Abstract

This white paper provides a technical overview of EMC® Avamar® backup and recovery software with integrated global, source data deduplication technology. It includes an in-depth look at the Avamar architecture, patented global data deduplication technology, key applications, and customer benefits.
Table of Contents

Executive summary ........................................................................................................... 4
Introduction ......................................................................................................................... 4
   Audience ......................................................................................................................... 4
EMC Avamar technology .................................................................................................... 5
   Global/Source deduplication ........................................................................................... 5
      Global elimination of redundant data at the source .......................................................... 5
      Variable vs. fixed-length data segments ........................................................................ 5
      Logical segment determination ..................................................................................... 5
      Not all data deduplication created equal ...................................................................... 6
      Benefits of Avamar’s efficiency ..................................................................................... 7
   General architecture ........................................................................................................ 7
      Avamar servers ................................................................................................................ 7
      Avamar Administrator ..................................................................................................... 8
      Avamar Enterprise Manager .......................................................................................... 8
      Avamar client software .................................................................................................. 8
      Avamar Replicator ......................................................................................................... 8
   Integration with platforms and applications ..................................................................... 9
      UNIX .............................................................................................................................. 9
      Linux .............................................................................................................................. 9
      MAC OS X ...................................................................................................................... 9
      NetWare ......................................................................................................................... 9
      Microsoft SharePoint ..................................................................................................... 9
      Lotus Notes / Domino .................................................................................................... 9
      Open Enterprise Server (OES) Linux for NSS file systems ........................................... 9
      Oracle databases ........................................................................................................... 9
      Microsoft SQL Server databases .................................................................................. 9
      Microsoft Exchange servers ........................................................................................ 10
      DB2 databases ............................................................................................................... 10
      NAS filers ..................................................................................................................... 10
      VMware environments ................................................................................................ 10
   Grid server architecture ................................................................................................ 12
      Global deduplication ..................................................................................................... 12
      Solving the index challenge .......................................................................................... 13
      Scalability and performance: Automated load balancing .............................................. 14
      Reliability and availability: RAIN, Replicator, and checkpoints .................................... 15
      Flexibility in deployment ............................................................................................... 15
      Recoverability verified daily ........................................................................................ 15
      Encryption ..................................................................................................................... 15
      Tape integration ............................................................................................................ 16
      Avamar File System (AvFS) .......................................................................................... 16
      Enterprise reporting capabilities and integration .......................................................... 16
   Integration with enterprise event management solutions .................................................. 17
      Flexible deployment options ........................................................................................ 17
      Secure data shredding .................................................................................................. 17
      Desktop/laptop backup and end-user recovery ............................................................... 18
Conclusion ........................................................................................................................... 19
Executive summary

Protecting critical data is a challenge for organizations of all sizes. According to the Enterprise Strategy Group (ESG), the amount of data requiring protection continues to grow at approximately 60 percent per year. Traditional backup solutions store data repeatedly, expanding total storage under management by five to 10 times. Customers need solutions to help manage the information explosion. In addition, government regulations and requests for legal discovery strain the resources and capabilities of traditional data protection solutions. Failure to comply or provide information in a timely fashion can result in significant costs and penalties. Furthermore, recent legislation has exposed the risk of shipping tapes—either encrypted or unencrypted—as one of the greatest security concerns in today’s IT infrastructure.

Traditional backup solutions require a rotational schedule of full and incremental backups, which move a significant amount of redundant data week after week. Due to the unnecessary data movement, enterprises are often faced with backup windows that roll into production hours, network constraints, and too much storage under management. Tape media, drives, and libraries are all prone to mechanical failure, and tape transport can be extremely unreliable.

Over the last decade, disk storage has been used to augment traditional backup approaches, but disk solutions that are designed to replace tape libraries and media solve only a fraction of the data protection challenges faced by enterprise organizations.

According to industry analyst Gartner Group, there are more than 4 million remote offices in the U.S. alone, at least 60 percent of all enterprise data resides at remote offices, and remote office data continues to grow at more than 50 percent per year. The key challenges of remote office data protection include a lack of trained local staff, limited WAN bandwidth, the high cost of obtaining additional bandwidth (if available), failure-prone equipment, manual processes, lack of centralized management, and high data growth rates. The risk of data loss or exposure from remote sites can be extremely high. As a result, it’s no surprise that remote office data protection is top of mind for most enterprise IT executives.

