IBM TotalStorage SAN File System
(based on IBM Storage Tank™ technology)

Planning Guide

Version 2 Release 2

GA27-4344-01

IBM TotalStorage SAN File System
(based on IBM Storage Tank™ technology)

Planning Guide

Version 2 Release 2

GA27-4344-01

Note

Before using this information and the product it supports, read the information in ″Notices.″

Second Edition (November 2004)

This edition applies to the IBM TotalStorage SAN File System and to all subsequent releases and modifications until otherwise indicated in new editions.

Order publications through your IBM representative or the IBM branch office servicing your locality. Publications are not stocked at the address below.

IBM welcomes your comments. A form for reader's comments is provided at the back of this publication. If the form has been removed, you may address your comments to:

International Business Machines Corporation
Design & Information Development
Department CGFA
PO Box 12195
Research Triangle Park, NC 27709–9990
U.S.A.

You can also submit comments by selecting Feedback at www.ibm.com/storage/support/.

When you send information to IBM, you grant IBM a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Contents

About this guide

This topic provides information about the contents of the *Planning Guide*.

This guide provides information that you can use to plan for the implementation of the IBM® TotalStorage® SAN File System.

Who should use this guide

This topic describes the audience for the *Planning Guide*.

This guide is intended for people who will be involved in planning for the installation of the SAN File System. The audience should have planning experience and skills in the following areas:

- Networking and network management
- Storage management
- SAN management
- Critical business issues, such as backup, disaster recovery, and security

The installer of SAN File System software should meet the following requirements:

- Knowledge and training in the technology of SAN File System and its functions
- Familiarity with the hardware on which the SAN File System will be installed
- Awareness of the procedures in this document
- Awareness of related installation and service publications

Notices in this guide

This topic describes the notices in the Installation and Configuration Guide.

The following notices are contained with the this guide and convey these specific meanings:

Note: These notices provide important tips, guidance, or advice.

Attention: These notices indicate possible damage to programs, devices, or data. An attention notice appears before the instruction or situation in which damage could occur.

CAUTION:
These notices indicate situations that can be potentially hazardous to you. A caution notice appears before the description of a potentially hazardous procedure step or situation.

DANGER

These notices indicate situations that can be potentially lethal or extremely hazardous to you. A danger notice appears before a description of a potentially lethal or extremely hazardous procedure step or situation.

Publications

This topic describes the publications in the SAN File System library and in related libraries.

SAN File System publications

This topic describes the publications in the SAN File System library.

The following publications are available in the SAN File System library. They are provided in softcopy on the *IBM TotalStorage SAN File System Publications CD* and at www.ibm.com/storage/support. To use the CD, insert it in the CD-ROM drive. If the CD does not launch automatically, follow the instructions on the CD label.

Note: The softcopy versions of these publications are accessibility-enabled for the IBM Home Page Reader.

- *IBM TotalStorage SAN File System Release Notes*

  This document provides any changes that were not available at the time the publications were produced. This document is available only from the technical support Web site: www.ibm.com/storage/support
- *IBM TotalStorage SAN File System Software License Information*

  This publication provides multilingual information regarding the software license for IBM TotalStorage SAN File System Software.
- *IBM TotalStorage SAN File System Administrator's Guide and Reference*, GA27-4317

  This publication introduces the concept of SAN File System, and provides instructions for configuring, managing, and monitoring the system using the SAN File System console and administrative command-line interfaces. This book also contains a commands reference for tasks that can be performed at the administrative command-line interface or the command window on the client machines.
- *IBM TotalStorage SAN File System Basic Configuration for a Quick Start*, GX27-4058

  The document walks you through basic SAN File System configuration and specific tasks that exercise basic SAN File System functions. It assumes that the physical configuration and software setup have already been completed.
- *IBM TotalStorage SAN File System Maintenance and Problem Determination Guide*, GA27-4318

  This publication provides instructions for adding and replacing hardware components, monitoring and troubleshooting the system, and resolving hardware and software problems.

  Note: This document is intended only for trained support personnel.
- *IBM TotalStorage SAN File System Installation and Configuration Guide*, GA27-4316

  This publication provides detailed procedures to set up and cable the hardware, install and upgrade the SAN File System software, perform the minimum required configuration, and migrate existing data.
- *IBM TotalStorage SAN File System Messages Reference*, GC30-4076

This publication contains message description and resolution information for errors that can occur in the SAN File System software.

- *IBM TotalStorage SAN File System Planning Guide*, GA27-4344

  This publication provides detailed procedures to plan the installation and configuration of SAN File System.

- *IBM TotalStorage SAN File System System Management API Guide and Reference*, GA27-4315

  This publication contains guide and reference information for using the CIM Proxy API, including common and SAN File System-specific information.

  Note: This document contains information and procedures intended for only selected IBM Business Partners. Contact your IBM representative before using this publication.

SAN File System related publications

These publications are related to SAN File System.

- *IBM TotalStorage* Subsystem Device Driver User's Guide, SC26-7637

Web sites

This topic discusses any Web sites that offer additional, up-to-date information about SAN File System.

The following Web sites have additional information about SAN File System:

- www.ibm.com/storage/support/sanfs/
- www.ibm.com/storage/software/virtualization/sfs/

The following Web site has information about the languages that have International Components for UNICODE (ICU) converters: oss.software.ibm.com/cgi-bin/icu/convexp/

Summary of changes in release 2.2

This section describes the enhancements made to SAN File System in release 2.2.

The following list describes the technical changes and enhancements made to SAN File System for release 2.2.

- **Heterogeneous file sharing —** SAN File System now supports heterogeneous file sharing by implementing user maps that identify equivalent UNIX® and Windows® domain-qualified users.
- **Additional client platforms —** SAN File System supports these additional client platforms:
  - SUSE Linux™ Enterprise Server 8 (32-bit)
- **Installation enhancements** — SAN File System has many installation enhancements, including:
  - **Rolling upgrade** — You can upgrade your metadata servers from SAN File System release 2.1 to 2.2 with minimal disruption in service.
  - **Client upgrade —** Upgrading the SAN File System clients to SAN File System release 2.2 is optional.
- **Management enhancements —** SAN File System has many management enhancements, including:

  - **Non-disruptive file movement —** SAN File System allows you to move a file, along with any FlashCopy® images for that file, to a new storage pool without disruption to the reader.
  - **File defragmentation —** You can defragment an individual file in a storage pool.
  - **File management —** You can create a policy to move files to a new storage pool or delete files to improve the use and balance of premium and inexpensive storage throughout the life cycle of that file.
- **Serviceability enhancements —** SAN File System has many serviceability enhancements, including:
  - **Client tracing —** The tracing utility has been improved to provide a robust, lightweight, buffered, tracing mechanism that is common among all SAN File System client platforms. In addition, log messages are now duplicated in the tracing buffer when tracing is enabled to enhance problem determination.
- **Usability enhancements —** SAN File System has many usability enhancements, including:
  - **Windows client configuration —** You can now use the SAN File System plug-in for the Microsoft® Management Console to configure settings for the Windows clients. This eliminates the need to manually modify the registry.
  - **Metadata checker status indicator —** A progress indicator now shows the progress and estimates the completeness of the metadata checker process.
  - **LUN details —** The LUN details now include the world-wide node name (WWNN) and the world-wide port name (WWPN) associated with each LUN.
  - **Unicode characters —** SAN File System supports both uppercase and lowercase non-ASCII Unicode characters in file names. SAN File System policies also support non-ASCII Unicode characters.
  - **Multibyte character set (MBCS) —** SAN File System now supports MBCS in the names of files used in policies and directories in the global namespace to which filesets can be attached:
    - **File names —** Policies accept rules with file names that use MBCS. Rule terms that support MBCS include CHARACTER_LENGTH, CHAR_LENGTH, LEFT, LIKE, POSITION, RIGHT, SUBSTRING and TRIM, and wild cards for zero, one or more characters.
    - **Fileset attach points —** The global namespace now accepts directory names that use MBCS; however, the root fileset attach point (for example, /sanfs) must be in ASCII.

Chapter 1. Introduction to SAN File System

This topic introduces the SAN File System.

What is IBM TotalStorage SAN File System?

This topic provides a brief overview of IBM TotalStorage SAN File System.

IBM TotalStorage SAN File System is a storage area network (SAN)-based, scalable, and highly-available file system and storage management solution for file aggregation and concurrent data sharing in an open, multi-platform environment. It uses SAN technology, which allows an enterprise to connect a large number of heterogeneous computers and share a large number of heterogeneous storage devices over a high-performance network.

With SAN File System, heterogeneous clients can access shared data directly from large, high-performance, high-function storage systems, such as IBM TotalStorage Enterprise Storage Server® (ESS) and IBM TotalStorage SAN Volume Controller. SAN File System is built on a Fibre Channel network and is designed to provide superior I/O performance for data sharing among heterogeneous computers. It also provides growth capability and simplified storage management.

SAN File System differs from conventional distributed file systems in that it uses a data-access model that separates *file metadata* (information about the files, such as owner, permissions, and the physical file location) from actual *file data* (contents of the files). The metadata is provided to clients by the metadata servers. Clients communicate with the metadata servers only to get the information they need to locate and access the files. Once they have this information, SAN File System clients can access data directly from the storage devices through the clients' own direct connection to the SAN. Direct data access eliminates server bottlenecks and provides the performance necessary for data-intensive applications.

SAN File System presents a single, global namespace in which clients can create and share data using uniform file names from any client or application. Data consistency and integrity are maintained through SAN File System's management of distributed *locks* and the use of *leases*. SAN File System provides locks that enable file sharing among SAN File System clients, and when necessary, provides locks that allow clients to have exclusive access to files. A lease determines the maximum period of time that a metadata server guarantees the locks that it grants to clients. A client must contact the metadata server before the lease period ends to retain its locks.

SAN File System also provides automatic file placement and management through the use of policies and rules. Based on the rules specified in centrally-defined and managed policies, SAN File System automatically stores, moves, and deletes data in *storage pools* that are specifically created to provide the capabilities and performance appropriate for how the data is accessed and used.

What are the major features?

This topic summarizes the major features of SAN File System.

Direct data access by exploitation of SAN technology

SAN File System uses a data-access model that allows client systems to access data directly from storage systems using a high-bandwidth SAN, without interposing servers. Direct data access helps eliminate server bottlenecks and provides the performance necessary for data-intensive applications.

Global namespace

SAN File System presents a single, uniform, global namespace view of all files in the system to all of the clients, without manual, client-by-client configuration by the administrator. A file can be identified using the same path and file name, regardless of the client platform from which it is being accessed. The global namespace shared directly by clients also reduces the requirement of data replication. As a result, the productivity of the administrator as well as the users accessing the data is improved.

Heterogeneous file sharing

All clients, regardless of operating system or hardware platform, have uniform access to the data stored (under the global namespace) in SAN File System. File metadata (such as last modification time) is presented to users and applications in a form that is compatible with the native file system interface of the platform.

Policy-based storage and data management

SAN File System is aimed at simplifying the storage-resource management and reducing the total cost of ownership by the policy-based automatic placement and management of files on appropriate storage devices. You can define storage pools based on specific application requirements and quality of services, and define rules based on data attributes to store the files on the appropriate storage devices automatically. SAN File System provides policy-based data management that automates the management of storage resources and the data stored on those resources.

Components

Figure 1 on page 3 illustrates the major components of SAN File System.

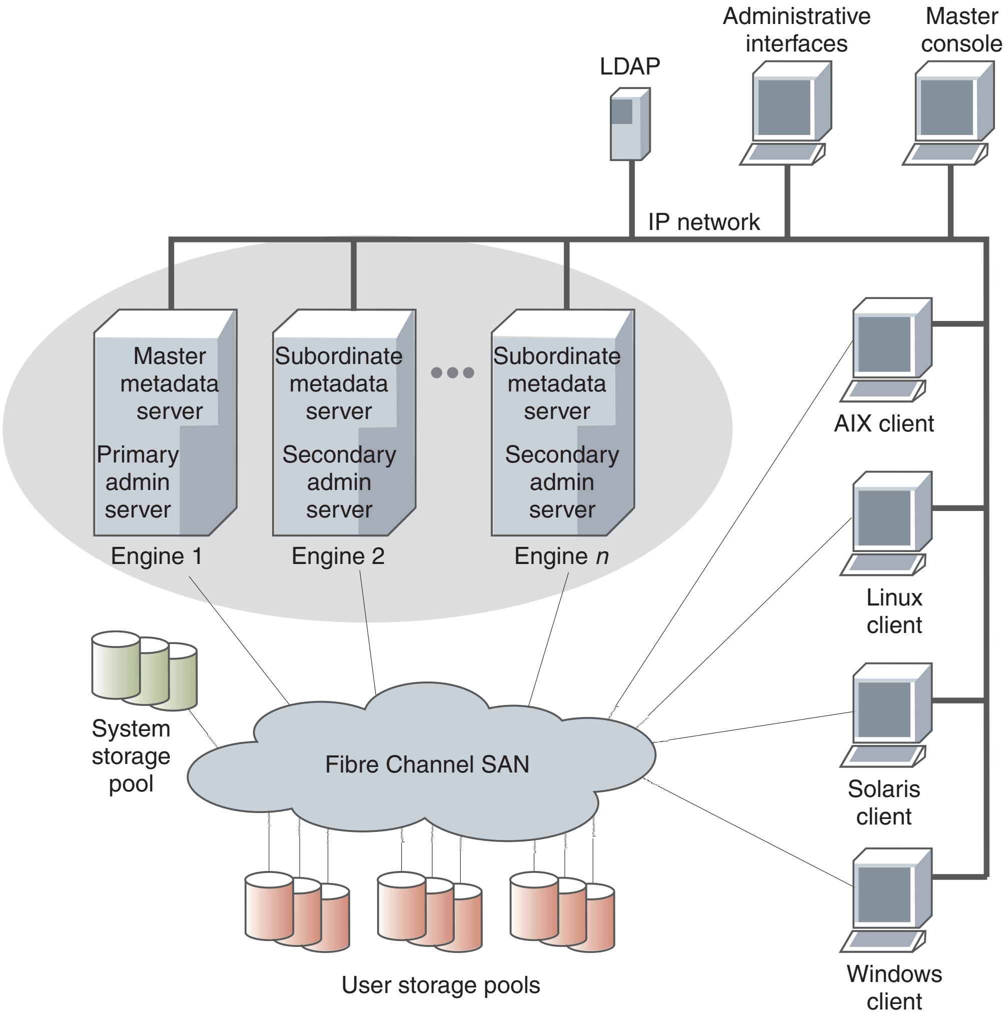


Figure 1. SAN File System components

The metadata servers and clients communicate over a private IP network and access data over a Fibre Channel storage attached network (SAN). SAN File System relies on networking hardware (including an IP network, SAN, network switches, and routers) that already exists in your environment.

The *metadata servers* run on separate physical machines (known as *engines*) and perform metadata, administrative, and storage-management services. The metadata servers are clustered for scalability and availability, and are referred to collectively as the *cluster*. In the cluster, there is one master metadata server and one or more subordinate metadata servers. Additional metadata servers can be added, as required, when the workload grows.

The metadata resides on private storage that is shared among all the metadata servers in the cluster. This storage is known as the *system storage pool*. A storage pool is a collection of SAN File System volumes in the SAN. The system storage pool contains the system metadata (such as system configuration and state information) and file metadata (such as file creation date and permissions). The actual file data is stored on the *user storage pools*, which may be shared among the clients.

The *administrative server* allows SAN File System to be remotely monitored and controlled through a Web-based user interface, called the *SAN File System console*. In addition, the administrative server processes requests issued from the administrative command-line interface, which can also be accessed remotely. The ability to access the SAN File System through these two types of interfaces allows you to administer SAN File System from almost any system with network connectivity. The machine that you use to access these interfaces is called the *administrative console*. The administrative server uses a *Lightweight Directory Access Protocol (LDAP) server* to look up authentication and authorization information about the administrative users. The primary administrative server runs on the same engine as the master metadata server. It receives all requests issued by administrators and also communicates with the administrative servers that run on each additional metadata server in the cluster to perform routine requests.

Computers that share data and have their storage centrally managed by SAN File System are known as *clients*. The SAN File System client software enables the clients to access a single, uniform global namespace through a virtual or installable file system. These clients can act as servers to a broader clientele, providing network file system (NFS) or common Internet file system (CIFS) access to the global namespace or hosting applications, such as database servers or Web-hosting services that use multiple servers.

The *master console* provides serviceability features, including the remote-support interface for remote access and service alert for call home capabilities. The master console is a required feature for SAN File System that can be shared with other IBM TotalStorage products, such as SAN Volume Controller.

Terminology

This topic provides an overview of the terminology that you need to understand before planning for SAN File System.

Cluster

Watch and learn

The SAN File System *cluster* is a set of metadata servers, each running on a separate hardware engine. The metadata servers communicate with each other and with SAN File System clients over your existing IP network. The cluster provides a single point of control for administrative and service operations.

The cluster has one master metadata server and one or more subordinate metadata servers. The master metadata server maintains the cluster state and is the focal point for most administrative services. The maximum number of metadata servers that SAN File System allows in the cluster is eight.

Note: Although SAN File System requires a minimum of two metadata servers, you can run a single metadata-server system if all other metadata servers in the cluster fail (for example, if you have only two engines, and one of them fails), or if you want to stop all of the metadata servers except one to perform scheduled hardware maintenance.

Engines

Within SAN File System, the hardware on which the metadata servers and administrative servers run are called storage *engines*. SAN File System supports from two to eight engines.

SAN File System is intended to run with a minimum of two engines; however, you can run a single-engine system if:

- All of the other engines fail (for example, if you have only two engines, and one of them fails)
- You want to bring down all of the engines except one before performing scheduled maintenance.
- One engine hosts a spare metadata server.

You can use the SAN File System console or administrative command-line interface to monitor and control the engines from any computer with a network connection to the cluster.

File placement

SAN File System provides automatic file placement at the time of creation through the use of policies and storage pools. You can create quality-of-service storage pools that are available to all users and define rules and policies that place newly created files into the appropriate storage pool automatically.

The file-placement policy tells a metadata server where to place the data for a newly created file in a specific storage pool if the attributes of that file meet the criteria specified in a rule. A rule can apply to any file being created or to only files being created within a specific fileset depending on how it is defined. Other criteria include these:

- Date and time when the file is created
- Fileset
- File name or extension
- User ID and group ID on UNIX clients

The rules in a file-placement policy are evaluated in order until the condition in one of the rules is met. The data for the file is then stored in the storage pool that is specified by the applicable rule. If none of the conditions specified in the rules is met, the data for the file is stored in the default storage pool.

Rules in a policy are evaluated only when a file is being created. If you switch from one policy to another, the rules in the new policy apply only to newly created files. Activating a new policy does not change the storage pool assignments for existing files. Moving a file does not cause a policy to be applied. You can create multiple policies, but only one policy can be active at a time.

After a file has been created, you can check its storage pool assignment using the **statfile** command from the administrative command-line interface (CLI). You can also use the **statpolicy** command from the administrative CLI to view the statistics about the file-placement policy rules.

Attention:

It is recommended that you do not use creation time, user ID or group ID to place file. If you do base any file-placement rules on creation time, user IDs, or group IDs, be aware of these restore and migration considerations:

- A rule that uses the creation date as the placement criteria assigns a file based on the time that the file was restored or migrated, not the original creation time.
- A rule that uses a user ID or group ID as the placement criteria assigns a file based on the effective user and group IDs of the restore process, not the original file's user and group IDs.

Filesets

In most file systems, a typical file hierarchy is represented as a series of folders or directories that form a tree-like structure. Each folder or directory could contain many other folders or directories, file objects, or other file-system objects, such as symbolic links and hard links. Every file system object has a name associated with it, and it is represented in the namespace as a node of the tree.

SAN File System introduces a new file system object, called a *fileset*. A fileset can be viewed as a portion of the tree-structured hierarchy (or global namespace). Filesets divide the global namespace into a logical, organized structure. They attach to other directories in the hierarchy, ultimately attaching through the hierarchy to the root of the SAN File System cluster mount point. The collection of filesets and their content in SAN File System along with the file system root combine to form the global namespace. Fileset boundaries are not visible to the clients; only the administrator of SAN File System is aware of them.

