IBM Tivoli Storage Manager

Integrating IBM Tivoli Storage Manager with IBM Elastic Storage

Document version 1.0

Dominic Müller-Wicke
Tivoli Storage Manager Development

IBM

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 Introduction

The integration of IBM Tivoli Storage Manager with IBM Elastic Storage addresses these major features: scalable backup processing, hierarchical storage management, and fast disaster recovery. When these products are integrated, the solution addresses enterprise requirements in data protection and data management for large GPFS file systems. This article describes the integration of the products on a high level, and provides an overview of the functions.

2 Scalable backup processing

File system backup processing in a client/server architecture can be divided in three different tasks. The first task is to capture the file system state. This means that the file system's directory tree and objects are scanned. The second task is to compare the scanned data to the existing backup data on the backup server. As a result of the comparison, objects are classified into the following groups and processed, which is the final task of the backup:

- Objects that have to be sent to the backup server for backup

- Objects that have to be deleted on the backup server

- Objects that have no change

Tivoli Storage Manager progressive incremental backup processing covers the tasks described above and can be used for GPFS file system backup. Because Tivoli Storage Manager progressive incremental backup processing uses standard interfaces to scan the file system in a single process, the parallelism of the GPFS cluster can't be used efficiently. Thus there is a scale limitation for this type of backup on the number of objects stored in the file system.

If the time it takes to process the daily backup exceeds the allowed backup duration for the user, use the GPFS command **mmbackup** to perform the file system backup.

The major advantage to using the **mmbackup** command, compared to Tivoli Storage Manager progressive incremental backup processing, is the ability to use the GPFS policy engine for fast file system scans. The command also ensures that all backup activities are run in parallel by using the GPFS cluster nodes to send backup data in parallel to the Tivoli Storage Manager server.

One feature of the integration of the products for scalable backup processing is the ability to reuse existing Tivoli Storage Manager configuration and backup data. You can switch seamlessly from Tivoli Storage Manager progressive incremental backup to backing up data by using the **mmbackup** command. Furthermore, many backup configuration parameters that are used in Tivoli Storage Manager enterprise environments are supported by the **mmbackup** command. In this way, the system can be configured to perform incremental backup processing with high performance and at scale.

The Tivoli Storage Manager progressive incremental scan method examines each object (directory, file, link, or special object) in the file system and compares its state to the

associated data in the Tivoli Storage Manager database. Differences found in the object's state triggers the appropriate backup activity. If a directory object is found to be different in the database, entries that are not found in the file system but listed in the backup database are now known to have been deleted. This scenario triggers the target object of the directory entry to be marked inactive and later to be expired from the Tivoli Storage Manager server. A new directory entry represents a new object because the last backup and the target of such an entry is considered for backup eligibility. Tivoli Storage Manager's progressive incremental backup processing adheres to a set of patterns known as the *include/exclude criteria*. These patterns are configurable and are used to limit backup processing to only those objects that the administrator has classified as appropriate for backup.

The **mmbackup** command functions in phases that are independently parallelized. The file system scan and backup processing are in separate phases. The following **mmbackup** command process is illustrated in figure 1:

- When the **mmbackup** command is used for the first time on a Tivoli Storage Manager protected file system, it queries the Tivoli Storage Manager server for backup data of previous backup procedures performed by the Tivoli Storage Manager backup archive client. Based on this information, it creates a *shadow* database. The shadow database is used to compute later backups performed by the **mmbackup** command and is updated accordingly. While the backup is in progress, the shadow database is updated by the **mmbackup** processing frequently to ensure the consistency and accuracy of the stored information over the whole backup process.

- To generate a list of file system changes, **mmbackup** processing uses the GPFS policy engine to scan the file system. The scan is guided by the include/exclude criteria in the Tivoli Storage Manager configuration. The scan result is next compared to the content stored in the shadow database. The comparison results in three actions:

  - File system objects that were removed from the file system since the last backup are expired.

  - File system objects are updated if metadata is changed

  - File system objects whose data content or ACL / EA information have changed are backed up

- Lists of the file system objects are created in a shared working directory. The file lists are processed by the GPFS policy system on select nodes in the GPFS cluster that have the Tivoli Storage Manager backup-archive client installed. On these nodes, the Tivoli Storage Manager backup-archive client performs either a file list-based selective backup (send file data), an incremental backup (update file metadata), or expires the data (delete files from Tivoli Storage Manager Server) for the named file system objects. The policy engine distributes the work by subdividing each large list into several smaller lists to be operated upon in a thread on a selected node. The number of threads, the nodes selected, and the size of each sub list are all controllable by using the **mmbackup** command options.

