EMC® AVAMAR®

5.0

OPERATIONAL BEST PRACTICES

P/N 300-008-815
REV A01

EMC CORPORATION
CORPORATE HEADQUARTERS:
HOPKINTON, MA 01748-9103
1-508-435-1000
WWW.EMC.COM

Copyright and Trademark Notices

Printed in the USA.

TABLE OF CONTENTS

FOREWORD

Scope and Intended Audience

Scope. This publication describes operational best practices for both single-node and multi-node servers in small and large heterogeneous client environments.

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

Product Information

For current documentation, release notes, software updates, as well as information about EMC products, licensing and service, go to the EMC Powerlink web site at http://Powerlink.EMC.com.

Your Comments

Your suggestions will help us continue to improve the accuracy, organization and overall quality of the user publications. Please send your opinion of this document to:

`SSGDocumentation@emc.com`

Please include the following information:

- Product name and version
- Document name, part number and revision (for example, A01)
- Page numbers
- Other details that will help us address the documentation issue

Typeface Conventions

The following table provides examples of standard typeface styles used in this publication to convey various kinds of information.

EXAMPLE	DESCRIPTION
Click **OK**. - or - Select **File > Close**.	Bold text denotes actual Graphical User Interface (GUI) buttons, commands, menus and options (any GUI element that initiates action). Also note in the second example that sequential commands are separated by a greater-than (**>**) character. In this example, you are being instructed to select the **Close** command from the **File** menu.
Type: **`cd /tmp`**	Bold fixed-width text denotes shell commands that must be entered exactly as they appear in this publication.
--logfile=FILE	All caps text often denotes a placeholder (token) for an actual value that must be supplied by the user. In this example, FILE would be an actual filename.
`Installation Complete.`	Regular (not bold) fixed-width text denotes command shell messages. It is also used to list code and file contents.

Notes, Tips and Warnings

The following kinds of notes, tips and warnings appear in this publication:

IMPORTANT: This is a warning. Warnings always contain information that if not heeded could result in unpredictable system behavior or loss of data.

TIP: This is a tip. Tips present optional information intended to improve your productivity or otherwise enhance your experience with our product. Tips never contain information that will cause a failure if ignored.

NOTE: This is a general note. Notes contain ancillary information intended to clarify a topic or procedure. Notes never contain information that will cause a failure if ignored.

OVERVIEW

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

Guide Organization

This best practices guide is organized as follows:

SUBJECT MATTER	CHAPTERS
Core Avamar system functions	*Designing Avamar to Maximize System Availability* (page 11) *Managing Capacity* (page 18) *Scheduling* (page 23) *Defining Domains, Groups, and Policies* (page 29) *Daily Monitoring of Backup Infrastructure* (page 33) *Daily Monitoring of Backup Operations* (page 35)
Tuning the Avamar system	*Tuning Performance* (page 38) *Understanding DPN Summary Reports* (page 49)
Avamar Desktop/ Laptop Clients	*Protecting Avamar Desktop/Laptop Clients* (page 56)
Other Avamar administration functions	*Other Avamar Administration Best Practices* (page 65)

Lifecycle Indicators

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

Planning and design. Topology and architecture options, risks and limitations and any other planning and design issues that must be considered prior to implementing the design.

Implementation. Installation options and directions for testing Avamar components after 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.

Best Practices Indicators

Throughout the guide, the following icons identify practices that are recommended, as well as those that are not.

YES ✓ Practice is recommended.

NO Practice is not recommended.

Assumptions and References

This guide does not attempt to provide introductory materials for basic Avamar technology or delivery methods. Refer to the following Avamar product documentation for additional information:

- *Avamar Release Notes*
- *Avamar Release Notes Addendum*
- *Avamar System Administration Guide*
- *Avamar Event Codes Listing*
- *Avamar Management Console Command Line Interface (MCCLI) Programmer Guide*
- *Avamar Backup Clients User Guide*
- *Avamar DB2 Client User Guide*
- *Avamar Exchange Client User Guide*
- *Avamar Lotus Domino Client User Guide*
- *Avamar NDMP Accelerator User Guide*
- *Avamar Oracle Client User Guide*
- *Avamar SharePoint Client User Guide*
- *Avamar SQL Server Client User Guide*
- *Avamar Product Security Guide*
- *Avamar Server Software Installation Guide*
- *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*

The documentation is available from http://Powerlink.EMC.com.

Top 10 Operational Best Practices

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

- Deploy the Avamar server with reliable, high-performance RAID arrays for back-end storage.
- Protect the data on the Avamar server by replicating the data to another Avamar server.
- 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.
- Enable the email home capability.
- Ensure every administrator logs in to the Avamar server with a unique username.

The chapters that follow provide more details on these best practices, and also provide additional best practices.

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.

This chapter also describes the following main redundancy methods for maintaining data integrity:

- RAID
- RAIN
- Checkpoints
- Replication

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 data 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 kinds 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 data nodes.

The Avamar server performs several functions throughout each day. The major operational functions are the following:

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 these 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, and so will adversely impact the performance and reliability of the Avamar server.

Best practices:

YES If the hardware is purchased separately by the customer, the customer must configure the disk arrays on each node by using RAID.

YES If the hardware is purchased separately by the customer, the customer must configure RAID rebuild as a low priority.

YES Set up log scanning to monitor and report hardware issues, and set up email home.

YES 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:

YES Always enable RAIN for any configuration other than single-node servers and 1x2’s (two active data nodes). Minimum RAIN configuration is a 1x3+1 (three active data nodes plus a utility node and spare node).

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.

YES When deploying non-RAIN 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.

YES Limit initial configurations to 12 to 14 active data 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. It 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:

YES Protect the data on the Avamar server by replicating the data to another Avamar server.

YES Use default standard replication, also known as “root-to-REPLICATE” replication, to do the following:

- Provide the flexibility to configure your replicated grids in a wide variety of ways
- Have full visibility into all the backups that have been replicated from one 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.

YES 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 percent of the total available bandwidth when WAN bandwidth is the performance bottleneck. The *Avamar System Administration Guide* contains more information about setting up replication to best use the system bandwidth.

YES 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.

YES Use the **--exclude** flag only if you decide that a high change-rate or low-priority client can be selectively excluded from the nightly replication.

YES 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 grid.

Checkpoints. Checkpoints provide redundancy across time. Checkpoints allow 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 or 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:

YES 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.

NOTE: During certain support actions, your Customer Support Representative might temporarily change the checkpoint retention policy to ensure that certain critical checkpoints are retained during the support action. Ensure the checkpoint retention policy is restored to the default setting when the support action is completed.

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 (centralized Avamar backup server), and replicate data to another large centralized Avamar server (centralized replication destination)?

When making this decision, refer to the factors described in the following table:

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 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 is important if a recovery time objective (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 data nodes, one utility node, and one spare) that functions as a centralized Avamar backup server.

FACTOR	DESCRIPTION
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, 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:

YES Unless you cannot meet your restore time objectives, 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.