From a client's perspective, a fileset appears as a regular directory or folder within which the clients can create their own regular directories and files. Clients cannot delete or rename the directories that represent filesets.

In addition to organizing the overall structure of the global namespace, SAN File System also uses filesets for these purposes:

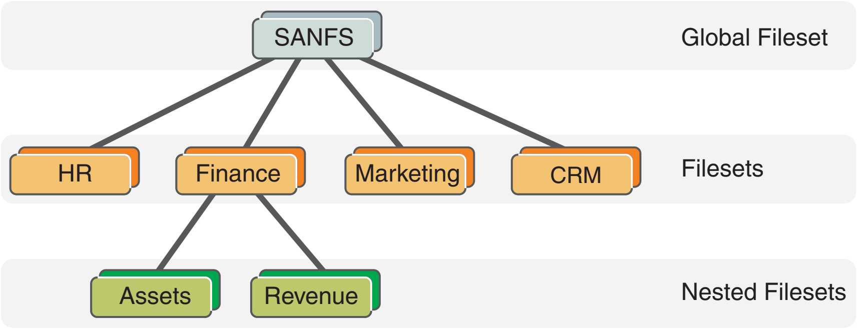
- Represent the workload for the metadata servers
- Provide a level of granularity for data replication (using FlashCopy images)
- Control the amount of space used by the clients (through hard and soft quotas)

A fileset has the following properties:

- A fileset name.
- A directory path leading to the directory within which the fileset is attached. The directory path for the global fileset is the same as the cluster name, *sanfs*.
- A directory name that the fileset is given at the end of the directory path.
- A hard or soft quota.

The root of the global namespace is the *global fileset*. The name of the global fileset is always ROOT. The directory path of the global fileset is specified when you set up the global namespace and is the same as the cluster name *sanfs*.

When you create a fileset, you attach it to a specific location in the global namespace, either to the global fileset or to another fileset. When a fileset is attached to another fileset other than the root fileset, it is called a *nested fileset*.



You can detach a fileset and reattach it at the same location or a different location. If a fileset is reattached at a different location, all the files contained in the fileset are rooted to the new location without any further operations. Before a fileset can be detached, any nested filesets must be detached first.

FlashCopy images

SAN File System has a FlashCopy function that creates an instantaneous copy (or image) of a fileset. The FlashCopy image is a read-only, space-efficient image of the contents of the fileset at the time that it was taken. You can use standard backup applications or utilities on SAN File System clients to back up the contents of FlashCopy images, rather than the actual fileset. Backing up the FlashCopy image avoids any issues with open files that might cause problems when backing up live data.

FlashCopy images are file-based, so SAN File System clients can see all of the files and directories in the image. The clients can use this image for quick restore of parts of the fileset if required, by simply copying the required files and folders back to the actual fileset. You can also quickly revert the entire fileset from a FlashCopy image.

Global namespace

The *global namespace* is the key to SAN File System. It gives all SAN File System clients common access to all files and directories, and ensures that all SAN File System clients have consistent access and a consistent view of the data and files managed by SAN File System. Having common access to all files reduces the need to store and manage duplicate copies of data and simplifies the backup process. Security mechanisms, such as permissions and access control lists (ACLs), restrict the visibility of files and directories.

Metadata server

Watch and learn

A *metadata server* is a software server that performs metadata, administrative, and storage-management services and provides clients with shared, coherent access to shared storage (or global namespace). The metadata servers are clustered for scalability and availability, and are often referred to as a cluster. In the cluster, there is one master metadata server and one or more subordinate metadata servers, each running on a separate storage engine. Additional metadata servers can be added, as required, when the workload grows.

All of the metadata servers, including the master metadata server, share the workload of the global namespace. Each is responsible for providing metadata and locks to clients for specific filesets assigned to them. They know which filesets belong to which metadata server, and when contacted by a client, can direct the client to the appropriate metadata server. They also manage distributed locks to ensure the integrity of all of the data within the global namespace.

In addition to providing metadata to clients and managing locks, metadata servers perform a wide variety of other tasks. They process requests to create and manage filesets, storage pools, volumes, and policies; enforce the policies to place files in appropriate storage pools; and send alerts when any threshold established for the filesets and storage pools are exceeded.

Policies and rules

This topic describes how SAN File System automates the management of files using policies and rules.

SAN File System enables you to automate the management of files using policies and rules. Properly managing your files allows you efficiently use and balance your premium and inexpensive storage. SAN File System supports these policies:

- File-placement policies are used to automatically place newly created files to a specific storage pool.
- File-management policies are used to manage files (move or delete) during its lifecycle by moving them to another storage pool or delete them all together.

Policies

A *policy* is a set of rules that determine where specific files are placed based on the file's attributes. You can define any number of policies, but only one policy can be active at a time. If you switch from one policy to another or make changes to a policy, that action has no effect on existing files in the global namespace. The new or changed policy is effective only on newly created files in SAN File System. Manually moving a file does not cause the policy to be applied.

A policy can contain any number of rules. There is no limit to the size of a policy.

SAN File System performs error checking for file-placement policies in the following phases:

- When you create a new policy, the master metadata server checks the basic syntax of all the rules in the policy.
- When you activate the policy, the master metadata server checks all references to filesets and storage pools. If a rule in the policy refers to a fileset or storage pool that does not exist, the policy is not activated and an error is returned.
- When a new file is created by a client, the rules in the active policy are evaluated in order. If an error is detected, the metadata server responsible for creating the file logs an error, skips all subsequent rules, and assigns the file to the default storage pool. If a default pool does not exist, the file is not created and the metadata server returns an error to the client application.

Currently, there is no error checking for file-management policies.

If your environment is set up in a non-uniform zone configuration (in which clients cannot access all volumes), you need to ensure that the rules in the active policy place files into volumes that are accessible to the clients that use them.

Tip: When SAN File System is first installed, a default file-placement policy is created and remains active until you create and activate a new one. The default file-placement policy assigns all files to the default storage pool. Although the default storage pool is created when SAN File System is first started, you must assign volumes to it before it can be used. If a user or application on a SAN File System client attempts to create new files that would be assigned to the default storage pool, and there are no volumes assigned to it, the user or application receives `No Space` errors.

Rules

A *rule* is an SQL-like statement that tells the metadata server what to do with the data for a file in a specific storage pool if the file meets specific criteria. A rule can apply to any file being created or only to files being created within a specific fileset or group of filesets.

Rules identify the conditions, such as these, that when matched causes that rule to be applied:
- Date and time when the file is created
- Date and time when the file was last accessed
- Fileset
- File name or extension
- File size
- User ID and group ID on UNIX clients

SAN File System evaluates rules in order, from top to bottom, as they appear in the active policy. The first rule that matches determines what is to be done with that file. For example, when a client creates a file, SAN File System scans the list of rules in the active file-placement policy to determine which rule applies to the file. When a rule applies to the file, SAN File System stops processing the rules and assigns the file to the appropriate storage pool. If no rule applies, the file is assigned to the default storage pool.

Attention:

It is recommended that you do not use creation time, user ID or group ID to place file. If you do base any file-placement rules on creation time, user IDs, or group IDs, be aware of these restore and migration considerations:
- A rule that uses the creation date as the placement criteria assigns a file based on the time that the file was restored or migrated, not the original creation time.
- A rule that uses a user ID or group ID as the placement criteria assigns a file based on the effective user and group IDs of the restore process, not the original file's user and group IDs.

Storage management

SAN File System provides automatic file placement and management through the use of policies. The policy *rules* cause newly created files to be placed in the appropriate storage pools and cause files of a certain age or size to be moved or deleted.

Storage pools

Watch and learn

A *storage pool* is a named set of SAN File System volumes that can be used to store either metadata or file data. A storage pool consists of one or more volumes that provide a quality of service that you want for a specific use, such as to store all files for a particular application or a specific business division. You must assign one or more volumes to a storage pool before it can be used.

SAN File System has two types of storage pools: system storage pool and user storage pool.

Storage pools and volumes

Typically, you assign volumes to storage pools based on their common characteristics, such as device capabilities (availability or performance level) and usage (business division, project, application, location, or customer).

Each storage pool manages its own volumes. File space is allocated to the volumes in a given storage pool in a round-robin algorithm (as shown in Figure 2) in logical partitions, or in blocks. Logical partitions are allocated to the system storage pool in 16-MB blocks. For user storage pools, including the default storage pool, you can allocate logical partitions in 16, 64, or 256-MB blocks. All logical partitions in the same storage pool must be the same size.

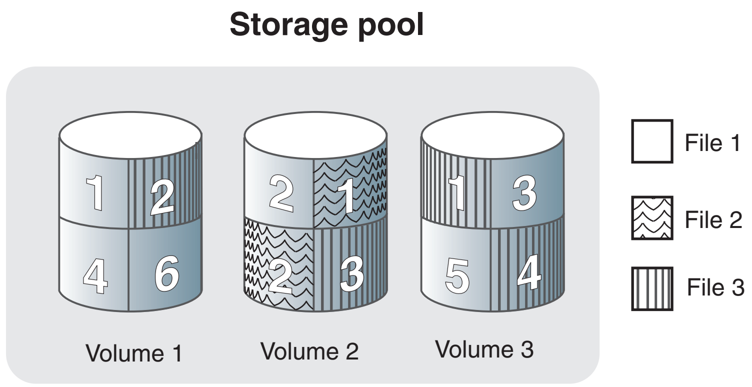


Figure 2. File space allocation

Tip: You can set a threshold to generate an alert when a storage pool reaches or exceeds a certain percentage of its maximum capacity. By default, an alert is generated when a storage pool becomes 80% full. An alert is logged every five minutes until one or more volumes are assigned to the storage pool. You can set configuration parameters to cause an SNMP trap message to be generated as well. An SNMP trap notifies you of this condition asynchronously.

System storage pool

The *system storage pool* contains the system metadata (system and file attributes, configuration information, and metadata server state) that is accessible to all metadata servers in the cluster. There is only *one* system storage pool that is created automatically when SAN File System is installed. The system storage pool contains the most critical data for SAN File System. The first volume that is assigned to the system storage pool, called the *master volume*, contains the most critical pages of metadata that SAN File System manages.

Important: Use highly-reliable and available logical unit numbers (LUNs) for the system storage pool (for example, mirroring or redundant array of independent disks (RAID), plus hot spares in the backend storage system) so that the cluster always has a robust copy of the system metadata.

Because the amount of metadata grows as the global namespace grows, you must monitor the system storage pool to ensure that there is always enough volumes assigned to it to accommodate the growth. The system storage pool typically requires approximately 2% to 5% of the total storage capacity that SAN File System manages, but this amount varies depending on your environment. Use the alert features on the system storage pool to ensure that you do not run out of space.

Tip: The minimum size of a system volume is 2 GB; therefore, the minimum size of the system storage pool is also 2 GB.

For security and reliability, the volumes that are assigned to the system storage pool should be accessible only to the cluster using a private SAN or a shared SAN with a combination of zoning, LUN masking, or special configuration. For reliability, the volumes should be virtualized RAID arrays (also known as *ranks* within IBM Enterprise Storage Server).

User storage pools

A *user storage pool* contains the blocks of data that make up user files. SAN File System stores the data that describes the files, called file metadata, separately from the actual file data. You can create one or more user storage pools, and then create policies that contain rules that cause metadata servers to store data for specific files in the appropriate storage pools.

The *default storage pool* is a special user storage pool. This optional storage pool is used to store the data for a file if the file is not assigned to a specific storage pool by a rule in the active policy. A default storage pool is created when SAN File System is installed. However, if you want to use the default storage pool, you must assign one or more volumes to it. There can be only one default user storage pool in SAN File System. You can designate any user storage pool that has volumes assigned to it to be the default storage pool. You can choose to disable the default storage pool. In this case, newly created files that do not match any rules in the active policy are not saved.

UNIX-based clients

A *UNIX-based client* is a SAN File System client that runs a UNIX operating system and has the SAN File System client code installed.

SAN File System supports clients running on these UNIX operating systems:

- AIX® 5.1 (32-bit only)
- AIX 5.2 (32-bit and 64-bit)
- AIX 5.3 (32-bit and 64-bit)
- Red Hat Enterprise Linux Advanced Server 3.0 running either SMP or Hugemem kernels in WS, ES or AS editions
- SUSE Linux Enterprise Server 8 (32-bit)
- Sun Solaris 9 (64-bit)

Restriction: SAN File System supports AIX client machines that have up to eight processors.

The SAN File System client code that is installed on a UNIX platform is called a Virtual File System (VFS). The VFS is a subsystem of the UNIX-based client's virtual file system layer. It directs all metadata operations to a metadata server and all data operations to storage devices that are attached to your SAN. The VFS makes the metadata that is visible to the client's operating system, as well as any applications that run on the client, look identical to metadata read from a native, locally-attached file system.

UNIX-based clients mount the global namespace on their systems. After the global namespace is mounted, you can use it just as you would any other file system to access data and to create, update, and delete files and directories. The following example shows an AIX mount point for SAN File System:

```
root@aix2:/# df
Filesystem  1024-blocks      Free  %Used  Iused  %Iused  Mounted on
/dev              32768     23024    30%   1413      9%  /
/dev/hd1         950272      8096   100%  29103     13%  /usr
SANFS          16728064  16154624     4%      1      1%  /sanfs
```

UNIX-based clients use standard UNIX permission semantics (such as read, write, and execute bits, and owner and group IDs) that make the global namespace appear as if it were a local UNIX file system.

User interfaces

There are two methods for managing SAN File System: an *administrative command-line interface* and a graphical user interface, called the *SAN File System console*. You can access the administrative command-line interface by either directly logging in to an engine or using a Secure Shell (SSH) client to remotely connect to the engine. You can access the SAN File System console using a Web browser.

SAN File System provides you with different levels of user access to perform administrative operations. The users and user roles are defined on your LDAP server. Therefore, you cannot access the SAN File System without a valid user ID that is defined in the LDAP server.

SAN File System provides the following user interfaces:

- A Web-based administrative user interface called the SAN File System console
- An administrative command-line interface
- A client command-line interface

Note: The administrative server does not lock administrative access in order to prevent simultaneous SAN File System console or administrative command-line interface sessions. You must manually coordinate the use of the administrative interfaces.

SAN File System console

The SAN File System console allows you to control and monitor SAN File System from a Web-based graphical user interface. For ease of monitoring, it provides a system overview that illustrates the status of the various SAN File System components. In addition, the SAN File System console provides inline messaging that assists with system configuration, performance tuning and troubleshooting tasks.

The SAN File System console also contains the Help Assistant, which provides panel-level help information as well as links to related topics in the SAN File System Information Center. The Information Center serves as an online, searchable repository for all of the product documentation.

Administrative command-line interface

You can use the administrative command-line interface to administer all aspects of SAN File System, including setting up and managing storage pools, volumes, and filesets. For security reasons, administrative command-line interface runs only on the engines in your cluster.

You can use the administrative command-line interface interactively using the **sfscli** utility. You can also embed administrative commands in scripts.

To access **sfscli**, you must log in to an engine that hosts any metadata server. The following figure illustrates how you access **sfscli**.

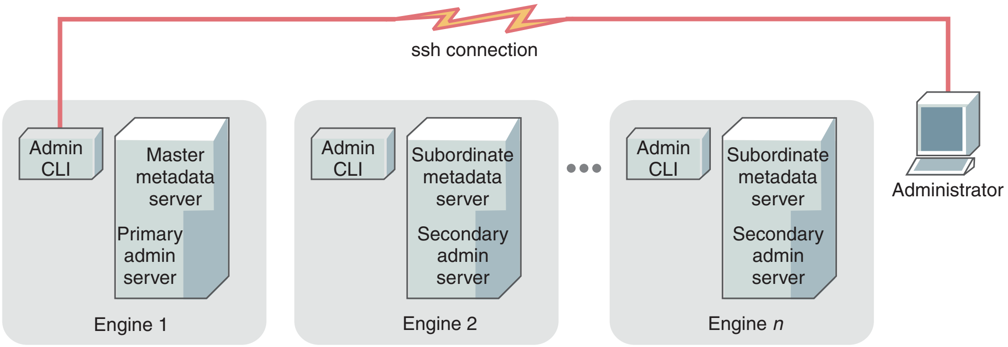


Figure 3. Accessing **sfscli**

Client commands

SAN File System provides a set of commands that are used to set up SAN File System clients and to perform planning, migration, and verification tasks for data. These commands are issued from the client machines.

Volumes

A *logical unit number* (LUN) is the logical unit of storage that a SAN or other disk subsystem can assign to metadata servers and clients. A *volume* is a LUN that is labeled by SAN File System for its use. Volumes are grouped together virtually to form storage pools, in which file data and metadata is stored.

An LUN becomes a SAN File System volume when you add it to a storage pool. It is automatically assigned a system-generated label that identifies it as a SAN File System volume. You must also give the volume a name that is unique among all the volumes used by a SAN File System cluster.

During startup, the metadata server scans all LUNs that it can access in the SAN, searching for the label that tells it that the LUN is a valid SAN File System volume. Clients perform this same search whenever they are started.

System-data LUN operations are performed by the metadata servers. All other data LUN operations are initiated from and coordinated by the metadata servers in the cluster but are actually performed by one or more clients; therefore, the metadata servers no longer need to see the data LUNs, and the clients only need to see the data LUNs that they need to access. This allows SAN File System to support a wide variety of SAN configurations, storage devices, and drivers, and also supports scaling to large numbers of storage devices and clients. This also allows SAN File System to support grouping clients and LUNs into SAN zones to provide enhanced security.

A volume must be empty to be removed from a storage pool. When you remove a volume, SAN File System moves the contents of that volume across other available volumes in the same storage pool. If the storage pool does not have sufficient space available in other volumes to move all of the data contained in the specified volume, the removal fails and the metadata server suspends the volume (the metadata server cannot allocate new data on that volume).

Tip: Keep the storage subsystem device driver's virtual path (vpath) configuration file current. If many LUNs are added and deleted from the metadata server, it is possible for the configuration file to contain references to LUNs that do not exist.

Restriction: A metadata server can access up to a combined total of 256 SCSI disk single-pathed and/or vpath multi-pathed LUNs. This is a limitation of the Linux operating system. When the number of entries in the storage subsystem device driver's vpath configuration file reaches 256, any new LUN configured on the metadata server will not be visible.

Volumes and storage pools

When you install SAN File System, there is a system storage pool, which is used by metadata servers to store system and file metadata, and a default storage pool, which can be used to store file data. You can create additional user storage pools for file data; however, no data can be stored in a storage pool until you assign one or more volumes to it. You can also remove the default storage pool if you choose.

The volumes added to the system storage pool are called *system volumes*.

As the amount of metadata that is generated for the server cluster and client files grows, you must ensure that the system storage pool always has enough volumes assigned to it so that it does not run out of space.

You must also ensure that the user storage pools, including the default storage pool, has a sufficient number of volumes. Each storage pool must have at least one volume assigned to it before any files can be stored in it.

To assist you in monitoring storage pool capacity, SAN File System provides a threshold option that you can specify when adding a volume to a storage pool or changing settings for a storage pool. A threshold is a specified percentage of the estimated maximum capacity of the storage pool. When a storage pool reaches or exceeds the percentage specified as its threshold, SAN File System generates an alert. This alert can also generate an SNMP trap message to notify you of the condition asynchronously, if you set the appropriate parameters for SNMP traps.

Limitations to volumes in the system storage pool

The volumes in the system storage pool have these limitations:

- All volumes in the system storage pool must be of the same type of backend storage device and must be one of the supported IBM storage subsystems. You can use IBM TotalStorage SAN Volume Controller to provide mixed storage as long as only the SAN Volume Controller virtual devices are visible to the cluster.
- All volumes in the system storage pool must be visible to all metadata servers in the cluster.

- Each volume in the system storage pool must be at least 2 GB in size.
- The system storage pool is limited to 126 dual-path volumes.

