Backing Up SAS Content In Your SAS®9 Enterprise Intelligence Platform

Considerations for Creating Backups of Your SAS Content

THE POWER TO KNOW®

Table of Contents

Introduction

SAS Enterprise Intelligence software creates and delivers enterprise intelligence through an integrated suite of software technologies and solutions, all based on the foundation of the SAS Intelligence Architecture. Specifically, this architecture integrates industry-leading SAS applications in data warehousing, business intelligence, and analytic intelligence in a single, cohesive platform. These capabilities provide the end-to-end intelligence infrastructure necessary to ensure consistent and reliable enterprise-wide intelligence that is needed for exploring, analyzing, and understanding your business.

SAS recognizes that your organization has made a considerable investment in your SAS software system. Information about the environment, reports generated by end-users, and analytical models that drive strategic decisions represents key corporate assets that directly impact business results. Therefore, your SAS assets must be included in your corporate backup strategy.

This document focuses on considerations for developing a content backup strategy for the SAS®9 Enterprise Intelligence Platform—including content from SAS Data Integration Server, SAS Enterprise BI Server, and analytic products such as SAS Enterprise Miner. SAS typically generates operating system files, so standard backup tools can be used to create backups of most content. However, there are considerations that can make your backup process more effective. This paper is not intended to provide a complete backup discussion, as we do not discuss back ups of configuration files, specific backup systems, or backup management processes for archiving and storage. Periodically, this paper will be updated to include additional topics.

This paper is written for SAS system administrators who are responsible for maintaining the SAS environment and ensuring that service levels are met for your internal constituents. You should be familiar with the SAS®9 architecture and understand how content is generated and stored in your environment.

Understanding the SAS Enterprise Intelligence Platform

The following discussion is a high-level overview of the SAS Enterprise Intelligence Platform. For more detailed information, see the *SAS Intelligence Platform: Overview* (SAS 2006).

The SAS Enterprise Intelligence Platform includes components in the following categories:

- Data Integration and ETL
  The data integration and extract, transform, and load (ETL) components enable you to consolidate and manage enterprise data from a variety of source systems, applications, and technologies. Components are provided to help you cleanse, migrate, synchronize, replicate, and promote your data. Metadata for all your intelligence resources is stored centrally and controlled through a single management interface.

- Business Intelligence
  The business intelligence components enable users with various needs and skill levels to create, produce, and share their own reports and analyses. Easy-to-use interfaces enable users to obtain answers to their business questions, but system administrators retain control over the quality and consistency of the data.

- Analytics
  SAS offers the richest and widest portfolio of analytic capabilities in the software industry. The portfolio includes solutions for statistical data analysis, data and text mining, forecasting, econometrics, quality improvement, and operations research. You can use any combination of these tools with the SAS Enterprise Intelligence Platform to add extraordinary precision and insight to your reports and analyses.

- Intelligence Storage
  The intelligence storage options are optimized for analytical processing. This enables you to quickly retrieve and report on large volumes of data. The options include simple relational databases, a threaded multidimensional database that supports online analytical processing (OLAP), and relational storage with a threaded multiple input/output (I/O) subsystem for intensive use by focused applications.

These components in the SAS®9 Enterprise Intelligence Platform leverage a common set of servers to generate data, analyses, and reports. The term **server** refers to a program or programs that wait for and fulfill requests from client programs for data or services. Each server is connected to content stores that need to be factored into the backup strategy for your SAS environment. Whenever you back up the SAS Metadata Server, you must also back up these content stores. (See Figure 1)

