Abstract

This white paper describes the use of EMC Avamar for NAS environments. It also explains how Avamar's fast, daily full backups compliment snapshots for optimized data protection across the enterprise.

July 2013
EMC Avamar for Fast, Efficient NAS Backup, Recovery, and Disaster Recovery

Table of Contents

Executive summary.................................................................................................................. 4
Audience................................................................................................................................. 4
Protecting NAS—today’s reality ............................................................................................. 4
EMC Avamar for NAS............................................................................................................. 5
Avamar NDMP Accelerator..................................................................................................... 5
Simplified recovery.................................................................................................................. 5
Daily replication for disaster recovery................................................................................... 6
Scalability, high availability, and reliability ........................................................................... 7
Snapshots ................................................................................................................................. 7
Snapshots and backups working together............................................................................... 9
Conclusion ............................................................................................................................... 10
Executive summary

Businesses of all sizes rely on NAS systems to share and store valuable information. However, ongoing storage consolidation and exponential data growth make it difficult to meet short backup windows. Traditional solutions that utilize weekly full and daily incremental backups, tape devices, with multi-step recovery are simply unable to keep up. This can lead to data loss, declining user productivity, constrained NAS performance, and overburdened IT staff.

EMC® Avamar® deduplication backup software and systems deliver an industry-leading, innovative solution for NAS backup, recovery, and disaster recovery. Unlike traditional methods, Avamar's integrated deduplication reduces the size of backup data before it is transferred across the network and stored to disk. As a result, Avamar provides fast, daily full backups via existing network links—without the need for a dedicated, high-speed NDMP backup network. And recovery is always a single-step.

Avamar also deduplicates backup data across sites and servers to reduce total disk storage by up to 50x, enabling cost-effective long-term retention on Avamar Data Store servers or EMC® Data Domain® systems. Backup data can be encrypted in-flight and at rest for security.

With fast, daily full NAS backups and single-step recovery, Avamar compliments other data protection methods (e.g. snapshots). EMC’s broad solution portfolio enables users to achieve the ideal recovery point and recovery time objectives, ensuring that data is readily available when needed.

Audience

This white paper is intended for backup administrators or technical staff seeking a more in-depth look at EMC Avamar for NAS systems. The paper assumes the reader has a general understanding of Avamar, the Network Data Management Protocol (NDMP), and snapshot technology.

Protecting NAS—today’s reality

Traditional NAS weekly full backups (level-0) often extend beyond the available backup window. And, in some cases, daily incremental (level-1) backups may take as long as level-0 full backups — especially if only a portion of a large file has changed, since the whole file must be moved. In many cases, NAS consolidation and usage are limited by the backup window, not storage capacity.

The use of tape is also challenging due to unreliable media and hardware, delays in offsite shipments, and the risk of losing unencrypted or sensitive data. And the cost of NDMP backup software licenses for NAS backups can be significant.

Similarly, the common practice of storing snapshots and primary data on the same NAS system can lead to data loss. And while WAN replication is desired, large amounts of data and limited network bandwidth can make it nearly impossible.
Traditional NAS recovery is a tedious, multi-step process, requiring the last good level-0 full backup and subsequent incremenitals to reach the desired recovery point. And retrieving tapes from offsite storage can take days or longer with no guarantee the data is recoverable.

**EMC Avamar for NAS**

Avamar delivers an industry-leading, innovative solution for NAS backup, recovery, and disaster recovery. Unlike traditional methods, Avamar’s integrated deduplication reduces the size of backup data before it is transferred across the network and stored to disk on an Avamar Data Store server or Data Domain system. As a result, Avamar provides fast, daily full backups via existing network links—without the need for a dedicated, high-speed NDMP backup network. And recovery is always a single-step.

Avamar also deduplicates backup data across sites and servers to reduce total disk storage by up to 50x, enabling cost-effective long-term retention. Backup data can also be encrypted in-flight and at rest for security.

For NAS backup and recovery, the Avamar NDMP Accelerator node is utilized in an Avamar environment. It is a dedicated device that includes an Avamar agent and it communicates with the NAS systems using the NDMP protocol.

**Avamar NDMP Accelerator**

When deployed as part of an Avamar solution, the Avamar NDMP Accelerator provides best-in-class backup and recovery for EMC storage systems (e.g. Isilon, VNX family and Celerra) and NetApp storage systems. Using an Avamar NDMP Accelerator node, a level-0 “dump” is performed only once, during the initial full backup.

Subsequent daily full backups are achieved by requesting only level-1 incremental dumps, enabling Avamar to dramatically reduce backup times and the impact on NAS resources and networks. The Avamar NDMP Accelerator maintains the “DUMP_DATE” information between subsequent backups. This information is stored with the backup and when the next NDMP backup is requested by Avamar, the NDMP Accelerator passes the last DUMP_DATE variable to the NAS system to be used as the “BASE_DATE” for the backup. As a result, only those files that have changed since the last backup will be identified and sent in the dump stream to the NDMP Accelerator. This dramatically reduces the number of files to be assembled by the NAS system, thus improving the overall productivity of the device and lowering the duration of the backup. The changed data is then deduplicated so that only new, unique sub-file variable length data segments are transferred across the network and stored to disk. This significantly improves backup performance, while lowering network, storage, and management costs.

