

# Understanding EMC Avamar with EMC Data Protection Advisor

*Applied Technology*

---

## **Abstract**

EMC® Data Protection Advisor provides a comprehensive set of features that reduce the complexity of managing data protection environments, improve compliance with business and regulatory requirements, and reduce the risk of data loss. This white paper outlines how Data Protection Advisor helps you gain control of your EMC Avamar® backup environment, by better understanding what is working well and what is not, giving you a proactive approach to managing your environment.

April 2011

---

---

Copyright © 2010, 2011 EMC Corporation. All rights reserved.

EMC believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” EMC CORPORATION MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any EMC software described in this publication requires an applicable software license.

For the most up-to-date listing of EMC product names, see EMC Corporation Trademarks on [EMC.com](http://EMC.com)

All other trademarks used herein are the property of their respective owners.

Part Number h6108.1

---

## Table of Contents

<b>Executive summary .....</b>	<b>4</b>
<b>Introduction .....</b>	<b>4</b>
Audience .....	4
DPA licensing with Avamar .....	4
<b>EMC Avamar .....</b>	<b>4</b>
<b>Avamar Data Store .....</b>	<b>5</b>
How does it work .....	5
Avamar Server component .....	7
<b>DPA 5.8 and Avamar 6.0 .....</b>	<b>7</b>
<b>Improved Avamar reporting .....</b>	<b>7</b>
<b>Conclusion .....</b>	<b>9</b>
<b>References .....</b>	<b>9</b>

---

## Executive summary

EMC® Data Protection Advisor (DPA) provides a unified data protection management window by collecting, monitoring, analyzing, and reporting on information from the entire data protection infrastructure. When EMC Avamar® is introduced into an environment, typically two or more backup solutions are being managed, so DPA eliminates the need for multiple interfaces.

EMC Avamar Data Store is a fully integrated, software/hardware, backup and recovery solution that also includes data deduplication. Avamar's deduplication begins at the client to be backed up. Avamar detects sub-file-level data changes and backs them up only if they have *not* already been backed up. This results in only the changed data being moved, enabling savings in the areas of time, processing, bandwidth, and storage requirements.

EMC Data Protection Advisor enhances the abilities of Avamar monitoring and reporting, through its central historical repository, analysis engine, and robust reporting.

## Introduction

This white paper describes the EMC Avamar technology, including its architecture and various components. It then details how EMC Data Protection Advisor can extend Avamar by showing some of the different options available.

### ***Audience***

This white paper is intended for Data Protection Advisor professionals to provide insight into what Avamar is all about and how Data Protection Advisor extends reporting of Avamar environments.

### ***DPA licensing with Avamar***

With DPA 5.8, DPA licenses Avamar environments based on the capacity of each grid. Customers who use DPA to monitor a variety of backup applications, including Avamar, will require both a DPA Backup Capacity license and Client license.

For example, a customer with a 2 TB Avamar grid will require a corresponding 2 TB DPA Backup Capacity license, regardless of the number of Avamar clients.

Avamar clients managed through EMC NetWorker® do not require a DPA Backup Capacity license as they are included as clients of the NetWorker server.

## EMC Avamar

As mentioned previously, Avamar's source-based deduplication begins at the client by detecting sub-file-level data changes and backing up only changed blocks of data. This results in only the changed data being transmitted over the network for backup and in huge savings with both the time taken to back up and with network bandwidth requirements. A backup client must be backed up in full once, and thereafter only changes are backed up to the Avamar Data Store.

Each backup job is treated as a full-level backup since Avamar can recover the client in its entirety from any backup, even though only incremental changes are saved from each backup.

Besides deduplicating data at the backup client level, Avamar also deduplicates data across clients and sites, a process that is known as global deduplication. This means that Avamar stores duplicate data/sub-file chunks of data only once. For example, since the majority of the C:\WINDOWS subdirectory and its contents will be exactly the same across all Windows backup clients, Avamar needs to store this only once but can allow any client to restore from it.

In environments where clients have been virtualized using VMware, Avamar also provides the same significant savings. With traditional backup solutions, virtual machines need to move weekly full and daily

---

incremental backups through the VMware host's physical resources (NIC, CPU memory, and so on), which can become overloaded and cause backup overruns and breached SLAs. With Avamar, sending only changed segments of data can potentially provide a 300 percent daily reduction in network resource usage compared to traditional backups. Avamar also does remote replication over standard wide area network (WAN) connections, thus providing disaster recovery replication without transporting tapes from one site to another. For full virtual machine backups, Avamar can also deduplicate data stored in virtual disks (VMware \*.vmdk files), significantly reducing storage consumption and enabling replication of virtual disks across congested WANs.