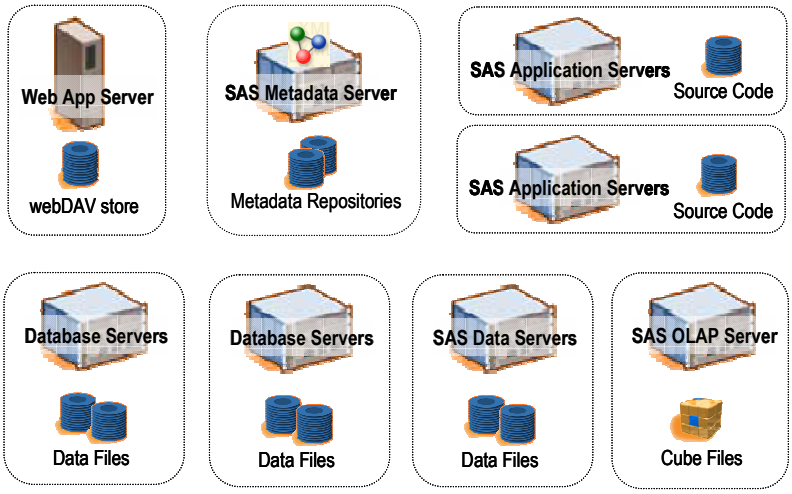

Figure 1. Content Storage in the SAS Enterprise Intelligence Platform

Web App Server – Web-based clients, such as SAS Web Report Studio and SAS Information Delivery Portal, leverage a webDAV store to house content that is surfaced via their interfaces. The recommended webDAV store in the SAS platform is Xythos.

Metadata Server – The SAS Metadata Server facilitates the sharing of content across SAS applications to provide a consistent representation of the business to all constituents in your organization. All SAS applications leverage the SAS Metadata Server to store information about their content in the metadata repositories.

Application Servers - The SAS application servers execute SAS analytical and reporting processes for distributed clients. The source code for these processes is stored in a physical location that is accessible to the server. Types of SAS application servers include the SAS Workspace Server and the SAS Stored Process Server.

Data Servers - Enterprise data includes SAS data (such as data sets, OLAP cubes, or SAS Scalable Performance Data (SPD) Server tables), relational database management systems, or Enterprise Resource Planning (ERP) system tables. These represent the actual physical data stores that feed the SAS platform.

Establish a Formal Backup Process

A formal backup process should be established and managed specifically as a key IT process. Management oversight should be provided to ensure that resources are available to support the backup process and to be responsible for corrective measures if a backup fails. There should be clear documentation regarding the steps required for completing and archiving backups. The following information should be included in the documentation:

- Stopping and starting software server processes
- Network locations of content to be backed up
- Locations, media, and restore facilities where backup copies will be archived
- Utilities and other software tools required to perform the backups
- Schedule of when backups should occur and related service levels appropriate for archive, restore, and scale requirements
- Amelioration and recovery processes for situations where backups fail or take longer than service levels allow

Define Specific Locations for Physical Content

In the SAS®9 environment, applications consume and create data and content that are stored physically in a database, a physical file, or a webDAV directory. Physical content must be backed up in conjunction with the SAS Metadata Repositories because they are dependent on one another and need to stay in synch.

As a best practice, we recommend that you specify standard locations for physical content in order to have well-defined locations for creating backups. These locations need to be accessible by the user ID that is running the backup process.

Physical content includes:

- Source code directories – stored processes, analytical models, SAS Data Integration Studio jobs
- SAS data files and relational database systems
- webDAV content – reports, documents
- the work area for Platform Computing Process Manager software for scheduled job flows

SAS source code directories and data files are standard operating files that can be backed up and recovered using existing backup tools. Relational database management systems should follow the recommended vendor backup strategies, employing vendor-recommended software and systems.

Schedule Backups

Backups of the environment should be a scheduled event. Because the metadata server and relational database servers must be paused to back up the repositories, you need to have a pre-set time when this happens, so that users will know when to expect a short period of down time.

How often you need to back up your environment depends on a number of factors.

- Activity – you should back up more often for a high update environment versus an environment that is more static or read-only.
- Importance to the operation of your organization – you should back up more often if it is critical to the operations of your organization to keep the SAS environment running.
- Service Levels – because you have to pause software services to complete the backup, there will be some SAS system downtime associated with the backup process. You should determine the allowable downtime to complete the backup process, and what schedule window it must reside in.
- Time needed to recover from a failure – if you need to recover quickly from a failure, it is best to make more frequent full backups. If you can take more time or do not have the network resources for multiple full backups, then you might want to supplement a less frequent full backup with incremental backups.

The more frequent the backups, the more difficult it might be to schedule them because backing up does interrupt service. However, you can use snapshot tools in conjunction with your backup process to keep down time to a minimum. This technique will require appropriate systems resources to utilize effectively.