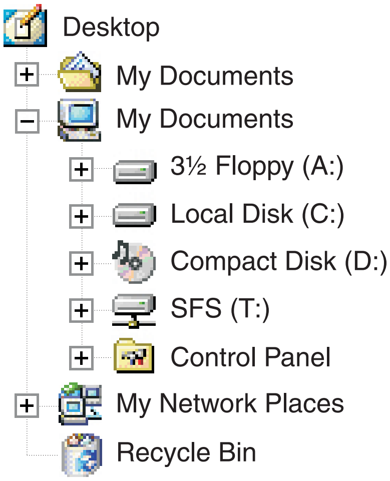
Windows-based clients

A *Windows-based client* is a client that runs a Windows operating system and has the SAN File System client code installed. In this release,

SAN File System supports clients that run on these Windows operating systems:
- Windows 2000 Advanced Server
- Windows 2000 Server
- Windows 2003, Standard Edition
- Windows 2003, Enterprise Edition

The SAN File System client code installed on a Windows-based client is an Installable File System (IFS). The IFS is a kernel-mode driver that extends the Windows I/O subsystem to support an additional file system. SAN File System client code directs all metadata operations to a metadata server and all data operations to storage devices attached to your storage area network (SAN). SAN File System client code makes the metadata that is visible to a client's operating system, as well as any applications that run on the client, look identical to metadata read from a native, locally attached file system.

Windows clients mount the global file system on their systems. After the global file system is mounted, users can use it just as they would any other file system to access data and to create, update, and delete files and directories. The following example shows the My Computer view from a Windows 2000 client. The T: drive (labeled SFS) is the attach point of SAN File System.



Desktop
My Documents
My Documents
3½ Floppy (A:)
Local Disk (C:)
Compact Disk (D:)
SFS (T:)
Control Panel
My Network Places
Recycle Bin

Windows-based clients use a subset of the Windows semantics. The allowed semantics are described to Windows as volume properties, which are visible, for example, as properties of the drive within Windows Explorer. The following volume properties are supported by SAN File System:
- NTFS-like access control lists (which requires all Windows-based clients to share a common Active Directory domain for users and groups)
- Long names and short names (eight-character names with three-character extensions)
- Unicode-based file names

- Case-sensitive file names

Planning checklist

This checklist provides a list of the activities that you need to perform to plan for the installation of the SAN File System. Use this worksheet to verify that you have performed all activities.

Use this checklist to ensure that you have completed all planning activities.

	Plan the client configuration.
	Plan the global namespace configuration.
	Plan the storage configuration.
	Plan the cluster configuration.
	Plan the file management strategy.
	Validate the existing SAN infrastructure.
	Plan the zoning configuration.
	Plan the security strategy.
	Plan the backup and recovery strategy.
	Plan the data migration strategy.
	Gather hardware and software prerequisites

Chapter 2. Planning the client configuration

This topic describes the how to plan which application servers or application groups will become SAN File System clients.

Perform these tasks to plan your client configuration:

- Identify the applications and application groups in your environment that you want to use SAN File System to store data. The machines running those applications and application groups will become clients to SAN File System.
- Identify the authentication mechanism that the clients will use to securely access the global namespace.
- Identify applications running on the clients that require direct I/O to the global namespace.
- Identify the clients that you want to have root access to the global namespace (called *privileged clients*).
- Gather information required to install IFS/VFS on the client machine.

Use these worksheets to collect information needed to install IFS/VFS:

- UNIX-based-client installation worksheet
- Windows-based-client installation worksheet

Antivirus software

If more than one SAN File System client is running antivirus software that scans directories and files, shared files only need to be scanned by one SAN File System client. It is unnecessary to scan shared files more than once. When you run antivirus scans from more than one client, schedule the scans to run at different times, to allow better performance of each scan.

Tips:

- Consider using a single, designated client machine to perform all virus scans.
- On Windows 2003 clients, use antivirus software that does not use the mini-filter or filter manager model.

Authentication and authorization

SAN File System performs authorization checking for file-system operations on client machines based on the native operating system's user authentication mechanism. SAN File System does not restrict how authentication is performed, but it does assume that all UNIX clients share a common definition of users and groups; specifically, it assumes that any given identity using SAN File System has the same numerical value on all UNIX clients.

Host-based clustering

This topic describes the cluster applications that you can run on SAN File System clients.

SAN File System works with clients that are in a clustered environment; however SAN File System is independent and not aware of any host-based clustering. SAN File System data volumes are owned and managed by SAN File System and must not be assigned as resources to the local operating system or cluster manager. Because cluster managers write on the volumes, configure the volumes as raw, unmanaged volumes to each member of the host-based cluster.

With SAN File System, you can use these clustering applications on the client machines:

- High availability cluster multi-processing (HACMP™) on AIX platforms
- Sun Solaris clustering

Restriction: You cannot use Microsoft clustering on Windows 2000 and Windows 2003 platforms at this time.

Data LUN configuration

You can configure the SAN so that all data LUNs are available to all clients (known as a *uniform configuration*) or so that only a subset of data LUNs are available to some clients (known as a *nonuniform configuration*). In nonuniform configurations, a client must be able to access all LUNs in any storage pool that has been used or can be chosen by a fileset that is used by that client.

If a client tries to read or write data on a LUN that it cannot access, SAN File System returns an I/O error to that client. If the client performs a file-system operation that only involves metadata (such as changing a directory, or listing or creating files), SAN File System does not return an I/O error because the operation does not involve the data LUN.

The active policy determines which storage pool is selected when a new file is created. In a nonuniform configuration, the policy must ensure that a newly-created file is allocated to a storage pool that is accessible to the clients that need the file. When you change the active policy, the new policy must meet this consistency property.

Note: There is no automatic consistency check for a nonuniform configuration; however, when a client identifies itself to a metadata server, the metadata server inspects the client volume list to ensure that no incomplete storage pools are visible to the client. If an incomplete storage pool is visible, the metadata server logs an error in the metadata server log.

Direct I/O considerations

This topic describes things to consider when your applications require direct I/O.

Some applications, such as database management systems, use their own cache management systems. For such applications, SAN File System provides a direct I/O mode, which allows these applications to bypass the data cache. In this mode, SAN File System performs direct writes to disk and does not cache data. This also allows distributed applications on different AIX or Linux client machines to write data to the same file at the same time. Using the direct I/O mode makes files act like raw devices. This gives database systems direct control over their I/O operations, while still providing the advantages of SAN File System, such as the FlashCopy feature and file-level backup and restore processing. Applications need to be aware of, and configured for, direct I/O.

UNIX-based clients use existing operating-system interfaces to use direct I/O. That is, you must set the O_DIRECT flag to open a file in direct I/O mode.

Windows-based clients enforce full, native direct I/O, or *unbuffered I/O*, semantics. You must specify the FILE_FLAG_NO_BUFFERING flag to open or create a file in direct I/O mode. When using this flag, your application must meet the following requirements:

- The I/O buffers, offsets and transfer size must be integer multiples of the volume's sector size.
- Buffer addresses for read and write operations are not required to be sector aligned; however, the target offsets must be sector aligned.

You receive a return code of 87 (ERROR_INVALID_PARAMETER) if the requirements are not met.

Restrictions:

- You cannot use direct I/O processing on Linux clients.
- Applications that use direct I/O are responsible for managing concurrent writes to the same file.
- A process cannot use direct I/O on a file that is being used in cache mode by another process on the same client machine. Similarly, a process cannot use a file in cached mode that is being used in direct I/O mode by another process on the same client machine. In either case, you will receive an EAGAIN error.

  However, a process can use direct I/O on a file that is being used in cache mode by a process on another client machine. Similarly, a process can use a file in cached mode that is being used in direct I/O mode by a process on another client machine.

Privileged clients

SAN File System includes a configurable list of privileged clients. A *privileged client* is a client on which root users in UNIX or users with administrator privileges in Windows are given those same privileges for the SAN File System global namespace. A root user that is logged in to a privileged UNIX-based client is granted full control over directories, files, and other file system objects that are created by UNIX-based clients. A user with administrator privileges who is logged in to a privileged Windows-based client is granted full control over the folders, files, and other file-system objects that are created by Windows-based clients.

If those same users log in to a client that is not a privileged client, their privileges for the global namespace are reduced to those of "everyone" for Windows users or "other" for UNIX users.

Chapter 3. Planning the global namespace configuration

This topic describes how to plan the global namespace configuration.

Perform these tasks to plan the global namespace configuration:

- Determine how to organize the global namespace by partitioning it into filesets
- Determine how to limit the amount of storage that clients can use, through hard and soft quotas
- Determine how much space is required for FlashCopy images.

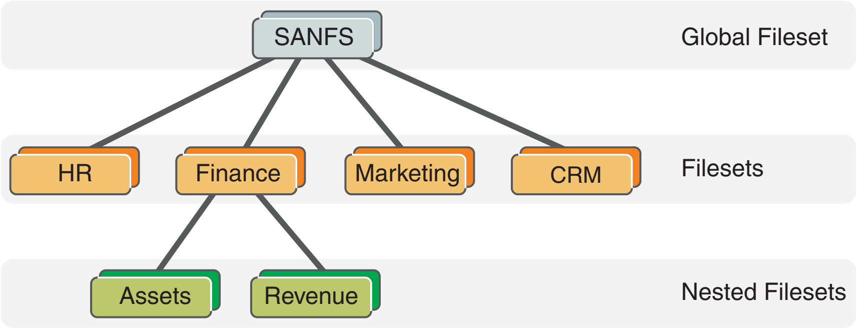
Use the Filesets worksheet to collect information about the global namespace.

How should I organize my global namespace?

This topic describes how to plan for organizing the global-namespace by partitioning it into filesets.

At the root of the global namespace is the *global fileset*. The name of the global fileset is always ROOT. The directory name of the global fileset is specified when you set up the global namespace, for example as sanfs.

When you create a new fileset, you attach it to a specific location in the global namespace, creating a hierarchy. You can attach the fileset to the global fileset or to another to another fileset. When a fileset is attached to another fileset, it is called a *nested fileset*.



Tip: In the SAN File System environment, you can create regular, non-fileset subdirectories only from the client machines. As a result, the metadata servers cannot recreated directory structures that contain a mix of filesets within subdirectories and regular subdirectories. To simplify disaster recovery, attach filesets only to the global fileset (root directory), not to regular subdirectories under another fileset. The **mkdrfile** command output can then be used to completely restore the top of the global namespace tree before using the client-based backup application to restore the rest of the global namespace.

These are some guidelines for partitioning the global namespace into filesets:

- Keep data used by an application in the same fileset. Do not split the data across multiple filesets.

- If you must split the data used by an application across multiple filesets, observe natural divisions in the application data to minimize cross-access.
- Create enough filesets so that workload can be easily balanced. It is easier to balance the workload with 10 filesets per server rather than two filesets per server.
- Define filesets so that fileset quotas is meaningful.
- Keep the number of file objects per fileset under 100 000 to 1 000 000.
- Keep number of fileset transactions per second for a single fileset under 1 000.

Fileset considerations

You can create filesets based on conditions in your environment (for example, workflow patterns, security, or backup considerations, all the files used by a specific application, or files associated with a specific application or client). Filesets are used not only for managing the storage space used, but also for creating FlashCopy images. Correctly defined filesets mean that you can take a FlashCopy image for all the files in a fileset together in a single operation, providing a consistent image for all of those files. The global namespace is partitioned into filesets that match the data-management model of the enterprise. Filesets can also be used as criteria when placing individual files in global namespace.

When you are creating filesets, consider the overall I/O loads on the cluster. Because each fileset is assigned to one (and only one) metadata server, you need to balance the load across all metadata servers in the cluster by assigning filesets appropriately. Also, when the number of filesets is greater than one thousand, response time will increase when you issue fileset commands.

To facilitate file sharing, you can optionally separate filesets by their *primary allegiance* of the operating system. Separating filesets also facilitates file-based backup methods (for example, utilities, such as **tar**, and Windows backup applications such as VERITAS NetBackup or IBM Tivoli® Storage Manager); full metadata attributes of Windows files can be backed up from a Windows backup client only and full metadata attributes of UNIX files can be backed up from an UNIX backup client only.

Fileset permissions

When you create and attach a new fileset to the global namespace, the fileset is owned by user *Anonymous*. A UNIX root user or a Windows administrator user must change the ownership and permissions of the fileset before the fileset is usable. (You must do this for the FlashCopy directory and the lost+found directory under the fileset root.) You need to make these changes only once in the lifetime of a fileset. The changed permissions are persistent across metadata server restarts and whenever the fileset is detached or attached.

Unlike the requirement for the global fileset, a UNIX or Windows user can own a fileset exclusively. The fileset is not required to have write permissions for both UNIX and Windows domains.

Tip: If you change the permissions of a fileset after you create a FlashCopy image and then revert back to that FlashCopy image, the permissions also revert to the settings at the time when the FlashCopy image was taken.

Nested fileset considerations

Consider the following circumstances when creating nested filesets:

- You cannot access a nested fileset if the metadata server that is hosting the parent fileset is unavailable. In other words, if the parent fileset becomes a rogue fileset and is unable to be failed over, then the nested filesets of that parent fileset would also, effectively, be unavailable.
- A FlashCopy image is created at the individual fileset level and does not include any nested filesets. You cannot make a FlashCopy image of a fileset and any nested filesets in a single operation. This can be of concern if you are required to have a consistent image of a fileset and its nested filesets. Making FlashCopy images in multiple operations could lead to ordering or consistency issues.
- To detach a fileset, you must first detach all of its nested filesets.
- It is not possible to revert to a FlashCopy image when nested filesets exist within the fileset. You must manually detach the nested filesets before reverting to the FlashCopy image. You can reattach the nested filesets after the fileset is reverted.
- When creating nested filesets, attach them only directly to other filesets. Do not attach filesets to client-created directories because a large-scale restore is more complex.

How much space can the clients use?

This topic describes how to plan for controlling the amount of space the clients can use in the global namespace by defining fileset quotas.

When creating a fileset, you can specify a maximum size for the fileset, called a *quota limit*, and specify whether SAN File System should generate an alert if the size of the fileset reaches or exceeds a specified percentage of the maximum size, called a *threshold*. For example, if the quota on the fileset is set to 100 GB, and the threshold is 80%, an alert is generated when the fileset contains 80 GB of data. (Note that the quota is based on space allocated to the fileset, not the data is contains.)

The action taken when the fileset reaches its quota size depends on whether the quota is defined as hard or soft. If you use a hard quota, once the threshold is reached, SAN File System denies new client requests to add more space to the fileset (by creating or extending files). If you use a soft quota, which is the default, SAN File System allocates more space but continues to send alerts. Once the amount of physical storage available to global fileset is exceeded, no more space can be used. You can set the quota limit, threshold and quota type individually for each fileset.

Note:

- The space used by a fileset includes the space used by FlashCopy images. It does not include the space used by any nested filesets.
- The metadata servers compute and track hard quota limits for filesets in multiples of the partition size. If a hard quota is not set as a multiple of the partition size, quota violation errors appear in the log file even though the size of the fileset has not reached the specified limit. To avoid this problem, specify hard quota limits as multiples of the partition size (for example, if the partition size is 16 MB, set the quota to multiples of 16).

How much space will I need for FlashCopy images?

This topic describes the FlashCopy images considerations related to the organizing global namespace.

FlashCopy images for each fileset are stored in a special hidden subdirectory, called .flashcopy, under the fileset's attachment point.

FlashCopy images consume space on the same volumes as the original fileset. Because FlashCopy uses a space-efficient method to make the image, the amount of space that is used by FlashCopy images is not possible to predict. If all blocks in the fileset are changed, the image takes up the same amount of space currently occupied by the non-FlashCopy objects within the fileset. If nothing in the fileset changes, the FlashCopy images takes up virtually no space (just pointers to the real fileset data). It is not possible to determine how much space is being occupied by a particular FlashCopy image at any particular time.

Therefore, when you plan your space requirements, include space for FlashCopy images. The amount of space you need to plan for flashcopy images correlates to the amount of changes you make to files with flashcopy images. Carefully monitor the user-storage-pool space threshold. Be aware that the space used by FlashCopy images count toward the fileset's quota.

SAN File System supports up to 32 FlashCopy images per fileset.

Chapter 4. Planning the storage configuration

This topic describes how to plan your storage configuration (storage pools and volumes).

Perform these tasks to plan your storage configuration:

- Determine how many storage pools are needed.
- Determine how many volumes to assign to the storage pools.
- Determine which volumes to assign to the system storage pool.

Use these worksheet to collect the storage configuration information:

- Storage pool worksheet
- System storage pool worksheet
- Volumes worksheet

How many storage pools do I need?

This topic provides consideration for determining how many storage pools you need in your environment.

Typically, you would assign volumes to storage pools based on common characteristics, such as device capabilities (or quality of service) or data security.

When you assign volumes based on device capabilities (for example, performance, reliability, and availability), you first need to classify your data and applications according to its value to the business. These are some examples of assigning volumes based on device capabilities:

- A storage pool that requires volumes that have the same RAID level, pathing redundancy, and controller caching behavior.
- A critical storage pool that requires volumes that are multi-pathed and RAID 5 with a large controller cache, such as with IBM Enterprise Storage Server (ESS).
- A storage pool that uses cheap storage.
- A storage pool that requires fast spindles for random I/O, such as in OLTP.
- A storage pool that requires volumes on a storage device that supports high transfer rates for serving music files to the web.

Storage pools defined in terms of data security would be used to limit access. For example, a medical-records database would be kept in a storage pool whose volumes are zoned to limit access.

Define your storage pool using factors from device capabilities and data-security policies. Creating too few storage pools indicates that there is not enough differentiation in quality of service or security. Because a reserve is typically kept in each storage pool, creating too many storage pools might cause wasted space.

Tip: To ease administration of the storage pools, make volumes in a pool homogenous so that they can be administered uniformly.

How many volumes should I assign to the storage pools?

This topic provides information about assigning volumes to storage pools.

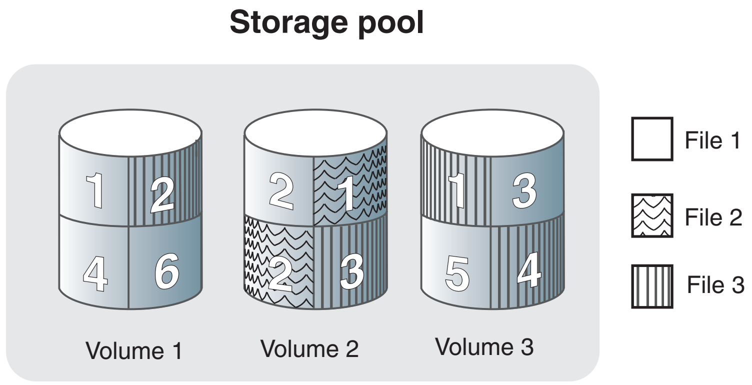
Files are distributed in partition-sized units across the available space in a given storage pool. The storage pools need volumes to distribute those file access across the volumes, thus distributing the I/O load.

Also consider the number of spindles available for volumes in the storage pool. The storage pools need enough spindles to provide adequate I/O parallelism.

Choose volumes from different storage subsystem ranks to allow more than one port to participate in I/O.

If you anticipate needing to remove volumes from a storage pool, the data on the volume to be removed will be distributed across other volumes in the storage pool. There must be enough room on the remaining volumes to contain the data.

Each storage pool manages its own volumes. File space is allocated to the volumes in a given storage pool in a round-robin algorithm (as shown in Figure 2 on page 10) in logical partitions, or in blocks. Logical partitions are allocated to the system storage pool in 16-MB blocks. For user storage pools, including the default storage pool, you can allocate logical partitions in 16, 64, or 256-MB blocks. All logical partitions in the same storage pool must be the same size.



Tip: You can set a threshold to generate an alert when a storage pool reaches or exceeds a certain percentage of its maximum capacity. By default, an alert is generated when a storage pool becomes 80% full. An alert is logged every five minutes until one or more volumes are assigned to the storage pool. You can set configuration parameters to cause an SNMP trap message to be generated as well. An SNMP trap notifies you of this condition asynchronously.