EMC® Avamar® software solves the challenges associated with traditional backup, enabling fast, reliable backup and recovery across the entire enterprise, including VMware environments, remote offices, LAN/NAS servers, and desktop/laptop systems. Avamar utilizes patented global, source data deduplication technology to identify redundant data segments at the source - before transfer across the network. By moving only new, unique subfile variable length data segments, Avamar reduces the required daily network bandwidth by up to 500x. This allows companies to utilize their existing physical and virtual infrastructure for fast, daily full backups and disaster recovery, despite network congestion and other bottlenecks. Data can be encrypted both in flight and at rest for added security, and centralized management makes protecting hundreds of remote offices easy. By storing just a single instance of each subfile data segment globally, Avamar also reduces total back-end backup storage by up to 50x for cost-effective, long-term, disk-based recovery. While EMC Avamar backs up data to disk, it can also work with existing tape and traditional backup software such as EMC NetWorker®. Avamar can also export deduplicated backup data to tape for cost-effective long-term retention via Avamar Data Transport. Avamar’s intuitive interface allows desktop and laptop users to quickly recover their own data, reducing the burden on IT staff. Finally, Avamar’s grid architecture provides online scalability, and patented redundant array of independent nodes (RAIN) technology provides high availability.

Introduction

This white paper provides a technical overview of EMC Avamar backup and recovery software. It provides a closer look at patented Avamar global, source data deduplication technology, Avamar’s architecture, supported platforms, high availability, and ease of management. Other topics include integrity checking, encryption, and integration with event management solutions.

Audience

This white paper is intended for backup administrators or technical staff seeking a more in-depth look at EMC Avamar’s underlying architecture and technology.
EMC Avamar technology

Global/Source deduplication

Enterprise data is highly redundant, with identical files or data stored both within systems and across systems (for example, OS files or documents sent to multiple recipients). Edited files also have tremendous redundancy with previous versions. Traditional backup software magnifies this redundancy by storing all of this redundant data over and over again. Avamar utilizes patented global, source data deduplication technology to eliminate redundancy at both the file and the subfile data segment level.

Global elimination of redundant data at the source

Avamar solves the challenge of redundancy in backup data at the source—before transfer across the LAN or WAN during a backup operation. Avamar backup agents are deployed on the systems to be protected (for example, servers, desktops, laptops) to identify and filter repeated data segments stored in files within a single system and across multiple systems over time. This ensures that each unique data segment is backed up only once across the enterprise. As a result, copied or edited files, shared applications, embedded attachments, and even daily changing databases generate only a small amount of incremental backup data.

By moving only new, unique subfile variable length data segments, Avamar reduces the required daily network bandwidth and storage by up to 500x. By storing just a single instance of each subfile data segment globally, Avamar also reduces total back-end storage by up to 50x for cost-effective, long-term, disk-based recovery.

Variable vs. fixed-length data segments

A key factor for eliminating redundant data at a segment (or subfile) level is the method for determining segment size. Fixed-block or fixed-length segments are commonly employed by snapshot and some deduplication technologies. Unfortunately, even small changes to a dataset (for example, inserting data into the beginning of a file) can change all fixed-length segments in a dataset, despite the fact that very little of the dataset has actually changed. Avamar uses an intelligent variable length method for determining segment size that looks at the data itself to determine logical boundary points, eliminating the inefficiency.

Logical segment determination

Avamar’s patented method for segment size determination is designed to yield optimal efficiency across all systems in an enterprise. Avamar’s algorithm analyzes the binary structure of a dataset (the 0s and 1s that make up a dataset) in order to determine segment boundaries that are context-dependent, so that Avamar’s client agents will be able to identify the exact same segments for any dataset, no matter where that dataset is stored in the enterprise. Avamar’s variable length segments average 24 KB in size and are then compressed to an average of just 12 KB.

By analyzing the binary structure, Avamar’s method works for all file types and sizes, including databases. For instance, if a paragraph is added to the beginning and the middle of a text file, Avamar’s algorithm will identify and back up only the new, modified segments, dramatically reducing the amount of backup data that needs to be sent and stored.
Figure 1. EMC Avamar software identifies the unique, subfile variable length data segments that comprise the data (in this case, a PowerPoint presentation). Only a single instance of each data segment is stored globally, across sites and servers.

