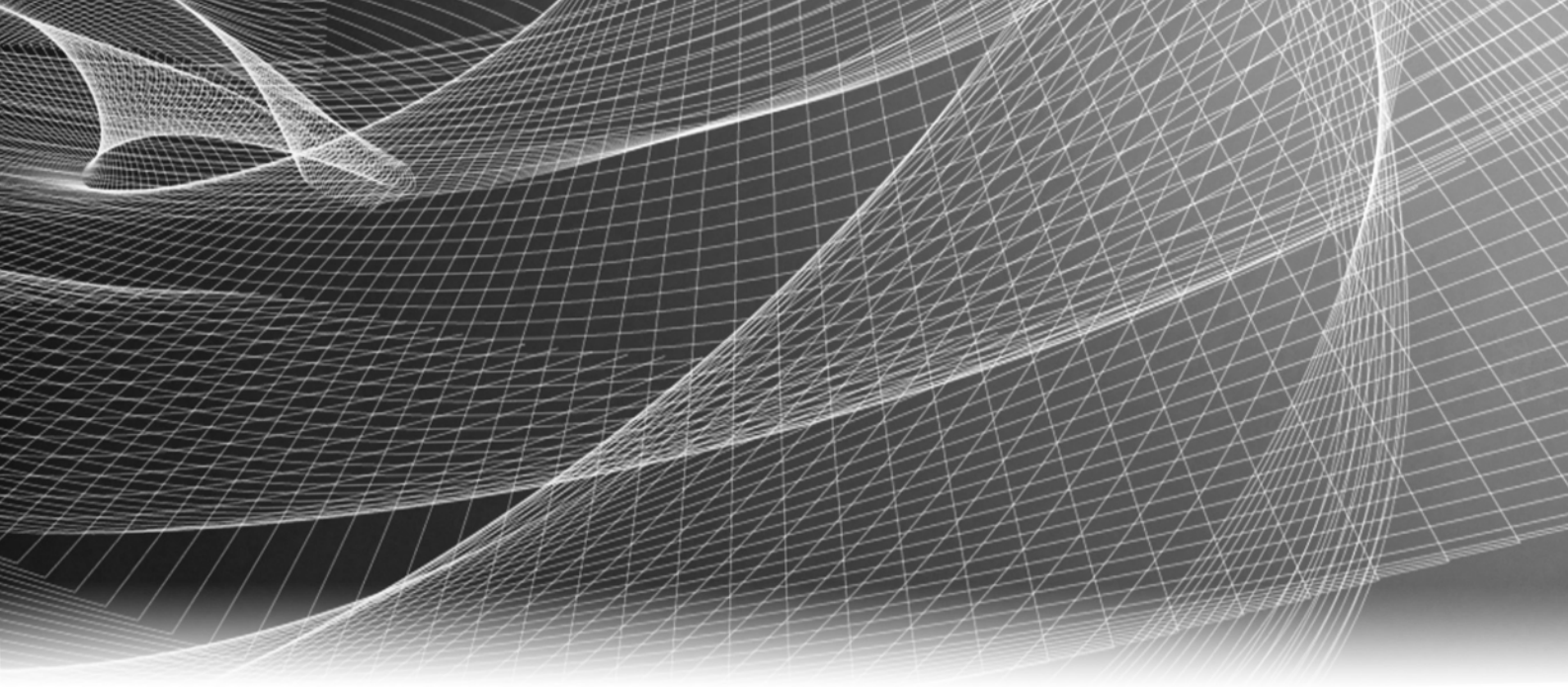


EMC® Avamar® 6.1
Operational Best Practices

P/N 300-013-346
REV 04

EMC²

Published March, 2013

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 regulatory document for your product line, go to the technical documentation and advisories section on the EMC online support website.

CONTENTS

TABLES

PREFACE

As part of an effort to improve its product lines, EMC periodically releases revisions of its software and hardware. Therefore, some functions described in this document might not be supported by all versions of the software or hardware currently in use. The product release notes provide the most up-to-date information on product features.

Contact your EMC representative if a product does not function properly or does not function as described in this document.

NOTICE

This document was accurate at publication time. New versions of this document might be released on the EMC online support website. Check the EMC online support website to ensure that you are using the latest version of this document.

Purpose

This guide describes operational best practices for both single-node and multi-node servers in small and large heterogeneous client environments. This guide does not provide introductory materials for basic Avamar technology or delivery methods.

Audience

The intended audience of this document is experienced UNIX, Linux, and Windows system administrators who will deploy and operate Avamar servers.

Revision history

The following table presents the revision history of this document.

Table 1 Revision history

Revision	Date	Description
04	March 1, 2013	Updated the definition of ModReduced in Chapter 8, "Understanding DPN Summary Reports."
03	July 31, 2012	Updated "Where to get help" on page 12 in the Preface.
A02	June 15, 2012	Revised "Steps to recover from capacity issues" on page 31
A01	April 25, 2012	First release of Avamar 6.1.

Related documentation

The following EMC publications provide additional information:

- *EMC Avamar Administration Guide*
- *EMC Avamar Backup Clients User Guide*
- *EMC Avamar Data Store Customer Service Guide*
- *EMC Avamar Data Store Single Node Customer Installation Guide*

- *EMC Avamar for Exchange VSS User Guide*
- *EMC Avamar for Hyper-V VSS User Guide*
- *EMC Avamar for IBM DB2 User Guide*
- *EMC Avamar for Lotus Domino User Guide*
- *EMC Avamar for Microsoft SharePoint Guide*
- *EMC Avamar for Oracle User Guide*
- *EMC Avamar for SAP with Oracle User Guide*
- *EMC Avamar for SQL Server User Guide*
- *EMC Avamar for Sybase ASE User Guide*
- *EMC Avamar for VMware User Guide*
- *EMC Avamar for Windows Server User Guide*
- *EMC Avamar Management Console Command Line Interface (MCCLI) Programmer Guide*
- *EMC Avamar NDMP Accelerator User Guide*
- *EMC Avamar Product Security Guide*
- *EMC Avamar Release Notes*
- *EMC Avamar Release Notes Addendum*
- White paper: *Efficient Data Protection with EMC Avamar Global Deduplication Software - Technology Concepts and Business Considerations*
- White paper: *Optimized Backup and Recovery for VMware® Infrastructure with EMC Avamar*

Conventions used in this document

EMC uses the following conventions for special notices:

⚠DANGER

DANGER indicates a hazardous situation which, if not avoided, will result in death or serious injury.

⚠WARNING

WARNING indicates a hazardous situation which, if not avoided, could result in death or serious injury.

⚠CAUTION

CAUTION, used with the safety alert symbol, indicates a hazardous situation which, if not avoided, could result in minor or moderate injury.

NOTICE

NOTICE is used to address practices not related to personal injury.

Note: A note presents information that is important, but not hazard-related.

IMPORTANT

An important notice contains information essential to software or hardware operation.

Typographical conventions

EMC uses the following type style conventions in this document:

Normal	Used in running (nonprocedural) text for: • Names of interface elements, such as names of windows, dialog boxes, buttons, fields, and menus • Names of resources, attributes, pools, Boolean expressions, buttons, DQL statements, keywords, clauses, environment variables, functions, and utilities • URLs, pathnames, file names, directory names, computer names, links, groups, service keys, file systems, and notifications
Bold	Used in running (nonprocedural) text for names of commands, daemons, options, programs, processes, services, applications, utilities, kernels, notifications, system calls, and man pages Used in procedures for: • Names of interface elements, such as names of windows, dialog boxes, buttons, fields, and menus • What the user specifically selects, clicks, presses, or types
Italic	Used in all text (including procedures) for: • Full titles of publications referenced in text • Emphasis, for example, a new term • Variables
`Courier`	Used for: • System output, such as an error message or script • URLs, complete paths, file names, prompts, and syntax when shown outside of running text
`Courier bold`	Used for specific user input, such as commands
`Courier italic`	Used in procedures for: • Variables on the command line • User input variables
< >	Angle brackets enclose parameter or variable values supplied by the user
[ ]	Square brackets enclose optional values
|	Vertical bar indicates alternate selections — the bar means “or”
{ }	Braces enclose content that the user must specify, such as x or y or z
...	Ellipses indicate nonessential information omitted from the example

Where to get help

The Avamar support page provides access to licensing information, product documentation, advisories, and downloads, as well as how-to and troubleshooting information. This information may enable you to resolve a product issue before you contact EMC Customer Service.

To access the Avamar support page:

1. Go to https://support.EMC.com/products.
2. Type a product name in the **Find a Product** box.
3. Select the product from the list that appears.
4. Click the arrow next to the **Find a Product** box.
5. (Optional) Add the product to the **My Products** list by clicking **Add to my products** in the top right corner of the **Support by Product** page.

Documentation

The Avamar product documentation provides a comprehensive set of feature overview, operational task, and technical reference information. Review the following documents in addition to product administration and user guides:

- Release notes provide an overview of new features and known limitations for a release.
- Technical notes provide technical details about specific product features, including step-by-step tasks, where necessary.
- White papers provide an in-depth technical perspective of a product or products as applied to critical business issues or requirements.

Knowledgebase

The EMC Knowledgebase contains applicable solutions that you can search for either by solution number (for example, esgxxxxxx) or by keyword.

To search the EMC Knowledgebase:

1. Click the **Search** link at the top of the page.
2. Type either the solution number or keywords in the search box.
3. (Optional) Limit the search to specific products by typing a product name in the **Scope by product** box and then selecting the product from the list that appears.
4. Select **Knowledgebase** from the **Scope by resource** list.
5. (Optional) Specify advanced options by clicking **Advanced options** and specifying values in the available fields.
6. Click the search button.

Live chat

To engage EMC Customer Service by using live interactive chat, click Join Live Chat on the Service Center panel of the Avamar support page.

Service Requests

For in-depth help from EMC Customer Service, submit a service request by clicking Create Service Requests on the Service Center panel of the Avamar support page.

Note: To open a service request, you must have a valid support agreement. Contact your EMC sales representative for details about obtaining a valid support agreement or with questions about your account.

To review an open service request, click the Service Center link on the Service Center panel, and then click View and manage service requests.

Facilitating support

EMC recommends that you enable ConnectEMC and Email Home on all Avamar systems:

- ConnectEMC automatically generates service requests for high priority events.
- Email Home emails configuration, capacity, and general system information to EMC Customer Service.

Your comments

Your suggestions help us to continue to improve the accuracy, organization, and overall quality of the user publications. Send your opinions of this document to:

BSGDocumentation@emc.com

Please include the following information:

- Product name and version
- Document name, part number, and revision (for example, 01)
- Page numbers
- Other details that will help us address the documentation issue

CHAPTER 1
Overview

This chapter provides an overview of operational best practices that apply to all EMC® Avamar® single-node and multi-node servers.

Topics in this chapter include:

Guide organization

The following table shows the best practices guide's organization.

Table 2 Best practices guide's organization

Subject matter	Chapters
Core Avamar system functions	• Chapter 2, "Designing Avamar to Maximize System Availability" • Chapter 3, "Managing Capacity" • Chapter 4, "Scheduling" • Chapter 5, "Defining Domains, Groups, and Policies" • Chapter 6, "Daily Monitoring of Backup Infrastructure and Operations"
Tuning the Avamar system	• Chapter 7, "Tuning Performance" • Chapter 8, "Understanding DPN Summary Reports"
Avamar Desktop/Laptop clients	• Chapter 9, "Protecting Avamar Desktop/Laptop Clients"
Other Avamar administration functions	• Chapter 10, "Other Avamar Administration Best Practices"
Data Domain systems	• Chapter 11, "Using Data Domain Systems"

The introduction to each chapter indicates which of the following Avamar server lifecycle phases it covers. The following table describes the phases.

Table 3 Lifecycle phases

Lifecycle phase	Description
Planning and design	Topology and architecture options, risks and limitations, and any other planning and design issues that must be considered before implementing the design.
Implementation	Installation options and directions for testing Avamar components after the installation is complete.
Daily operations	Regular management of Avamar server capacity, performance optimization of backups and replication, and daily monitoring of the Avamar infrastructure and operations.

Most important operational best practices

Here are the most important best practices to understand and follow:

- Check EMC Online Support (https://support.EMC.com/products) for the most current version of the *EMC Avamar Operational Best Practices*.
- Set up a correct daily operational schedule for the Avamar server.
- Proactively assess and correct systematic issues.
- Deploy the Avamar server with reliable, high-performance RAID arrays for back-end storage.
- Deploy a uninterruptible power supply (UPS) for the Avamar server hardware to protect against data loss caused by unplanned power outages.
- Understand how to monitor and manage the storage capacity of the Avamar server on a daily basis.
- Minimize the number of groups used to back up clients. Schedule backups during the server's backup window so that they do not overlap with daily maintenance jobs.
- Monitor the Avamar server on a daily basis. Interpret all system warnings and errors.
- Investigate all failed backups, missing clients, and backups that completed with exceptions.
- Protect the Avamar server from the Internet by providing full firewall protection.
- Change all factory default passwords except the passwords for the backuponly, restoreonly, and backuprestore software application users.
- Ensure every administrator logs in to the Avamar server with a unique username.
- Enable the Email Home capability.
- Check the network bandwidth before adding a Data Domain system to an Avamar configuration.

The chapters that follow provide more details on these best practices and others.

CHAPTER 2
Designing Avamar to Maximize System Availability

This planning and design chapter includes a description of Avamar architecture, details on planning, considerations for design, recommendations for approaches and practices, and notes on data collection and documentation.

Topics in this chapter include:

Avamar architecture

To ensure the long-term reliability, availability, and supportability of the Avamar server, you must design it carefully.

Several processes run on the Avamar server nodes. Key processes include:

- Avamar Administrator server and the Avamar Enterprise Manager server on the utility node.
- Avamar data server on all active storage nodes.

The Avamar data server is also known as GSAN (Global Storage Area Network).

The Avamar data server stores, processes, and manages the variable-sized chunks that the client sends during a backup. An average size chunk is about 10 KB depending on the customer data. Through the patented deduplication technology, only unique data chunks are sent to the Avamar data server.

Stripes

The term "stripe" refers to the container an Avamar data server uses to manage the data in the system. Stripes are files of various sizes that are based on the kind of stripe.

Each stripe has a unique name, and the Avamar server can identify and access a stripe by name only. The following table describes four types of stripes.

Table 4 Types of stripes

Stripe	Description
Atomic data	Contains data that originates on the customer system and is read during a backup.
Composite	Contains references to other composite or atomic stripes, and provides the means to build trees that can arbitrarily represent large amounts of data. References are SHA-1 hashes.
Index	Maps a hash to the stripe that contains corresponding data. This is the essence of a "content addressed" store.
Parity	Provides simple XOR parity that can be used to reconstruct data when a failure occurs. If RAIN is used, every stripe belongs to a parity group that protects it. A protected stripe is called a "safe" stripe.

Avamar data server functions

The Avamar data server is a high-transaction-rate database-like application that is optimized to store and manage billions of variable-sized objects in parallel across all active storage nodes.