In addition to your regular, full backup schedule, you should perform a full interim backup after any of the following occurs:

- You change the structure of your metadata, such as adding a repository, deleting a repository, or changing the dependencies between two repositories. Because there is content integration between repositories, it is critical that you keep the structures in synch with a full backup.

- You make significant changes to the metadata contents, such as bulk loading users and groups from an external, enterprise data source, conducting a full repository promotion, or implementing significant BI application-driven updates.

- You change the structure of your SAS data files because of changes in your data integration processes.

For metadata and its associated content, we recommend that full backups always be performed. This would include backing up all metadata repositories, webDAV stores, and source code files associated to stored processes, SAS Data Integration Studio jobs, SAS Enterprise Guide projects, and SAS Enterprise Miner models. For example:

- Scenario 1: Initial development in a new SAS environment
    - Back up all content daily
- Scenario 2: Production environment with limited ad-hoc changes
    - Back up metadata and metadata-related content weekly
    - Back up SAS data files weekly
    - Back up all content bi-weekly
- Scenario 3: Production environment with active ad-hoc changes
    - Back up metadata and metadata-related content daily
    - Back up SAS data files daily
    - Back up all content weekly

Store Backups

Backups should be archived by date and time, so that you can restore to the state close to when the failure occurred. You should be able to leverage existing backup storage systems for archiving SAS backups. A complete backup of SAS content including data sources could be very large, so you will need to ensure that your system has enough memory and CPU resources to facilitate the backup.

Once the backup has been created, relevant information should be captured, such as the owner ID, the date, and a description of the contents. In addition, there should be documentation about how the backup was generated. The documentation should include the following information:

- The name of the metadata server and its repositories
- The physical content that is contained, including file path locations
- The scripts and processes that were used to make the backups
- The location of the actual backup files

Backing Up the SAS®9 Environment

The SAS®9 environment creates both metadata and physical content that must remain in synch. As a result, it is critical to incorporate the three key components into the content backup strategy: metadata, webDAV content, and physical files (such as source code files and data files). The natural inclination is to back up each component separately, but because of the integrated nature of the SAS environment, you must back up all components within the same backup window. For SAS metadata and metadata-related content, you should always perform a full backup versus an incremental backup. This will ensure that all dependencies and cross-references stay in synch. Incremental backups can be effectively leveraged for databases, such as relational databases or SAS Scalable Performance Data Server tables.

To back up the content in the SAS®9 environment, follow these general steps:

1. Stop or pause the SAS Metadata Server.
2. Lock down write access to physical content locations, such as source code directories and database tables.
3. Back up metadata repositories.
4. Resume the SAS Metadata Server with its repositories in read-only state, so that applications can read information but not create or edit it.
5. Back up webDAV locations and physical content, such as source code files and SAS data files.
6. Resume the SAS Metadata Server to active status when all backups are completed.
7. Restore access to physical content locations.

Backing Up Metadata

At the core of the SAS®9 Enterprise Intelligence Platform is the SAS Metadata Server. All applications depend on the metadata server to create and access shared resources and content. Therefore, you should plan your backup strategy around what is needed for the metadata server.

As a best practice, you should always back up the entire metadata server. Full backups include the Foundation SAS repository and all custom and project repositories that are registered on the metadata server. By performing full backups, you ensure that dependencies and cross-references between repositories are correctly maintained and that the SAS Metadata Server can be restored to a consistent state.

Performing the Metadata BackUp Process

The metadata repositories that underlie the SAS Metadata Server are built on a set of SAS data sets. When the SAS Metadata Server is active, these data sets are in a "locked" state; they are essentially inaccessible for copying. In order to back up those SAS data sets, you must unlock access to them by pausing the SAS Metadata Server process.

If the server is not stopped or paused, the resulting backup might be unusable. If no backup exists or the backup was not performed correctly, completely rebuilding the metadata repository might be required. For detailed backup instructions, see "Backing Up and Restoring the SAS Metadata Server" in the *SAS® 9.1.3 Intelligence Platform: Administration Guide* (SAS 2006)[1].

Although you can manually back up the SAS Metadata Server, we recommend automating the process by using the %OMABAKUP macro. The %OMABAKUP macro is a SAS batch interface that enables you to perform a complete backup of the SAS Metadata Server. The macro is part of the SAS autocall libraries and runs in a dedicated SAS session.