Not all data deduplication created equal

For each 24 KB segment, Avamar generates a unique 20-byte ID, using the SHA-1 encryption algorithm. This unique ID is like a fingerprint for that segment. Avamar’s software then uses the unique ID to determine whether a data segment has been stored before. Files, directories, entire file systems, and even databases can be quickly and efficiently stored with a hierarchical map of these unique IDs.

As shown in Figure 2, EMC Avamar’s patented global, source data deduplication technology significantly reduces the amount of backup data transferred and stored during Avamar’s fast daily full backups.

<table>
<thead>
<tr>
<th>Data Type</th>
<th>Amount of Primary Data Backed Up</th>
<th>Amount of Data Moved Daily</th>
</tr>
</thead>
<tbody>
<tr>
<td>Windows file systems</td>
<td>3,573 GB</td>
<td>6.1 GB</td>
</tr>
<tr>
<td>Mix of Windows, Linux, and UNIX file systems</td>
<td>5,097 GB</td>
<td>11.7 GB</td>
</tr>
<tr>
<td>Engineering files on NAS (NDMP backups)</td>
<td>3,265 GB</td>
<td>24.2 GB</td>
</tr>
<tr>
<td>Mix of 20% databases, 80% file systems (Windows and UNIX)</td>
<td>9,583 GB</td>
<td>80.0 GB</td>
</tr>
<tr>
<td>Mix of Linux file systems and databases</td>
<td>7,831 GB</td>
<td>104.2 GB</td>
</tr>
</tbody>
</table>

Source: EMC

Figure 2. While results will vary by data type and mix, Avamar can dramatically improve backup performance and efficiency.
Benefits of Avamar’s efficiency

Avamar’s superior backup efficiency translates into many important customer benefits, including:

- Reduces required daily network bandwidth and backup storage by up to 500x.
- Provides fast, daily full backups across existing LAN/WAN bandwidth and virtual infrastructure.
- Up to 10x faster daily full backup performance.
- Reduces total back-end disk or tape backup storage by up to 50x.
- Up to 85 percent reduction in total client CPU utilization. Avamar agents run in low priority or “nice” mode, so they do not contend with other applications vying for CPU resources on client systems. While Avamar agents typically use 15 percent more CPU than traditional backup agents during backup operations, Avamar backups complete up to 10x faster. As a result, Avamar reduces total client CPU utilization by up to 85 percent over a seven-day period compared to traditional backup methods.
- Immediate, single-step recovery. Avamar stores all backups as virtual full images, which can be immediately recovered in a single step to any system running the Avamar agent. There is no need to restore from the last good full and subsequent incremental backups to reach the desired recovery point.

General architecture

Avamar’s software consists of a number of components, including the Avamar server, Avamar Administrator, Avamar Enterprise Manager, and Avamar client software. Avamar servers can be deployed in either single-node or scalable multinode configurations, depending upon the amount of performance and disk capacity required. Avamar Replicator efficiently and securely replicates Avamar servers across the WAN between sites.

Avamar servers

EMC offers flexible Avamar server deployment options including Avamar Data Store, the Avamar Virtual Edition virtual appliance, and the option to leverage existing certified servers. The easiest way to deploy a physical Avamar server is via the Avamar Data Store – a scalable, all-in-one packaged solution consisting of Avamar software preinstalled and preconfigured on EMC-certified hardware to simplify purchasing, deployment, and service while minimizing onsite setup. For virtual environments, Avamar Virtual Edition for VMware enables an Avamar server to be quickly deployed as a virtual appliance, leveraging an existing ESX server and its attached disk storage. Another option is to install Avamar software on industry-standard Intel-based servers (certified on Dell, HP, IBM) running Red Hat Enterprise Linux.

Avamar servers store client backups and manage the policies for scheduling, determining datasets, and retention periods. An Avamar backup provides a point-in-time full copy of data that can be restored on demand from an Avamar server. Multinode Avamar servers segregate components of the Avamar server across multiple hardware servers for scalability and performance.

