Paper 3497-2019

A Collection of Pre-installation Resource Documentation for SAS® 9.4

Jim Kuell, SAS Institute Inc.

ABSTRACT

Many requirements and guidelines need to be taken into consideration when you are setting up the compute infrastructure for SAS® 9.4. Tracking down these requirements can be a daunting task because they are located across several discipline areas of the support.sas.com website. This paper serves as a collection of the pre-installation requirements and guidelines, and presents a general overview of the topics as well as links to additional documentation.

INTRODUCTION

SAS has become an integral part of day-to-day operations for many organizations worldwide. It provides powerful software that helps transform a world of data into a world of intelligence. But, to achieve this, the software must be coupled with a well-performing infrastructure.

Before implementing an infrastructure, you must fully understand every aspect of the environment and its expectations: specific SAS applications, number of users, data sizes, anticipated growth, and so on. You can then architect and build an infrastructure that meets all these requirements.

There are many components that make up an infrastructure and choosing from all the options can be a daunting task. This paper provides requirements and guidelines for this process, as well links to additional resources.

GUIDELINES

The following sections provide a general overview of SAS 9.4 pre-installation requirements and guidelines. Each section also includes links to additional information that can be used during the architecture planning and deployment phases for both new installs and hardware upgrades.

A question that often arises is whether it is better to have a single compute node with many cores, or multiple compute nodes with smaller core counts. This answer depends on the expected usage of the environment: the number of concurrent users, their usage patterns, data sizes, high availability requirements, and so on. Each approach has its pros and cons, so they must carefully be considered.

VIRTUALIZATION

Virtualization is a very common practice in SAS shops around the world. Virtualized environments can function very well with minimal overhead if they are configured and provisioned correctly. SAS does not work well when it's located on a thinly provisioned virtualized server farm with shared everything (cores, memory, and storage arrays), so it is important to provide it with dedicated hardware resources. Analysis must be done with the specific application or applications to determine the appropriate sizing and configuration of the infrastructure.

For more information about this deployment methodology, see SAS Note 62240.

For more information about SAS' virtualization support policy, see SAS Product Support for Virtualization Environments.

PUBLIC CLOUD

Many SAS customers are deciding to move SAS deployments from on-premises data centers to a public cloud infrastructure. As with setting up any infrastructure for SAS, it is extremely important to plan and architect cloud infrastructures to ensure that they can handle the application and user usage requirements.

The same hyper-thread/SMT thread and dedicated resource guidelines listed in the Virtualization section above also apply to the public cloud. This is especially true in regard to securing dedicated network interfaces for SAS.

For more information about SAS in the public cloud, see SAS Note 62239.

OPERATING SYSTEM TUNING

Operating system tuning recommendations vary depending on the specific operating system and hardware configurations chosen. These guidelines can be found in SAS Note 53873.

CPU

Each new generation of CPUs generally includes significant performance improvements over the previous generation. Because of this, we always recommend using the most recent generation of processors possible.

The number of cores that are needed for each system largely depends on what applications it will be hosting and their expected usage. Only specific SAS applications and functionality can take advantage of hyper-threading, so this must be taken into consideration when deciding on the CPU arrangement or arrangements.

A standard recommendation of ours is to avoid processors with more than two sockets. Without careful planning and configuration, processors with more than two sockets can run into performance issues caused by NUMA (non-uniform memory access).

Another recommendation is to disable hyper-threads in x86 systems. While other tasks on the system might benefit from these threads, the opposite typically happens with heavyweight SAS tasks. SAS runs primarily on physical cores, and enabling hyper-threads can result in significant performance overhead. There is only one floating-point unit (FPU) on a processor core, and hyper-threads must wait for it to be free on context switching. Because this can be a long time in SAS, disabling these threads avoids the blockage with floating points.

MEMORY

Like all other aspects of your hardware infrastructure, the amount of memory that is needed for each system depends on the specific SAS solutions and usage requirements. Here is the *minimum* amount of RAM that SAS recommends for each tier of the environment:

- SAS compute tier: A minimum of 8 GB RAM per physical CPU core.
- SAS middle tier: A minimum of 24 GB **or** 8 GB RAM per physical CPU core--whichever is larger.
- SAS metadata tier: A minimum of 8 GB RAM per physical CPU core.

The amount of virtual memory in a system can greatly affect the performance of SAS. Virtual memory is the amount of RAM plus the amount of swap space (where pages are stored when they are inactive--typically located on disk). SAS recommends that virtual

memory be 1.5 to 2 times the amount of physical RAM in the system.

If normal activity on the system results in a high paging rate, you will likely need to either increase the amount of physical RAM in the system or upgrade the I/O infrastructure to provide a higher bandwidth to the swap space. One or both of these might be necessary.

I/O INFRASTRUCTURE

One of the most common, but most easily preventable, causes of performance issues that are seen by SAS Technical Support is insufficient I/O infrastructure bandwidth. SAS applications are customarily very I/O-intensive, often handling extremely large amounts of data. The necessary I/O throughput rates vary greatly from SAS shop to SAS shop, depending on the data sizes and usage characteristics.

The SAS I/O pattern is also predominantly large-block, sequential reads and writes. This is very different from the small block, IOPS- (input/output operations per second) based throughput patterns of many other applications. Because of this, you will need to work with your administrators and storage vendors to ensure that any storage arrays, file systems, interconnects, and so on, that make up the I/O infrastructure are carefully configured and optimized for the way that SAS does I/O.

It is also important to note that your I/O infrastructure is only as fast as its slowest component. You should consider the peak I/O throughput requirements of the environment and work with your storage vendors to ensure that each segment of your I/O infrastructure can meet those demands.

