White Paper

Addressing NAS Backup and Recovery Challenges

By Terri McClure

September, 2010

This ESG White Paper was commissioned by EMC and is distributed under license from ESG.

Contents

Introduction

Meeting service level agreements (SLAs) for full Network Data Management Protocol (NDMP) backup and recovery is often a challenge. Performing a full NDMP backup is more likely to exceed the window of time allocated to performing backup—potentially disrupting application availability and end-user productivity. The process also consumes significant network and system resources.

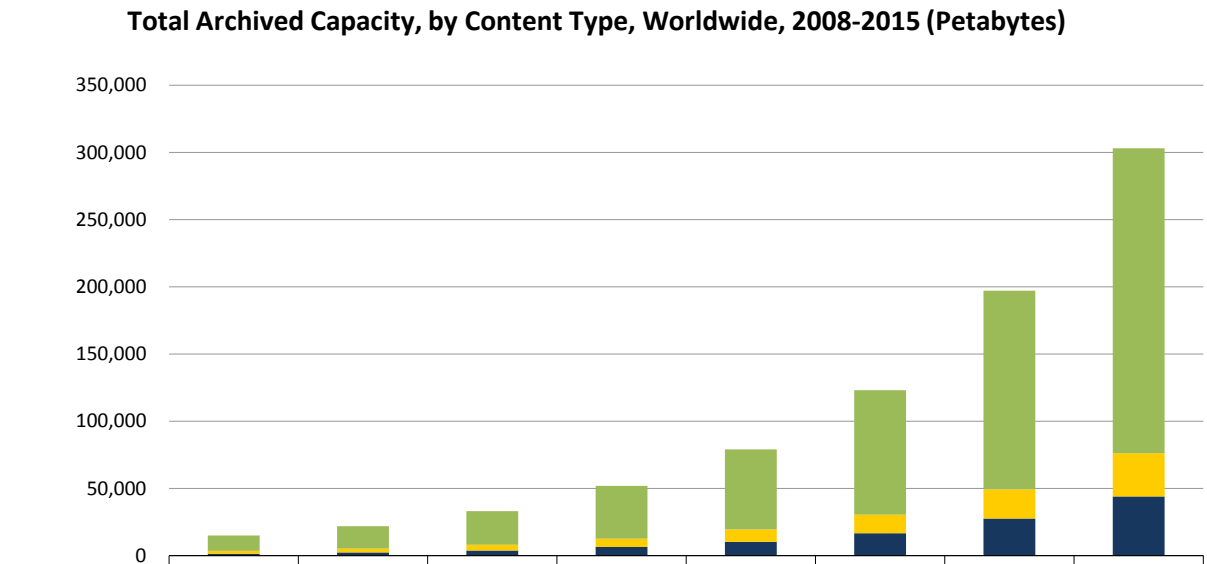
EMC utilizes integrated global, client-side data deduplication and replication to back up Network Attached Storage (NAS) NAS systems in hours versus days. This creates tremendous efficiency over more traditional approaches to protecting NAS systems, saving time in backup and recovery operations and minimizing the use of network bandwidth and storage capacity.

NAS and Data Protection Challenges

The majority of capacity under management in the commercial sector will be born as file-based, rich digital content. File data encompasses a wide range of document types, including Word, Excel, PDF, PowerPoint, scanned images, CAD/CAM, source code, check images, and x-rays as well as Internet-era rich digital content such as video, audio, blogs, and wikis.

Unstructured data and file shares are growing exponentially. ESG research indicates that data growth in this area is exceeding that of other data types—estimating 226 exabytes of archived file data by 2015, dwarfing database- and e-mail-based archive data (see Figure 1).[1]

Figure 1. Projected Archive Data Growth By Type



	2008	2009	2010	2011	2012	2013	2014	2015
Unstructured	11,430	16,737	25,127	39,237	59,600	92,536	147,885	226,716
Database	1,952	2,782	4,065	6,179	9,140	13,824	21,532	32,188
E-mail	1,652	2,552	4,025	6,575	10,411	16,796	27,817	44,091

Source: Enterprise Strategy Group, 2010.