There are two primary node types for a multinode Avamar server:

- **Storage Node**—Stores the deduplicated backup data. Multiple Storage Nodes are configured with multinode Avamar servers based upon performance and capacity requirements. Storage Nodes can be added to an Avamar server over time to expand performance with no downtime required. Avamar clients connect directly with Avamar Storage Nodes; client connections and data are load balanced across Storage Nodes automatically without any downtime.
- **Utility Node**—Node dedicated to scheduling and managing background Avamar server jobs. One utility node is configured per multinode Avamar server. Data on the Utility Node is protected by the Avamar server. Note: Utility Nodes are not single points of access for an Avamar server; backups and restores can still complete to connected clients, even without a Utility Node.

Other optional nodes include:

- **NDMP Accelerator Node**—Specialized node that works with NDMP in order to provide data protection for NAS filers (EMC Celerra® and NetApp filers).
• Access Node—Used to run the Avamar File System (AvFS).

Avamar Administrator
The Avamar Administrator is a graphical management application that includes the following:
• Management of backup policies, including datasets, schedules, and retention
• Management of users and clients
• Centralized on-demand backups and restores
• Detailed monitoring and reporting

The Avamar Administrator can be launched directly from the Web-based Enterprise Manager user interface without any software installation or can be installed locally on any Windows or Linux system.

Avamar Enterprise Manager
Avamar Enterprise Manager is a Web-based management interface that provides the ability to monitor and manage a distributed Avamar deployment. It delivers an “at-a-glance” dashboard that allows backup administrators to quickly assess both the status of backups across the enterprise as well as the status of all Avamar systems in a distributed Avamar deployment. The dashboard also enables users to quickly drill down into a particular Avamar server to review granular system and backup status or to perform detailed server management.

As shown in Figure 3, the Web-based management interface provides an at-a-glance view of the Avamar backup environment — including success and warning indicators.

![Figure 3. Avamar Enterprise Manager](image)

Avamar client software
Avamar supports the automated protection of industry-leading operating systems and applications. Avamar clients filter out redundant data before sending backup data over networks, making it possible to protect systems even over congested LANs or WANs.

Avamar Replicator
Avamar Replicator enables efficient, encrypted, and asynchronous replication of data stored in an Avamar server to another Avamar server deployed in a remote location without the need to ship tapes. Avamar Replicator is a scheduled process between two independent Avamar servers, providing a higher level of
reliability for stored backups. Avamar Replicator can be scheduled to run at off-peak hours to minimize bandwidth impact.

**Integration with platforms and applications**

Avamar client software supports the following industry-leading platforms and applications.

Avamar supports automated protection of all FAT or NTFS data, and also provides the ability to back up and restore Windows system state. The Avamar client for Windows is often used to protect remote desktops and laptops over the WAN.

**UNIX**
Avamar supports a variety of UNIX operating systems, including Solaris, HP-UX, SCO, Free BSD, and IBM AIX.

**Linux**
Avamar supports Red Hat and SUSE Linux operating systems.

**MAC OS X**
Avamar supports Mac OS X on the PowerPC and Intel platforms.

**NetWare**
Avamar supports Novell NetWare on Intel platforms.

**Microsoft SharePoint**
Avamar supports Microsoft SharePoint environments.

**Lotus Notes / Domino**
Avamar supports Lotus Notes / Domino environments.

**Open Enterprise Server (OES) Linux for NSS file systems**
Avamar supports Novell OES 2 on Intel platforms.

**Oracle databases**
Avamar utilizes Oracle Recovery Manager to provide fast and automated protection of Oracle databases, while maintaining online availability. The Avamar client for Oracle can be used to provide full daily backups of an active Oracle database while generating only a small amount of incremental storage.

**Microsoft SQL Server databases**
The Avamar client for Microsoft SQL Server enables administrators to perform centralized, hot backups of SQL Server for an entire server or for individual databases.
Microsoft Exchange servers
The Avamar client for Microsoft Exchange enables administrators to perform centralized online backups of Exchange at a server, mailbox, or message level.

DB2 databases
The Avamar client for DB2 provides fast and automated protection of DB2 databases, while maintaining the online availability of the databases.

