
</tr>
</tbody>
</table>
As you can see, the last line in this output denotes the snapshot that now exists on the CLARiiON disk array, and the amount of metadata stored on the NetWorker server that references this snapshot is only 2KB. The second line from the bottom is the metadata stored on the NetWorker for the deduplication portion of the backup. Although the actual content of drive W:\ was 16MB, the NetWorker server only needs 7 KB to reference this data. The actual data was sent to the Avamar grid.

You can also view the backup on the Avamar grid itself. This is accomplished by logging into the Avamar administrator GUI, and selecting “Backup Management”.

In this illustration, you can see the client name as2, and the amount of data “protected.” There is a difference between the amount of data protected and the amount of data transmitted to the Avamar server. Usually the amount of data transmitted on a well-seeded Avamar grid is very small due to the commonality factoring that deduplication backups take advantage of. The Size column in this view will always show the amount of data that was backed up, even though a small fraction was actually sent.
Performing Recovers

There are several options to recover from this solution. The most common is to use the NetWorker User client software to browse the backup data to select the files and folders for the recovery. There are more advanced restore capabilities such as rollback which will synchronize the snapshot internal to the array back onto the source LUN. This makes the data revert back to the time at which the backup was performed.

For more details on rollback, please consult with the Powersnap administrator guide. Here is an illustration of how to perform a restore.
The restore is now complete, and the selected file is restored back into as2.
General Configuration Tips

- Do not include deduplication clients in the same group as non-deduplication clients.

- Once a deduplication node for a given client has been selected for an initial full backup, continue to use the same deduplication node for the client's backups to take advantage of the data already stored on the server.

- Deduplication backups must be scheduled to avoid the Avamar server read-only periods. Avamar servers spend a percentage of their time running in maintenance mode and may not be available for backups or have limited bandwidth. If NetWorker deduplication backups run during this period, they may hang until the Avamar server resources become available.
Conclusion

With rapidly decreasing cost and increasing storage capability of disk drives, coupled with the explosive growth in data deduplication technology that is now quickly becoming the standard in today’s data centers, tape technology as the primary means of daily backup operations is becoming a less lucrative business proposition then ever before.

For high-end data centers in which production data resides on sophisticated disk arrays such as EMC CLARiiON, it is now possible to merge the best of all these disk based backup technologies into a single backup solution that not only provides array based backup and restore capabilities but backup commonality factoring off store data backup to an Avamar data grid, all through a single point of management tool.

This solution provides the answer to several data center challenges such as shrinking backup windows, increased amounts of data needing to be protected, all while minimizing the impact on the production application servers through server-less backup functionality.
Glossary

**Application host** Computer hosting the target application to be backed up. This application stores its data on the production storage of this host.

**Avamar Administrator** is a graphical management console software application that is used to remotely administer an Avamar system from a supported Windows or client computer.

**Avamar Server** The server component of the Avamar client/server system. Avamar server is a fault-tolerant, high-availability system that efficiently stores the backups from all protected clients. It also provides essential processes and services required for data restores, client access and remote system administration. Avamar server runs as a distributed application across multiple networked storage nodes.

**Backup** An operation that saves data to a volume.

**Backup unit** All save sets which finally result in one snapshot are referred to as a backup unit. Generally, this is a volume group if managed as such, but it could also be a file system or a raw disk.

**Backup volume** A volume used to store backup data. Backup data cannot be stored on an archive volume or a clone volume.

**EMC CLARiiON array** A network storage device connected to a host, either in a SAN environment over Fibre Channel or direct-attach SCSI. Allows for the creation of logical units of varying sizes that can be accessed and used by production hosts as a remote disk device.

**Client** A computer, workstation, or fileserver whose data can be backed up and recovered.
**Client resource**  A NetWorker server resource that identifies the save sets to be backed up on a client. The Client resource also specifies information about the backup, such as the schedule, browse policy, and retention policy for the save sets.

**Command line**  The line on a display screen, also known as a command prompt or shell prompt, where you type software commands.

**Console server**  The software program used to manage NetWorker servers and clients. The Console server also provides reporting and monitoring capabilities for all NetWorker processes.

**Copy-on-write (COW)**  is one of the methods that EMC CLARiiON SnapView uses to perform snapshot technology. It is implemented through EMC SnapView by delaying writes to a file system block until its contents have first been transferred to the snapshot storage area.

