Preparing Your Computer for SAS® Grid

Margaret Crevar, SAS Institute Inc., Cary, NC, USA

Tony Brown, SAS Institute Inc, Dallas, TX USA

ABSTRACT

Have you ever wondered if you are really prepared to start the installation process of SAS software on your hardware? Perhaps you have read the System Requirements sheets for your SAS release and version, and the appropriate SAS platform administration guide for your operating system. Are there other crucial items to consider before the installation process that are targeted toward your company's expected performance of SAS?

INTRODUCTION

When you are configuring and tuning your computer hardware resources to best support your SAS users, it helps to understand how SAS will be used at your site. This paper offers guidelines about important aspects of hardware architecture, configuration, and setup in order to best support SAS application usage. A clear and measured understanding of how SAS will be used at your site is required to leverage these guidelines effectively. The more information you can access prior to the installation of SAS software, the fewer things you might have to change as SAS usage increases at your site.

Preparing your computer hardware systems <u>prior</u> to the installation of SAS software should include the following tasks:

- Conduct a technical assessment.
- Estimate the required system resources.
- Perform a review of software requirements (both SAS and third-party).
- Plan the storage and file-system layout.
- Develop an on-going monitoring process.

For the purpose of this paper we will be talking about a production system to support your SAS users. However, the guidelines mentioned are all applicable to your development and testing systems. The only additional issue that needs to be discussed is how to move SAS applications or data between these systems.

TECHNICAL ASSESSMENT

Performing a technical assessment of how SAS is (or will be) used at your site is a critical first step toward understanding the resource requirements. The overall assessment needs to involve someone who understands the business goals and tasks; however, the technical assessment focuses more on metrics such as how many people are doing what tasks with how much data simultaneously.

Before establishing a new environment to support SAS, it is imperative to understand how SAS will be used within the organization. A good first step is to work with your SAS account team to have a sizing analysis guideline done by the SAS Enterprise Excellence Center (EEC). This will help you get a good understanding of the SAS workload. Not having a complete understanding of how SAS will be used can make this difficult. A typical SAS workload involves several aspects, including fixed (ETL and standard reporting) and ad hoc workloads (usually ad hoc analysis, studies and reporting). There are many components to understanding the resources needed to support your SAS workload. A few of these components are listed below:

- number of concurrent SAS sessions

- major SAS tasks to be used

- data volumes (number and size of SAS data files)

- data origination points (local SAS data files, remote RDBMS, external flat files, networked file systems, and so on.)

- service level run-times required or expected for individual jobs

- individual SAS job sessions (for example, batch programs or individual interactive jobs) versus pooled SAS server sessions

- required safety, security, performance, and authentication protocols (for example, encryption, government compliance, confidential data, data availability and recovery, and so on.)

- disaster recovery strategies

- backup strategies

It is critical to know what demands and expectations are being placed on a computer system to support the tasks of provisioning it physically and logically for SAS usage. But it is also important to know how mission critical these SAS applications are. Is it going to impact your company's business if the computer system is down for a day or an hour? If so, then good disaster recovery and backup strategies need to be put into place during the design of the computer system infrastructure. Information about these two concepts can be found in the "Considerations for Implementing a Highly Available or Disaster Recovery Environment" and "Backing Up SAS Content in Your SAS9 Enterprise Intelligence Platform"—papers that are listed in the reference section of this paper.