**Simplified recovery**

Although Avamar only requests this differential incremental backup (level 1 + the daily updated BASE_DATE) from the NAS system, every backup is indexed and completely available to the administrator as a full backup. That means that every file, directory, subfolder, filesystem and volume to be backed-up will be visible and
available in the restore interface. Using Avamar, recovery is always a single-step, eliminating the tedious process of restoring the last good full and subsequent incrementals to reach the desired recovery point.

File-level granularity of backups and restores are possible with NAS or file-based data. When LUNs are backed-up and restored directly from the storage system, the entire LUN is captured; files within the LUN cannot be identified using NDMP. For LUN backups, it is best to install the Avamar agent on the respective client/host server where the LUN is provisioned.

Restores are simplified because the administrator is interacting directly with the disk based Avamar Data Store or Data Domain server. It is similar to interacting with an online database – data begins streaming back to the NAS system right away. There is no waiting for tape media or drives to stream and seek.

Figure 1: Avamar delivers fast, daily full NAS backup via existing IP networks.

The NDMP Accelerator is uniquely capable of backing up many terabytes of data and tens of millions of files within the backup window. The Avamar solution is superior to other NAS backup methods, which often require lengthy recurring full backups (level-0 dumps), that often exceed the available backup window.

With Avamar, backup data is deduplicated at the NDMP Accelerator with typical daily deduplication rates of 96 to 99.9%, depending on the data type and daily change rate. By only sending the new, unique sub-file data segments across the network, Avamar enables the use of existing IP LAN/WAN links for daily full backups and replication for disaster recovery. This makes the NDMP Accelerator ideal for deployment in remote offices or secondary datacenters too. In addition, administrators can restore any NDMP backup file or directory to a native Linux or Windows host that is running an Avamar agent for added flexibility.

**Daily replication for disaster recovery**

Avamar enables efficient, encrypted, asynchronous replication of data stored in an Avamar server to another Avamar server deployed in a remote location, eliminating the need to ship tapes. Replication can be scheduled to run at off-peak hours to minimize network impact. In the event of a disaster scenario where an Avamar system
becomes unavailable, data can be recovered directly from the replication target, providing a high level of availability.

**Scalability, high availability, and reliability**

Unlike most solutions, Avamar uses a scalable grid architecture that provides linear performance and on-demand capacity increases by simply adding storage nodes. Each incremental node increases CPU, memory, I/O, and disk capacity for the entire environment. Existing backup data is automatically load-balanced across the newly added storage node while online.

Avamar utilizes a patented redundant array of independent nodes (RAIN) architecture that provides high availability and fault tolerance across Avamar servers. In addition, Avamar system integrity is verified twice daily via internal system checkpoints. Avamar also verifies the recoverability of all backup data daily, and backups can be efficiently replicated offsite to another Avamar server via existing IP WAN links.

**Snapshots**

In many cases, vendors position snapshots as a replacement for regular backups. This is frequently suggested by vendors that only offer snapshot capabilities within their NAS storage arrays and do not offer backup solutions. While snapshots are viable for specific applications, they are not a substitute for recurring backups.

Utilizing snapshots involves retaining some number of disk-based or volume-based snapshots on the storage array for recovery in the event of data loss or corruption. The following graphic provides a high level overview of snapshot methodology:

![Figure 2: Example - snapshots at a source disk volume, replicated to a target secondary system, and retained for extended period of time.](image-url)
Although many NAS systems provide the ability to quickly create and store snapshots, using those snapshots for extended backup retention can create a number of management (operational) and financial issues. These include:

**Operational Differences:**

- Volume or filesystem corruption can result in the loss of all snapshots or "backups".
- Replication to another location requires additional expensive software.
- Fixed quantity (number of snaps) and disk capacity constraints for snapshots within volume containers limits retention periods.
- In some cases, recovery to a specific snapshot deletes all intermediary snapshots.
- “Overfilling” the target volume requires a variety of actions to be taken such as migration, volume splitting or data deletion.
- Snapshots kept in volumes will become unavailable if the containing volume goes offline for any reason. Volume-checking maintenance is time-consuming and has unpredictable results.
- Requires a deep understanding of the storage array, snapshot space management, performance and capacity planning.

**Financial Challenges:**

- Some vendors’ snapshots are kept in the same volume space as the production data. This becomes expensive as inactive blocks captured in snapshots have the same cost per GB as active production blocks.
- Data deduplication may only occur within a given volume or file system container, not globally. Any blocks referenced by snapshots of the volume prior to deduplication are not candidates for deduplication, limiting the value.
- Replication requires 100% more capacity plus the incremental changes. There is a requirement that the baseline, active volume must be transferred and reside on the target prior to the transfer of subsequent snapshots.
- Snapshots are not vendor agnostic and in many cases, other vendor's storage is not supported. This can create vendor lock-in for application protection and the inability to utilize the most cost effective storage.

