

EMC Avamar for NAS - Accelerating NDMP Backup Performance

John Webster
June, 2011

Technology Insight Series

Evaluator Group

enabling you to make the best storage decisions



Copyright 2011 Evaluator Group, Inc. All rights reserved.

Introduction/Executive Summary

It almost goes without saying that storage administrators continue to grapple with managing ever increasing volumes of data. On average, Evaluator Group sees a 50% annual growth rate in the amount of data stored among its user clients. Certainly one of the first places we see the stress that rapid data growth creates is on data protection processes.

The limitations of traditional NAS backup processes that require lengthy, recurring level-0 full backups and daily incrementals via NDMP can leave data unprotected when backup windows are exceeded and prevent organizations from optimizing their NAS environments. Here we review a unique solution from EMC Avamar for NAS system backup. Avamar for NAS implements data deduplication coupled with appliance-resident software that accelerates backups using the NDMP protocol.

NAS environments can exhibit a relatively high degree of data redundancy. Eliminating duplicate data from the backup stream will accelerate data protection processes and reduce the capacity required to store backup copies. Use of NDMP coupled with technology that accelerates NDMP backup processes will drive further data protection efficiency and effectiveness.

Bottom line: Faster and more reliable backup and recovery processes promote a more efficient and effective data protection environment. In addition, they allow IT administrators to consolidate and grow NAS file shares beyond their current limits.

The NAS backup problem

The limited scalability of traditional single system NAS platforms has been well documented as an issue. Less well documented are the issues that can arise when the backup and recovery process can't keep up with the growth.

We have seen that, when conducting NAS weekly full backups (dump level-0) under NDMP, the full backup process can take days to complete depending on the backup application and network bandwidth. In one instance we saw that dump level-0 backups were taking seven days to fully complete. Doing "the math" makes it evident that that this is not a sustainable practice.

In some cases, conducting daily incremental (dump level-1) backups may be just as problematic when protecting large files. Even if only a fraction of a large file has changed, a copy of the whole file must be moved to the data protection storage system. On a per file basis, this process may take as long as a level-0 full backups and is very inefficient. NAS consolidation and usage are often limited by the backup window, not the NAS storage capacity.

Good data protection operational practice also includes knowing that the data recovery process will run smoothly and efficiently when called upon and that SLA's can be met. Recoveries that require retrieval of the last good level-0 full backup and subsequent incrementals to reach the desired recovery point can be time consuming and error prone. Recovery times are elongated and the potential for recovery problems is amplified if backup tapes must be retrieved from offsite storage.

Evaluator Group recognizes that storage administrators often use NAS-based snapshots to provide data protection capabilities. The snapshots are combined with other NAS-resident data replication features for a combination of local and remote replication as well as application-specific modules to provide on and off-site data retention requirements. However, snapshots should be considered as a complement to backup, since they offer more granular recovery points for some applications, and not a replacement for backup processes within the overall business continuance plan. Also, snapshots often reside on the same disk array as the primary storage, leaving data at risk in case of an array failure. And replication licenses along with the cost of primary storage can be expensive. Depending on the environment, the number of retained snapshots may be limited, introducing issues when snapshot copies must be retained for some predefined time period.

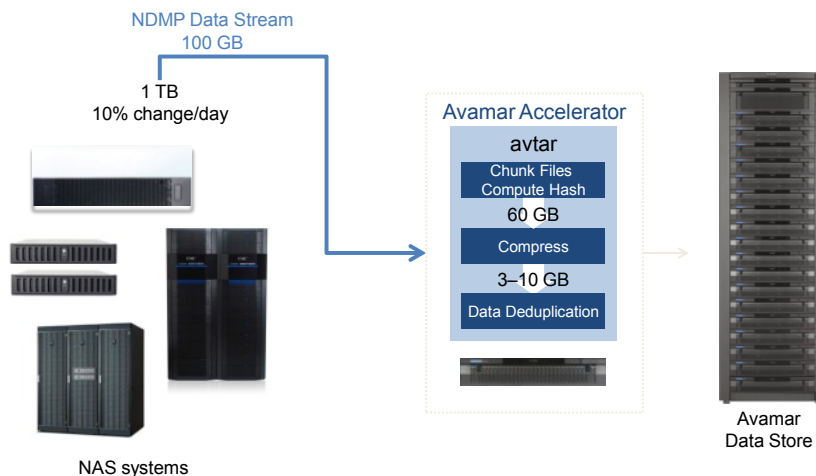
EMC Avamar addresses these issues with data deduplication, efficient WAN replication, and disk-based backup and recovery systems.

