

Paper SAS4312-2020

Important Performance Considerations When Moving SAS® to a Public Cloud

Margaret Crevar, SAS Institute Inc.

ABSTRACT

When choosing a hardware infrastructure for your SAS® applications, you need a solid understanding of all the layers and components of the SAS infrastructure. You also need to not just successfully run the software but to optimize its performance. Finally, you need an administrator to configure and manage the infrastructure. This paper discusses important performance considerations for SAS®9 (both SAS® Foundation and SAS® Grid Manager) and for SAS® Viya® when hosted in any of the available public clouds—Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform, and so on. It also provides guidance on how to configure the cloud infrastructure to get the best performance with SAS.

Disclaimer: We strongly encourage you to take the advice in this paper and work with your local public cloud teams to make sure the instances you decide to use are available in the closest region and that you understand their costs. In addition, any advice in this paper is based on the information we have at the time of publishing this paper (March 2020).

INTRODUCTION

Many SAS customers are making the decision to move their current SAS applications from their on-premises data centers to a public cloud. The hype around public clouds portrays this as a very simple task that saves SAS customers a lot of money.

The information discussed in this paper is based on what is available from the public clouds and our experience with the public clouds at the time of its writing. Public cloud offerings are constantly changing. Therefore, it is in your best interest to understand the rationale used in the selection process and to consider what was done as a point-in-time design.

However, there is a lot of planning needed and, depending on the architecture requirements from the SAS customer, the price might not be cheaper than on-premises hosting. You might need to provision more cores, enhanced networking, and disk space capacity to ensure the success of SAS public cloud deployments. This is particularly true if IO throughput is crucial to the success of your SAS applications in the public cloud. The reasons for this are explained in this paper.

BEFORE YOU START

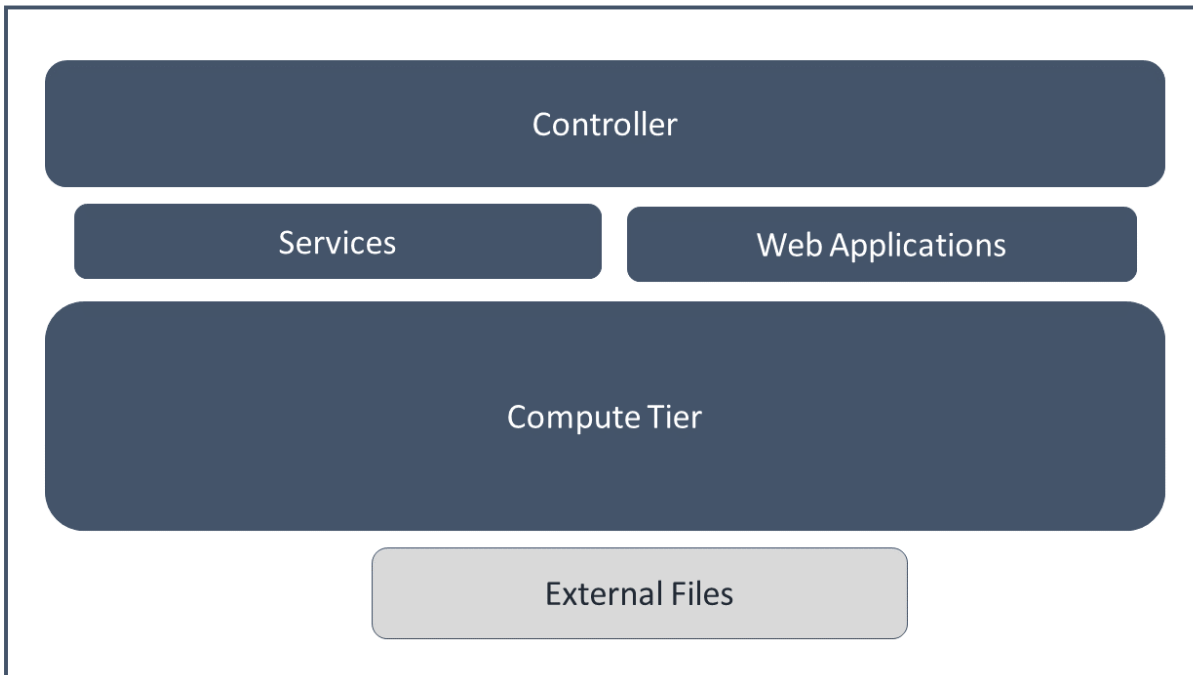
As mentioned in the introduction, a good understanding of the SAS workload requirements, along with the hardware infrastructure required to meet the service objectives (SLAs), specifically the time to complete the task, is crucial. For existing SAS customers, the following questions help guide that examination:

- 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 frame—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. It must be determined if this same IO throughput can be achieved in the public cloud.

