

Lab Validation Report

EMC Avamar Revolutionizing Backup and Recovery

By Brian Garrett
With Heidi Biggar, Claude Bouffard and Lauren Whitehouse

May, 2008

Table of Contents

Table of Contents	i
Introduction.....	1
<i>Background</i>	<i>1</i>
<i>EMC Avamar.....</i>	<i>2</i>
ESG Lab Validation.....	3
<i>Protecting Remote Offices</i>	<i>3</i>
<i>Source-based Global Data De-duplication.....</i>	<i>5</i>
<i>Solving VMware Backup Challenges.....</i>	<i>8</i>
<i>NetWorker and Avamar Integration.....</i>	<i>13</i>
<i>Solving Data Center Challenges</i>	<i>15</i>
<i>Scalability.....</i>	<i>15</i>
<i>High Availability, Reliability and Security</i>	<i>15</i>
ESG Lab Validation Highlights.....	16
Issues to Consider.....	16
ESG Lab’s View.....	17
Appendix.....	18

ESG Lab Reports

The goal of ESG Lab reports is to educate IT professionals about emerging technologies and products in the storage, data management and information security industries. ESG Lab reports are not meant to replace the evaluation process that should be conducted before making purchasing decisions, but rather to provide insight into these emerging technologies. Our objective is to go over some of the more valuable feature/functions of products, show how they can be used to solve real customer problems and identify any areas needing improvement. ESG Lab’s expert third-party perspective is based on our own hands-on testing as well as on interviews with customers who use these products in production environments. This ESG Lab report was sponsored by EMC.

All trademark names are property of their respective companies. Information contained in this publication has been obtained by sources The Enterprise Strategy Group (ESG) considers to be reliable but is not warranted by ESG. This publication may contain opinions of ESG, which are subject to change from time to time. This publication is copyrighted by The Enterprise Strategy Group, Inc. Any reproduction or redistribution of this publication, in whole or in part, whether in hard-copy format, electronically, or otherwise to persons not authorized to receive it, without the express consent of the Enterprise Strategy Group, Inc., is in violation of U.S. Copyright law and will be subject to an action for civil damages and, if applicable, criminal prosecution. Should you have any questions, please contact ESG Client Relations at (508) 482.0188.

Introduction

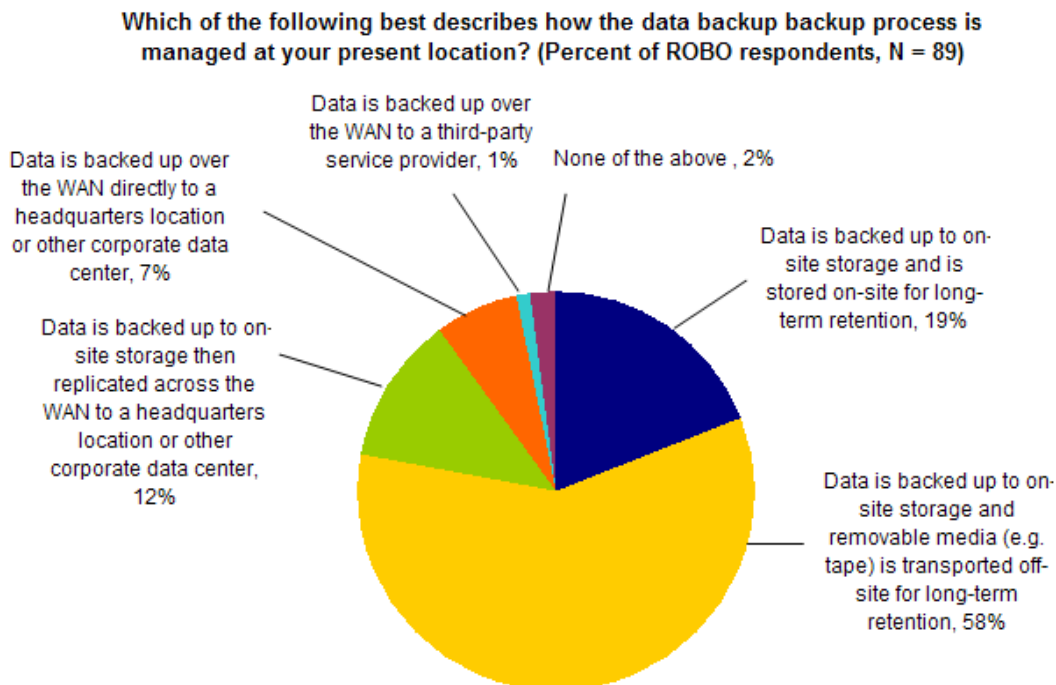
Rapid adoption of virtual server technology and the challenges associated with the backup and recovery of ever-growing stores of information located within branch and remote offices is causing a number of IT managers to reevaluate their data protection strategies. New backup and recovery methods which use data de-duplication technology to reduce capacity and network bandwidth requirements are being deployed to keep up with explosive data growth, shrinking backup windows, compliance initiatives and security concerns.

This ESG Lab report presents the results of hands-on testing of EMC Avamar backup and recovery software with integrated global, source data de-duplication technology. Avamar identifies redundant sub-file data segments at the source (client), reducing the size of backup data before it is sent over the network and stored on disk. Areas of focus include fast and secure backup and recovery of remote office data, resource efficient backup and recovery of VMware virtual server environments and integration with EMC NetWorker. Features designed to solve data center problems—including system scalability, high availability, reliability and secure encryption—are also explored.

Background

Data residing within remote and branch offices (ROBOs) outside the data center accounts for a significant portion of enterprise information stores, yet this data is often protected by inefficient backup processes or not protected at all, increasing the risk of data loss. As shown in Figure 1, ESG research indicates that 89% of organizations currently back up ROBO data using on-site storage and personnel at a remote or branch office.¹ The volume of data and the high cost of WAN bandwidth associated with these legacy methods lead most organizations (58%) to transfer copies of ROBO backup data to a remote facility for disaster recovery using tapes and trucks. Non-technical staff is often left with the responsibility of troubleshooting backup failures and handling tapes.

FIGURE 1. LEGACY REMOTE AND BRANCH OFFICE DATA PROTECTION STRATEGIES



¹ Source: ESG Research Report, *Branch Office Optimization*, January 2007

In addition to the challenges associated with remote and branch office data, a growing number of IT managers are searching for ways to better protect virtual server data (e.g. VMware environments), since traditional backup and recovery methods can cause backup performance bottlenecks and limit server virtualization levels. Legacy backup methods often lead to missed backup windows, application performance issues and an inability to meet service level agreements. As a result, ESG research indicates that the deployment of server virtualization is causing a growing number of IT managers to rethink their existing backup and recovery methods. Twenty percent of respondents indicate that they've already adopted new backup tools and processes to protect virtual server data.²

EMC Avamar

EMC Avamar is backup and recovery software with integrated global, source data de-duplication technology. Avamar significantly reduces the cost and complexity associated with the backup and recovery of ROBO, virtual server environments, and datacenter LAN servers. Lightweight Avamar agents are installed on application servers to eliminate redundant backup data before it is sent over a local or wide area network and stored on disk. These agents are supported for a number of popular operating systems and applications including AIX, HP-UX, Linux, Mac OS, Solaris, Windows, VMware, NetWare, Microsoft Exchange, Microsoft SQL Server, IBM DB2, Oracle and NDMP (for NetApp filers and EMC Celerra). The Avamar agents work with Avamar software running within a centrally managed Avamar data repository (server) to eliminate duplicate data on a global scale throughout the organization. A logical full/incremental forever backup policy—combined with source-based global data de-duplication, compression and encryption—provides fast and efficient backup and recovery.

Avamar offers deployment flexibility to meet the needs of enterprise organizations. For smaller remote offices, only Avamar agents are required (no extra hardware), while larger offices or those requiring faster recovery can deploy a local Avamar server. EMC offers several methods to deploy an Avamar server, including the Avamar Data Store, which provides a turnkey pre-configured solution consisting of Avamar software running on EMC hardware. An Avamar Data Store single-node configuration is available in 1 TB or 2TB capacities to store de-duplicated data (which, under a typical backup schedule, could require tens of terabytes of traditional cumulative disk or tape storage, depending on the backup method and retention period). For larger implementations, a scalable Avamar Data Store multi-node configuration with self-healing redundancy can be used to retain the equivalent of petabytes of daily full backups. Avamar can also be deployed on industry standard EMC certified servers. And for VMware environments, Avamar Virtual Edition enables a complete Avamar server to be deployed as a virtual appliance by running as a Guest on an existing ESX server and leveraging existing disk storage.

Avamar was designed to solve a number of data center problems. 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.

ESG Lab first evaluated Avamar in 2005 and again in 2007 shortly after its acquisition by EMC. A number of compelling new capabilities have been added to the solution since the first ESG Lab validation. The ability to quickly and efficiently backup VMware environments via Avamar agents at the Guest, Service Console, or VMware Consolidated Backup (VCB) proxy server level provides a resource-efficient method for the backup and recovery of virtual servers. Also, integration with EMC NetWorker provides a path for NetWorker customers who want to leverage the benefits of data de-duplication with the familiar look and feel of the NetWorker user interface. The balance of this report presents the results of ESG Lab hands-on testing of EMC Avamar.

² Source: ESG Research Report, *The Impact of Server Virtualization on Storage*, December 2007, N=365

ESG Lab Validation

ESG Lab first performed hands-on testing of EMC Avamar in Irvine, California in 2005. This first round of testing was focused on the use of source-based, global data de-duplication to protect remote and branch offices.