Which volumes to use for the system storage pool?

This topic provides considerations for determining which volumes to assign to the system storage pool.

Because the system storage pool contains the system data and file metadata, it is important that the volumes you assign to it provide high reliability, availability, redundancy and performance, especially if the file system has update workload. Use volumes that have a large write cache to reduces transaction latency.

The system storage pool typically requires approximately 2% to 5% of the total storage capacity that SAN File System manages. As a rule of thumb, assign 2% to 5% of the total number of volumes in the user storage pools to the system storage pool. For example, if you have 100 volumes (assigned to user storage pools) over 48 spindles, you would plan for 5 system volumes over 6 spindles.

Remember to take into account the characteristics of the application workload when you estimate the size of the system storage pool. Your IBM representative can help you assess your applications and determine the metadata workload.

These are additional considerations for assigning volumes to the system storage pool:

- For security, the volumes that are assigned to the system storage pool should be accessible only to the metadata servers in the cluster using a private SAN or shared SAN with a combination of zoning, LUN masking, or special configuration.
- For reliability, the volumes should be virtualized RAID arrays (also known as *ranks* in IBM Enterprise Storage Server).
- For availability, use highly reliable and available volumes (for example, volumes with mirroring or redundant array of independent disks (RAID), plus hot spares in the storage subsystem) so that the metadata servers always have a robust copy of the system metadata.
- All volumes in the system storage pool must be of the same type of backend storage and must be an IBM storage subsystem (for example, IBM Enterprise Storage Server or IBM SAN Volume Controller).

  Note: You can use SAN Volume Controller introduce mixed storage provided that the cluster only sees the SAN Volume Controller virtual devices.
- All volumes in the system storage pool must be visible to all metadata servers in the cluster.
- Each volume in the system storage pool must be at least 2 GB in size.
- You can assign up to 127 dual-path volumes to the system storage pool.

Chapter 5. Planning the cluster configuration

This topic describes how to plan the cluster configuration.

The SAN File System cluster consists of from two to eight metadata server engines that communicate with each other over an IP LAN. Each metadata server engine contains metadata server software and administrative server software. One of the metadata server engines is defined as the *master* metadata server

Perform these tasks to plan the cluster configuration:

- Determine how many metadata servers you will need based on the application workload or the amount of storage managed by the SAN File System.
- Determine how much storage is needed for the metadata server cache.
- Determine how to balance the workload across the metadata servers.

How many metadata servers do I need?

This topic helps you determine the number of metadata servers that are needed to handle the workload in your environment.

Estimating the number of metadata servers requires a strong understanding of your environment and application workloads. If you do not have these statistics, it is recommended that you start with two to three metadata servers. Metadata servers can be added dynamically at a later time based upon actual observed performance. Note that three metadata servers would allow for additional metadata-server capacity and higher availability in case a metadata server needs service.

There are two methods for estimating the number of metadata servers, depending on the level of understanding you have with your environment and application workload: available storage and application workload. The application-workload method generally gives a more accurate picture of what to expect, but it requires a strong understanding of the I/O pattern for the application workload. The available-storage method, based on the number of disk drives, generally provides a simpler approach that is based on the physical attributes of the environment. The approach might not be as accurate for workloads that have very high or low metadata-transaction rates. Typically, workloads that perform a high number of file creations or deletions or space allocation activities result in high amounts of metadata traffic, whereas workloads that perform I/O operations over a few, fixed number of file system objects are generally not metadata intensive.

The number of metadata servers needed is proportional to the sum of all metadata transactions that all the connected clients generate for any workload. However, because there is an intermediate metadata cache on the SAN File System clients, under typical working conditions the volume of metadata transactions or metadata server operations per second (OPs) might be relatively few compared to the volume of file operations per second (FOPs) produced from a given workload of application operations per second (APPOPS).

The cache on the SAN File System client plays an important role in the SAN File System operation and generally operates like any other least-recently-used object cache. The cache has the typical characteristics of a cache in the sense that the larger the cache, the higher the performance, or the larger the hotset size (number of objects), the greater the potential for lower performance.

Applications that use few file system objects and are I/O intensive over a few file system objects tend to have the smallest hotset size. Hence, they could potentially perform the best under SAN File System.

Sizing the metadata servers based on application workload

This topic helps you determine the number of metadata servers that are needed to handle the workload based on the application workload.

It is importance to understand the right parameters for loading and sizing the SAN File System to help determine how many metadata servers are needed for a given client workload. Important factors to consider are:

- Number and the mix of file system objects, such as files, directories and symbolic links that would be involved in the combined workload as seen by the Metadata server cluster, and how many filesets those file system objects are partitioned into.
- Size and mix of the hotset of those objects and filesets that each client would expect to operate upon with their respective applications.
- Typical file operations that a client application might generate. This exercise requires expertise with the application under consideration. However, FOPs for many well-known and standard workload classifications can be used to estimate this information.
- An estimate for the type of caching that the metadata will have for the given application load.

This estimate can be based on the mix of FOPs that the application generates, the hotset of objects and the size of the metadata cache. However, various client installable file systems might have varying amounts of memory that can be used for caching the metadata.

Perform the following steps to size the metadata servers based on application workload and characteristics is based on the following steps:

1. Understand the type of workload that the application generates. A reference table for a variety of standard workload types has been presented to help map your application to one or more of the standard application types.
2. Gather specific workload and environment information by answering a set of questions. Because it may be difficult to get answers to all the questions for all the environments, we have divided the questionnaire into mandatory and optional parts.
3. Calculate the number of file operations that will be generated per second for the specific workload and environment, based on the cache effectiveness measure provided for various workload, and, based on that, calculate the number of metadata server operations that will be generated.
4. calculate the number of metadata servers required based on the number of metadata server operations generated per second and the sustained rate of transaction serving capacity that you assume for the metadata server.

The next few sections take you through each of the steps in detail.

Use the Metadata servers — application workload method worksheet to help you size the metadata servers.

Understanding the workload

This topic helps you determine the workload generated by various application types.

The first step in sizing is to get a feel for how many objects you have to deal with at a steady state workload or a peak workload. Some sizing exercises are better done by using peak application workloads rather than steady state workloads. The choice is up to you. The application should be sized with respect to name, space, or file system objects. You should obtain the hotset information because this number can have a direct impact on the metadata cache performance. The hotset is the number of active objects that the application will be accessing during its peak or steady state operation. The exact determination of the hotset depends upon the type of application that you are using. For example, for a mail server application, although the installation may have millions of user accounts, if at any time only ten percent of the users are active, then the hotset is ten percent of the total number of objects.

To classify an application, you should essentially attempt to break down the application‘s interaction with a generic file system. Having detailed information about the application makes it easier. From a sizing perspective, it is important to understand the mix of file operations generated by the application. For your reference, the following table provides a sample of typical application types and the mix of file operations that they typically generate. Try to place your application in one of the categories listed. If it does not fall in any of the categories listed, try to gather similar data for your application type.

Op Type	Spec 1997	Web server	Web proxy	Database (OLTP)	Peer-Peer	Mail server	News server	D bench	Warehouse (DB2®) using Direct I/O and DMS	Warehouse (DB2) using Direct I/O and DMS
lookup	27%	14%	14%	0%	1%	27%	1%	61%	1%	13%
read	18%	28%	6%	61%	54%	14%	22%	3%	15%	0%
read direct	0%	0%	0%	0%	0%	0%	0%	0%	32%	78%
write	9%	0%	23%	31%	35%	24%	64%	16%	32%	4%
getattr	11%	55%	18%	3%	1%	3%	0%	7%	0%	0%
readlink	7%	0%	0%	0%	0%	0%	0%	8%	0%	0%
readdir	2%	1%	1%	0%	0%	0%	1%	0%	0%	0%
create	1%	0%	11%	0%	1%	0%	1%	1%	0%	0%
remove	1%	0%	11%	0%	1%	0%	1%	1%	0%	0%
mkdir	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
fsstat	1%	1%	1%	0%	1%	1%	1%	0%	0%	0%
setattr	1%	0%	0%	0%	0%	4%	0%	0%	0%	0%
readdir plus	9%	0%	0%	0%	0%	0%	0%	0%	0%	0%
access	7%	1%	4%	1%	1%	3%	1%	1%	0%	1%
commit	5%	0%	1%	4%	5%	24%	8%	0%	0%	0%
map/ unmap	0%	0%	0%	0%	0%	0%	0%	0%	0%	3%

Gathering application and environment characteristics

This topic helps you determine the characteristics of the applications and their workload

The following questions will help you gather important information about your applications and environment. Note that it is not necessarily true that all clients generate the same and uniform workloads, though in many cases you might see such deployments. Provide complete and accurate answers to as many questions as possible before moving on to the next step. The answers to these questions provide information about key workload characteristics that are specific to your environment and that will be used for adjustment to the final sizing numbers.

- What are the primary applications (for example, DB2, sendmail)?
- What is the average file size for each application?
- What is the size of the hotset (number of objects) for each application?
- What are the client types, and number of each?
- How many clients are running each application (for example, how many are running DB2, IBM WebSphere®, or both)?
- What is the average directory size?
- Does the application do a dirwalk?
- Do multiple clients share objects with read and write operations?
- What is the application transaction rate (peak and steady state) for each application? Express this in terms native to the application (for example, Web connects/second or orders/second).

Calculating the number of file operations

This topic helps you determine the calculate the number of file operations per second.

For various application types, the following table provides mapping between application transaction rate (APPOP) and file operations (FOP), that is, the number of expected file operations for each application transaction, and file operations (FOP) to metadata server operation (the number of expected metadata server operations for each file operation). The values in this table are based upon IBM internal testing and estimates based upon those test results.

Application type	Typical metric	Average FOP/APPOP	Average metadata server OP/FOP
Mail server	Mails per second	20	3%
Oltp database back end	Transactions per second	2	5%
Data Warehouse database back end	Tuples per second	0.1	10%
Office workgroup	Documents per second	10	1%
Web server	Connections per second	5	10%
Web proxy	Pages per second	15	10%
Peer-to-peer	Files per second	12	10%
Network File System Serving	Megabytes per second	30	10%
Common Internet File System Serving	Megabytes per second	30	10%
Compile Build (Development)	Files built per second	10	5%
User Folder Serving	Files used per second	5	3%

This table can be used to help estimate the rate of metadata server operations generated by your application, given your application‘s transaction serving rate. For example, if you have a mail server and you expect to serve 50 mails per second, the expected number of metadata server operations generated would be 30.

After the number of metadata server operations is determined for your application, adjust the estimate using the previous table and answers to the application and environment questionnaire. Use the adjustment factors in the following table, as applicable.

	Adjustment factor
Average file size larger than 500 Mb	+ 20%
File hotset > 10K	+ 20% for each 10K
Windows client	+ 20%
Number of clients (Uniform)	Multiply By # of Clients
Average directory size larger than 100 objects	+ 10%
Special application dirwalk	+ 20%
Object-sharing R/W	+ 20%

Calculating the number of metadata servers

This topic helps you determine the calculate the number of metadata server that are needed for your SAN File System.

The final step is to calculate how many metadata servers you need in your SAN File System environment. This is calculated by dividing the metadata server operations generated per second by the sustained transaction serving rate of the metadata server, which is 1 500.

Sizing the metadata server based on available storage

This topic helps you determine the number of metadata servers that are needed to handle the workload based on the number of hard disk drive available to SAN File System.

This method for determining the number of metadata servers looks at the number of physical data disk drives in the SAN File System configuration and performs a calculation from this number. This method relies on the fact that hard disk drives can provide a fixed amount of I/O per drive. The more drives, the more I/O throughput that can be obtained. As the number of I/Os is scaled up, the number of metadata server transactions is also scaled up, which has a direct correlation to how many metadata servers are required.

The calculation for this method is:

```
metadata servers = ( hard disk drives / 50 ) * application factor
```

where:

Hard disk drives

Number of physical hard disk drives for the data (all drives in all storage pools except for the system pool). This includes physical drives but not logical drives. If there is a RAID 5 LUN that consists of 5 drives, this should be counted as 5 drives. Hot spares or standby drives should not be included in this number. If you are using RAID 5 and a 5 + P array, you must count all six drives because the data is striped across all six drives.

Application factor

A factor from 0.5 to 2.0. If the application is cache friendly for the SAN File System client, use 0.5 as the application factor. When the application factor is 0.5, we use one metadata server for each 100 hard disk drives. TPC-H, data warehouse and similar applications are in this category. If you know the application is client cache unfriendly, create-and-delete intensive, or if you want to be conservative, you might use 2 as the application factor. In this case, each metadata server supports about 25 hard disk drives. Otherwise, use 1 for the application factor.

Use the Metadata servers—available storage worksheet to help you size the metadata servers.

Sizing the system storage pool

This topic helps you determine the amount of storage that is required for the system storage pool.

SAN File System stores the system and file metadata on the system storage pool. The total space required to store the metadata depends on the total number of objects in each fileset and the exact mix of the objects, such as directories, files, and symbolic links.

Generally, a local file systems requires approximately 3 MB of metadata for every 100 MB of actual data that a file system can hold, which is about three percent. However, SAN File System generally requires approximately 5 MB of metadata for every 100 MB of actual data, which is about five percent. If there are FlashCopy images to be maintained, plan on having at least an extra four percent for every FlashCopy image to be maintained; that is, four percent of the space of the fileset or filesets used with FlashCopy images. As a rough guide, using five percent of the current file system space (which includes metadata) for the metadata storage pool will typically provide adequate space.

The System storage pool worksheet provides calculations for more accurate estimate of the space requirements for the system storage pool.

The total space required depends primarily on the amount of the data and the exact mix of objects in each fileset. For each object, there are two components to the storage requirement: a variable component and a fixed component. The variable component depends on the size of the object and the name length. The following table has the figures to be used for the calculations.

Object type	Fixed storage	Variable component based on object size	Variable component based on name length
Files	560 bytes	415 bytes/MB	3 * name length
Directories	560 bytes	–	3 * name length
Symbolic links	560 bytes	–	3 * name length
Hard links	90 bytes	–	3 * name length
FIFO objects	560 bytes	–	3 * name length

The fixed component requires 128 MB for each metadata server.

Note: SAN File System supports the FlashCopy function, which performs a copy on write (COW) when updates occur. This require additional space in the system storage pool. Because the design is based on COW, the additional space required depends on the number of times a FlashCopy image has been taken and the percent of data that has been modified since the last FlashCopy image. For example, if a fileset requires 100 MB for its metadata component and 80 MB for its data component, and if 50 percent of the data has been modified since the last FlashCopy image was created, the additional space required would be 100 MB + 50% of 80 MB, which equals 140 MB.

How do I balance the workload?

This topic provides information to plan for balancing the workload (filesets) among the metadata servers.

Each metadata server in a cluster, including the master metadata server, is assigned a workload. The *workload* is the amount of processing that is required to manage the locks, leases, and metadata on behalf of the clients when they request access to data that resides in a fileset in the global namespace. To the client, the fileset appears to be another directory, but to the metadata server, a *fileset* is an amount of workload. For example, a fileset that contains data that is frequently accessed by clients will have a higher workload than a fileset that stores archived data.

During client setup, a client is given the address of one of the metadata servers for initial contact and metadata server cluster discovery. When the client issues a request to access data, it is automatically directed to the appropriate metadata server to obtain the metadata and locks required to access the data.

Workload-balance methods

There are two methods for assigning fileset to metadata servers:

- You can choose to *statically* assign filesets to metadata servers. The master metadata server will make the assignment if the assigned metadata server is running. If it is not running, the master metadata server will temporarily assign the fileset to another metadata server. Once the assigned metadata server is running, the master metadata server will automatically reassign the fileset back to its assigned metadata server.
- You can choose to have the SAN File System *dynamically* assign a fileset to a metadata server. The master metadata server will choose a metadata server to be assigned based on a workload distribution algorithm. This algorithm evenly distributes the filesets among the metadata servers in the cluster. Dynamic-fileset assignment works best when there are many filesets with uniform workloads.

High-availability considerations

You need enough spare workload capacity such that a metadata server can be taken offline and its filesets can be distributed to other metadata servers without overloading them. To ensure that there is enough spare workload capacity:

- Ensure that in an *N*-server cluster configuration, each metadata server runs at most (*N*–1)/*N* percent of capacity. This way, if one metadata server goes offline, the remaining *N*–1 metadata servers do not run at 100%.

- Because you cannot control which metadata server is the master, reserve spare capacity (such as 5%) on each metadata server for the master workload.
- If you use static-fileset assignments, reserve a spare metadata server in the cluster to take on the workload of another metadata server that goes offline and preserve the workload balance. A *spare metadata server* is an idle metadata server that has no statically assigned filesets. It can take on the workload of any metadata server without impacting the remaining metadata servers in the cluster.

Chapter 6. Planning the file management strategy

This topic describes how to plan the file management strategy.

Perform these tasks to plan the file management strategy:
- Determine whether you need a default storage pool.
- Define the file-placement policy rules you need.

Use the Policies worksheet to define the file placement rules for your installation.

Do I need a default storage pool?

This topic provides information to help you determine whether you need to a default storage pool and what storage pools can be used.

If you intend to configure all SAN File System clients to see all storage volumes, you can set one storage pool to be the default storage pool. Files created by a client that do not meet any of the rules in the active policy set will be stored in this storage pool.

If you configure your SAN clients and storage volumes into zones so that clients can see only a subset of all storage volumes, you will not be able to set a default storage pool. In this case, you can choose between two options:
- Configure a single storage pool that is accessible from all client zones (each of the volumes in that storage pool must be accessible from all clients). Then, you can define a default rule in the active policy set so that if a file does not meet any of the other rules in the policy set, it is stored in the default storage pool.
- Create a single storage pool for each client zone that can be used like a default storage pool. Again, you must define a default rule in the active policy set for each set of clients in each zone so that if, for a particular zone, a file does not meet any of the other placement rules, it will be placed in this storage pool.

If a client attempts to create a file and there is no policy set rule for placing that file into a storage pool, the SAN File System will generate an error.

What file-placement policies do I want?

This topic describes considerations for determining what file-placement policies you want to implement.

The SAN File System uses policy rules to determine how files are placed in the user storage pools. Each rule in a policy specifies the fileset to which the rule applies and the storage pool in which user data will be stored. In addition, it includes conditions for storing the files. These conditions are typically based on file attributes, such as file name, owner, group ID of owner, or the system creating the file.

Although defining rules for file placement based on creation time, user ID, or group is supported, use these clauses with care. When files are restored from a backup or migrated, these attributes are changed. For example, the following changes occur when files are restored from backup:

- Creation time is set to the time the files were restored.
- The user and group are set to the user and group of the application used to restore the files.

You can define multiple policies for the SAN File System. However, only one policy can be active as any one time.

Important: In a non-uniform zone configuration, the clients can access only the storage pools in their zone. You must verify which filesets each client uses, and ensure that the policy rules use only storage pools that the clients can access.

Chapter 7. Validating the existing SAN infrastructure

This topic describes how to validate the existing SAN infrastructure to ensure that it supports the implementation of the SAN File System.

Before you begin planning for the installation of the SAN File System, you should first ensure that your existing SAN infrastructure can support the SAN File System. You need to validate:

- IP network
- Cluster engine connections
- Lightweight Directory Access Protocol (LDAP) server
- Storage subsystems
- Switches

IP network

This topic describes the IP network considerations for implementing the SAN File System.

SAN File System supports two to eight metadata servers, each running on a separate engine. Each engine has at least two IP addresses, one for the metadata server and one for the RSA:

- Copper or fiber Ethernet adapter is used by the metadata servers to communicate with each other and the clients through an existing IP network.
- RSA, which uses copper Ethernet cabling, is used to establish a remote console to the engine from the master console.

To set up an IP network for SAN File System:

- Assess the requirements for the interaction of your existing LANs, with the SAN File System clients and metadata servers. Your requirements should include administrative LAN connections to new SAN hardware, storage devices, and SAN File System engines. SAN File System requires a low latency network for metadata-server-to-metadata-server and metadata-server-to-client communication.
- Determine if the existing LAN has enough switch ports for the metadata servers.
- Determine which software modifications will be needed to support the resulting LAN topology (clients, servers, and switches).
- Determine which hardware modifications, including network interface cards, cabling, and switches, are needed to support the required LAN topology.

