Paper SAS5520-2016

When the Answer to *Public or Private?* Is *Both*: Managing a Hybrid Cloud Environment

Ethan Merrill, Bryan Harkola, SAS Institute Inc.

ABSTRACT

For many organizations, the answer to whether to manage their data and analytics in a public or private cloud is going to be "both." "Both" can be the answer for many different reasons: common sense logic not to replace a system that already works just to incorporate something new; legal or corporate regulations that require some data, but not all data, to remain in place; and even a desire to provide local employees with a traditional data center experience while providing remote or international employees with cloud-based analytics easily managed through software deployed via Amazon Web Services (AWS). In this paper, we discuss some of the unique technical challenges of managing a hybrid environment, including how to monitor system performance simultaneously for two different systems that might not share the same infrastructure or even provide comparable system monitoring tools; how to manage authorization when access and permissions might be driven by two different security technologies that make implementation of a singular protocol problematic; and how to ensure overall automation of two platforms that might be independently automated, but not originally designed to work together. In this paper, we share lessons learned from a decade of experience implementing hybrid cloud environments.

INTRODUCTION

In this paper, we focus on some of the business considerations of running an enterprise SAS® solution in a hybrid environment, gained from SAS Solution OnDemand's experience as it expanded into the public cloud. A complementary paper, *Getting There from Here: Lifting Enterprise SAS® to the Amazon Public Cloud*, dives into the technical aspects of how SAS Solutions OnDemand implements and maintains SAS Solutions in the public cloud for its customers.

Customers have been successfully running workloads of all types in the cloud for many years. The hybrid model, if implemented correctly, allows for additional flexibility, such as decreased implementation times, regional deployments, and geographically separated disaster recovery (DR) models. The deciding factors for moving to the cloud reduce to matching the applicable use case with the right technology.

WHEN IS THE HYBRID CLOUD RIGHT FOR ENTERPRISE SAS?

The concept of a hybrid environment is not new. For the sake of this paper, we define a hybrid cloud as an extension of an internal network. More specifically, it joins internally accessible resources with public cloud resources and meets or exceeds corporate security guidelines.

Implementing SAS in a hybrid cloud environment does not automatically mean that the application takes advantage of cloud-based advancements, such as elasticity or bursting. Nor should it be expected that performance is better in the cloud; typically, it is not.

The following use cases often define the requirements that move customers to run enterprise SAS in a hybrid cloud model.

COMPLIANCE

In some cases, it is easier or more cost effective to adopt a cloud provider's ability to meet regulatory frameworks than it is to meet the same requirements internally. For example, as of this writing, AWS has released no fewer than 13 *Assurance Programs*, including FDA, HIPAA, NIST, and FedRAMP. If needed, AWS even provides an entire region called GovCloud that is constructed to meet ITAR compliance. Remember, putting your workload into AWS does not mean you are *compliant*; rather, it just means your workload is running on compliant infrastructure – this is the *Shared Responsibility* model.

REGIONAL OR GLOBAL DEPLOYMENTS

The most apparent need for geographical separation is usually for disaster recovery (DR) sites. The top-tier public cloud providers have a global presence, making it easier for them to satisfy DR geographical separation requirements. A hybrid cloud environment provides the flexibility for DR sites ranging from those that simply keep critical SAS data safe up to those providing 100% capacity and scalability of an on-premises system.

When there is a need to keep data close to users on the other side of the world, a public cloud might be the solution. Public clouds might also be applicable for data sovereignty; for example, SAS has customers in other countries that prefer to not store or host data in the US. If an organization devotes the time needed to build robust automation around a hybrid cloud strategy, it is possible to leverage a public cloud to establish a global presence very quickly and cost effectively.

COST AND TIME

Most organizations today realize that shifting assets to the cloud changes the way for which they are paid. There is also a new level of instant gratification offered by cloud providers, allowing the near instant spin-up of resources. This is particularly suited to test environments: spin up the required resources, install SAS, run some tests, and then shut everything down when the job is done. Instead of spending thousands of dollars on hardware and waiting weeks for rack, stack, ping, and power, the test can be completed for a fraction of the cost and in a fraction of the time.

Running an enterprise SAS stack in the cloud is generally more expensive than running a similar stack within an established, on-premises data center. The associated costs vary widely from company to company so our guidance is to run the budgetary numbers for both footprints. The main feature already realized by most people is that they only pay for what they use in the cloud. For example, a production instance, which runs 24x365, can be installed on-premises and a corresponding development instance can be installed in AWS. Simple automation can shut down the development instance overnight and on weekends. This can lead to notable reduction in the total cost of ownership.