In its default mode, %OMABAKUP performs the following actions:

- Pauses the repository manager, which changes the access state to read-only status so that repositories cannot be added or deleted
- Pauses each metadata repository. The pause operation does the following:
    - retains all active client connections
    - temporarily changes the repository's access state to offline status so that it stops receiving client requests to read or write metadata. If a client application submits a new request, the client displays a message that tells the user that the server is paused and that the user should try the request later.

[1] In November 2006, the guide will be called *SAS 9.1.3 Intelligence Platform: System Administration Guide.*

    - flushes updated metadata from the server's memory and writes the updated metadata to the metadata repository
    - closes the SAS data sets that serve as repository containers
- Uses PROC COPY to copy the metadata server's repository manager and all its registered metadata repositories to a network path that you specify.
- Creates control files that contain information about the backup.

The user ID that executes %OMABAKUP must have administrative user status on the SAS Metadata Server, as well as full operating system access to the Metadata Server directory, the rposmgr directory, and all repository and backup directories. The user ID should also be compatible with the owner of the repository files. Typically, you should use the installation user ID. If you need to restore files later, the metadata server process user ID will need to have access to the files.

Note: The %OMABAKUP macro does not back up SAS Metadata Server configuration files such as:

- omaconfig.xml
- adminUsers.txt
- trustedUsers.txt
- repository audit trails

You must include these files as part of a full backup or full restore. Use operating-system commands or a third-party backup solution to back up these files.

Because the SAS Metadata Server must be paused, SAS applications will not be able to access metadata during the backup process. To minimize the down time of the metadata server, you can use a third-party tool that can take a quick snapshot of the metadata repositories instead of using the %OMABAKUP macro. The steps to follow are:

- Use PROC METAOPERATE to pause the repositories to an offline state.
- Snapshot the metadata repositories.
- Use PROC METAOPERATE to resume the metadata repositories.
- Back up the snapshot after the SAS Metadata Server is back online.

For an example, the VxFS file system in HP-UX provides a mechanism for taking snapshot images of mounted file systems, which is useful for making backups. The snapshot is an exact image of the file system at the time the snapshot is made. Selected files can be backed up from the snapshot (by using standard utilities such as cpio or cp) or the entire file system image can be backed up (by using the vxdump or fscat utilities). By making a file system that only contains the

SAS Metadata Server backup files, you can quickly take a snapshot of this file system for archiving and maintain the maximum up time for the SAS Metadata Server.

For storage area network (SAN) users, making a snapshot of the data to a spare area on the SAN is a way to ensure that the backup process happens quickly and minimizes the down time of the SAS Metadata Server. The backup of the data can then be done after the SAS Metadata Server is re-started. The tradeoff, if you use this technique, is that you need enough extra disk space to support the second copy of the SAS data store.

Changing the State of the Metadata Repository

After running the %OMABAKUP macro, the repositories may be returned to an active state. However, you must continue to restrict access to the metadata in order to back up webDAV and physical content before returning to an active state. You can change the state of the metadata repositories to be read-only, so that client applications can still surface content but cannot edit or create new metadata.

A SAS Metadata Repository can exist in one of three modes: full access, read-only, and offline. A repository's intended access mode is set when the repository is created and is recorded in the repository's Access attribute. For example, if a repository was configured to be read-only at creation, then it will have an Access value of OMS_READONLY. The default access mode is full access.

A repository's access mode can be temporarily changed to a more restrictive state by issuing a Pause action using PROC METAOPERATE. For example, a repository that was created with an access mode of full access can be paused to a read-only or offline state. A repository that was created with an access mode of read-only can be paused to an offline state; it cannot be paused to full access.