The Avamar server performs several functions throughout each day. The following table describes the major operational functions.

Table 5 Avamar server operational functions

Function	Description
Backup	Supports the backup operation by receiving, processing, and storing the backup data that Avamar clients send to it. During this process, the Avamar server interacts with the client to ensure that only unique data chunks are sent from the client to the server.
Restore	Restores the data stored on the Avamar server to the Avamar client.
Checkpoint	Creates consistent point-in-time images (checkpoints) every day. Checkpoints are used as rollback points to recover from various issues, such as sudden power loss.
hfscheck	Validates one of the checkpoints every day through a process called hfscheck.
Garbage collection	Deletes the orphaned chunks of data that are no longer referenced within any backups stored on the system.
Replication	Supports daily replication of the backups.
Precrunching	Prepares stripes throughout the day to be reused during backup. During this process, the server selects the emptiest stripes, those that contain more empty space than the data partitions (by percentage), and defragments them.This precrunching process leaves contiguous space for new data.

The Avamar server requires adequate CPU, memory, and I/O resources to perform these functions throughout the day. Avamar performs extensive qualification testing of all approved platforms to ensure that the resources available are adequate to meet long-term reliability, availability, and supportability requirements.

RAID, RAIN, replication, and checkpoints

The Avamar system provides up to four levels of systematic fault tolerance: RAID, RAIN, replication, and checkpoints.

Redundant Array of Independent Disks (RAID)

All standard Avamar server node configurations use RAID to protect the system from disk failures. RAID provides the capability to hot swap the hard disk drives that have been the highest failure rate hardware items in Avamar servers.

Failed drives impact I/O performance and affect Avamar server performance and reliability. Further, RAID rebuilds can significantly reduce the I/O performance, which can adversely impact the performance and reliability of the Avamar server.

Best practices

- If the hardware is purchased separately by the customer, the customer must configure the disk arrays on each node by using RAID.
- If the hardware is purchased separately by the customer, the customer must configure RAID rebuild as a low priority.
- If the customer purchases the hardware separately, the customer must regularly monitor and address hardware issues promptly.

Redundant Array of Independent Nodes (RAIN)

RAIN provides the means for the Avamar server to continue to operate even when a node fails. If a node fails, RAIN is used to reconstruct the data on a replacement node. In addition to providing failsafe redundancy, RAIN is used when rebalancing the capacity across the nodes after you have expanded the Avamar server (added nodes). This is a critical element to being able to manage the capacity of the system as the amount of data added to the system continues to increase. Except for two-node systems, RAIN protection is enabled in multi-node Avamar servers. Single-node servers do not use RAIN.

Best practices

- Always enable RAIN for any configuration other than single-node servers. Minimum RAIN configuration is a 1x3 system (three active storage nodes plus a utility node and optionally, a spare node).

  NOTICE

  Spare nodes are optional in ADS Gen4 systems that run Avamar 6.1.

  Double-disk failures on a node, or a complete RAID controller failure can occur. Either of these failures can corrupt the data on a node. Without RAIN, the only recourse is to reinitialize the entire system and replicate the data back from the replication target.

  NOTICE

  New installs of 1x2 systems are not supported for ADS Gen4 with Avamar 6.1.

- When deploying single-node servers, you must replicate the data on them to ensure that the data is protected. Non-RAIN servers have no data redundancy and any loss of data requires that the system be re-initialized.
- Limit initial configurations to 12 to 14 active storage nodes so that nodes can be added later if needed to recover from high-capacity utilization situations.

Replication

The Avamar system can efficiently replicate data from one Avamar server to another on a scheduled basis. This ensures complete data recovery if the primary backup Avamar server is lost.

Replication is useful for more than recovering a single client. Replication moves data to another system that can be used for data recovery in the event of an unexpected incident. Replication is, by far, the most reliable form of redundancy that the system can offer

because it creates a logical copy of the data from the replication source to the destination. Replication does not create a physical copy of the blocks of data. Any corruptions, whether due to hardware or software, are far less likely to be propagated from one Avamar server to another. In addition, multiple checks of the data occur during replication to ensure that only uncorrupted data is replicated to the replication target.

Therefore, if maximizing the availability of the backup server for backups and restores is important, you should set up a replication system as quickly as possible.

Best practices

- Protect the data on the Avamar server by replicating the data to another Avamar server.
- Use default standard replication, also known as "root-to-REPLICATE" replication, to do the following:
  - Provide the flexibility to configure replicated grids in a wide variety of ways
  - Have full visibility into all the backups that have been replicated from one Avamar grid to another

  Standard replication also supports the ability to replicate the contents of many replication source grids to a single large replication destination (many-to-one), or to cross-replicate the contents of a couple of grids to each other. At any time, you can browse the contents of the /REPLICATE domain on the replication destination and see all the backups that have been replicated for each account.
- Ensure that available network bandwidth is adequate to replicate all of the daily changed data within a four-hour window so that the system can accommodate peaks of up to eight hours per day. The replicator can use 60% to 80% of the total available bandwidth when WAN bandwidth is the performance bottleneck. The *EMC Avamar Administration Guide* contains more information about setting up replication to best use the system bandwidth.
- When defining daily replication, avoid using the **--include** option. This option should be used to perform only selective replication under certain conditions. Specifying clients that must be replicated by listing them with the **--include** option is prone to error. Every time you add a new client to the active Avamar server, the client data is not replicated unless you edit the repl_cron.cfg file to add a new **--include** option for that client.
- Use the **–exclude** option only if you decide that a high change-rate or low-priority client can be selectively excluded from the nightly replication.
- When configuring replication, always set the **--retention-type** option to replicate all retention types (none, daily, weekly, monthly, and yearly).

  If you leave out retention type "none" from the replication, then hourly Avamar Administrator server backups or the Enterprise Manager backups are not replicated. These system backups are required to perform a full disaster recovery of the replication source Avamar grid.

Checkpoints

Checkpoints provide redundancy across time. Checkpoints enable you to recover from operational issues. For example:

- Attempting to back up a client that is too large to fit in the available remaining capacity.
- Accidentally deleting a client and all of the associated backups.

In addition, checkpoints enable you to recover from certain kinds of corruption by rolling back to the last validated checkpoint.

Although checkpoints are an effective way to revert the system back to an earlier point in time, checkpoints are like all other forms of redundancy and therefore, require disk space. The more checkpoints you retain, the larger the checkpoint overhead.

Best practice

Leave the checkpoint retention policy at the default values. The default is set to retain the last two checkpoints, whenever created, and the last validated checkpoint.

NOTICE

During certain support actions, EMC Customer Support might temporarily change the checkpoint retention policy to ensure that certain critical checkpoints are retained during the support action. After the support action is completed, restore the checkpoint retention policy to the default setting.

Backing up clients in remote offices

When you back up clients in a remote office, consider the following options:

- Option 1 — Is it better to back up remote office clients to a small Avamar server that is located in a remote office (remote Avamar backup server), and replicate data to a large centralized Avamar server (centralized replication destination)?
- Option 2 — Is it better to back up those clients directly to a large centralized Avamar server, and replicate data to another large centralized Avamar server (centralized replication destination)?

The following table provides factors to consider when deciding a remote backup strategy.

Table 6 Remote office backups

Factor	Description
Recovery time objective (RTO)	When the Avamar system performs a restore, all data that must be restored is compressed and sent from the Avamar server to the Avamar client, where it is uncompressed. However, no deduplication is performed on the restored data. The primary advantage of backing up data to a remote Avamar backup server (Option 1) is that the restore can be done directly from that server across the local area network to the client. This advantage is important if an RTO requirement must be satisfied.
Server administration	The amount of administration and support required is roughly proportional to the number of Avamar servers deployed in an environment. For example, 10 single-node servers deployed as remote Avamar backup servers require considerably more administration and support than a single 1x8+1 multi-node configuration of 10 nodes (eight active storage nodes, one utility node, and one spare) that functions as a centralized Avamar backup server.
IT resources	Even if a remote Avamar backup server is deployed at a remote office, adequate IT resources for performing disaster recovery restores might not be available at the remote office. In this case, Option 2 might be appropriate, in which case, a centralized IT staff can perform disaster recovery restores to replacement hardware at the central site and then ship the fully-configured replacement client to the remote site.
Exchange Server	If a Microsoft Exchange Server is located in the remote office, and depending on the bandwidth, the only practical way to restore the large amount of data typically associated with this kind of server's storage group or database might be Option 1.
Large multi-node servers	If large multi-node servers are required to back up all data in a remote office, there might not be a significant reduction in the number of Avamar servers that are deployed, even if Option 1 is selected. In this case, the cost of deploying, managing, and supporting the Avamar servers is roughly the same, regardless of whether these Avamar servers are deployed as remote Avamar backup servers or as centralized Avamar backup servers.

If the deployment environment's WAN throughput is a bottleneck, the time required to perform nightly replication in Option 1 is roughly the same as the time required to perform backups in Option 2. The trade-off then becomes RTO compared to the additional cost of deploying, managing, and supporting multiple Avamar server instances.

Best practice

Unless you cannot meet the RTO, design the system so that clients first back up directly to a large, active, and centralized Avamar server. Then replicate the data to another large centralized Avamar server.

CHAPTER 3
Managing Capacity

This daily operations chapter focuses on the kinds of activities and behaviors one can reasonably expect during the first several weeks in the Avamar server lifecycle.

Topics in this chapter include:

Impact of storage capacity when deploying a new Avamar system

When a new Avamar system is initially deployed, the server typically fills rapidly for the first few weeks. This is because nearly every client that is backed up contains unique data. The Avamar commonality feature is best leveraged when other similar clients have been backed up, or the same clients have been backed up at least once.

After the initial backup, the Avamar system backs up less unique data during subsequent backups. When initial backups are complete and the maximum retention periods are exceeded, it is possible to consider and measure the ability of the system to store about as much new data each day as it frees during the maintenance windows.

This ability is known as achieving steady state capacity utilization.

Successfully achieving steady state capacity utilization is especially important for the single-node and non-RAIN server because these are fixed-capacity systems.

Impact of storage capacity on system performance

When managing an Avamar server, you can significantly improve the long-term reliability, availability, and manageability of the Avamar server if you do either of the following:

- Minimize the average daily data change rate of the clients that are being protected. The *EMC Avamar Administration Guide* contains details about daily data change rate.
- Reduce the per-node capacity that is utilized within the Avamar server by doing one or more of the following:
  - Reducing backup retentions
  - Ensuring daily maintenance jobs run regularly
  - Adding more nodes to the Avamar server

NOTICE

Many of the operational best practices described throughout this document are targeted at understanding the average daily change rate or managing the per-node capacity.

Definitions of Avamar server capacities

Storage subsystem (GSAN) capacity is the total amount of commonality factored data and RAIN parity data (net after garbage collect) on each data partition of the server node. The GSAN process measures and reports this amount. The administrator of the Avamar server can control this reported capacity:

- First, by changing the dataset definitions, retention policies, or even the clients that are backed up to this server.
- Secondly, by ensuring that a garbage collect operation runs regularly to remove expired data.

Operating system capacity is the total amount of data in each data partition, as measured by the operating system. This amount is not particularly useful to an external observer because the server manages disk space itself.

Avamar capacity thresholds

The GSAN changes behavior as the various capacities increase. The following table describes the behavior of key capacity thresholds.

Table 7 Capacity thresholds

Threshold	Default values	Capacity used for comparison	Behavior
Capacity warning	80% of read-only threshold	GSAN	The Management Console Server issues a warning event when the GSAN capacity exceeds 80% of the read-only limit.
Healthcheck limit	95% of read-only threshold	GSAN	If the GSAN capacity reaches the healthcheck limit, existing backups are allowed to complete, but all new backup activity is suspended. A notification is sent in the form of a pop-up alert when you log in to Avamar Administrator. The system event must be acknowledged before future backup activity can resume.
Server read-only limit	100% of read-only threshold, which is set to a prespecified percentage of available hard drive capacity	GSAN	If the GSAN capacity on any data partition on any node exceeds the read-only threshold, the Avamar server transitions to read-only state to prevent new data from being added to the server. This value is reported as server utilization on the Server Management tab (Avamar Administrator › Server › Server Management). The reported value represents the average utilization relative to the read-only threshold.
System too full to perform garbage collect	85% of available hard drive capacity	Internal GSAN calculation	If the GSAN determines that the space available on any data partition on any node exceeds the **disknogc** configuration threshold, a garbage collect operation does not run. The operation fails with the error message MSG_ERR_DISKFULL.

Impact of capacity on various operations

Another key consideration when managing the Avamar server is that many of the maintenance operations take longer to complete as the amount of data stored in the Avamar server increases. This behavior is most notable in the garbage collection (GC) activity.

Any variations with incoming data or daily maintenance routines can lead to the system becoming read-only or to additional maintenance routines failing.

Best practices for capacity management

- Understand how to monitor and manage the storage capacity of the Avamar server on a daily basis.
- Limit storage capacity usage to 80% of the available GSAN capacity.
- Monitor all variations with incoming data to prevent the system from becoming read-only.
- Monitor all variations with maintenance jobs to prevent these jobs from failing.

Proactive steps to manage capacity

You receive a warning when the GSAN capacity exceeds 80% of the read-only threshold. If this occurs:

1. Stop adding new clients to the system.
2. Reassess retention policies to see if you can decrease the retention, and therefore, reduce the capacity use.
3. Investigate the possibility that backups are preventing a garbage collect operation from starting:
   a. Use **dumpmaintlogs --types=gc** to view logs for the garbage collection operation.

      Look for either of the following error messages in the garbage collection log:

      ```
      MSG_ERR_BACKUPSINPROGRESS
      garbage collection skipped because backups in progress
      ```

   b. Use **capacity.sh** to:
      - Assess the data change rate in the environment.
      - Assess the garbage collect effectiveness.
      - Ensure that the system is running in steady state.
      - Identify the three highest change rate clients.