Tip:

- The metadata server IP addresses must be on a common subnet.
- The IP network over which the metadata servers and client communicate must be secure, using physical security or IPSec.

Cluster engine connections

This topic describes the several ways that cluster engines are connected in SAN File System.

The SAN File System cluster engines connect with each other in three ways:

- Ethernet LAN connections of the engine network interfaces or ports. These can be the copper interfaces on the motherboard of the xSeries® 345 or fiber interfaces on the optional fiber GigE card.
- Ethernet LAN connections of the network interfaces on the RSA IIs in each engine.
- The RS-485 or serial connection between the RSA IIs.

These connections are used in the following ways:

- The Ethernet port in an metadata server engine is used to:
  - Communicate with other metadata server nodes. This communication consists of cluster heartbeats as well as request-response messages that occur as part of metadata transactions in the metadata server nodes. For example, if a subordinate metadata server needs some space from a storage pool, it sends a request to the master metadata server which allocates a partition to the subordinate. This results in messages going back and forth between the two metadata servers through their Ethernet ports.
  - Enable communication between SAN File System clients and the metadata servers.
  - Send SNMP traps generated by SAN File System.
- The RSA II Ethernet connection is used to:
  - Send SNMP traps from the RSA II to the master console.
  - Define the KVM of the master console as the console of the metadata server engine in which the RSA II resides.
  - Update the RSA II firmware through its Web interface.
- The RSA II serial connection (RS-485) is used when the SAN File System administrative agent , running on the metadata servers, requests that the local RSA II obtain information from a remote RSA II. This flow of information occurs over the RS-485 connection. This connection is also used for the requests generated by one metadata server to stop another that is not functioning properly.

The Ethernet connections use a customer LAN that does not need to be isolated. For performance reasons, the metadata server engines and the SAN File System clients should all be on the same subnet; that is, there should be no routers or gateways in their paths. If you deploy a metadata server cluster on a heavily loaded IP-network, it is possible that there could be delays in getting heartbeat traffic across the network. This could result in unnecessary metadata server fail-overs. A possible solution to situations such as these would be a dedicated VLAN.

LDAP considerations

This topic describes the Lightweight Directory Access Protocol (LDAP) requirements for SAN File System.

SAN File System uses an LDAP server to authenticate and authorize each administrative operation. Each SAN File System administrative user must have an entry in the LDAP server database.

SAN File System requires that a Lightweight Directory Access Protocol (LDAP) server be installed in your network. LDAP servers that can be used with the SAN File System include:

- IBM Directory Server 5.1 for Windows
- IBM Directory Server 5.1 for Linux
- OpenLDAP/Linux
- Microsoft Active Directory

Storage subsystem considerations

This topic describes the storage subsystems that are supported by the SAN File System.

SAN File System supports heterogeneous, simultaneously-connected Fibre Channel storage subsystems on clients with host bus adapter (HBA) sharing, subject to the limitations of the client platform, drivers, and storage vendors.

SAN File System supports an unlimited number of LUNs for user data storage. However, the amount of user data storage that you can have in your environment is determined by the amount of storage that is supported by the storage subsystems and the client operating systems.

For more information about supported storage subsystems, refer to the following Web site:

www.ibm.com/storage/support/sanfs

System storage pool

Currently, SAN File System supports only these storage subsystems for use in the system storage pool:

- The IBM TotalStorage Enterprise Storage Server (ESS), models 2105-F20 and 2105-800
- The IBM TotalStorage SAN Volume Controller, model 2145 with storage subsystems that are supported by SAN Volume Controller
- IBM TotalStorage DS4300 Turbo (IBM TotalStorage FAStT600T), DS4400 (FAStT700), and DS4500 (FAStT900) running firmware version 8.4 on the storage device and software version 8.41 on the client platforms

Ensure that the IDs for any LUNs that are used by the system storage pool starts with 0. Refer to your storage documentation for information about assigning LUN IDs.

Note: If you are using a DS4000 Series storage subsystem, all DS4000 LUNs that are used in the system storage pool must be in a separate partition from the LUNs that are used by the user storage pools.

Refer to the IBM TotalStorage Web site for the supported code levels of these storage subsystems.

www.ibm.com/storage/support

User storage pool

For user storage pools, SAN File System is designed to work with FCP-compliant storage subsystems that meet the following qualifications:
- Conforms to SCSI standards for device driver interface, including unique device identification.
- Supports the required device drivers and operating-system stack.
- Are SAN-attached to the client machines.

Tip: Consider any restrictions imposed by the storage subsystem, host bus adapter (HBA) cards, device drivers, and client platforms that are used in your SAN File System environment to ensure that they are all compatible.

DS4000 LUNs assigned to user storage pools can be shared by multiple SAN File System clients as long as the clients are aware of one operating system type. LUNs within a DS4000 partition can only be used by one operating system. If you use DS4000 firmware lower than version 8.41, you cannot have more than 32 LUNs per partition.

Storage subsystems other than ESS or SAN Volume Controller may require additional, manual configuration to be detected and used by SAN File System.

Refer to the platform support documentation for a list of storage subsystems that are supported for each client platform in your environment.

Switch considerations

This topic provides the considerations for switches in the existing SAN infrastructure.

Consider the following items when setting up the switches in the SAN infrastructure:
- All interconnected switches in each fabric within the SAN should be homogeneous; that is, all switches that are interconnected must be from the same vendor. Note that if you have more than one fabric in the SAN, each fabric may consist of switches from different vendors.
- You need to make sure that there are enough switch ports, GBICs, and fiber cables to support the SAN File System engines and any new storage subsystems that you are adding.
  - SAN File System requires each engine to have one dual-port or two single-port host bus adapters (HBAs). Each HBA port requires a port on a switch or fabric. Therefore, you must have two switch ports for each engine.

    Note: HBAs on SAN File System metadata servers have LC-type fiber cable connectors.
  - Each storage subsystem that is connected to the SAN must have a minimum of four ports for multi-pathing, failover, and high availability: two ports for engine connectivity and two ports for SAN File System client connectivity. Consult the storage subsystem documentation to determine if additional ports are needed.
- The SAN configuration for the SAN File System must not have a single point of failure. This means that connectivity must be guaranteed in case there is a loss of an HBA, switch, small form factor pluggable (SFP) transceiver, fiber cable, or storage controller. To ensure that there is no single point of failure, consider using dual switch fabrics and separate HBAs within each platform.

- It is recommended that you separate the fabrics between the HBA ports within the engines. By separating the fabrics, you will avoid a single path of failure for the fabric services, such as the name server. Also, by creating separate fabrics and using strict switch zoning, you can limit the number of ″paths″ to each LUN to two, thereby allowing more LUNs to be assigned to SAN File System. Currently, the SAN File System supports a maximum of 127 dual-path LUNs in the system storage pool. Note that this number takes into consideration the two LUNs that are reserved for the boot disk and local service processor.

  This number of LUNs is further reduced by half if four paths to each LUN are used. You may wish to use fabric or switch zoning to limit the number of paths to the system storage pool to two paths per LUN to maximize LUN usage while providing a redundant connection to each LUN.
- Multi-pathing is required for volumes in the system storage pool.
- DS4000 (FAStT) may have additional connectivity requirements. Refer to your DS4000 (FAStT) documentation for more information.
- A user storage pool that is in use by a Linux Client is limited to 254 single-path LUNs. Consider creating multiple user storage pools to maximize the LUNs used by non-Linux clients. Also consider whether using a single path to the Linux client would suffice because dual pathing reduces the number of LUNs that may be used by a Linux client by half.
- When setting up switch zoning in your SAN, use hard zoning whenever possible. It is a best-practice to zone the switches such that no HBA port resides in the same switch zone as any other HBA port. It is a SAN File System requirement that hard zoning be used to isolate the system storage pool from access by any systems other than the metadata servers.

Chapter 8. Planning the security strategy

This topic describes how to plan the security strategy

You can control access to data by implementing LUN masking, switch zoning, and file-permission semantics that are native tot the client platforms.

You can also control administrative access to the SAN File System through an LDAP server. Administrative users log in to the SAN File System (through either the SAN File System console or the Administrative command-line interface) and are authenticated through the LDAP server. SAN File System uses roles to determine the level of access that each administrator has.

Perform these tasks to plan the security strategy:

- Determine a strategy for zoning storage, clients, and metadata servers.
- Determine a strategy for masking LUNs, if appropriate.
- Determine a strategy for limiting access to specific files or sets of files using file-permission semantics.
- Gather information needed to configure LDAP.

How do I set up zones in SAN File System?

This topic provides the considerations for setting up zones in the SAN infrastructure.

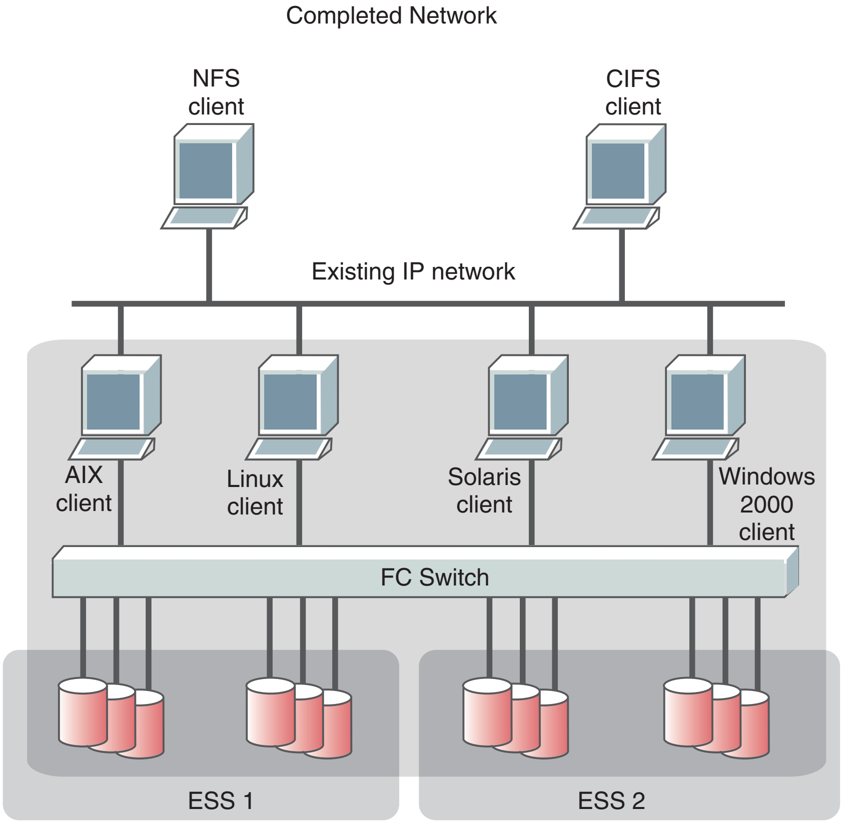
Client zoning

This topic provides the considerations for setting up zones for the clients and the LUNs in the user storage pools.

When a file is created or modified from a SAN File System client, user data is stored in user storage pools that are made up of volumes (or LUNs). Each SAN File System client needs access to all volumes that comprise the user storage pool where data from a fileset that can be accessed by that client will be stored. There are two ways to zone SAN File System clients: uniform and non-uniform.

Uniform zone configuration

In a *uniform zone* configuration, you create a single zone in which all clients have access to all SAN File System volumes. The following figure show an example of a uniform zone configuration.



The advantages of a uniform zone configuration are:

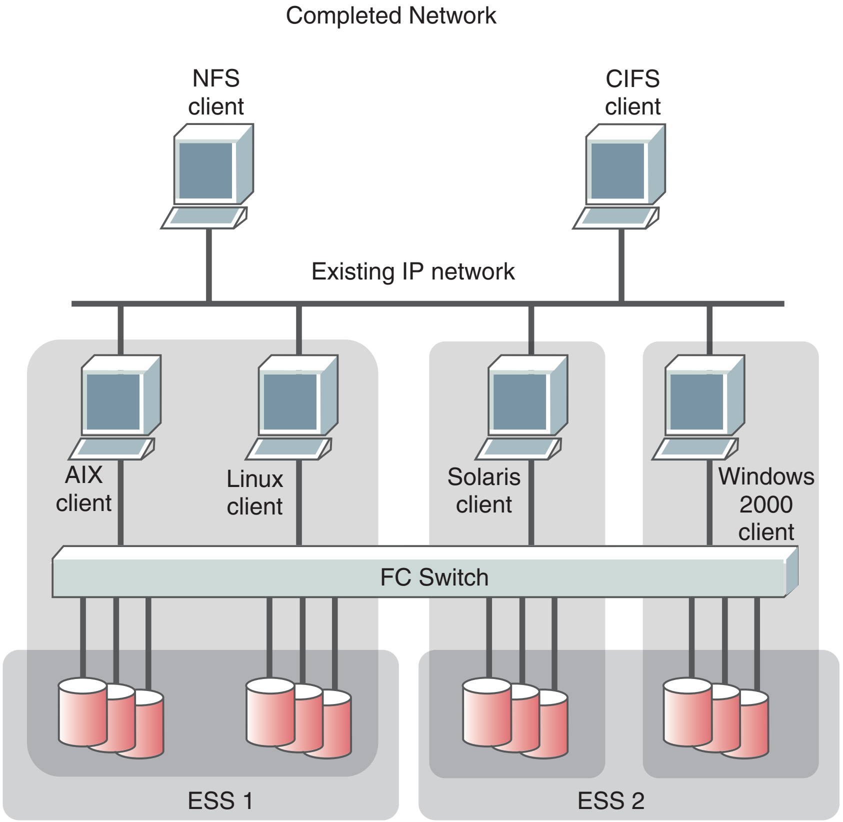
- Simplifies the management of policies, filesets, and user storage pools because all clients can access all volumes.

The disadvantages of a uniform zone configuration are:

- Any client could potentially access sensitive data unless file-permission semantics are used to control access at a file level.

Non-uniform zone configuration

In a *non-uniform zone* configuration, you create multiple zones in which clients have access to only the volumes that they will actually need. The following figure show an example of a non-uniform zone configuration.



You must ensure that all clients in a non-uniform zone configuration can access all of the volumes in any user storage pool that can be used by filesets in use by that client. If a client tries to read or write data on a volume that it cannot access, SAN File System will return an I/O error. File system operations that involve only metadata, such as changing the current directory or listing files, will not receive an I/O error because those functions do not require access to the user storage pool.

These are the advantages of a non-uniform zone configuration:

- Provides another layer of security for sensitive data by configuring LUNs to be accessed only by those clients that need to see that data.
- Avoids configuring clients across multiple vendor storage subsystems simultaneously.
- Allows great scaling because not all of the volumes must be seen by all of the clients.

These are the disadvantages of a non-uniform zone configuration:

- Complicates the management of policies, filesets, and user storage pools because you must ensure that the clients can access all of the volumes through the active policy.

Considerations

These are considerations for planning your client zones:

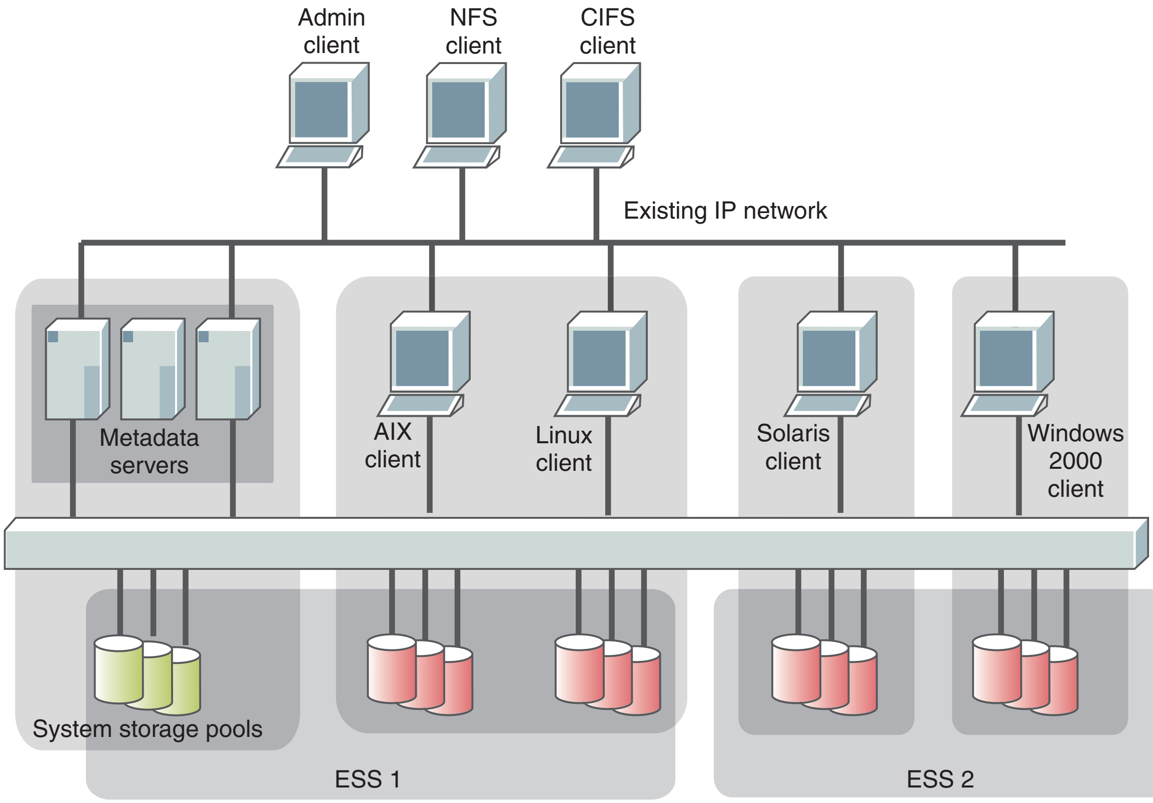
- The clients need access only to the user storage pools that they use. They must not have access to the system storage pool. Configure the client zones to encompass only those LUNs associated with the user storage pools.

- All clients can access all storage subsystems that are attached to SAN File System, unless you use zoning to limit the clients' access to specific devices or LUNs. This enables data sharing among heterogeneous clients.
- Ensure that the storage subsystem you are using allows you to mask a single LUN to different operating system types.
- LUNs should be masked in such a way that only the clients that are intended to use that LUN have access.
- With DS4000 (FAStT), a specific client platform can only be defined to one ″host group″ so they cannot be masked to access more than one user storage pool.
- All volumes in a storage pool need to be accessible by all the clients that will use that storage pool.
- All system storage pool volumes must be seen by all metadata servers and only by metadata servers.
- SAN File System volumes must be masked or zoned so that access by non-SAN File System clients or application servers is denied.

Metadata server zoning

This topic provides the considerations for setting up a zone for metadata servers and the LUNs in the system storage pool.

It is not necessary for metadata servers to access the user storage pools. Although data access is coordinated by the metadata servers, data operations are always performed by a client. The metadata servers only need access to the volumes in the system storage pool, and should be zoned to limit their access to only those volumes.



Considerations

SAN File System allows you to zone the metadata servers and clients so that they access only the LUNs that they need during LUN discovery and normal operation.

- The metadata servers need access only to the system storage pools. They do not need access to the user storage pools. Configure the metadata server zone to encompass only those LUNs associated with the system storage pool.
- All system storage pool volumes must be seen by all metadata servers and only by metadata servers.
- No user storage pool volumes should be seen by the metadata servers.
- Zone the metadata servers so that the number of paths from each metadata server to each LUN is limited to two. This maximizes the number of dual-access LUNs that are accessible to the cluster.

What LUNs do I want to mask?

This topic describes how to decide whether you want to mask client LUNs.

LUN masking gives you the ability to exclusively assign LUNs to one or more clients. Only those clients that are assigned to the LUN can access it. All LUNs that are not assigned to a client are hidden from that client.