How important is the security of the SAS applications and data? Below is a detailed example of all the various SAS components for a SAS Intelligence Platform infrastructure from a security perspective.
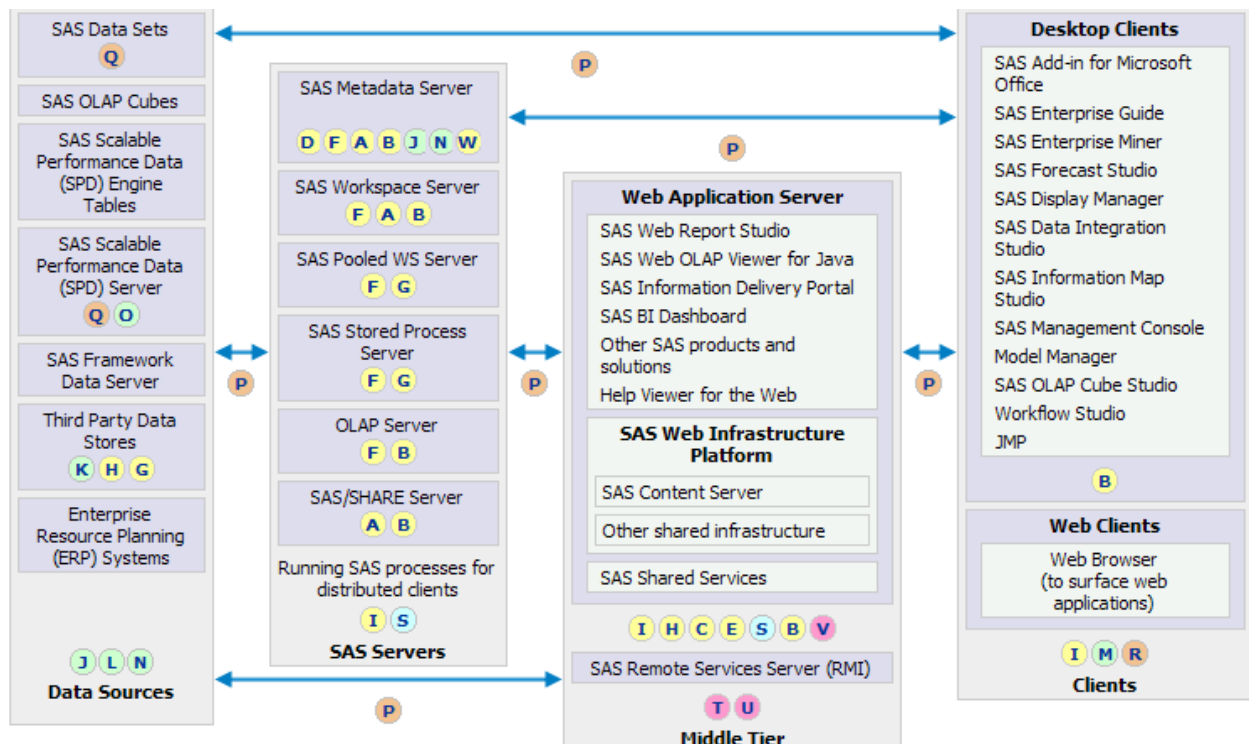
Authentication
- (A) Host Authentication
- (B) Integrated Windows Auth. (IWA)
- (C) JAAS
- (D) LDAP Authentication
- (E) Reverse Proxy Servers
- (F) SAS Token Authentication
- (G) Shared Credentials
- (H) Single Sign-On
- (I) Web Authentication
- (W) SAS Internal Authentication

Authorization
- (J) Metadata Authorization
- (K) Third Party Authorization
- (L) OS Authorization
- (M) Role-Based Security
- (N) Row/Member-Level Security
- (O) SPDS Authorization

Encryption
- (P) Encryption (in-transit)
- (Q) Encryption (at-rest)
- (R) Encrypted ODS PDFs

Web Applications
- (T) Web App Security
- (U) Portal Integration
- (V) Java 2 Security

Security Reporting
- (S) Security Reporting

Additional Security Reporting Topics
- Auditing
- Logging
- Audit and Performance Measurement Package
- Glossary of security terms

Additional Authentication Topics
- Pluggable Authentication Modules (PAM)
- Glossary of authentication related terms

Additional Authorization Topics
- Glossary of authorization terms

Additional Encryption Topics
- Secure Shell (SSH)
- Secure Socket Layer (SSL)
- SFTP (Secure File Transfer Protocol)
- Federal Information Processing Standard (FIPS) 140-2

Additional Security Information

Figure 1: Conceptual Overview -- Security and the SAS Architecture

Going into the details of security is beyond the scope of this paper. Please ask your SAS account team for a copy of the "Security within the SAS Architecture – A Conceptual Overview" paper that accompanies the above diagram.

SAS TASKS

For purposes of this document we will address two common categories of SAS usage. Batch tasks tend to run at off hours, and prime time or information consumer tasks run during business hours. Let's talk about the characteristics of each task.