- Where is the source data for the SAS jobs located? Does this data already reside in the public cloud of choice? If not, the amount of time required to move data to the cloud space where SAS is executing must be determined. This added time will affect the SLA of the jobs that consume off-cloud data.
- How much network bandwidth is required between your SAS servers? We are finding a minimum of a dedicated 10 Gbit network connection (NIC) is needed for the communications of all the SAS servers – both 9.4 and SAS Viya.
- **Is the customer’s IT staff willing to do stand-up authentication in the public cloud?**
- What security is needed for the data and/or SAS code?

The answers to these questions and fact-finding need to be fully understood so that the correct hardware and storage are selected from the available public cloud offerings.

Before we discuss what instance types to use, let’s go over the major parts of SAS 9.4 and SAS Viya 3.x. Please note between the major parts and the external data files, there needs to be a robust network bandwidth – we strongly recommend a dedicated 10-gigabit or faster network.



In addition to the robust network, there are some other things that need to be known to help you with your instance selections. The controller and compute nodes need to be the same processor family (i.e. Skylake or Broadwell). For SAS Viya, this includes the CAS Controller and CAS Workers. For SAS 9.4, this would be the SAS Grid Manager and SAS Grid nodes.

There are many different hardware and storage types. Some are hardware equipped for the heavy analytical and large sequential IO that SAS 9 does. Others are better equipped for the in-memory needs of SAS Viya. It is important to understand the workload profile of the **customer’s SAS applications to ensure that correct hardware and storage selections (cloud server and storage types)** are made for the best performance. Please note that to get the best achievable performance, the least expensive hardware and storage types from the public cloud offerings might not be suitable. For example, the customer might require server and storage instances with more physical cores than required for computing needs and/or

more storage capacity than the initial sizes needed to acquire the maximum IO bandwidth available for their SAS applications.

Now let's talk about what needs to be considered to ensure that you can configure the hardware infrastructure in the public cloud to perform as optimally as possible. These things include the following:

- what server instance type to use for your SAS 9.4, SAS Viya, or hybrid SAS infrastructure
- what storage type (for both persistent and nonpersistent storage) to use
- if you are deploying SAS® Grid Manager, what shared file system to use
- where to place temporary (SASWORK/UTILLOC and CAS_Disk_Cache) and permanent (SASDATA and CASDATADIR) data to be used by SAS
- where to place the SAS clients that will be used
- where to place authentication tools
- whether high availability and security are required

WHAT INSTANCE TYPE TO USE

In SAS 9 and SAS Viya infrastructures, there are several SAS server types and uses. Each has different and specific requirements for CPU, IO throughput, and memory provisioning. We will list each SAS server type and discuss its provisioning requirements. Please remember that most public cloud instances list CPUs as virtual CPU(s). These CPUs might be hyperthreaded (two threads per physical core). You need to understand if the vCPU includes hyper threads so that you can ensure you have the correct number of physical cores for SAS. For example, Oracle Cloud Infrastructure (OCI) instances list CPUs as Oracle Compute Units (OCPU). An OCPU is defined as the CPU capacity equivalent of one physical core of an Intel Xeon processor with hyper threading enabled.

You might have to use an instance with more physical cores in it than your workload requires. This is because a higher CPU count machine might be required to obtain a dedicated network interface card (NIC) of enough bandwidth (minimum 10 Gbit dedicated) to maximize IO to and from off-server cloud storage. Server instance types come in set models, with set CPU counts, set NIC installation, RAM, and so on. To get a dedicated NIC or the IO bandwidth required for your SAS workload through the NIC to the off-server storage, you might have to upgrade to a larger server type (with more CPU and RAM than needed for computing). When calculating NIC IO bandwidth capacity to drive storage, be aware that sharing a NIC in a cloud server host with other multi-tenant applications residing on the same physical server might result in inferior performance than occurs randomly. This is especially true for virtualized cloud host instances.

When setting up the instances, make sure that all the instances and components of your SAS infrastructure (fat clients, data (including data stored in external data stores), authentication software, SAS products, and so on) are in the same cloud space (region, area zone, placement group, and so on). Failure to do so results in an additional WAN connection that severely impacts performance.