      The **capacity.sh** command displays output similar to the following:

```
admin@avamar-1:~/>: capacity.sh
Date       New Data    #BU    Removed    #GC   Net Change
---------- ---------- ----- ---------- ----- ----------
2012-02-15  12044 mb  9     0 mb              12044 mb
2012-02-16  5365 mb   9     -1038 mb   1      4327 mb
2012-02-17 5997 mb    9     -989 mb    1     5008 mb
2012-02-18 3675 mb    10    -87211 mb  1     -83535 mb
2012-02-19 10939 mb   8     -63369 mb  1     -52429 mb
2012-02-20 3036 mb    8     -32857 mb  1     -29820 mb
2012-02-21 2461 mb    8     -41344 mb  1     -38782 mb
2012-02-22 5150 mb    9     -6459 mb   1     -1308 mb
2012-02-23 24735 mb   9      -1044 mb  1     23691 mb
2012-02-24 5217 mb    10     -12653 mb 1      -7435 mb
2012-02-25 8238 mb    9     -46075 mb  1     -37836 mb
2012-02-26 5372 mb    10    -58718 mb  1     -53345 mb
2012-02-27 3322 mb    7     -10372 mb  1     -7049 mb
2012-02-28 3418 mb    8     -3140 mb   1     278 mb
2012-03-01 4529 mb    8     -9884 mb   1     -5354 mb
2012-03-02 18807 mb   8     25770 mb   1     6062 mb
2012-03-03 4590 mb    8     -41766 mb  1     -37175 mb
2012-03-04 3676 mb    7     -16630 mb  1     -12953 mb
2012-03-05 4759 mb    6     -13746 mb  1     -8986 mb
2012-03-06 4106 mb    6     -74316 mb  1     -70209 mb
2012-03-07 2186 mb    7     -82198 mb  1     -80011 mb
```

```
2012-03-08  6558 mb     8      -22672 mb   1      -16113 mb
2012-03-09  7425 mb     4      -14921 mb   1      -7495 mb
2012-03-10  10755 mb    4      -11171 mb   1      -415 mb
2012-03-11  9744 mb     4      -22178 mb   1      -12433 mb
2012-03-12  15586 mb    3      -19787 mb   1      -4200 mb
2012-03-13  277 mb      3      -14355 mb   1      -14077 mb
2012-03-14  4714 mb     4      -26187 mb   1      -21472 mb
2012-03-15  9995 mb     3      -24524 mb   1      -14528 mb
2012-03-16  16267 mb    3      -42623 mb   1      -26355 mb
2012-03-17  12736 mb    6      -51306 mb   1      -38569 mb
----------  ----------  -----  ----------  -----  ----------
Average     7606 mb            -28364 mb          -20758 mb

 Top 5 High Change Clients:
 --------------------------
Total for all clients 235798 mb  100.0%
MyServer-1            151950 mb   64.4%  0.003%
MyServer-2            74982 mb    31.8%  0.005%
MyServer-3            3249 mb     1.4%   0.003%
MyServer-4            2008 mb     0.9%   0.000%
MyServer-5            1722 mb     0.7%   0.016%
```

4. Extend the blackout window to improve garbage collection performance.

   NOTICE

   You can decrease the GSAN capacity by deleting or expiring backups and running garbage collection.

   Deleting backups or clients (and therefore, all the backups associated with those clients) does not free space until garbage collect has run, possibly several times. Garbage collect finds and deletes the unique data associated with these backups.

Steps to recover from capacity issues

Once the Avamar server capacity exceeds the warning threshold and approaches the diskreadonly limit, take one or more of the following actions:

- Follow the steps described in "Proactive steps to manage capacity" on page 30.
- If the Avamar server reaches the healthcheck limit, all new backup activity is suspended until you acknowledge this event.
- If the Avamar server transitions to a read-only state, you must contact EMC Customer Support.
- In an extreme case, consider replicating the data to another server temporarily, and then replicating the data back after reinitializing the server. Because replication creates a logical copy of the data, this compacts all the data onto fewer stripes.
- If the Avamar server is a multi-node server that utilizes RAIN, consider adding nodes and rebalancing the capacity. If the server has eight or more active storage nodes, add two nodes at a time, rather than adding just one node, to noticeably reduce the capacity per node.

- Extend the blackout window to slowly reduce the GSAN capacity utilization of the Avamar grid.

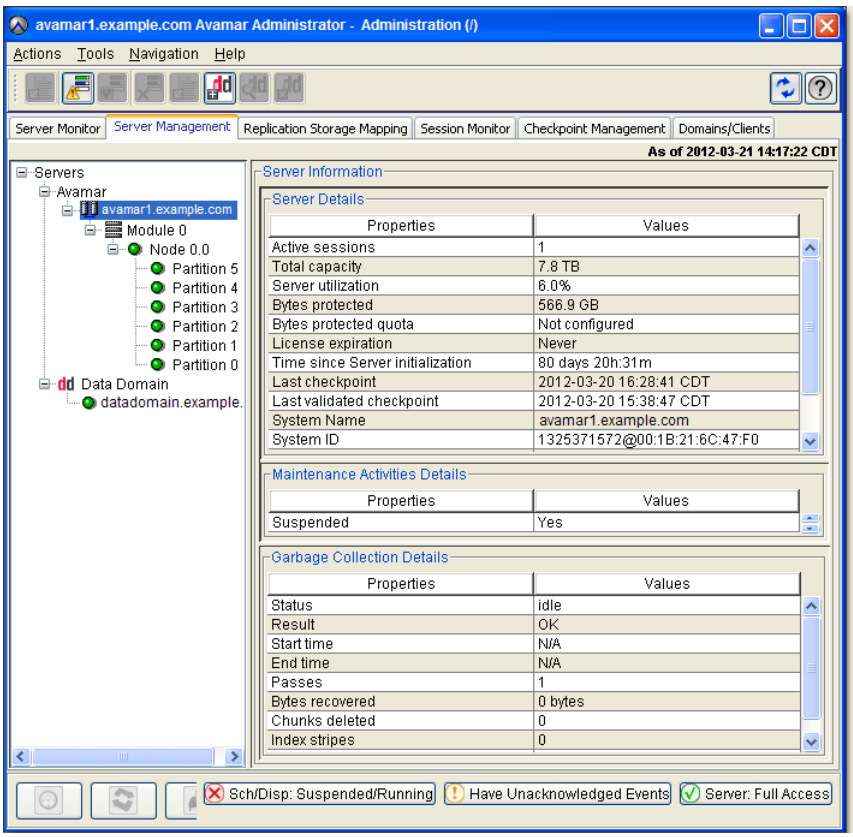
  This process requires some analysis to properly size the blackout window. After you achieve the correct blackout window to reduce the GSAN capacity to approximately 1% per week, the system then runs in steady state for an extended period of time.

Steady state system

Typically, an Avamar system achieves steady state shortly after the longest retention period for the backups. For example, if you retain all daily backups for 30 days and all monthly backups for three months, the system begins to operate in steady state about 3 1/2 to 4 months after the last client has been added to the system. A slight delay occurs before achieving steady state because the garbage collect process requires several passes before it reaches the bottom of the file system tree. Garbage collect finds orphaned chunks in the upper levels first before removing orphaned data in the lower levels of the file system.

After the system has achieved steady state, do the following:

1. Ensure that activities are scheduled so that all backups and maintenance tasks run successfully.
2. Verify that Server utilization, as shown in the following figure, is at or below 80%:



CHAPTER 4
Scheduling

This planning and design chapter focuses on scheduling activities, important steps in designing, and setting up a new Avamar system.

Topics in this chapter include:

Avamar client details

Avamar client agents are applications that run natively on the client systems. The Avamar client software comprises at least two executable programs: **avagent** and **avtar**.

The **avagent** program runs as a service on the client and establishes and maintains communication with the Avamar Administrator server.

When the Avamar Administrator server queues a work order (for example, a backup), the Avamar Administrator server pages the client **avagent**. If the client is nonpageable (the Avamar Administrator server cannot establish a connection with the client), the client **avagent** polls the Avamar Administrator server at a regular interval to check for a work order.

When the Avamar Administrator server queues a work order for the client, the client **avagent** retrieves the work order.

The **avagent** program runs the **avtar** program with the parameters specified in the work order. The **avtar** program executes the backup based on the set of parameters related to the backup task. The **avtar** program performs a backup by making a connection to the Avamar server over the LAN or a remote connection over the WAN. TCP/IP is the base protocol used to make the connection.

Restores are executed in a similar manner to backups. A restore work order is created containing the parameters necessary to complete a restore of all or a subset of the files of a specific backup.

Restrictions and limitations

The following table lists known restrictions and limitations to consider during planning and design. In addition to this table, review the known limitations in the *EMC Avamar Release Notes*.

Table 8 Known restrictions and limitations for planning and designing the Avamar system

Restrictions and limitations	Impact
Recovery time objective (RTO)	RTO involves processes, communication service levels, regular testing, and people. The time to restore data is only one of several critical components needed to achieve a given RTO. Also, the RTO for any individual client is typically limited by the performance capabilities of the client or network, and not the capability of the Avamar server to restore the data.
5 to 10 million files per Avamar client	Backup scheduling could be impacted when an Avamar client has several million files. The actual amount of time required to back up the Avamar client depends on the following: • Total number of files on that client • Hardware performance characteristics of the client The Avamar system can accommodate file system clients with significantly more than 10 million files, but this might require additional configuration or tuning.
500 GB to 2 TB of database data per Avamar client	Backup scheduling could be impacted when an Avamar client has large databases that need to be backed up. The actual amount of time required to back up the Avamar client depends on the following: • Total amount of database data on the client • Hardware performance characteristics of the client The Avamar system can accommodate database clients with significantly more than 2 TB of database data, but this might require additional configuration or tuning. **Notice:** If the amount of scanned data, for example, new data in the file system plus the size of all the databases combined, is greater than 1TB, use cache prefixes.
2 to 5 TB of fileserver data per Avamar client	Backup scheduling could be impacted when an Avamar client is a fileserver that is protecting a large amount of data. The actual amount of time required to back up the Avamar client depends on the following: • Total number of files on the client • Hardware performance characteristics of the client The Avamar system can accommodate clients with significantly more than 5 TB of file system data, but this might require additional configuration or tuning.

Best practices

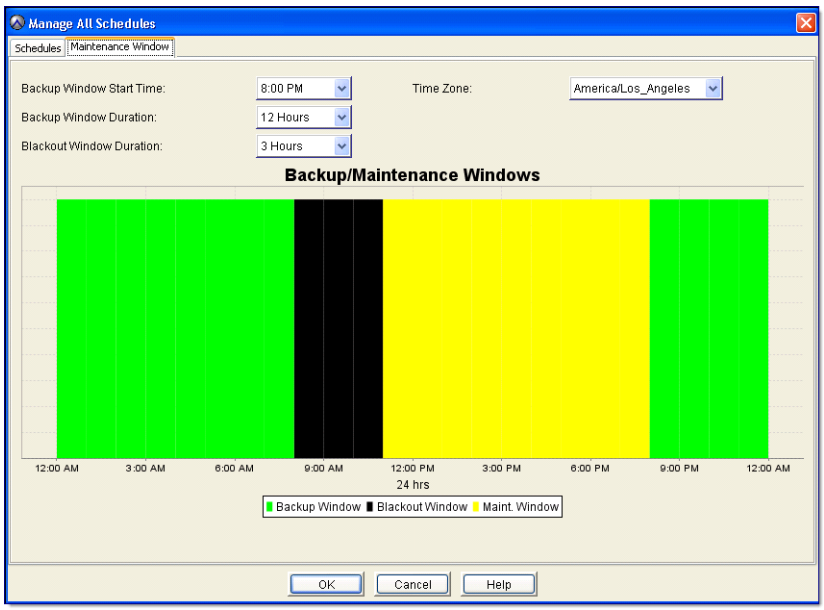
- Carefully review the client support matrix with the presales technical engineer.
  - Ensure that Avamar software supports the clients and applications you want to protect.
  - Verify that Avamar software supports all details of the deployment, such as revisions, clusters, third-party plug-ins, and add-ons.
- Consider storing certain data types (very large databases with very high change rates) on Data Domain® systems. Avamar clients that support backup and restore to and from Data Domain systems include:
  - Avamar for VMware
  - Avamar Plug-in for Exchange VSS
  - Avamar Plug-in for Oracle
  - Avamar Plug-in for SAP with Oracle
  - Avamar Plug-in for SQL Server
  - Avamar Plug-in for SharePoint VSS
  - Avamar Plug-in for Sybase ASE
- Consider multi-streaming backups to improve performance for:
  - Avamar NDMP Accelerator
  - Avamar Plug-in for Exchange VSS
  - Avamar Plug-in for SharePoint VSS
  - Avamar Plug-in for Oracle

The *EMC Avamar and Data Domain Integration Guide* provides more information about the use of Data Domain systems as storage for Avamar backups.

Scheduling activities during the course of a day

Typically, the longest running activities throughout the day are hfscheck, backups, and replication. During the planning and design stage, proper scheduling of activities throughout the day is one of the most important factors that influences the system reliability, availability, and supportability.

Each 24-hour day is divided into three operational windows, during which various system activities are performed. The following figure shows the default backup, blackout, and maintenance windows:



Backup window

The backup window is the portion of each day reserved for performing normal scheduled backups.

- Operational impact — No maintenance activities are performed during the backup window.
- Default settings — The default backup window begins at 8 p.m. local server time and continues uninterrupted for 12 hours until 8 a.m. the following morning.
- Customization — You can customize the backup window start time and duration to meet specific site requirements.

Blackout window

The blackout window is the portion of each day reserved for performing server maintenance activities, such as checkpoints and garbage collection, which require unrestricted access to the server.

- Operational impact — No backup or administrative activities are allowed during the blackout window. You can perform restores.
- Default settings — The default blackout window begins at 8 a.m. local server time and continues uninterrupted for three hours until 11 a.m. that same morning.
- Customization — You can customize the blackout window duration to meet specific site requirements.

Any changes to blackout window duration also affect maintenance window duration. For example, changing the blackout window duration from three hours to two hours, extends the maintenance window duration one hour because it begins one hour earlier. The backup window is not affected.

NOTICE

If the blackout window is too short, garbage collection might not have enough time to run. If you shorten the blackout window, be sure to closely monitor server-capacity utilization and forecasting regularly (at least weekly) to ensure that adequate garbage collection is taking place.

Maintenance window

The maintenance window is the portion of each day reserved for performing routine server maintenance activities such as checkpoint validation.

- Operational impact — There might be brief periods when backup or administrative activities are not allowed.

  Although backups can be initiated during the maintenance window, doing so impacts both the backup and maintenance activities. For this reason, minimize any backup or administrative activities during the maintenance window. You can, however, perform restores.

  Although hfscheck and backups can overlap, doing so might result in I/O resource contention, which can cause both activities to take longer to complete and possibly even to fail.

- Default settings — The default maintenance window begins at 11 a.m. local server time and continues uninterrupted for nine hours until 8 p.m. that evening.

- Customization — Although the maintenance window is not directly customizable, its start time and duration is derived from backup and blackout window settings.

  The maintenance window starts immediately after the blackout window and continues until the backup window start time.

Replication

When replicating data from the local server to a replication target:

- All other maintenance jobs can start.
- All backup work orders are queued immediately.

When receiving replicated data from the replication source:

- The garbage collect operation cannot start. All other maintenance jobs, such as checkpoint and hfscheck, can start.
- All backup work orders are queued immediately.

If replication is bottlenecked by WAN throughput, overlapping replication with backup activities is unlikely to affect the amount of time required to perform replication and has only a slight impact on backup performance.

Two reasons some clients take a long time to back up:

- The backup throughput for the clients is limited by the WAN bandwidth. In this case, because the activity level on the Avamar server is relatively low, it is acceptable to overlap replication with the end of the backup window.
- The backup window for these clients is long because the clients are large. The time required to perform the backup is directly proportional to the type and amount of data on the clients being backed up.