**Data mover**  The client system or application, such as NetWorker, that moves the data during a backup, recovery, or snapshot operation.

**Data zone**  A group of computers administered by a NetWorker server.

**Directed recovery**  A method of recovery that recovers data that originated on one client computer and re-creates it on another client computer.

**Disk subsystem**  An integrated collection of storage controllers and/or HBAs, disks, and any required control software that provides storage services to one or more computers, such as EMC CLARiiON arrays.

**Group**  A client or group of client computers that are configured to back up files at a designated time of day.

**Host**  A computer on a network.
Host name The name or address of either a physical or virtual host computer that is connected to a TCP/IP network, including the interface. Each host has a unique IP address.

Manual backup A backup that a user performs from the client, also known as an unscheduled backup or an ad hoc backup. The user specifies the files, file systems, and directories to back up.

Media database A database that contains indexed entries of storage volume location and the life cycle status of all data and volumes managed by the NetWorker server.

NetWorker A network-based EMC software product that backs up and recovers file systems.

NetWorker server Computer on a network running the NetWorker software, containing the online indexes, and providing backup and recover services to the clients on the same network.

PIT copy (point-in-time copy) A fully usable copy of a defined collection of data, such as a consistent file system, database, or volume, that contains an image of the data as it appeared at a single point in time. A PIT copy is also called a shadow copy or a snapshot.

Pool A feature to sort backup data to selected storage volumes.

PowerSnap EMC technology that provides point-in-time snapshots of data to be backed up. Applications that are running on the host system continue to write data during the snapshot operation, and data from open files is included in the snapshots.

PowerSnap Module A software module that exports services of a storage subsystem by interfacing with vendor-specific APIs. This module is independent of applications and backup and recovery interfaces.
**Proxy client** A surrogate client that performs the NetWorker save operation for the client that requests the backup. A proxy client is required to perform a serverless backup.

**Restore** The process of retrieving individual datafiles from backup storage and copying the files to disk.

**Rollback restore** The process by which a specific point-in-time copy (snapshot) of data is restored to the source location by using the hardware's particular capabilities. A rollback restore is a destructive save set restore.

**Save set** A group of files or a file system from a single client computer that is backed up on storage media.

**Scheduled backup** A type of backup that is configured to start automatically at a specified time for a group of one or more NetWorker clients. A scheduled backup generates a bootstrap save set.

**Serverless backup** A backup method that uses a proxy client to move the data from primary storage on the application host to secondary storage on another host. Serverless backups free up resources on the application server by offloading the work of processing shadow copies to a secondary host.

**Snapshot** A point-in-time, read-only copy of data created during an instant backup.

**Snapshot policy** is a set of rules that control the lifecycle of a snap set. The snapshot policy specifies the frequency of snapshots, and how long snapshots are retained before recycling.

**SnapView** EMC CLARiiON software that allows you to obtain a copy of a LUN by creating a clone or snapshot. The clone or snapshot can serve for backup, decision support scenarios, or as a base for temporary operations on the production data without damaging the original data on the source LUN.
SnapView session The process of defining the point-in-time designation by tracking the session with physical memory and invoking copy-on-first-write activity for updates to the source LUN. Starting a session assigns a reserved LUN to the source LUN, and each source LUN can have up to eight sessions.

SnapView snapshot Instantaneous point in time copies of EMC CLARiiON LUNs created by EMC CLARiiON SnapView software. SnapView snapshots are views of a point-in-time image of a source LUN(s). A snapshot occupies no disk space, but appears like a normal LUN to secondary servers. Replication Manager can make SnapView snapshots available to another host through mount commands or restore using the restore capabilities of the product.

Source LUN The LUN containing production data on which the user starts at least one SnapView session, and at least one snapshot.

Storage processor The physical hardware that controls access to the logical units (LUNs) that are created on the EMC CLARiiON array. There are two per array, allowing for failover. Often abbreviated as SP.

Referenced Documentation

-EMC® NetWorker® Release 7.5 Administrators Guide
-EMC® NetWorker® Release 7.4.x NetWorker De-Dupe Node Design and Implementation
-EMC® NetWorker® PowerSnap™ Module for CLARiiON® Installation and Administration Guide
-EMC® Avamar SYSTEM ADMINISTRATION MANUAL