## Avamar Data Store

Avamar's deduplication begins at the client to be backed up. Avamar detects sub-file-level data changes and backs it up only if that sub-file (chunk) has *not* already been backed up. This results in only the "changed" data being transmitted over the network for backup, and also in sizable reductions on the time taken to conduct backup coupled with meeting a highly efficient bandwidth utilization requirement. This implies that a backup client needs to be "backed up" in FULL at least once, and thereafter only "changes" are backed up to the Data Store.

This implies a traditional INCREMENTAL backup; however, each backup "job" is treated, and is, a FULL level backup as Avamar has the ability to recover the client in its entirety.

Besides deduplicating at the backup client level, Avamar also deduplicates across clients and sites – this is also known as global deduplication. This means that duplicate data/sub-file chunks of data will only be stored *once* on the Avamar Data Store or grid. A typical example would be the C:\WINDOWS sub-directory and its contents; this global deduplication benefit will be the same across *all* Windows backup clients and thus Avamar only needs to store this directory *once* but the effect also allows *any* client to restore from it. As an example, a company logo and confidentiality statement that are stamped on every customer-facing presentation are stored only once on the grid but are available for all restores of that presentation no matter which business unit or division. In summary, global deduplication has no digital borders – global commonality is restricted only to the size of the Avamar grid.

## How does it work

Avamar eliminates redundancy in backup data at the source, that is, before the data is transferred across the LAN or WAN during backups. Avamar agents are deployed on the systems to be protected (for example, physical or virtual servers, Network Data Management Protocol (NDMP) filers, desktops, laptops) to identify and filter over time duplicate data segments stored in files within a single host and across multiple systems. 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 day-to-day backup data.

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

A key factor for eliminating redundant data at a sub-file (or segment) level is the method for determining segment size. Snapshot or replication technologies commonly use fixed-block/fixed-length segments. Unfortunately, even small changes to a dataset (for example, inserting data at the head of the file) can change all fixed-length segments after that insertion thus wiping out all deduplication in that dataset; this condition is called offset. In contrast, Avamar uses an intelligent method for determining segment size that inspects each segment in the data.

Avamar's patented method for determining segment size is designed to yield optimal efficiency across all systems in an enterprise. Avamar's algorithm analyzes the binary structure of a dataset to determine segment boundaries that are context-dependent, so that Avamar's client agents can identify the exact same segments for any dataset, no matter where that dataset is stored in the enterprise. Avamar's segments average 24 KB in size and are then compressed prior to LAN or WAN transmission to an average of just 12 KB. By analyzing the binary structure, Avamar's method works for all file types and sizes.

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 software then uses this 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.

In summary, the benefits of Avamar are:

- A reduction in daily network bandwidth and backup storage by up to 500x.
- Daily *full* level backups across existing LAN/WAN bandwidths.
- Up to 10x faster backup performance.
- Total back-end disk backup storage reduction of up to 50 percent.
- Up to 85 percent reduction in total client CPU utilization. Avamar clients run in low priority (or “nice” mode in UNIX) to avoid resource contention. While Avamar clients typically use 15 percent more CPU than traditional backup agents during backup operations, Avamar reduces the window required for backup operations by up to 10x, thus reducing overall CPU utilization.
- Immediate single-step recovery. Avamar stores all backups as virtual full images, which can be immediately recovered in a single step to either their original destination or any other IP-addressable node that role-based access control permits. It is not necessary to restore data from full and incremental backups to reach the desired recovery point.

Avamar Data Stores can be deployed in either single-node or scalable multinode configurations depending on the performance and capacity required. Built-in software includes a replication software module for efficient replication of Avamar Servers across WAN sites or any IP-based circuits.

Avamar Data Store multinode configurations use a Redundant Array of Independent Nodes (RAIN) architecture, thus providing scalability, optimization, and resilience to individual node failure – RAIN does *not* support single-node configurations but is available in three Data Store and the previously mentioned configurations.

Avamar Data Stores are offered in two primary classes: 3.3 and 2 TB high-density nodes and 1 TB standard density nodes (high and standard density nodes cannot be mixed in a single Avamar Data Store to ensure that RAIN reconciles all disk geometries of each potential data store that it protects in the grid).

**Table 1. Comparison of single- and multinode configurations**

	<b>Maximum configuration (16 storage nodes)</b>	<b>Single node configuration (standard density/high density)</b>
Licensed capacity	16 TB / 32 TB / 52.8 TB	1.0 TB / 2.0 TB / 3.3 TB
Equivalent cumulative, traditional backup storage	.55 PB / 1.1 PB / 2.4 PB	35 TB / 70 TB / 105 TB

Gigabit Ethernet is used with TCP/IP protocols. Avamar 6.0 bonds these interfaces for additional performance.