File Systems

The three main file systems used by the SAS compute tier are SAS DATA, SAS WORK, and SAS UTILLOC. SAS DATA is where permanent SAS data files are stored. SAS WORK and SAS UTILLOC are both used as scratch working space for temporary data files. With these, a separate working space is created for each SAS session, and is destroyed when the session ends. With modern, fast SSD storage, SAS WORK and UTILLOC can be combined into a single file system. Here are the *minimum* throughput requirements for each of these file systems:

- SAS DATA: A minimum of 100-125 MB/sec per physical core. If this file system is used primarily for reads, the minimum throughput can be reduced to 50-75 MB/sec per physical core.
- SAS WORK and SAS UTILLOC: A minimum of 150 MB/sec per physical core.

File System Types

There are two general groups of file systems: non-shared (local) and shared (clustered). Local file systems are independent to each node. Clustered file systems house the common data that is shared between all the nodes. SAS has extensively tested many file systems across both and has developed sets of recommendations and best practices.

For more information about the various types of file systems for use with SAS, see the Best Practices for Configuring Your I/O paper.

Please note that SAS does not recommend the use of NFS file systems. Not only do these file systems have a history of file locking issues, but accessing data files stored here commonly negatively affects performance because of insufficient network bandwidth.

Non-Shared File Systems

SAS recommends the following non-clustered file systems:

- Linux: XFS
- Windows: NTFS
- AIX: JFS2

Best practices and configuration recommendations for these can be found in SAS Note 53873.

Shared File Systems

There are numerous clustered file system options available for use with SAS. Here are recommendations, test results, and best practices:

- Shared File Systems: Determining the Best Choice for Your Distributed SAS Foundation Applications
- SAS Note 53875

Storage Arrays

Storage arrays are one of the most important parts of the I/O infrastructure. The SAS Performance Lab is constantly working with storage vendors to test storage arrays and assemble papers with tuning guidelines and configuration recommendations. These papers can be found in SAS Note 53874.

HIGH AVAILABILITY AND DISASTER RECOVERY

You never know when a server will crash, or when a disaster will strike and take out an entire data center. The world is unpredictable, and because of that, high availability and disaster recovery should be essential parts of every SAS infrastructure. For more information about these, see SAS Note 62238.

MONITORING

After the above factors have been taken into consideration and the infrastructure has been optimally implemented, it is very important to continuously monitor the environment. Proactively monitoring usage patterns enables you to be prepared when additional resources or resource upgrades are needed. For more information about appropriate monitoring guidelines, see SAS Note 53877.

CHECKLIST FOR SAS PLATFORM ADMINISTRATION TASKS

In addition to the recommendations listed throughout this paper, the SAS Global Enablement and Learning team has compiled a checklist of administration tasks that need to be followed for SAS 9.4 deployments.

CONCLUSION

The hardware requirements for modern SAS environments vary greatly depending on the specific SAS applications, number of users, data sizes, anticipated growth, and so on. The overabundance of factors makes it impossible to create a one-size-fits-all recipe.

It is very important to have a complete understanding of the application and usage expectations before an infrastructure is implemented, and to continuously monitor the environment after. The recommendations and guidelines in this paper are meant to be used as a starting point for architecting an efficient infrastructure.

REFERENCES

SAS Institute Inc. 2014. SAS Note 42197. "A list of papers useful for troubleshooting system performance problems." Available http://support.sas.com/kb/42/197.html.

SAS Institute Inc. 2014. SAS Note 53873. "Troubleshooting system performance problems: Operating System Tuning papers." Available http://support.sas.com/kb/53/873.html

SAS Institute Inc. 2014. SAS Note 53875. "Troubleshooting system performance problems: Shared/clustered file systems." Available http://support.sas.com/kb/53/875.html

SAS Institute Inc. 2014. SAS Note 53877. "Performance Monitoring and Troubleshooting." Available http://support.sas.com/kb/53/877.html

SAS Institute Inc. 2015. "Checklist of SAS Platform Administration Tasks." Available http://support.sas.com/resources/papers/Platform-Administration-Tasks.pdf

Brown, Tony, and Margaret Crevar. 2016. "Best Practices for Configuring Your I/O Subsystem for SAS®9 Applications. *Proceedings of the SAS Global Forum 2016 Conference*. Cary, NC: SAS Institute Inc. Available http://support.sas.com/resources/papers/proceedings16/SAS6761-2016.pdf

SAS Institute Inc. 2017. SAS Note 53874. "Troubleshooting system performance problems: I/O subsystem and storage papers." Available http://support.sas.com/kb/53/874.html

SAS Institute Inc. 2018. SAS Note 62238. "Troubleshooting system performance problems: High-availability and disaster recovery papers." Available http://support.sas.com/kb/62/238.html

SAS Institute Inc. 2018. SAS Note 62239. "Troubleshooting system performance problems: SAS on the Public Cloud." Available http://support.sas.com/kb/62/239.html

SAS Institute Inc. 2018. SAS Note 62240. "Troubleshooting system performance problems: SAS on virtualized infrastructure papers." Available http://support.sas.com/kb/62/240.html

SAS Institute Inc. "SAS Product Support for Virtualization Environments." Available http://support.sas.com/techsup/pcn/virtualization.html. Accessed March 13, 2019.

ACKNOWLEDGMENTS

A special thank you to Margaret Crevar and Tony Brown for their contributions to this paper.

CONTACT INFORMATION

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

Jim Kuell
SAS Institute Inc.
Jim.Kuell@sas.com
http://www.sas.com