Batch tasks tend to be a long-running job or jobs that run during off hours. These can consist of Extract Transform and Load (ETL) processes, post-ETL generation of standardized analysis and updates to data marts or cubes. These sessions typically execute nightly or over a weekend during a limited window of time. This batch workload can run off-hours in batch (as is common with many ETL and scheduled-reporting tasks) or during the day to meet business and update requirements. Batch workloads are things that are run on a very regular basis, and their workload resource demand is fairly well known and predictable. BATCH sessions tend to be extremely I/O intensive as they update and populate new and existing data stores. For large batch workloads that have many concurrent job streams, significant CPU, memory, and I/O resources can be required, often to the limits of a given server. Batch reporting workloads are generally less I/O intense, but again, large parallel work streams can represent a significant resource demand. The batch workloads are often run as part of a daily schedule, and their resource demand can shift as data, processes, or concurrent resource competition increase. This workload set tends to be easier to monitor, measure, and predict for resource planning.

The information consumer tasks range from basic data reporting and queries to data analysis and mining to interactive SAS solutions that are run during the day by the SAS users who are exploiting the data created by the off-hours tasks or pulling data from your centralized data warehouse.

- Basic data reporting and queries tend to be the most common, generally coming from using the SAS® Enterprise Guide® (EG) client to access the source data, analyze it, and then produce reports. SAS EG is a very powerful tool, but it can also cause lots of data movement to happen on your computer system.

- Data analysis and mining jobs tend to be legacy SAS jobs created to do tasks. It is best to run these during off peak hours of the day. Many times these are long processes that run for hours doing lots of different tasks. They also tend to be very I/O in nature.

- SAS solutions are SAS tasks bundled in a manner to do a particular type of reporting or industry analysis. Most SAS solutions have a specialty interface that enables the end users to submit sessions against the backend SAS compute servers. The processing done on these backend SAS compute servers tends to mimic ad hoc reports. Almost all SAS solutions require that the SAS Enterprise Business Intelligence infrastructure is in place or it is part of the Solution. The SAS EBI infrastructure includes the SAS Metadata

server, the mid-tier server (which includes the necessary Web applications server), and the back-end SAS compute servers.

The primary information to capture with the above tasks is how many concurrent SAS sessions (that is, SAS servers or individual SAS sessions) will be running at any given time, and what their I/O, CPU, and memory demand requirements are.

In addition to the two list of SAS tasks above, understanding the expected or required response time of each of these tasks is critical. Knowing all of this will help determine how to provision your computer system hardware.

NUMBER OF SAS SESSIONS

As discussed above, understanding the number of concurrent SAS sessions that will be running at any given time is essential. Will the number be constant every day, or will there be peaks during specific times of the week, month, or year? Are there weeks during a year (holiday shopping day) that will require more SAS sessions to be active or the volume of data to increase significantly? If so, how will the additional workload be accommodated? Will you have underutilized resources during non-peak periods, or will you look to add computer resources during these peak times, or will you limit who has access on the computer to only crucial business processing during these peak times.

What is the anticipated growth over the next six months to a year? How do you plan to accommodate for his growth? Historically, IT departments would "scale up" by bringing in a larger SMP computer system with more cores; however, this is becoming a very expensive strategy. The current trend is to "scale out" by setting up multiple smaller commodity computers (typically with four to eight cores) and then increase the number of these computers as the SAS usage demands. SAS supports this trend, but there are additional SAS components that need to be put in place at the start if you plan to go from a single back-end compute server to a cluster of back-end compute servers. Establishing the decision to scale up or scale out is an important factor in the technical assessment.

DATA VOLUMES

Understanding how much data will be used, where the data persists, and where it will be processed affects the amount of storage, the IO throughput, and network capacity needed. What is the source of the data for the SAS tasks listed above? Is it existing SAS data files, is it a SAS data mart created from a corporate SAS warehouse, or is it a data warehouse in an external RDBMS? If it is the external RDBMS, how often will the data be pulled from it for the analysis and how often will reporting be done by the SAS applications?

What is the overall size of the SAS data files that will be stored on the computer system? Will they be stored in a single directory on the system or in multiple directories?

What is the anticipated data growth over the next six months to a year? Will you need to keep all historical data online at all times, or will you be able to archive a month off as a new month rolls on? Again, this information needs to factor in to the overall size of the SAS data files that need to be stored.