MANAGING CAPACITY

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

When your new Avamar system is initially deployed at your site, the server typically fills rapidly for the first few weeks. This is because, at least initially, nearly every client that is backed up contains relatively large amounts of 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 significantly 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 is referred to as achieving a steady state of 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, keep in mind that 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 *Avamar System 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

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. This is the total amount of commonality factored data and RAIN parity data (net after garbage collect) on each data partition of the server node. This amount is measured and reported by the GSAN process. 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. This 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. You need to understand the behavior of key thresholds as described in the following table:

THRESHOLD	DEFAULT VALUE	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	When the GSAN capacity reaches this 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 into Avamar Administrator, and the system event must be acknowledged before future backup activity can resume.

THRESHOLD	DEFAULT VALUE	CAPACITY USED FOR COMPARISON	BEHAVIOR
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. This prevents new data from being added to the server. This value is reported as server utilization on the Server Management tab (Avamar Administrator > Server > Server Manangement). 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 is most notable in the garbage collection activity.

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

Best practices:

YES ✓ Understand how to monitor and manage the storage capacity of the Avamar server on a daily basis.

YES ✓ Limit storage capacity usage to 80 percent of the available GSAN capacity.

YES ✓ Monitor all variations with incoming data to prevent the system from becoming read-only.

YES ✓ Monitor all variations with maintenance jobs to prevent these jobs from failing.

Proactive Steps to Manage Capacity

You will get a warning when the GSAN capacity exceeds 80 percent of the read-only threshold. If this occurs, you must perform the following additional best practices:

YES Stop adding new clients to the system.

YES Reassess retention policies to see if you can decrease the retention, and therefore, reduce the capacity use.

YES Investigate the possibility that backups are preventing a garbage collect operation from starting. If this is the case, the following error message is written to the garbage collection log:

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

You can use **dumpmaintlogs --types=gc** to view logs for the garbage collection operation.

IMPORTANT: You can decrease the GSAN capacity by deleting or expiring backups and running garbage collection, but you cannot decrease the amount of data in the /data?? partitions.

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

Reactive Steps to Recovering from Capacity Issues

Once the Avamar server capacity significantly exceeds the warning threshold and approaches the diskreadonly limit, you must do one or more of the following:

1. Follow the steps described previously for an Avamar server that starts to issue warnings.
2. If the Avamar server reaches the healthcheck limit, all new backup activity is suspended until you acknowledge this event.
3. If the Avamar server transitions to a read-only state, you must contact EMC Technical Support. You will be expected to delete backups, change retention policies, and suspend backups for at least two days while the system performs an aggressive garbage collect operation.
4. 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.

5. 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 data nodes, add two nodes at a time, rather than adding just one node, to noticeably reduce the capacity per node.

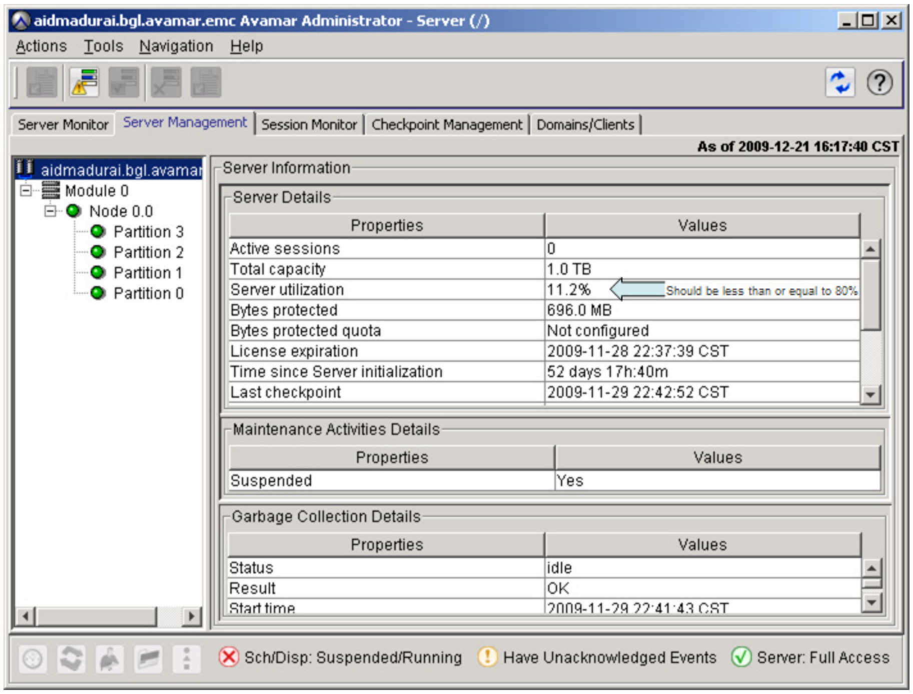
Steady State System

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 percent:



As of 2009-12-21 16:17:40 CST

Server Details

Properties	Values
Active sessions	0
Total capacity	1.0 TB
Server utilization	11.2%
Bytes protected	696.0 MB
Bytes protected quota	Not configured
License expiration	2009-11-28 22:37:39 CST
Time since Server initialization	52 days 17h:40m
Last checkpoint	2009-11-29 22:42:52 CST

Should be less than or equal to 80%

Maintenance Activities Details

Properties	Values
Suspended	Yes

Garbage Collection Details

Properties	Values
Status	idle
Result	OK
Start time	2009-11-29 22:41:43 CST

SCHEDULING

This planning and design chapter focuses on scheduling activities, important steps in designing, and setting up a new Avamar system. This chapter discusses several of these activities, including:

- Avamar server maintenance activities
- Backups

Avamar Client Details

Avamar's 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 up a work order (for example, a backup), the Avamar Administrator Server pages the client avagent. If the client is nonpageable (that is, 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.

If the Avamar Administrator Server queues a work order for the client, then 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 with 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. Refer to the most recent release notes and documents listed in *Assumptions and References* (page 9) for updates to this information.

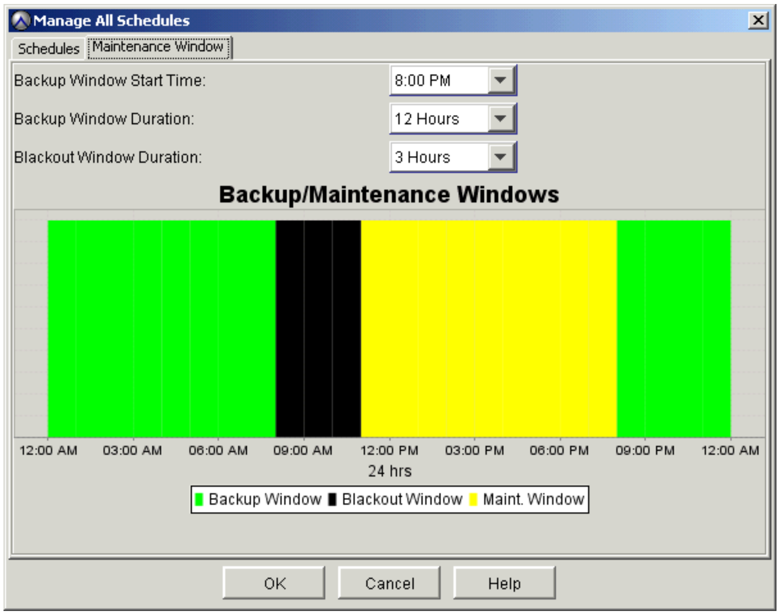
RESTRICTIONS AND LIMITATIONS	IMPACT
Carefully review the client support matrix with the presales technical engineer.	Ensure that clients and applications you want to protect with Avamar software are fully supported. In particular, verify that the specific details of the deployment, such as revisions, clusters, third-party plugins and add-ons, are supported.
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 filesystem 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.