Protecting Remote Offices

Remote and branch offices (ROBOs) contain a large amount of enterprise data, yet most still rely upon traditional tape-based backup methods manned by non-IT staff. Tape backup, snapshots, mirroring, staging and archiving all significantly increase the amount of storage under management. Non-IT staff handling backups and tapes increases the cost of data protection and the risk of data loss. In a distributed environment, it is impractical to centralize backup operations with traditional backup architectures and existing network capacity. Moving even daily incremental backup data sets requires so much network bandwidth and time that even this simple process can become prohibitively expensive and inefficient.

Avamar attacks this issue by improving backup processes in several ways. It employs capacity reduction technology and global data de-duplication across sites and servers with a lightweight agent running on the client systems it protects. Avamar requires a one-time-only full backup where redundant data is eliminated and after that, only new, unique sub-file data segments are backed up daily across the WAN to the data center or DR site. The software identifies and filters repeated variable length data segments stored in files within a single system and across multiple systems over time, capturing only net new segments. In addition, data compression algorithms are applied to eliminate space and redundant file patterns, further shrinking data volumes.

FIGURE 2. PROTECTING REMOTE AND BRANCH OFFICES

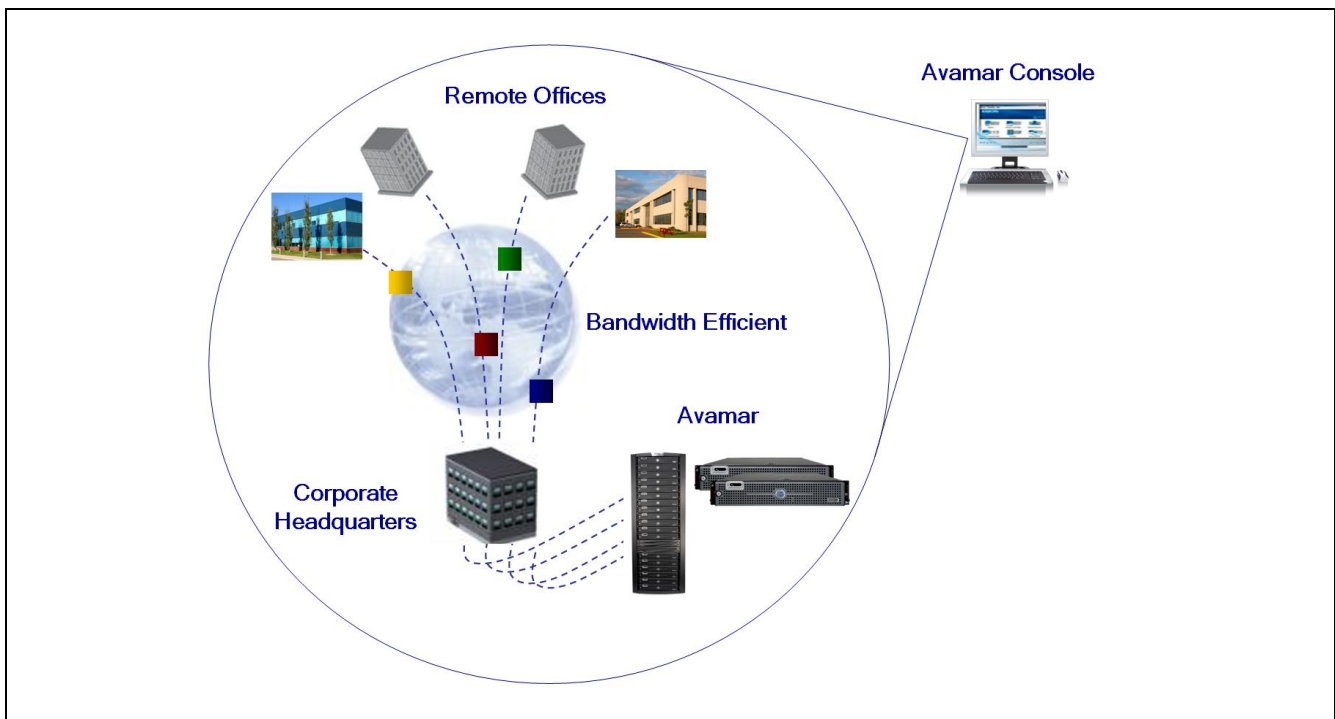


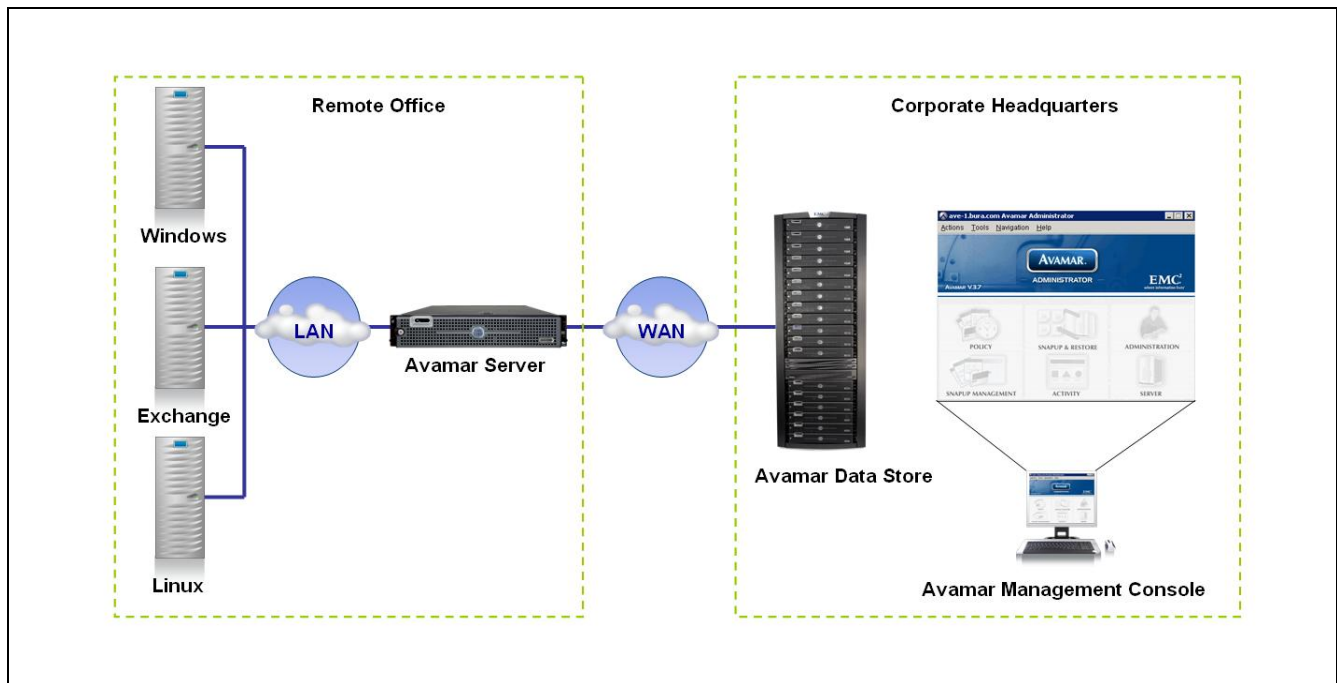
Figure 2 illustrates how remote and branch offices using EMC Avamar can reduce the amount of data being transferred during backup (labeled “Bandwidth Efficient”) by sending only unique blocks of data to a centralized data center. At smaller remote offices, only Avamar software agents are deployed on the systems, while larger remote offices and data centers typically deploy a local Avamar server to improve recovery performance.

Avamar also supports disaster recovery and offsite archival by replicating backup data over a wide area network to a remote data center. Data de-duplication and replication capabilities ensure that only changes since the last backup, comprised of unique and compressed sub-file variable length data segments, are replicated. This is extremely valuable, reducing the cost of network bandwidth and improving the performance of backups and replication over the WAN.

ESG Lab Testing

A single-node Avamar system was configured to test remote office backup to a corporate data site (see Figure 3). An internal Ethernet network simulated a wide area network during this phase of testing. In this test, ESG Lab used Avamar to protect a Microsoft Windows, Exchange and Linux server at a simulated remote office.