Please note that AWS has a clustered placement group (CPG) designed for low network latency and high network throughput for instances that communicate with one another via the network. This sounds like a good feature for SAS, but you need to also note that AWS recommends that all the instances you spin up in the CPG be the same instance type. You can find more details on CPG here

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/placement-groups.html#placement-groups-cluster>

Please note that Google has the ability for you to place all your SAS servers on a single physical server. This can be done by creating sole-tenant nodes. You can find more details on sole-tenant nodes from GCP here

<https://cloud.google.com/compute/docs/nodes/#overview>

Google also has the ability to create an instance template that can be used to create a group of instances that all have the same configuration, for the SAS 9.4 Grid compute nodes and SAS Viya CAS workers, for example. You can find more details on Instance Templates here <https://cloud.google.com/compute/docs/instance-templates/>

SAS 9 SYSTEMS

These are general guidelines. We suggest that you have a detailed workload assessment done to determine which hardware is needed to support your planned usage of SAS 9 in the public cloud.

SAS Compute Tiers Including SAS Grid Node

These servers need fast CPUs for processing data, a minimum of 8 GB of physical RAM per physical core, and robust IO throughput (especially to SAS WORK and SAS UTILLOC):

- Amazon
 - I3en family – The primary reason is the high internal IO bandwidth from striped NVMe SSD drives for SASWORK and UTILLOC file systems. Like the I3 family, but with much larger NIC cards.
 - I3 family
 - R5d family
 - M5n family
 - C5d family – This family does not have the amount of RAM per physical core that is recommended for SAS.
- MS Azure – Azure Accelerated Networking must be enabled.
 - ESv3 series - this series can be Broadwell or Skylake processors. You cannot choose which processor is used when creating an instance.
 - FSv2 series
 - Lsv2 series (AMD chip)
 - M series – if you would like the largest amount of RAM per physical core
 - Mv2 series
 - DSv3 series – these have the smallest ephemeral disks. This series can be Haswell, Broadwell, or Skylake processors. You cannot choose which processor is used when creating an instance.
- Google
 - N1-standard or N1-highmem series – be sure to choose a Broadwell or Skylake processor.
 - N2-standard series
 - C2-standard series
- Oracle Cloud Infrastructure – this series can provide NVMe SSD drives that can be striped together in a RAID0 configuration to give the best IO bandwidth for SASWORK and UTILLOC file systems.
 - VM.DenseIO2.x series
 - BM.DenseIO2.x series
 - BM.HPC2.x series

Shared File System Storage Required for SAS Grid

These servers need robust IO throughput to the permanent storage that will support shared file systems. These instances will also need a minimum of 8 GB of RAM per physical core:

- Amazon
 - M5/M5d family
 - R5/R5d family
 - C5N and C5D family
- MS Azure - Azure Accelerated Networking must be enabled.
 - ESv3 series - this series can be Broadwell or Skylake processors. You cannot choose which processor is used when creating an instance.
 - DSv3 series – these have the smallest ephemeral disks. This series can be Haswell, Broadwell, or Skylake processors. You cannot choose which processor is used when creating an instance.
- Google
 - N1-standard series
- Oracle Cloud Infrastructure
 - BM.Standard2.52 – **recommended, since it's a bare metal compute** instance, they come with 2 physical NICs, each of 25 Gbps. One NIC can be fully utilized to access block storage for best IO throughput. Oracle Cloud Infrastructure provides a service-level agreement (SLA) for network throughput between bare metal instances in the same availability domain in a virtual cloud network (VCN).
 - BM.DenseIO2.52 – DenseIO shapes are bare metal compute instances with local NVMe SSD storage. Local NVMeS are not ephemeral, which means data survives reboot. Network attached Block storage can also be used in addition to Local NVMe SSD drives.
 - BM.HPC2.x series
 - VM.Standard2.x series
 - VM.DenseIO2.x series

SAS Mid-Tier and Metadata Servers

These servers do not require computing-intensive resources or robust IO bandwidth, but they do require access to more memory than the SAS computing tiers. The recommendation is a minimum of 24 GB of physical RAM or 8 GB of physical RAM per physical core— whichever is larger:

- Amazon
 - R5/R5d family
- MS Azure - Azure Accelerated Networking must be enabled.
 - M series - if you would like the largest amount of RAM per physical core
 - Mv2 series
 - ESv3 series - this series can be Broadwell or Skylake processors. You cannot choose which processor is used when creating an instance.
 - DSv3 series – these have the smallest ephemeral disks. This series can be Haswell, Broadwell, or Skylake processors. It does not appear you can choose which processor to use you create the instance.
- Google
 - N1-highmem series

- Oracle Cloud Infrastructure
 - VM.Standard2.x series. (VM.Standard2.2 or higher)
 - VM.Standard.E2.x series. (VM.Standard.E2.4 or higher)

SAS VIYA SERVERS

Here is a list of several machines to use for the most robust SAS Viya offering. Some of the SAS Viya offerings (single, non-distributed node going against small data from a source other than Hadoop) will not need all the servers listed here. These are general guidelines. We suggest that you have a detailed workload assessment done to determine which hardware is needed to support your planned usage of SAS Viya in the public cloud.

CAS Nodes

For an SMP CAS deployment, only one CAS node is required. For an MPP CAS deployment, a minimum of three CAS nodes are required, one of which will be the CAS Controller.

These nodes require fast CPUs for processing data, enough physical RAM to hold all the data files to be analyzed by all the concurrent SAS Viya users, and robust IO throughput (especially to CAS_Disk_Cache and in some use cases CAS Data Directory). If you are not sure how much data will be accessed at any given time, but you know that your SAS users will be accessing files in the 100s of gigabytes range, we recommend a minimum of 16 GB of RAM per physical core (more if you have large files that will reside in memory):

- Amazon
 - I3en family
 - R5d/R5ad family
 - C5n/C5d family – this family does not have the amount of RAM per physical core that is recommended for SAS.
- MS Azure - Azure Accelerated Networking must be enabled.
 - ESv3 series - this series can be Broadwell or Skylake processors. You cannot choose which processor is used when creating an instance.
 - FSv2 series
 - Lsv2 series (AMD chip)
 - M series – if you would like the largest amount of RAM per physical core
 - Mv2 series
 - DSv3 series – these have the smallest ephemeral disks. This series can be Haswell, Broadwell, or Skylake processors. It does not appear you can choose which processor to use you create the instance.
- Google
 - N1-standard or N1-highmem series – be sure to choose a Broadwell or Skylake processor.
 - N2-standard series
 - C2-standard series
- Oracle Cloud Infrastructure – below are shapes with local NVMe SSDs for fast access to CAS_Disk_Cache:
 - Bare Metal DenseIO shapes
 - Virtual Machine DenseIO shapes
 - Bare Metal HPC shapes

MicroServices and Web Interfaces Nodes

These nodes do not require high computational speed or power. To run all SAS® Viya products, you will need at least 96 GB of RAM in your MicroServices Node and your web interfaces node:

- Amazon
 - R5/R5d/R5a family
- MS Azure - Azure Accelerated Networking must be enabled.
 - M series
- Google
 - N1-highmem series
- Oracle Cloud Infrastructure
 - BM/VM.Standard Shapes
 - BM/VM.DenseIO Shapes
 - BM.HPC shapes

SAS® Programming Run-Time Node

This node will run SAS 9 code based on the application and requires all the file systems that a standard SAS 9 deployment requires. This node needs fast CPUs for processing data (at least 16 GB of RAM per physical core) and robust IO throughput (especially to SASWORK and SAS UTILLOC):

- Amazon
 - I3en family – the primary reason is the high internal IO bandwidth from striped NVMe SSD drives for SAS WORK and UTILLOC file systems. Like to the I3 family, but with much larger NIC cards.
 - I3 family
 - R5d I3 family
 - C5n/C5d I3 family – this family does not have the amount of RAM per physical core that is recommended for SAS.
- MS Azure - Azure Accelerated Networking must be enabled.
 - ESv3 series - this series can be Broadwell or Skylake processors. You cannot choose which processor is used when creating an instance.
 - DSv3 series – these have the smallest ephemeral disks. This series can be Haswell, Broadwell, or Skylake processors. You cannot choose which processor is used when creating an instance.
- Google
 - N1-standard series
- Oracle Cloud Infrastructure
 - BM/VM.DenseIO Shapes – multiple local NVMe SSD drives for high IO bandwidth for SASWORK and UTILLOC file systems.