RESTRICTIONS AND LIMITATIONS	IMPACT
2 to 5 TB of file server data per Avamar client	Backup scheduling could be impacted when an Avamar client is a file server 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.

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:



Manage All Schedules

Backup Window Start Time: 8:00 PM

Backup Window Duration: 12 Hours

Blackout Window Duration: 3 Hours

Backup Window. 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

Both backup window start time and duration can be customized to meet your specific site requirements.

Blackout Window. The portion of each day reserved for performing server maintenance activities (primarily checkpoint and garbage collection) that 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

Blackout window duration can be customized to meet your 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, will extend the maintenance window duration one hour because it begins one hour earlier. The backup window is not affected.

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

Maintenance Window. The portion of each day reserved for performing routine server maintenance activities (primarily checkpoint validation).

Operational Impact

There might be brief periods when backup or administrative activities will not be allowed. Although backups can be initiated during the maintenance window, doing so will impact both the backup and maintenance activities. For this reason, you should 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. This 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. That is, maintenance starts immediately after the blackout window and continues until the backup window start time.

Replication — Operational Impact

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

- All other maintenance jobs can start.
- All backup work orders will be 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 will be 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 will have 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, since 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:

YES 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.

YES Use the default maintenance window schedule and do not deviate from this schedule unless absolutely necessary.

YES 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.

For example, if a customer wants to back up clients from around the globe, it might be best to set up 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 do the following:

- 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 carefully observing the following limitations:

- Maximum number of files per client.
- Maximum amount of database data per client.
- Maximum amount of data per file server.

Typically, 80 to 90 percent of the clients will, on a daily basis, complete backups within the first hour or two of the backup window. This is the reason to schedule 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 percent 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.

DEFINING DOMAINS, GROUPS, AND POLICIES

This planning and design chapter describes initial backup management policy decisions that must be made after the overall daily schedule has been defined. These decisions include:

- 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?
- How should retention policies be set up to retain the backup data for the required period?
- When should backups be scheduled?
- How long should client backups be allowed to run?

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 is known as hierarchical management.

Another possible reason for segregating clients by domain is 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, then you should register all clients in the /clients domain.

Best practice:

YES ✓ Minimize the number of domains you create. Ideally, you will 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), and retention policy.

Best practices for defining groups:

YES Minimize the number of groups you define. Each dataset policy can include separate dataset definitions for various client plugins. In other words, 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.

YES 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, then any clients that are activated with the Avamar server will automatically be 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

Best practices for defining datasets:

YES Minimize the number of datasets required.

NO In general, do not attempt to back up a large client by defining multiple subsets of data that run every night. This 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. Refer to *Tuning Client Caches to Optimize Backup Performance* (page 39) for additional information.

Defining Schedules

The default schedule runs nightly during the server’s backup window. Depending on the amount of data in your 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 activities on a daily basis.

Best practices for defining schedules:

YES 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.

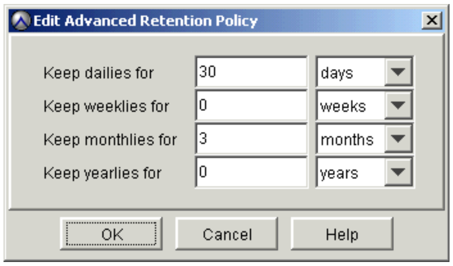
YES Minimize the number of clients that need to be backed up outside of the server's backup window. When setting up backup schedules, remember that mobile laptop clients might need to be scheduled to back up during the day when they are connected to the network. The system can handle a small number of exceptions. In this case, you will want to overlap the backup of this handful of exception clients with the server’s maintenance window.

Defining Retention Policies

Best practices for setting up retention policies:

YES 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.



Edit Advanced Retention Policy

Keep dailies for	30	days
Keep weeklies for	0	weeks
Keep monthlies for	3	months
Keep yearlies for	0	years

For example, if you retain three months of backups by keeping 30 days of daily backups, and by keeping three months of monthly backups, the total amount of data stored on the Avamar server for the clients is equivalent to the initial unique data plus 42 equivalent days of backups. This requires less back-end capacity than storing three months of daily backups, which is equivalent to the initial unique data plus 91 equivalent days of backups. The *Avamar System Administration Guide* contains more information about advanced retention policies.

YES

Set the minimum retention period to at least 14 days. Remember that, when selecting the maximum retention period, the Avamar server does not retain the last unexpired backup. Therefore, if the retention period is relatively short (for instance, 7 days), you must monitor the backup operations closely enough to ensure that the last unexpired backup does not expire before the system completes another backup. If all backups for a client expire before you correct the issue that prevented the client from completing a backup, the next backup will be equivalent to an initial backup.

Daily Monitoring of Backup Infrastructure

This daily operations chapter focuses on a number of Avamar features and functions that generate notifications when specific events occur. 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.

Monitoring the Avamar System

You should monitor the event notification system for warning and error events every day.

Best practice:

YES ✓ Monitor the Avamar server on a daily basis and understand how to interpret all system warnings and errors.

The following table describes possible ways to monitor Avamar systems:

METHOD	DESCRIPTION
syslog or SNMP event notification	If your network management infrastructure supports syslog or SNMP event notification, enable the syslog or SNMP event notification subsystem through Avamar Administrator. Refer to the *Avamar System Administration Manual* for instructions on enabling syslog or SNMP notification.
Email notification system	You can set up email notification in either of the following ways: • To batch email notifications that are sent twice daily according to the default notification schedule. • To send emails as the selected events occur.

METHOD	DESCRIPTION
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**.
Unacknowledged events	At least once a day, review and clear any Unacknowledged Events queued on the Avamar Administrator > Administration > Event Management tab > Unacknowledged Events tab. 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 on the Avamar Administrator > Administration > Event Management tab > Event Monitor tab.

DAILY MONITORING OF BACKUP OPERATIONS

This daily operations chapter focuses on a number of Avamar features and functions that generate notifications when specific backup, restore, or replication operation events occur. 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.

Monitoring the Avamar System Backup Operations

You should monitor the event notification system for warning and error events related to backup operations every day.

Best practice:

YES ✓ Monitor the Avamar Activity Monitor on a daily basis 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 status messages typically associated with these failures are the following:

- 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 the 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 as follows:

  - Somebody rebooted the client in the middle of the backup.
  - A network communication outage lasted longer than one hour.

  The Administrator server will automatically queue 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 will continue to fail with the same status, unless some change is made, such as tuning the client caches.

