Paper 3609-2019

Top SAS® Architecture Pitfalls: Lessons Learned the Hard Way

Margaret Crevar, SAS Institute Inc.

ABSTRACT

Are you a SAS® infrastructure architect? Do you want to know what has worked for other SAS customers and what things should be avoided? Attend this discussion to learn from others' experiences with SAS. We also encourage you to share your experiences with us.

INTRODUCTION

The intent of this presentation is to do a review of the top SAS architecture pitfalls collected by various SAS Architects, along with information on how to avoid these. This presentation is intended to be more of a discussion with the audience than a straight lecture.

BEFORE YOU START

A good understanding of the SAS workload requirements and hardware infrastructure needed to meet the service objectives (SLAs), specifically time to task complete, is crucial. For existing SAS customers, the following questions help guide that examination:

- Are there SAS jobs that need to execute within a certain time frame? Are you expecting your SAS jobs to execute in the same time as—or faster than—they are currently running in your existing data center? If so, a determination of the IO throughput required for each file system being used must be made.
    - If moving to the public cloud, it must be determined if this same IO throughput can be achieved in the given cloud infrastructure.
- What contention is on the infrastructure? You must consider the extent of concurrent SAS and non-SAS jobs at peak times which may impact your SLAs.
- Where is the source data for the SAS jobs located?
    - If moving to the public cloud, does this data already reside in the public cloud of choice? If not, the amount of time to move data to the cloud space where SAS is executing in needs to be determined. This added time will affect the SLA of the jobs that consume off-cloud data.
- What are your security requirements? You should have a good understanding of how you'll be handling authentication of users and authorization to access data, content, and functionality within the SAS environment.

TOP ARCHITECTURE PITFALLS

NOT ALL THE COMPONENTS OF THE SAS INFRASTRUCTURE ARE IN THE SAME LOCATION

What do we mean by this statement? We mean that the source data files, the storage, or some of the systems where SAS is doing computations are not in the same data center on the same subnet. Here is a blog that discusses this in more detail: https://communities.sas.com/t5/Administration-and-Deployment/Does-It-Matter-Where-the-Various-Components-of-Your-SAS/m-p/483426

PROVISIONING THE UNDERLYING HARDWARE

It is very important when you are creating your virtual systems for SAS that you understand what the underlying hardware is. You need to make sure that every component of the underlying network and hardware infrastructure can meet the IO demands that will be happening between the SAS components, their attached storage, and any other network communication (backups to storage) that will be happening.

Another important aspect that you need to understand is whether the underlying CPUs are housed in multiple sockets (NUMA nodes). A good practice is to keep all the cores of your virtual system on a single socket within the underlying hardware. With today's sockets containing 8, 10, 14, 20, etc cores per socket, you need to understand their socket placement before you assign CPU cores to your virtual systems for SAS.

ENSURING YOU HAVE ENOUGH NETWORK BANDWIDTH FOR SAS COMPONENTS

Over the years, we have learned that the communication between the SAS 9.4 components (SAS Compute servers, SAS Metadata server, SAS mid-tier servers) can be quite chatty. Because of this, we strongly encourage that each SAS 9.4 system has a dedicated 10 Gigabit network interface card (NIC) and underlying fabric/switches, minimum, for this communication between the SAS 9.4 systems. Larger NICs may be required depending on the usage patterns and data sizes.

We are also seeing a very similar need for network communications between the SAS Viya systems (CAS Controller, CAS workers, SAS Programming Runtime node, CAS Microservices node, RabbitMQ/Postgres node). Because there is lots of communication between these systems, we are also strongly encouraging a minimum of a 10 Gigabit channel for the communication between SAS Viya systems.

ENSURING YOU HAVE ENOUGH NETWORK BANDWIDTH TO EXTERNAL STORAGE

For SAS 9.4, we recommend a minimum of 100 MB/sec/physical core for the various SAS file systems. For SAS WORK, we increase that IO throughput suggestion to a minimum 150 MB/sec/core. These are general suggestions and your SAS applications may need more or less than the stated speeds.

For SAS Viya, there are file systems that may need similar IO throughput speeds. These speed recommendations will be driven by the time you need to uplift the input data for CAS actions from disk into memory.

In both scenarios, for network attached storage, you may want the NIC card that will be used to transfer to/from external storage to be separate from the NIC card used for communication between the SAS servers.

ENSURING YOU HAVE ENOUGH HARDWARE CAPACITY TO SUPPORT YOUR SHARED FILE SYSTEM

If you plan to use a shared file system (SFS), you will need to work with the SFS vendor to understand the compute, memory, and IO bandwidth resources (more than just disk capacity) that are needed to support the SFS. In some cases, the SFS manages its own page caching space to help improve IO transfers. This caching uses both memory and CPU cycles, so you will need to make sure that these hardware requirements are part of the hardware requirements for SAS.

UNDERSTANDING VIRTUALIZATION AND SAS

SAS runs on virtualized systems without issues. However, you need to understand how your virtualized systems and storage are configured. Are they thinly provisioned (the underlying hardware is over committed) or thickly provisioned? If your VM farm is thinly provisioned, then a SAS application that is accessing large volumes of data (over 100 GB) in a large block sequential manner may not perform well on this farm.

Here are two papers to review on this subject:

- "Moving SAS applications from physical to virtual VMware environment" http://support.sas.com/resources/papers/MovingVirtualVMware.pdf
- "Deploying SAS Grid Systems on VMware ESXi Virtually Provisioned Storage" (https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2018/1931-2018.pdf)

ENSURING YOUR STORAGE IS CONFIGURED CORRECTLY

Accessing data on external storage arrays is required by all SAS applications. How fast this access is depends on the timeframe required to complete the SAS task. It is important to understand if the storage will be shared by other applications.

Please review the best practices for configuring storage with SAS. These can be found here: http://support.sas.com/kb/53/874.html

CONCLUSION

As you can see from this paper, it is very important to understand the SAS workload for which you are designing a hardware architecture, what the proposed hardware will be, and where the components of the hardware architecture and data sources will resid. The choice of hardware resources, data stores, and application architecture placement are crucial to achieving the best performance for your SAS users. It is also very important to follow all the best practices in configuring the hardware architecture, storage and data sources, whether your are planning to run on bare metal, a VM system, or using container technology. These apply when deploying SAS both on-premises and in the public cloud.

There may be times where you have to use slower systems or storage because that is all you have the budget for. You must carefully consider the performance ramifications of various hardware architecture decisions along with your budget.

REFERENCES

All papers listed on this support.sas.com site. http://support.sas.com/kb/42/197.html

ACKNOWLEDGMENTS

Many thanks to Tony Brown, Jim Kuell and Mark Schneider for reviewing the content of the paper.

CONTACT INFORMATION

Feel free to contact Margaret with your comments or questions after reading this paper:

Margaret Crevar
SAS Institute Inc
Margaret.Crevar@sas.com