- After all the backup processing has finished, the updated shadow database is sent to the Tivoli Storage Manager server for backup to ensure the availability of the information as baseline for the next backup in case something happens to the online copy kept in the root.
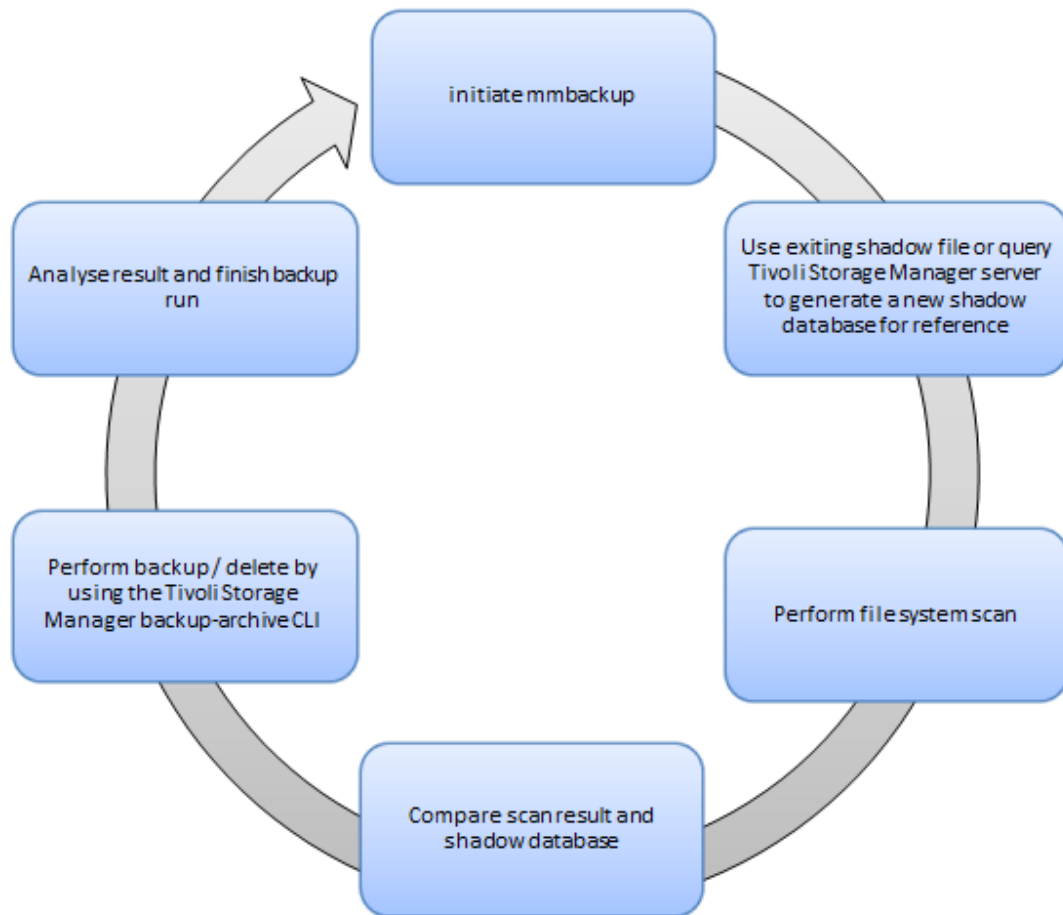
*Figure 1:   Illustration of the **mmbackup** command backup procedure*

Most of the functions of Tivoli Storage Manager progressive incremental backup processing are implemented by using the **mmbackup** command, too. . Because processing is different, some changes in the environment that are detected from Tivoli Storage Manager progressive incremental backup processing cannot be detected by **mmbackup** command processing. Table 1 compares the backup methods and shows the differences.