Best practices

- Minimize the number of groups used to back up clients, and schedule backups during the server's backup window so that they do not overlap with daily maintenance jobs.
- Use the default maintenance window schedule and do not deviate from this schedule unless absolutely necessary.
- If there are a large number of clients that must be backed up outside of the server's backup window, set up a separate Avamar server that backs up those clients.
- To back up clients from around the globe, consider setting up the Avamar servers as follows:
  - One server to back up the clients in the Americas.
  - Second server to back up the clients in Europe, Middle East, and Africa (EMEA).
  - Third Avamar server to back up the clients in Asia Pacific and Japan (APJ).
- Limit the amount of time required to perform checkpoint, hfscheck, and so forth by carefully managing the capacity on the node. The most effective way to do this is to:
  - Limit the clients being backed up.
  - Reduce the retention policies.
  - Back up clients with lower daily change rates.
  - Ensure that the garbage collect operation runs every day.
- Limit the amount of time required to perform backups by monitoring:
  - Maximum number of files per client.
  - Maximum amount of database data per client.
  - Maximum amount of data per fileserver.

Typically, 80% to 90% of the clients complete daily backups within the first hour or two of the backup window. You can, therefore, consider scheduling replication to start two hours after the start of the backup window. The Avamar server is typically the bottleneck for backup operations only during the first one to two hours of the backup window. The remaining 10% to 20% of the clients might take several hours to complete backups, depending on the number of files or amount of data that needs to be backed up on these clients.

CHAPTER 5
Defining Domains, Groups, and Policies

Topics in this planning and design chapter include:

Management policy decisions

Initial backup management policy decisions must be made after you define the overall daily schedule.

Best practices for making management policy decisions

- What domains should be set up with designated domain administrators to take advantage of the hierarchical administration capability?
- What groups (which include dataset, backup schedule, and retention policies) should be created to back up clients effectively and manage the backups?
- When should backups be scheduled?
- How long should client backups be allowed to run?
- How should retention policies be set up to retain the backup data for the required period?

Defining domains

Domains are distinct zones within the Avamar server accounting system that are used to organize and segregate clients. The real power of domains is that they provide the ability for a domain-level administrator to manage clients and policies within that domain. This ability is known as hierarchical management.

You might consider segregating clients by domain for billing other internal organizations for backup services. Segregating clients by department or work group might be a convenient way to bill them.

If you are not going to use hierarchical management, register all clients in the /clients domain.

Best practice

Minimize the number of domains you create. Ideally, register all clients in the /clients domain.

Defining groups

A group defines the backup policy for the clients assigned to it, and includes the following three policy elements:

- Dataset policy (including the source data, exclusions, and inclusions)
- Backup schedule (or backup window)
- Retention policy

Best practices for defining groups

- Minimize the number of groups you define. Each dataset policy can include separate dataset definitions for various client plug-ins. For example, a single dataset policy can define independent datasets for Windows, Linux, Solaris, and other clients. You do not need to define separate groups to back up various kinds of operating system clients.
- Leave the default group disabled. By default, all new clients that have been activated with the Avamar server are automatically added to the default group. If you enable the default group, any clients activated with the Avamar server automatically are backed up according to the default group policy.
- To help manage the capacity on the Avamar server and to avoid being surprised by unexpected clients, leave the default group disabled.
- To back up any clients in the default group to the Avamar server, you can add the client to an enabled group and remove the client from the default group.

Defining datasets

In general, do not back up a large client by defining multiple subsets of data that run every night. Defining multiple subsets of data is a good practice in only three instances:

- When you want to define different retentions for different subsets of data on the client.
- When you are breaking up the dataset so that different subsets of the data are backed up on different days of the week.
- When you do not have enough memory to accommodate an appropriate sized file cache or hash cache for the entire dataset. "Tuning client caches to enhance performance" on page 57 provides more information.

Best practice for defining datasets

Minimize the number of datasets required.

Defining schedules and retention policies

The default schedule runs nightly during the server's backup window. Depending on the amount of data in the largest clients, this might not be enough time, and you might need to extend the server's backup window. Before extending the backup window, you must evaluate the time required for checkpoint, garbage collection, and hfscheck to determine that extra time is available after completing these daily activities.

Best practices for defining schedules

- Set appropriate expectations for how long the longest client backups should run every night, and validate that the long-running client backups meet the expectations.
- Minimize the number of clients that need to be backed up outside of the server's backup window.

  When setting up backup schedules, remember that mobile laptop clients might require a backup schedule that runs during the day when they are connected to the network. The system can handle a small number of exceptions. In this case, overlap the backup of this handful of exception clients with the server's maintenance window.

Best practices for setting up retention policies

- Use the advanced retention policy whenever possible. This helps to reduce the total amount of back-end storage consumed on the Avamar server. Typically, the following applies:
  - Weekly backups are equivalent, in the amount of unique data, to three daily backups.
  - Monthly backups are equivalent, in the amount of unique data, to six daily backups.

  For example, you can configure a retention policy to keep 30 days of daily backups and three months of monthly backups. The following figure shows the retention policy settings in the Edit Advanced Retention Policy dialog box:

  When this retention policy is used, the amount of client data stored on the Avamar is equivalent to the initial unique data plus 42 equivalent days of backups. You use less back-end capacity than the amount you would use if you stored three months of daily backups. (Three months of daily backup is equivalent to the initial unique data plus 91 equivalent days of backups.

  The *EMC Avamar Administration Guide* contains more information about advanced retention policies.

- Set the minimum retention period to at least 14 days.

  When you select the maximum retention period, the Avamar server does not retain the last unexpired backup. For a short retention period (7 days or less), closely monitor the backup operations to ensure that the last unexpired backup does not expire

before the system completes another backup. If all client backups expire before you correct the issue that prevented the client from completing a backup, the next backup is equivalent to an initial backup.

CHAPTER 6
Daily Monitoring of Backup Infrastructure and Operations

This daily operations chapter focuses on Avamar features and functions that generate notifications.

Topics in this chapter include:

Monitoring the Avamar system

The system reports all Avamar system activity and operational status as events to the administrator server. Examples of Avamar events include offline server nodes, failed or missing server maintenance jobs, and hardware issues.

Monitor the event notification system for warning and error events every day.

Best practice

Monitor the Avamar server daily and understand how to interpret all system warnings and errors.

The following table describes possible ways to monitor Avamar systems.

Table 9 Ways to monitor the Avamar system

Method	Description
syslog or SNMP event notification	If the network management infrastructure supports syslog or SNMP event notification, enable the syslog or SNMP event notification subsystem through Avamar Administrator. The *EMC Avamar Administration Guide* provides instructions for enabling syslog or SNMP notification.
Email notification system	You can set up email notification to: • Batch email notifications that are sent twice daily according to the default **notification schedule**. • Send emails as the selected events occur.
Avamar Enterprise Manager dashboard	To manually monitor the Avamar system, check the overall health of the Avamar backup infrastructure through the Avamar Enterprise Manager dashboard. Avamar server issues are immediately obvious because they are flagged by a red "X" under **Server Status**. **Notice:** Avamar Enterprise Manager can monitor Avamar 4.1 and 5.0 servers, in addition to 6.x servers.
Unacknowledged events	At least once a day, review and clear any Unacknowledged Events queued: 1. From Avamar Administrator, click the **Administration** launcher button 2. Select the **Event Management** tab 3. Select the **Unacknowledged Events** tab. **Notice:** On any Avamar Administrator view, click **Have Unacknowledged Events** to be redirected to the **Unacknowledged Events** page.
Avamar Administrator Event Monitor	At least once a day, review the event monitor: 1. From Avamar Administrator, click the **Administration** launcher button. 2. Select the **Event Management** tab. 3. Select the **Event Monitor** tab.

Monitoring the Avamar system backup operations

The system reports all Avamar system activity and operational status as events to the administrator server. You can then use client logs to investigate backup or restore issues. Monitor the event notification system for warning and error events related to backup operations every day.

Best practice

Monitor the Avamar Activity Monitor daily and understand how to interpret all activity warnings and errors.

Closely monitor daily backup activities

To create consistent backups, you must closely monitor daily backup activities.

The following factors may interfere with backups:

- Network issues

  These issues can cause backup failures.

- Client I/O errors

  These errors can prevent all files from being backed up (also known as Completed with Exceptions status).

- High client activity levels

  These levels can prevent all files from being backed up, or can prevent backups from completing within the backup window.

- Operator intervention

  Such as rebooting the client during the backup, or canceling the backup.

- Incomplete or incorrect dataset definitions
- Inadequate or incorrect retention periods

When you examine the activities, resolve all exceptions and failures.

The most obvious issues are the ones where the clients did not create a restorable backup.

The following table describes status messages typically associated with these failures.

Table 10 Client messages for client backups

Status message	Description
Failed	The client failed to perform the activity. The activity ended due to an error condition. Refer to the associated client log.
Canceled	The activity was cancelled, either from the client or from Avamar Administrator. Refer to the associated client log.
Dropped Session	The activity was successfully initiated but, because the Administrator server could not detect any progress, the activity was cancelled. The two most common causes are: • Somebody rebooted the client in the middle of the backup. • A network communication outage lasted longer than one hour. The Administrator server automatically queues a rework work order if the client backup fails due to a dropped session.
Timed Out - Start	The client did not start the activity in the scheduled window. This failure is most likely because the client is not on the network.
Timed Out - End	The client did not complete the activity in the scheduled window. This failure requires special attention because there is a lot of system activity with no restorable backup. Typically, if this is the case, subsequent backups continue to fail with the same status, unless some change is made, such as tuning the client caches.

A less obvious failure, but one that still requires attention, is a backup that reports the Completed with Exceptions status. In this case, the backup completed but with errors. Typically, the errors are due to open files that could not be backed up. Do not ignore this status. Some missing files, such as .PST files, can be significant.

You should examine all backups that completed successfully to ensure that the dataset definitions and retentions are appropriate.

The primary tool for monitoring daily backups is the Activity Monitor in Avamar Administrator. The Activity Monitor is described in the *EMC Avamar Administration Guide*.

Avamar Administrator can email reports to you that can help you monitor client backups that failed or completed with exceptions. The following table describes these reports.

Table 11 Avamar Administrator reports for client backups

Report	Description
Activities - Exceptions	This report lists all activities in the specified period that completed with exceptions.
Activities - Failed	This report lists all activities in the specified period that failed due to errors.
Clients - No Activities	This report lists all clients that did not have any activities in the specified period.

The *EMC Avamar Administration Guide* provides descriptions for these and other reports that are available.

Best practices for monitoring daily backup activities

- Monitor backups every day and investigate all failed backups, missing clients, and backups that completed with exceptions.
- Enable the advanced statistics report during all backups. This information is useful for addressing performance issues.
- Enable debugging messages when investigating backup or restore failures.
- Enable various activity report email messages, such as:
  - Activities - Exceptions
  - Activities - Failed
  - Clients - No Activities

Closely monitor nightly replication

Ensure that nightly replication successfully completes. The Avamar Administrator Activity Monitor displays a list of all clients that completed replication activities.

CHAPTER 7
Tuning Performance

This implementation and daily operations chapter focuses on Avamar system tuning activities.

Topics in this chapter include:

Avamar client caching overview

At the beginning of a backup, the Avamar client process, avtar, loads two cache files from the var directory into RAM.

- f_cache.dat
- p_cache.dat

The f_cache.dat cache file stores a 20-byte SHA-1 hash of the file attributes, and is used to quickly identify which files have previously been backed up to the Avamar server. The file cache is one of the main reasons subsequent Avamar backups that occur after the initial backup are generally very fast. Typically, when backing up file servers, the file cache screens out approximately 98% of the files. When backing up databases, however, the file cache is not effective because all the files in a database appear to be modified every day.

The p_cache.dat hash cache stores the hashes of the chunks and composites that have been sent to the Avamar server. The hash cache is used to quickly identify which chunks or composites have previously been backed up to the Avamar server. The hash cache is very important when backing up databases.

The client cache files help to reduce:

- The amount of time required to perform a backup.
- The processing load on the Avamar client and server.

The difference in backup performance with properly-sized caches is dramatic. Experience has shown that after sizing the client caches properly, backups that regularly required over 20 hours to complete suddenly complete in four hours every night.

A typical backup should take about one hour for every million files in a file server or about one hour for every 100 GB of data in a database server. If backups take more than 30% longer than these metrics, you should investigate whether client caches are properly tuned.

As with many operational best practices, you must carefully consider the trade-offs. Arbitrarily increasing the size of client caches consumes more memory. That could cause swapping and slow overall client performance. Ensure that you do the math required to size client caches appropriately.

Impact of caches on memory

Some stem administrators might be concerned about the amount of memory that the avtar process uses during a backup.

The avtar binary itself requires memory when performing a backup. The amount of memory consumed by the avtar process is generally in the range of 20 to 30 MB. This amount depends on which operating system the client is running, and also fluctuates during the backup depending on the structure of the files that are being backed up by avtar.

The file cache and hash cache can increase to maximum sizes of one-eighth and one-sixteenth of the total RAM in the system, respectively. For a client that has more than one-half GB of RAM, for example, the file and hash caches contribute more to the overall memory use than the rest of the avtar process. This is because both caches are read

completely into memory at the start of the **avtar** backup. Also, by default, the overall memory that client caches use is limited to approximately three-sixteenth of the physical RAM on the Avamar client.

Cache information in the avtar logs

The sizes of the file and hash caches are printed near the beginning of the **avtar** logs. For example, refer to the following output:

```
avtar Info <5573>: - Loaded cache file C:\Program
   Files\Avamar\var\f_cache.dat (5767712 bytes)
avtar Info <5573>: - Loaded cache file C:\Program
   Files\Avamar\var\p_cache.dat (25166368 bytes)
```

The file cache is 5.5 MB and the hash cache is 24 MB.

1 MB = 1048576 bytes

5767712 bytes/1048576 bytes = 5.5 MB

25166368 bytes/1048576 bytes = 24 MB

The end of the **avtar** log contains the following set of messages:

```
avtar Info <5587>: Updating cache files in C:\Program Files\Avamar\var
avtar Info <5069>: - Writing cache file C:\Program
   Files\Avamar\var\f_cache.dat
avtar Info <5546>: - Cache update complete C:\Program
   Files\Avamar\var\f_cache.dat (5.5MB of 63MB max)
avtar Stats <6151>: File cache: 131072 entries, added/updated 140,
   booted 0
avtar Info <5069>: - Writing cache file C:\Program
   Files\Avamar\var\p_cache.dat
avtar Info <5546>: - Cache update complete C:\Program
   Files\Avamar\var\p_cache.dat (24.0MB of 31MB max)
avtar Stats <6152>: Hash cache: 1048576 entries, added/updated 1091,
   booted 0
```

You can see that the file cache has room to increase in size:

```
Files\Avamar\var\f_cache.dat (5.5MB of 63MB max)
```

But the hash cache is at its maximum allowable size:

```
Files\Avamar\var\p_cache.dat (24.0MB of 31MB max)
```

If the file cache is undersized, the “booted” value is nonzero, and the log includes a warning that the cache is undersized. This information is very important because the size of the cache has a huge influence on the overall performance of the system.