INDEPENDENCE FROM IT

Sometimes business stakeholders want to take things into their own hands, control their own destiny… after all how hard can it be? In the right situation, having a hybrid cloud can help. The above situation of a quick test installation of SAS is a good example of when this might make sense. There are also a handful of reasons why this is a bad idea, most notably that successfully running enterprise SAS in the cloud for production workloads requires both IT and SAS expertise. The higher the expectations around performance and uptime, the more apparent this becomes.

HYBRID CLOUD

After an organization decides to build a hybrid cloud environment, it must address additional considerations. As discussed previously, underlying cloud technologies do not add new functionality or increase performance. For example, a hybrid cloud does not allow SAS to burst to a public cloud when it runs out of on-premises resources. What it does allow is for very simple data movement between internal resources and those resources in the cloud, using standard technology like `rsync`. It also allows for the resources in the public cloud to leverage domain resources such as DNS, AD or LDAP. Most important, the cloud resources follow the same, or better, security controls.

The following sections outline additional considerations that organizations should address when moving enterprise SAS into a hybrid cloud environment, notably data classification and operational complexity.

DATA CLASSIFICATION

Before moving any data into the cloud, it is important for organizations to understand applicable data classification policies. Classifications are often based on business criticality, sensitivity, and legal requirements. Based on these guidelines, the data is assessed and categorized, to determine whether storing the applicable data in the cloud is appropriate, and to provide a framework for any required controls to ensure data security.

A hybrid model that aligns with the highest attainable data classification requirements is generally recommended. When the architecture incorporates a hybrid model, cloud resources sitting in the public cloud can often be treated in the same way as internal resources.

Organizations must also ensure that they take the proper steps to meet or exceed established data backup and retention policies. The most simplistic method of backup in AWS is to use snapshots. Snapshots work best for small environments with few storage volumes that have short retention policies. As complexity increases, organizations might outgrow the snapshot method. At that point, we recommend using one of the many commercially available backup and recovery tools.

OPERATIONAL COMPLEXITY

Moving a SAS workload to a hybrid cloud can create a great deal of additional operational complexity, if treated like a separate system. To avoid this complexity, SAS Solutions OnDemand aligned its hybrid cloud architecture with established hosting policies and procedures. Doing so provided our cloud enablement team the ability to leverage over a decade's worth of established automation tools, ITIL processes, security best practices, and other leveraged systems within enterprise hosting. We partnered with internal SAS IT experts and worked side-by-side to build software-defined Virtual Private Clouds (VPCs), automated the spin-up of cloud-based resources, and implemented a robust security model to protect our customers' data.

Many of our support teams are not even aware that they are troubleshooting an instance of SAS running in the public cloud. They connect the same way as they would any other internal system, the compute resources and file systems are set up in an identical manner, and monitoring and event management are the same. Of course, our configuration management database (CMDB) tracks the hosting location and we follow rigid tagging guidelines so that identifying asset location requires only a simple query. Where possible, we leverage existing infrastructure and process to avoid unnecessary operational complexity.

AUTHORIZATION AND AUTHENTICATION

Leveraging established systems is especially useful when deciding how to securely access public cloud deployments. An example of this is our use of Active Directory (AD). We leverage our existing on-premises AD system by promoting domain controllers in each public cloud region that we deploy to. We then configure all of our deployed SAS systems to connect to their region's AD domain controllers. For our Windows servers, this is a straightforward AD domain connection; for our Linux servers, we leverage a tool called *VAS* that allows Linux operating systems to join an AD domain to provide user administration. This allows us to provide an authentication and authorization model that is identical to what we have used in on-premises SAS Solutions OnDemand hosting for the past decade. Subsequently, we achieve the following:

- Security, using an industry-standard authentication tool.
- Compliance, with a clear process around authorization and group membership.
- Uniformity, in that SAS administrators access AWS deployments the same way they access on-premises deployments.

VIRTUAL PRIVATE CLOUDS: AT THE CENTER OF A HYBRID CLOUD ENVIRONMENT

By leveraging the AWS Direct Connect and hardware virtual private network services, we are able to marry the SAS corporate network with the AWS network. Specifically, this allows us to create AWS VPCs that are reachable from our corporate network. This is the only way we could have accomplished the reuse of leveraged hosting services discussed earlier. Figure 1 contains a high-level diagram of how this works for SAS.