Table 1:   Backup methods comparison

	Tivoli Storage Manager progressive incremental backup (dsmc incremental)	GPFS policy-driven backup (mmbackup)
Detects changes in files and sends a new copy of the file to the server.	Yes	Yes
Detects changes in metadata and updates the file	Yes	Yes

metadata on the server or sends a new copy of the file to the server (in terms of ACL/EA changes).		
Detects directory move, copy or rename functions, and sends a new copy of the file to the server.	Yes	Yes
Detects local file deletion and expires the file on the server.	Yes	Yes
Detects Tivoli Storage Manager file space deletion or node/policy changes, and sends a new copy of the file to the server.	Yes	No*
Detects file deletion from the Tivoli Storage Manager server and sends a new copy of the file to the server.	Yes	No*
Detects additions of new exclude rules and expires the file on the server.	Yes	Yes
Detects policy changes to new include rules and rebinds the file to the new storage pool.	Yes	No**
Detects copy mode and copy frequency configuration changes.	Yes	No*
Detects migration state changes (Tivoli Storage Manager for Space Management) and updates the server object.	Yes	Yes
Detects that a file wasn't processed successfully during a backup operation and attempts again at the next backup.	Yes	Yes
* The **mmbackup** command queries the Tivoli Storage Manager server only once at the time of the first backup. Changes that are performed on the Tivoli Storage Manager server by using the Tivoli Storage Manager administrative client cannot be detected by **mmbackup** processing. You must rebuild the mmbackup shadow database if the Tivoli Storage Manager server file space changes.		

> ** Tivoli Storage Manager includes rules with associated management class bindings that cannot be detected by **mmbackup** processing. Therefore mmbackup processing does not rebind a file if a management class changes include rules.

.

3   Hierarchical storage management

A hierarchical storage management (HSM) solution is typically a virtual expansion of the file system space because it introduces additional hierarchical storage tiers. The reason for introducing storage tiers is to store large amounts of data at lower cost and higher efficiency than traditional rotating magnetic disks. In many cases, HSM-tiered environments are composed of different SSD classes, HDD classes, and tape as the last tier in the hierarchy. The process of moving files between tiers is called *migration and recall*. The term *migration* refer to moving the file data to the next lower tier and replace the data on the higher tier with only the metadata that is required to identify the new location of the file data in the lower tier for later recall. The metadata is typically called a *stub file* and consumes nearly no space. The term *recall* refers to moving the data back to the higher tier in the storage hierarchy. The physical location of the file data is transparent. An HSM solution does not distinguish between different file versions. A change to file data requires the recall from the lower tiers to a higher tier.

Tivoli Storage Manager for Space Management (also known as UNIX HSM) is the solution provided by Tivoli Storage Manager for the GPFS file system Tivoli Storage Manager for Space Management can be installed on multiple compute nodes in the GPFS cluster. All nodes that have UNIX HSM installed participate in file system management. HSM activities like migration and recall of files are distributed to all cluster nodes to use the parallelism inherent in the GPFS cluster. Furthermore, the processing is parallelized within each cluster node as well.

Elastic Storage supports the industry standard mechanism for coupling HSM functions to the file system's internal data. This standard is called the Data Management application programming interface (DMAPI). DMAPI is provided by GPFS and Tivoli Storage Manager for Space Management links to the GPFS DMAPI library. This integration point permits the UNIX HSM to be a qualified data management application to GPFS.

In terms of migrating files from a higher tier to a lower tier, UNIX HSM has the following file migration states:

- **Resident**: The file data exists only on the live file system (tier 0). No valid copy of the file resides in the Tivoli Storage Manager server (tier 1). This is the migration state for new or changed files.

- **Premigrated**: The file data has a valid copy in both the file system (tier 0) and also on the Tivoli Storage Manager server (tier 1). If a file was opened for read and then recalled from the Tivoli Storage Manager server, the pre-migrated state also applied.

- **Migrated**: The file in tier 0 was replaced by a stub file and exists only on the Tivoli Storage Manager server (tier 1). This is the migration state persists until a data access is attempted on the file data.

When a file data access is performed a recall is initiated. The data is staged to tier 0 while the data access system call is blocked, waiting for it to complete. You can recall files from a lower tier to a higher tier by either explicitly invoking an HSM command to recall a file or list of files, or by using the GPFS **mmapplypolicy** command to recall files whose attributes match a set of policy rule criteria. UNIX HSM provides several ways to recall a file. Three recall styles can be used for either manual or transparent recall. The tape-optimized recall method can be used only for a user-initiated list-based recall.