Since I/O throughput is crucial to the success of SAS, and it represents the largest contributor to inadequate performance, understanding the data volumes that will be accessed and when the heaviest SAS workload will be happening is crucial. With new SAS deployments, it can be very difficult to get a realistic estimate of how much I/O throughput is needed. But one needs to take the time to understand the size of the SAS data files used by each SAS session and how often will it be touched during a SAS task. Touching a SAS file can mean any of the following:

- reading it for data processing
- writing out a new file after doing the data processing
- sorting or doing analysis on the file (which creates extra copies of most files "behind the scenes")
- merging multiple files together to create a new file

Knowing this process flow for each SAS session running and the desired response time for the SAS tasks will help you determine the I/O throughput rate required for SAS on your computer system. Please remember that there will be multiple SAS sessions running simultaneously, so you will need to add up all the I/O throughput demand for each SAS session to get the overall I/O throughput needs for each operating system instance.

There is more information that we need to discuss here regarding the number of file systems needed by SAS, but let's first discuss the number of distinct operating system instances that will be needed to support the different components of the SAS Intelligence Platform infrastructure.

ESTIMATING THE REQUIRED RESOURCES

Resource decisions include:

- How many (real/physical) computer systems (and now many cores/CPUs per system) do I need?

- Should I use clustering or a grid or a large SMP machine to support the SAS workload?
- What type of file system and data sharing options should I consider?
- Should I use virtualization?

NUMBER OF DISTINCT OPERATING SYSTEM INSTANCES REQUIRED

With the new SAS Intelligence Platform applications, the minimum SAS footprint is comprised of three distinct operating system instances, one for each of the required SAS and WAS servers (SAS Metadata Server, Web Applications Server, and SAS compute servers. Because each type of server has different computer resource needs, SAS strongly recommends that each of these server types be placed on their own instance of an operating system (referred to as system for the rest of this paper).

1) SAS Metadata Server System – needs to provide at least two cores and 12GB of physical RAM to support this system, and more cores and RAM if the SAS applications will be doing numerous and frequent updates to the underlying SAS Metadata Repository. Please note that this repository is stored in memory for faster access times. There will be a delay at the initial start of the SAS Metadata server as the repository is instantiated from disk to memory.

2) Mid-Tier Server System – needs to provide at least two cores and 12GB of physical RAM to support the Web Application Server (WAS) that will be used by the SAS applications. Please note that WAS tend to be very memory intensive, which is why we state the 6GB of physical RAM per core requirement.

3) SAS Compute Server System – There needs to be at least one system of at least four cores and 16GB of physical RAM in place to support the back-end SAS compute servers and miscellaneous SAS sessions that will be run to support all the SAS applications. As the number of concurrent SAS sessions grows to support the number of SAS users, and thereby the total I/O throughput demand, one has to choose either a single large SMP computer system - or to branch out into a grid or clustered infrastructure. Either method is supported and has its pros and cons.