WHICH STORAGE TYPE TO USE

The public clouds have multiple types of permanent and temporary data storage offerings for use with SAS. Discussed here are findings from general field experience and lab testing with SAS in public cloud spaces. Please note that the authors of this paper are not public cloud storage experts, we are just sharing our general field experience. Work with the public cloud provider to make sure the storage being used by SAS is configured optimally.

Permanent SAS Data Storage Will Persist Through a Reboot/Restart

Permanent SAS data storage is used for SAS 9.4 SAS data files and SAS Viya CAS tables:

- Amazon
 - Elastic Block Storage (EBS) – Our recommendation is to stripe together a minimum of 4 EBS volumes for IO bandwidth aggregation. Please see the [EBS Storage guide](#) on the Amazon website for more details.

- EBS st1 (Throughput Optimized HDD) Storage – Preferred. It is designed for large block sequential IO. A 12.5 TB volume can sustain 500 MB/second. If your volume size is less than this, you will only get 500 MB/second total bandwidth during your burst window.
 - EBS io1 (Provisioned IOPS SSD) storage can also be used. The above EBS Storage guide states there is a 250 MB/second maximum IO throughput per IO1 volume. Customers can also choose to have EBS IO1 (provisioned storage). However, costs would increase as IO1 volumes are charged by storage and by provisioned IOPS. For ex – 32K IOPS can yield as much as 500 MB/sec for which customers would pay an additional amount for the desired provisioned IOPS.
 - Other EBS storage types - like general-purpose SSD (gp2), and cold storage (sc1) - should not be used for permanent SAS 9 data files.
 - S3 storage – please review the SAS Usage Note 63001 (<http://support.sas.com/kb/63/001.html>) that discusses ways to get additional functionality and better performance to S3 with SAS.
- MS Azure
 - Premium Storage Disk Type. You will need to review the IO throughput per storage disk type (<https://docs.microsoft.com/en-us/azure/virtual-machines/linux/premium-storage-performance#premium-storage-disk-sizes>) to determine how many disks to stripe together. We suggest that for a 16 physical core instance, you will need to stripe eleven P20s, eight P30s, seven P40s/P50s, four P60s, or three P70s/P80s together. Please review the MS Azure web site to determine the maximum IO throughput that the instance type you plan to use can sustain to Premium storage.
- Google
 - Use the storage with the highest IO throughput. You should stripe at least 4 (preferably 8) together for each file system.
- Oracle Cloud Infrastructure
 - Block Volume Service. **Either the “Higher Performance” Block volumes or “Balanced” Block volumes.** Our recommendation is to stripe together enough volumes to provide the minimum of 100 MB/second/physical core IO bandwidth. Please see the [Block Volume performance](#) on Oracle Cloud’s website for more details.
 - Local NVMe SSDs. Several compute shapes (DenseIO) have non-ephemeral local NVMe storage, which survives reboot.

Temporary SAS Data Storage Will Not Persist Through a Reboot/Restart

Temporary storage is most commonly used for SAS WORK, SAS UTILLOC, and CAS_Disk_Cache because this data does not need to persist through reboots and restarts. When placed on striped NVMeS, SAS WORK and SAS UTILLOC can share a single file system:

- Amazon
 - Internal SSD devices striped together with RAID0. We would prefer you use instances with NVMe (ephemeral) devices with high bandwidth, low latency, and sequential IO - all of which are ideal for temporary SAS data.

- MS Azure
 - Internal SSD devices, at the present time, *cannot* be striped together with RAID0. However, you can use a single SSD device for SAS WORK. *
- Google
 - Internal SSD devices striped together with RAID0. *
- Oracle Cloud Infrastructure
 - Internal SSD devices striped together with RAID0. Use DenseIO Compute shapes with local NVMe SSD storage, which are not ephemeral, meaning they survive reboot. They are available for bare metal and virtual machine Compute shapes.

* Note: If there is not enough capacity using internal drives, please follow the advice for permanent SAS data files listed earlier.

IF DEPLOYING SAS GRID MANAGER, WHICH SHARED FILE SYSTEM TO USE

SAS Grid Manager requires a shared file system for the permanent files being shared by all the SAS Grid compute nodes. Here are some shared file systems that have been tested with SAS Grid Manager in several public cloud infrastructures:

- DDN Lustre - For optimal performance, you will need a *minimum* of four IO robust instances for your Lustre shared file system. A paper from April 2015 that discusses how we configured Lustre to work in AWS can be found here: <https://support.sas.com/rnd/scalability/grid/SGMonAWS.pdf>.