- **Normal**: The application that triggered the recall by accessing the file must wait until the complete file data was recalled from the Tivoli Storage Manager server to the file system. The type of the access (which can be read, write or truncate) has no impact on the recall mode.

- **Partial**: Only the portion of the file that was accessed is recalled. Depending on the access type, this might be a single block of the file or a number of collocated blocks. The recall mode can be applied only for read access to the file. For write or truncate access, the recall mode switches back to normal recall automatically.

- **Streaming**: The application can access the file in streaming mode beginning at byte 0. The recall occurs in the background while the application is allowed to access the file sequentially. The recall mode can be applied only for read access to the file. For write or truncate access, the recall mode switches back to normal recall automatically.

- **Tape optimized**: You generate a list of files to be recalled from tape. The recall command preprocesses the file list and creates ordered file lists, one for each tape. The tape ordered file lists can then be submitted to the HSM **recall** command and the files are recalled.

Each file system that is enabled for UNIX HSM has one management node. In cases of node or network failure, the file system management automatically fails over to another UNIX HSM node in the cluster. Reconciliation methods ensure that the file system content and the content at the Tivoli Storage Manager server are synchronized.
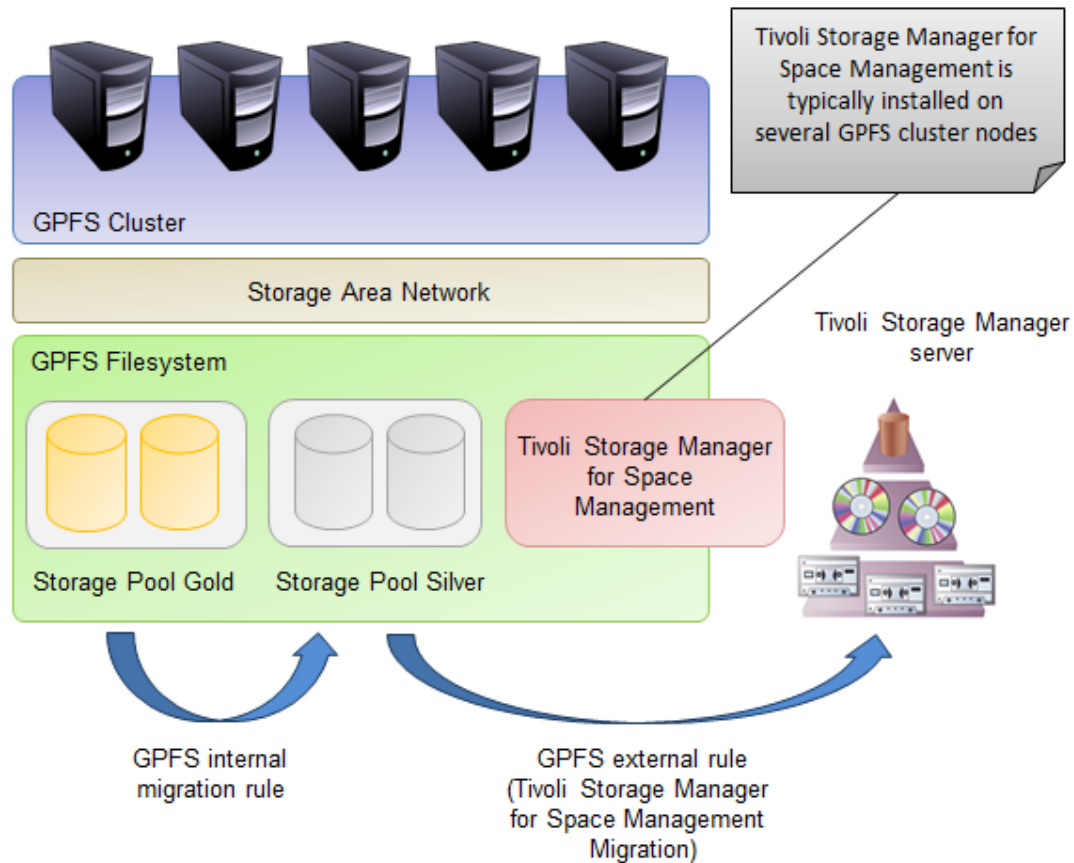
Figure 2:   Integrating Tivoli Storage Manager for Space Management with IBM Elastic Storage