You would use LUN masking, for example, if you were to install a user application in the global namespace to limit the use of that application.

Tip:

- Ensure that your host bus adapter (HBA) driver supports LUN masking.
- IDs for any LUNs that are used by system storage pool must start with 0.

Do I want to limit client access to some files?

This topic describes considerations for limiting file access to only certain clients for security reasons.

You can prevent clients from accessing sensitive data using the file-permission semantics that are defined by the client platform.

In the current release of SAN File System, it is recommended that you separate files in the filesets for each client platform; that is, a Windows client should create files only within filesets dedicated to Windows files, and a UNIX-based client should create files only within filesets dedicated to UNIX. This is referred to as the *primary allegiance* of a fileset, either Windows or UNIX. The different client platforms can, however, share files in a common fileset if the permissions allow. Therefore, it is important to set up your ACLs on the clients to accomplish this goal.

How do I ensure consistency in the Access Control Entries when migrating

This topic describes how to ensure consistency in the Access Control Entries (ACEs) during migration.

The migration utility copies all of the security information, including the ACEs, from the source to the destination tree. Because of how the operating system handles inheritance of ACEs and dynamic propagation, it is possible that the destination tree ACE list is not the same as the source ACE list.

The migration tool disables the inheritance feature on the destination folder so this tree does not inherit any ACEs from its parents. This also prevents the destination folder from passing ACEs down to its children. If the source data being migrated was inheriting ACEs from its parents in the source, these inherited ACEs will not be applied in the destination folder. In this case, you will see an ACE mismatch error during the verify phase of migration.

This condition might occur when the newly-created destination folder does not have the same ACEs as the source had before the migration was run. To avoid this condition, you can manually create the destination folder and manually apply the source ACEs to the destination before starting the migration.

What do I need to configure LDAP?

This topic describes the information that you need to configure LDAP.

Some configuration of the LDAP server is required by the SAN File System for it to use the LDAP server to authenticate SAN File System administrators. For example, the SAN File System requires an authorized LDAP username that can browse the LDAP tree where the users and roles are stored.

The requirements to configure the SAN File System for LDAP include:

- You must be able to create four objects under one parent distinguished name (DN), one for each SAN File System role.
- Each role object must contain an attribute that supports multiple DNs.
- You must be able to create an object for each SAN File System administrative user.
- Each administrative user object must contain an attribute that can be used to log in to the SAN File System console or CLI, and a userPassword attribute.
- If you are accessing the LDAP server over Secure Sockets Layer (SSL), a public SSL authorization certificate (key) must be included when the truststore is created during installation.

Use the LDAP configuration worksheet to gather the information that you need to provide during the installation of the SAN File System.

What users and roles do I need to create?

This topic provides an overview of users and roles as they relate to the SAN File System.

A SAN File System *administrator* is the same as a *user* in the LDAP database entries. A user can use the administrative command-line interface and the SAN File System console.

SAN File System administrative users must have an assigned role. The role determines the scope of commands which an administrator can execute. SAN File System supports a predefined set of roles. These roles are Monitor, Operator, Backup, and Administrator. You need to define these roles in your LDAP server.

You also need to define all of the administrative users that will need to access the SAN File System. The role that you define for each user determines the level of access for that user.

Some configuration of your LDAP server is required for SAN File System to use LDAP for authenticating SAN File System administrators. SAN File System requires an authorized LDAP user name that can browse the LDAP tree where the users and roles are stored. If a secure LDAP connection is required, then the SSL certificate is needed. Fill in the tables below with your values.

You will also need to know the name of your LDAP certificate, which is used to create the Truststore. Get this information from your LDAP administrator.

Use the LDAP configuration worksheet to collect information for setting up users and roles in the LDAP server.

User roles

SAN File System provides different levels of user access that are assigned to specific administrative tasks in your environment. These access levels, or *user roles*, are one way to provide security. The following table describes the SAN File System user roles.

Table 1. SAN File System user roles

Role	Level	Description
Monitor	Basic level of access	Allows you to obtain basic status information about the cluster, display the message logs, display the rules in a policy, and list information regarding SAN File System elements such as storage pools, volumes, and filesets.
Backup	Monitor + backup access	Allows you to perform backup and recovery tasks in addition to all operations available to the Monitor role.
Operator	Backup + additional access	Allows you perform day-to-day operations and tasks requiring frequent modifications, in addition to all operations available to the Backup and Monitor roles.
Administrator	Full access	Provides you with full, unrestricted access to all administrative operations.

At least one user with Administrator access is required. You can also choose to define other roles as appropriate for your organization.

Chapter 9. Planning the backup and restore strategy

This topic provides an overview of how to plan the backup and restore strategy for the SAN File System.

SAN File System supports the use of backup tools that are already present in your environment. For example, if your enterprise currently uses a storage management product such as Tivoli Storage Manager (TSM), SAN File System clients can use the functions and features of that product to back up and restore files that reside in the SAN File System global namespace.

For backing up in a normal, available environment, you can use the FlashCopy image feature of SAN File System.

To prepare for disaster recovery in situations where SAN File System becomes unavailable, you can perform LUN-based backups using the instant copy features that exist in the storage subsystems that SAN File System supports. If your SAN storage subsystems do not offer copy services, you must back up for disaster recovery using third-party backup and restore applications.

Backup and restore

It is important to have a process for backing up your environment so that you can easily recover from a storage device failure or loss of data. In SAN File System, you must save both the file data and metadata together when you back up the global namespace. These are used to recreate your user data. You must also backup the system metadata, which is used to recreate the SAN File System configuration.

SAN File System does not provide backup and restore functionality; instead, it supports backup tools that are already present in your SAN environment. Depending on the type of failure, you might need to restore a single file, an older version of a file, a directory, a volume, or the entire system. SAN File System supports various options for protecting the system, including:

- Creating FlashCopy images
- Backing up files using third-party backup and restore applications that are already present in your environment, for example IBM Tivoli Storage Manager (TSM), Legato NetWorker, or VERITAS NetBackup
- Using copy services that exist in the underlying storage device (for example, FlashCopy and Peer-to-Peer Remote Copy (PPRC) functions in the IBM TotalStorage Enterprise Storage Server or IBM TotalStorage SAN Volume Controller
- Saving the cluster configuration (system metadata)

There are two basic methods available for backing up and restoring your data:

- File method
- Volume method

The file-based method saves and restores data at the file level. It uses the FlashCopy function or other third-party backup and restore application in your environment to back up or restore your user data. Use the file-based approach when files have been lost but the overall system remains healthy.

The volume-based method saves and restores data at the device level (that is, a "just-a-bunch-of-bytes" approach). To adopt the volume method, however, the actual copying and restoring of data must be provided as a service by the underlying storage subsystem. Use the volume-based approach when disaster strikes and the system, as well as the FlashCopy images, are unusable.

Tip: Your backup and restore process does not have to be centralized and homogenous, covering the entire SAN, although such a process simplifies the procedure. You can use the volume method even for a fragmented SAN that requires a piecemeal volume copy across two or more storage subsystems. In such a scenario, you would be responsible for manually managing those multiple backup sets as though they were a single backup set.

File-based data backup and restore

In a file-based backup, the smallest unit that you can restore is an individual file. There are two basic methods for backing up files:

- SAN File System FlashCopy, which backs up at the fileset level, but provides the ability to restore parts of the fileset, such as directories, groups of files, or individual files.
- Operating system utilities and vendor-provided backup and recovery tools and applications. These include utilities such as **tar**, **cpio**, **xcopy**, and applications such as Windows Backup, IBM Tivoli Storage Manager, VERITAS NetBackup, and Legato NetWorker. All these should be able to access the SAN File System global namespace exactly as they would a local drive.

In a situation where files have been lost but the overall system remains healthy, the first line of defense for restoring files is to have previously used the FlashCopy function to create an image of the files. The FlashCopy function is available in SAN File System. The FlashCopy function provides a space-efficient image of the contents of part of the global namespace at a particular moment in time. The FlashCopy image contains read-only copies of the files in a specific fileset as they exist at a specific point in time.

The FlashCopy image is stored in a special subdirectory named .flashcopy under the fileset's root attach point. After you create a FlashCopy image of a fileset, you can use standard backup tools on the SAN File System client to back up the files by specifying the path to the FlashCopy image instead of the path to the actual files. This allows users and applications to continue working with the actual files while the backup occurs.

You can use the FlashCopy image to restore the entire fileset or restore a single file.

When using a file-based backup method other than the FlashCopy function, be aware of the associated file metadata backup, which includes all the permissions and extended attributes of the files. This file metadata for Windows-created files can only be backed up completely from a Windows client. Similarly, file metadata for UNIX files can only be backed up completely from an UNIX client. Therefore, if it is important for you to preserve full file-attribute information, create separate filesets by primary allegiance. In other words, have certain filesets that only contain Windows-created files, and other filesets that only contain UNIX-created files. In this way, you can back up the filesets from the appropriate client operating system.

Limitations to file-based backup and restore

You must be aware of limitations that apply when backing up files that are used by both UNIX and Windows clients.

To request a backup from a UNIX client, you must have read permissions on all files and search permissions on all directories (typically a root user). To request a backup from a Windows client, you must have read permissions on all files and folders.

To avoid losing security metadata, files created on Windows operating systems must be backed up on Windows and files created on UNIX operating systems must be backed up on UNIX. Here are some sample solutions:

- Designate directories, including the contents of their child directories, as used for files created in either UNIX or Window. Then, you can back up the directory and all its descendents from its associated platform.
- Designate filesets as containing either UNIX files or Windows files, and back up filesets as units. This method is compatible with backing up from the FlashCopy image of a fileset.

You might have used special naming conventions to create files or directories. SAN File System does not interpret special naming conventions in any way. In addition, SAN File System cannot always prevent an administrator or any client user from creating a file that violates a naming convention.

Therefore, be aware of the naming conventions when performing backups. Ensure that backups of files in specially named directories are performed only from the same type of client that created the files. So, in an environment that has both UNIX and Windows clients, divide the backup process into multiple parts to prevent the loss of security attributes for files. The number of parts can be as few as two or as many as the number of directories in the global namespace, depending on the capabilities of the backup utility that a client uses.

Fileset considerations for backup and restore

From the SAN File System client perspective (and therefore from the backup application perspective), a fileset looks exactly like a regular subdirectory. From the metadata server and administrative server perspective, however, this is a fileset that is attached to an arbitrary subdirectory in the global namespace.

When a fileset resides within a directory, backup applications automatically create subdirectories when attempting to restore files to a directory that does not exist. In SAN File System, the subdirectory being created might have originally been part of another directory.

Tip: Restore the system metadata backup before restoring any subdirectories and files.

The **mkdrfile** command saves state information that you can use later to recreate this portion of the SAN File System if there is a disastrous loss of data. Because you can create regular subdirectories only from a client machine, the metadata server cannot recreate directory trees that contain a mix of filesets within subdirectories, and regular subdirectories. To simplify disaster recovery, attach filesets only to the global fileset (root directory) or to each other, but not to regular subdirectories. You can then use the disaster-recovery file created by the **mkdrfile** command to completely restore the top of the global namespace tree before using the client-based backup application to restore the rest of the global namespace.

FlashCopy image considerations for backup and restore

When you make a FlashCopy image, the .flashcopy subdirectory is created as read-only. The client backup application typically backs up the .flashcopy subdirectories along with all other directories or files. At restore time, however, the same backup application attempts to copy the original files back into the same subdirectories and fails. (You would not want the subdirectories there because they would appear to be valid FlashCopy images from the client perspective, when in reality, the metadata needed for the original FlashCopy images would be missing.)

One consequence of this process is that there is no way to restore your original FlashCopy images if you have lost your metadata in a disaster scenario. Only the original files are restored. However, if your backup application has the ability to restore files to a directory other than the files' original location (that is, to the grandparent directory two levels above the ./.flashcopy/*flashcopy_name* directory), then you have all the ingredients for a highly efficient backup, which leverages the FlashCopy image feature.

Tip: Periodically create FlashCopy images. They are the most efficient method for quickly backing up and restoring files in scenarios where the metadata is still available.

General considerations for backup and restore

These are some backup and restore considerations to keep in mind while planning your backup process:

- UNIX-based clients and Windows-based clients should be used to back up and restore files only within filesets that are dedicated to their respective client type.
- When using the Veritas NetBackup application's standard backup option, do not use the Windows NT® File System (NTFS) to ensure proper backups of data. When using Veritas NetBackup to restore data, use the override option, not the default option.
- To simplify disaster recovery, attach filesets only to the global fileset or to other fileset, and not to a directory.
- Backup to a SAN File System disk cannot be done with Windows-based clients using the Veritas Backup Exec 9.0 application; however, SAN File System data can be backed up to a tape device or a local disk, and data can be safely restored back to a SAN File System disk.
- SAN File System does not support the change-journal mechanism for Windows-based clients. Therefore, differential backups using Veritas Backup Exec do not work with SAN File System.
- Veritas Backup Exec's Advance Open File Option package does not work with SAN File System.

System metadata backup and restore

SAN File System manages data and metadata separately. When you back up a file, only the file's data and attributes are backed up. For disaster recovery purposes, you must also back up system metadata (which includes information about fileset attachment points, storage pools, volumes, and policies) separately.

You can create a file that contains a backup copy of system metadata either from the SAN File System console or administrative command-line interface using the **mkdrfile** command . The file, which is stored in the /usr/tank/server/DR directory on the master metadata server's local disk, contains everything that is required to recreate the system metadata. When needed, you can use the contents of this file (along with normal restore processes for file data) to recreate the state of the cluster.

To restore system metadata, you process the information that is contained in the system metadata backup file using the **builddrscript** command. This command creates several scripts that you must first review in order to obtain a restore scenario, and then run to recreate the SAN File System configuration. After the system configuration information is restored, you can then restore the user data files from the SAN File System clients.

You should run the **mkdrfile** and **builddrscript** commands often to ensure that any configuration changes are reflected in the output of these commands. You should store copies of the output of the **mkdrfile** and **builddrscript** commands in an easily recoverable location on backup media where critical system and application files are kept for backup and restore purposes.

Note: To assist in protecting against the corruption of metadata and other metadata failures, you can check the metadata from the SAN File System console by using the **startmetadatacheck** command. This command performs a consistency check on the system metadata, and optionally repairs any problems it finds. It allows you to check file metadata for one or more filesets, the system metadata, or both. There is also an option to check only the metadata structure, or to check the metadata structure and its contents.

There are three cases when you might need to perform a consistency check or repair operation:

- As part of a regularly scheduled cycle of preventive maintenance
- In response to an alert that recommends that this operation be performed (extra detail might be supplied that specifies the restore option that you must use in order to salvage the metadata)
- If metadata corruption (or any other SAN File System corruption) is suspected

If the check-metadata operation cannot resolve the problem, you must perform a full restore of SAN File System, beginning with restoring the metadata. It is critical that **mkdrfile** is run in order to recover from such a situation.

Volume-based data backup and restore

The volume-based approach backs up the entire global namespace, at the device level, in a single operation and restores the global namespace as a complete namespace. This approach uses the copy services features that exist in the storage subsystems (for example, the FlashCopy feature of the IBM TotalStorage Enterprise Storage Server).

When performing a volume-based backup, you must back up the volumes used as volumes in both user storage pools and system storage pool at the same time. All of the volumes in the user storage pools and the system storage pool must be in a static, consistent state to ensure a static state of the volumes both for the metadata and the user data.

Before performing a volume-based backup, you must quiesce the SAN File System clients and the cluster to ensure a consistent backup. You can also stop the cluster before performing a volume-based backup.

Tip: Use the cluster-transition timestamp to ensure that no unintended cluster transitions occurred during a volume-based backup. The **lscluster** command displays the cluster statistics.

Advantages to volume-based backup and restore

Using an volume-based backup and restore process has these advantages:

- The backup and restore is performed at the storage subsystem layer, so the storage engines are not involved in the backup process.
- The backup and restore deals with data at the byte level, and has the ability to back up and restore the entire SAN File System global namespace in a single operation.
- The backup and restore of the complete file system (metadata and file data) is performed as it happens at the volume level.

Limitations to volume-based backup and restore

These are some of the limitations to a volume-based backup and restore:

- It is not granular, and does not provide individual file or volume restore capability.
- You must save and restore all the volumes, including those containing metadata and file data. The volumes in the system storage pool and the user storage pools form a consistency group (that is, they must be backed up and restored together).
- The SAN File System clients and the cluster must be quiesced before performing a volume-based backup. In a fully-quiesced system, all file system activity stops and all buffers are flushed to disk. Volume-based backups are guaranteed to get a time-consistent view of the entire system in this state. In a partially-quiesced system, all file system activity stops and only the metadata buffers are flushed to disk. Volume-based backups are guaranteed to get a time-consistent view of the metadata, but not file data because the file buffers are not flushed to disk. Unless your applications can recover from incomplete data writes, your system should be in the fully-quiesced state.

What files should I include in my backup?

This topic lists the SAN File System files that you need to include when backing up the SAN File System.

- The metadata server disaster recovery file (which is generated using the **mkdrfile** command), which allows you to reconstruct filesets and their attach points. This file resides on the master metadata server boot disk.
- The SFSLCM ILM policy rule file, if it was created.
- The LDAP or NIS configuration files.
- The metadata server configuration files that define details such as cluster configuration, administrative server configuration, and so forth. These files reside on the master metadata server boot disk.

Note: Some of the cluster configuration may alternately be recreated from the metadata disk known as the Master Volume.

- The contents of the SAN File System metadata, which record where the client data is located on the SAN.
- The SAN File System client file data.

Chapter 10. Planning the data migration strategy

This topic describes how to plan the data migration strategy for your existing data.

If you are migrating a large amount of data, thoroughly plan the migration in advance, to minimize downtime to your organization.

The SAN File System data migration process copies file-system objects from an existing file system to the SAN File System global namespace. It uses the active policy to place files in the appropriate user storage pool.

You must perform data migration from the client machine. Data from a Windows-based application must be migrated from a Windows-based client. Likewise, data from an UNIX application must be migrated from an UNIX client. To migrate the data, the client must be able to access the source storage subsystem, the target storage pools, and the cluster during the migration process.

Note: The SAN File System must be installed and operational on the client machines and the metadata server engines before migrating data to the global namespace. In addition, filesets, storage pools, and policies must be set up.

Perform these tasks to plan the data migration strategy:
- Review the data migration considerations
- Determine the source data that you want to migrate to the SAN File System.
- Determine how you want to migrate your data
- Determine a schedule for migrating data.

You can use the Data Migration worksheet to help you plan your data migration strategy.

What data do I want to migrate?

This topic provides considerations for determining the data that should be migrated to the SAN File System.

All data that you want to access through SAN File System must be migrated to the SAN File System global namespace. SAN File System supports the following types of file-system objects:
- Regular files
- Directories
- Symbolic links (only for UNIX-based clients)
- Named pipes (FIFO objects)

All other objects in the existing file system are ignored.

If there are multiple hard links to a source file, SAN File System will attempt to preserve all the links; however, the SAN File System does not support hard links across filesets. If these hard links need to traverse across filesets, the hard link will be replaced with a symbolic link, and SAN File System will generate a warning.

If you migrate files with POSIX access control lists (ACLs), such as journaled file system (JFS), you will lose the ACLs from those files.

Data migration considerations

This topic describes considerations for the source storage type from which you are migrating data and the target storage type to which you are migrating data.

- Make sure that the storage pools to which the data will be migrated contains sufficient space for the migration:
  - You should have at a minimum twice the amount of space as the total amount of data being migrated.
  - Compressed files are expanded during the data migration process. Sufficient space must be available in the SAN File System to store the expanded files. Refer to the documentation for your operating system to determine the compression ratio and estimate the amount of space required.
  - Sparse files become dense, or full, files during data migration. Sufficient space must be available in the SAN File System to store the dense files.
- When migrating data from a Windows-based client using the **migratedata** utility, create the destination directory in the global namespace and verify that the security attributes of the destination directory match that of the source directory. If the security attributes do not match, the verification phase will fail and the migrated data will have incorrect permissions.