MULTIPLE GRID NODES VERSUS LARGE SMP SYSTEM

As the need to support additional SAS compute servers arises, a choice needs to made between a single large SMP system with 16+ cores and several commodity computer systems with four to eight cores each. The total cost of ownership needs to be considered. A single instance of an operating system is much easier to support, but the current cost of large SMP computer systems is very high compared to smaller commodity systems. Multiple instances of an operating system spread over the various nodes within your grid infrastructure enable you to use less expensive hardware and gives you the advantage of higher availability and reductions of single points of failure. So we are seeing a growing trend of using a grid of multiple commodity systems to support all the back-end SAS compute servers.
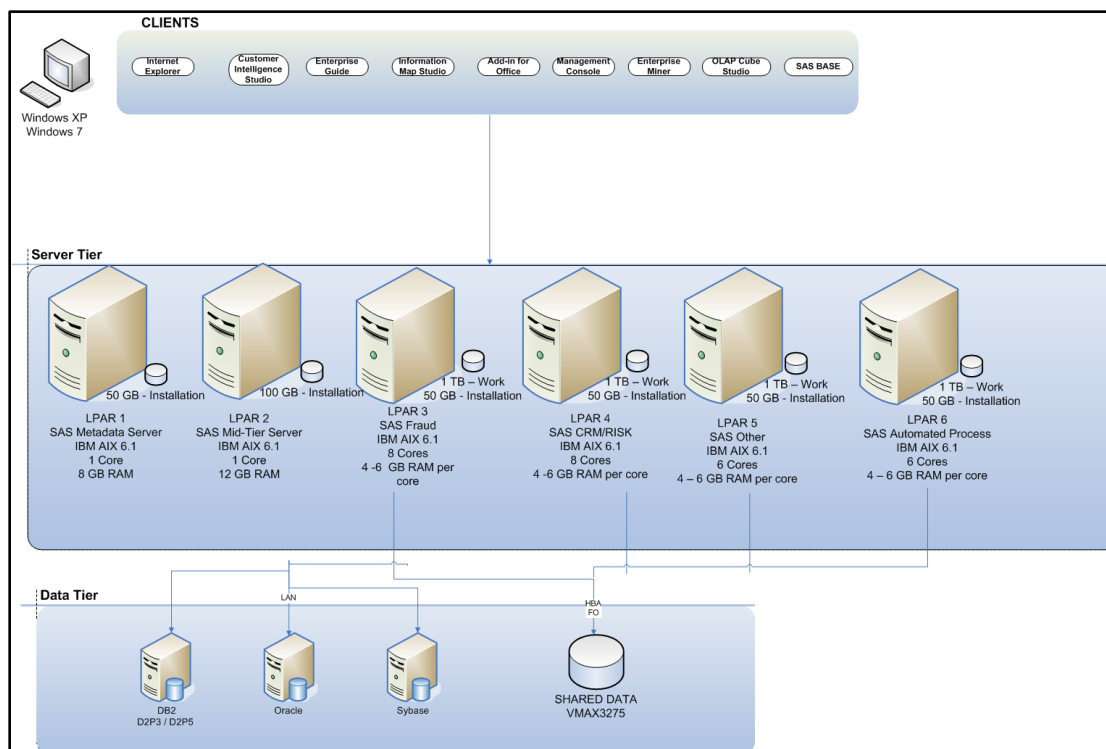
Figure 2. SAS Intelligence Platform Infrastructure Example

Once the decision has been made to implement a grid of SAS compute servers, the next step is to determine an algorithm to distribute the workload across all the systems. SAS offers several options to do this distribution. One approach is to automate the distribution of SAS sessions to different SAS compute servers using round-robin scheduling such as the SAS Workspace Servers Load Balancing routines or using the more sophisticated SAS Grid Manager. The pros and cons of the various methods are beyond the scope of this paper.

Another way to distribute SAS session workloads across different SAS compute servers is to divide the workload by different departments and/or SAS applications. This approach is sometimes used to enhance security and separate the data accessed by different groups.

If your environment uses multiple SAS compute servers, then you need to review your strategy for sharing data. Since adding compute servers is one approach to handling growth, it is helpful to consider your data sharing strategy even if your initial implementation will use a single SAS compute server. If a lot of data needs to be shared in just a READ fashion, then you might be able to just use NFS mounts between the nodes. However, if a lot of data needs to be shared with read and write access, then you will need to use a shared or clustered file system across all the SAS compute servers. Please note that shared or clustered file systems generally require an additional license and are not bundled with an operating system, so this is an extra expense that needs to be placed in the total cost of ownership category.

PHYSICAL VERSUS VIRTUAL

Along with the use of smaller commodity systems is the trend to use these systems in a shared environment with a virtualization tool such as VMware. The topic of running SAS on physical versus virtual system is coming up regularly in discussions with SAS customers. Please note that SAS is supported in both scenarios, and as long as you have the physical computer resources present at peak times, you should have no issues running SAS on a virtual system.

We often find that most IT shops tend to limit the size of virtual systems to just a handful of cores with no more than 2GB of physical RAM per core. As you can see from the above computer resource recommendations, this is far below what SAS needs for any of the tiers associated with a SAS Intelligence Platform environment.

Other things to consider before deciding on going the virtual route are what applications will be run in the other virtual partitions within your computer system and what (if any) computer resources (including NIC cards and HBA switches) will be shared. Please note that applications that perform only small, random I/O accesses do not "share" computer resources nicely with SAS applications that perform large sequential I/O accesses.

Lastly, from a performance perspective, we are seeing that customers who run in virtual partitions tend to get the best performance if their storage arrays are attached locally to these partitions. Inconsistent performance issues are more prevalent when customers are using virtual storage arrays, or storage arrays connected via a virtual hypervisor.

Now that we have gone over some common ways of setting up the hardware to support your SAS Intelligence Platform environment, let's go over the various requirements to support SAS.