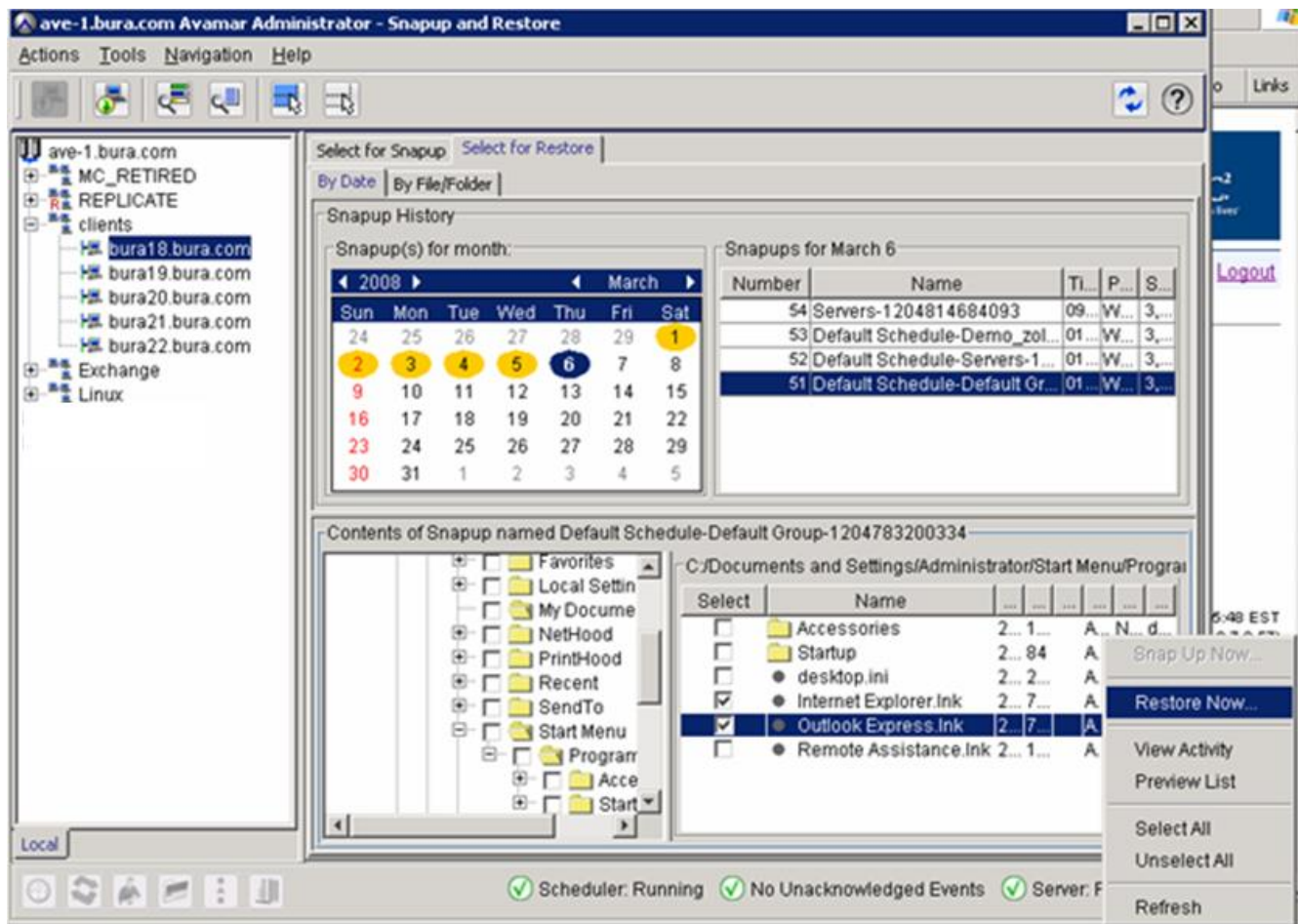
FIGURE 3. REMOTE BRANCH OFFICE TEST BED



The installation of the remote system with approximately 1 TB of de-duplicated backup capacity began with an Avamar-specific Linux distribution and was completed in about 30 minutes. Within an hour, ESG Lab had configured a replication policy to run at 6:00 PM nightly. The Avamar Management console was used to select the servers and data for an on-demand backup. A remote restore was tested after the nightly replication job has run.

ESG Lab found that managing backups and restores using the Avamar management console is extremely intuitive. As shown in Figure 4, the backup history for a server appears in an intuitive monthly calendar format. After choosing the backup to recover from (in this case, the most recent backup on March 6), a directory tree interface that looks and feels like Windows Explorer was used to select the files to restore. During ESG Lab remote recovery testing, a single log file (NTUSER.DAT) was recovered at the simulated remote office from a replica at corporate headquarters in less than three minutes.

FIGURE 4. THE AVAMAR MANGEMENT CONSOLE INTERFACE



Source-based Global Data De-duplication

During backups, the Avamar client agent examines the user's file system and applies a data de-duplication algorithm that identifies redundant data sequences and breaks that file system into variable-sized sub-file objects. Each object is assigned a unique ID. The Avamar client software then determines whether or not this unique ID has already been stored on the Avamar server.

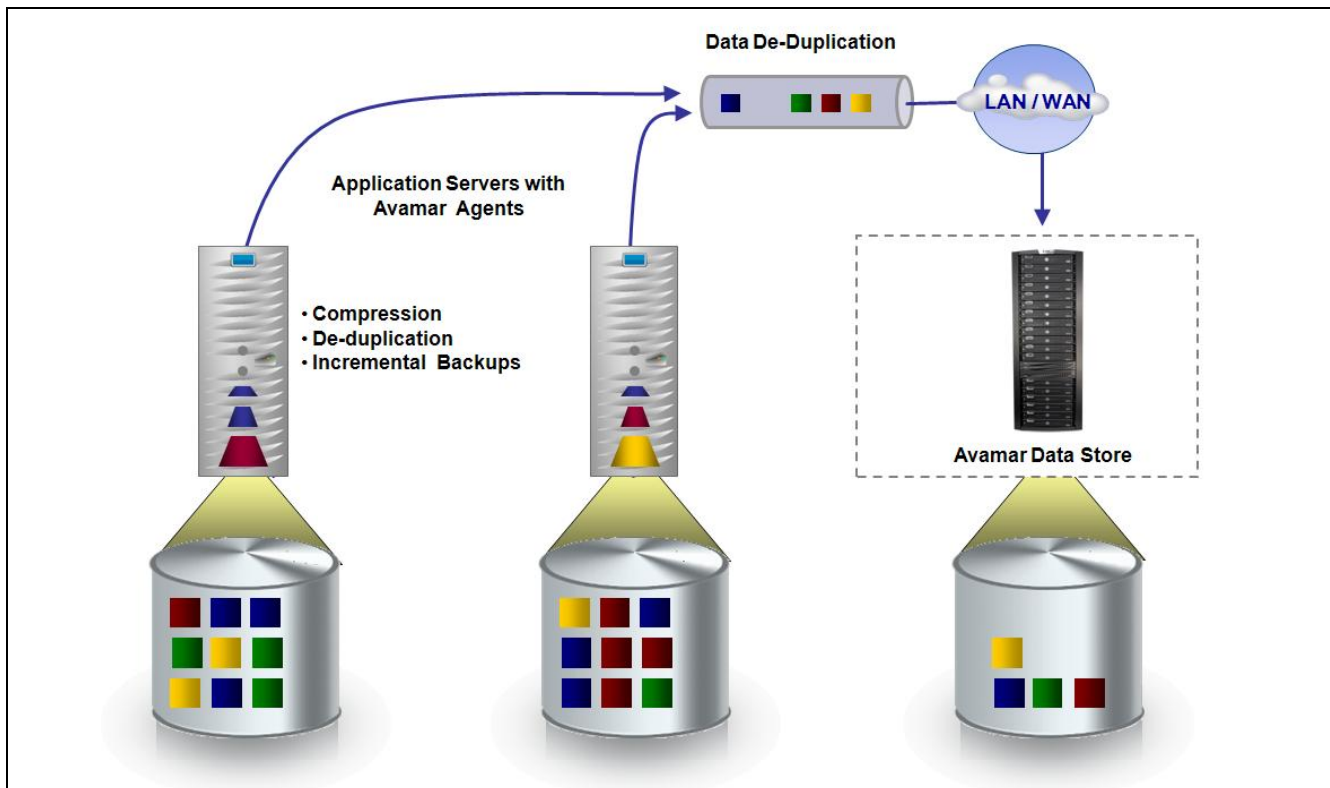
If the object does reside on the Avamar server, a link to the stored object is referenced in this backup. Once an object has been stored on the server, it never has to be resent over the network, no matter how many times it is encountered throughout the organization. This way, Avamar customers only need to perform an initial, full backup once as subsequent daily backups consist of only new, unique sub-file (block-level) data segments that are continuously applied—forever. This feature greatly reduces network traffic and provides enhanced storage efficiency during backups and enables fast, one-step recovery of daily full backups.

To better understand Avamar's data de-duplication technology, consider a PowerPoint presentation attached to an e-mail. If the e-mail is sent to multiple recipients—including remote office locations and then forwarded to yet another set of recipients at a different remote office—Avamar's data de-duplication technology will still store the presentation only once, which is an example of data de-duplication at work at the file level. If one of the recipients modifies a slide in the presentation and again forwards it to a group of colleagues, Avamar will only transmit the unique blocks of the PowerPoint file that has changed—which is sub-file de-duplication. If each remote office backs up to a centralized corporate data center, Avamar ensures that each block of data is unique globally or it won't be transferred. This is referred to as global data de-duplication—as it provides the greatest value of capacity saving and bandwidth optimization. With global data de-duplication, the actual capacity required to

protect and store user data is significantly less than the virtual capacity it is protecting.

As shown in Figure 5, Avamar software filters primary data at the source—before it is backed up across the network—to ensure that only unique sub-file data segments (represented by the colored boxes) are transferred and stored to disk. The unique data blocks are compressed (represented by the compacted color boxes) and then backed up over an Ethernet LAN or WAN to an Avamar Data Store appliance, an industry-standard server running Avamar software. This dual-process minimizes the amount of data ultimately backed up over the LAN to a back-end data repository: There are fewer colored boxes on the right side of the diagram than on the left, indicating the capacity savings with data de-duplication.

FIGURE 5. AVAMAR SOURCE-BASED GLOBAL DATA DE-DUPLICATION



This technique also eliminates the issues and delays associated with using one or more incremental backup images during a restore because Avamar provides the equivalent of daily full backups that can be immediately recovered in a single step. Avamar's space efficient logically complete backup method also eliminate the need to periodically perform full backups, which makes backups run much faster because they read and move significantly less data than a full backup.

ESG Lab Testing

ESG Lab performed hands-on tests and audited historical production backup data to validate capacity and bandwidth efficiency of the EMC Avamar solution. According to EMC, an initial backup of Microsoft Office file data stored to disk is typically reduced by 70% in real-world deployments. Even more impressive is what happens when backups are retained over time: As new daily backups occur, the amount of common data across backup data sets increases, ultimately reducing the required network bandwidth and backup storage required by up to 500x compared to traditional daily full backup methods.

