Paper 459 -2013

Virtualized Environment for SAS® High-Performance Analytics

Thomas N. Keefer Jr, SAS; Rich Pletcher, SAS; Daniel Zuniga, SAS

ABSTRACT

Companies are adopting virtualization technologies to support entire IT infrastructures. There is an increase in requests for virtualization deployment reference architectures and best practices. A SAS staff built reference architecture was tested in a virtualized environment for SAS® 9.3, SAS® Grid Manager, SAS® High-Performance Analytics, and SAS® Visual Analytics. This paper outlines performance testing results and best practices for planning, managing and deploying a successful enterprise-class, virtualized environment.

INTRODUCTION

Customers are building out internal virtual environments or renting time on external “cloud” environments to host software applications. There are increasingly more products and solutions that are moving into virtualized environments. For years, SAS solutions have been used in virtualized environments for customer testing, development and production deployments. The move to virtualization has been primarily to reduce IT infrastructure costs. However, other benefits have been achieved as well and are highlighted in the design goals section of this paper.

This paper shares best practices and test results from a virtual environment deployment that was designed to support SAS High-Performance Analytics (SAS Grid Manager, SAS High-Performance Analytics Server, and SAS Visual Analytics). Virtualization software and shared storage were core components used to build out this scalable virtualized environment. The design and methods used in this project are applicable to small and enterprise sized deployments.

The target reader for this paper needs a general understanding of virtual computing terms and SAS High-Performance Analytics.

DESIGN GOALS

Before designing an environment, it is important to understand the detailed business and technology goals.

Design goals addressed in the described architecture include the following:

COST REDUCTION

Effective hardware usage, less idle resources reduces cost. Shutting down unused hardware reduces electricity usage and heat in the data center.

FLEXIBILITY AND SCALABILITY

- Provide the ability to increase and reduce the number of active system resources as needed (for example: end of quarter processing).
- Fast reallocation of system resources to other servers and projects.
- Built-in replication and or backup of servers (for example: quick creation of development or test servers from production).
- Ability to easily leverage and manage large amounts of data (many tables greater than 1TB each).

AVAILABILITY AND MAINTAINABILITY

- Apply patches or do maintenance on hardware without shutting down the software service.
- Provide some level of fault tolerance and/or highly available service if needed.

RESOURCE MANAGEMENT

Restrict and control access to resources and prioritize important tasks.

TEST DESCRIPTIONS

SAS High-Performance Analytics uses resources across a pool or cluster of servers. These solutions use large amounts of RAM, CPU, and I/O to solve complex business challenges. A multi-user test scenario was created to simulate activity within a virtualized environment. A combination of shell and HP Loadrunner scripts were used to control and launch the various workloads.

Test scenarios include the following:

SAS VISUAL ANALYTICS

The goal of this scenario is to demonstrate CPU usage characteristics and server response time to ad hoc analytical requests. The initial in-memory data tables in this multi-user Loadrunner driven test vary in size depending on the number of users and systems resources available. An example of a small table used in testing is; 115 million records, 46 columns, and 44GB stored in Hadoop. The number of users ramp up. This simulates the type of activities that would occur during a monthly, quarterly or annual reporting cycle. A mixture of tasks including reporting and analytic research tasks are executed during the scenario. During task execution, CPU and RAM resources across the cluster of servers are simultaneously leveraged. The critical measurement used during this project is SAS Visual Analytics report and task response times. The response time for individual tasks is longer on systems with increased activity. Best practices are used to help maintain acceptable response times during the project.

SAS HIGH PERFORMANCE ANALYTICS SERVER

A typical SAS programming task is manipulating data and executing a series of analytic processes. This test scenario simulates one or more users executing these types of tasks. Processes execute against data stored in Hadoop. Data is lifted into memory for analytic processing during each task execution. These processes execute across the cluster of servers in parallel by leveraging the same CPU and RAM resources used by the Hadoop environment. Batch scripts are used to launch these jobs and simulate work coming into the system from users running SAS® Enterprise Miner™, SAS® Enterprise Guide®, or display manager. Source data volumes used during the testing varied. Initial test tables were 18GB.

SAS GRID MANAGER

Scripts are used to interactively launch multiple processes to simulate users coming and going from the system during scenario execution. Jobs are spawned across the grid nodes simultaneously. Task types included in this multi-user scenario are designed simulate reporting, research analytics, and batch ETL workloads.

SAS HIGH-PERFORMANCE ANALTYICS COMBINED

The final test used a combination of all three scenarios executed at the same time. The SAS Visual Analytics workload is launched first. Then the SAS High-Performance Analytics Server and SAS Grid Manager scenarios are started in the background on other virtual machines in the environment. In all virtual environments there are multiple virtual machines and applications running simultaneously.