The Tivoli Storage Manager for Space Management client is integrated with Elastic Storage to provide additional hierarchical storage tiers like disk pools and tape pools, which are provided by the Tivoli Storage Manager server. Tivoli Storage Manager for Space Management acts as an external storage pool and integrates seamlessly with the GPFS storage pool concept. The integration between GPFS and Tivoli Storage Manager for Space Management addresses the following functions:

- **Threshold Migration:** Tivoli Storage Manager for Space Management uses the GPFS policy engine to monitor the file system thresholds when the user-defined thresholds are reached, the GPFS policy engine scans the file system and generates candidate lists for migration based on user-defined, fine-grain policy rules. Typically, the rules nominate the large and less frequently used files. Compared to standard file-system scan methods, the policy engine significantly increases the overall performance of the threshold migration.

- **Reconciliation:** Data is synchronized between the file system and the Tivoli Storage Manager server. The Tivoli Storage Manager for Space Management client uses the GPFS policy engine to generate a list of all migrated files in the file system and performs a comparison of the files listed and the files stored on the Tivoli Storage Manager server.

High Availability: The Tivoli Storage Manager for Space Management failover function ensures the high availability of the HSM services. Typically, one node in the Tivoli Storage Manager GPFS cluster is responsible for the HSM activities for a file system. If this node fails (for example,

because of network failures or if the system is powered off) the HSM service automatically fails over to another node in the cluster. GPFS user exit callbacks are used for this function of the Tivoli Storage Manager for Space Management Disaster recovery

4   Disaster recovery

The Elastic Storage Scale Out Backup And Restore (SOBAR) feature is a fast disaster recovery solution for GPFS file systems. It requires the Tivoli Storage Manager for Space Management client and the Tivoli Storage Manager backup-archive client. The solution reduces the recovery time objective (RTO) of a GPFS file system after a disaster.

To be able to recover a GPFS file system from a disaster, four types of information must be copied or backed up before a data loss. Figure 3 illustrates the types of information to be backed up.
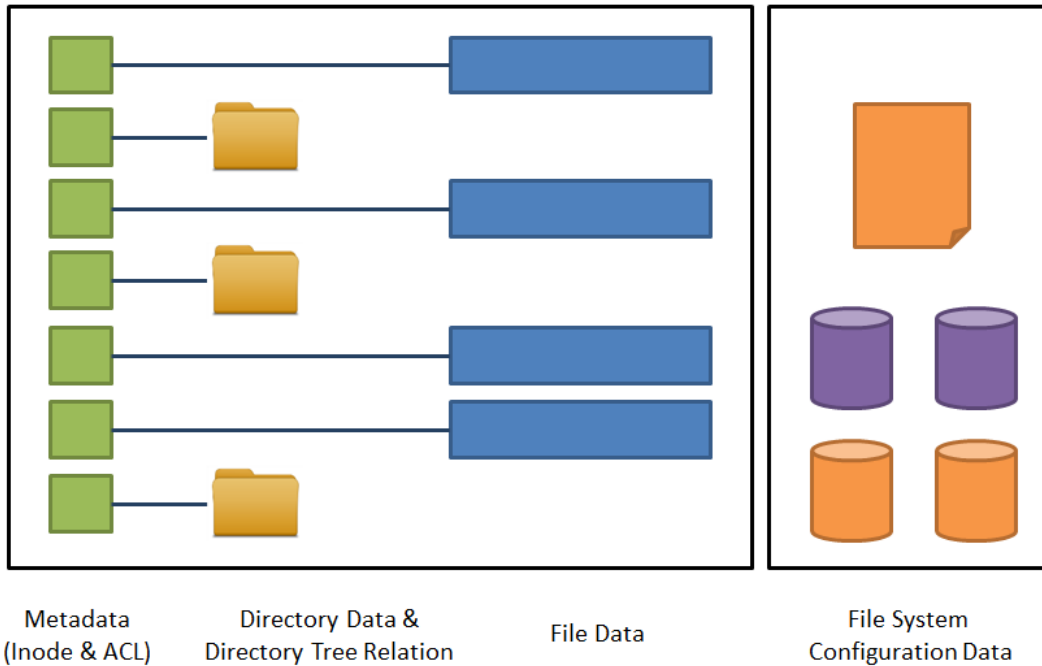

| Metadata (Inode & ACL) | Directory Data & Directory Tree Relation | File Data | File System Configuration Data |

Figure 3:   Types of protected data in a GPFS file system

- **Metadata**: The metadata includes the inode information of the file system including all extended attributes and access control lists. The metadata information includes both file metadata and directory metadata.

