Paper SAS565-2017

Do You Have a Disaster Recovery Plan for Your SAS® Infrastructure?

Margaret Crevar, SAS Institute Inc.

ABSTRACT

Are you prepared if a disaster happens? If your company relies on SAS® applications to stay in business, you should have a Disaster Recovery Plan (DRP) in place. By a DRP, we mean documentation of the process to recover and protect your SAS infrastructure (SAS binaries, the operating system that is tuned to run your SAS applications, and all the pertinent data that the SAS applications require) in the event of a disaster. This paper discusses what needs to be in this plan to ensure that your SAS infrastructure not only works after it is recovered, but is able to be maintained on the recovery hardware infrastructure.

INTRODUCTION

For your DRP to be successful, it is very important that you have a very good understanding of the SAS® applications that your SAS users are using. For example, you should be familiar with which SAS data files and jobs are associated with mission-critical SAS applications, and which ones are just used for ad hoc reporting. Of these mission-critical input data files, are they only stored in SAS data sets or will they be re-created from other external data sources? If they are only stored in SAS data sets, how often are they updated? What times of the day, week, or month are these files being accessed or not being accessed? All of this information helps determine how often and when this data needs to be backed up and duplicated to the SAS disaster recovery systems.

In addition, you need to be familiar with the *SAS Position Statement Regarding Disaster Recovery* paper. This paper lists what needs to be done for SAS Technical Support to support your DRP.

This paper lists considerations of what needs to be backed up and makes suggestions for how often. How you accomplish this at your site depends on decisions made in your DRP. You need to make sure your IT administrators and SAS users are in agreement about the DRP and associated recovery objectives.

Now that we have established the purpose of the paper, let's start by defining several terms we use in this paper.

DISASTER RECOVERY

For this paper, we use the definition from Wikipedia:

> *Disaster Recovery (DR) involves a set of policies and procedures to enable the recovery or continuation of vital technology infrastructure and systems following a natural or human-induced disaster. Disaster recovery focuses on the IT or technology systems supporting critical business functions, as opposed to business continuity, which involves keeping all essential aspects of a business functioning despite significant disruptive events. Disaster recovery is therefore a subset of business continuity.*

DISASTER RECOVERY PLAN (DRP)

As we all know, some disasters cannot be avoided. However, with careful planning, the impact of a disaster on your business operations can be minimized. Documenting and regularly validating your DRP and having agreement and buy-in among all parties is the first step in this process. The second step is understanding what needs to be backed up, methods that can be used to back it up, and how often. In addition to these steps, to minimize downtime and data loss, you need to measure in:

- The recovery time objective (RTO): This is the time in which the SAS application must be restored after a major incident (MI) has occurred to meet the demands of the customer.
- The recovery point objective (RPO): This is the age of the files that must be recovered from backup storage for normal operations to resume as a result of a MI.

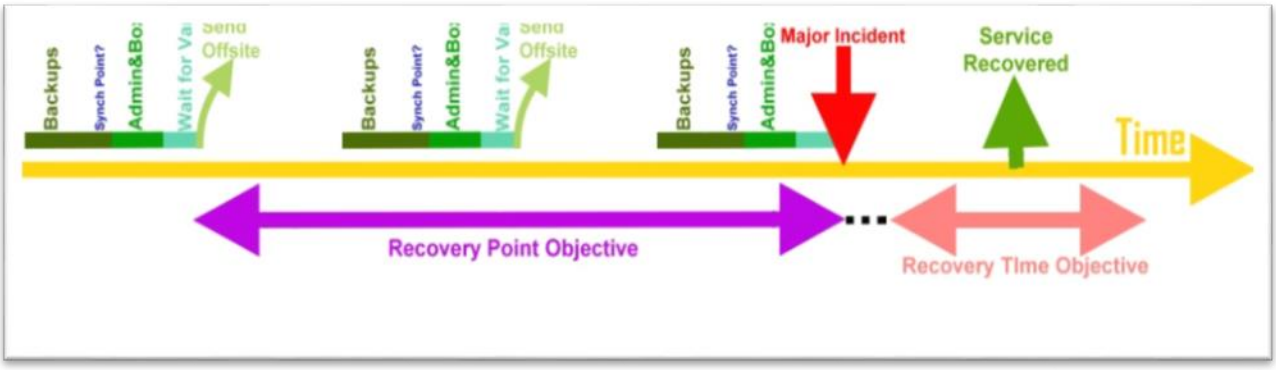

Figure 1. A DRP (Source: Wikipedia.org)