How do I want to migrate my data?

This topic describes the different ways you can migrate your data.

You can use standard copy commands or utilities that are provided by the client operating system to migrate your data (for example, **cp**, **cpio**, or **tar** commands on UNIX and the **xcopy** command or Explorer on Windows). You can also use backup applications to restore data from the latest backup into the SAN File System as the destination. These methods work best when migrating large numbers of small files.

For migrating large files, SAN File System provides a data migration utility, called **migratedata**, to help migrate your data quickly and efficiently, while preserving the file attributes (such as owner, group, and creation time) of your files. This utility is optimized for bulk data movement of a small number of large files. It includes these features:

- A plan phase to estimate, in advance, the amount of time that the migration operation should take
- A copy phase, when the actual data is copied
- A verify phase, which verifies that the data was successfully migrated
- A transaction-based logging and checkpoint process that allows the migration to be restarted

Migrating data using the **migratedata** utility is a *disruptive process*. This means that, to guarantee data integrity, you must stop all applications and users from modifying the data being migrated (including database and application servers) until the migration is complete. Only the data being migrated must remain unchanged. To minimize the impact of a migration, a service technician can migrate your data in chunks rather than all at one time. If your environment cannot handle a disruption in service, the **migratedata** utility might not be the best tool for your migrating your data.

If you are migrating an IBM DB2 environment, the procedures vary depending on whether your environment is file-system based or contains raw configuration devices. For a file-system-based environment, a service technician can use the **migratedata** utility to migrate your files, and then reconfigure DB2 to reflect the data movement. For raw configuration devices, the service technician must use the DB2 unload command to move data out of the raw devices to a temporary holding location, and then perform a load operation to place the data in its location in the global namespace. After the data is loaded into the global namespace, it is file based.

If you are migrating a Microsoft Exchange database from NTFS to SAN File System, use Exchange-supplied tools rather than the **migratedata** utility to ensure that the user configuration data and parameters are also migrated.

Note:

- The **migratedata** utility does not prevent an application or user from modifying (for example, editing, moving, or deleting) the data being migrated. Make sure that no one modifies this data.
- The **migratedata** utility writes different parts of a single file in parallel; therefore, there is no guarantee that the file data will be densely allocated. Operating system utilities that copy data sequentially (such as **cp** and **xcopy**) do result in densely allocated file data.
- On Windows-based clients, migrating from the root of any volume to SAN File System using the **migratedata** utility can set the destination directory to system and hidden. This is as designed because any root volume on NTFS is set to system and hidden.
- During migration, if there are no other applications running on an AIX-based client, shut down the operating system daemon *syncd* before you start the migration process.

What is my data migration schedule?

This topic describes considerations for determining your data migration schedule.

The amount of time needed to migrate data from the source storage subsystem to the SAN File System will depend on several factors, such as the amount of data to migrate, the amount of disk space on the SAN File System client, and data transfer rate between the source storage system and the SAN File System client.

You should take into consideration all of these factors when determining your data migration schedule. In addition, you can run the **migratedata** command with the plan parameter to have the SAN File System provide an estimate for you. You can use this estimate to validate that your schedule is valid.

Note: As you develop a data migration schedule, keep in mind that all applications that modify the data being migrated (including database and application servers) must be stopped during the migration process to guarantee data integrity.

Chapter 11. Gathering the hardware and software prerequisites

This topic helps you gather the hardware and software prerequisites that you need to obtains for SAN File System.

Administrative console

This topic provides an overview of the SAN File System administrative console.

The *administrative console* is a machine that you use to administer and manage SAN File System, using either the SAN File System console or the administrative command-line interface. To access the SAN File System console, this machine requires hypertext transfer protocol (HTTPS) access to the engines that hosts master metadata server. To access the administrative command-line interface, this machine requires secure shell (SSH) or telnet access to the engines that host the metadata servers.

Tip: You can use the master console as an administrative console.

Supported browsers

This topic discusses the Web browsers which SAN File System supports.

Web browser support

You access the SAN File System console and the online documentation through a standard Web browser. SAN File System supports the following Web browsers (others may work, but have not been tested):

- Microsoft Internet Explorer 6.02 with Service Pack 1 or higher
- Netscape 6.2 or higher

  Note: Although you can use Netscape 6.2, Netscape 7.0 or higher is preferred.
- Mozilla 1.1 or higher

Limitations

The **Back**, **Forward**, **Refresh** or **Reload** functions of either browser are not supported and may cause unexpected rendering problems. Additionally, opening a hyperlink into a separate browser window is not supported.

Client prerequisites

This topic defines the hardware and software prerequisites that you need to obtain for the SAN File System clients.

Host bus adapter considerations

This topic describes considerations for the existing host bus adapters (HBAs) considerations.

Verify that the combination of HBAs in your client machines are supported by SAN File System. Refer to the compatibility matrix on this Web site:

www.ibm.com/storage/support/sanfs/

Supported device drivers

This topic describes the device drivers that are supported by the storage subsystems in the SAN infrastructure.

For the clients, SAN File System requires either a single-pathing or a multi-pathing device driver to communicate with the storage subsystems.

Single-pathing device drivers

A *single-pathing device driver* allows for basic communication between the client machine and a storage subsystem. You can use any standard single-pathing device drivers that are supported by your storage subsystems. Optionally, you can use a single-pathing device driver on any SAN File System client machine.

Multi-pathing device drivers

A *multi-pathing device driver* allows multiple Fibre Channel paths to be connected to the storage subsystem and to be managed for functions such as redundant-path failover and load balancing. The multi-pathing device drivers you need on the clients depends on the types of storage subsystems you have in the SAN environment:

IBM Enterprise Storage System (ESS) 800 and F20
Subsystem Device Driver (SDD) version 1.5.1

IBM TotalStorage SAN Volume Controller 2145
Subsystem Device Driver (SDD) version 1.5.1

IBM TotalStorage DS4300 Turbo (FAStT600T), DS4400 (FAStT700), and DS4500 (FAStT900) running firmware version 8.40
RDAC for FAStT firmware version 9.0

Note:

- All metadata servers that are attached to a FAStT in the same storage pool must run the RDAC multi-pathing device driver for coordinated controller failover.
- RDAC is supported for all SAN File System client platforms except Red Hat Enterprise Linux Advanced Server 3.0. Consider using a single-path device driver on the client that runs on the Red Hat Linux platform. For Solaris 9 (64-bit), RDAC is available through a Request for Price Quote (RPQ) and limits the LUN partitions to 32.
- Using RDAC firmware lower than version 8.41 has these limitations:
  - You are limited to 32 LUNs per partition.
  - You cannot run SDD and RDAC simultaneously on Windows platforms.

Refer to the following Web site for current information about multi-pathing device drivers, including release and firmware levels:

www.ibm.com/storage/support/sanfs/

Supported client platforms

This topic describes the platforms that are supported by the SAN File System clients.

SAN File System supports client machines that run either Windows or UNIX-based operating systems.

UNIX-based clients

SAN File System supports connectivity with RISC-based machines that run these operating systems:

- AIX 5.1 (32-bit only) uniprocessor or multiprocessor with maintenance level 3, including bos.perf.tools

  Note: The bos.mp (multiprocessor) or bos.up (uniprocessor) packages must be at level 5.1.0.58 or higher. High availability cluster multi-processing (HACMP) environments are supported at the specified maintenance level.
- AIX 5.2 (32-bit and 64-bit)

  Note: The bos.mp (multiprocessor) package must be at level 5.2.0.18 or higher. The bos.up (uniprocessor) package must be at level 5.2.0.18 or higher. High availability cluster multi-processing (HACMP) environments are supported at the specified maintenance level.
- AIX 5.3 (32-bit and 64-bit)

  Note: The bos.mp (multiprocessor) package must be at level 5.3.0.0 or higher. The bos.up (uniprocessor) package must be at level 5.3.0.0 or higher. High availability cluster multi-processing (HACMP) environments are supported at the specified maintenance level.
- Red Hat Enterprise Linux Advanced Server 3.0, 2.4.21-9.ELhugemem or SMP config for i686
- Sun Solaris 9 (64-bit)

Tip: SAN File System supports up to 8 processors on AIX clients.

Windows-based clients

SAN File System supports connectivity with Intel-based machines that run these operating system:

- Microsoft Windows 2000 Server, with Service Pack 4 or higher
- Microsoft Windows 2000 Advanced Server, with Service Pack 4 or higher
- Microsoft Windows 2003 Server

Master console prerequisites

This topic defines the hardware and software prerequisites that you need to obtain for the master console.

Hardware prerequisites

SAN File System supports a single master console that can be shared with other IBM TotalStorage products, such as SAN Volume Controller. If you do not already have a master console, you must obtain a rack-mounted, high-performance, highly-reliable Intel™ server (such as the IBM eServer™ xSeries 305 or equivalent) with the following options:

- One Pentium® 4 processor, minimum 2.6 GHz
- Minimum of 1 GB of system memory
- Two IDE hard disk drives, minimum 40 GB each
- CD-ROM and diskette drives
- Two GB Ethernet ports (fiber or copper)
- Two 2 GB Fibre Channel host bus adapter (HBA) ports, such as QLogic 2342 or QLogic 2340 FC-2 cards or equivalent. The HBA must be compatible with the Linux operating system and the storage subsystems in your SAN environment.
- Keyboard, such as the Space Saver NLS keyboard or equivalent
- Monitor, such as Netbay 1U Flat Panel Monitor Console kit without keyboard or equivalent.
- Keyboard, video, monitor (KVM) switch

Note: The master console must be in the same proximity as the SAN File System hardware engines. If such a location is not possible, you need an additional keyboard and monitor for SAN File System to access the master console.

Consult your IBM sales representative or product specialist for direction in hardware configuration and ordering.

Example hardware configuration

- IBM xSeries 305 server (1U)
- Intel Pentium 4 2.8 GHz processor
- 1.2 GB memory DIMMs (254 MB comes with base unit)
- Two 40 GB IDE hard disk drives (one comes with base unit)
- Two 10/100/1000 Copper Ethernet ports on planar
- Two 1-port 1/2 GB/s Fibre Channel host bus adapters
- NetBay 1U Flat Panel Monitor Console Kit with US keyboard and KVM switch

Software prerequisites

The master console requires that you obtain the following software:

- Antivirus software
- Microsoft Windows Internet Explorer version 6.0
- Microsoft Windows 2000 Server Edition with Service Pack 3 or higher

  You can obtain the Microsoft Windows operating system by going to the following Web site and then clicking **How to Buy**:

  www.microsoft.com/windows2000/server/
- Microsoft Windows 2003 Standard Server Edition
- J2SE Java™ Runtime Environment (JRE) 1.4.2

  You can obtain JRE 1.4.2 by going to the following Web site and then clicking **Downloads**, **Java & Technologies**, **Java 2 Platform, Standard Edition 1.4**, and then **Download J2SE JRE**:

  www.sun.com/

Metadata server prerequisites

This topic defines the hardware and software prerequisites that you need to obtain for the metadata server.

Hardware prerequisites

SAN File System supports from two to eight metadata servers running on hardware known as storage engines. For each engine, you must obtain a rack-mounted, high-performance, and highly-reliable Intel server (such as IBM eServer xSeries 345 or equivalent) with the following options:

- Two processors, minimum 3 GHz each
- Minimum of 4 GB of system memory
- Two internal hard disk drives with mirroring, minimum 36 GB each
- Two power supplies (optional but recommended)
- One Gb port for Ethernet connections (fiber or copper)
- Two 2 Gb Fibre Channel host bus adapter (HBA) ports (must be compatible with the Linux operating system and the storage subsystems in your SAN environment)

  Important: Verify that the HBA card is compatible with the switches in your environment.
- CD-ROM and diskette drives
- Remote Supervisory Adapter II card (must be compatible with the Linux operating system and the storage subsystems in your SAN environment)
- Remote Supervisory Adapter II external power supply (optional)

Consult your IBM sales representative or product specialist for direction in hardware configuration and ordering.

Example hardware configuration

- IBM xSeries 345 server (2U)
- Two 2-way 3.067 GHz Intel processors (one comes with the base unit)
- Four 4 GB memory DIMMs
- Two 36.4 GB 10K-rpm hot swap U320 hard disk drives with RAID 1 mirroring
- Two redundant hot swap 514 Watt power supplies (one comes with base unit)
- Two redundant hot swap fans (included in base unit)
- Two 10/100/1000 Copper Ethernet ports on planar
- Two QLogic QLA2340 or QLA2342L 1-port 1/2 Gb/s Fibre Channel host bus adapters
- Remote Supervisory Adapter II

Software prerequisites

The metadata servers requires that you obtain the following software for each metadata server:

- SUSE Linux Enterprise Server 8, with United Linux Service Pack 3

  You need a licensed copy of SUSE Linux for each engine. You can obtain SUSE Linux operating system by going to the following Web site and then clicking **Online Store**:

www.suse.com/us/business/products/server/sles/

- QLogic QLA2340 or QLA2342L version 6.06.64

  You can obtain appropriate version of the QLogic driver by going to the following Web site:

  www.qlogic.com/support/oem_detail_all.asp?oemid=22.

- Lightweight Directory Access Protocol (LDAP) server

  An LDAP server, which is available and configurable by the SAN File System administrator, is required to provide authentication and authorization. If you do not already have an LDAP server in your environment, install it on a computer in the network that is accessible to the metadata servers and master console. Do not install the LDAP server on the metadata server or master console hardware.

  You can download a free copy of the OpenLDAP server from the following Web site:

  www.openldap.org/software/download/

- Time synchronization software (optional)

  Time synchronization software, such as Network Time Protocol (NTP), ensures that the system clocks on each engine in the metadata-server cluster are synchronized. Having synchronized clocks on each engine will ensure that the log files on each engine accurately reflect the sequence of events in SAN File System.

Chapter 12. Worksheets

This topic provides an overview of each of the planning worksheets that is available.

Data migration worksheet

Use this worksheet to plan the migration of your existing data into the SAN File System global namespace.

Source directory	Destination directory	Size (MB)	Application	Client name	Migration date

Filesets worksheet

Use this worksheet to you plan the fileset configuration.

Fileset name	Parent fileset	Attach point	Metadata server (for static fileset assignments only)	Usage threshold (%)	Quota type	Quota size (MB)
ROOT	n/a	/	(master)			

Fileset name	Parent fileset	Attach point	Metadata server (for static fileset assignments only)	Usage threshold (%)	Quota type	Quota size (MB)

Hardware prerequisites worksheet

Use this worksheet to determine the hardware that you need to purchase.

Master console

Part	Part description	Part number	Quantity
Machine			
Processor			
Memory			
Hard disk drives			
CD-ROM drive			
Diskette drive			
Ethernet ports			
Host bus adapters			
Keyboard, mouse, monitor			

Metadata server

Part	Part description	Part number	Quantity
Machine			
Processor			

Part	Part description	Part number	Quantity
Memory			
Hard disk drives			
Power supplies			
Ethernet ports			
Host bus adapters			
RSA II			
RSA II external power assembly (optional)(When each of the two power supplies on the metadata servers is connected to a separate and independent power circuit, then any single point of failure is eliminated. The external RSA II power supply is not necessary. If you decide to use the external RSA II power supply, ensure that it is on a separate and independent circuit.)			

LDAP planning worksheet

Use this worksheet to plan your LDAP configuration.

LDAP configuration

Description	Recommended value	Your Value
IP address	n/a	
Subnet	n/a	
Authorized LDAP username	n/a	
Authorized LDAP password	n/a	
LDAP certificate	n/a	

Roles

Each of the four roles must have an entry in the LDAP database. All must have the parent DN, and all must have the same objectClass. Each must have an attribute containing the string that describes its role; ″Administrator,″ ″Backup,″ ″Operator,″ or ″Monitor″. Finally, each must support an attribute that can contain multiple values; one value for each role occupant's DN.

Description	Recommended value	Your Value
Role parent DN	ou-SANFS Roles... objectclass: organizationalUnit	
Attribute containing role name	cn	

Description	Recommended value	Your Value
Attribute for role occupants	roleOccupant	

Users

Each user must have an entry in the LDAP database. All users must have the same parent DN, and the same objectClass. They must contain a ″user ID″ type of attribute.

Description	Recommended value	Your Value
Attribute containing login userid	uid	

Metadata server worksheet

Use this worksheet to plan your metadata server configuration.

Setting	Metadata server 1	Metadata server 2	Metadata server 3
Engine serial number			
Host name			
Engine IP address			
Engine subnet			
RSA IP address			
RSA subnet			
HBA 1 WWN			
HBA 2 WWN			
Gateway			
DNS address (optional)			
BIOS level			
HBA model/driver			

Metadata servers — application workload method worksheet

Use this worksheet to calculate the number of metadata servers that are required in your environment, based on the number of hard disk drives that are available to SAN File System.

Application workload characteristics

Primary applications types (for example, database or mail server)	
Average file size, per application	
Size of the hotset (number of objects), per application	
Client platforms	
Number of clients, per application	
Average directory size	

Does the application perform a dirwalk?	
Will objects be shared (read/write) among multiple clients?	
Transaction rate (for example, connects per second, or orders per second), per application	

Average metadata server operations per second

Use the following table to estimate the average FOP and metadata server OP percentage:

Application type	Typical metric	Average FOP/APPOP	Average metadata server OP/FOP
Mail server	Mails per second	20	3%
OLTP database back end	Transactions per second	2	5%
Data Warehouse database back end	Tuples per second	0.1	10%
Office workgroup	Documents per second	10	1%
Web server	Connections per second	5	10%
Web proxy	Pages per second	15	10%
Peer-to-peer	Files per second	12	10%
Network File System Serving	Megabytes per second	30	10%
Common Internet File System Serving	Megabytes per second	30	10%
Compile Build (Development)	Files built per second	10	5%
User Folder Serving	Files used per second	5	3%

Calculate the average metadata server operations per second:

(Transaction rate * average FOPs) * metadata server OPs	

Adjustment factor

Average file size larger than 500 Mb + 20%	
File hotset less than 10 K	
Windows clients + 20%	
Number of clients (uniform) * number of clients	
Average directory size larger than 100 objects	
Special application dirwalk + 20%	
Object sharing read/writes + 20%	

Number of required metadata servers

highest metadata server OP / 1 500	

Metadata servers — available storage worksheet

Use this worksheet to calculate the number of metadata servers that are required in your environment, based on the number of hard disk drives that are available to SAN File System.

Tip: If the application is cache-friendly (for example, TPC-H and data warehouse), use an application factor of 0.5. If the application is not cache-friendly and is create and delete intensive, use a application factor of 2. Otherwise, use a application factor of 1.

Application factor	
(Number of disk drives / 50) * application factor	

Policies worksheet

Use this worksheet to plan your policy configuration.

Policy name:

____

Rule name	Storage pool	Fileset names	Condition

Rule name	Storage pool	Fileset names	Condition

Sample policy sets

This topic provides sample policy sets.

Distribute files based on fileset

```
VERSION 1
RULE 'rule1' SET STGPOOL 'pool1' FOR FILESET('fileset1','fileset2')
RULE 'rule2' SET STGPOOL 'pool2' FOR FILESET('fileset3')
```

Distribute files based on file extension

```
VERSION 1
RULE 'documents' SET STGPOOL 'pool1' WHERE
   UCASE(NAME) LIKE '%.DOC' OR
   UCASE(NAME) LIKE '%.LWP' OR
   UCASE(NAME) LIKE '%.TXT'
RULE 'executables' SET STGPOOL 'pool2' WHERE
   UCASE(NAME) LIKE '%.EXE' OR
   UCASE(NAME) LIKE '%.COM' OR
   UCASE(NAME) LIKE '%.BAT' OR
   UCASE(NAME) LIKE '%.SH' OR
   UCASE(NAME) LIKE '%PL'
```

Distribute files based on the day of the week

Note:

1. The file placement resulting from this policy set cannot be restored from backups.
2. This policy set assumes placement based on coordinated universal time (UTC).

```
VERSION 1
RULE 'documents' SET STGPOOL 'pool1' WHERE
   UCASE(NAME) LIKE '%.DOC' OR
   UCASE(NAME) LIKE '%.LWP' OR
   UCASE(NAME) LIKE '%.TXT'
RULE 'executables' SET STGPOOL 'pool2' WHERE
   UCASE(NAME) LIKE '%.EXE' OR
UCASE(NAME) LIKE '%.COM' OR
UCASE(NAME) LIKE '%.BAT' OR
UCASE(NAME) LIKE '%.SH' OR
UCASE(NAME) LIKE '%PL'
```

Software prerequisites worksheet

Use this worksheet to determine the software that you need to obtain.