The 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. This tool is described in the *Avamar System Administration Guide*.

Avamar Administrator can email reports that you can use to monitor clients that failed backup or completed backup with exceptions. The following client reports are helpful:

- 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.

Refer to the *Avamar System Administration Guide* for descriptions of these and other reports that are available.

Best practice:

YES ✓ Monitor backups every day and investigate all failed backups, missing clients, and backups that completed with exceptions.

YES ✓ Enable the advanced statistics report during all backups. This information is useful for addressing performance issues.

YES ✓ Enable debugging messages when investigating backup or restore failures.

YES ✓ 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.

TUNING PERFORMANCE

This implementation and daily operations chapter focuses on important Avamar system tuning activities, such as:

- Optimizing backup performance
- Setting up and using replication
- Tuning caches

Refer to *Managing Capacity* (page 18) for additional information.

IMPORTANT: 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 4 hours every night.

A typical backup should take about 1 hour for every million files in a file server or about 1 hour for every 100 GB of data in a database server. If backups take more than 30 percent 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.

Tuning Client Caches to Optimize Backup Performance

This section describes how the client caches work, and how you can tune the client caches appropriately to optimize backup performance. This section also describes best practices that should be considered when setting up and installing clients.

The Avamar client process (avtar) loads two cache files into memory when performing a backup. These client caches are used to:

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

Client Caches

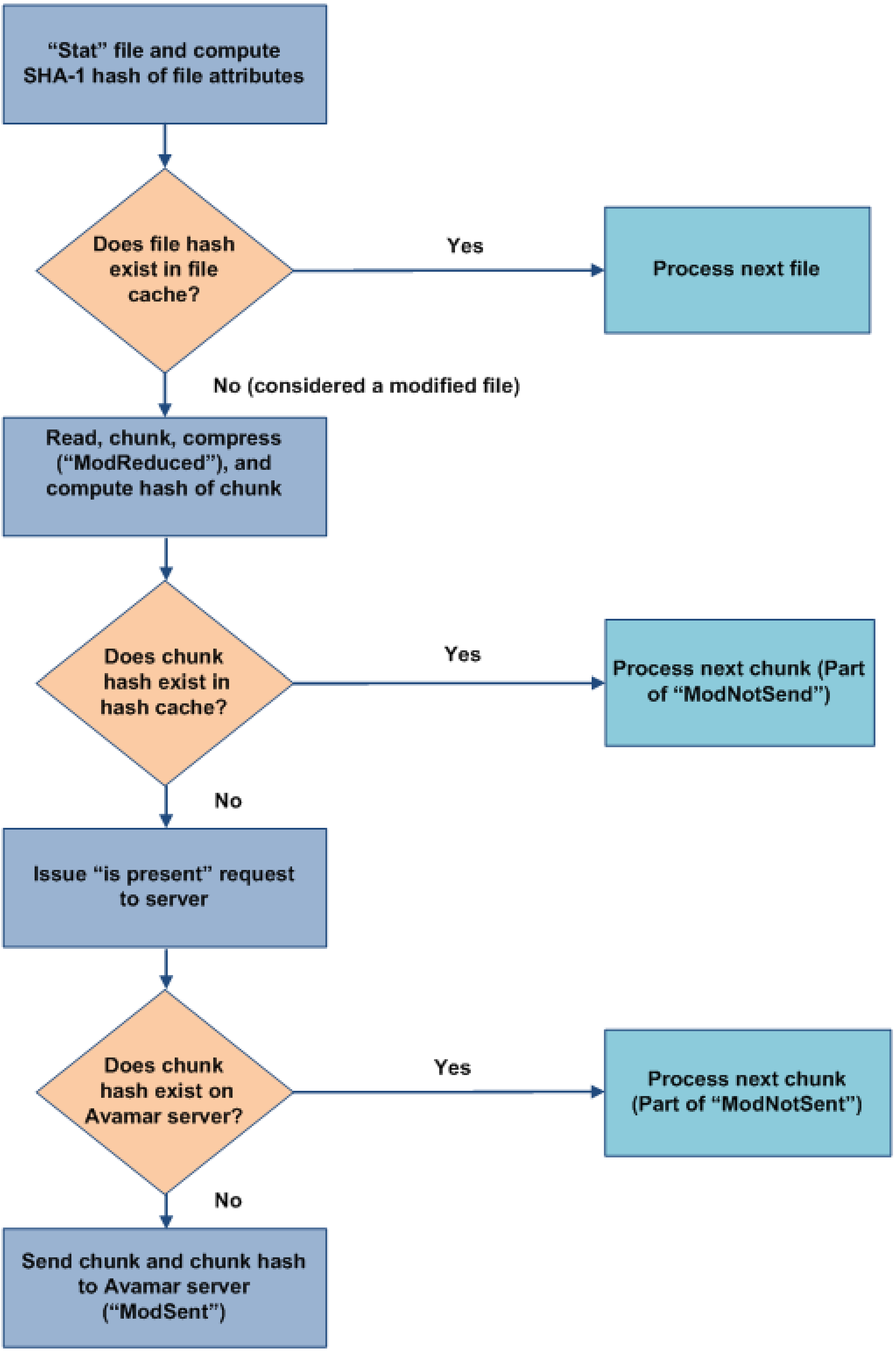
At the beginning of a backup, the avtar process loads two cache files from the var directory into memory. The var directory is found in the Avamar installation path.

The first of the cache files is the file cache (f_cache.dat). The file cache 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 percent of the files. When backing up databases, however, the file cache is not effective since all the files in a database appear to be modified every day.

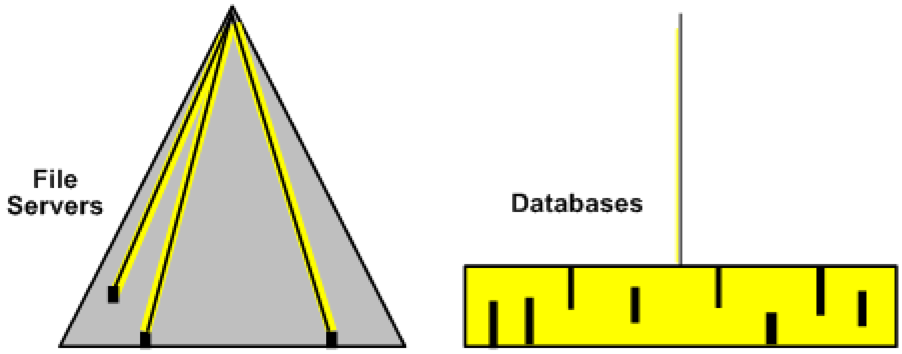
The second cache is the hash cache (p_cache.dat). The hash cache stores the hashes of the chunks and composites that have been sent to the Avamar server. The hash cache is used to quickly identify which chunks or composites have previously been backed up to the Avamar server. The hash cache is very important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

Overview of Cache Operation

The following flowchart shows the process that avtar uses to filter out previously backed-up files and chunks. This image shows the values that are incremented and reported in the avtar advanced statistics option (for instance, filebytes, filecache, and so forth).