Avamar NAS Data Protection

EMC's Avamar for NAS environments is an integrated data protection (backup, recovery, disaster recovery) software and hardware solution optimized for NAS data protection. It consists of:

- Avamar NDMP Accelerator Node – an Avamar device that receives daily NDMP backup streams from NAS systems and performs data deduplication.
- EMC Avamar Data Store—a disk-based, scalable server that stores deduplicated backup data from multiple locations, and can be configured with RAIN (Redundant Array of Independent Nodes) for high availability and reliability
- Avamar management console—handles backup scheduling, user-defined data protection policies, data retention periods, and recovery.
- Existing IP LAN/WAN links

Example: How NDMP Data Reduction Works



In the example above, if a 1 TB NAS system has a daily change rate of 10 percent, then 100 GB of new data would be sent through NDMP during daily level-1 dumps. The Avamar NDMP Accelerator analyzes this data and breaks it into sub-file, variable length data segments. Each segment is assigned a unique ID, and only new, unique segments are transferred to an Avamar server during daily full backups. Data is compressed and deduplicated for efficiency, enabling faster transfer across existing WAN/LAN links.

The dramatic reduction in backup completion times as reported by users is the result of a number of Avamar for NAS features noted above including data deduplication, plus the elimination of recurring level-0 full backups.

NAS backup data deduplication

For NAS environments, Avamar deduplicates NDMP backup data streams at the Avamar NDMP Accelerator node. By comparing sub-file data segments to those already stored on a central Avamar Data Store, only new and unique segments are transferred during daily full backups via IP LAN/WAN links. This method reduces the size of backup data before it is transferred to the Avamar Data Store. As a result, users continue to leverage existing IP LAN/WAN connectivity for data protection and achieve a dramatic reduction in backup completion times.

Daily full backups

A full (level-0) NAS backup is performed only once during the initial setup of the Avamar NDMP Accelerator node. Subsequent daily backups only request level-1 incremental dumps from the NAS systems. Avamar's technology deduplicates the backup data and also creates daily full backup images that can be recovered in one-step. Avamar eliminates the need for recurring, lengthy level-0 full backups and the tedious process of restoring from the last good full plus subsequent incremental backups to reach the desired recovery point.

Daily replication for disaster recovery

Avamar also enables 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.

NDMP

When configured with the Avamar NDMP Accelerator node, users can take advantage of NDMP's direct NAS-to-data protection device connectivity.

An Avamar NDMP Accelerator node presents an NDMP-capable target to a NAS system configured for NDMP backup. It acts as a pass-through conduit from one or more NAS systems to the Avamar Data Store.

Connectivity to the NAS system can be over existing or dedicated Ethernet LAN connections. However, IT administrators should consider bandwidth requirements for both the backup and restoration processes when deciding which route to go.

Supported NAS systems include EMC VNX series, EMC VNXe series and EMC Celerra, along with NetApp NAS systems. Avamar also supports backup for NAS systems based on Microsoft Windows Storage Server (WSS). Backup for WSS systems utilizes an embedded Avamar agent and does not require an Avamar NDMP Accelerator node.

The Avamar for NAS solution with NDMP acceleration and data deduplication dramatically improves backup and recovery performance and efficiency vs. traditional implementations. It enables daily full backups and one-step recovery processes via existing IP networks, greatly reducing the stress that storage administrators normally experience when managing data protection environments (see real world customer experience sidebar). And it enables even greater NAS usage and data consolidation in environments where backup bottlenecks limit NAS scalability.

One Storage Administrator's Experience

One Avamar user we spoke to described the following scenario: His computing environment supports a significant amount of processing related to US Department of Defense contracts. Two NetApp NAS systems running ONTAP 7.1 serve as storage for these applications. Total NAS primary storage capacity is 25TB.

Backup processes were managed by IBM Tivoli TPC. Backups were performed at the sub-volume level over a dedicated Ethernet LAN with sub volumes being either 250 GB or 2TB in size.

The business environment requires that all backups be retained for 3 years so the company is using tape to satisfy this requirement.

Prior to installing Avamar, a full backup would take seven days to complete as a best case. Job failures or situations where the job simply ran out of tape elongated the backup operation. As a result, full backups were running continuously and primary data was at risk. By implementing an Avamar NDMP Accelerator Node between each NAS system and an EMC Avamar disk-based Data Store, full backup completion times were reduced to between six and nineteen minutes.

Thirty days worth of full backups are now retained on the Avamar Data Store before being rolled-off to tape for compliance purposes. Recoveries are now processed from disk and are nearly instantaneous. Primary data is no longer at risk because of an inability to completely process a full backup. Implementing the Avamar NDMP Accelerator provided the company with its first experience with both a data protection solution that included data deduplication to disk as a backup target, and NDMP processes that were further accelerated by EMC Avamar. Hence, the very dramatic reduction in full backup completion times and recovery performance.