DISASTERS

For the purposes of this paper, there are two primary types of disasters that you need to make plans for– natural and man-made. A natural disaster includes floods, tornadoes, hurricanes or cyclones, earthquakes, heat waves, landslides, tsunamis, and volcanic eruptions. A man-made disaster includes industrial accidents, power failures, terrorism, explosions, fires, or acts of war.

PLANNING METHODOLOGY

An article in the *Disaster Recovery Journal* listed 10 areas that need to be considered when developing your DRP. Here is a short description of each area. You can read the article (listed in the **Reference** section) if you would like more details.

1. Obtaining Top Management Commitment: For any plan to be successful, you need a sponsor in upper management. Resources that management must allocate include both financial and personnel considerations.
2. Establishing a Planning Committee: This committee is responsible for overseeing the development and implementation of the plan.
3. Performing a Risk Assessment: A risk analysis and business impact analysis need to be part of the DRP.
4. Establishing Priorities for Processing and Operations: Critical needs of the users and consumers of your SAS applications need to be determined and prioritized. The amount of time each SAS user can operate without their SAS application is determined and added to the RTO.
5. Determining Recover Strategies: Setting up SLAs is very important. These SLAs include the amount of time to get SAS applications back up and running and to what level (for example, not losing more than a day's worth of updates).
6. Collecting Data: Collect information about what SAS applications and data files need to be backed up and how often.
7. Organizing and Documenting a Written Plan: In this phase, the DRP is developed and shared with others.
8. Developing Test Criteria and Procedures: Best practices specify that DRPs be thoroughly tested and evaluated on a regular basis.
9. Testing Plan: A dry run of the plan needs to be executed to make sure that all the information that is needed after a disaster has been identified. Any missing information needs to be added to the plan, and the documentation needs to be updated. Also, bring up the SAS® Deployment Wizard to make sure it is working and that it can see all the files it needs so that you will be able to update the SAS deployment if you have to on your restored disaster SAS disaster recovery system.

10. Obtaining Plan Approval: The plan needs to go back to the folks who are sponsoring the plan for their final approval.

It is very important to get all of the above decided on and agreed upon so that in the case of a disaster, there will be no surprises.

Now that we have all the definitions in place, we need to talk about the specifics of what needs to be kept in sync between your original SAS production system and your SAS disaster recovery system from a SAS perspective.

WHAT NEEDS TO BE REPLICATED ON YOUR SAS DISASTER RECOVERY SYSTEMS?

There are three distinct areas that need to be replicated to ensure your SAS infrastructure and associated source data files function and are able to be updated post recovery. These are:

1. SAS Deployment, Third-Party Applications, and Operating System (OS) Files: The SAS deployment and binary files, any third-party applications, the home directory of the user ID that was used to install SAS, and all OS files should be backed up after the initial install of SAS **and** every time **any** of these files are updated. The best way to do this is to create a clone of:

    - the source/production system, including the operating system kernel settings and the registry
    - all SAS deployment files (both configuration files and SAS binary files)
    - all third-party applications used by SAS
    - all home directories

    On a UNIX or Linux system, this translates to a copy of the file system from `/` down, except for various temporary directories like `/usr/tmp` and `/tmp` and the SAS WORK and UTILOC file systems.

    The cloning process needs to be performed again after any action that modifies any of the above systems. Examples include applying operating system updates; changing a registry or kernel setting; changes from SAS patches, hot fixes, or maintenance releases; changes to the SAS configuration that have been applied to SAS; and changes to the operating system (these changes include when a customer runs any of the numerous tasks in the SAS Deployment Wizard to change the SAS configuration (for example, but not limited to, updating host name references, updating configuration, removing configuration, running deployment agent configuration tasks, applying hot fixes, changing passwords, etc.)).