Please note that you must mount your Lustre file systems with this option (*flock*) in order for Lustre to work optimally with SAS, especially if you plan to host your SAS binaries on Lustre.

Also note that you need to understand that you are unable to expand the size of your Lustre file system. You will need to create a larger Lustre file system and copy data from the old system to the new one.

You can use your own copy of Lustre, or you can use a copy from the public cloud marketplace. Links to the marketplace offerings are listed here:

- Amazon –
 - Premier support: https://aws.amazon.com/marketplace/pp/B07FTX9GBV?qid=1548776061090&sr=0-1&ref_=srh_res_product_title
 - Self-support: <https://aws.amazon.com/marketplace/pp/B07FTSL9VX>
- Microsoft Azure – <https://wiki.whamcloud.com/display/PUB/Creating+a+Lustre+Cluster+on+Azure>
- Google – <https://wiki.whamcloud.com/display/PUB/Cloud+Edition+for+Lustre+on+Google+Cloud+Platform>
- Oracle Cloud Infrastructure – Install open-source Lustre using automated Terraform deployment template. <https://github.com/oracle-quickstart/oci-lustre>.
- IBM Spectrum Scale (formerly known as GPFS) – For optimal performance, you will need a *minimum* of four IO robust instances for your Spectrum Scale shared file system. A paper from April 2018 that discusses how to configure Spectrum Scale in AWS can be found here:

<https://support.sas.com/resources/papers/performance-tuning-considerations-amazon-web-services-sas-9-4-ibm-spectrum-scale.pdf>.

You can use your own copy of Spectrum Scale, or you can use a copy for the public cloud marketplace. Links to marketplace offerings are listed here:

- o Amazon - https://aws.amazon.com/marketplace/search/?filters=vendor_id&vendor_id=2f01b596-7972-4714-be52-b3709033137e
- o Microsoft Azure - not available at the writing of this paper
- o Google – not available at the writing of this paper
- o Oracle Cloud Infrastructure – Install IBM Spectrum Scale using automated Terraform deployment template - <https://github.com/oracle-quickstart/oci-ibm-spectrum-scale>.

- Amazon FSx for Lustre – This is a new shared file system only available on AWS. Information about this file system can be found here: <https://aws.amazon.com/fsx/lustre/>.

Please note that AWS sets up the Lustre file systems with these mount options (*rw, seclabel, lazystatfs*). As you can see from the information in this paper, these are not the mount options that SAS prefers. You will need to unmount the FSx for Lustre file systems that were set up for you by Amazon and remount them with the *flock* parameter.

Also note that you need to understand that you are unable to expand the size of your Lustre file system. You will need to create a larger Lustre file system and copy data from the old system to the new one.

- Amazon Elastic File System (EFS) – This file system is only available in AWS. Please note that the maximum IO throughput is currently 250 MB per second per instance, refer to Amazon EFS limits (<https://docs.aws.amazon.com/efs/latest/ug/limits.html#limits-efs-resources-per-account-per-region>) for the latest information on limits. You will need to have multiple EFS file systems per instance to overcome this IO throughput limitation, but you also need to understand that there is a limitation on the single NIC in each AWS EC2 instance. These file systems *cannot* be striped together.

Please note that AWS recently increased the number of hard file locks from 87 to 512 for these shared file systems.

- Azure NetApp Files** – This NFS-based file system is only available in Azure. When we tested ANF, we used these NFS mount options:

```
bg,rw,hard,rsize=65536,wsize=65536,vers=3,noatime,  
nodiratime,rdirplus,acdirmin=0,tcp,_netdev
```

- Google Cloud Filesystem** – This NFS-based file system is only available in Google. <https://cloud.google.com/filestore/>
- Oracle Cloud Infrastructure File Storage service (FSS)** – This file system is only available in OCI. FSS supports Network File System (NFS) protocol version 3 along with Network Lock Manager (NLM), and its throughput scales with the amount of underlying storage. <https://www.oracle.com/cloud/storage/file-storage.html>

** Please note that all the concerns with heavy WRITES to an NFS file system apply here. For more details about NFS and SAS 9.4 can be found starting on page 4 in this paper:

<http://support.sas.com/resources/papers/proceedings16/SAS6761-2016.pdf>.