The proliferation of file data has created challenges for management and backup/recovery. In fact, improving backup and recovery is a "top three" 2010 IT investment[2] as reported by respondents to a recent ESG research survey. Complete, reliable, high-performance backup and recovery is a persistent problem for IT organizations and it's often exacerbated by data growth. The top data protection challenges are tied to data growth: keeping pace

[1] Source: ESG Research Report, *Digital Archiving: End-User Survey & Market Forecast 2010-2015*, July 2010.
[2] Source: ESG Research Report, *2010 IT Spending Intentions Survey*, January 2010.

with the capacity of data to protect and reducing backup and recovery times as well as covering backup hardware costs.[3]

NAS environments contribute to growth in storage capacity since NAS environments have a high degree of redundancy in file shares. Backup solutions that don't filter and eliminate duplicate data risk consuming a lot of storage and impacting network links, as well as not completing backup processes within the allotted timeframes.

Another issue is that many NAS systems are remotely located. Local backups are difficult as these locations are often not staffed with backup administrators. Centralized backup is impractical due to the potential impact of daily backup capacity on network bandwidth.

NAS Market Overview

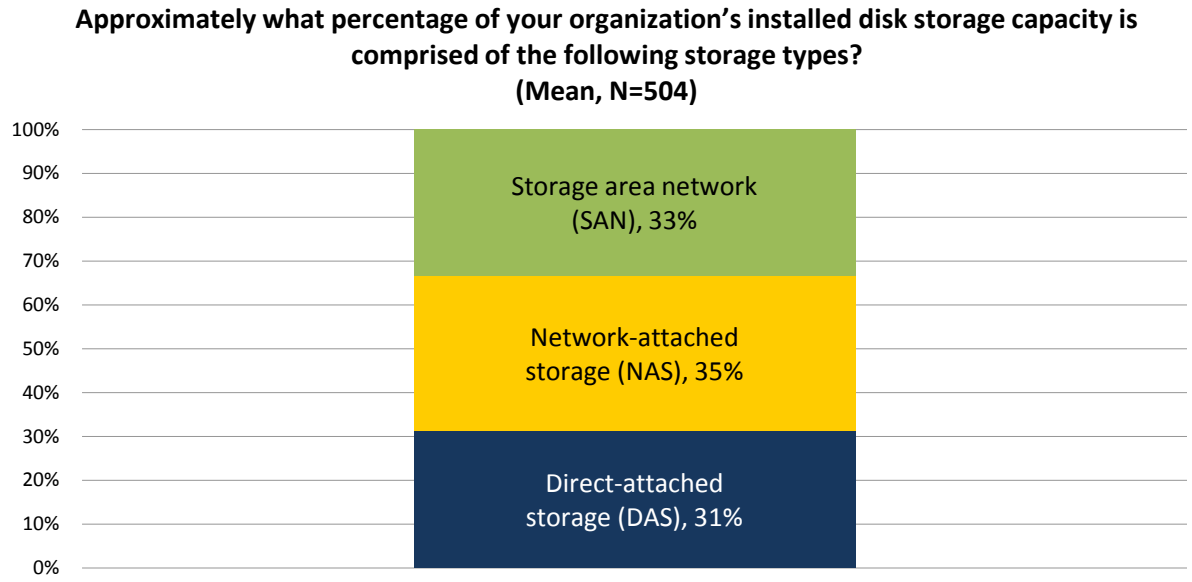
Storage and management of file-based data gave rise to technology optimized for that function. A NAS device is purpose-built to store files and perform file serving tasks. NAS systems are typically dedicated, high-performance data storage systems running embedded proprietary operating systems optimized for disk IO. NAS device hard drive space accessed via an Ethernet connection can be used as a primary storage device by any or all host devices on that network.

NAS Adoption

The massive growth of file data is driving increased adoption of NAS. ESG research revealed that 55% of midmarket organizations and 70% of enterprise organizations are using NAS for primary storage.[4] Furthermore, ESG found that new NAS storage system purchases ranked in the "top five" storage investment initiatives for 2010.[5]

As shown in Figure 2, the capacity of NAS storage (35% of ESG research respondents' total disk storage capacity) outweighs storage area network (SAN) storage (33%) and direct attached storage (DAS) (31%). SAN systems transfer data over the network in the form of disk blocks whereas NAS systems transfer file data. SAN and NAS systems are networked whereas DAS is dedicated to the file server to which it's attached.