2. SAS Infrastructure and Configuration Files: Files associated with SAS applications and components that are updated on a daily basis. This includes:

    - SAS® Metadata Repository
    - contents of the Data, SASEnvironment, and SAS server configuration directories
    - SAS® Content Server repository, the databases managed by the SAS® Web Infrastructure Platform Data Server
    - the lsf.cfg file from the LSF install directory
    - additional directories under *`SAS-configuration-directory`*`/Levn` as specified by the administrator
    - all support files for the SAS solutions being used

3. SAS Data Files: These are the files used by SAS applications, which can be both data files and SAS code files.

WHAT BACKUP TOOLS TO USE

Now that you understand what files need to be backed up, we need to discuss what backup and recovery tools you can use for your disaster recovery activities to perform the backup and to move it to an off-site location designated in your DRP.

- Full-System Backups

  Use any disk cloning or disk imaging of all disks to create and maintain a full-system backup or clone of the SAS production system. By full-system backup, we mean, at a minimum, the operating system (including user and group identifiers, environment variables, mount points, kernel settings); all user home directories; all third-party applications; SAS deployment (SAS binaries and all SAS configuration directories); SAS data files; and all external data files that are needed by your SAS applications (if they do not have their own DRP in place).

- SAS Infrastructure and Configuration Files

  The SAS® Deployment Backup and Recovery Tool (SAS DBRT) can be used to back up the content associated with several (but not all) SAS services. The SAS Metadata Server; the contents of the Data, SASEnvironment, and SAS server configuration directories; the SAS Content Server repository; the databases managed by the SAS Web Infrastructure Platform Data Server; and additional directories under *`SAS-configuration-directory`*`/Levn` as specified by the administrator are backed up with the SAS DBRT. Other files and directories can be added to what this tool backs up by your SAS administrator. Please note that this tool was designed to back up data and restore the data to a single system. It does not have any tools to change the host names that are part of the above content.

- SAS Data Files

  Many SAS customers have 100s of GBs to a few PB of SAS data files. It is important to determine the SAS data files that are unique (for example, that cannot be re-created from external databases) to back up. Please note that the SAS WORK and UTILLOC file systems do not need to be backed up because they are temporary SAS data files.

HOW OFTEN TO BACK UP THE VARIOUS FILES?

- SAS Deployment, Third-Party Applications, and Operating System (OS) Files: These need to be cloned after the initial installation, and then a new clone after any updates and changes to files such as hot fixes, patches, maintenance releases, password changes, etc.
- SAS Infrastructure Files: These are files associated with SAS applications or components that are updated on a daily basis. These files can be updated on a daily basis, so how often you make a backup will depend on the RPO you have established in your DRP.
- SAS Data Files: These are the files used by SAS applications. Like the SAS infrastructure files, some of these files are updated on a daily basis, but some are archived versions of data from previous years used as input into your analysis. So, how often you make a backup depends on the RPO you have established in your DRP and which category the data files fall into. Here are some things to consider:
  - What is the source of these SAS data files? If the source is from a corporate data warehouse in an external RDBMS that is already covered by your DRP, do you need to back up the SAS version of this data?
  - Is this a newly created SAS data file? Or, is it one that has been around for several weeks or months and is already in an existing backup?

For bullets two and three, you need to do these at scheduled times during the day, not continuously like some of the storage devices and shared file systems allow you to do. This is because the files associated with SAS must be in an unlocked (closed) state before you back them up, especially the data files associated with our in-memory servers. If you just copy or back up the files without closing them, these files, when restored, will present themselves as corrupt or incomplete.

For bullet 3, you should check with your business users to see what is the most mission-critical data stored only in SAS data files. Data stored in external databases should have their own DRP. Also, if any of the SAS data files are static, you should consider excluding them from your weekly backups. You can determine when any of your SAS files were last touched using UNIX commands. Or, you can use a SAS macro to see how often they are accessed and updated. The macro is available in *Ensuring that Your SAS Infrastructure Is Able to Meet Your SAS Users' Demands*. Here are the UNIX commands you can run:

- *find /xxx -atime 90* This command finds files accessed in the last 90 days in the `/xxx` directory.
- *find /xxx -mtime 90* This command finds files modified in the last 90 days in the `/xxx` directory.

