Planning for the Worst—SAS® Grid Manager and Disaster Recovery

Glenn Horton and Doug Haigh, SAS Institute Inc.

ABSTRACT

Many companies use geographically dispersed data centers running SAS® Grid Manager to provide 24/7 SAS® processing capability with the thought that if a disaster takes out one of the data centers, another data center can take over the SAS processing. To accomplish this, careful planning must take into consideration hardware, software, and communication infrastructure along with the SAS workload. This paper looks into some of the options available, focusing on using SAS Grid Manager to manage the disaster workload shift.

INTRODUCTION

We will consider several options for Disaster Recovery (DR) strategies. In making the best choice for your organization you must consider several factors such as:

- Cost, including hardware, software and administrative costs

- Complexity, including network and storage infrastructure

- Available technologies, including data replication and virtual machines (VMs)

- Recovery Time Objective (RTO)

- Recovery Point Objective (RPO)

The following sections cover various strategies and compare the strategies in terms of the factors listed above. The strategies we will cover are:

- Active-active SAS Grid Manager instances, with and without VMs

- Active-passive SAS Grid Manager instances, with and without VMs

- A single SAS Grid Manager instance spread across two sites

A common requirement in all of these strategies is data replication between sites. That topic is not covered in this presentation, because it is very specific to the selected storage technology, the vendor used, and the Recovery Point Objective. Also, the topic of licensing implications is not covered in great detail here. Licensing should be discussed with your SAS account team based on the strategies you choose.

Another factor to consider is your Recovery Time Objective (RTO), or the amount of time it takes for the DR site to become fully functional as the various strategies may support different RTOs.

ACTIVE-ACTIVE SAS GRID MANAGER INSTANCES

An active-active configuration includes two SAS Grid Manager instances, which means two IBM Platform LSF (LSF) clusters, that are running in distinct data centers. For example, you may have your production SAS Grid Manager instance running in one data center and your development SAS Grid Manager instance running in another data center. In the event of a disaster at the production site the production SAS Grid Manager instance can be started at the DR site. The simplest SAS deployment strategy requires the DR site to have the capability of bringing up the operating system instances that were running in the failed site.

Two strategies for bringing up the failed OS instances at the DR site are:

- VMs and the associated VM high availability (HA) technology

- network booting (in a configuration that does not use VMs)

These strategies are the most transparent, because whenever production fails over to the DR site, the OS instances are exactly what was running prior to the DR event.

The advantages of this approach are:

- The SAS configuration can be replicated from the failed site to the DR site. No initial configuration is required at the DR site and no manual actions are required to keep the sites synchronized. The configuration can be replicated using a storage replication technology, because the machine names will not change during a DR event.

- Because machine names do not change, no client reconfiguration is required when a disaster takes place or when the DR system is tested. Obviously, this assumes the network can make the failover transparent to clients.

Another approach would be to mount the replicated production file systems on the development OS instances. However, there are several complications that would need to be addressed for this to be an effective strategy. Some of those complications are:

- Both the production and the development machines need to be configured into SAS metadata for features such as SAS Workspace Server load balancing. This isrequired so that the configuration data in the production SAS Metadata Server can used on the development operating systems.

- Both the production and the development machines need to be configured into LSF. This is required so that SAS load balancing will work when using the production LSF on the development operating systems.

- Network accommodations need to be made (for example Virtual IP addresses put in place) to allow clients that typically connect to the production site to connect to the development site in a DR situation.

This active-active strategy might make sense if, for example, it is acceptable to the business for the development environment to be down for a longer time than it is acceptable for the production environment to be down.

ACTIVE-PASSIVE SAS GRID MANAGER INSTANCES

The primary difference between the active-active approach and the active-passive approach is that there are no SAS license costs for the passive site. In the previous section, we made the assumption that it was acceptable for the development site to be unavailable in the case of a failure at the production site. If that is not the case, then a better choice might be to have more than one active site and a dedicated passive DR site. The DR site could be used as the DR site for either the production site or for the development site. This removes the limitation of having to operate for some period of time at reduced capacity. One obvious drawback to this approach is storage cost, because both data and SAS installation and configuration information need to be replicated from both the production site and the development site to the DR site. Another drawback is the requirement for three physical sites, because we need to plan for a disaster at either the production or development site.

A SINGLE SAS GRID MANAGER INSTANCE SPREAD ACROSS TWO SITES

You likely noticed that SAS Grid Manager isn't being used to help meet any of the DR requirements in the first two scenarios. That's going to change as we consider a single SAS Grid Manager instance that is spread across two physical sites.

In this approach, although we have a single SAS Grid Manager instance across two physical sites (and therefore a single LSF instance across two sites), the two sites are used for different purposes. For example, one of the sites could be used for production workload and the other site used for development workload. In a configuration like this, the SAS data for a given site is typically local to that site for performance reasons. However, the production data must be replicated to the development site to support a case such as failover of the production system to the development site. Other considerations that are required for this HA configuration but are atypical for a single-site SAS Grid Manager environment include:

- Although a single LSF instance is used, each site has its own identical copy of the LSF installation and configuration. This is required because if the site hosting the LSF file system goes down, that failure does not take down the other site due to the LSF directory being unavailable. This consideration implies that when a LSF configuration change is made, that change must be replicated in the mirrored directory at the alternate site.

- IBM Platform Enterprise Grid Orchestrator (EGO) services for a given environment must be defined so that they fail over from their primary machine to other machines in their primary site, and to the DR site only if all of the primary site is down. This prevents services from failing over to the DR site in situations such as taking a machine down for maintenance in the primary site, or a single machine failure in the primary site. These services include components such as SAS Metadata Server, SAS Middle Tier and Platform Process Manager.

- An EGO service could be created so that when it fails over to the DR site, it mounts the replicated file systems that are required by the SAS environment that is failing over.

- Network components, such as Virtual IP Address switches, must be capable of handling failover from the primary to the DR site. This allows clients to find SAS components such as SAS Metadata Server and SAS Middle Tier without requiring the clients to be reconfigured.

CONCLUSION

You need to consider many criteria when you create a DR plan,including:

- What is your RTO?

- Will you use an active-active or an active-passive configuration?

- If active-active, will there be a single SAS Grid Manager instance or multiple SAS Grid Manager instances?

- If active-passive, will you use VMs, network boot or some manual approach to getting the system running at the DR site?

- Which systems need a DR plan?

- What is your RPO?

- What is your data replication strategy?

- What is your test plan and test frequency?

- What are your IT standards?

Each of these decisions can have implications for cost, ease of implementation and ease of recovery in the case of a disaster. However, not making a decision can be more costly. Of businesses that experience a disaster and do not have a DR plan, 80% are out of business in one year or less. Can you afford to wait?

REFERENCES

Spectra. 2011. "Three Key Requirements of a Sound Disaster Recovery Strategy." Availablehttps://www.spectralogic.com/index.cfm?fuseaction=home.displayFile&DocID=3664. Accessed January 12, 2015.

HotLink. 2014. "Easy VMware Disaster Recovery & Business Continuity in Amazon Web Services" Available http://www.hotlink.com/resources/DR_EXP_WP01.pdf . Accessed February 8, 2015.

CONTACT INFORMATION

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

Glenn Horton
100 SAS Campus Drive
Cary NC 27513
SAS Institute Inc.
Glenn.Horton@sas.com
http://www.sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.