Having witnessed the benefits of data reduction over the course of two backups, ESG Lab audited a series of Avamar backup logs to determine the cumulative data reduction benefits that can be achieved over longer periods of time based on real-life application data. Avamar software is used internally by the Avamar team at

EMC to protect all of its data. The data stored on servers used by the engineering, finance and operations groups, as well as on desktops and even laptops, is protected using Avamar software.

Backup logs were audited for a variety of server and data types. The statistics presented in Avamar backup logs (Data Protected, Common, Changed) were used to calculate the data reduction factors (see Table 1).

TABLE 1. AVAMAR DATA REDUCTION BENEFITS		
Data Type	Changed Data Daily	Data Reduction Factor
Laptop	0.28%	363x
Engineering Server	0.2%	611x
Exchange Server	0.5% to 4%	25 to 200x

Armed with hands-on experience and an audit of the Avamar backup logs, ESG Lab interviewed an Avamar customer to validate the data reduction benefits of Avamar software over time. The customer worked for a construction firm with about 12 TB of Oracle, e-mail and file data. Before the introduction of the Avamar software, Veritas NetBackup and tape libraries had been used. Weekly, full backups were beginning to take all weekend long and these backups frequently failed, meaning that portions had to be re-run on Monday. Restore requests coming in over the weekend or on Monday often had to be deferred until Tuesday. Handling over 60 TB of on- and off-site DLT and LTO tape cartridges was becoming untenable—a fact that wasn't fully realized until tape was phased out after the installation of Avamar software.

Avamar software was installed and initially deployed to protect project data residing on a file server. All of the company's data is now backed up using Avamar. A month's worth of backups are currently kept on disk for quick and reliable restores and the user interface made a great first impression. As a matter of fact, ESG Lab was told that "getting it going was quite a bit of fun"—not something one normally hears about backup software. "It's easy and it works reliably. Running full backups is as simple as a mouse click. Backups that run in parallel every night are done in a couple hours. Restores of lost or corrupt file are fast and simple. Now that backups aren't running all weekend, Saturdays can be used for maintenance."

Avamar logs are checked each morning to verify that the night's backup completed successfully. The percentage of common data across all data types is typically 99.7%, which implies that a daily effective full backup of 12 TB takes only 36 GB of capacity on disk, for an effective data reduction factor of 333x.

Based on the results of ESG Lab testing, interviews with EMC Avamar customers and an audit of backup logs, ESG is confident that customers can run daily full backups and achieve a daily data reduction factor in the 300-500x range, while cost effectively retaining backups on disk for six months or longer.

Why This Matters

Information critical to the success and efficiency of an organization is not just found in corporate data centers, it also resides at remote and branch offices. Protecting remote offices using traditional backup methods can be costly and risky—especially when performed by non-technical local staff at remote offices. Replicating data over a WAN instead of shipping tapes can be cost prohibitive because of the significant amount of data that needs to be transferred across the network.

ESG Lab has confirmed that EMC Avamar source-based global data de-duplication significantly reduces the amount of data sent over the WAN, enabling fast, daily full backups for remote office data. Backups are centrally managed from a corporate data center using the intuitive Avamar Management console and data can be encrypted for security. Avamar can also be used to solve a number of data center challenges including the resource efficient backup and recovery of virtual server data (e.g. VMware environments).

Solving VMware Backup Challenges

Server virtualization creates the perfect opportunity for organizations to consolidate servers, centralize management, increase storage capacity utilization, improve availability, enhance data protection and reduce backup windows. Server virtualization can, however, introduce a new set of backup and recovery challenges: Increased storage requirements, longer backup jobs, shrinking backup windows and application performance degradation.

ESG research indicates that among current server virtualization users, 75% report that their virtualized and non-virtualized environments share the same backup processes and tools. ESG believes there are two ways to look at this data. On one hand, it implies that the majority of organizations have not yet developed comprehensive strategies when it comes to virtual server backup and data protection.

Server virtualization adoption is on the rise within organizations of all sizes. A recently completed ESG Research Survey indicates that 52% of organizations are currently utilizing server virtualization technology and 48% plan to do so.³ Twenty percent of the early adopters indicate they have deployed different backup processes and/or tools for virtual machines. This is evidence that a change in data protection processes is afoot as organizations roll out expanded server virtualization implementations and move from thinking about backing up and restoring *data* to backing up and restoring *servers* (i.e., virtual machine images).

Virtual server images, which contain operating system and application data, are great candidates for data deduplication. Consider the system boot and configuration files for a virtual server. Each virtual machine accesses the same set of files, yet operationally, each acts as if it were its own physical server. Traditional backup schemes would make copies of each virtual server's environment and send them off to the backup store. Avamar simplifies the backup and recovery of virtual machines as it reduces the amount of data backed up, which in turn minimizes the impact on host servers, network bandwidth and storage requirements.

EMC Avamar can quickly backup VMware environments via the Guest, Service Console, or VMware Consolidated Backup (VCB) proxy server to provide a resource-efficient method for the backup and recovery of virtual server data. Unlike traditional backup methods, Avamar enables fast, daily full Guest backups and individual file or .vmdk recovery. An Avamar server can also be deployed as a virtual appliance within a virtual machine (Avamar Virtual Edition), leveraging an existing ESX server and storage for simplified deployment. Whether the Avamar agent is deployed via a Guest within a virtual machine, on an ESX Server managed via the VMware Infrastructure Service Console, or at the VCB Proxy Server, Avamar's sub-file variable-length data deduplication provides significant resource and capacity savings as it eliminates redundant backup data within and across virtual machines.

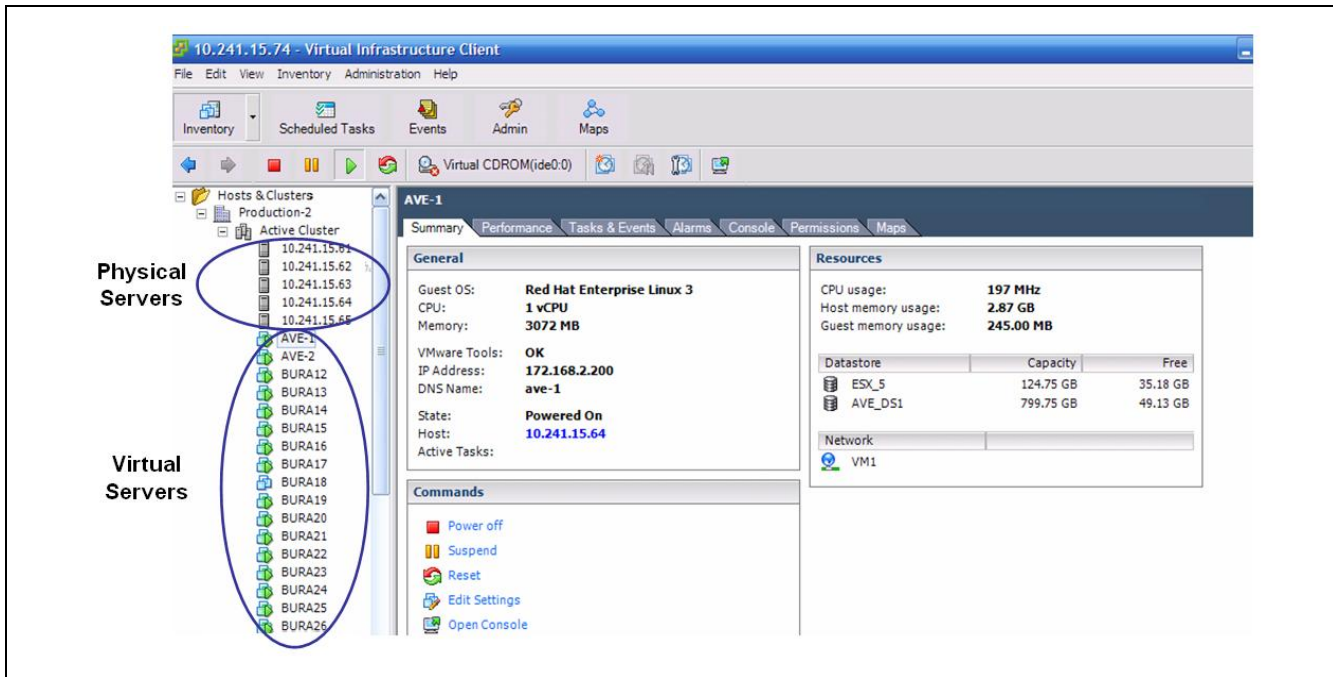
ESG Lab Testing

ESG Lab evaluated the capacity and resource savings that can be achieved with Avamar in an EMC customer proof of concept lab. The lab has been using Avamar to back up hundreds of virtual servers for more than 60 days. Data culled from EMC's production environment is merged into the customer proof of concept lab on a weekly basis to simulate the changes and growth in a production VMware environment.

The lab environment consisted of five Dell 6850 servers running VMware ESX version 3.0.2. Each server was equipped with 4-way 3.2 GHz dual core processors and 64 GB of memory. The servers were SAN-attached to a pair of EMC CLARiiON CX3-40 disk arrays. A total of 280 virtual machines were deployed on the five physical servers. Most of the virtual machines (262) were configured with Windows operating systems and a random sample of production office data culled from EMC home directories (e.g., documents, spreadsheets and presentations). The balance of the virtual machines ran Microsoft applications (Exchange, SQL Server, Sharepoint and Active Directory), Linux applications (e.g., Apache) and EMC applications (ECC Infrastructure, StorageScope and InfoScape). A view of the VMware configuration from the perspective of VMware Virtual Infrastructure console is shown in Figure 7.