Demand-page cache overview

The demand-page cache feature implements file cache management in a new way. When you enable demand-page cache, the **avtar** process loads portions or “pages” of f_cache2.dat into RAM instead of the entire file. The **avtar** process loads pages of the cache files into RAM on an “as needed basis.”

The cache files used by the demand-page feature are named:

- f_cache2.dat
- p_cache2.dat

Because the demand-page cache files have unique names, the use of the demand-page cache feature does interfere with the default cache files (f_cache.dat and p_cache.date). The demand-page cache feature is disabled by default. To enable the demand-page cache feature, use the **-paging-cache** option with the **avtar** command.

Demand-page file cache

The default file cache method uses approximately 1 GB of disk space and 1 GB of RAM to track approximately 10 million files when **avtar** is running. This method imposes limits on the maximum size of file systems that Avamar can backup. The demand-page cache feature removes this size limitation.

Although approximately 1 GB of disk space is necessary to track approximately 10 million files during a backup, the RAM consumption that **avtar** uses is substantially reduced. The demand-page cache feature enables RAM utilization to remain fairly flat at approximately 100 MB regardless of the number of files in the file system.

Demand-page hash cache

The default hash cache has restrictions similar to those of the file cache. For example, a 1-GB hash cache (which is considered to be very large) can track approximately 22 million hashes or approximately 260 GB of data. For situations in which the database size is 1 TB or more, the hash cache size is clearly insufficient. The demand-page cache feature enables a virtually unlimited number of hashes to be tracked in the cache within a fixed amount of RAM by:

- Using many fixed-sized caches arranged in pages rather than a single table. The default cache page size is 4 MB.
- Keeping a small number of cache pages active in memory at any given time. The default maximum number of pages is 20, which yields an approximate limit of 80 MB of RAM-resident cache pages: 20 x 4 M - 80 MB.
- Using a selection heuristic to load the “top candidate” cache pages in RAM. The heuristic relies on a RAM-resident table of selected hashes (coined “champions”) to statistically determine which cache pages should be loaded in RAM at any point in time. The default ratio of 1-in-100 consumes approximately 20 KB of RAM per cache page.
- Grouping entries within a page by disk locality to reduce page thrashing.

Example 1 Demand-page hash cache

A cache of 1000 pages requires approximately 100 MB of RAM: 20 MB (1000 x 20 KB) for the table of champions and 80 MB (20 x 4 MB) for cache pages.

This cache is the equivalent to a 4 GB monolithic cache (used in the default cache implementation) and can track approximately 40 million files.

Tuning client caches to enhance performance

This topic describes how the client caches work, and how you can tune the client caches appropriately to optimize backup performance. This topic also includes best practices to consider when setting up and installing clients.

The topics in this section are not relevant for backups to a Data Domain system.

Rules for tuning the maximum cache sizes

The most important rule is to ensure that the caches do not increase to the extent that the client ends up swapping because it has insufficient physical RAM to handle all the processes. Swapping is the movement of memory pages between RAM and disk.

Best practices for tuning the maximum cache sizes

- Never allow the total combined cache sizes to exceed one-fourth of the total available physical RAM.
- Set the maximum file and hash cache sizes to a fraction of the total available physical RAM. Specify the file and hash cache sizes by using negative integers.
- Limit the total cache sizes to approximately one-fourth of the physical RAM.
- Set one of the caches to be -5 (20%), and set the other cache to be -32 (3%).
- For example, for a large database client use the following settings:

  --filecachemax=-32

  --hashcachemax=-5
- If you use something other than the default cache sizes, include the customized maximum cache settings in the avtar.cmd file on the client.
- Sometimes the only choice may be to increase the amount of physical RAM on the client. You might also be able to back up the client by using multiple smaller datasets.
- If you need to limit the sizes of the caches below the optimum values:
  - For a typical file server, first allocate the required RAM to the file cache.
  - For a typical database client, first allocate the required RAM to the hash cache.

Tuning the file cache

If the file cache is deleted, unused, or undersized, every file that is not a hit in the file cache must be read, chunked, compressed, and hashed before the **avtar** process finds that the hashes were previously sent to the Avamar server. When a file is a hit in the file cache, the file is never read, which saves significant time and CPU.

By default, the file cache could consume up to one-eighth of the physical RAM on the Avamar client. For example, if the client has 4 GB of RAM, the file cache is limited to 4 GB divided by 8, or 512 MB maximum.

The file cache doubles in size each time it needs to increase. The current file cache sizes are in megabytes: 5.5 MB, 11 MB, 22 MB, 44 MB, 88 MB, 176 MB, 352 MB, 704 MB, and 1,408 MB.

Because the **avtar** program is a 32-bit application, the maximum file cache size that **avtar** can use is limited to less than 2 GB. In an example where a client has 4 GB of RAM, the maximum size of the file cache is 352 MB.

Each entry in a file cache comprises a 4-byte header plus two 20-byte SHA-1 hashes (44 bytes total):

- SHA-1 hash entry of the file attributes.

  The file attributes include: file name, file path, modification time, file size, owner, group, and permissions.

- SHA-1 hash entry for the hash of the actual file content, independent of the file attributes.

File cache rule

If the client comprises *N* million files, the file cache must be at least *N* million files x 44 million bytes/million files. This means that the file cache must be at least *N* x 44 MB, where *N* is the number of millions of files in the backup.

Example 2 File cache

When a client has 4 million files, the file cache must be at least 176 MB (4 x 44 MB). The file cache must be allowed to increase to 176 MB to accommodate all the files.

Best practice

The file cache must be a minimum of *N* x 44 MB, where *N* is the number of millions of files in the backup.

The file cache doubles in size each time it grows. To adequately size the file cache:

1. Set the **--filecachemax** value as follows:

   `--filecachemax = 2 x` *`N`* `x 44`

   where *N* is the number of millions of files in the backup.

2. Set the **--hashcachemax** to a small value, such as:

   `--hashcachemax=30`

Tuning the hash cache

If the **avtar** process finds that a hash of a chunk is not contained in the hash cache, it queries the Avamar server for the presence of the hash.

By default, the hash cache could consume up to one-sixteenth of the physical RAM on the Avamar client. Using the same client with 4 GB of RAM described in "Tuning the file cache" on page 57, the hash cache is limited to 4 GB/16, or 256 MB maximum.

The hash cache also doubles in size each time it needs to increase. The current hash cache sizes are in megabytes: 24 MB, 48 MB, 96 MB, 192 MB, 384 MB, 768 MB, and so forth. In this example where a client has 4 GB of RAM, the maximum size of the hash cache is 192 MB.

Each entry in a hash cache comprises a 4-byte header plus one SHA-1 hash per chunk or composite, which is the hash of the contents of the chunk or composite.

Hash cache rule

If the client comprises *Y* GB of database data, the hash cache must be at least *Y* GB/average chunk size x 24 million bytes/million chunks. Use 24 KB as the average chunk size for all backups. The hash cache must be at least *Y* MB, where *Y* is the number of gigabytes of database data in the backup.

Example 3 Hash cache

When a database client has 500 GB of database data, the hash cache must be allowed to increase to at least 500 MB. The hash cache must be allowed to increase to the next incremental size (768 MB) to accommodate the hashes for all the chunks in a database backup.

Best practice

The hash cache must be a minimum of *Y* MB, where *Y* is the size of the database being backed up in gigabytes.

The hash cache doubles in size each time it grows. To adequately size the hash cache, set the **--hashcachemax** value as follows:

--hashcachemax = 2 x *Y*

where *Y* is the size of the database to be backed up in gigabytes.

Using cacheprefix

When a client does not have enough memory to accommodate the cache files of appropriate sizes, you can back up the client and get the full benefit of appropriately-sized cache files by doing one of the following:

- Breaking the client file system into multiple smaller datasets.
- Ensuring that the maximum file and hash caches assign a unique **cacheprefix** attribute for each dataset.

Example 4 Using cacheprefix

Assume a client has 5.5 million files but only 1.5 GB of RAM. One volume has 2.5 million files and three other volumes have 1 million files each. You can break this client file system into four datasets. A volume with 2.5 million files requires a file cache of at least 110 MB (2.5 x 44 MB). The next increment that accommodates this is 176 MB.

You can define other datasets as shown in the following table.

Table 12 Segregating data into separate datasets

Drive	Attribute settings
C:\ drive (2.5 M files)	filecachemax=220 hashcachemax=30 cacheprefix=driveC
E:\ drive (1.0 M files)	filecachemax=88 hashcachemax=30 cacheprefix=driveE
F:\ drive (1.0 M files)	filecachemax=88 hashcachemax=30 cacheprefix=driveF
G:\ drive (1.0 M files)	filecachemax=88 hashcachemax=30 cacheprefix=driveG

Configure **cacheprefix** in the dataset by setting **Enter Attribute = cacheprefix** and **Enter Attribute Value = driveC**.

The following cache files are located in the Avamar /var directory on the client:

```
driveC_f_cache.dat
driveC_p_cache.dat
driveE_f_cache.dat
driveE_p_cache.dat
driveF_f_cache.dat
driveF_p_cache.dat
driveG_f_cache.dat
driveG_p_cache.dat
```

Ensure adequate disk space is available to accommodate the additional file and hash caches.

When specifying various **cacheprefix** values, ensure that new cache files are excluded from the backups. The cache files are large and have extremely high change rates.

Custom hash settings for Microsoft

For a Microsoft Exchange Server database backup, configure the maximum hash cache in the dataset by adding attributes and values:

1. From the **Edit Dataset** or **New Dataset** dialog box, select the **Options** tab.
2. Click **More**.
3. Type the attribute/value pair:
   a. In the **Enter Attribute** field, type **[avtar]hashcachemax**.
   b. In the **Enter Attribute Value** field, type **200**.

For a Microsoft SQL Server database backup, configure the maximum hash cache in the dataset by adding attributes and values:

1. From the **Edit Dataset** or **New Dataset** dialog box, select the **Options** tab.
2. Click **More**.
3. Type the attribute/value pair:
   a. In the **Enter Attribute** field, type **[avtar]hashcachemax**.
   b. In the **Enter Attribute Value** field, type **200**.

Tuning replicator

Work with EMC Customer Support Services to configure and tune the replicator. EMC Customer Support Services performs the following tasks:

1. Computes the bandwidth-delay-product (BDP) to determine whether the BDP is high enough to require customized tuning.
2. Verifies that the expected bandwidth is available between the replicator source utility node and the replicator destination storage nodes.
3. Tests the WAN link with the Avamar system components to verify that the Avamar system can utilize about 60% to 80% of the available bandwidth.
4. Sets up the appropriate replication parameters to optimize utilization of the available bandwidth.
5. Tests the replicator to verify its performance.

CHAPTER 8
Understanding DPN Summary Reports

This daily operations chapter describes the DPN Summary Reports.

Topics in this chapter include:

DPN Summary reports

Use DPN Summary reports to determine how well an Avamar system performs once it has achieved steady state. The DPN Summary report helps you to determine:

- Daily change rate for each individual client
- Daily change rate across the overall system
- High change rate clients that contribute the most to overall system change rate
- Amount of data that is protected per client and across the system
- Number of clients that are protected
- Abnormal client behavior such as:
  - Days with unusually high change rates
  - Unusually long backups
  - Frequent backup failures
- Amount of data that moved across the network with Avamar instead of incremental tape backups
- Benefit associated with the combined effect of commonality factoring and compression, when compared with commonality factoring or just compression

To access the DPN Summary report:

1. From Avamar Administrator, select **Tools** › **Manage Reports**.
2. Select **Activities - DPN Summary** from the navigation tree and click **Run**.
3. Select a date range and click **Retrieve**.

Example DPN Summary entry

Example 5 DPN Summary entry

This example uses the following excerpt from a DPN Summary report:

Host	StartValue	OS	StartTime
Avamar.corp.emc.com	1169366400	Windows Server 2008 R2 Standard...	"2012-01-21 08:07:39.36"

Root	Seconds	NumFiles	NumModFiles	ModReduced	ModNotSent
/EMC IT Windows Dataset	2,777	517,023	1,908	55,023,086,382	4,833,745,400

TotalBytes	PcntCommon	Overhead	WorkOrderID
451,940,965,688	99	60,055,981	EMC IT Windows Schedule – 1169348400105

ClientVer	Operation	Status	SessionID
6.1.100-276	Scheduled Backup	Activity completed successfully.	9116934840011000

The following table describes each DPN Summary column heading. "Background on backups" on page 67 provides more information.

Table 13 DPN Summary column descriptions (page 1 of 2)

Column heading	Description
Host	The client hostname as defined in DNS. • During backups, the hostname is the client that backs up data to the Avamar server. • During restores, the hostname is the client that receives the restored data. **Notice:** This is not the client that sourced the data.
StartValue	The UNIX start time of this activity. The UNIX start time is in the local time of the Avamar server.
OS	The client operating system.
StartTime	The date and time this activity was initiated. The StartTime is in Coordinated Universal Time (UTC)/Greenwich Mean Time (GMT).
Root	The name of the dataset that was used during the activity, if applicable.
Seconds	The duration, in seconds, of the activity.
NumFiles	The total number of files scanned during the activity less those files that were excluded through exclusion rules.
NumModFiles	The total number of modified files associated with the activity.
ModReduced	The amount of modified data that is reduced due to compression during commonality processing.
ModNotSent	The amount of bytes in modified files that do not have to be sent to the Avamar server because of subfile-level commonality factoring.
ModSent	The amount of new bytes sent to the Avamar server.
TotalBytes	"Summary of key DPN summary terms" on page 69 provides a description for TotalBytes.
PcntCommon	Commonality percentage during the activity.
Overhead	The number of bytes for COMPOSITEs and DIRELEMs used to store data. Overhead is the amount of nonfile data sent by the client to the server for the following items: • Indexing information • Requests from the client to the server for the presence of specific data chunks • ACLs • Directory information • Message headers On any relatively active file system, overhead is generally a very small percentage of the file data that is sent to the Avamar server.

Table 13 DPN Summary column descriptions (page 2 of 2)

Column heading	Description
WorkOrderID	The unique identifier for the following activities: • For scheduled backups, the work order ID is formatted as: *SCHEDULENAME-GROUPNAME-UNIX* time in milliseconds where *SCHEDULENAME* is the name of the Avamar schedule and *GROUPNAME* is the name of the Avamar group. • For on-demand backups initiated from the Policy window **Back Up Group Now** command, the work order ID is formatted as: *GROUPNAME-UNIX* time in milliseconds • For on-demand backups or restores initiated from the **Backup and Restore** window, the work order ID is formatted as: *MOD-UNIX* time in milliseconds • For on-demand backups initiated from the systray icon on a Windows Avamar client, the work order ID is formatted as: *COD-UNIX* time in milliseconds • For command-line backups or restores, the work order ID is formatted as: *NAH-UNIX* time in milliseconds • For replication activities, the work order ID is formatted as: *COD-NAH-UNIX* time in milliseconds
ClientVer	The Avamar client software version.
Operation	Operation is one of the following kinds of activities: • On-demand backup • Scheduled backup • Restore • Validate • Replication source • Replication destination
Status	The FINAL status of the client activity is one of the following: • Activity completed successfully • Activity completed with exceptions • Activity cancelled • Activity failed - timed out before starting • Activity failed - timed out before completion • Activity failed - client was given a workorder, but did not acknowledge its receipt • Activity failed - client error(s) • Activity failed - timed out before completion • Activity failed - client has no data specified by dataset • Dropped Session - No progress reported
SessionID	The SessionID is a unique identifier for the client activity.