REVIEW SOFTWARE REQUIREMENTS

Before the actual installation of SAS software is downloaded from the SAS Software Depot, review the operating system requirements, third-party software requirements, and operating system tuning guidelines that are available on the support.sas.com Web site.

OPERATING SYSTEM REQUIREMENTS AND TUNING GUIDANCE

The Install Center portion on the support.sas.com Web site provides a list of the operating system requirements for supported versions of SAS. These can be found at http://support.sas.com/resources/sysreq/index.html. These documents describe the operating system release level and the amount of disk space required to install the various SAS products.

The document also describes the minimum memory required for a single SAS instance to come up to a prompt. It is not the total amount of memory that a SAS instance might use (that is, memory for the SAS application and operating system file cache to support the SAS data files). As always, you will need to remember that there will be multiple SAS instances running simultaneously at a time.

In addition to these system requirement sheets, SAS continuously works with the hardware partners to review and recommend operating system tuning guidelines. These papers can be found on the various external Web sites as well:

- AIX Tuning Papers (AIX5, AIX6, AIX7): http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101529
- Windows 2008 Tuning Paper: http://support.sas.com/resources/papers/WindowsServer2008ConfigurationandTuning.pdf
- RHEL Tuning Papers (RHEL5, RHEL6): http://support.sas.com/resources/papers/tnote/tnote_performance.html
- Solaris Paper: http://www.oracle.com/technetwork/database/focus-areas/bi-datawarehousing/sas/sas9-on-solaris10-superheroes-paper-511997.pdf
- Moving SAS Applications from a Physical to a Virtual VMware Environment: http://support.sas.com/resources/papers/MovingVirtuaVMware.pdf

THIRD-PARTY SOFTWARE REQUIREMENTS

For the SAS Intelligence Platform, there are several third-party software products needed depending on the SAS application. These dependencies are listed at http://support.sas.com/resources/thirdpartysupport/index.html. In most cases, you will need to ensure that the third-party software is installed before starting your SAS software installation or update.

FILE SYSTEM LAYOUT GUIDELINES

Providing guidelines that describe how to layout your file system is one of the more difficult areas to address because there are so many dependencies to consider. Factors include the following: how SAS will be used at the customer location, what type of storage media (disks, array) will be used, whether you are running on a single computer system or multiple ones, data and throughput volumes, and many other issues. We have, however, put together some general guidelines to help your storage administrator with this task.

To begin with, you need to determine how many separate file systems will be needed to support your SAS applications and how big those file systems need to be. At a minimum you should have one file system for your permanent SAS data files and another one for SAS WORK (the temporary files stored in SAS WORK are created during a SAS session and then deleted when SAS terminates). If you are sorting large data files within SAS on a system that will be used heavily, then you should have a third file system to store the utility files created by the SORT procedure. Again, many SAS solutions require multiple file systems to store their permanent SAS data files, so understanding what is needed by your SAS users is a must.

The next step becomes to determine the I/O throughput required for each of these file systems along with the RAID level that will best support them. In general, we ask that each file system is able to sustain at least 100 MB/sec I/O throughput; more if you will be working with very large data files (over 100 GB in size) and have over 10 concurrent

SAS sessions running.

STORAGE ARRAY SETUP GUIDELINES

The difficult part now becomes how to achieve the needed throughput and redundancy with the storage array your company currently has in place or that is within budget. As our friends in the storage array business say, "Fast, Cheap or Reliable: choose two." Reliable storage is a given, so you now have to decide between fast storage arrays or cheap ones.

In many cases, to achieve the required I/O throughput for your SAS applications you will end up with a lot of extra disk space per file system since you will need to have a minimum of seven physical disk drives in a RAID5 in order to achieve the 100 MB/sec I/O throughput minimum. If these physical disk drives are all 500GB in size, you will have approximately 3 TB of disk space available to you.

Other things to mention here are the use of locally attached (or internal) disks or a classic SAN or one of the new virtual SANs. Locally attached drives are great for the SAS WORK area, but it can be hard to attach enough physical disk drives to support the I/O throughput rate required. The use of SANs is becoming the de facto standard with most SAS customers. Classic SANs setup used LUNS (groups of physical disk drives) striped onto separated disk ranks, into the volumes (referred to as file systems above) needed by SAS. In this fashion the physical disk drives were often dedicated to the LUNs comprising a single file system and not shared by other SAS file systems or applications.