WHERE TO PLACE VARIOUS SAS COMPONENTS

For optimal performance, it is very important to place all the components of your SAS infrastructure in the same public cloud region, placement group, and availability zone. This includes the clients, SAS binaries, SAS tiers, source data, authentication tools, and so on. If you do not place all the components so that they are on a LAN within the public cloud region, placement group, and availability zone, you will subject your SAS users to slow WAN speeds that can greatly impact the performance of their SAS applications. A blog, "Does it Matter Where the Various Components of Your SAS Infrastructure Are Installed", on this subject can be found here: <https://communities.sas.com/t5/Administration-and-Deployment/Does-It-Matter-Where-the-Variou-Components-of-Your-SAS/m-p/483426>.

CLIENTS

The most common placement of SAS clients like SAS® Enterprise Guide®, SAS® Data Integration Studio, and SAS® Studio (even though this is a web-based client, the issue with its placement in comparison to the SAS processing is very important) is within the public cloud infrastructure (either a Windows server or Windows virtual desktops). The question is where to put these Windows systems – in the public cloud in the same availability zone and placement group, on a system within your current data center, or on a desktop/laptop?

The answer depends on the volume of data being transferred back to the SAS client. If there is a lot of data being transferred to populate drop-down windows or to view tables, then having the clients and the backend SAS servers on the same LAN will yield the fastest results. For public clouds, this would mean standing up an instance with a Windows server on it and placing your SAS clients on that instance. Your SAS users would then have to access their SAS clients on the Windows server inside the public cloud infrastructure.

In that case, it is important that the Windows server is in the same region, placement group, and availability zone as the SAS deployment.

SAS INFRASTRUCTURE

SAS has several tools that allow sharing of SAS data files on-premises with SAS applications that run in a public cloud and vice versa. While these tools function well, please note that the IO throughput between your data center and the public cloud might be as low as 500 KB/second. Rarely have we seen it more than 20 MB/sec. If the SLAs for your SAS applications in the public cloud can be met with this slow data transfer rate, then it is acceptable to keep your data in a different physical location than where SAS is running.

Our best practice is to have SAS applications and the data associated with them residing in the same data center – whether on-premises or in the same availability zone and placement group in the public cloud.

SOURCE DATA FILES

As we mention above, it is best to have the source data files in the same location as the systems that will be using these files. If you must keep your source data files in a different location on the WAN, please note you will have to incorporate the slow WAN performance to the SLA for using SAS in the public cloud. Note that many of the client GUI pull data from the source data files when populating screens, so having the source data files across a slow WAN will greatly impact the performance of the SAS client.

AUTHENTICATION TOOLS

The considerations needed when deciding where to place your authentication tools are like those needed when deciding where to place SAS clients or data used by SAS: in the public cloud in the same availability zone and placement group or on a system within your current data center?

This depends on how often your SAS application will interact with your authentication tools for permission to use a file. If this is a high number, it is recommended that you move your authentication tools and associated data into the public cloud. Again, these need to run on a system in the same region, placement group, and availability zone.

AVAILABLE QUICK STARTS FOR SAS

To ease your deployment of SAS, there are several SAS Quick Starts available to you:

SAS 9.4 GRID MANAGER

- Amazon: <https://aws.amazon.com/quickstart/architecture/sas-grid/>
- Azure: not available at the writing of this paper
- Google: not available at the writing of this paper
- Oracle Cloud Infrastructure: <https://github.com/oracle-quickstart/oci-sas-grid>

SAS VIYA

- Amazon: <https://aws.amazon.com/quickstart/architecture/sas-viya/>
- Azure: <https://azure.microsoft.com/en-us/resources/templates/sas-viya/>
- Google: <https://github.com/sassoftware/quickstart-sas-viya-gcp>
- Oracle Cloud Infrastructure: not available at the writing of this paper

ARE HIGH AVAILABILITY AND SECURITY REQUIRED?

Depending on your need for high availability (HA), you might require that processes be in place to quickly create a new cloud host instance if one of your existing instances fails. This is more of a failover HA practice. The ability to quickly spin up a new instance is one of the benefits of running SAS in a public cloud.

For SAS Grid Manager customers, please note that a shared file system (for example, DDN Lustre, IBM Spectrum Scale) will remain operational if one of the nodes associated with the shared file system goes down. However, any data that is associated with that node will not be available until the node is restored. This is because only one copy of the data is stored by default. **It's possible to** enable replication services with these shared file systems so that two or three copies of your data are stored on multiple nodes of the shared file system. This does drive up the cost, especially if you have hundreds of terabytes of data.