The following SAS code example connects to a metadata server and issues a PAUSE action to change client activity on all repositories and change the repository manager to a READONLY state. The State parameter is omitted from the options string because the default value for the Pause state is READONLY.

```
PROC METAOPERATE
   SERVER="machine.us.company.com"
   PORT=8561
   USERID="myuserid"
   PASSWORD="mypassword"
   PROTOCOL=BRIDGE

   ACTION=PAUSE
   OPTIONS="<Repository Id=""ALL""/>
            <Repository Id=""REPOSMGR""/>"
   NOAUTOPAUSE;

RUN;
```

The metadata server connection options can be handled via parameters, so that you do not have to hard-code the values into your program. More information about PROC METAOPERATE is available in online documentation at:
support.sas.com/onlinedoc/913/getDoc/en/omaref.hlp/procmetaserver.htm

Backing Up WebDAV

A webDAV store contains the content that is used in Web applications. SAS Web Report Studio and SAS Information Delivery Portal use a webDAV server to store content. You should back up the webDAV store whenever you back up the metadata server to keep the content in synch.

The SAS®9 platform supports two webDAV providers: Apache Group and Xythos. The Apache webDAV store is straightforward; it uses standard file system support. You can use system tools to back up the files. SAS can also provide a webDAV dump utility (similar to the WFSDump utility discussed later). If you want access to this utility, contact SAS Technical Support.

The Xythos WFS webDAV architecture uses a relational database and the file system, and both components must be backed up. Your backup and recovery strategy will vary depending on the number of Document Stores being used, as well as, whether External Storage is enabled. For a discussion of backup and recovery considerations, see the "Database Recovery Guide", which is bundled in the installation:

- WFS 4.0: <xythos_home>wfs-4.0.x/doc/wfs_dbrecovery.html
- WFS 4.1: <xythos_home>wfs-4.1.x/doc/en/wfs_admin.html#_DB_Recovery
- WFS 4.2: <xythos_home>wfs-4.2.x/doc/en/wfs_admin_file_system.html#_backup_restore

File System

The file system component is a set of files written to disk. This disk can be a network attached storage (NAS), a storage area network (SAN), or directly attached. You can use operating-system commands or third-party tools to back up the series of file systems that are used by Xythos WFS. Xythos WFS adds a temporary storage area that works much like a database redo log: when a file is added to the system, the file is written to its primary disk and a temporary disk. If the primary disk crashes, a sysadmin can restore the primary disk using a backup. After restoration, Xythos WFS will move the files that are missing (files that were written after the backup but before the primary disk crash) from the temporary storage to the primary storage. The temporary storage is emptied on a time schedule that corresponds to your file system backup schedule. In this way, both the database and the file system can always be restored to the same point in time of any system crash.

Relational Database for Xythos

Because Xythos leverages a transactional database (postgres, SQL Server, DB2, or Oracle), the transactional database can be configured to provide backups and replay logs. If the database crashes, it can be recovered by replaying the logs and safely reconstructing the database from a backup to the time it was last running. If external storage is not being used, follow the database software instructions to recover from a database crash.

External Storage Backup

All files in the WFS system are stored in one or more Document Stores. If external storage is being used, each document store writes files across both a database and file system(s) (external storage locations). The file metadata (name, location, permissions, and so on) are stored in the database while the file contents (bytes, blobs, and so on) are stored on external storage locations. Both of these systems must be backed up.

If external storage is being used, use of the database as an intermediary still guarantees that the file bodies can be recovered with the same reliability. Xythos recommends taking advantage of the two storage locations that are available per Document Store when configuring your backup strategy. In order to reduce the amount of data in the database, Xythos recommends configuring both primary storage (the Storage Location parameter on the Document Store administrative page) and temporary storage as external storage locations.

The primary storage location should be backed up at some interval. The amount of time between backups of the primary storage location drives how long files should be kept in the temporary storage location (Temporary Storage Period). The Temporary Storage Period, which is configured on the storage page of the Xythos WFS Admin, should be about twice your primary storage backup interval. If you have a crash on primary storage, you can recover the primary storage to its most recent backup, and then use the External Storage Recovery feature in the Xythos WFS Admin to replay the files that were written since the last primary backup and the time of the crash. IE files are moved from your temporary storage to the primary. The temporary storage location acts as your "redo/recovery" buffer and captures new file content that hasn't been backed up yet.

WFSDump Utility

SAS also provides a WFSDump utility, which can be found in the Xythos install directory (<xythos_home>/custom/bin). This utility requires that the SAS Metadata Server be available to respond to requests. You can pause the metadata repositories to a read-only state (as discussed earlier) to prevent any content from being modified or added during the backup process.