The new trend in the SAN marketplace is a virtual SAN array. In a virtual storage system, most or all of the physical hard drives are striped together and then all of the LUNs (or volumes) are partitioned across this stripe of many disks. This is often referred to as a "striped everything system". It can result in very fast file systems, but can also cause issues if you have two applications doing completely different types of I/O to the same set of disk drives competing with each other under a heavy load.

Other things to consider about the setup of your storage array are the types of physical disk drives that are being used within storage arrays currently. The following list explains the different types of physical drives that are available and how SAS works with them.

- Serial Advanced Technology Attachment (SATA) drives: SATA drives are quite popular because they have a large footprint (for example, 500 GB, 1 TB, or 2 TB per device) and they are relatively inexpensive. However, their slower spin rate and higher seek time typically do not make them a good fit for heavy SAS workloads, especially for the SAS WORK area.
- Serial Attached SCSI (SAS) drives: SAS drives are faster and tend to be a better fit for SAS applications. This is especially true when budget or architecture limits the number of disks – such as internal in a chassis, or directly attached or low-end storage arrays.
- Dynamic RAM (DRAM) devices: These devices store data in memory and yield as fast as nanosecond-range performance for data access and writing, which works extremely well with current CPU speeds. SAS has had good field performance experience with these devices, but their cost can be hard to justify.
- Solid-State drives (SSD, also referred to as flash drives): Solid-state drives are popular as well. Technology with these has improved since they were first introduced and SAS has seen very good read and random-write rates to these devices. The changes to these drives make them very attractive in terms of very high performance for SAS WORK file systems.
- Network Storage Appliances: These devices are self-contained storage arrays that come preconfigured with a considerable amount of physical disk space. Our experience in working with customers has shown that these appliances are a good fit for smaller, very static application and data profiles. Generally, they do not perform as well with the high volume of I/O of many concurrent (10+) SAS users that demands fast I/O with large volumes of data (files of 25 GB or more) with shifting workloads. Be aware that you must have enough network bandwidth to support these arrays.

More information about how to setup storage arrays for SAS can be found in the two papers on storage listed in the Recommended Reading section of this paper.

MONITOR, MONITOR, MONITOR

It is almost guaranteed that the use of SAS will grow at a customer site over time. The reasons behind this growth can any of the following: more concurrent SAS sessions running on the computer; the size of the data being analyzed or reported getting larger; the types of SAS applications increasing over time. This is why the guidelines for setting up your hardware to run SAS are only guidelines, not detailed instructions. No two SAS sites are the same, and we anticipate that your usage will change over time; therefore, we highly recommend that you monitor your hardware on a regular basis to ensure you do not run out of a computer resource and, as a result, cause SAS sessions to perform poorly or terminate unexpectedly.

There have been several papers written that discuss ways to monitor the hardware:

- "Ensuring You Have the Proper Resources for your SAS®9 Applications" (Crevar, 2007) discusses what computer resources need to be monitored for various types of SAS applications.
- A methodology for solving performance problems with your SAS jobs or applications is documented in the SAS white paper "A Practical Approach to Solving Performance Problems with the SAS System: (Brown, 2001).
- A follow-up paper entitled "Solving SAS Performance Problems: Employing Host-Based Tools" (Brown, 2006) goes into more details about how you can solve performance problems by using standard monitoring tools that are shipped with most of the commonly used operating environments.

If you want to monitor specific usage for a few minutes at a time, there are some very simple tools that enable snapshot monitoring included with most operating system. On UNIX, use the commands: `top`, `prstat`, or `topas`; on Windows, use TaskManager, PerfMon, or Process Monitor from Microsoft Sysinternals. There are also some third-party tools that provide insight into system and application behavior (for example, HP OpenView GlancePlus and Solaris Resource Management from Sun Microsystems). These tools have very nice graphical user interfaces, and they have the capability to define "rules" that make the task of monitoring the hardware fairly easy and repeatable. Usually, these tools run interactively and do not produce log files.