ARCHITECHURE OVERVIEW

Architecture design was done in partnership, using hardware and software partners. It meets the design goals described earlier in this paper.   However, it is important to note that each customer situation is unique. The architecture described in the paper is just one example of a virtual environment. Other products and configurations can be leveraged to achieve a robust virtual environment.

Critical components of a successful virtual environment are the virtualization layer and the shared file system. VMware was the virtual layer used in this effort. It provides the ability to carve the physical environment into multiple independent virtual machines. These virtual machines run an independent operating system copy and appear to users and applications to be a dedicated physical system. The shared file system provides the ability to migrate virtual machines and data quickly between all of the physical resources. EMC Isilon and Network File System (NFS) were used as the shared file system in this effort. The availability, flexibility, and scalability design goals are not possible without the shared storage component. Figure 1 is a diagram of the reference architecture.

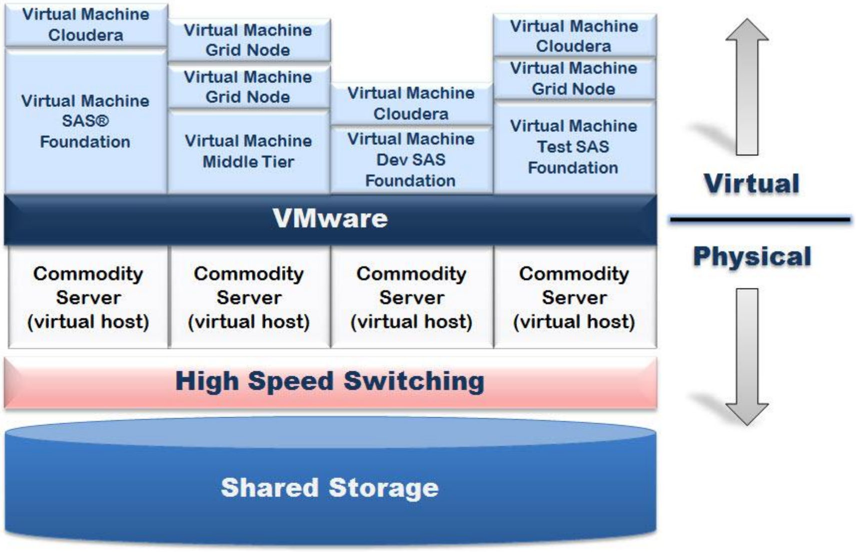

Figure 1. High Level Virtual Machine Architecture

The production environment used during testing is located at SAS Headquarters in Cary, NC. Other background activities on other virtual machines and servers ran in the environment during testing. However, the baseline test for each scenario was run during periods of reduced activity (off hours).

Figure 2 shows how the individual virtual machines were balanced across the physical hosts running VMware. The most active virtual machines in this test scenario were spread evenly across the physical resources.

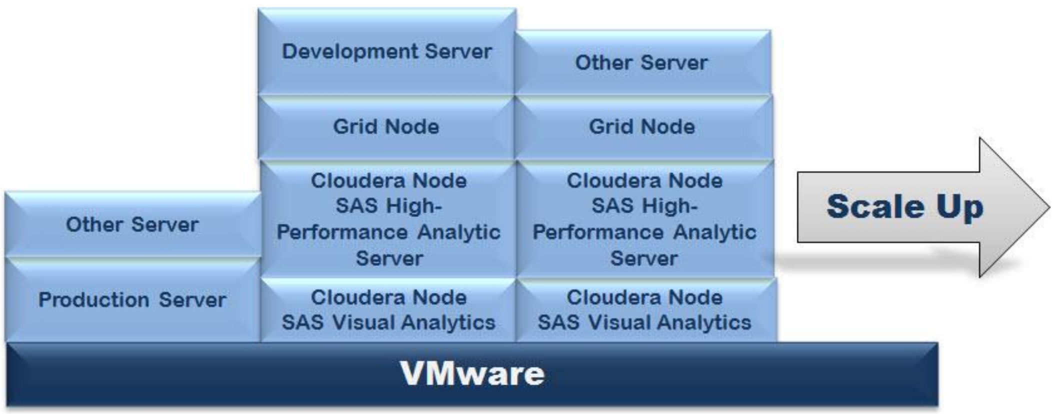

Figure 2. Virtual Machine Layout Used during Tests

HARDWARE COMPONENTS

Commodity x64 Servers used as virtual hosts (physical machines):

- Dual 10Gbit and 1Gbit Ethernet
- 12 CPU cores
- 48GB RAM
- EMC Isilon x400 Storage
- 10Gbit Ethernet Switches