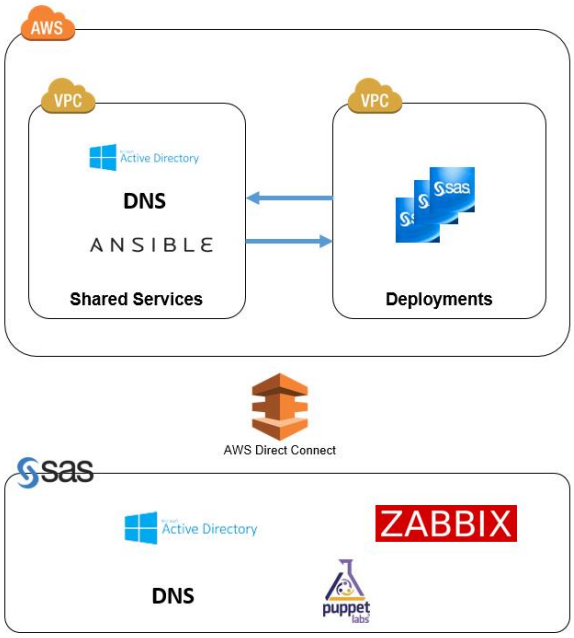

Figure 1: High-Level Network Diagram

Many of the lower-level details, including subnets, security group rules, Network Access Control Lists (NACLs), and peering rules, are not included here, to simplify the picture for purposes of this topic.

In AWS, we use one VPC, called Shared Services, which provides the services we leverage from SAS IT. Specifically, we have promoted AD servers and DNS servers. The reason to put these services into AWS is to prevent latency for the SAS instances that require them. Services having to go all the way back to the SAS on-premises network to resolve DNS or authenticate users would result in excessive latency, and this is especially true as we roll our hybrid footprint out to AWS regions around the globe.

The Deployments VPC is where we deploy hosted SAS solutions. It is peered with the Shared Services VPC so that deployed solutions can leverage the shared services. Specific security group rules are in place to allow only the access needed across the peered connection. This is also true for the hybrid network provided via AWS Direct Connect – specific, least-privilege SAS firewall and AWS security group rules are in place to allow only the traffic needed across the network boundaries.

In the end, the SAS-corporate-to-AWS hybrid cloud is basically just a hybrid network, without which we would have been unable to deploy our SAS Solutions OnDemand hosted customers to the public cloud in 2015.

PERFORMANCE IN THE PUBLIC CLOUD

Obtaining a benchmark for the expected performance of an enterprise SAS implementation *prior* to implementing it in a hybrid environment can also be helpful. This helps establish the desired performance targets for the cloud environment. Based on the way SAS behaves under load, it is typical to see bottlenecks shift between CPU, network, and disk I/O. We recommend first benchmarking an enterprise SAS system with known good performance, then rerunning the same tests on the compute instances located in the public cloud. Start outside of SAS with OS-level benchmarking, then resolve any issues prior to installing and benchmarking within SAS. The end goal is to complete benchmark testing using real workloads within the SAS environment.

After using established benchmarks to help determine the architecture of the cloud instances, the focus then becomes monitoring. We recommend extending the same tools that are used to manage the on-premises SAS implementation to those running in the public cloud. In addition, we recommend collecting metrics at both the hypervisor and host levels. For example, in AWS, best practices are to collect and analyze metrics from the SAS instances, in addition to the metrics from AWS CloudWatch. This provides

a holistic view into overall system performance and can help during performance tweaking and troubleshooting sessions.

CONCLUSION

Organizations are becoming more comfortable with running enterprise workloads in the public cloud. The value of leveraging a public cloud like AWS in this fashion is clear. However, installing SAS into a hybrid cloud environment requires a well thought-out architecture to ensure the system meets end user expectations. The first important considerations are whether and how to leverage existing, proven on-premises systems that can cross the hybrid cloud boundary, to provide out-of-the-box security, compliance, monitoring, and uniformity. After choosing these systems and extending them to the public cloud, the focus should shift to the SAS workload itself, to consider performance, benchmarking, backups, and recovery. It is important at this stage to compare workloads running in the public cloud to known, well-performing workloads on-premises. Finally, consider how best to leverage public cloud services, like snapshots, to make maintaining and recovering workloads as easy and automated as possible. By taking the time to plan a hybrid cloud architecture in this fashion, you can be confident in delivering a well performing, compliant enterprise-grade SAS environment that runs in a public cloud.

REFERENCES

Amazon Web Services. “AWS Cloud Compliance.” Accessed March 1, 2016. https://aws.amazon.com/compliance/

RECOMMENDED READING

- *Performance and Tuning Considerations for SAS® Grid® Manager 9.4 on Amazon (AWS)Cloud using Intel Lustre File System, https://support.sas.com/rnd/scalability/grid/SGMonAWS.pdf*
- *Monitoring 101: New Features in SAS 9.4 for Monitoring Your SAS Intelligence Platform, https://support.sas.com/resources/papers/proceedings13/463-2013.pdf*
- *Getting There from Here: Lifting Enterprise SAS® to the Amazon Public Cloud*, Paper SAS5501-2016

CONTACT INFORMATION

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

Ethan Merrill
SAS Institute, Inc.
+1 (919) 531-2241
Ethan.merrill@sas.com
http://www.sas.com

Bryan Harkola
SAS Institute, Inc.
+1 (919) 531-5604
Bryan.Harkola@sas.com
http://www.sas.com