Using the SAS® Deployment Backup and Recovery Tool in the Third Maintenance Release of SAS® 9.4

Bhaskar Kulkarni, SAS Institute Inc., Cary, NC

ABSTRACT

The SAS® Deployment Backup and Recovery Tool in the third maintenance release of SAS® 9.4 helps SAS administrators to collectively take backups of important data artifacts in SAS deployments, which include SAS® Metadata Server, SAS® Content Server, SAS® Web Infrastructure Platform Data Server, and physical data in the SAS configuration directory. The tool supports all types of deployments, from single-tier to multi-tier clustered heterogeneous host deployments. The new configuration options in the tool give administrators more control over the data that is being backed up, from the SAS tiers to the individual directory level. The new options allow administrators to filter out old log directories and to choose the databases to back up. This paper talks about not only how to use the configuration options but also how to mediate the effects that the SAS deployment configuration changes have on the backups and how to optimize the backups in terms of size and time.

INTRODUCTION

The SAS® Deployment Backup and Recovery Tool takes a snapshot of the data spread across various data sources. The typical data sources include the following:

- SAS® Metadata Server

- SAS® Content Server

- all instances of SAS® Web Infrastructure Platform Data Server

- physical data in the SAS configuration directory spread across multiple tiers

The backups are scheduled by default and run weekly. All the data sources that need backup are auto-discovered. If a newer release or solution is added to the current deployment and results in an additional data source, that source is detected and added to the backups automatically.

The third maintenance release of SAS® 9.4 introduced various configuration options to control what goes into a backup. A new graphical user interface is also included as a plug-in to SAS Environment Manager.

The new backup configuration options introduced abilities to filter the data as well as the data sources from the backup. The new backup configuration options can be added by using the command-line utilities.

BACKUP CONFIGURATION

Backup configuration provides a view of the data sources that are included in the backup and details about the location of the data sources in the SAS deployment. This view not only provides information about the data being backed up, but also provides a topological view of the deployment.

The Deployment backup plug-in within SAS Environment Manager shows a topology diagram of the backup configuration. The diagram can be viewed in two modes, a machine view and a backup data source view.
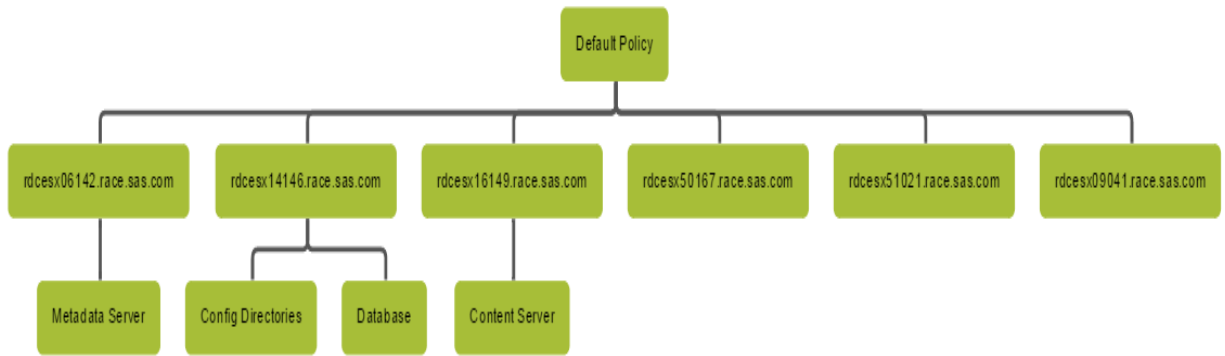
Figure 1. Backup Configuration Machine View
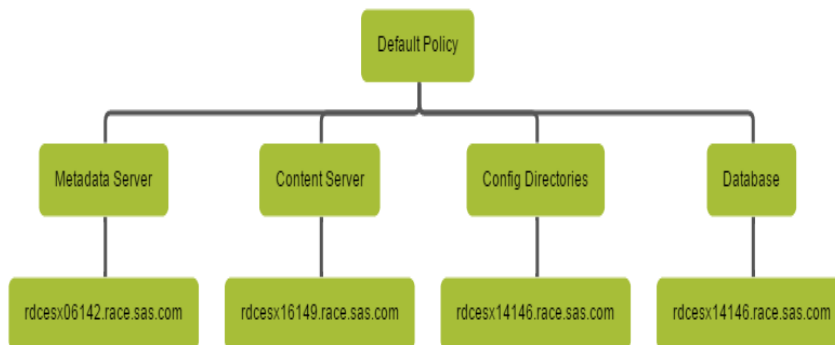
Source View

Figure 2. Backup Configuration Backup Source View

2

PREPARING A BACKUP CONFIGURATION STRATEGY

The SAS Deployment Backup and Recovery Tool by default includes all the data sources in the backup and takes the backup of all the data. But as the data inside a system grows, it becomes important to identify and choose the exact data that should be backed up. The new configuration options allow you to filter SAS tiers, SAS Web Infrastructure Platform Data Server, and physical data in the SAS configuration directory.