Figure 2. Storage Capacity by Storage Type



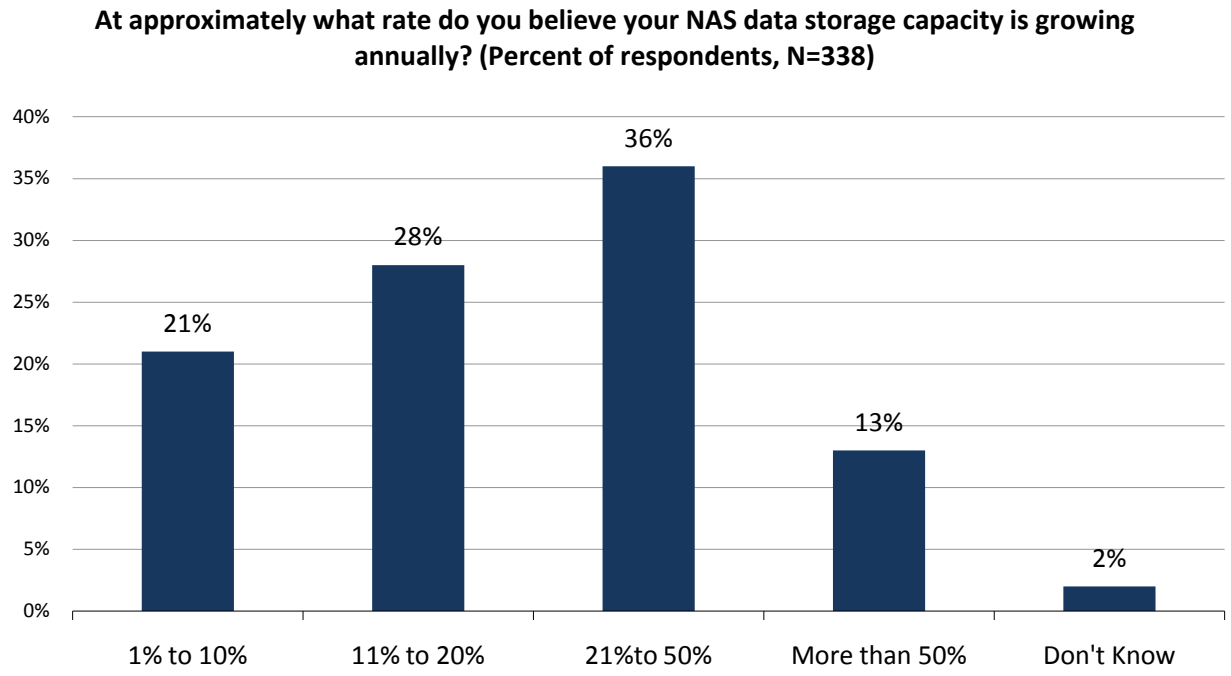
Source: Enterprise Strategy Group, 2010.

[3] Source: ESG Research Report, *2010 Data Protection Trends*, April 2010.
[4] Source: ESG Research, *iSCSI Market Update*, 2009.
[5] Source: ESG Research Report, *2010 IT Spending Intentions Survey*, January 2010.

Many organizations use NAS systems because they are easy to install/deploy, are affordable, and are reliable, ensuring support time is kept to a minimum. Other factors driving adoption include ease of capacity expansion and ease of administration, especially when it comes to serving up files to heterogeneous clients.

ESG research respondents using NAS storage cite consistent and persistent growth in the storage medium. Forty-nine percent of research respondents cite growth rates of NAS storage capacity at more than 20% per year (see Figure 3).[6]

Figure 3. Storage Capacity by Storage Type



Source: Enterprise Strategy Group, 2008.

Backup and Recovery in NAS Environments

Traditional file/server backup architectures utilize a software agent running on each server hosting data. The backup client agent facilitates data capture and transfer, and communicates with the central backup engine. Since a NAS device runs a proprietary operating system and is dedicated to serving files, it doesn't handle backup software components such as backup agent technology that is typically needed to facilitate file- or application-level backup/recovery.

Backup and Recovery Alternatives

Proxy Backup

One of the simplest methods to back up NAS environments is to maintain a proxy backup system on the network where the backup agent is installed. Each NAS volume requiring backup can then be mounted via Common Interface File System (CIFS) or Network File System (NFS) sharing protocols to the server hosting the backup agent. This approach is less efficient due to excessive use of network bandwidth to move data between systems, which has the potential to introduce performance degradation for backup operations.

[6] Source: ESG Research Report, *2008 Enterprise Storage Systems Survey*, November 2008.