NAS filers
Avamar supports NDMP backups via the Avamar NDMP Accelerator Node to provide reliable and high-performance backup and recovery for NAS filers (for example, EMC Celerra and NetApp filers). Avamar provides fast, daily full backups for filers while requesting only a level-one (incremental) dump of data from the filer itself, dramatically reducing backup times and network utilization (see Figure 4).

Fast NAS NDMP Backups with Fewer Resources Consumed

- Source and global deduplication at the Avamar NDMP accelerator node
- Only modified segments moved from NAS device
- Centralized management of NAS backups with simple one-step recovery
- Benefits
  - Backup times significantly reduced
  - Reduced demand on network and filer resources
  - Reduce tape infrastructure, media, and handling costs
  - Increase NAS utilization rates, lower TCO

Figure 4. Avamar delivers fast, daily full backups for NAS filers

VMware environments
Avamar software quickly and efficiently protects VMware environments by reducing the size of backup data within and across virtual machines. This eliminates traditional backup bottlenecks caused by the large amount of data that must pass through the same set of shared resources—the physical server’s CPU, NIC, memory, and disk storage. Avamar reduces the traditional backup load—up to 200 percent weekly—to as little as 2 percent over the same seven-day period, dramatically reducing backup times and resource utilization.

Avamar can quickly protect VMware environments by installing agents on the virtual machine Guests, proxy server. In all cases, Avamar’s powerful deduplication technology provides fast, daily full backups while reducing required network/infrastructure bandwidth and storage (Figure 5). In addition, unlike traditional backup solutions, Avamar can deduplicate the data stored in virtual disks (*.vmdk files), significantly reducing storage utilization and enabling replication of virtual disks across congested WANs.
The EMC Avamar Virtual Edition for VMware is the industry’s first deduplication virtual appliance for backup, recovery, and disaster recovery. Avamar Virtual Edition enables users to deploy Avamar’s deduplication technology easily, effectively, and in a repeatable fashion on VMware ESX Server hosts. Each virtual appliance supports up to 2 TB of deduplicated backup capacity (which under a typical traditional backup schedule would require approximately 70 TB of tape or disk storage), and can leverage the existing VMware shared server and storage infrastructure to lower costs and simplify IT management. Avamar Virtual Edition supports VMotion for deployment flexibility, and up to two Avamar Virtual Edition virtual appliances per ESX server provides scalability (Figure 6). Replication between Avamar virtual appliances or from Avamar virtual appliances to physical Avamar servers (for example, Avamar Data Store) eliminates reliance on offsite tape shipments and the risk of losing unencrypted data.

Avamar Efficiently Protects Virtual Machines

- Up to 95% reduction in data moved
- Up to 90% reduction in backup times
- Up to 50% reduction in disk impact
- Up to 95% reduction in NIC usage
- Up to 80% reduction in CPU usage
- Up to 50% reduction in memory usage
- All backups stored as “virtual full backups,” ready for immediate restore
- Maintain effective consolidation ratios without over-taxing CPU utilization

Figure 5. Avamar deduplicates backup data within and across virtual machines, at the source, and globally, providing fast and reliable daily full backups
Avamar Server Software Deployed as a Virtual Appliance

Figure 6. Avamar Virtual Edition for VMware is the industry's first deduplication virtual appliance for backup, recovery, and disaster recovery

**Grid server architecture**
Avamar software solutions are deployed on Intel-based servers running Red Hat Enterprise Linux in a grid architecture. This grid server architecture provides a number of significant benefits including reliability, availability, scalability, performance and flexibility in deployment.

**Global deduplication**
Avamar stores only a single instance of each unique subfile variable length data segment for all protected servers, desktops, and laptops. Each segment of data is assigned a unique ID, using the SHA-1 encryption algorithm. This 20-byte ID is unique to the data segment. Whenever an Avamar client encounters a new data segment, it generates the unique ID and sends the ID to the Avamar server to determine if the segment has been previously stored. The segment will only be sent to an Avamar server if it is new and unique.