"Stat" file and compute SHA-1 hash of file attributes
Does file hash exist in file cache?
Yes
Process next file
No (considered a modified file)
Read, chunk, compress ("ModReduced"), and compute hash of chunk
Does chunk hash exist in hash cache?
Yes
Process next chunk (Part of "ModNotSend")
No
Issue "is present" request to server
Does chunk hash exist on Avamar server?
Yes
Process next chunk (Part of "ModNotSent")
No
Send chunk and chunk hash to Avamar server ("ModSent")

The following figure shows the effectiveness of the file cache when backing up file servers in contrast to databases. In file servers, the file cache is typically 98 percent effective in filtering previously backed-up files.



Note: Yellow area represents data read during backup

- Assume 1 TB file system
- Find modified files
  - File cache ~98% effective reducing client load
  - 980+ GB screened out
- Read, chunk, compress and hash modified files (20 GB) to find unique chunks
- Send 0.3% (3 GB) to Avamar server

- Assume 1 TB database
- Find modified files
  - File cache 0% effective in reducing client load
  - 0 GB screened out
- Read, chunk, compress and hash modified files (1 TB) to find unique chunks
- Send 3% (30 GB) to Avamar server

File Cache

If the file cache is deleted, unused, or is 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. If a file hits in the file cache, then 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 will be limited to 4 GB/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. Since the avtar program is a 32-bit application, the maximum file cache size that can be used 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 will be 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: filename, filepath, 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.

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

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 *File Cache* (page 41), the hash cache will be 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 will be 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.

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

Impact of Caches on Memory Consumption

Some customers 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 (shown in bold) has room to grow:

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

But the hash cache (shown in bold) 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 is very important because the size of the cache has a huge influence on the overall performance of the system.

Changing the Maximum Cache Sizes

You can override the default limits on the size of the file and hash caches by using the following two options with the avtar command:

```
--filecachemax=VALUE
```

```
--hashcachemax=VALUE
```

where VALUE is an amount in megabytes or a fraction (a negative value is a fraction of RAM).

Default values:

```
--filecachemax=-8
```

```
--hashcachemax=-16
```

As another example, the following option limits the file cache to 100 MB. Because the file cache doubles in size every time it needs to grow, the file cache actually increases to a maximum of 88 MB if the following option is set:

```
--filecachemax=100
```

If you decide to limit either of the two cache sizes to a limit lower than the default, and if the cache size is already beyond your specified value, you need to delete the cache for the new limit to become effective. Cache sizes increase monotonically. There is no way to shrink the cache files without deleting them and building them back to a new limit.

Another implementation consideration is that if you decide to limit the cache size on a set of clients, you should add the appropriate parameters to each client's avtar.cmd file. Limits are applied every time the client performs a backup, even an on-demand backup. If you exclude the flag in the avtar.cmd file, and an on-demand backup occurs without the appropriate option, the file cache or hash cache could increase to the default values.

Tuning the Maximum Cache Sizes

To optimize performance, sometimes you need to increase the cache sizes from the default values. These conditions could exist in the following two opposed cases: millions of small files and large databases.

Millions of Small Files. If the client has millions of small files, then you might need to increase the file cache from the default size. Generally, for every one million files on the Avamar client, the client requires 512 MB of physical RAM. If a client has one million files, then, by using the formula in *File Cache* (page 41), a minimum of 44 MB (1 x 44 MB) is required just to store all the file hashes for a single backup. Since the file hashes must be stored for several backups, more than 44 MB is required.

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.

By default, since one-eighth of the physical RAM of 512 MB is used, cache can increase to a limit of 64 MB, which means that the default value of one-eighth of RAM for the file cache is adequate.

You must be particularly alert to this situation when configuring a Windows client to back up a large file server (for example, a NAS filer) through a network mount.

Large Databases. If the client has a few large files, then usually the default of one-sixteenth for the hash cache is insufficient. For example, for a 240 GB database, a minimum of 240 MB is required when using the formula in *Hash Cache* (page 42). This amount can only store the hashes for a single backup.

The hash cache also doubles in size each time it needs to grow. The current hash cache sizes are in megabytes: 24 MB, 48 MB, 96 MB, 192 MB, 384 MB, 768 MB, and 1,536 MB.

The next increment available is 384 MB. Therefore, if this client has 4 GB of RAM, the hash cache must increase to one-eighth of the RAM. If the default of one-sixteenth of the RAM is used, the hash cache will be limited to 192 MB, and an undersized hash cache will result. In the case of databases, since very few files are backed up, the file cache will be considerably smaller, so the net memory use is still about one-eighth to three-sixteenth of the RAM.

Rules for Tuning the Maximum Cache Sizes

The most important rule is to ensure that the caches do not grow so large that the client ends up swapping (excessive movement of memory pages between RAM and disk) because it has insufficient physical RAM to handle all the processes.

Best practice:

YES 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 percent), and set the other cache to be -32 (3 percent). 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 your 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, then remember the following:

- 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

The *File Cache* (page 41) section makes the following assertions:

- 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.

Therefore, 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

The *Hash Cache* (page 42) section makes the following assertions:

- 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.

Therefore, 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

If the 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. To do so:

- Break the client file system into multiple smaller datasets.
- For each dataset, ensure that the maximum file and hash caches assign a unique `cacheprefix` attribute.

Example

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. For the volume with 2.5 million files, a file cache of at least 110 MB (2.5 x 44 MB) is required. The next increment that accommodates this is 176 MB. The other datasets could be defined as follows:

- 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 **Attribute = cacheprefix**, and **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.

Customizing Maximum Hash Cache Settings for Exchange and SQL Servers

For an Exchange Server database backup, configure the maximum hash cache in the dataset definition by setting the following:

- **`Attribute = hashcachemax`**
- **`Attribute Value = 200`**

For a SQL Server database backup, configure the maximum hash cache in the dataset definition by setting the following:

- **`Attribute = [avtar]hashcachemax`**
- **`Attribute Value = 200`**

Tuning Replicator

Work with your EMC Practice Consultant to configure and tune the replicator. The Practice Consultant will perform the following tasks:

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

UNDERSTANDING DPN SUMMARY REPORTS

This chapter describes usage of the DPN Summary report.

To access the DPN Summary report:

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

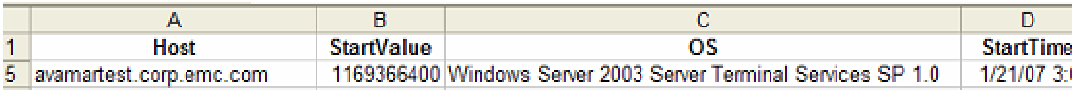
Use this report to determine how well an Avamar system performs once it has achieved steady state. The Avamar system can generate this report that can be used to determine:

- 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

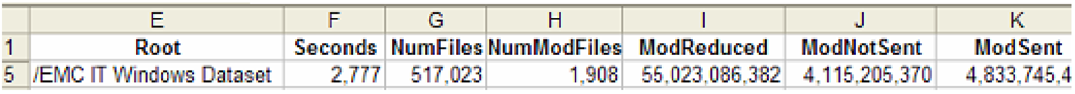
The purpose of this section is to help you understand the information in the DPN Summary report.