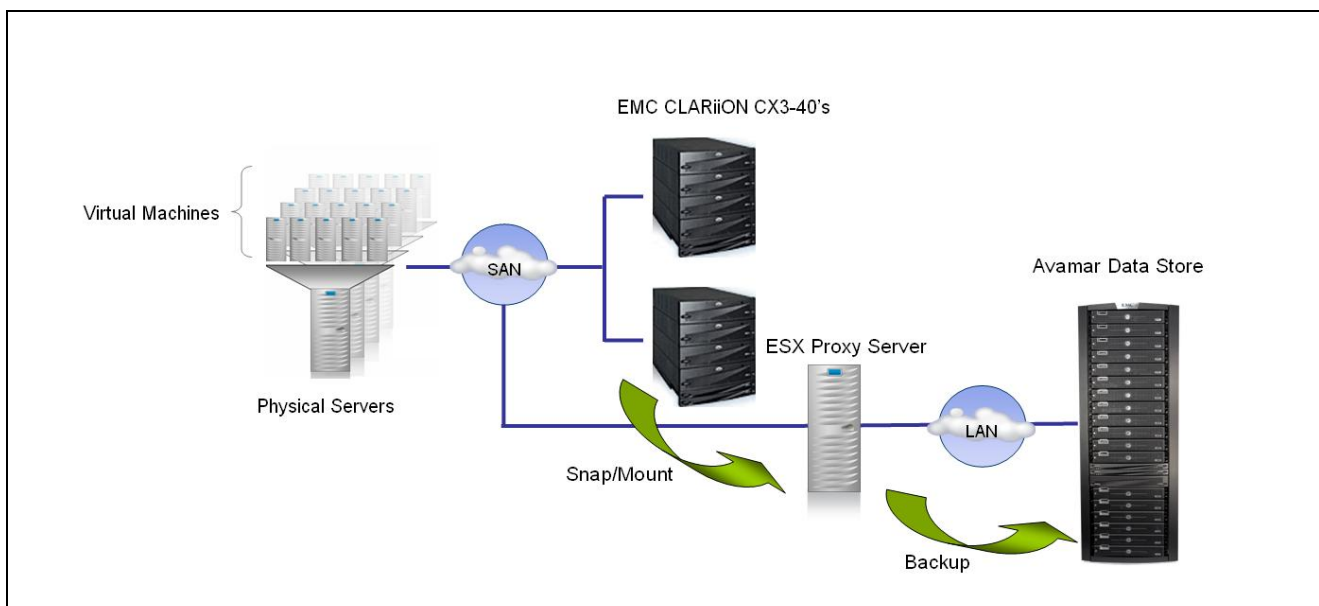
³ Source: ESG Research Report, *The Impact of Server Virtualization on Storage*, December 2007

FIGURE 7. VMWARE INFRASTRUCTURE VIEW OF THE EMC CUSTOMER PROOF OF CONCEPT LAB



ESG Lab performed a backup of 16 virtual servers. Figure 7 shows what was happening behind the scenes as Avamar software worked in concert with VMware Consolidate Backup. An EMC CLARiiON snapshot⁴ of production virtual server data was mounted by the ESX proxy server (a Dell 1950 dual core 3.2 GHz with 64 GB of memory) and backed up using Avamar software running on the proxy server. Backup data was stored on a LAN-attached Avamar Data Store.

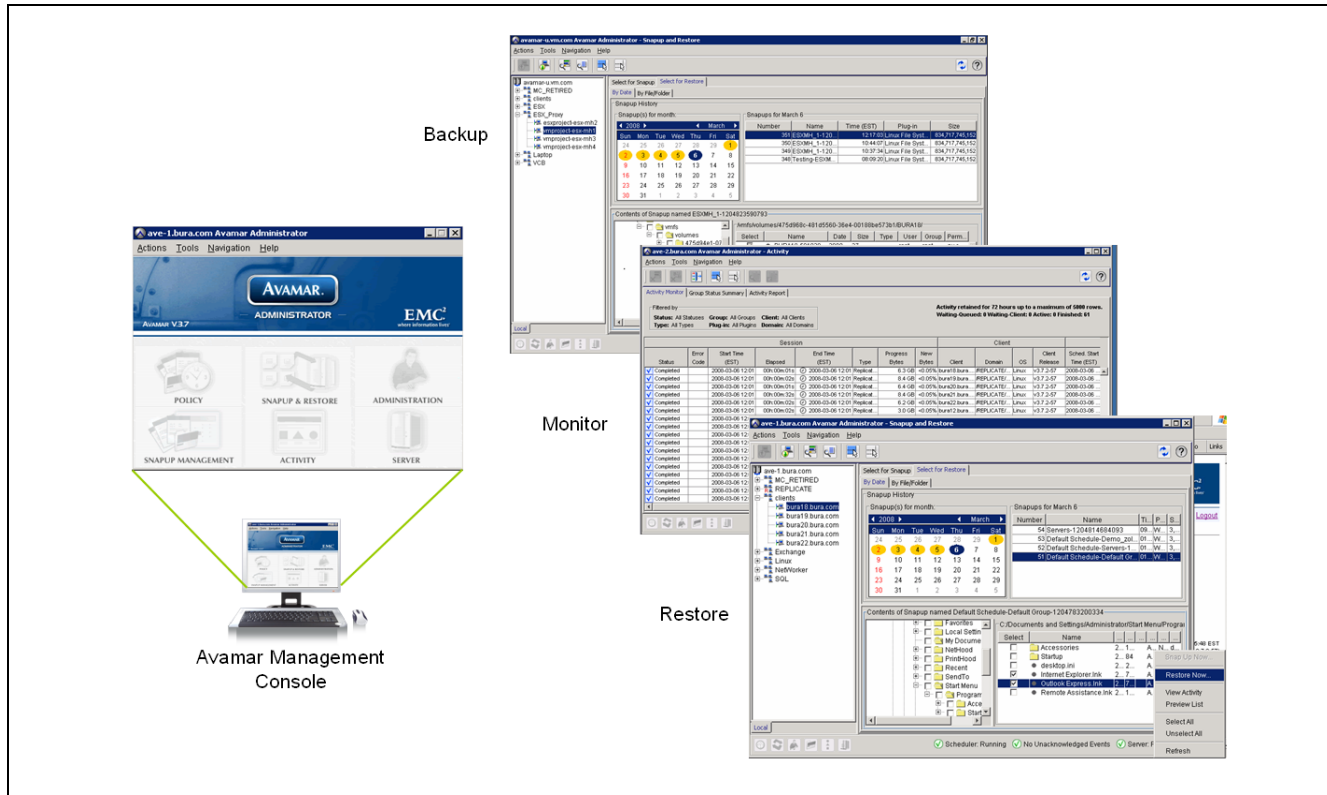
FIGURE 8. AVAMAR AND VMWARE CONSOLIDATED BACKUP



⁴ Microsoft Volume Shadow Copy Services (VSS) works together with the EMC CLARiiON behind the scenes to coordinate the creation of consistent snapshots.

The backup of five virtual Windows servers with 17 GB of data, most of which had been backed up over the past 60 days, completed in 39 seconds. The balance of the backup jobs for seven Linux servers (44 GB of data) and four Microsoft SQL Servers (16 GB of data) finished nine minutes later. A 2 MB document was restored in less than a minute. The process used to define, start and monitor the backup and recovery of virtual machine data was the same as the intuitive Avamar user interface used in physical server environments as shown in Figure 9.

FIGURE 9. AVAMAR BACKUP TESTING



ESG Lab audited Avamar log console data to determine the space reduction and backup acceleration benefits that can be achieved when Avamar is integrated with VMware Consolidated Backup running on an ESX proxy server. The results for the first three backups are shown in Table 2.

TABLE 2. AVAMAR VCB BACKUP RESULTS

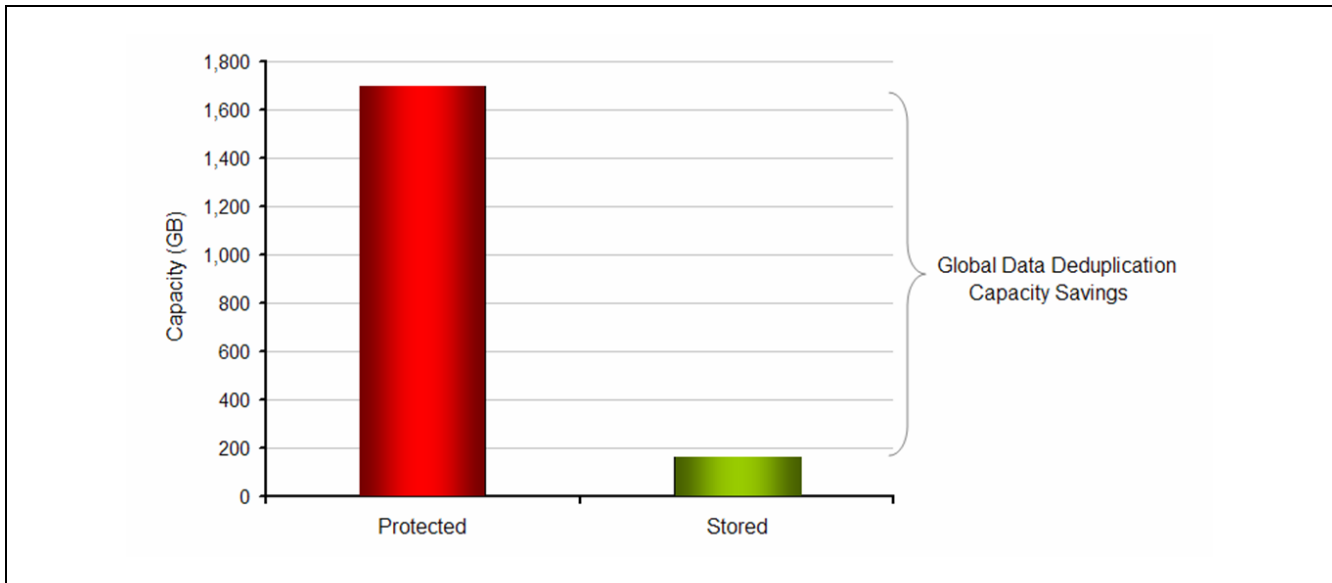
	Elapsed Time (HH:MM:SS)	Protected (GB)	Stored (GB)	Reduction Factor (protected/stored)
First Backup	07:36:00	1,700	164.6	10x
Second Backup	01:29:00	2,000	8	250x
Third Backup	00:13:26	2,000	0.2	10,000x