Avamar Data Store is available in various models:

- A scalable multinode (6-18 nodes, including spare and utility using RAIN technology, expandable in single same-size node increments) providing upgrade and high-availability options.
- One- and two-node, fixed capacity models (ideal for remote offices or smaller data center environments)

Avamar Servers are delivered with a supported operating system pre-installed (primarily a certified Intel x86 server running Red Hat Enterprise Linux 4 for Avamar versions 3-5, SUSE LINUX Enterprise 11.1 for Avamar 6.0), all with fully tested hardware from the factory. The Avamar software is installed onsite after hardware installation using kick-start media that has all the approved drivers, RPMs, and so on. Supportability features include remote access and management, email-home capabilities, and enabled SNMP. Avamar 6.0 goes one step further by including EMC Secure Remote Services (ESRS) with a

---

secured VPN configured to alert both customer and EMC of disk failures, security patches, and remote support. ESRS also delivers the ability to move patches and updates to the server and install them.

Avamar supports all major operating systems and provides dedicated agents for protecting databases and network-attached storage with NDMP for both EMC Celerra® (including the new EMC VNX™ system) and NetApp systems.

## **Avamar Server component**

Avamar Servers store client backups and manage the policies for scheduling and for determining datasets and retention periods. An Avamar backup provides a point-in-time full copy of data that can be restored on demand. In multinode Avamar Server environments, the components of the Avamar Server are segregated across multiple servers for scalability and performance.

The two primary node types configured in a multimode Avamar Server are:

- Storage node – stores the actual backup data. Storage nodes can be added to an Avamar Server over time with no downtime required. Avamar clients connect directly with the storage node, and client connections and data are load balanced across storage nodes.
- Utility node – schedules and manages background Avamar Server jobs. One utility node is configured per Avamar Server.

The NDMP Accelerator node is an optional generic utility-like node running Red Hat/SLES that works with NDMP in order to provide protection for NAS filers.

For integration with reporting tools, Avamar uses a Postgres SQL database to store information about backup activities and policies. This database can be probed by standard ODBC calls using a published schema.

## **DPA 5.8 and Avamar 6.0**

To gather data from EMC Avamar, DPA connects directly to the Avamar database via the **mcldb** database on the default port (5555) for Avamar. If these parameters were modified, change the Avamar Config and Avamar Job Monitor requests to override these DPA defaults to match the new Avamar parameter settings. When DPA connects to the database, it uses the **viewuser** account to log in to the database. If the Avamar installation was modified so that “viewuser” does not have permission to log in to the database, or the password for this user has been modified, change the user and password in the Default Avamar Credentials to reflect the username and password that should be used to connect to the database.

The Collector must be installed on a host that *is in the same time zone as the Avamar server.*

## **Improved Avamar reporting**

DPA 5.8 includes new features that enhance the reporting capabilities for Avamar environments:

- Backup Job Report – The following figure is an example of the updated Backup jobs report. This report now includes fields such as backup path, error summary code, and status of the code. This allows administrators to manage backups more effectively.

Error Code Summary	Status Code Summary	Effective Path	Plugin Name	Bytes Modified Sent (MB)	Bytes Modified Not Sent (GB)
Administrative code	Activity failed - timed out before starting	Infrastructure\Sharepoint\Sharepoint 2007 Full AVAMAR020	Microsoft Office SharePoint Server	0.0	0.0
Administrative code	Activity failed - timed out before starting	MS SQL Dataset	Windows SQL	0.0	0.0
Command completed with errors, client log should be examined	Activity completed with exceptions	/Default Dataset	Windows File System	628.8	6.4
Command completed with errors, client log should be examined	Activity completed with exceptions	/Default Dataset	Windows File System	36.4	0.4
Administrative code	Activity completed successfully	/Client On-Demand Data	Macintosh File System	35.9	5.3
Administrative code	Activity failed - timed out before starting	/DTLT General Atwood	Windows File System	0.0	0.0
Administrative code	Activity completed successfully	/DTLT General Atwood	Windows File System	44.4	0.1
Administrative code	Activity failed - timed out before starting	/DTLT General Davis	Windows File System	0.0	0.0
Administrative code	Activity failed - timed out before starting	/DTLT General Emrlev	Windows File System	0.0	0.0

**Figure 1. Example updated Backup Jobs Report**

- Used Capacity Report - The following figure is a sample Used Capacity Report, which contains the total used capacity of the Avamar grid. In addition it also indicates utilization of each Avamar Server.