Example DPN Summary Entry

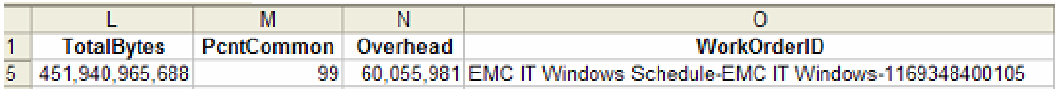
This example uses the following excerpt from a DPN Summary report:



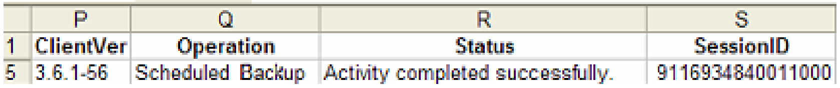
	A	B	C	D
1	Host	StartValue	OS	StartTime
5	avamartest.corp.emc.com	1169366400	Windows Server 2003 Server Terminal Services SP 1.0	1/21/07 3:



	E	F	G	H	I	J	K
1	Root	Seconds	NumFiles	NumModFiles	ModReduced	ModNotSent	ModSent
5	/EMC IT Windows Dataset	2,777	517,023	1,908	55,023,086,382	4,115,205,370	4,833,745,4



	L	M	N	O
1	TotalBytes	PcntCommon	Overhead	WorkOrderID
5	451,940,965,688	99	60,055,981	EMC IT Windows Schedule-EMC IT Windows-1169348400105



	P	Q	R	S
1	ClientVer	Operation	Status	SessionID
5	3.6.1-56	Scheduled Backup	Activity completed successfully.	9116934840011000

Definitions of each DPN Summary column follow. Refer to *Background on Backups* (page 52) for additional information.

Host. The client hostname as defined in DNS. During backups, this is the client that backs up data to the Avamar server. During restores, this is the client that receives the restored data. 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 UTC (or 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 this activity (less those files that were excluded through exclusion rules).

NumModFiles. The total number of modified files associated with this activity.

ModReduced. Refer to *Background on Backups* (page 52).

ModNotSent. Refer to *Background on Backups* (page 52).

ModSent. Refer to *Background on Backups* (page 52).

TotalBytes. Refer to *Background on Backups* (page 52).

PcntCommon. This is the commonality percentage during this activity.

Overhead. The number of bytes for COMPOSITEs and DIRELEMs used to store data. It is the amount of nonfile data sent by the client to the server, and includes such things as indexing information, requests from the client to the server for the presence of specific data chunks, ACLs, directory information and message headers. On any relatively active file system, this is generally a very small percentage of the file data that is sent to the server.

WorkOrderID. The unique identifier for the following activities:

- For scheduled backups, the work order ID is formatted as: <Schedule name>-<Group name>-<Unix time in ms>.
- For on-demand backups initiated from the Policy window **Back Up Group Now** command, the work order ID is formatted as: <Group name>-<Unix time in ms>.
- For on-demand backups or restores initiated from the **Backup and Restore** window, the work order ID is formatted as: MOD-<Unix time in ms>.
- For on-demand backups initiated from the systray icon of the Windows Avamar client, the work order ID is formatted as: COD-<Unix time in ms>.
- For command-line backups or restores, the work order ID is formatted as: NAH-<Unix time in ms>.
- For replication activities, the work order ID is formatted as: COD-NAH-<Unix time in ms>.

ClientVer. The Avamar client software version, where the Avamar client is as defined previously under **Host**.

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

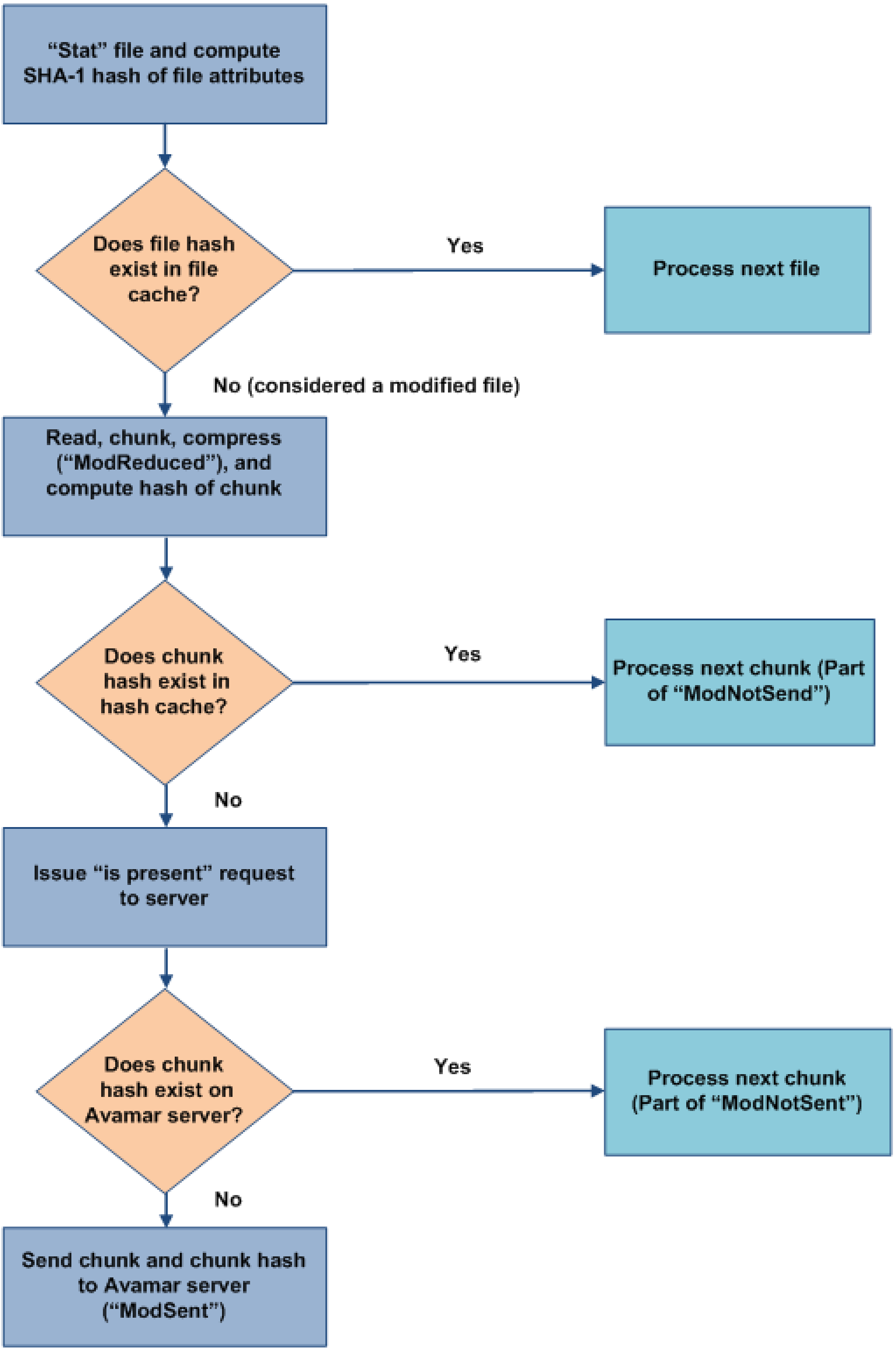
This section provides background information about how the Avamar client performs backups, including key statistics.