What the Numbers Mean

- The first backup of more than 200 virtual machines with a total of 1.7 TB stored on a pair of SAN-attached EMC CLARiiON disk arrays took seven hours and thirty six minutes to complete.
- Avamar source-based global data deduplication identified and eliminated more than 1.5 TB of redundant data that was common amongst the virtual machines.
- During the first backup, Avamar reduced the disk capacity required to store logically full backup images of more than 200 virtual machines by a factor of 10x as shown in Figure 10.

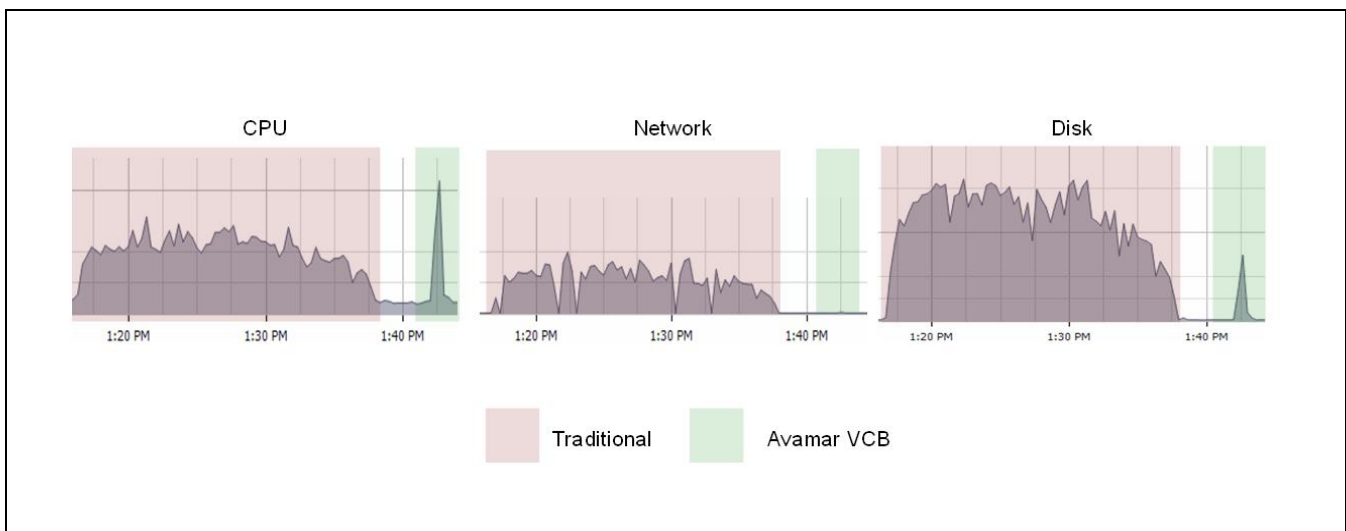
- A second backup after 300 GB of new data composed of general purpose office data pulled from production home directories completed in just under an hour and a half (01:29:00).
- The second backup provided a logically full restore image as it processed only the changes since the first backup. Only 8 GB of new and unique data was identified and stored in the Avamar Data Store for an effective capacity reduction of 250x.
- A third full backup after a trivial amount of new data had been added completed in 13 minutes and 26 seconds and added only 200 MB of new data to the data store.

FIGURE 10. VIRTUAL SERVER CAPACITY SAVINGS AFTER THE FIRST AVAMAR VCB BACKUP



Having witnessed the capacity and bandwidth savings that can be achieved with Avamar and VMware Consolidated Backups, ESG Lab analyzed server resource utilization as backup jobs were running. Using the same EMC customer proof of concept lab data set, the resource usage within one of the five physical servers as it ran a first full backup with traditional backup agents running within the guest operating system was compared to Avamar-enabled VCB method running on the ESX proxy server. The results are shown in Figure 11.

FIGURE 11. ESX SERVER UTILIZATION COMPARISON (FULL BACKUP)



What the Numbers Mean

- High disk utilization slows applications and leads to end-user complaints. The VCB backup method, one of three methods that Avamar supports for the backup and recovery of virtual server data, uses a VMware proxy server to backup snapshot images of production data. This drastically reduces the magnitude and duration of disk utilization during backups. The disk utilization impact is reduced to a level that will not be noticed by end-users.
- Except for a momentary spike in CPU utilization, the VCB method consumed less server resources than traditional methods with backup agents running at the guest operating system level.
- The duration of the backup job, and the potential impact on application performance, was significantly less with the VCB method.
- The VCB method, working in conjunction with the space and bandwidth efficiency of Avamar de-duplication, reduces the resource impact and cost of VMware backups.

Customer Feedback

ESG spoke with a backup administrator who has standardized on Avamar for the backup and recovery of VMware virtual server data. He's deployed Avamar for daily full backups of the VM Guests. Different from the VCB method presented previously where Avamar agents run within a VMware ESX proxy server, Avamar agents are deployed at the Guest level within virtual machines.

Avamar is being used to protect 123 servers and more than 3 TB of primary data. Ninety percent of the servers have been virtualized with VMware. Eighty percent of the servers are running Windows operating systems, with the balance running Linux. Prior to using Avamar, the customer's traditional backup software with tape solution was being stressed to the limit. Key backup challenges included weekly full backups that no longer fit within a single nightly backup window; full backups for groups of servers were being staggered throughout the week, while restores from tape took at least two hours and were becoming unreliable.

Using the Avamar Guest backup method, nightly full backups are running in parallel. Most are completed in less than 20 minutes and all are finished in less than three hours. Avamar de-duplication has reduced the required nightly backup disk storage capacity (and network traffic) by 99% compared to traditional full backup methods. Ad hoc restores which used to take a minimum of two hours with tape can now be completed in seconds or minutes. As summarized by the backup administrator, "Avamar de-duplication is the way to go for virtual servers – it's easy to use, it's less expensive and it's fast".

Why This Matters

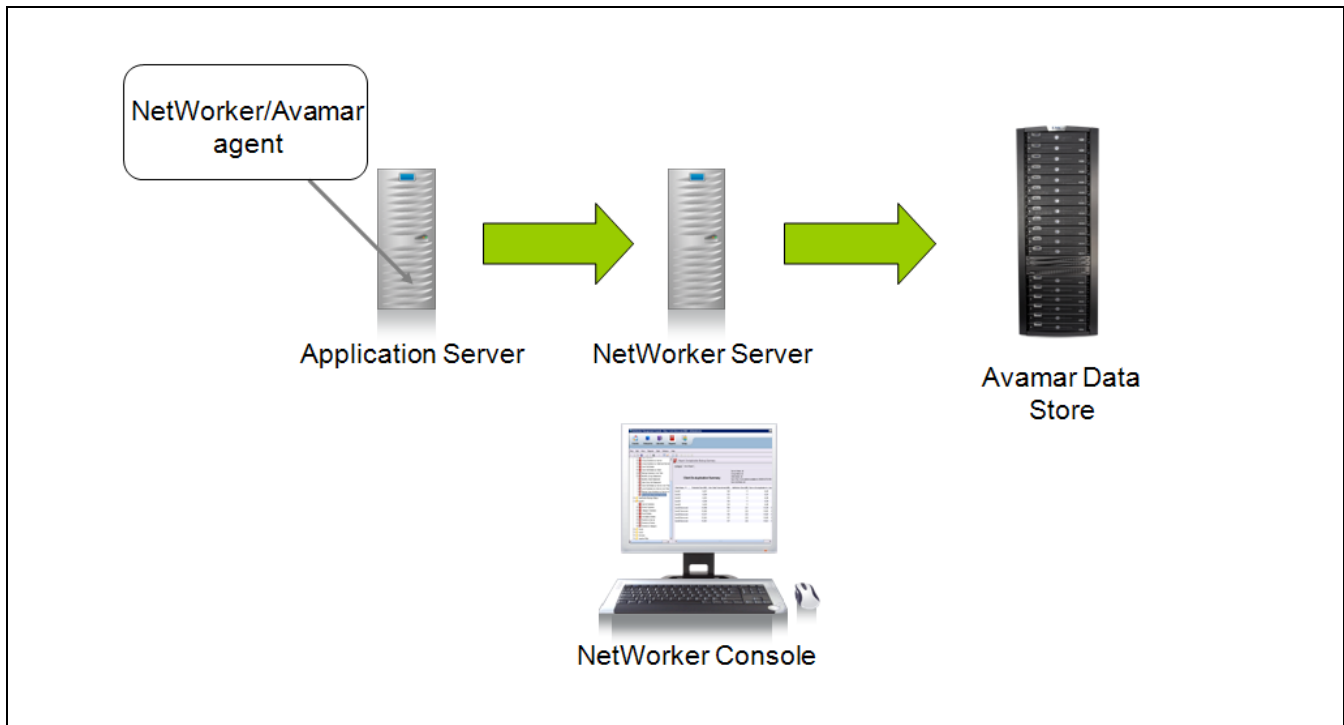
A growing number of organizations have embraced server virtualization technology to consolidate servers and reduce infrastructure costs while improving the mobility and availability of applications. Backing up virtual servers using traditional backup and recovery methods spikes server resource utilization and makes backup jobs run longer. ESG Lab has confirmed that Avamar source-based, global data de-duplication can be used to make backups run faster as it drastically reduces resource utilization within virtualized servers. The bottom line benefit is fast daily full backups, quick and reliable restores, cost savings as Avamar identifies and eliminates duplicate data within and across virtual server backup images, and potentially greater levels of server consolidation as traditional backup method bottlenecks are eliminated.