Server	Utilization (%)	Used Capacity (GB)
avamar220.lss.emc.com	2.8	208.3

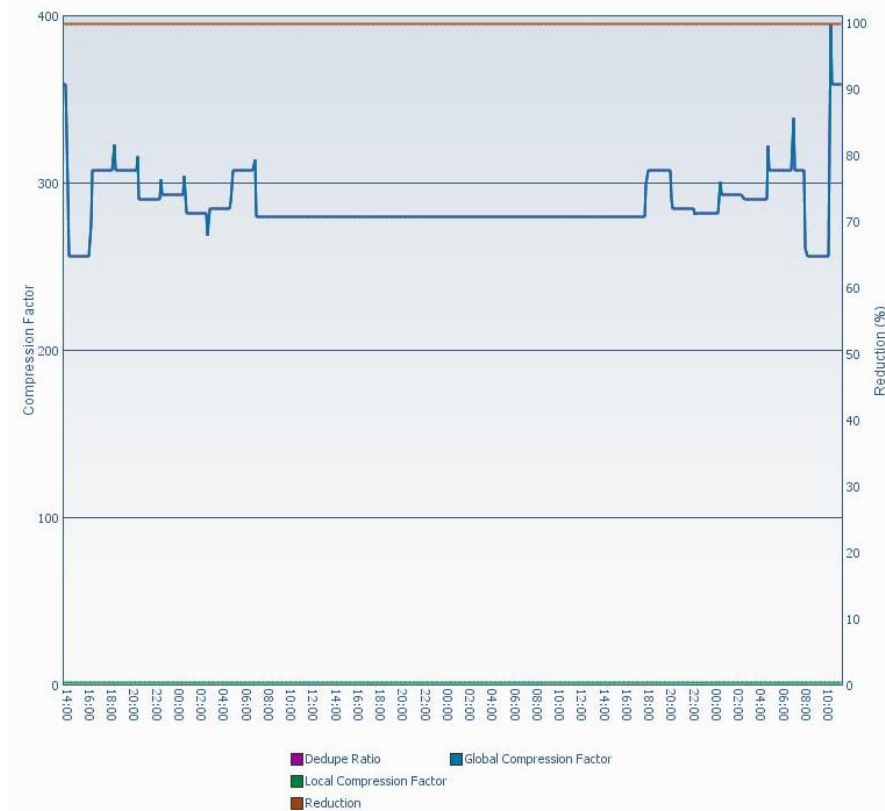
**Figure 2. Used Capacity Report**

- Restore Job Report - The following figure is a sample Restore Job Report, which now lists Job ID, backup number, backup label, recorded data time, number files and bytes scanned.

App Job Id	Job	Backup Number	Number Of Files	Scanned (MB)	Plugin Name
9129964982108809	Windows File System-Backup #50 label "General Field DTLT Policy-1297643977813" timestamp 2011-03-09 05:52:49.568, 1 files, 4 Mbytes	50	1	4.0	Windows File System
9129964983293509	Windows File System-Backup #50 label "General Field DTLT Policy-1297643977813" timestamp	50	1	4.0	Windows File System

**Figure 3. Restore Job Report**

- Daily Compression Rate – Lastly Avamar now has a report that can identify Daily Compression for the following functions, as shown in Figure 4:
  - Dedup Ratio
  - Local Compression Factor
  - Global Compression Factor
  - Reduction



**Figure 4. Daily Compression Rate**

## Conclusion

With the addition of these new reports to DPA, Avamar allows customers to more effectively manage their environments. In addition administrators can accurately understand consumption of each Avamar grid and appropriately make retention and management decisions more accurately. While DPA can be useful in an Avamar-only environment, its value increases when an environment has multiple backup solutions, multiple sites, or multiple business units. This paper does not cover all reports that are available within DPA, but it does highlight areas of interest when using DPA with EMC Avamar.

## References

The following documents can provide additional information and can be found on the EMC Powerlink® website.

- *EMC Data Protection Advisor Version 5.8 Architecture Overview*
- *EMC Data Protection Advisor Version 5.8 Compatibility Matrix*
- *EMC Data Protection Advisor Version 5.8 Installation Guide*
- *EMC Data Protection Advisor Version 5.8 Administration Guide*
- *EMC Data Protection Advisor Version 5.8 Reference Guide*
- *EMC Data Protection Advisor Version 5.8 Release Notes*
- *EMC Data Protection Advisor Version 5.8 User Guide*