Background on backups

This topic provides background information about how the Avamar client performs backups, including key statistics.

Dataset size

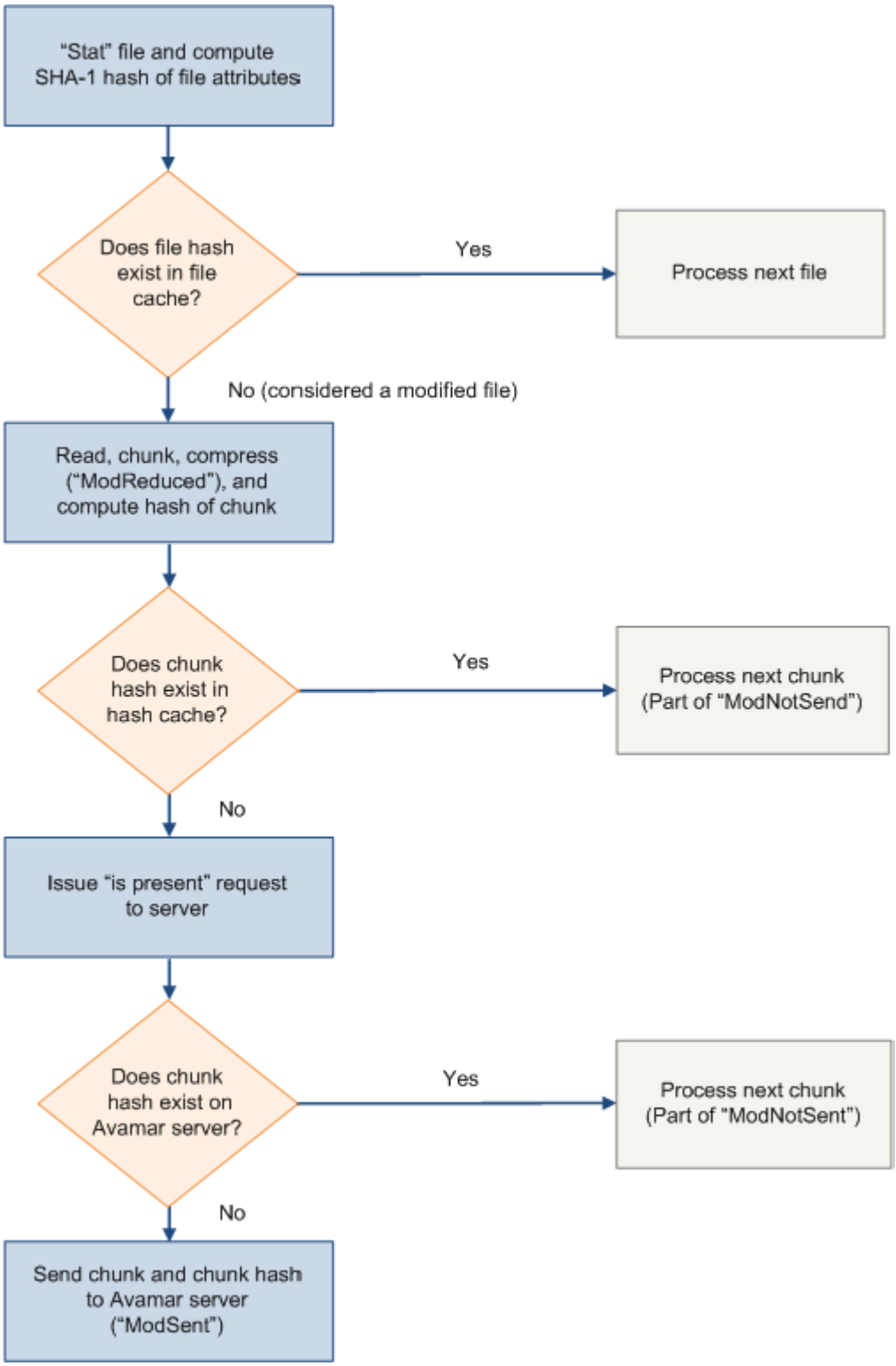
Begin with the value in the TotalBytes column as shown in Example 5 on page 64. This value, 451,940,965,688 bytes (or 421 GB), represents the dataset size. Any files excluded due to exclusion rules are not counted in this total. This total does include open files that could not be backed up, perhaps because the file system was not frozen.

Modified files

When scanning through a file system, obtain file metadata and compute the SHA-1 hash of the metadata. Then look up SHA-1 hash in the file cache on the client. If the hash is present, the opening and reading of the files is not necessary. Therefore, a high percentage of hits in the file cache makes the overall backup proceed very quickly.

Any file whose metadata hash gets a miss in the file cache is considered a modified file (a file that was modified since the last backup). Therefore, the Mod bytes in NumModFiles (column H), ModReduced (column I), ModNotSent (column J), and ModSent (column K) are really shorthand for bytes associated with modified files (that is, files you must open and read so that all the data in the file can be chunked, compressed, and hashed).

The following figure shows the roles of the file cache, the hash cache, and the server **is_present** requests in determining which data to send to the server.



The cache and hash flowchart references the following terms:

- ModReduced — When Avamar backs up modified files, the data is chunked, compressed, and then hashed. Because the compression takes place on the client, the amount of compressed data is reported as ModReduced.

  In Example 5 on page 64, the ModReduced = 55,023,086,382 (51 GB).

- ModNotSent — When subfile level commonality exists, the data is not sent. ModNotSent is shorthand for bytes in modified files that do not have to be sent to the Avamar server because of subfile-level commonality factoring.

  ModNotSent = 4,115,205,370 (3.8 GB) in Example 5 on page 64 means 3.8 GB of compressed chunks were already on the Avamar server.

- ModSent — When new bytes must be sent to the server, they would be reported as ModSent.

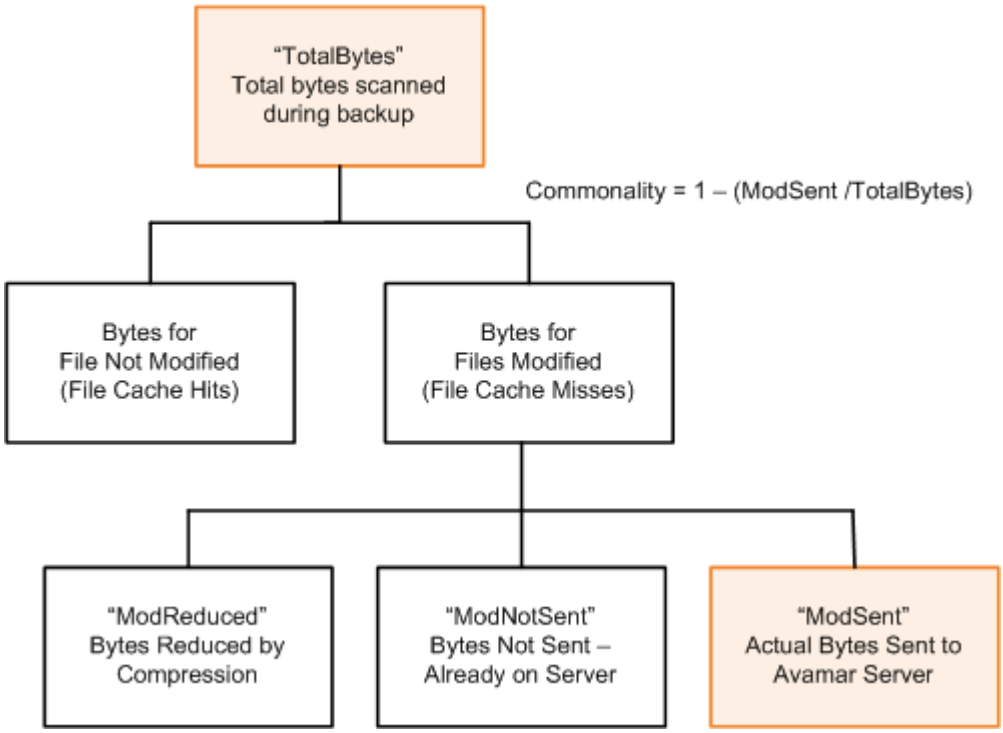
  In this case, ModSent = 393,498,485 (0.37 GB).

Summary of key DPN summary terms

The following describes the relationships between DPN summary terms:

- TotalBytes = (Total bytes in the dataset, including open files that were not backed up) - (Subtotal bytes excluded by exclusion rules)
- TotalBytes = (Subtotal bytes for files not modified) + (Subtotal bytes for files modified since previous backup)
- Subtotal bytes for files modified since previous backup = ModReduced + ModNotSent + ModSent

The relationship between these values is shown in the following figure.



Definition of commonality

Avamar change rate is equivalent to the ModSent / TotalBytes.

The Avamar commonality = 1 - (change rate).

Avamar backup compared to incremental tape backup

During an incremental tape backup, the amount of data sent across the network (if the backup data was not already compressed on the client) is equal to the Subtotal bytes for files modified since the previous backup. The efficiency associated with the Avamar commonality factoring when compared to incremental tape backups is as follows:

ModSent / (ModReduced + ModNotSent + ModSent)

On this particular date as shown in Example 5 on page 64, the Subtotal bytes for files modified since the previous backup = 51 GB + 3.8 GB + 0.37 GB = 55 GB. If this is divided by the TotalBytes, the result is 55 / 421 = 13%.

Typically, in day-over-day backups of file servers, this value is expected to be in the 2% range.

When backing up databases, expect this value to be 100% because every file is touched every day. Therefore, the total bytes associated with modified files is equal to the total bytes.

Definition of terms during restore activities

During restore and validate activities, ModReduced is the amount that the data expanded during the restore or validate operation. ModSent is the actual amount of data sent from the Avamar server to the Avamar client during the restore or validate operation. During restore or validate, TotalBytes = ModSent + ModReduced.

CHAPTER 9
Protecting Avamar Desktop/Laptop Clients

This implementation and daily operations chapter focuses on best practices for Avamar environments with Avamar Desktop/Laptop clients.

Topics in this chapter include:

About Avamar Desktop/Laptop

Avamar Desktop/Laptop is client/server software that extends data backup and recovery to end users who are on the LAN, in remote offices, or connected to the corporate network through a VPN.

When end users log in during normal backup windows, Avamar Desktop/Laptop backs up data from desktop and laptop computers to the Avamar server by using existing network links. Users can also initiate backups from the desktop user interface.

The Avamar Desktop/Laptop client is installed as part of an Avamar Client for Windows, Avamar Client for Mac OS X, or Avamar Client for Linux installation. The Avamar Desktop/Laptop server is installed as part of every Avamar server installation.

NOTICE

The *EMC Avamar Administration Guide* provides more information about Avamar Desktop/Laptop.

Deploy additional Avamar servers for Desktop/Laptop clients

When deploying Avamar Desktop/Laptop to a location with existing Avamar servers, use an additional Avamar grid to support the desktop and laptop clients. Backups of Desktop/Laptop clients must be run when users are online (usually during the day). "Adjust runtime for daily maintenance tasks" on page 82 provides more information. Scheduled backups for fileservers and database clients normally run during the night. "Scheduling activities during the course of a day" on page 36 contains more information about backup windows, blackout windows, and maintenance windows.

Best practice

When deploying Avamar Desktop/Laptop to a location with existing Avamar servers, use an additional Avamar grid to support the desktop and laptop clients.

Create a dataset to back up only user data

The use of Avamar Desktop/Laptop to back up users' desktop and laptop computers can significantly impact Avamar storage capacity depending on the following factors:

- Number of desktop and laptop computers to back up
- Amount of data on each computer

To best manage storage capacity, back up only user files and folders, and exclude common data such as application and operating system files.

Best practices for creating a dataset

- Create a backup dataset that specifies the files and folders for the backup.
- Exclude certain file types from desktop and laptop backups.
- If practicable, minimize the number of entries you define in exclude and include lists.
- In an environment that contains both Windows XP and Windows Vista or Windows 7 clients, add the **--x18=256** option to the dataset to prevent the "Path not found" error.

The following table lists folders to include in a Desktop/Laptop dataset:

Table 14 Desktop/Laptop file types to include in a dataset

OS	Tab	Files and folders
Windows	Source Data	• #USERDOCS#*\Desktop • #USERDOCS#*\Documents • #USERDOCS#*\My Documents • #USERDOCS#*\Favorites **Notice:** A change in the default location of user directories was made between the Windows XP release, and the Windows Vista and Windows 7 releases. To handle this change, the Windows Desktop/Laptop plug-in uses #USERDOCS# as a variable that translates to the default location based on the specific Windows operating system.
Mac	Source Data	• /Users • /Users/*/Desktop • /Users/*/Documents • /Users/*/Library/Safari

Exclusions for Windows computers

The following table lists folders to exclude from the Desktop/Laptop dataset for a Windows system.

Table 15 Desktop/Laptop file types to exclude from a dataset (page 1 of 2)

File type	Files and folders
Link files in each user's Recent folder	*\Recent*.lnk
Google Desktop Search folder	*\Local Settings\Application Data\Google\Google Desktop Search
Windows Indexing and Search services	• *\catalog.wci • *\windows.edb

Table 15 Desktop/Laptop file types to exclude from a dataset (page 2 of 2)

File type	Files and folders
Other nonbusiness files such as, personal music, pictures, video, and so forth	• *.avi • *.cdr • *.dmg • *.iso • *.m4v • *.mov • *.mp3 • *.mp4 • *.mpeg • *.jpeg • *.rar • *.r[0-9][0-9] • *.tgz • */iTunes/ • *.wma • *.wmv
Anti-virus software quarantine files	Check the anti-virus vendor documentation to determine the folder used to store quarantine files.
Recycle bin files	• ‹*SYSTEM DRIVE*›\%Recycle.bin (Windows Vista and Windows 7) • ‹*SYSTEM DRIVE*›:\RECYCLER (Windows XP)

Exclusions for Mac computers

Exclude the following files and folders in the Desktop/Laptop dataset for a Mac system:

- */.Trash/
- */Library/Caches/
- */Library/Cookies/
- */Library/Logs/
- */Library/PubSub/Feeds/
- */Library/Application Support/SyncServices/Local/

Minimize the number of exclude and include lists

Avamar must compare every file selected for a backup with each entry in both lists to determine whether the file is to be backed up. This comparison process adds overhead and potentially increases the duration of each backup.