Clients

	Single-path or multi-path device drivers

Master console

	Antivirus software
	Microsoft Windows 2000 Server Edition with Service Pack 3 or higher
	Microsoft Windows Internet Explorer version 6.0

Metadata server (for each)

	LDAP server
	Host bus adapter (QLogic driver QLA2340/QLA2342L)
	SUSE Linux Enterprise Server 8, with United Linux Service Pack 3 or higher

Storage access worksheet

Use this worksheet to verify that the metadata servers and clients have access to all volumes that they use to store data.

System storage pools

Use this table to record the volumes associated with the system storage pool, which is accessible from all of the metadata servers.

Zone name	Volumes

User storage pools

Use this table to record clients that are affected by each rule, and the filesets, storage pools, volume, and zone associated with that rule to verify that the clients have access to all volumes that they use to store data.

Client	Rule name	Fileset	Storage pools	Volumes	Zone name

Storage pools worksheet

Use this worksheet to plan your storage pool configuration.

Storage pool name	Logical partition size (16, 64 or 256 MB)	Allocation size (auto, 4 or 128 KB)	Usage threshold (%)	Enable alerts?	Volume names
SYSTEM					

System storage pool worksheet

Use this worksheet to calculate the amount of storage space required for the system storage pool.

Fileset measurements

	Fileset 1	Fileset 2	Fileset 3	Total
Total data size (byte)				
Number of files				
Number of directories				
Number of symbolic links				
Number of hard links				
Number of FIFO objects				
Average object-name length				

File-metadata storage space

Total data size * 415 / 1 000	bytes
Total number of files * (560 + (3 * name length))	bytes
Total number of directories * (560 + (3 * name length))	bytes
Total number of symbolic links * (560 + (3 * name length))	bytes
Total number of hard links * (90 + (3 * name length))	bytes
Total number of FIFO objects * (560 + (3 * name length))	bytes
Total / 1 000 000	MB

System-metadata storage space

Number of metadata servers * 128 MB	MB

Total storage space required

File-metadata storage + system metadata storage	MB

UNIX-based-client installation worksheet

Use this worksheet to collect information necessary to install a UNIX-based client.

Setting	Client 1	Client 2	Client 3
Client name			
Client IP address			

Setting	Client 1	Client 2	Client 3
Subnet			
Metadata server connection host			
Metadata server port			
Transport protocol			
Device-candidates list directory			
Mount file system read-only			
Display verbose messages			

Legend

Client name
The host name of the client machine.

Client IP address
The IP address of the client machine.

STFS kernel module
The client loads the file-system driver as a kernel extension. Specify the path to the location of the client kernel module file. The default is: /usr/tank/client/bin/stfs.o (This is /base/client/bin/stfs.o , where base is the base directory.)

Metadata server connection host
The fully-qualified host name or IP address of one of the Metadata servers in the cluster, in dotted decimal format (for example, 9.47.101.01).

Metadata server port number
The UDP port number of the Metadata server connection host, in dotted decimal format (for example, 10190).

Transport protocol
The transport protocol that you want the client to use to connect to the Metadata server. Specify either TCP/IP or UDP.

Device-candidates list directory
The client determines which disks to use as volumes by searching the SAN for a list of available disks, called device candidates. The device-candidate list consists of those devices that have device special files. (Device special files are UNIX files that reference hardware. The device-candidate list is not viewable to the user.) Specify the directory that contains the device special files (for example, /dev/stfsdisk/)

Mount point
The mount point (directory) from which the file system appears on the client. The default mount point is /mnt/tank.

Mount file system read-only
Specify whether you want to view, but not modify, data and metadata in the file system. The default is no.

Display verbose messages
Specify whether you want to display information messages from the commands. The default is no.

Windows-based-client installation worksheet

Use this worksheet to collect information necessary to install a Windows-based client.

Client name	Metadata server IP address	Metadata server port	Drive letter

Legend

Metadata server IP address
The IP address of one of the metadata servers in the cluster, in dotted decimal format (for example, 9.47.101.01).

Metadata server port number
The port number of the metadata server, in dotted decimal format. The default port is 1700.

Client name
The name that you want to use for the client (for example, st.ibm.com)

Drive letter
The drive letter you want to use for SAN File System storage.

Note: Windows clients should have Service Pack 4 installed.

Volumes worksheet

Use this worksheet to plan your volume configuration.

LUN ID	Volume name	Size (MB)	Storage device

Appendix A. File placement policy syntax

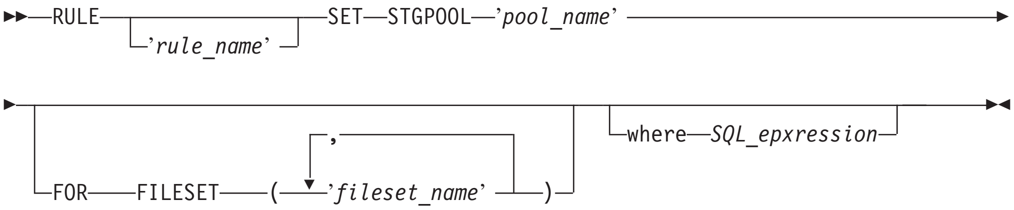
This topic describes the syntax conventions for file-placement rules.

You can create a file containing policy rules for placing newly created files. You can then use this rule file when creating a policy using the **mkpolicy** command from the administrative CLI. You can also edit the policy rules that you create using the SAN File System console.

Important:

1. Every policy file must start with VERSION 1.
2. A policy is not required to contain any rules, in which case it would be equivalent to the default policy.
3. The maximum size of a policy is 32 KB.

You can also add comments to the policy. All comments must start with /* and end with */ (for example, /* comment */).



Parameters

RULE
Initiates the rule statement.

'rule_name'
Identifies the rule. This parameter is optional.

SETSTGPOOL *'pool_name'*
Identifies the pool in which you want to place all files that match the rule criteria (fileset and SQL expression).

FOR FILESET (*'fileset_name'*)
Identifies one or more filesets in which the file is created to determine where the file is to be placed. In the case of nested filesets, the rules apply if the file is created in the innermost fileset.

where *SQL_expression*
Compares the file attributes specified in the rule with the attributes of the file being created to determine where the file is to be placed. The *SQL_expression* can be any combination of standard SQL-syntax expressions, including comparison predicates, between predicates, in predicates, like predicates, mathematical value expressions, and boolean, string and numeric literals.

Restriction: Case expressions and compared-when clauses are not allowed.

With SAN File System, you can use built-in functions that can be used in comparison predicates, between predicates, in predicates, and like predicates.

These functions are organized in three categories: date and time manipulation, numeric calculations, and string manipulation.

Attributes

You can use any of these attributes in the expression:

NAME
Name of the file. You can use a percent (%) wildcard in the name to represent zero or more characters and use the underscore (_) wildcard to represent one single-byte or multibyte character.

CREATION_DATE
Date and time that the file was created.

GROUP_ID
Numeric group ID. This attribute is valid only for UNIX clients.

USER_ID
Numeric user ID. This attribute is valid only for UNIX clients.

String functions

You can use these string-manipulation functions on file names and literals.

Important: You must enclose strings in single-quotation marks. You can include a single-quotation mark in a string by using two single-quotation marks (for example, 'a''b' represents the string a'b).

CHAR(*x*)
Converts an integer *x* to a string.

CHARACTER_LENGTH(*x*)
Determines the number of characters in string *x*. Both single-byte and multibyte characters count as one character in a string.

CHAR_LENGTH(*x*)
Determines the number of characters in string *x*. Both single-byte and multibyte characters count as one character in a string.

CONAT(*x*,*y*)
Concatenates strings *x* and *y*.

HEX(*x*)
Converts an integer *x* in hexadecimal format.

LCASE(*x*)
Converts string *x* to lowercase.

LEFT(*x*,*y*,*z*)
Left justifies string *x* in a field of *y* characters, optionally padding with character *z*.

LENGTH(*x*)
Determines the length of the data type of string *x*.

LOWER(*x*)
Converts string *x* to lowercase.

LTRIM(*x*)
Removes leading blank characters from string *x*.

POSITION(*x* IN *y*)
Determines the position of string *x* in string *y*.

POSSTR(*x*,*y*)
Determines the position of string *y* in string *x*.

RIGHT(*x*,*y*,*z*)
Right justifies string *x* in a field of *y* characters, optionally padding with character *z*.

RTRIM(*x*)
Removes the trailing blank characters from string *x*.

SUBSTR(*x* FROM *y* FOR *z*)
Extracts a portion of string *x*, starting at position *y*, optionally for *z* characters (otherwise to the end of the string).

SUBSTRING(*x* FROM *y* FOR *z*)
Extracts a portion of string *x*, starting at position *y*, optionally for *z* characters (otherwise to the end of the string).

TRIM(*x*)
Trims blank characters from the beginning and end of string *x*.

TRIM(*x* FROM *y*)
Trims blank characters that are *x* (LEADING, TRAILING, or BOTH) from string *z*.

TRIM(*x* *y* FROM *z*)
Trims character *y* that is *x* (LEADING, TRAILING, or BOTH) from string *z*.

UCASE(*x*)
Converts the string *x* to uppercase.

UPPER(*x*)
Converts the string *x* to uppercase.

Numerical functions

You can use these numeric-calculation functions to place files based on either numeric parts of the file name, numeric parts of the current date, and UNIX-client user IDs or group IDs. These can be used in combination with comparison predicates and mathematical infix operators (such as addition, subtraction, multiplication, division, modulo division, and exponentiation).

INT(*x*)
Converts number *x* to a whole number, rounding up fractions of .5 or greater.

INTEGER(*x*)
Converts number *x* to a whole number, rounding up fractions of .5 or greater.

MOD(*x*,*y*)
Determines *x* % *y*.

Date and time functions

You can use these date-manipulation and time-manipulation functions to place files based on when the files are created and the local time of the metadata server serving the directory within which the file is being created.

Important: Universal Time is used for all date and time functions.

CURRENT DATE
Determines the current date on the metadata server.

CURRENT_DATE
Determines the current date on the metadata server.

CURRENT TIME
Determines the current time on the metadata server.

CURRENT_TIME
Determines the current time on the metadata server.

CURRENT TIMESTAMP
Determines the current date and time on the metadata server.

CURRENT_TIMESTAMP
Determines the current date and time on the metadata server.

DATE(*x*)
Creates a date out of *x*.

DAY(*x*)
Creates a day of the month out of *x*.

DAYOFWEEK(*x*)
Creates the day of the week out of date *x*, where *x* is a number from 1 to 7 (Sunday=1).

DAYOFYEAR(*x*)
Creates the day of the year out of date *x*, where *x* is a number from 1 to 366.

DAYS(*x*)
Determines the number of days since 0000-00-00.

DAYSINMONTH(*x*)
Determines the number of days in the month from date *x*.

DAYSINYEAR(*x*)
Determines the day of the year from date *x*.

HOUR(*x*)
Determines the hour of the day (a value from 0 to 23) of time or timestamp *x*.

MINUTE(*x*)
Determines the minute from date *x*.

MONTH(*x*)
Determines the month of the year from date *x*.

QUARTER(*x*)
Determines the quarter of year from date *x*, where *x* is a number from 1 to 4 (for example, January, February, and March is quarter 1).

SECOND(*x*)
Returns the seconds portion of time *x*.

TIME(*x*)
Displays *x* in a time format.

TIMESTAMP(*x*,*y*)
Creates a timestamp (date and time) from a date *x* and optionally a time *y*.

WEEK(*x*)
Determines the week of the year from date *x*.

YEAR(*x*)
Determines the year from date *x*.

Time and dates formats

Use any of these formats when specifying times and dates.

Note: All date and time attributes in these rules are based in coordinated universal time (UTC).

Timestamp
Use one of the following formats to specify a timestamp:

- *date time*
- *date*

There must be exactly one space between the date and time.

You can mix formats for the date and time. For example, you can specify ISO format for the date and international format for the time.

Date Use one of these formats to specify a date:

European
DD.MM.YYYY

ISO *YYYY–MM–DD*

USA *MM/DD/YYYY*

You can leave off leading zeros from *MM* (month) and *DD* (day). You can use a two-digit year, in which case 1900 is added if the year is greater than 50 and 2000 is added if the year is 50 or less.

Important: The MONTHNAME() and DAYNAME() functions produce English names with no internationalization.

Time Use one of these formats to specify a time:

International
HH:*MM*[*SS*[.*UUUUUU*]]

USA *HH*[:*MM*[:*SS*]] [A|P|AM|PM]

You can leave off leading zeros from any field except subseconds. The international format uses a 24–hour clock. The USA format uses a 12–hour clock followed by A, P, AM, or PM.

You can substitute commas or periods for colon delimiters in the international format.

Examples

The following example shows a sample file:

```
VERSION 1

rule 'stgRule1' set stgpool 'pool1' for fileset ('cnt_A')
rule 'stgRule2' set stgpool 'pool2' where NAME like '%.doc'
rule 'stgRule3' set stgpool 'pool3' where DAYOFWEEK(CREATION_TIME) == 1
rule 'stgRule4' set stgpool 'pool4' where USER_ID <= 100
```

Appendix B. Accessibility

This topic provides information about the accessibility features of SAN File System and its accompanying documentation.

Accessibility features help a user who has a physical disability, such as restricted mobility or limited vision, to use software products successfully.

Features

These are the major accessibility features in SAN File System:

- You can use screen-reader software and a digital speech synthesizer to hear what is displayed on the screen.

  Note: The SAN File System Information Center and its related publications are accessibility-enabled for the IBM Home Page Reader.
- You can operate all features using the keyboard instead of the mouse.

Navigating by keyboard

You can use keys or key combinations to perform operations and initiate many menu actions that can also be done with a mouse. You can navigate the SAN File System console and help system from the keyboard by using the following key combinations:

- To traverse to the next link, button or topic, press Tab inside a frame (page).
- To expand or collapse a tree node, press Right Arrow or Left Arrow, respectively.
- To move to the next topic node, press Down Arrow or Tab.
- To move to the previous topic node, press Up Arrow or Shift+Tab.
- To scroll all the way up or down, press Home or End, respectively.
- To go back, press Alt+Left Arrow
- To go forward, press Alt+Right Arrow.
- To go to the next frame, press Ctrl+Tab. There are quite a number of frames in the help system.
- To move to the previous frame, press Shift+Ctrl+Tab.
- To print the current page or active frame, press Ctrl+P.

Appendix C. Getting help, service, and information

If you need help, service, technical assistance, or just want more information about IBM products, you can find a wide variety of sources available from IBM to assist you.

Services available and telephone numbers listed are subject to change without notice.

Software Maintenance Agreement

All distributed software licenses include Software Maintenance Agreement(software subscription and technical support) for a period of 12 months from the date of acquisition providing a streamlined way to acquire IBM software and assure technical support coverage for all licenses. You can elect to extend coverage for a total of three years from date of acquisition. While your Software Maintenance is in effect, IBM provides you assistance for your 1) routine, short duration installation and usage (how-to) questions; and 2) code-related questions. IBM provides assistance by telephone and, if available, electronic access, only to your information systems (IS) technical support personnel during the normal business hours (published prime shift hours) of your IBM Support Center. (This assistance is not available to your end users.) IBM provides Severity 1 assistance 24 hours a day, every day of the year.

Before you call for service

This topic provides information you need to know before you call for service.

Some problems can be solved without outside assistance. You can use the online help by looking in the online or printed documentation that comes with the SAN File System, or by consulting the IBM Support Home Web site. Also, be sure to read the information in any README files and release notes that come with the SAN File System.

Getting help online

IBM maintains pages on the World Wide Web where you can get information about IBM products and services and find the latest technical information.

Table 2 lists some of these pages.

Table 2. IBM Web sites for help, services, and information

www.ibm.com/	Main IBM home page
www.ibm.com/storage/	IBM Storage home page
www.ibm.com/storage/support	IBM Support home page

Getting help by telephone

With the original purchase of the SAN File System, you have access to extensive support coverage. During the product warranty period, you can call the IBM Support Center (1 800 426-7378 in the U.S.) for product assistance covered under the terms of the software maintenance contract that comes with SAN File System purchase.

Have the following information ready when you call:

- SAN File System software identifier, which can be either the product name (SAN File System) or the Product Identification (PID) number
- Description of the problem
- Exact wording of any error messages
- Hardware and software configuration information

If possible, have access to your master console when you call.

In the U.S. and Canada, these services are available 24 hours a day, 7 days a week. In the U.K., these services are available Monday through Friday, from 9:00 a.m. to 6:00 p.m. In all other countries, contact your IBM reseller or IBM marketing representative.[1]

[1] Response time varies depending on the number and complexity of incoming calls.

Appendix D. Purchasing additional services

During and after the warranty period, you can purchase additional services, such as support for other IBM and non-IBM hardware, operating systems, and application programs; network setup and configuration; extended hardware repair services; and custom installations. Service availability and name might vary by country.

Appendix E. Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

IBM World Trade Asia Corporation
Licensing
2-31 Roppongi 3-chome, Minato-ku
Tokyo 106-0032, Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION ″AS IS″ WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Corporation
MW9A/050
5600 Cottle Road
San Jose, CA 95193
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks

The following terms are trademarks of International Business Machines Corporation or Tivoli Systems Inc. in the United States or other countries or both:

AIX
Enterprise Storage Server
HACMP

AIX 5L
eServer
IBM

DB2
FlashCopy
IBM logo

Storage Tank
Tivoli
TotalStorage
WebSphere
xSeries

Intel and Pentium are trademarks of Intel Corporation in the United States, other countries, or both.

Java and all Java-based trademarks are registered trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, and Windows NT are trademarks or registered trademarks of Microsoft Corporation.

Red Hat and all Red Hat-based trademarks and logos are trademarks or registered trademarks of Red Hat, Inc., in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product, and service names may be trademarks or service marks of others.

Index

Readers' Comments — We'd Like to Hear from You

IBM TotalStorage SAN File System
(based on IBM Storage Tank™ technology)
Planning Guide
Version 2 Release 2

Publication No. GA27-4344-01

Overall, how satisfied are you with the information in this book?

	Very Satisfied	Satisfied	Neutral	Dissatisfied	Very Dissatisfied
Overall satisfaction	□	□	□	□	□

How satisfied are you that the information in this book is:

	Very Satisfied	Satisfied	Neutral	Dissatisfied	Very Dissatisfied
Accurate	□	□	□	□	□
Complete	□	□	□	□	□
Easy to find	□	□	□	□	□
Easy to understand	□	□	□	□	□
Well organized	□	□	□	□	□
Applicable to your tasks	□	□	□	□	□

Please tell us how we can improve this book:

Thank you for your responses. May we contact you? □ Yes □ No

When you send comments to IBM, you grant IBM a nonexclusive right to use or distribute your comments in any way it believes appropriate without incurring any obligation to you.

Name	Address
Company or Organization	
Phone No.	

Readers' Comments — We'd Like to Hear from You

GA27-4344-01

Cut or Fold
Along Line

Fold and Tape **Please do not staple** Fold and Tape

NO POSTAGE
NECESSARY
IF MAILED IN THE
UNITED STATES

BUSINESS REPLY MAIL

FIRST-CLASS MAIL PERMIT NO. 40 ARMONK, NEW YORK

POSTAGE WILL BE PAID BY ADDRESSEE

IBM Corp
Dept. CGFA
PO Box 12195
Research Triangle Park, NC 27709-9990

Fold and Tape **Please do not staple** Fold and Tape

Cut or Fold
Along Line

Printed in USA

GA27-4344-01