Identify Tiers to Back Up

In most cases, the default tiers included in the backup are required, because those tiers hold backup data sources. However, in a few situations, especially in a grid environment where the worker nodes do not have anything to back up, tiers can be excluded from the backup.

Identify Physical Data to Back Up

Data and configuration within configuration directories are part of the backup. Typically, the backup includes the data from the following directories:

```
{SASConfiguration}/{ApplicationContext}/Data
{SASConfiguration}/{ApplicationContext}/SASEnvironment
{SASConfiguration}/{ApplicationContext}/WorkspaceServer
{SASConfiguration}/{ApplicationContext}/StoredProcessServer
{SASConfiguration}/OLAPServer
{SASConfiguration}/ObjectSpawner
```

Identify where crucial data resides. If the data is not in one of the directories shown above, the deployment backup tool supports adding additional directories from SAS configuration directories. Not all data inside the directories is essential, so some data can be left out of the backup. Typically, log files, work directories, and temporary files are good candidates for filtering. Over time, the data size increases and so does the time to take the backup. In some cases, filtering such data from the backup makes the backup faster.

Identify Scheduling Requirements

The default backup schedule is Sunday at 1.00 AM, with a default retention period of 30 days. This means that at any given time you have the last four backups. The scheduling frequency and the retention period go hand in hand. If you plan to change your schedule from weekly to daily, for example, this will affect the number of backups in your backup vault. For a daily backup with retention period of 30 days, there will be 30 backups sitting in the backup vault. So if you want to change the default schedule, don't forget to change the retention period accordingly.

CONFIGURATION SETUP THROUGH COMMAND-LINE UTILITIES

The backup command-line utility *sas-update-backup-config* is the primary backup utility for setting updates or removing the backup configuration.

The backup configuration command-line utility requires following parameters:

1. profile – specifies a file containing the connection URL and credential information for the SAS environment.

2. sourcetype – specifies the backup source type.

3. inputdata – specifies the actual configuration filter information in JSON format.

FILTER SAS TIERS

There are situations when there will not be any data to back up on SAS tiers. In a typical grid environment, the worker nodes do not have any data to back up. In such cases, the SAS tier can be filtered from the backup. There are other situations when it is necessary to filter entire SAS tiers from the backup. The Deployment Backup and Recovery Tool relies on the configuration metadata to identify the SAS tiers to include in the backup , but sometimes stale configuration metadata can cause the backup operation to include a non-existent SAS tier, causing backup failure.

Run the sas-update-backup-config utility to exclude the tiers from backup:

```
sas-update-backup-config –profile environment.properties –sourcetype
                              sastiers –inputdata update-sastiers.json
```

Use the following JSON format to exclude a SAS tier:

```
{
    "tierName": "tier2",
    "included": false
}
```

Run the sas-dislay-backup-config command to find out the tier name:

```
sas-display-backup-config –profile environment.properties –sourcetype
                                                            sastiers
```

```
sas-display-backup-config
        sasTiers
                id:       tier1
                tierName:        tier1
                sasHome:         C:\Program Files\SASHome
                sasConfig:       C:\SAS\Config\Lev1
                included:        true
        tierHost
                        hostType:       wx6
                        hostName:       host.fyi.com


        id:     tier2
        tierName:       tier2
        sasHome:        /install/SAS/Home
        sasConfig:      /install/SAS/Config/Lev1
        included:       true
tierHost
                hostType:       lax
                hostName:       host1.fyi.com
```

Output 1. Output from sas-display-backup-config Command

In the above example output, If **host.fyi.com** is the SAS tier to be excluded, then use **tier1** as **tierName**.

FILTER SAS WEB INFRASTRUCTURE PLATFORM DATA SERVER

The database filtering can be done at two levels: at the individual database server level and at the database instance level inside each database server. It is not recommended to filter the database

servers, but in some cases, especially in a test environment where there is a need to replicate a specific database server on some other system, all the other database instances can be excluded so that the backup is taken on an individual server. The configuration can then be reverted to the original state.