Dataset Size. Begin with the dataset size (TotalBytes - column L) equal to 451,940,965,688 bytes (421 GB). This is the total bytes in the dataset. Any files that are excluded due to exclusion rules are not counted in this total. However, this total includes those 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 it 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 (that is, 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 flowchart 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.


"Stat" file and compute SHA-1 hash of file attributes
Does file hash exist in file cache?
Yes
Process next file
No (considered a modified file)
Read, chunk, compress ("ModReduced"), and compute hash of chunk
Does chunk hash exist in hash cache?
Yes
Process next chunk (Part of "ModNotSend")
No
Issue "is present" request to server
Does chunk hash exist on Avamar server?
Yes
Process next chunk (Part of "ModNotSent")
No
Send chunk and chunk hash to Avamar server ("ModSent")

ModReduced. When the modified files are backed up, the data is chunked, compressed, and then hashed. Since the compression takes place on the client, the amount that the data was compressed is reported as ModReduced. In the example row, the ModReduced = 55,023,086,382 (51 GB).

ModNotSent. If subfile level commonality exists, then the data is not sent. Therefore, 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. In the example row, ModNotSent = 4,115,205,370 (3.8 GB), which means that 3.8 GB of compressed chunks were already on the Avamar server.

ModSent. If 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

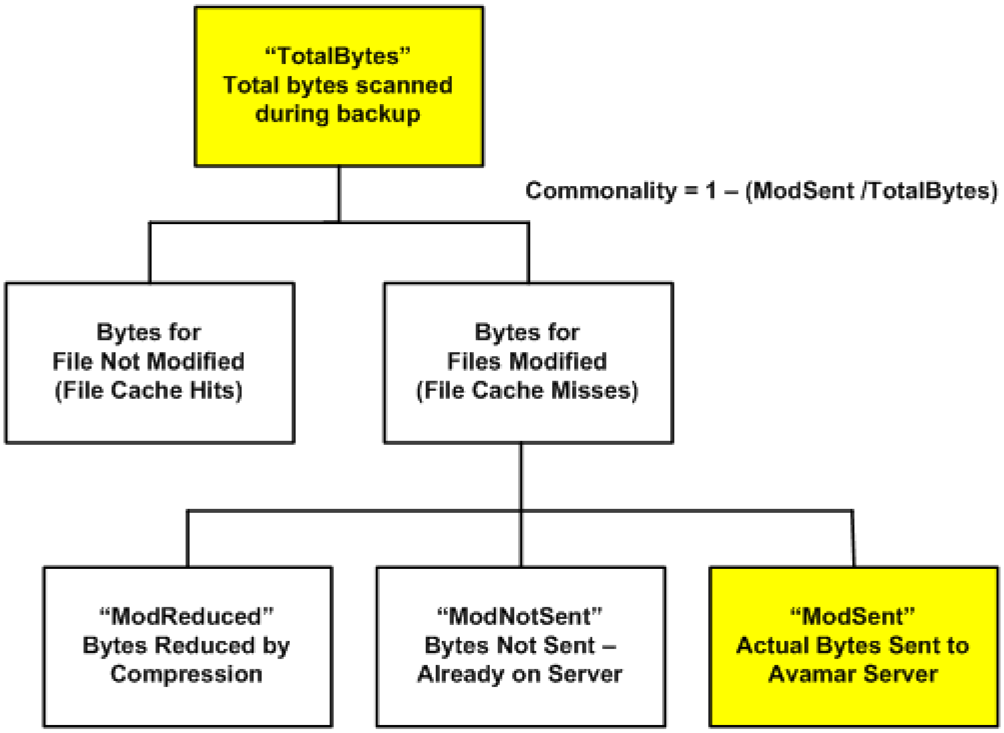
The following describes the relationships between terms:

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 following diagram.


"TotalBytes"
Total bytes scanned during backup
Commonality = 1 – (ModSent /TotalBytes)
Bytes for File Not Modified (File Cache Hits)
Bytes for Files Modified (File Cache Misses)
"ModReduced" Bytes Reduced by Compression
"ModNotSent" Bytes Not Sent – Already on Server
"ModSent" Actual Bytes Sent to Avamar Server

Definition of Commonality

Avamar change rate is equivalent to the ModSent / TotalBytes.

The Avamar commonality = 1 - (change rate).

Data Moved During 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 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, the Subtotal bytes for files modified since previous backup = 51 GB + 3.8 GB + 0.37 GB = 55 GB. If this is divided by TotalBytes, the result is 55 / 421 = 13%.

Typically, in day-over-day backups of fileservers, this value is expected to be in the 2 percent range.

When backing up databases, expect this value to be 100 percent 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.

PROTECTING AVAMAR DESKTOP/LAPTOP CLIENTS

This chapter about implementation and daily operations focuses on best practices for Avamar environments with Avamar Desktop/Laptop clients. Avamar Desktop/Laptop is client/server software.

- The Desktop/Laptop client is a feature that is installed as part of an Avamar Windows Client or Avamar Mac OS X Client installation.
- All code necessary to support Desktop/Laptop is automatically installed when you install the Avamar 5.0 server.

Avamar Desktop/Laptop 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.

This chapter includes the following best practices for environments in which Avamar Desktop/Laptop has been deployed:

YES ✓ *Deploying Additional Avamar Servers for Desktop/Laptop Clients* (page 57)

YES ✓ *Creating a Dataset to Back Up Only User Data* (page 57)

YES ✓ *Keeping the Initial Client Backups to a Manageable Number* (page 60)

YES ✓ *Determining the Backup Window* (page 61)

YES ✓ *Scheduling Backups to Complete within the Backup Window* (page 62)

NO ⊘ *Running System Utilities During the Backup Window* (page 62)

YES ✓ *Running Backups More Frequently Than the Retention Policy* (page 62)

YES *Ensuring Adequate Initialization Time for Wake-on-Lan Backups* (page 63)

YES *Adjusting Runtime for Daily Maintenance Tasks* (page 63)

Deploying Additional Avamar Servers for Desktop/Laptop Clients

When deploying Avamar Desktop/Laptop to a location with existing Avamar servers, use an additional grid to support the desktop and laptop clients. Backups of Desktop/Laptop clients must be run when users are online (usually during the day). Refer to *Adjusting Runtime for Daily Maintenance Tasks* (page 63) for more information. Scheduled backups for file servers and database clients normally run during the night. *Scheduling Activities During the Course of a Day* (page 25) contains more information about backup windows, blackout windows, and maintenance windows.

Creating a Dataset to Back Up Only User Data

The use of Avamar Desktop/Laptop to back up end-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.