NAS System-Specific Data Protection

NAS systems may offer data protection features to facilitate backup and recovery of data. For example, a snapshot capability provides local or remote block-level point-in-time copies. However, the number of snapshots may be limited (up to 250, for example), which could introduce retention issues. Replication provides local or remote, synchronous or asynchronous mirroring.

These native tools may require added license fees and will likely be limited to operating on homogenous storage systems. Snapshot copies typically capture only changed blocks after the first full copy, minimizing the capacity of data captured, transferred, and stored. With greater frequency of snaps, recovery point objectives (RPOs) can be improved. One of the hidden costs of this approach is that the NAS device may pre-allocate disk space for snapshots (estimated 20% of total volume) on more expensive primary storage.

NDMP Backup

The Network Data Management Protocol (NDMP) is an industry-standard interface that provides best practice backup and recovery for NAS systems. An NDMP approach enables the backup server to communicate directly with the NAS appliance and transmit data to the specified backup storage device.

The protocol eliminates the need for backup vendors to write device-specific code for NAS devices to facilitate backup. Any third-party backup application running on another server on the network has access to a set of commands to read data and write it to media (for backup) or read data from media and write an NDMP data stream (for restore). The advantage of this approach is that it eliminates the load from production servers and keeps backup traffic off the production LAN since the NAS device sends individual files and metadata directly to the backup media.

There are some tradeoffs associated with this technique, including additional costs for NDMP licenses from the backup software vendor. Most traditional backup solutions leveraging NDMP require a weekly full backup (level 0) and daily incremental backups (level 1). Completing full backups within the prescribed window of time is more often a challenge, given the massive amount of data on NAS systems today. Full backups that run too long can hurt end-user productivity and application availability. Difficulty in completing the weekly full backup will impede NAS consolidation goals, which means IT organizations can't maximize their NAS investment.

Traditionally, scans of large file systems add hours to the incremental backup process because the entire file system (many millions of files and directories) must be traversed to check for changes (defined as new, modified, or deleted files). Any change to a large file requires backup of the entire file (not just the changes). Redundant data will also impact backup time and storage capacity.

Recovery processes are also affected. Recovering from a backup failure is a much slower process since checkpointing is not supported. Backup software must read sequentially through the backup image to locate the requested information. Recovery could require multiple steps to restore a directory, individual file, or set of files unless direct access restore (DAR) is supported by the backup vendor.

EMC Avamar's Innovative Solution for NAS Backup and Recovery

Avamar Backup and Recovery

EMC Avamar is backup and recovery software with integrated data deduplication technology. It identifies and eliminates redundant sub-file data segments at the client system, reducing the size of backup data before it is sent over the network, providing fast, daily full backups even across slow or congested links. It also reduces required backup disk storage via deduplication across sites and servers globally—making disk-to-disk backup more financially feasible than a disk-to-tape approach.

The creation of offsite copies of data for disaster recovery purposes is facilitated via site-to-site deduplicated replication over a WAN connection. Data can be encrypted in flight and at rest for added security and privacy. Avamar backup data can be exported to physical tape media as well.

The solution is ideal for fast and secure backup and recovery of not only NAS environments; EMC Avamar is well-suited for VMware virtual server environments, desktop and laptop systems, and remote offices.

How Avamar Works

Deduplication at the client is the technology that drives Avamar's advantages and is the reason that Avamar is able to solve key enterprise data protection challenges. Enterprise data is highly redundant, with identical files and sub-file data segments stored within systems and across systems company-wide. Traditional backup software magnifies the problem by storing redundant data over and over again. Avamar solves this problem by viewing primary data in sub-file chunks called data segments. The software generates and assigns each data segment a unique ID based on its content, which is used to compare it with other data segments that have already been backed up. Only new, unique data segments are transferred during a backup operation, ensuring that only a single instance of each segment is stored in a central location. Avamar eliminates redundant backup data at the client—before that data is sent across the network. It also deduplicates across sites and servers, hence the term "global." As a result, Avamar provides daily full backups and dramatically reduces backup times, the impact on the backup network, and the required back-end disk storage for cost-effective retention on disk and fast, single-step recovery.

Avamar Improves NAS Backup and Recovery

To address many of the aforementioned challenges of protecting NAS-resident data, EMC offers its Avamar NDMP Accelerator. A single, dedicated Avamar NDMP Accelerator node uses the NDMP protocol and acts as a pass-through conduit from one or more NAS devices to the Avamar server. Deduplication at the Avamar NDMP Accelerator Node creates efficiency by reducing demand on network and system resources and improving performance significantly over traditional NDMP full backups.