This is a good way to monitor how often the files are accessed and modified so that you can understand whether they need to be backed up and on what schedule.

The next discussion is how to achieve the above. Everything we are asking can be done with standard backup tools or the snap features of storage arrays and clustered file systems, especially for the SAS data files that are used by the SAS applications. However, for the updates to the SAS deployment files, to the operating system, or to patches/hot fixes/maintenance releases, you might want to enlist a third-party tool like Puppet (https://puppet.com/product/how-puppet-works) that keeps track of all operating system and application deployment files.

DOES ANYTHING NEED TO BE DONE AFTER A RESTORE TO MAKE SAS FUNCTION?

Because you have backed up the operating system (which has the host name embedded in it), you might need to change that host name after you restore all the data. This depends on what tool you use to back up the SAS configuration files. If you use the SAS® Migration Utility to do a migration or promotion, this tool makes the host name change for you.

DOES SAS HAVE A TOOL TO DO THE BACKUPS?

SAS does not have tool that can back up all the above files, but for a subset of the SAS infrastructure files, you can use the SAS Deployment Backup and Recovery Tool (DBRT). The SAS DBRT, new with SAS 9.4, provides an integrated method for backing up your SAS content on a single machine and recovering these files on the same machine (or a new system using the same host name). The tool is installed on the middle tier as part of the SAS® Web Infrastructure Platform. It connects with the SAS® Deployment Agent on each middle-tier and server-tier host machine and backs up the following components:

- The SAS Metadata Server, including all registered metadata repositories, the repository manager, and the server's configuration directory.
- The contents of the Data directories, SASEnvironment directories, and server configuration directories for each server on the SAS server tier. (If symbolic links in these directories point to other locations, the referenced locations are not backed up.)
- The SAS Content Server repository.
- The databases that are managed by the SAS Web Infrastructure Platform Database. By default, all of the databases are backed up. You can modify the backup configuration so that only selected databases are backed up.
- Additional directories under *`SAS-configuration-directory`*`/Levn` as specified by the SAS administrator.

More details about the DBRT can be found in Chapter 29 of the *SAS® 9.4 Intelligence Platform: System Administration Guide, Fourth Edition* available at http://support.sas.com/documentation/cdl/en/bisag/68240/PDF/default/bisag.pdf.

Note: As a reminder, before you start your backups, all the files associated with SAS must be in an unlocked (closed) state, especially the data files associated with our in-memory servers. If you just copy or back up the files without closing them, these files, when restored, will present themselves as corrupt or incomplete.

WHERE THINGS COULD GO WRONG

When doing a full backup, your business is responsible for ensuring that the operating environment of the SAS disaster recovery system is a clone of production ***at all times***. Subtle differences between the SAS disaster recovery system and the SAS production system will cause either the recovery process to fail or the deployment tooling to fail to be able to maintain, administer, and back up the recovered system. If the administration tooling does not function in the disaster recovery environment, you will not be able to recover your users back to a new production system when it becomes available. Examples of subtle differences that will result in failures to recover or execute include:

- deltas in the configuration of the kernel security modules (for example, SELinux or Solaris Trusted Extensions)
- deltas in Discretionary Access Controls (DACs)
- deltas in the way file systems such as tmp are mounted (for example, noexec)
- differences in UIDs/GIDs of any SAS accounts (or user accounts)
- different drive-letter mappings on Windows operating systems
- different operating system patch levels
- different environment variables
- different PATH variables
- different versions or locations of Java
- different versions, configurations, or locations of Java or third-party products and drivers used by SAS or SAS users

When canonical names or aliases are used, it is important to use domain name system (DNS) names, rather than local host names throughout the SAS environment. This includes the SAS® Deployment Wizard, the SAS® Deployment Manager, SAS® Management Console, and any manual edits to configuration files.

Copying only the SAS deployment (the SASHome or SAS configuration directories) from the SAS production system to another system is not supported. In this scenario, SAS Deployment Manager tasks fail to be able to properly manage and maintain the deployment and the SAS® Deployment Agent (necessary to perform future backups and middle-tier clustering).