Avamar’s grid server architecture, shown in Figure 7, provides scalable performance and capacity. Every Avamar client can connect to every Storage Node for both backup and restore, which eliminates potential performance bottlenecks.
Solving the index challenge

As data volumes increase, a centralized index becomes increasingly complex and difficult to manage, often introducing a bottleneck to backup operations. In addition, corruption of the centralized index may result in the inability for an organization to recover data, since the index can no longer be used to identify which tapes contain the particular set of data that must be recovered. Avamar uses an elegant, distributed indexing architecture to eliminate the indexing challenge. Avamar uses segment IDs in a manner similar to a phone number for land lines. In a phone number, the area code provides the first general area where a call needs to be routed, and the number itself determines the exact location where the call is targeted. Avamar uses a portion of each unique ID (like an area code) to determine which Storage Node will store a specific segment of data. Avamar uses another portion of the unique ID (like a phone number) to determine where that segment of data will be stored inside that Storage Node. As a result, just by looking at the unique ID of a segment, Avamar can determine exactly where to store or retrieve that segment—even across a large number of Storage Nodes.

Avamar utilizes an elegant, distributed indexing architecture that streamlines access to data (Figure 8). Automatic load balancing distributes data across all available Storage Nodes and enables linear performance increases by simply adding nodes when needed.
Furthermore, Avamar takes groups of unique IDs (for instance, all the IDs for a set of segments that make up a file) and generates a new unique ID for that group. A request for that group ID will cascade into requests for all the segments that make up that group. This process continues hierarchically, so Avamar can quickly store and retrieve files, directories, entire file systems, and even databases, without the need for any centralized index or database that can become a bottleneck to performance or scalability. It is important to note that the Avamar Utility Node is not used as a database or index for storage or access of unique segments. Unlike some other deduplication architectures, Avamar does not use access nodes or metadata nodes, which can become bottlenecks to performance. Every Avamar client can connect to every Storage Node in an Avamar Grid Server for both backup and restore. Avamar’s elegant index structure eliminates redundancy even for indices and metadata, ensuring that the indexing component of an Avamar Grid Server remains approximately two percent of total data storage.

**Figure 8. Avamar indexing architecture**

![Avamar Indexing Architecture Diagram](image)

Furthermore, Avamar uses the SHA-1 encryption algorithm to determine unique IDs. The SHA-1 encryption algorithm provides a flat distribution of unique IDs across the full potential range of outputs. As a result, Avamar automatically load balances data across all available Storage Nodes for optimized scalability and performance for backup and restore.

New Storage Nodes can be added hot to an Avamar server with no need for scheduled downtime, allowing an Avamar server to accommodate data growth or increased retention periods. Since Avamar can be easily scaled by the addition of Storage Nodes, customers can purchase only the capacity they need (just-in-time purchasing) and add Storage Nodes as they grow, reducing IT costs. In addition, when new Storage Nodes are added to an existing Avamar server, data is actively load balanced across the newly added nodes. Many disk-based data protection solutions available today are deployed as a “single, large server” attached to a large pool of external disk storage. As the amount of primary data that must be protected grows, these single, large servers quickly become bottlenecks to performance and scalability. Avamar’s load balancing enables linear performance increases by simply adding Storage Nodes. Each incremental node increases CPU, memory, I/O and disk capacity for the entire grid. This scalable performance is particularly critical in...
disaster recovery situations when multiple servers must be restored simultaneously to meet recovery-time objectives.

**Reliability and availability: RAIN, Replicator, and checkpoints**

When traditional backup solutions fail, enterprises are exposed to windows of potential data loss. Avamar employs patented redundant array of independent nodes (RAIN) technology (U.S. Patent: 6,826,711), in order to provide failover and fault tolerance across the nodes in an Avamar server grid. Avamar can continue to provide reliable data protection and access, even if a server node fails or becomes unavailable, since data stored on any node can be reconstructed from the other nodes.

In addition to protecting enterprise systems, Avamar also protects itself with twice daily, internal checkpoints — consistent snapshots of the entire Avamar system that can be verified for integrity. If an integrity check fails due to an inconsistency, and the inconsistency cannot be fixed, the system can be quickly rolled back to a prior checkpoint. And RAID protects data stored on disk (Figure 9).

Avamar Replicator provides the ability to logically replicate the data stored within an Avamar system to another Avamar system deployed in an off-site location. In the event of a disaster scenario where a complete Avamar system becomes unavailable, data can be recovered directly from the replication target, providing a high level of availability.