NetWorker and Avamar Integration

The latest release of Avamar has been integrated with EMC's flagship NetWorker backup product. As a result, NetWorker users can now take advantage of Avamar data de-duplication using familiar NetWorker interfaces, workflows and backup policies.

NetWorker clients can be configured to take advantage of EMC Avamar data de-duplication to dramatically decrease the amount of time, network bandwidth and disk capacity required to perform backup jobs. This is accomplished by installing a version 7.4 or later NetWorker agent with built-in Avamar support on an application server as shown in Figure 12. The agent performs source-side global data deduplication before forwarding backup data to a NetWorker server and an Avamar Data Store.

FIGURE 12. NETWORKER DATA DE-DUPLICATION IMPLEMENTATION



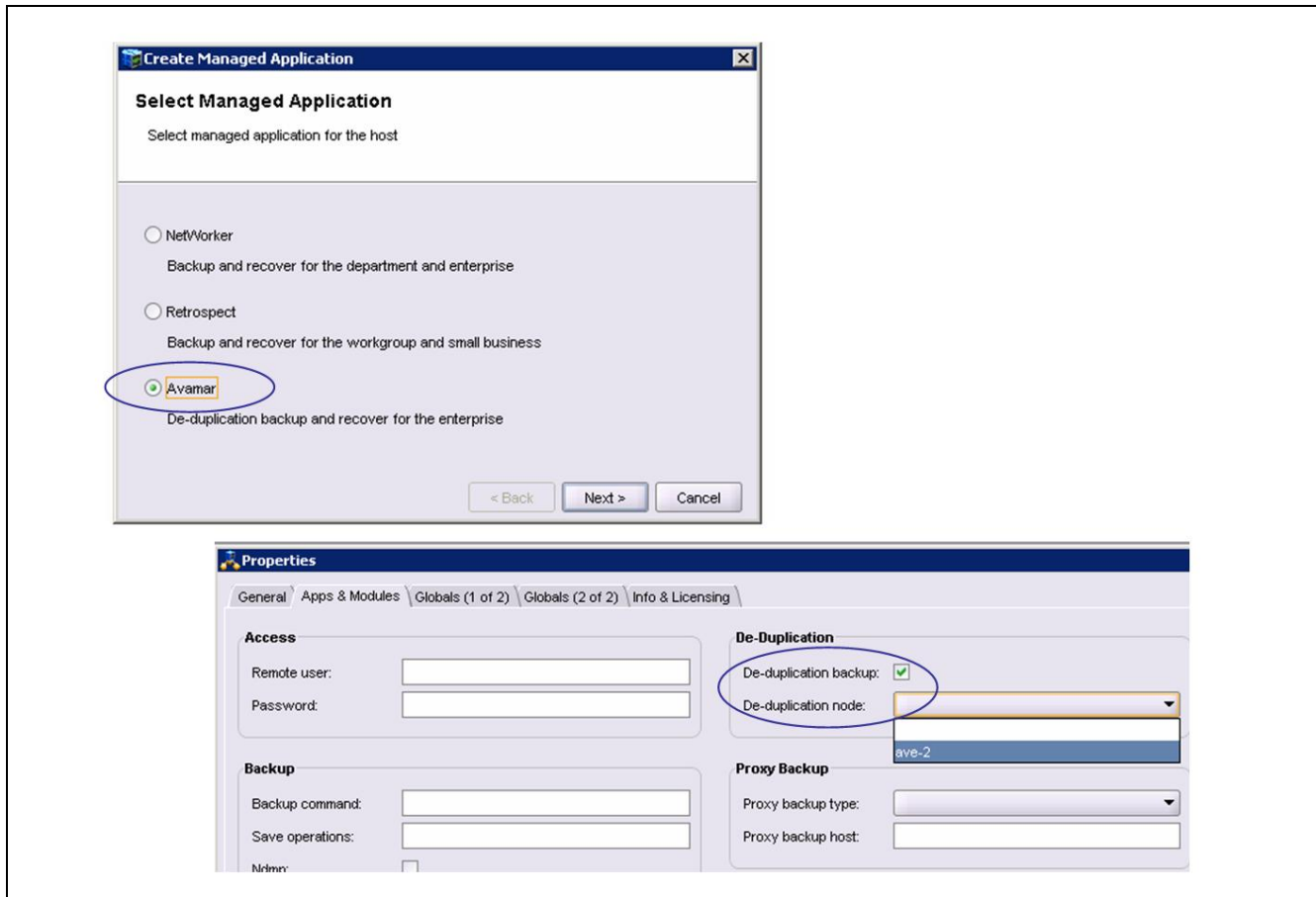
ESG Lab Testing

ESG Lab performed an Avamar-enabled backup using EMC NetWorker on a test bed with a NetWorker version 7.4 agent installed on a Windows server running within a VMware-enabled virtual machine. Configuring NetWorker to de-duplicate data was performed in three simple steps as shown in Figure 13:

1. Select Avamar as a NetWorker managed application
2. Choose an automatically discovered Avamar de-duplication node (ave-2)
3. Select the de-duplication backup checkbox

From this point on, configuring and executing backup and recovery jobs felt exactly that same as if Avamar had never been installed.

FIGURE 13. CONFIGURING NETWORKER FOR AVAMAR ENABLED DATA DE-DUPLICATION



ESG Lab used the NetWorker with Avamar-enabled de-duplication to backup a Windows virtual server. The backup took 5 minutes 31 seconds to backup up 15 GB of data using an existing NetWorker backup policy. A restore of a 10 MB log file took two seconds. A NetWorker de-duplication backup summary report was used to learn that the latest full backup of the Windows virtual server protected 15,261 MB of data, yet only 127 MB of data was transferred.

Why This Matters

A growing number of organizations are considering the benefits of data de-duplication to reduce the cost of disk-based backup and recovery. As a matter of fact, a recent ESG Research Survey indicates that 11% of respondents are currently using data de-duplication and 32% plan on doing so.⁵ Until recently, the only way to introduce data de-duplication into an existing backup and recovery environment was to swap tape or disk-based recovery hardware for an appliance that supports data de-duplication. While these hardware-based solutions can de-duplicate backup data stored on disk, they do nothing to reduce the amount of data flowing over the network and can create islands of de-duplication. ESG Lab has confirmed that EMC NetWorker can be easily deployed, leveraging Avamar de-duplication technology to create a global pool of de-duplicated data that drastically reduces disk and network requirements while preserving the familiar look and feel of NetWorker.

⁵ Source: ESG Research Report, *Data Protection*, January 2008

Solving Data Center Challenges

Over the past three years, ESG Lab has looked at a number of Avamar capabilities that can be used to solve data center challenges, including N-way grid scalability, high availability, reliability, and encryption.

Scalability

Avamar servers can be deployed in a grid architecture that can be upgraded by simply adding another server to the grid. Hard drives within each additional server increase the capacity of the total system. Adding servers also adds processors, bandwidth and memory, all of which increase total system performance. As servers are added, existing backup data is automatically rebalanced and spread evenly across all nodes, which also increases system performance.

Avamar supports a concept that ESG refers to as a “single level of management.” Regardless of the size of an Avamar grid, whether the system is built from 3 nodes or 20 nodes, it can be managed from a single user interface.

ESG Lab upgraded a three server Avamar grid to a four server grid. The fourth server added approximately 1 TB of internal de-duplicated storage capacity. Once the installation had completed and the new server had joined the cluster, existing used capacity was automatically re-balanced over the cluster. Re-balancing progress was observed on the user interface.

Why This Matters

Managing many systems is harder than managing a single system. This fact is known all too well by backup administrators struggling with the management of too many backup media servers. A centrally managed Avamar system with scalable processing power, bandwidth and capacity can be used to replace many media servers. Consolidation of multiple legacy media servers into a centrally managed, globally de-duplicated repository also reduces complexity, saves time and reduces cost.

High Availability, Reliability and Security

Avamar utilizes patented redundant array of independent nodes (RAIN) technology to deliver high availability and eliminate single points of failure. For reliability, Avamar protects itself with internal checkpoints that are verified twice daily for integrity and data recoverability. Avamar data nodes are arranged in a cluster and have been architected to survive a data node (server) or hard drive failure. After a hardware failure, the Avamar system remains available for backups and restores. A data node was powered off and a drive was failed during ESG Lab testing. In each case, the Avamar system remained available for backups and restores.