Dataset caveat

In an environment that contains Windows XP desktop or laptop computers along with Windows Vista or Windows 7 desktop or laptop computers, backups can appear to fail if both Windows XP clients and Windows Vista or Windows 7 clients use a single dataset that specifies the My Documents folder and the Documents folder. A backup in such an environment displays a backup failure on the status bar and writes an error similar to the following to the log file:

```
Path not found
```

In an environment that contains both Windows XP and Windows Vista or Windows 7 clients, add the **--x18=256** option to the dataset to prevent the "Path not found" error.

To add the **--x18=256** option to the dataset:

1. From Avamar Administrator, select **Tools** › **Manage Datasets**.

   The Manage All Datasets dialog box appears.

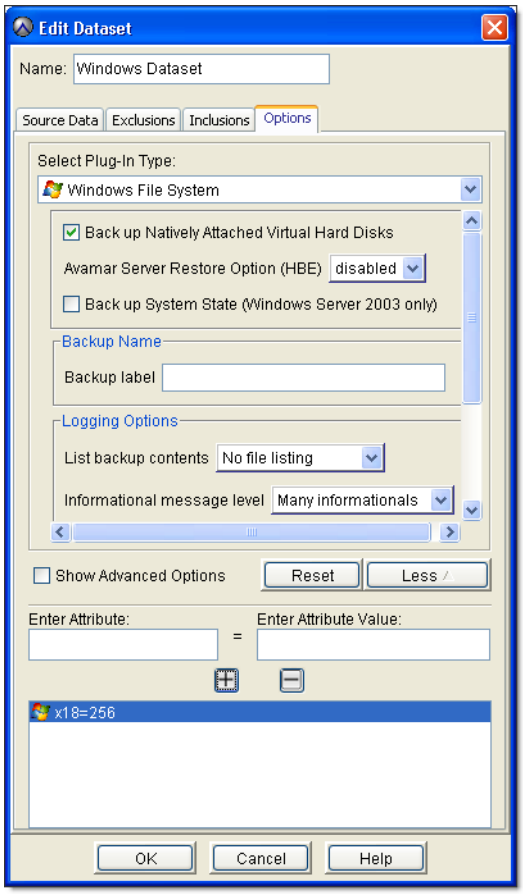
2. Select the dataset from the list and click **Edit**.

   The Edit Dataset dialog box appears.

3. Click **Options** and select the plug-in from the **Select Plug-In Type** list.
4. Click **More**.
5. Type **x18** in the **Enter Attribute** text box.
6. Type **256** in the **Enter Attribute Value** text box.

7. Click ⊞.

   The attribute/value pair (--x18=256) appears in the large text box as shown in the following figure.



8. Click **OK** twice.

Keep the initial client backups to a manageable number

Avamar Desktop/Laptop environments can support up to 5000 clients for each Avamar server. However, simultaneously running first-time backups for hundreds or thousands of clients can create network throughput issues. These issues can prevent the completion of the backups within the backup window. Throughput issues caused by large amounts of data transfer are normally only an issue when running first-time backups. This is because the savings realized through data deduplication is at its lowest when a system is first backed up and so the amount of data that must be transferred is at its greatest.

Best practices for first-time backups

- Keep the initial client backups to a manageable number.
- To minimize the impact of first-time backups, bring clients online in smaller groups.

Strategy for performing first-time backups

All new Avamar client computers require a first-time backup of all the data specified by the dataset. These first-time backups require significantly more time and storage space than subsequent backups that only back up changed data.

Use the first few groups to discover information about the capabilities of the network. Start with smaller groups of clients and, after each group is successfully added, increase the size of the next group.

On the first day, start with activation and a first-time backup of clients equal to no more than 10 times the number of storage nodes deployed.

For example, if you have 5 storage nodes, back up 50 clients:

10 x 5 storage nodes = 50 backup clients

If all backups for the first day complete within the scheduled backup window, double the amount of clients in the group on day two. Continue adding more clients on subsequent days until all initial backups are complete. Reduce the number of clients if any backups fail to complete within the backup window.

Strategy for setting up backup groups

As the amount of data archived in an Avamar system increases, the benefit of global deduplication increases. This means that throughput problems decrease exponentially as more clients are added and the number of common data objects in the system increases. To achieve these results:

1. Place clients with the smallest burden on the network infrastructure in the first backup groups that are brought online.
2. Place clients with the greatest burden on the network in the last groups brought online.
3. Set up backup groups following these guidelines:
    a. First backup groups should consist of computers that traverse the shortest logical distance to the Avamar server.

       Logical distance is increased by the following factors:

       – Routers
       – Firewalls
       – VPN tunnels
       – Physical distance

b. First backup groups should consist of computers that utilize the fastest network transmission rates.

   An illustrative, nonexhaustive list, from fastest to slowest:

   – Gigabit Ethernet
   – 802.11n
   – Fast Ethernet
   – 802.11g
   – Ethernet
   – V.92
   – V.34

Activating clients by using the Avamar Client Manager

The recommendation to bring clients online in sequentially targeted groups can best be achieved by using the directory information tree or search capability of Avamar Client Manager. You can select appropriately sized and situated organizational units by using the tree structure. Or, you can use search terms to define a group with the target number and type of clients. Then, using its drag-and-drop capability, you can activate and initiate first-time backups of correctly sized and connected groups of clients. This way, you avoid the problems associated with overtaxing the throughput capabilities of the network.

Consider node size when configuring Avamar servers

In certain situations, the use of lower capacity nodes might be advisable. The use of lower capacity nodes increases the connection count, which in turn, increases the potential number of concurrent backups.

Best practice

Consider node size when configuring Avamar servers.

Determine the backup window

Avamar Desktop/Laptop environments typically include more clients than traditional Avamar systems. The Avamar Administrator server allows a maximum of 27 concurrent backup connections for each active storage node. One connection for the overall Avamar grid is reserved for restores.

To determine how many backups can complete within an hour, consider the following two examples.

Example 6 Six backups per storage node

Backup criteria:

- 10 storage nodes
- Average backup time = 10 minutes
- 6 backups per storage node connection per hour (60 min./10 min. = 6)

Formula:

10 nodes x 6 backups per connection x 27 concurrent connections = 1,620 backups per hour

Example 7 Two backups per storage node

Backup criteria:

- 10 storage nodes
- Average backup time = 30 minutes
- 2 backups per storage node connection per hour (60 min./30 min. = 2)

Formula:

10 nodes x 2 backups per connection x 27 concurrent connections = 540 backups per hour

In both examples, variables such as the following can affect the total backup time:

- Total size of the backup dataset
- Amount of daily changes to data
- Network speed and amount of network traffic
- Concurrent processes that run on the Avamar server

For instance, backing up data from a LAN-connected laptop usually takes less time to complete than backing up the same computer when it is connected to the corporate network by an 802.11G wireless connection.

Best practice for determining the backup window

Use the number of backups per hour to help you determine a backup window, which allows enough time to back up all desktop and laptop computers.

Schedule the backup window

The backup window for desktop and laptop clients is often the opposite of traditional server clients. Avamar Desktop/Laptop backups must run while the desktop and laptop computers are online. The backup window, therefore, is typically during the work day.

When determining the backup window, ensure that it is flexible enough for users who are offline due to traveling, meetings, and so forth.

Start with a backup window of 12 hours and increase or decrease it as necessary. Depending on the location of remote clients, backing up all clients might require multiple Avamar servers.

Best practice for scheduling backups

Schedule the backup window to back up desktop and laptop computers when they are most likely to be online.

Adjust runtime for daily maintenance tasks

It is important that Avamar daily maintenance tasks complete successfully every day. Failures of these tasks quickly results in capacity problems for the Avamar server.

The timing for daily maintenance tasks for Avamar Desktop/Laptop is different from the timing of a standard servers-as-clients deployment of Avamar.

In a standard deployment, the servers are backed up at night when they are least active. To accommodate this, the Avamar daily maintenance tasks run during the day.

For Avamar Desktop/Laptop, the client backups usually run during the day, when the clients are most likely to be powered-up and connected. The Avamar daily maintenance tasks often must be changed to run during the night to avoid conflicts with client backups and restores. "Scheduling activities during the course of a day" on page 36 provides more information.

The daily maintenance task of garbage collection requires a quiet system. Garbage collection cannot start when backups are running. Because garbage collection reclaims space on the Avamar server, the failure to successfully complete this task can quickly create capacity problems.

Best practice for scheduling daily maintenance tasks

Adjust the maintenance window so that it allows daily maintenance tasks to complete without overlapping with the backup window.

Do not run client utilities during the backup window

Avoid running multiple system utilities on the user's PC or Mac at the same time as backups. For instance, do not schedule antivirus scans or disk defragmenter jobs during the backup window.

Run backups more frequently than the retention policy

Backup retention policies specify how long to keep a particular backup in the system. Backups are automatically marked for deletion after they expire. You must, therefore, be sure to run backups more frequently than the amount of time specified by the retention policy. For example, if the retention policy is 30 days, ensure that you run backups before the 30-day retention period expires. If you fail to back up data within the retention period, the data is no longer part of the backup. The data is still available on the hard drive unless it has been deleted.

Frequent backups ensure that data is always available from the backup.

The following resources provide more information about retention policies:

- *EMC Avamar Administration Guide*
- "Defining Domains, Groups, and Policies" on page 41

As an alternative to running backups more frequently than what the retention policy specifies, you can configure the Avamar server to never delete the client's last backup. To enable this feature you set the keep_last_backup key to true in the mcserver.xml file on the Avamar server. This setting prevents Avamar from deleting a client's last backup, which can be beneficial for backup clients with short retention policies (for instance, less than 2 weeks), or backup clients that are offline for extended periods of time when users are out of the workplace for vacations or for other reasons. Configuring the Avamar server to never delete the client's last backup, therefore, might be preferable to running backups more frequently than what the retention policy specifies. The tradeoff in configuring the keep_last_backup key in the mcserver.xml file on the Avamar server is that more space is consumed on the Avamar server because the last backup for all clients is never deleted. The *EMC Avamar Administration Guide* provides more information about setting the keep_last_backup key.

Prevent backups on a wireless connection

In some locations, users might pay exorbitant data transmission fees when their backups run over a wireless connection. To avoid these exorbitant data transmission fees, you can disable backups from running over a wireless connection by clearing the Back Up On Wireless option.

On Windows, to clear the Back Up On Wireless option:

1. Right-click the Avamar icon in the system tray.

   The context menu appears.

2. Select the **Settings** menu.
3. Clear the **Back Up On Wireless** option.

On Mac, to clear the Back Up On Wireless option:

1. Click the Avamar icon on the menu bar.

   The context menu appears.

2. Select the **Settings** menu.
3. Clear the **Back Up On Wireless** option.

Manage storage capacity for Desktop/Laptop clients

The most important consideration in successfully maintaining Avamar Desktop/ Laptop is capacity management. A properly managed Avamar server achieves steady state when storage capacity is well below the capacity warning threshold, which is 80% of storage capacity.

Steady state operation is achieved when the average data sent to the Avamar server is less than or equal to the average data removed from the multi-node server. "Steady state system" on page 32 provides more information.

As a multi-node server enters the range of 85% to 100% of storage capacity, performance degradation occurs. If storage capacity reaches 100%, the Avamar server transitions to a read-only state. This transition protects the integrity of the data already stored on the server. "Avamar capacity thresholds" on page 29 provides more information.

A server cannot achieve steady state and will exceed storage capacity if:

- Clients back up more than the initial backup size limit
- Clients exceed the daily data change rate limit
- Garbage collection fails

Best practice

Manage storage capacity for Avamar Desktop/ Laptop clients.

The *EMC Avamar Administration Guide* provides more information about capacity management. This guide provides detailed descriptions of the features, functions, and tools available to assist you with properly monitoring and managing server storage capacity.

Ensure adequate initialization time for Wake-on-Lan backups

Both Windows and Mac OS X offer power management features to reduce power consumption and save energy. If you use Wake-on-Lan (WoL) network technology to remotely power on or wake up a computer before a scheduled backup starts, make sure client systems have adequate initialization time.

Best practices

- Ensure that power management settings for client computers do not return the client to a powered-down or sleep state before the backup request is received.
- Depending on the number of clients to be backed up, clients may be queued while waiting for processing.
- Schedule WoL backups so that clients are powered on or awake before a connection is available.

CHAPTER 10
Other Avamar Administration Best Practices

This planning, design, and implementation chapter contains information about the following administrative best practices.

Topics include:

Protecting the Avamar server

Deploy an Avamar server in a protected network and not one exposed to the Internet. Even when you deploy an Avamar server in an internal network, protect the server by a firewall to prevent unauthorized access to the nodes that comprise the server.

Best practice

Protect the Avamar server from the Internet by providing full firewall protection.

Use of an uninterruptible power supply with Avamar

Under some circumstances, an unclean shutdown of an Avamar grid can result in data inconsistencies, which a rollback can recover from, but results in losing any backups completed after the last successful checkpoint. There are various ways Avamar grids can experience unclean shutdowns, many of which are preventable. They include cases like unexpected site-wide power outages, not performing a clean shutdown before planned power outages, not connecting the Avamar Data Store redundant power system to independent power sources, or using incorrect shutdown procedures.

Avamar systems are not synchronously protected against unexpected power loss. Avamar nodes have dual power supplies, so use an uninterruptible power supply (UPS) to achieve full protection against power loss. EMC recommends the use of redundant power supplies to the Avamar system that are connected to separate sources, with one of them being protected with a UPS.

After a power outage has occurred and while the system is being protected by the UPS, you should run a checkpoint in preparation for shutting down the system. Use the proper shutdown procedures as documented in the Knowledgebase article esg112243, “Avamar Data Store Single- and Multi-node Shutdown/Startup Procedures,” available at https://support.EMC.com/products.

Best practices for using a UPS with Avamar

- Ensure that redundant power is actually connected to separate sources, one of them being backed by UPS.
- Use proper shutdown procedures.
- Ensure daily integrity checks are successful, should a rollback be required.
- Deploy a UPS for the Avamar server hardware to protect against data loss caused by unplanned power outages.

Changing passwords

If you have not changed Avamar passwords from the factory default values, use the **change-passwords** utility to change them. The *EMC Avamar Administration Guide* provides more information about changing Avamar passwords.

The following table lists Avamar user accounts and SSH keys that require password changes.

Table 16 Avamar user accounts and SSH keys

Type	Username
Operating system user accounts	• root • admin • dpn
SSH keys	• admin_key • dpnid
Root-level software application user accounts	• root • MCUser • repluser

Change the passwords for all of these user accounts.

Changing only the passwords for the operating system users does not sufficiently prevent someone from logging in to Avamar server nodes. If you have not changed the two SSH keys, someone could use the factory-default SSH keys to log in to the Avamar server.

Best practice

Change all factory default passwords except the passwords for the backuponly, restoreonly, and backuprestore software application users.