Create a backup dataset that specifies the files and folders for the backup. The following table contains pathnames for Windows XP, Windows 7, and Vista, and MAC:

OS	TAB	FILES AND FOLDERS
Windows XP	Source Data	C:\Documents and Settings
	Exclusions	C:\Documents and Settings*
	Inclusions	C:\Documents and Settings*\Desktop C:\Documents and Settings*\My Documents C:\Documents and Settings*\Favorites

OS	TAB	FILES AND FOLDERS
Windows 7, Vista	Source Data	C:\Users
	Exclusions	C:\Users\*
	Inclusions	C:\Users\*\Desktop C:\Users\*\Documents C:\Users\*\Favorites
Mac	Source Data	/Users
	Exclusions	/Users/*
	Inclusions	/Users/*/Desktop /Users/*/Documents /Users/*/Library/Favorites

Additional Exclusions for Windows Computers

In addition to the files and folders specified in the table, exclude the following files and folders from Windows desktop and laptop backups:

- 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

- Other nonbusiness files such as, personal music, pictures, video, and so forth:

  *.avi

  *.mp3

  *.mp4

  *.mpeg

  *.jpeg

  *.wma

  *.wmv

Dataset Caveat

In an environment that contains Windows XP desktop or laptop computers along with Vista or Windows 7 desktop or laptop computers, backups can appear to fail if both Windows XP clients and Vista or Windows 7 clients use a single dataset that specifies the C:\Documents and Settings folder and the C:\Users 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 Vista or Windows 7 clients, add the `--x18=256` flag to the dataset to prevent the `Path not found` error.

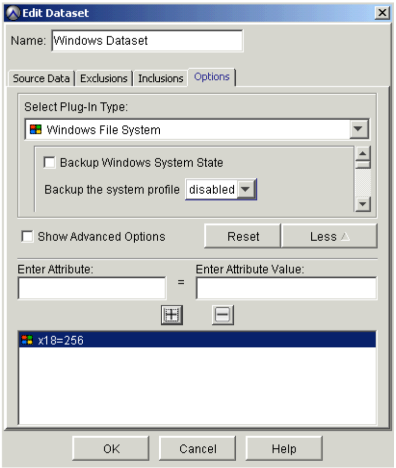
To add the --18=256 flag to the dataset, perform the following:

1. Start Avamar Administrator.
2. Select **Tools** > **Manage Datasets**.

   The Manage All Datasets dialog box appears.
3. Select the dataset from the list and click **Edit**.

   The Edit Dataset dialog box appears.
4. Click **Options** and select the plug-in from the Select Plug-In Type list.
5. Click **More**.
6. Type **--x18** in the Enter Attribute text box.
7. Type **256** in the Enter Attribute Value text box.
8. Click ⊞.

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



Name: Windows Dataset

Select Plug-In Type: Windows File System

☐ Backup Windows System State

Backup the system profile: disabled

Enter Attribute: = Enter Attribute Value:

x18=256

9. Click **OK** twice.

Keeping 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.

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.

To minimize the impact of first-time backups, bring clients online in smaller groups. Use the first few groups to provide information about the capabilities of your 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.

As the amount of data archived in an Avamar system increases, the benefit of deduplication increases. This means that throughput problems will 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 your 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

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 Activation 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 your 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 your network.

Determining the Backup Window

Avamar Desktop/Laptop environments typically include more clients than traditional Avamar systems. The Avamar Administrator server allows up to an average of 18 concurrent backup connections per active storage node (less one for the overall grid). 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.

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

Example 1 Backup criteria:

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

Formula:

10 nodes x 6 backups per node x 18 concurrent connections = 1,080 backups per hour

Example 2 Backup criteria:

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

Formula:

10 nodes x 2 backups per node x 18 concurrent connections = 360 backups per hour

Use the number of backups per hour to determine the backup window, which is the total amount of time required to back up all desktop and laptop computers.

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 a backing up the same computer when it is connected to the corporate network by an 802.11G wireless connection.

Scheduling Backups to Complete within 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 end 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, multiple Avamar servers might be required to back up all clients.

Running System Utilities During the Backup Window

Avoid running multiple system utilities on the end-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.

Running 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 set. 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 set.

Refer to the following resources for more information on retention policies:

- Avamar System Administration Guide
- *Defining Domains, Groups, and Policies* (page 29)

Ensuring 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. To ensure that system initialization completes before a scheduled backup starts:

- 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.

Managing 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 percent 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. Refer to *Steady State System* (page 22) for 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. Refer to *Avamar Capacity Thresholds* (page 19) for more information.

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

- Clients backing up more than the initial backup size limit
- Clients exceeding the daily data change rate limit
- Garbage collection failures

Extensive information about capacity management is provided in the *Avamar System Administration Guide*. In that document you will find detailed descriptions of the various features, functions, and tools available to assist you with properly monitoring and managing your server storage capacity.

Adjusting Runtime for Daily Maintenance Tasks

It is important that Avamar daily maintenance tasks complete successfully every day. Failures of these tasks will quickly result 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 must 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. Refer to *Scheduling Activities During the Course of a Day* (page 25) for more information.

The daily maintenance task of garbage collection requires a quiet system. If any backups are running, the task will not start. Since garbage collection reclaims space on the Avamar server, the failure to successfully complete this task can quickly create capacity problems.

OTHER AVAMAR ADMINISTRATION BEST PRACTICES

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

- Passwords
- Internet and firewalls
- Email home

Protecting Your Avamar Server

Deploy an Avamar server in a protected network and not exposed to the Internet. Even when an Avamar server is deployed in an internal network, the server should be protected by a firewall to prevent unauthorized access to the nodes that compose the server.

Best practice:

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

Changing Passwords

If your Avamar passwords have not changed from the factory default values, change them by running the **`change-passwords`** utility. Refer to your *Avamar System Administration Guide* for additional information.

There are three operating system users (root, admin, and dpn), two SSH keys (admin_key and dpnid), and three root-level software application users (root, MCUser, and repluser). You must change the passwords for all of these users. Just changing the passwords for the operating system users is not sufficient to prevent someone from logging in to the Avamar server nodes. If you have not changed the two SSH keys, then someone could use the factory-default SSH keys to log in to your Avamar server.

Best practice:

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

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 Technical Support once daily, and critical alerts in near-real time on an as-needed basis.

Enable this feature on your servers. Refer to your *Avamar System Administration Guide* for additional information.

The email home capability 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 the email home messages. The customer must assume primary responsibility for monitoring the Avamar systems.
2. Only issues that affect the backup infrastructure (Avamar server) will have support cases automatically opened (for instance, failed hfscheck). Cases are not opened for issues associated with backup operations (for instance, failed backups).
3. If email home messages are not being sent home, EMC Technical Support does not proactively alert customers of an issue that is causing the email home messages not to be sent. This could be because the server is down, because the scheduler has been disabled, and so forth.

Best practice:

YES Enable the email home capability.

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

To fully leverage the audit logging capability, follow these best practices:

YES 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.

YES Work with your EMC Corporation Practice Consultant or EMC Technical Support Engineer 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.

INDEX