If you want to monitor specific usage or resources for longer periods of time, the paper "Solving SAS Performance Problems: Employing Host-Based Tools" (Brown 2006), describes the tools that you can use. In the paper you will find information about setting parameters for various tools, such as the Windows Performance Monitor (PerfMon) and the UNIX tools that use the commands: `sar`, `iostat`, and `vmstat`, to monitor system resources as well as interpret the information you have collected. There are sample scripts that can be run where you pass in the collection interval and how many collections you would like, and this information is written out to several log files. The log files are interpreted by manually scrolling through the lines of data. IBM has a nice tool for AIX called *nmon* that takes all the standard UNIX monitor data and puts it into high-quality graphs.

CONTINUOUS HARDWARE MONITORING

If you want to monitor your hardware continuously to make sure that you do not run out of a computer resource or that there are no bottlenecks affecting an important computer resource, third-party monitoring tools will make this task easier.

SAS provides a SAS Audit and Performance Measurement (SAS APM) package to help with managing and monitoring SAS®9 EBI environments. The SAS APM package helps administrators collect and analyze data about the SAS behavior, and to route selected information to infrastructure management and monitoring tools. Features include the following:

- Packaged suite of scripts, tools, monitors, and documentation that describes the installation and integration components for SAS 9 EBI
- Facilities to start, stop, pause, and restart each of the key SAS servers: SAS Metadata Server, SAS Stored Process Server, SAS OLAP Server, SAS Object Spawner
- Availability to manage facilities to verify the status of key servers and generate events as necessary for server status
- Ability to monitor key SAS server metrics and generate events into the infrastructure management environment— metrics such as processor utilization, memory allocation, and storage utilization
- Ability to analyze SAS server logs for key actions and generate events into the infrastructure management environment—actions such as server initialization, user connection events, user ID authentication, and server error events

For more details about this package, please reference to SAS APM Web site http://support.sas.com/rnd/emi/EbiApm92/index.html.

CONCLUSION

It is strongly recommended that a detailed technical assessment regarding how SAS will function, the volumes of data that will be processed, analyzed, or manipulated, and the concurrent number of SAS sessions running is done before you start to set up your computer system infrastructure and install SAS software to support the planned SAS applications. This assessment (which will take time) will help determine the number of file systems required, the I/O throughput needed for each file system, and the number operating systems instances needed to support the multiple SAS tiers. It will also contribute to your plan for SAS growth (in both data volume and SAS usage). Taking the time

to perform this assessment at the start of the project will help you avoid re-architecting your system if you run into unacceptable performance of SAS.

REFERENCES

Hatcher, Diane. 2011. "Considerations for Implementing a Highly Available or Disaster Recovery Environment" Proceedings of the SAS Global Forum 2011 Conference. Available at http://support.sas.com/resources/papers/proceedings11/358-2011.pdf.

SAS Institute Inc. SAS Institute white paper. "Backing Up SAS Content in Your SAS®9 Enterprise Intelligence Platform". Cary, NC. Available at http://support.sas.com/resources/papers/contentbackup.pdf.

ACKNOWLEDGMENTS

Our thanks to Donna Bennett, Simon Williams, Tom Keefer, and Clarke Thacher from SAS Institute Inc. for their review of this paper.

RECOMMENDED READING

- How to Maintain Happy SAS Users: http://support.sas.com/resources/papers/proceedings09/310-2009.pdf
- Best Practices for Configuring IO for SAS Applications: http://support.sas.com/rnd/papers/sgf07/sgf2007-iosubsystem.pdf
- Frequently Asked Questions Regarding Storage Configurations: http://support.sas.com/resources/papers/proceedings10/FAQforStorageConfiguration.pdf
- Best Practices for Data Sharing in a Grid Distributed SAS Environment http://www.sas.com/rnd/scalability/grid/Shared_FileSystem_GRID.pdf

CONTACT INFORMATION

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

Name:	Margaret Crevar
Enterprise:	SAS Institute Inc.
Address:	SAS Campus Drive, Room R-2427
City, State ZIP:	Cary, NC 27531
Work Phone:	919.531.7095
Fax:	919.677.4444
E-mail:	Margaret.Crevar@sas.com
Web:	www.sas.com

Name:	Tony Brown
Enterprise:	SAS Institute Inc.
Address:	15455 Dallas Parkway, Suite 1300
City, State ZIP:	Addison, TX 75001
Work Phone:	214.977.3916 ext=52155
Fax:	919.677.4444
E-mail:	Tony.Brown@sas.com
Web:	www.sas.com