**Flexibility in deployment**

Enterprises have a tremendous amount of flexibility in planning long-term deployments of an Avamar server. Avamar’s grid architecture enables an Avamar system to be expanded one node at a time—even across dissimilar hardware platforms from different manufacturers or with different capacity disk drives. This allows customers to take advantage of new, more cost-effective hardware as their Avamar systems grow over time.

**Figure 9. RAIN architecture**

**Recoverability verified daily**

Avamar runs daily, scheduled integrity checks of all data stores. The integrity checking application uses a 160-bit checksum to verify that all data backups can be restored to their original state, an exacting process that would be very difficult and expensive to accomplish with removable-media solutions. As a result, Avamar eliminates the risk of “silent” data loss or corruption, which is common with tape archives. If integrity is compromised in an Avamar server, Avamar can recover the corrupted data using RAIN, internal checkpoints, or a replicated Avamar server.

**Encryption**

Avamar provides comprehensive encryption capabilities, including the ability to encrypt backup data while in transit and at rest. For enhanced security during client/server data transfers, Avamar supports SSL encryption. SSL encryption utilizes the 128-bit or 256-bit Advanced Encryption Standard (AES) algorithm and should be used for any external network communications, where security is a significant concern. The choice of encryption method can be made on a client-by-client basis or for an entire group of clients.

Avamar also supports the option to enable encryption of data at rest using 128-bit Blowfish encryption. By encrypting data at rest, organizations are further protected from backup data theft or unauthorized access.
**Tape integration**

There are several options available when export to tape is required. Avamar Data Transport enables deduplicated data to be exported to tape for cost-effective, long-term retention. It utilizes an intuitive interface with policy-driven processes for export and recovery of deduplicated Avamar backup data to and from tape. A searchable file-level catalog eases and speeds recovery, while leveraging existing backup software (for example, EMC NetWorker). Avamar Data Transport significantly reduces the required number of tapes by up to 50 times (Figure 10). Avamar is also tightly integrated with EMC NetWorker, and utilizes the existing NetWorker management interface that includes the ability to backup to tape.

![Diagram](image)

**Figure 10.** Avamar Data Transport exports deduplicated backup data to tape for cost-effective, long-term retention

**Avamar File System (AvFS)**

Avamar provides a Windows and Linux file system interface to all backups stored in an Avamar server. As a result, administrators can utilize the file system view of backups to perform direct restores of individual data files or volumes without the need to install Avamar clients on target systems.

**Enterprise reporting capabilities and integration**

Avamar provides a number of standardized reports. For organizations that prefer customized reporting capabilities, Avamar integrates with standard enterprise reporting solutions such as Crystal Reports or Actuate, or with specialized backup reporting applications such as EMC Data Protection Advisor. To facilitate integration of Avamar with external reporting tools, Avamar uses a PostgreSQL database to store information about backup activities (for example, successful and failed backups) as well as backup policies. Information in the database is accessible through any PostgreSQL-compliant Open Database Connectivity (ODBC) interface. Avamar also publishes comprehensive information about all available database views that customers can leverage to create custom reports.

---

Efficient Data Protection with EMC Avamar Global Deduplication Software

Applied Technology  16
Integration with enterprise event management solutions

Avamar integrates with popular enterprise management solutions that can pull SNMP information from an Avamar SNMP server, or accept SNMP traps from the Avamar server. Avamar system activities and operational status are events that can be integrated into enterprise management solutions. Examples of various Avamar events include client registration and activation, successful and failed backups, hard disk status, and so on. Avamar provides the ability to flexibly configure events so they can be monitored via:

- **E-mail**—Events can be configured on an event-by-event basis to send an e-mail message to a designated list of recipients. E-mail notifications can be sent immediately or in batches at regularly scheduled times.
- **Syslog**—Events can be configured on an event-by-event basis to log information to local or remote syslog files based on filtering rules configured for the syslog daemon receiving the events. Third-party monitoring tools and utilities capable of examining log entries can access the syslog files and process them in order to integrate Avamar events into larger site activity and status reports.
- **SNMP**—Avamar supports two SNMP methods to access events and activity completion status. Avamar provides a mechanism for SNMP management applications to “pull” information from Avamar’s SNMP server. In addition, SNMP traps provide a mechanism for the Avamar server to “push” information to SNMP management applications whenever designated Avamar events occur.