Data Protection Operational Impact

We note that orders of magnitude reductions in backup completion times have been experienced by NAS users who replaced traditional LAN-based data protection processes with EMC Avamar. We believe these benefits come as a result of three factors working in combination:

- Avamar data deduplication at the sub-file level, which reduces the amount of data transferred during daily full backups and the amount of data stored to disk-based Avamar servers
- The ability to request daily incremental NDMP data dumps after the initial full backup is completed on the first day, and present all backups as daily fulls to the IT administrator

- The ability to leverage NDMP for direct NAS system to Avamar NDMP Accelerator node connectivity, avoiding the use of enterprise LANs that support application users and experience periodic congestion (see NDMP sidebar). Also avoided are traditional NAS backup processes that deploy client-side agents and require recurring level-0 full backups that often exceed available backup windows.

NDMP

Network Data Management Protocol (NDMP) is an open standard that is currently administered by the IETF standards body. NDMP enables two very important functions in the context of NAS data protection:

NDMP allows for the separation of backup and restore data streams from the control information emanating from data protection applications that manage these data streams. This addresses the need for storage administrators to manage data protection processes in a heterogeneous storage environment from a centralized location.

NDMP supports “server-free” backup—backup data streams flow directly from the NAS system to the backup target, typically a VTL. Therefore, congesting the enterprise LAN with backup traffic is avoided.

Additionally, NDMP supports:

- Three-way backup from a NAS system via LAN to another NAS system with a locally attached data protection system
- Backup from a UNIX or Windows server via LAN to a NAS system with a locally attached data protection system
- Backup from a NAS system via LAN to a UNIX or Windows backup server with a locally attached data protection system

Additionally, Avamar for NAS is a more efficient alternative to recovering from data loss or corruption. The Avamar methodology begins with the creation of a full backup as a starting point, and then requires only level-1 incrementals on a daily basis. When a recovery is needed, the Avamar backup copy is presented as a daily full backup, which significantly simplifies and speeds recovery of NAS data whether a storage administrator is restoring an entire file system or just individual files.

By reducing or eliminating the storage administrators existing backup bottlenecks and limitations, storage architects and other IT administrators (DBAs for example) can manage the NAS environment in more efficient ways. When managing individual NAS system capacity, administrators generally set limitations on growth within individual NAS systems based on overall backup performance requirements. NAS capacity is an important consideration here because backup performance is impacted by the total amount of data contained within the NAS system and the communications bandwidth between the NAS system and the target storage device. Removing backup bottlenecks can allow storage administrators to grow the number of files or volumes per system. As a result, administrators can set the capacity bar higher on individual NAS file shares, allowing for more efficient use of NAS resources and/or allowing for some degree of NAS device consolidation.

Recommendations for IT Administrators

Evaluator Group recommends NAS Users leverage NDMP along with a solution that eliminates the need for recurring level-0 weekly full backups to address long back up times and improve their data protection. Those storage administrators using traditional backup applications and who have not yet leveraged NDMP will likely see an order of magnitude reduction in backup completion time when switching to the Avamar NDMP Accelerator.

For administrators already using NDMP for NAS data protection, we note that the cost of NDMP backup software licenses for NAS backups can be significant using traditional methods. NDMP users should review their software licenses related to NDMP-based data protection and the use of snapshots for ways to reduce cost. EMC Avamar does not charge for NDMP licenses when protecting EMC and NetApp NAS systems.

We are seeing NAS increasingly deployed in virtualized server environments. Here, the performance of data protection (backup and recovery) systems becomes even more important as exposure to data loss from a single NAS system can impact multiple applications and an increasing number of users. This combined with file data growth at magnitudes which at times seem incomprehensible, necessitate a diligent review and update of data protection approaches.

Conclusion

The introduction of NDMP was a significant step forward in increasing the performance and reliability of enterprise data protection capabilities. EMC Avamar's NDMP Accelerator solution provides a combination of data deduplication and backup acceleration for even greater NAS backup performance, efficiency and reliability.

Avamar deduplication software and systems provide a single backup, recovery, and disaster recovery solution for the entire enterprise. For VMware, remote offices, enterprise applications, and desktop/laptop systems Avamar deploys lightweight agents to deduplicate data at the client, delivering fast, daily full backups across existing network links, and one-step recovery. And as we have seen in enterprise data center environments, it allows storage administrators to maintain the data protection umbrella over NAS file shares as they take on growing volumes of business-critical data.

About Evaluator Group

Evaluator Group Inc. is dedicated to helping [IT professionals](#) and vendors create and implement strategies that make the most of the value of their storage and digital information. Evaluator Group services deliver [in-depth, unbiased analysis](#) on storage architectures, infrastructures and management for IT professionals. Since 1997 Evaluator Group has provided services for thousands of end users and vendor professionals through product and market evaluations, competitive analysis and [education](#). www.evaluatorgroup.com Follow us on Twitter @evaluator_group