Using Avamar Client Manager

Avamar Client Manager is a web-based management application that provides centralized Avamar client administration capabilities for larger businesses and enterprises. You start Avamar Client Manager from the Avamar Enterprise Manager menu bar.

For environments that include a large number of clients, use Avamar Client Manager to simplify the following tasks:

- Activating clients
- Moving clients to new domain on same Avamar grid
- Removing a client from a group on same Avamar grid
- Moving a client to a new group on same Avamar grid
- Moving client to new Avamar grid
- Reporting with backup and restore summary

- Retiring clients
- Deleting clients
- Upgrading client software (requires the 6.1 plug-in)

Best practice

Use Avamar Client Manager to facilitate the management of large numbers of Avamar clients.

The *EMC Avamar Administration Guide* provides more information about Avamar Client Manager.

Enabling the Email Home feature

When configured and enabled, the Email Home feature, including ConnectEMC, automatically emails configuration, capacity, and general system information to EMC Customer Support once daily, and critical alerts in near-real time on an as-needed basis.

Enable this feature on all Avamar servers. The *EMC Avamar Administration Guide* provides more information.

The Email Home feature is offered as part of the server maintenance plan. However, it is offered with the following understanding:

1. There is no guaranteed service level agreement for monitoring Email Home messages. You must assume primary responsibility for monitoring the Avamar systems.
2. Support cases are automatically opened for issues that affect the backup infrastructure (Avamar server) such as a failed hfscheck. Support cases are not opened for issues associated with backup operations such as failed backups.
3. EMC Customer Support will not proactively alert you of problems, such as a server down or a disabled schedule, that prevent Email Home messages from being sent.

Best practice

Enable Email Home.

Using EMC Secure Remote Support solution

EMC Secure Remote Support (ESRS) IP Solution is an IP-based automated connect home and remote support solution. The ESRS IP solution creates both a unified architecture and a common point of access for remote support activities performed on EMC products.

To simplify remote support of Avamar servers, install the EMC Secure Remote Gateway. It is integrated with ConnectEMC.

The use of the ESRS IP solution enables EMC Customer Support to:

- Log in through the EMC Secure Remote Gateway to troubleshoot problems, which eliminates the need for WebEX sessions.
- Begin work on critical issues soon after a ConnectEMC notification of a problem is received.

Best practice

Enable ESRS.

The *EMC Secure Remote Support IP Solution Security Management and Certificate Policy: Frequently Asked Questions*, which is available from EMC Online Support (https://support.EMC.com/products), provides more information.

Assigning users

The Avamar software includes access audit logging capability. The broad intent of this feature is to maintain a log of every action taken on vital system components/objects.

The data in this log enables enterprises deploying Avamar to do the following:

- Enforce security policies
- Follow up on security breaches or deviations from policies
- Hold appropriate users accountable for those actions

Best practices

- Assign each Avamar administrator, operator, or user a unique login credential. Ensure that all users log in to the Avamar system by using those unique login credentials rather than the default Avamar application root and MCUser users.
- Work with EMC Customer Support to set up External Authentication so that all Avamar administrators, operators, and users can log in to the Avamar server with Active Directory, LDAP, or NIS login credentials. The *EMC Avamar Administration Guide* provides more information.

CHAPTER 11
Using Data Domain Systems

This planing chapter contains best practices information for Data Domain systems to be used as backup storage in an Avamar configuration.

Topics include:

Network bandwidth recommendations

Before you add a Data Domain system to an Avamar configuration, ensure that the infrastructure provides adequate network bandwidth for backups and Avamar maintenance activities.

Network bandwidth in an Avamar configuration has the most impact on Avamar client backups to a Data Domain system and Avamar server maintenance activities. The process that sends Avamar client metadata to the Avamar server has less impact on the network bandwidth.

To measure the network bandwidth between the Avamar server and Data Domain system, use the **iperf** utility. The **iperf** utility is available on the Avamar server, on the Data Domain system, and from the Internet:

- On an Avamar server, the Linux operating system includes the **iperf** utility in /usr/local/avamar/bin.
- On a Data Domain system, the Data Domain Operating System (DD OS) includes the **iperf** utility.
- For Avamar clients, download the **iperf** utility from the Internet.

Use the iperf utility to test the network bandwidth

For the most comprehensive results, run the **iperf** utility in server mode on the Avamar server and in client mode on the Data Domain system. Then run the **iperf** utility in server mode on the Data Domain system and in client mode on the Avamar server. Run the **iperf** utility several times to verify the consistency of the results.

To run the **iperf** utility:

1. Run the **iperf** utility in server mode on the Avamar server by typing:

   `iperf -s -w 256k`

2. Run the **iperf** utility in client mode on the Avamar server by typing:

   `iperf -c` *`iperf-server-name`* **`-w 256k`**

   where *iperf-server-name* is the iperf server.

   To view statistics for every second, add the **-i 1** option.

3. Run the **iperf** utility in server mode on the Data Domain system by typing:

   `net iperf server window-size 256K`

4. Run the **iperf** utility in client mode on the Data Domain system by typing:

   `net iperf client` *`iperf-server-name`* **`window-size 256K`**

   where *iperf-server-name* is the iperf server.

   To view statistics for every second, add the **interval 1** option.

Recommended network bandwidth

For a 1-gigabit connection to the Data Domain server, a network bandwidth of 800 Mbps/sec or greater is sufficient for client backups. A number less than 800 Mbps/sec might cause a network bottleneck and limit the throughput to a Data Domain system.

For a 10-gigabit connection to the Data Domain server, the network bandwidth of 5 Gbps/sec or greater is sufficient for client backups. A number less than 3 Gbps/sec might cause a network bottleneck on certain Data Domain systems models.

If the network bandwidth results are insufficient for the Avamar client and Data Domain system, review the information on the **Status** › **Stats** page in the Data Domain Enterprise Manager. This page shows network, nfs, and disk throughput. The *EMC DD OS 5.1 Administration Guide* provides more information about viewing system statistics.

Example iperf utility sessions

The following examples show output from the **iperf** utility when it is used to test bidirectional-network bandwidth between a Data Domain system and an Avamar client.

Example 8 The **iperf** utility output from a Data Domain as the **iperf** client and an Avamar client as the **iperf** server.

Example criteria:

- iperf server — Avamar client
- iperf client — Data Domain system
- Connection — 1 gigabit

```
sysadmin@datadomain1# net iperf client clidev02.lab.com window-size 256K
------------------------------------------------------------
Client connecting to clidev02.lab.com, TCP port 5001
TCP window size:       512 KByte (WARNING: requested 256 KByte)
------------------------------------------------------------
[  3] local 192.168.0.10 port 56276 connected with 192.168.0.11 port 5001
[  3] 0.0-10.0 sec     1.09 GBytes          938 Mbits/sec
sysadmin@datadomain1#

root@clidev02# iperf -s -w 256k
------------------------------------------------------------
Server listening on TCP port 5001
TCP window size: 264 KByte (WARNING: requested 256 KByte)
------------------------------------------------------------
[  4] local 192.168.0.11 port 5001 connected with 192.168.0.10 port 56276
[ ID] Interval         Transfer            Bandwidth
[  4] 0.0-10.0 sec     1.09 GBytes          938 Mbits/sec
```

Example 9 The **iperf** utility output from a Data Domain as the **iperf** server and an Avamar client as the **iperf** client.

Example criteria:

- iperf server — Data Domain system
- iperf client — Avamar client
- Connection — 1 gigabit

```
sysadmin@datadomain1# net iperf server window-size 256K
------------------------------------------------------------
Server listening on TCP port 5001
TCP window size:       512 KByte (WARNING: requested 256 KByte)
------------------------------------------------------------
[  4] local 192.168.0.10 port 5001 connected with 192.168.0.11 port 52347
[  4] 0.0-10.0 sec     1.09 GBytes          937 Mbits/sec

root@clidev02# iperf -c datadomain1.lab.com -w 256k
------------------------------------------------------------
Client connecting to datadomain1.lab.com, TCP port 5001
TCP window size:  264 KByte (WARNING: requested  256 KByte)
------------------------------------------------------------
[  3] local 192.168.0.11 port 52347 connected with 192.168.0.10 port 5001
[ ID] Interval         Transfer             Bandwidth
[  3] 0.0-10.0 sec     1.09 GBytes          936 Mbits/sec
```

Example 10 The **iperf** utility output from a Data Domain as the **iperf** client and an Avamar client as the **iperf** server.

Example criteria:

- iperf server — Avamar client
- iperf client — Data Domain system
- Connection — 10 gigabit

```
sysadmin@datadomain4# net iperf client clidev02.lab.com window-size 256K
------------------------------------------------------------
Client connecting to clidev02.lab.com, TCP port 5001
TCP window size: 512 KByte (WARNING: requested 256 KByte)
------------------------------------------------------------
[  3] local 192.168.0.12 port 37368 connected with 192.168.0.11 port 5001
[  3] 0.0-10.0 sec     7.82 GBytes          6.71 Gbits/sec
sysadmin@datadomain4#

root@clidev02# iperf -s -w 256k
------------------------------------------------------------
Server listening on TCP port 5001
TCP window size:  264 KByte (WARNING: requested 256 KByte)
------------------------------------------------------------
[  4] local 192.168.0.11 port 5001 connected with 192.168.0.12 port 37368
[ ID] Interval         Transfer             Bandwidth
[  4] 0.0-10.0 sec     7.82 GBytes          6.71 Gbits/sec
```

Example 11 The **iperf** utility output from a Data Domain as the **iperf** server and an Avamar client as the **iperf** client.

Example criteria:

- iperf server — Data Domain system
- iperf client — Avamar client
- Connection — 10 gigabit

```
sysadmin@datadomain4# net iperf server window-size 256K
------------------------------------------------------------
Server listening on TCP port 5001
TCP window size:        512 KByte (WARNING: requested 256 KByte)
------------------------------------------------------------
[  4] local 192.168.0.12 port 5001 connected with 192.168.0.11 port 52351
[  4]  0.0-10.0 sec     9.70 GBytes          8.33 Gbits/sec
root@clidev02# iperf -c datadomain4.lab.com -w 256k
------------------------------------------------------------
Client connecting to datadomain4.lab.com, TCP port 5001
TCP window size: 264 KByte (WARNING: requested 256 KByte)
------------------------------------------------------------
[  3] local 192.168.0.11 port 52351 connected with 192.168.0.12 port 5001
[ ID] Interval          Transfer             Bandwidth
[  3]  0.0-10.0 sec     9.70 GBytes          8.32 Gbits/sec
root@clidev02#
```

Configuration best practices

The following topics cover configuration best practices for Data Domain systems and Data Domain Archivers.

Use fully qualified domain names

Assign fully qualified domain names (FQDN) to the Data Domain system before you add it to an Avamar configuration. When you configure the Data Domain system:

- Do not use the IP address of the Data Domain system. Use of an IP address instead of the hostname can limit the ability to route duplication traffic.
- Use forward and reverse DNS lookup for the Avamar server, Avamar clients, and the Data Domain system.
- Use DNS to resolve hostnames to routable IP addresses.
- Use host files to resolve hostnames to non-routable IP addresses.
- Do not create secondary hostnames for alternate or local IP interfaces.

Review the amount of files in the MTree

The DD Boost and DD OS are not designed to handle backups of large amounts of small files or backups of thousands of clients. The best use for Avamar integration with Data Domain systems is in data centers that back up a few very large files.

When an MTree contains many small files, the following known issues might occur:

- Avamar checkpoint and rollback processes on a Data Domain system might fail if the Data Domain system contains more than 400,000 directories or files below the cur directory or in any checkpoint directory.
- A DD OS 4.9 conversion of an LSU to an MTree might fail if there are over 400,000 objects in the LSU.

The expiration of a two-minute timeout period causes both of these issues.

Do not modify the MTree

An MTree is protected by the DD OS so that only the DD Boost account has read/write privileges to the MTree. Do not modify the MTree structure by using SSH from the DD OS command line, NFS, CIFS, or the Data Domain Enterprise Manager. Modifications to the MTree by using any one of these methods might result in failed backups, restores, or Avamar server maintenance operations.

The Avamar software controls additions, modifications, and deletions to the Data Domain system by using the DD Boost library. Do not manually add, modify, or delete the contents of an MTree. Doing so might cause irreversible consistency problems between the Avamar server and Data Domain system.

Specify the maximum number of data streams

Avamar clients that support Data Domain systems as a storage device can use multiple data streams during backups and restores. For example, the Avamar Plug-in for SQL Server, the Avamar Plug-in for SharePoint VSS, and the Avamar Plug-in for Exchange VSS use one stream for each backup or restore. The Avamar Plug-in for Oracle can use 1 to 6 streams. The number of streams you specify depends on the number of backups you plan to run simultaneously.

When you add a Data Domain system to an Avamar configuration:

- Specify the total amount of streams that the Data Domain system model supports if only one Avamar server uses the Data Domain system.
- Specify a subset of streams that the Data Domain system model supports if multiple Avamar servers or other third-party applications share the Data Domain system.

Evaluate storage requirements

When you configure Avamar to use a Data Domain system that third-party applications also use, carefully evaluate the amount of storage you need for Avamar data. Ensure that enough storage is available for both Avamar data and the third-party applications.

Synchronize the time on the Avamar server and Data Domain system

Use an Network Time Protocol (NTP) to synchronize the system time on the Avamar server and the Data Domain system.

Restore SQL Server backups by using the Use SQL REPLACE option

When you restore SQL Server backups from a Data Domain system, always select the Use SQL REPACE option checkbox and clear the Tail-log backup checkbox. The *EMC Avamar for SQL Server User Guide* provides more information about these options.

Space requirements for replication configurations

Avamar replication configurations that include Data Domain systems or Data Domain Archivers must have enough storage space to accommodate replication. After replicating data to a Data Domain system or Data Domain Archiver, ensure that you have at least 10% of free space.

Fully understand the data movement policy before you configure one

To best use the storage space on a Data Domain Archiver, do not configure a data movement policy unless you thoroughly understand this feature. The following Data Domain Archiver use case provides more information.

Use case for Data Domain Archiver

A user wants to store Avamar backups that have long retention periods on a Data Domain Archiver. The user performs the following steps:

1. Configures an Avamar dataset to expire in seven years.
2. Creates a data movement policy on the Data Domain Archiver for the MTree that hosts the Avamar backups.
3. Specifies an age threshold of six months in the data movement policy for the Avamar backup files.

The six-month age threshold setting causes Data Domain Archiver to move Avamar backups on the active tier to the archive tier. When the archive tier is full, Data Domain Archiver seals the archive tier. Sealed archive tiers are read-only.

When the Data Domain Archiver seals an archive tier that contains Avamar backups, the backups expire on the Avamar server. The Data Domain Archiver, therefore, cannot reclaim the space that was allocated for the Avamar backups.

The *EMC DD860 Archiver Administration Guide DD OS 5.1* provides more information about configuring the data movement policy.

INDEX

E

F

G

H

I

K

L

M

N

O

P

R

S

T

U

V

W