For security, Avamar provides comprehensive encryption capabilities, including the ability to encrypt backup data while in transit and at rest. ESG Lab noted that Avamar can encrypt data using a light-weight Avamar-specific protocol (AVS) or via SSL encryption, which utilizes the 128-bit Advanced Encryption Standard (AES) algorithm. 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.

Why This Matters

ESG Lab believes that it is important that the solution used to protect your data is extremely reliable and secure. While this may seem self-evident, many customers who invest in fault tolerance for primary application servers tend to ignore fault tolerance and security when architecting a backup and recovery solution. Avamar delivers built-in fault tolerance and security, using patented RAIN technology, commodity RAID technology, daily integrity checks and two levels of encryption, including industry standard AES.

ESG Lab Validation Highlights

- ☑ Daily full backups of data residing at a simulated remote office were performed over a WAN connection and managed from a centralized location with Avamar.
- ☑ Configuring and executing backups and restores was extremely easy and intuitive.
- ☑ ESG Lab verified that the combination of data de-duplication at the source and globally, compression and incremental forever algorithms can reduce the amount of backup data transferred across the network and stored to disk by up to 500x daily, compared to a traditional daily full backups to disk.
- ☑ Based on hands on testing and conversations with customers, ESG Lab is confident that Avamar can be used to reduce total backup storage requirements up to 50x over time.
- ☑ Avamar integration with VMware Consolidated Backup drastically reduced virtual server resource overhead (CPU, network and disk) while providing dramatic capacity savings due to the high degree of common data within virtual machines. Capacity was reduced by a factor of more than 10 to 1 for a first full backup of 280 virtual machines. Subsequent logical full backups consumed more than 10,000 times less capacity and bandwidth than traditional full backup methods.
- ☑ ESG Lab has confirmed that lightweight Avamar agents deployed within VM Guests enabled fast, daily full backups that could not be easily accomplished using traditional backup methods.
- ☑ EMC NetWorker backup and recovery software was configured in three easy steps to take advantage of Avamar source-based global data de-duplication technology.
- ☑ An upgrade of an Avamar cluster was performed (from 3 data nodes to 4). A single web-based interface into the upgraded cluster was used to restore files created before the upgrade had occurred.

Issues to Consider

- ☑ For companies that want copies of backup data on tape, legacy backup software and policies can be used in combination with Avamar to make full recovery images of backup data on tape. For those organizations that haven't yet adopted a policy of backing up to disk and then tape (D2D2T), it should be noted that Avamar does not currently support backup directly to tape, but this is planned for a future release.
- ☑ While integration with EMC NetWorker backup software can be used to provide the resource and performance benefits of Avamar global data de-duplication while preserving existing policies and the familiar look and feel of the NetWorker user interface, this level of integration is not supported with non-EMC backup software solutions (e.g., Veritas NetBackup, CommVault Simpana, HP Data Protector, CA ARCserve).
- ☑ ESG Lab recently heard from a happy Avamar customer who said that, "Even though my nightly backup jobs finish up to ten times faster than full weekly backups to tape, I'd like to see multiple backup jobs on the same server processed in parallel instead of one at a time". Currently if one server has multiple logical drives to backup (e.g. an internal C: drive, a NAS attached N: drive and a SAN attached S: drive), then Avamar processes the drives in sequence instead of parallel. For this particular administrator with more than 100 servers being backed up nightly it is expected that the longest running nightly backup job which currently completes in about two hours would complete closer to the majority of servers which are finished in 20 minutes. The backup administrator who made this request understands that this enhancement would generate increased network traffic that needs to be monitored. EMC has confirmed that this capability is available today through scripting and will be incorporated in a future product release.

ESG Lab's View

A majority of IT organizations are struggling with one or more inefficiencies with their backup and recovery ecosystems. Existing backup methods can't keep up with the explosive growth of data. Legacy tape-based methods can be error-prone and unreliable. The risks associated with tapes being used for offsite archival and the potential exposure of confidential data is a real concern. With these challenges as a backdrop, the protection of ROBO and virtual server data is causing a number of IT managers to rethink the way they do backup and recovery.

ESG Lab first evaluated Avamar technology in 2005. With the most intuitive graphical user interface that ESG Lab has ever evaluated, ESG believes that Avamar has revolutionized the way that backup and recovery is done. Lightweight Avamar agent software running on application servers eliminates redundant data at the source, before it is sent over a local or wide area network. Compressing and encrypting the data before it is forwarded to a global data store, Avamar drastically reduces the disk capacity and network bandwidth needed to do fast and efficient backup and recovery operations. ESG Lab retested Avamar shortly after it was acquired by EMC in 2007. Bolstered by the reputation and reach of the EMC sales and support organization, Avamar adoption has been growing rapidly—especially within enterprises looking for a hands-off, WAN-efficient solution for the protection of remote and branch office data.

After this most recent testing, ESG Lab is impressed by Avamar's newly introduced integration with VMware Consolidated Backup (VCB). Disk-based snapshots accessed from a VMware proxy server eliminate the backup window and resource problems that can plague legacy backup methods in virtual server environments. Testing within an environment where Avamar had been used for 60 days to protect 280 virtual servers running on five physical servers indicates that the Avamar method for VCB backup drastically reduces server resource utilization (CPU, memory and disk). Mature EMC snapshot technology provides a nearly instantaneous backup window. And last but not least, the high degree of commonality between virtual server images provides an outstanding reduction in the amount of disk capacity needed for fast and reliable backup and recovery operations. Similar levels of storage and resource savings can be achieved with the Avamar ESX Guest and Service Console backup methods.

A number of small to medium sized businesses, and a growing number of larger organizations, have switched from legacy backup and recovery methods to an all-Avamar approach. ESG has spoken with a number of these users and the feedback is unanimously positive. Avamar customers have consistently told ESG that they are able to reduce backup and restore times to just a fraction of what they were previously.

While the feedback from Avamar customers is positive, some enterprises have been reticent to adopt an all-Avamar approach in the data center due to investments made in legacy backup software over the years. A growing number of enterprises are, however, deploying Avamar in problem areas to protect ROBOs, virtual servers and large file servers and then expanding into problem areas in the data center. ROBO is a particular sweet spot for early adopters, as there is often no incumbent backup solution due to problems with tape management and lack of WAN bandwidth. Leveraging WAN-efficient source-based global data de-duplication, Avamar is being used for hands-off protection of ROBO data that is managed from a corporate data center.

For organizations that have standardized on EMC NetWorker backup and recovery software, the latest release of EMC Avamar provides a new and powerful hybrid approach. A single NetWorker agent within an application server can be used to provide the network and capacity savings of Avamar with the familiar look and feel of NetWorker. ESG Lab tested this capability and found that after installing the NetWorker agent, a one time click of checkbox within the NetWorker user interface is all that is needed to achieve the capacity and network bandwidth savings provided by Avamar's patented source-based global data de-duplication technology.

IT organizations have been struggling with backup window and recovery reliability issues for decades. ESG believes that the backup and recovery market is ripe for a revolution and that EMC Avamar is a compelling solution that is leading the way.

Appendix

The equipment used during the ESG Lab validation is listed below along with the types of application data used for backup and restore testing.

ROBO Test Bed

Corporate Data Center	
Avamar software revisions	4.0
Avamar Utility Node	1U Avamar server, 4x750GB SATA
Avamar Data Node 1	2U HP DL380, 6x300 GB SCSI
Avamar Data Node 2	2U Dell PE 2850 , 6x300GB SCSI
Avamar Data Node 3	2U SuperMico Optima, 6x300 GB SCSI
Avamar Data Node 4	2U IBM x346,, 6x300 GB SCSI
GigE switches	Dell PowerConnect 48 port, Cisco Catalyst 4506 6-slot.
Remote Office	
Single Node Avamar server	2U Dell PE 2850 , 6x300GB SCSI
GigE switch	ProCurve 2848 48-port
Windows server	Xeon 3.0 GHz, 4 GB RAM, W2K3
Linux server	Dual 3.0 GHz Xeon, 4 GB RAM, RHEL 4
Solaris server	UltraSparc-IIIi, 2 GB RAM, Solaris 9

Source-Based Global Data De-duplication Test Bed

Server	Amount of Data	Server Type	O/S	Application
Linux file system	8.5 GB	Dual 3.0 GHz Xeon, 4 GB RAM	RHEL 4.0	ext3
Windows file system	8 GB	Xeon 3.0 GHz, 4 GB RAM	W2K3	NTFS
Windows/Oracle	30 GB	Dual 2.4 GHz Xeon, 2 GB RAM	W2K3	Oracle 10 R1
Windows/Exchange	12 GB	2.8 GHz Pentium 4, 1 GB RAM	W2K3	Exchange 2003
Solaris file system	10 GB	Dual UltraSparc-IIIi, 2 GB RAM	Solaris 9	ZFS

Integration with EMC NetWorker

EMC NetWorker	Version 7.4, SP1, build 335
EMC Avamar	Version 3.7.2.57

Integration with VMware Consolidated Backup Test Bed

Physical Servers	Six Dell 6850 4x3.2GHz dual core CPU,64 GB RAM
Storage	Two EMC CLARiiON CX3-40 disk arrays
Virtual Server Software	VMware ESX version 3.0.2
Virtual Machines	280 in total including Windows, Linux, Exchange, SQL Server, Active Directory, ECC Infrastructure, Storage Scope, InfoScape and SharePoint
ESX Proxy Server	Dell 1950, dual 3.2 GHz CPU. 64 GB RAM
Avamar Data Store	Avamar version 3.7.1, 5+1 server configuration (four active)



20 Asylum Street
Milford, MA 01757
Tel: 508-482-0188
Fax: 508-482-0218

www.enterprisestrategygroup.com