- **File data**: The user data that is stored in files in the directory tree of the file system.

- **Directory data**: The path names of all objects and the relation of the directories in the file system directory tree.

- **File system configuration data**: For GPFS, the file system setup information and the GPFS network shared disk (NSD) information for all storage devices. File set

definitions, quotas, and storage pool definitions. The GPFS cluster configuration is not part of this category and must be handled separately.

During disaster recovery of a GPFS file system, the file system is re-created in the same configuration as it was before the disaster occurred. Furthermore, the file system content (all file sets, storage pools, files, directories, access control lists and extended attributes) are re-created as part of disaster recovery. The required processing is called "SOBAR backup," and consists of the following steps as illustrated in figure 4:

- **Meet requirements**: The GPFS file system must be HSM-managed from Tivoli Storage Manager for Space Management. Furthermore, the Tivoli Storage Manager backup-archive client must be set up back up the file-system metadata image to the Tivoli Storage Manager server.

- **Protect file data**: All files in the file system must be HSM-premigrated or migrated to the Tivoli Storage Manager server so that a valid copy of the file data exists for recall after disaster recovery.

- **Protect configuration**: The file system configuration must be dumped to a file. This configuration file must be backed up to the Tivoli Storage Manager server.

- **Protect metadata**: A file-system metadata image must be created. By using the commands provided from GPFS, several files are created to hold this information. These metadata image files are backed up to the Tivoli Storage Manager server.
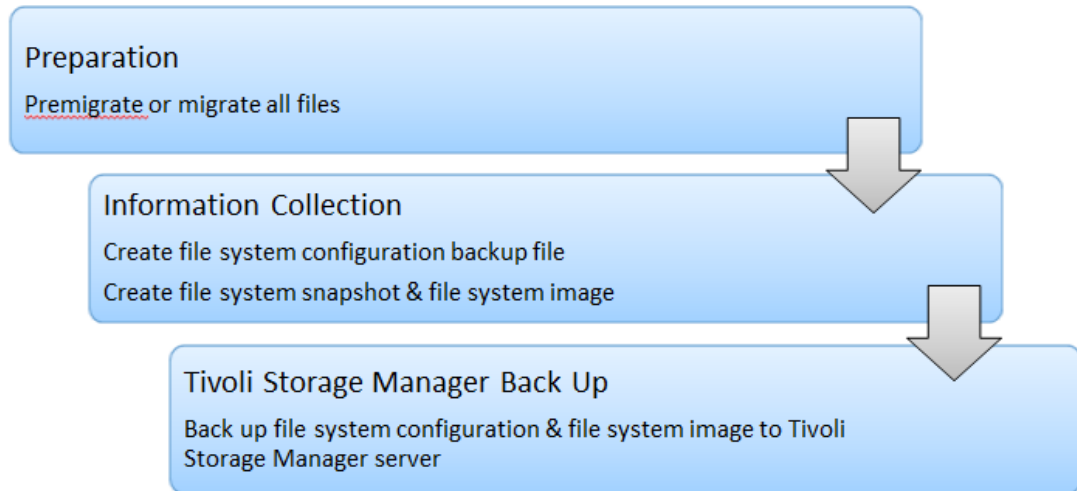

Figure 4:   SOBAR backup process

To recover from a disaster, the process called "SOBAR restore" must be applied. This process consists of the following steps as illustrated in figure 5:

- **Restore configuration and metadata**: Use the Tivoli Storage Manager restore function to restore the file-system configuration data from the Tivoli Storage Manager server backup. The metadata image files must be restored to a location that is accessible by GPFS nodes that will have access to the file system that must be recovered.

- **Recover configuration**: The configuration information is used to re-create the GPFS file system with the configuration as it was before the disaster occurred. GPFS file sets and storage pools are re-created in an empty file system that is formatted on the disk devices.

- **Recover metadata**: The file system metadata must be extracted in the new file system. The extraction process re-creates the inode table of the file system including all extended attributes and access control lists. The directory tree is re-created. All file inodes are re-created in UNIX HSM stub-file format, and the migration state of the files are set to "migrated."