In addition, replication may involve the process of propagating a snapshot to another storage array. If the structure of the snapshot is compromised, the replicated target will follow suit. Furthermore, all these technologies rely on the same underlying file system that can experience propagating corruption issues. This can become an issue
since any one failure at the root of the dependency tree (in this case the local NAS snapshot) has the potential to propagate to its codependents.

A data protection strategy can only be considered reliable when it incorporates technology and methods that decouple and isolate the backup data ("last line of defense") from the primary data under protection. This requires that the software used to protect the storage is completely different from the software running the storage. This isolation allows the backup storage to be physically and logically different than the primary storage, hence decreasing the risk of data loss.

So "are snapshots true backups?" If simple inode pointers to blocks with previous time stamps are considered a "backup", then perhaps, but the majority of users who rely on the resiliency and consistency of their backups disagree. Snapshots, themselves, do not execute integrity checking. This is an activity that must be performed by the host system. And in the case of snapshots, disk arrays do not perform this activity for file-level data. Sometimes file system or volume-level integrity checking can be initiated, but this is a CPU-intensive process that requires manual execution and scheduling – it is generally not performed due to this reason. The lack of routine consistency checking can compromise data integrity.

In addition, there are no automated processes to defragment a volume or file system. In some online storage arrays, executing data deduplication will leave duplicate blocks (scattered around the disk pool) freed for future use. Similarly, snapshots will create pointers to potentially scattered blocks and the impact of this fragmentation causes degraded performance over time.

**Snapshots and backups working together**

For specific applications, snapshots can provide point-in-time copies of blocks that can be leveraged to quickly revert corrupted or deleted data blocks. And snapshots can be configured to occur many times per day in order to provide granular Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO). These point-in-time copies allow for near-immediate restores from data blocks maintained with very short retention periods. But as noted above, snapshots also have significant operational and financial considerations.

An ideal solution improves SLAs, NAS system utilization, and productivity while reducing costs and the risk of data loss. EMC offers an ideal solution that can be applied to a variety of datacenter and remote office locations. For example, some offices may have local NAS systems servicing local clients via NFS exports and CIFS shares. For those locations the NDMP Accelerator is deployed with the NAS system to perform deduplication and daily full backups directly from the NAS system. Other offices may not have NAS systems and therefore only the Avamar agent needs to be installed directly on the local servers. No ongoing intervention from local staff is required, nor are there any NDMP or client licenses required. It is a centrally managed, extremely network-efficient model that is well suited for any sized environment.
The use of tiered “protection offerings” to support a variety of SLAs is recommended, since not all data volumes are equally valuable. Sample tiers may include:

- **Tier 1:** local snapshot retention with aggressive schedule + near-time DR replication + Avamar backup with NDMP Accelerator, replicating to DR site
- **Tier 2:** local snapshot retention with less aggressive schedule and fewer retained copies + daily or weekly replication to DR site + Avamar backup with NDMP Accelerator, replicating only critical data to DR site
- **Tier 3:** No local snapshots, Avamar backup with NDMP Accelerator, replicating only critical data to DR site
- **Tier 4:** No local snapshots, only routine backup to Avamar using NDMP Accelerator, no replication of backups

The immediate cost and operational efficiencies realized with the combined architecture and the tiered protection model include:

- Improved SLAs by leveraging the Avamar NDMP Accelerator for fast, daily full backups, one-step recovery, and disaster recovery
- Increased NAS productivity and use by reducing the required backup window
- Sub-file backup data deduplication to reduce required storage, network bandwidth, datacenter floor space, power, and cooling
- Tiered service offerings can reduce the amount of data requiring additional protection (such as replication, multiple snapshots and long-term retention)
- Less reliance on proprietary vendor snapshot management solutions for maintaining backup data reduces vendor lock-in and provides flexibility

**Conclusion**

Avamar deduplication software and systems are unique in the industry and best in class for NAS backup, recovery, and disaster recovery. Avamar provides fast, daily full backups and enables efficient daily replication to offsite locations using existing IP networks. And by eliminating backup bottlenecks, Avamar provides the freedom to consolidate storage and optimize NAS systems—without the hassle of limiting the number and size of files or volumes due to backup performance limitations.

As a result, Avamar enables even greater levels of NAS consolidation and facilitates larger NAS volumes without the risk of leaving data unprotected. And, there is no charge for NDMP backup software licenses. With Avamar, IT managers can reduce the amount of time and resources spent on NAS backup, while reducing their total cost of ownership. Avamar also compliments snapshots, enabling users to optimize their desired recovery point and time objectives, while ensuring data is readily available.

Developed to solve the challenges associated with traditional backup, Avamar is a proven enterprise solution, ideal for protecting NAS systems, virtual environments, remote offices, desktop/laptop systems, and business critical applications.