The result? With Avamar, weekly full backups are a thing of the past. The initial Avamar NDMP backup of the NAS environment is a full backup (level 0). Thereafter, Avamar only requests daily incremental (level 1) backups, which take significantly less time to complete, but they are stored as full backups on the Avamar system. Since each Avamar backup is stored as a full backup, recovery becomes a simpler and less time consuming one-step process. Avamar eliminates the need to perform multi-step restore (a full restore plus one or more incremental restores) to reach the desired recovery point, delivering immediate one-step recovery to the original or an alternate location with a single operation.

No Compromises

EMC Avamar and its NDMP Accelerator Node enable faster NAS backup and recovery without compromises. There are several key advantages to Avamar-driven NDMP backup and recovery:

Fast, daily full backups. Because Avamar requests only modified (level 1) data for daily backup and deduplicates this data locally, the vendor claims that backups are up to 10x faster than traditional NDMP backups. This means that daily full backups occur in hours versus days, and the need for recurring, lengthy level 0 backups is eliminated.

Single-step full recovery. Avamar always creates a full backup image, so there is never a need to reconstruct a backup image from the last full and one or more incremental backups. This simplifies and speeds recovery of NAS data. Avamar provides flexible options for recovery, including the ability to restore the entire file system or just individual files. And the restore can be directed to any other Avamar client on any supported operating system.

Minimized impact on network and backup storage resources. By dealing only with modified data, Avamar significantly reduces the volume of backup data transferred and stored. This makes it more practical to back up remote NAS systems centrally.

Supports NAS consolidation. Avamar enables greater levels of NAS consolidation. The optimization provided with the Avamar Accelerator Node removes backup bottlenecks.

Availability and integrity ensure recoverability. Built-in high availability and daily data integrity checking provide a high degree of confidence in data recoverability. Replication of encrypted data to a remote Avamar server provides a simple and secure disaster recovery solution.

No limitation on retention. Unlike a NAS snapshot-only protection strategy, Avamar has no limitations on the number of point-in-time copies (backups for Avamar) that can be retained.

Integrated central policy engine, interface, and index. Application-specific backup is integrated from a single and central policy engine, managed from a single user interface, and tracked via a single index.

The Bigger Truth

Data growth—especially file data—is driving greater adoption of NAS. The NAS architecture creates challenges for traditional client/server backup solutions to efficiently perform backup and recovery of NAS-resident data. It has forced IT organizations to mount each NAS volume from a server on the network which hosted a backup agent. The performance degradation and operational overhead associated with this approach led to the development of the NDMP protocol.

The NDMP protocol streamlined backup and recovery operations, but still has limitations today—specifically, meeting prescribed timeframes for performing backup and recovery. NAS-specific snapshot and replication technologies offer an alternative to NDMP backup operations; however, they too have tradeoffs.

EMC Avamar NDMP Accelerator Node works with Avamar software running within a centrally managed Avamar data repository (server) to eliminate duplicate backup data on a global scale throughout the organization. A logical full/incremental thereafter backup policy—combined with client-based global data deduplication and compression—provides fast and efficient backup and recovery.

Its N-way grid architecture provides extreme capacity and performance scalability by simply adding nodes (servers) to the grid. And Avamar's redundant array of independent nodes (RAIN) architecture provides high availability by eliminating single points of failure and providing fault tolerance across nodes. Avamar also performs daily integrity checks of the Avamar server and backup data, ensuring that data can be quickly recovered when needed.

Organizations that rely on NAS and are challenged to meet backup and recovery objectives for NAS-based data need only to evaluate EMC Avamar and its NDMP Accelerator technology. Avamar provides end-to-end efficiency for NAS environments, capturing significantly less data at the source, which means lower bandwidth requirements, shorter backup timeframes, and lower capacity requirements. Avamar's single full backup and incremental-thereafter approach ensures the availability of a rapid, single-step recovery. As a result, EMC Avamar offers a more efficient backup and recovery solution for NAS environments, without consequences or compromises.

Enterprise Strategy Group | **Getting to the bigger truth.**

20 Asylum Street | Milford, MA 01757 | Tel:508.482.0188 Fax: 508.482.0218 | www.enterprisestrategygroup.com