CUSTOMER EXAMPLES

Let's walk through a couple of customer examples:

SAS Customer 1: A SAS customer has a three-tier SAS infrastructure in a data center in NC. They would like to create a disaster recovery site in a data center in CA. In working with the SAS users, they can be without SAS for one business day in the case of a disaster and are able to work from the full backup of their SAS application and data, which is done every weekend. In this scenario, each Saturday, a full backup of the disaster recovery system for each of the three tiers is created. The full backup includes the operating system, SAS binaries or configuration files, and all SAS data files. These backups are transferred to the data center in CA to be used in the case of an emergency.

SAS Customer 2: This SAS customer is running all their SAS servers and sessions on a single system in a data center in NY. They would like to create a disaster recovery system in a data center in CA. This is for a mission-critical SAS application, so the disaster recovery site will be live in order to support the one-business day SLA for recovery. In this scenario, after the SAS grid infrastructure has been configured, a

full backup of all the SAS production systems are created and applied to the SAS disaster recovery systems in the CA data center. They use a tool called Puppet to keep the operating system and SAS applications in sync (for example, any changes made to the SAS production system are automatically made to the SAS disaster recovery system) between the two data centers.

On a daily basis, the SAS® Migration Utility is run on the SAS production system to create a SAS Migration Utility package of the SAS configuration files. At the same time, because the SAS servers are stopped, they use a storage array tool to make a snapshot of their SAS data and application files. This snapshot and SAS Migration Utility package are moved to the SAS disaster recovery system and applied.

Please note with this SAS customer, most of their persistent data is stored in an external RDBMS that has its own DRP, so not many permanent SAS data files need to be replicated to the target system.

SAS Customer 3: This SAS customer is running a SAS grid infrastructure in a data center in NY. They would like to create a disaster recovery SAS grid infrastructure in a data center in CA. This is for a mission-critical SAS application, so the disaster recovery site will be live in order to support the two-hour SLA for recovery. This customer will follow the same DRP as SAS customer 2. The main difference is that they need to keep SAS Migration Utility packages for each of the different SAS tiers. There is more work because a SAS grid infrastructure has many independent nodes with different SAS installs and purposes.

Understanding what your SAS users need from a DRP and making sure it is executed correctly will give everyone a piece of mind.

CONCLUSION

Create a Disaster Recovery Plan (DRP) for your SAS infrastructure that meets the needs of your SAS users and test it several times each year. The more you understand about your SAS applications, the easier the creation of the DRP will be. Once you have finalized the DRP, please share it with SAS Technical Support so that they can validate that what you plan to do is supported by SAS. During this validation process, any concerns or potential issues will be discussed.

Test this plan multiple times each calendar year. During the test, make sure that SAS not only comes up, but the data that is used by your mission-critical SAS applications is available and current. Also, bring up the SAS Deployment Wizard to make sure it is working and can see all the files it needs so that you will be able to update the SAS deployment if you have to on your restored disaster recovery system.

If you run into any issues, modify the DRP and retest.

Note: As a reminder before you start your backups, all the files associated with SAS must be in an unlocked (closed) state, especially the data files associated with our in-memory servers. If you just copy or back up the files without closing them, these files, when restored, will present themselves as corrupt or incomplete.

REFERENCES

Crevar, Margaret (2016) Ensuring that Your SAS Infrastructure Is Able to Meet Your SAS Users' Demands SAS Global Forum 2016

SAS® 9.4 Intelligence Platform: System Administration Guide, Fourth Edition

"Disaster Recovery Plan" https://en.wikipedia.org/wiki/Disaster_recovery_plan#Planning_methodology

Wold, Geoffrey H. (1997). "Disaster Recovery Planning Process". *Disaster Recovery Journal*. Adapted from Volume 5 #1. Disaster Recovery World. Retrieved 8 August 2012.

"SAS Position Statement Regarding Disaster Recovery" http://support.sas.com/techsup/disaster-recovery-position-statement.html

ACKNOWLEDGMENTS

Thank you to Jawna Gardner and Tony Brown for their help with this paper.

CONTACT INFORMATION

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

Margaret Crevar
SAS Institute Inc
+1 919.531.7095
Margaret.Crevar@sas.com
www.sas.com