- **Recover file data**: After the file system is created and the file system metadata is extracted, the file system can be mounted and UNIX HSM management can be restored to the file system. Because the files are now in UNIX HSM stub format, they can be accessed immediately and are recalled on demand. In addition to the transparent recall of files, a job can be set up that recalls files in tape-optimized order in the background.
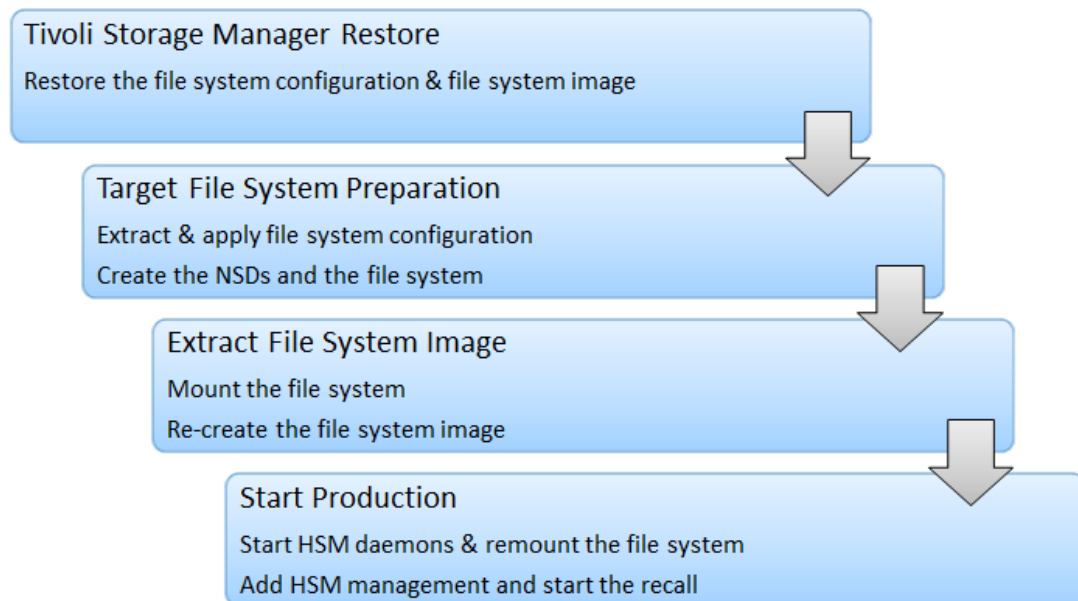
Tivoli Storage Manager Restore

Restore the file system configuration & file system image

Target File System Preparation

Extract & apply file system configuration

Create the NSDs and the file system

Extract File System Image

Mount the file system

Re-create the file system image

Start Production

Start HSM daemons & remount the file system

Add HSM management and start the recall

Figure 5:   SOBAR recovery process

Compared to typical file restore processing of all files that are stored in a file system, SOBAR-based recovery processing takes only a fraction of the usual time to return the file system back to production. Therefore, file system disaster protection by using Elastic Storage SOBAR in combination with Tivoli Storage Manager for Space Management and the Tivoli Storage Manager backup-archive client can be applied as disaster recovery solution for an enterprise business.

References

IBM Tivoli Storage Manager on IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/SSGSG7/welcome

IBM Elastic Storage on IBM Knowledge Center:

http://www.ibm.com/support/knowledgecenter/SSFKCN/gpfs_welcome.html

Tivoli Field Guide - TSM for Space Management for UNIX-GPFS Integration

http://www.ibm.com/support/docview.wss?uid=swg27018848

Tivoli Field Guide - Tivoli Storage Manager for Space Management in a mixed GPFS cluster environment:

http://www.ibm.com/support/docview.wss?uid=swg27028178

Tivoli Field Guide - Performance of GPFS/HSM - Tivoli Storage Manager for Space Management: Part 1: Automigration:

http://www.ibm.com/support/docview.wss?uid=swg27020987

Data Storage Management (XDSM) API:

http://pubs.opengroup.org/onlinepubs/9657099/

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 character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokyo 103-8510, 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 PAPER "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 may be made periodically to the information herein; these changes may be incorporated in subsequent versions of the paper. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this paper at any time without notice.

Any references in this document 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
2Z4A/101
11400 Burnet Road
Austin, TX 78758
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.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

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.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

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

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide.
 Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

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