If your definition of high availability includes disaster recovery, then you will need to look at mirroring your SAS deployment, SAS files, and data store to another region of the public cloud or even to a separate cloud. The SAS application considerations for disaster recovery will apply to the public cloud in the same manner as on-premises infrastructures. Details on what needs to be considered in a Disaster Recovery SAS implementation can be found in ["Do You Have a Disaster Recovery Plan for Your SAS® Infrastructure?"](#) (in *Proceedings of the SAS Global Forum 2016 Conference*).

CONCLUSION

This paper has been written to help SAS customers understand the public cloud instance types that best meet the needs of their SAS workloads. Many of the topics discussed in this paper are based on real-world experiences by SAS customers standing up SAS workloads in public cloud offerings. These customers have learned that there are very few things from a hardware and storage perspective that can be tuned, but the ones that can are listed in this paper. The chief takeaway is that IO throughput is crucial and, many times, is a limiting factor in the success of SAS public cloud deployments.

The intent of this paper is to raise awareness of SAS workload requirements (not only storage but also IO throughput and network bandwidth needs) and how best to meet these

requirements in a public cloud. The choice of hardware resources, data stores, and application architecture placement are crucial to achieving the best performance the cloud can offer. The maximum IO throughput to persistent data storage that most public cloud instances can offer is usually much lower than the typical SAS minimum recommended IO throughput of 100 MB/s per core (or more for SAS WORK file systems). You must carefully consider the performance ramifications of moving on-premises SAS workloads to the public cloud, unless you can suffice with a decreased IO bandwidth performance for persistent data activity.

As mentioned in the introduction, the information discussed in this paper is based on what is available from the public clouds and our experience with the public clouds at the time of writing. Public cloud offerings are constantly changing. Therefore, it is in your best interest to understand the rationale used in the selection processes and to consider what was done as a point-in-time design. We also encourage you work closely with your public cloud provider to insure you are using the best instance and storage that is available to you at the time or your deployment.

REFERENCES

Amazon Cloud. "Amazon EBS-Optimized Instances." Available

<https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EBSOptimized.html#ebs-optimization-support>.

Amazon Cloud. "Amazon EC2 Instance Types." Available

<https://aws.amazon.com/ec2/instance-types/>.

Google Cloud. "Machine Types." Available <https://cloud.google.com/compute/docs/machine-types>.

Microsoft Azure. "High Performance Compute VM Sizes." Available

<https://docs.microsoft.com/en-us/azure/virtual-machines/windows/sizes-hpc>.

Oracle Cloud Infrastructure. "Compute Shapes." Available

<https://docs.cloud.oracle.com/iaas/Content/Compute/Concepts/computeoverview.htm#InstanceTypes>

Oracle Cloud Infrastructure. "Block Volumes Performance." Available

<https://docs.cloud.oracle.com/iaas/Content/Block/Concepts/blockvolumeperformance.htm>

Red Hat. "Optimizing SAS on Red Hat Enterprise Linux (RHEL) 6 & 7." Available

http://support.sas.com/resources/papers/proceedings11/342794_OptimizingSASonRHEL6and7.pdf.

Crevar, Margaret. April 2016. "Do You Have a Disaster Recovery Plan for Your SAS® Infrastructure." Available

<https://support.sas.com/resources/papers/proceedings17/SAS0565-2017.pdf>.

Crevar, Margaret. August 2018. "Does It Matter Where the Various Components of Your SAS Infrastructure are Installed?" BLOG. Available

<https://communities.sas.com/t5/Administration-and-Deployment/Does-It-Matter-Where-the-Variou-Components-of-Your-SAS/m-p/483426>

Smith, Ben. April 2018. "Performance and Tuning Considerations on Amazon Web Services with SAS® 9.4 Using IBM Spectrum Scale™". Available

<https://support.sas.com/resources/papers/performance-tuning-considerations-amazon-web-services-sas-9-4-ibm-spectrum-scale.pdf>.

Crevar, Margaret. April 2015. "Performance and Tuning Considerations for SAS Grid Manager 9.4 on Amazon Cloud using Intel Cloud Edition for Lustre File System." Available

<http://support.sas.com/rnd/scalability/grid/SGMonAWS.pdf>.

ACKNOWLEDGMENTS

Many thanks to everyone who helped review this paper. There are too many to list.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact:

Margaret Crevar
SAS Institute Inc
Margaret.Crevar@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.