The WFSDump utility is primarily used to migrate content from one webDAV server to another. However, you can use the same utility to provide backups. The utility creates an XML file that describes the content and one file for every content resource that is found in the server. The files that are created by the WFSDump utility will probably be a little larger than the database and file storage location backup.

Backing Up Physical Content

Because SAS also leverages physical content that is stored on the file system, your backup strategy should include the following information:

- source code files for stored processes and stored process results that are stored in permanent packages
- packages that have been published to a channel's persistent store location (if using the file system or FTP location)
- data sources, including the following:
    - SAS data files, such as SAS data sets, SAS catalogs, and OLAP structure files
    - SAS Scalable Performance Data Server data
    - relational database sets
- XML files and XML maps that are contained in libraries
- custom code, processes, and post-processes that are associated with SAS Data Integration Studio jobs
- source files for SAS Data Integration Studio jobs that have been scheduled or deployed for scheduling, which could include the work area for Platform Computing Process Manager software.
- generated and user-written SAS code, reports, sample and work data sets, and project properties for SAS Enterprise Miner models

These physical content files are stored as standard operating system files. You can use operating-system commands or third-party tools to back up these files. You should ensure that all in-memory changes have been flushed to disk before performing backups. You can do this by issuing a "sync" command, which is supported by both Windows and UNIX environments.

Considerations for SAS Data Files

SAS data files, such as SAS data sets and OLAP data files are locked if there is an active server process that is accessing them. In order to back up these files, you must stop the SAS processes. This may require stopping all application servers before starting the backup process, which includes Workspace Servers, Stored Process Servers, and OLAP Servers. Once these files are available, they can be backed up using standard backup tools.

Considerations for SAS Scalable Performance Data Server

The SAS Scalable Performance Data Server (SAS SPD Server) tables can be backed up using standard backup tools and integrated into existing backup processes. Full backups are recommended if the SPD Server data is partitioned, as restoration time would be fairly quick. However, if the data tables are large, an incremental backup strategy may be more efficient.

SAS SPD Server also provides backup and restore utilities to alleviate resource problems that are associated with backing up large data tables. Instead of backing up the entire table, the utility backs up only the records that changed after the previous table backup date.

For more details about backing up your SAS SPD Server tables, see the *SAS Scalable Performance Data Server: Administrator's Guide* (Second Edition) (SAS 2006).

Conclusion

To safeguard your investment in developing SAS content, it is imperative to establish a clear strategy to manage backups of content. Because backups must entail metadata, webDAV content, and physical content, you must have a coordinated process to handle all components. To maintain the integrity of the backups, you must retain existing dependencies and cross-references.

The focus of this paper is dealing with a single metadata server configuration. With more complex environments that involve multiple metadata servers and 24x7 service level requirements, there are additional considerations that are beyond the scope of this paper. If you would like to have more detailed discussions about how to manage backups in a more complex environment, please contact SAS Consulting Services.

With a combination of system tools, SAS tools, and internal processes, you can establish a solid strategy for backing up SAS content. These backups can be used for auditing and recovery purposes if a system failure occurs that requires re-staging the environment. Although we do not discuss recovery in this paper (there will be a paper about recovery forthcoming), many of the tools that are discussed here provide facilities for recovery as well. See the product documentation for details about this functionality.

References

SAS Institute Inc. 2006. *SAS® 9.1.3 Intelligence Platform: Administration Guide*, Fifth Edition. Available at support.sas.com/documentation/configuration/bicag.pdf.[2]

SAS Institute Inc. 2006. *SAS® Intelligence Platform: Overview.* Available at support.sas.com/documentation/configuration/biov.pdf.

SAS Institute Inc. 2006. *SAS ® Scalable Performance Data Server® 4.4: Administrator's Guide*. Available at support.sas.com/documentation/onlinedoc/91pdf/sasdoc_913/scalable_ag_9719.pdf.

[2] In November 2006, the guide will be called *SAS 9.1.3 Intelligence Platform: System Administration Guide*

SAS INSTITUTE INC. WORLD HEADQUARTERS 919 677 8000

U.S. & CANADA SALES 800 727 0025 SAS INTERNATIONAL +49 6221 416-0 **WWW.SAS.COM**