To filter a database server, the name of the SAS Infrastructure Platform Data Server is required.

Run the sas-update-backup-config utility to exclude a specified database server from the backup:

```
sas-update-backup-config –profile environment.properties –sourcetype
                         database –inputdata update-db.json
```

Use the following JSON format to exclude database server:

```
{
    "name": "Sample Platform Data Sever1",
    "included": false
    "sasTierId": "tier2"
}
```

Run the sas-dislay-backup-config command to find out the tier name:

```
sas-display-backup-config –profile environment.properties –sourcetype
                                                        database
```

```
sas-display-backup-config
        sourceType:    database
        configurableServers
               name:    Sample Platform Data Server 9.4
               sasTierId:       tier2
               included:        true
```

Output 2. Output from sas-display-backup-config Command

Use **sasTierId** as **tier2** and the **name** as **Sample Platform Data Server 9.4** to filter the database server.

FILTER THE DATABASE INSTANCE

A typical database server holds multiple databases. In some cases, the database can hold temporary data, and it can re-create itself at the start-up after a restore. SAS Environment Manager uses the SAS Web Infrastructure Platform Data Server database to store its data. Most of the SAS Environment Manager data is about system monitoring and the current state of the system. Such databases are good candidates for exclusion if time and space constraints play a major role in backup strategy.

Run the sas-update-backup-config utility to exclude a specified database from backup:

```
sas-update-backup-config –profile environment.properties –sourcetype
                         database –inputdata update-db.json
```

Use the following JSON format to exclude the database EVManager:

```
{
    "name": "Sample Platform Data Sever1",
    "included": true
    "sasTierId": "tier2"
    "filter": {
        "excludes": [
                "EVManager"
         ]
    }
}
```

FILTER PHYSICAL DATA FROM THE SAS CONFIGURATION DIRECTORY

As the data size grows, the time and size parameters for backup also grows. It then becomes necessary to filter out the physical data. To filter out the physical data, the SAS Deployment Backup and Recovery Tool uses ANT-like regular expression syntax to define the filter paths on the file system. Log files, working directories, and derived data such as cubes and large physical files are good candidates for physical data filters.

Run the sas-update-backup-config utility to exclude single or multiple files using regular expression-like syntax to define the filter paths:

```
sas-update-backup-config –profile environment.properties –sourcetype
                            configdirectories –inputdata configdir.json
```

Use the following JSON format to exclude all temporary files having the extension *.tmp* from the SAS configuration directory:

```
{
    "name": "tier1",
    "included": true
    "sasTierId": "tier1"
    "configurablePath": "/sas/config/Lev1"
    "filter": {
        "includes": [
                "/sas/config/Lev1/SASApp/Data",
                "/sas/config/Lev1/SASApp/SASEnvironment"
         ]
         "excludes": [
                "**/*.tmp"
          ]
    }
}
```

The following table shows the examples of regular expressions that can be used to filter the file system paths.

Regular Expression	Description
**/*.log	All log files in the current directory and all its subdirectories
parent/**/*.*	All subdirectories of **parent**
parent/child/cubes*.txt	All the files starting with the name **cubes** and having the extension **txt** inside the **parent/child** directory
Parent/**/child1/**	Any directory named **child1** at any depth inside the **parent** directory
Parent/**/child1/child2/*.log	All log files from the path **child1/child2** at any depth inside the **parent** directory

Table 1. Example Regular Expressions Supported to Filter Physical Data

CONCLUSION

The SAS Deployment Backup and Recovery Tool at the third maintenance release of SAS 9.4 provides a rich set of features to define a backup strategy by applying configuration options. It also provides a graphical user interface as part of the SAS Environment Manager plug-in. The network graphs on the user interface not only provide the backup source view but also help in identifying the SAS deployment topology.

The backup configuration updates can easily be reset by removing the applied filters on the backup data sources.

Define a backup strategy through the new backup configuration options to save on time and space requirements for backup.

RECOMMENDED READING

SAS Institute Inc. 2015. *SAS® 9.4 Intelligence Platform: System Administration Guide, Fourth Edition.*
    Cary, NC: SAS Institute Inc. Available
    http://support.sas.com/documentation/cdl/en/bisag/68240/PDF/default/bisag.pdf.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author:

Bhaskar Kulkarni
Level 2A & Level 3, Cybercity, Tower 5, Magarpatta city, Hadapsar
Pune, Maharashtra India – 411013
SAS Institute Inc.

+91 20 4911 8700
Bhaskar.kulkarni@sas.com

http://www.sas.com