Flexible deployment options

**Agent-only option** - For smaller remote offices, EMC Avamar software agents can be installed directly on the systems to be protected, without the need for additional local hardware. This enables data to be backed up directly over existing WAN connections to a central EMC Avamar server at the data center.

To protect larger remote offices and data centers, data can be backed up to a local Avamar server for faster recovery, and then replicated to another Avamar server located at the data center or remote disaster recovery site. When an Avamar server is required, there are several deployment options available:

**EMC-certified server** - Avamar software installed on an EMC-certified server running Red Hat Enterprise Linux from vendors including Dell, HP, and IBM.

**EMC Avamar Virtual Edition for VMware** - The industry’s first deduplication virtual appliance for backup, recovery, and disaster recovery. Enables Avamar to be deployed easily, effectively, and in a repeatable fashion on VMware ESX Server hosts, leveraging the existing server CPU and disk storage.

**EMC Avamar Data Store** – An all-in-one packaged solution consisting of EMC Avamar software running on preconfigured EMC-certified hardware, the Avamar Data Store is available in two models – a scalable multi-node model and a single-node model. This approach simplifies purchasing, deployment, and service while minimizing on-site setup.

The multi-node Avamar Data Store is designed for the data center where backup data is being consolidated from multiple remote locations or to protect VMware environments and LAN/NAS servers. It can efficiently retain the equivalent of up to several petabytes of traditional cumulative daily full backups.

The single-node Avamar Data Store is ideal for deployment at remote offices that require faster local recovery performance. It provides up to 1 TB, 2 TB, or 3.3 TB of deduplicated backup capacity, which under a typical traditional backup schedule could require tens of terabytes of disk or tape storage, depending on the backup method and retention period. In addition, both models support replication, either from the remote office to the data center for consolidation, or between data centers for disaster recovery.

Secure data shredding

Data shredding support helps facilitate the secure deletion of data from an Avamar server when classified data is accidentally backed up to a non-classified server. Typically, this scenario is applicable to
Department of Defense installations, but some enterprise companies may wish to perform the same actions. Avamar’s data shredding implementation is in accordance with the DoD 5220.22-M standard.

**Desktop/laptop backup and end-user recovery**
Avamar’s fast, daily full backups and integrated global, source data deduplication are also available for desktop and laptop systems. Avamar’s intuitive interface allows desktop and laptop users to quickly recover their own data, reducing the burden on IT staff (Figure 11). End users can restore files and folders directly to the original location with no loss of ACLs.

---

Figure 11. Avamar’s intuitive, easy-to-use interface enables desktop and laptop users to quickly recover their own data, reducing the burden on IT staff.
Conclusion

EMC Avamar software solves the challenges associated with traditional backup, enabling fast, reliable backup and recovery for VMware environments, remote offices, LAN/NAS servers, and desktop/laptop systems. Avamar utilizes patented global, source data deduplication technology to identify redundant data segments at the source, reducing daily backup data by up to 500x before it is transferred across the network and stored to disk. This allows companies to utilize existing physical and virtual infrastructure for fast, reliable backup, recovery, and disaster recovery. Data can be encrypted for added security, and centralized management makes protecting hundreds of remote sites easy and efficient. By storing just a single instance of each subfile variable length data segment globally, Avamar also reduces total back-end storage by up to 50x for cost-effective, long-term, disk-based recovery. Avamar can also work with existing tape and traditional backup software environments. Avamar Data Transport enables deduplicated backup data to be exported to tape for cost-effective, long-term retention. Avamar’s intuitive interface allows desktop and laptop users to quickly recover their own data, reducing the burden on IT staff. And Avamar is also tightly integrated with EMC NetWorker, utilizing the existing NetWorker management interface and policies.

EMC Avamar’s open software solutions can be quickly deployed via a variety of flexible deployment options. Built on a grid architecture, Avamar provides enterprises with scalability, just-in-time capacity provisioning, and high performance. High availability is delivered via patented RAIN technology to eliminate single points of failure and provide fault tolerance. With Avamar, enterprises can protect more data, dramatically reduce media and network costs, and enjoy fast single-step recovery.

Efficient Data Protection with EMC Avamar Global Deduplication Software

Applied Technology