SAS® SOLUTIONS AND THIRD-PARTY SOFTWARE COMPONENTS

- SAS Visual Analytics 6.1 (SAS 9.3)
- SAS High-Performance Analytics Server 12.2 (SAS 9.3)
- SAS Grid Manager (SAS 9.3)
- Cloudera CDH42 (Hadoop)
- Red Hat Linux 6.3
- VMware vSphere 5

TEST RESULTS

DEPLOYMENT HIGHLIGHTS

The ability to quickly clone or expand clusters of servers was the primary features evaluated with this testing architecture. During environment setup and testing virtual machines for both SAS software and Hadoop were replicated in order to create a second instance. This feature is valuable in a production environment where it is common to create a development, test, and production instance of a server or set of servers. This allows for a quick way to create a backup of an entire environment in order to allow for recovery to a known state following an issue. This feature was used to add nodes to the Hadoop instances in the virtual environment in order to increase performance. Nodes were added to the Hadoop environment dynamically without taking down servers or software components.

Server migration across physical resources was used during project deployment and setup. The tested environment was an active virtual environment with other projects running on the same physical resources. Server migration (VMware vMotion) was used to move virtual machines to different physical machines to minimize the impact to other projects.This feature can also be automated if desired.

PERFORMANCE TEST RESULTS

The SAS Visual Analytics test scenario was used as the primary tool to measure system performance during the project. Response times were effectively used as the primary measure for determining system effectiveness. The SAS Grid Manager and SAS High-Performance Analytics Server scenarios were used as background workload to add additional competition for the physical resources in order to simulate a real world environment.

The primary performance issue with shared virtual environments is maintaining predictable performance for critical applications and processes. Implementing resource management helps protect resources for priority applications and processes to achieve predictable performance. Testing focused on protecting SAS Visual Analytics with the use of virtual resource management. Task response time measurement was used to determine the effect of resource management. Table1 shows testing results. Average response times are shown for overall transactions. Additionally, the average response time for the resource intensive transaction open table is listed.

Test	Average Transaction Time	Average Open Table Time
One No other system activity	7.0 seconds	23 seconds
Two With background activity	8.1 seconds	40 seconds
Three With background activity and virtual resource management enabled	6.9 seconds	22 seconds

Table 1. SAS Visual Analytics response times captured during testing

The testing shows that SAS Visual Analytics baseline and final test execution times t with background workload and resource management enabled were similar. This result highlights that resource management was able to protect the more critical application and help provide for a predictable run time. SAS High-Performance Analytics Server run times (not reported here) were affected by the priority settings for SAS Visual Analytics in test three, by a few percentage points.

BEST PRACTICES

CAREFULLY PLAN RESOURCE REQUIREMENTS

Virtual resources do require physical resources, even though it “feels” like it is unlimited, to create new virtual machines at any time. A complete understanding of the CPU, RAM, and I/O requirements for all virtual machines and applications and physical resource usage is crucial. Failure to do this leads to performance issues when virtual machines fight to share limited physical resources.

VMware and other virtualization products have the capability to isolate and monitor applications during execution. It is important to monitor how physical resources are leveraged by an application before deployment n a virtual environment.

AVOID RESOURCE OVERCOMMITMENT

RAM

SAS Visual Analytics and SAS High-Performance Analytics Server are sensitive to RAM availability. When RAM is overcommitted (the servers and or applications need more physical memory than is available) application performance is reduced. These SAS solutions load data into memory for processing and are targeted primarily at larger data volumes. Guaranteeing there is adequate physical RAM is essential. This is done by ensuring the total RAM allocated for all virtual machines on a physical machine (virtual host) does not exceed the physical RAM in that box. When this is not possible, resource management can be used to prioritize the RAM allocation for the virtual machines that are intolerant to reduced RAM access.

For any SAS application that does significant I/O (storage access), RAM is critical for supporting the operating system file cache. When accessing large data volumes, file cache is using all the available, unallocated memory. I/O performance can suffer when RAM is not available.

I/O

Ensure there is adequate I/O throughput available to the virtual machines when needed. Design the storage subsystem with the ability to handle the peak I/O requirement from all active virtual servers and applications. The storage subsystem needs to handle simultaneous access from multiple servers and applications without hindering performance. Over commitment of shared storage I/O throughput capability is a common issue.

IMPLEMENT RESOURCE MANAGEMENT

In VMware, there are several ways to automatically or manually protect physical resources assigned to and shared by all virtual machines in the environment. During testing resource pools were used to set priority for CPU and RAM resources for critical systems in order to guarantee SAS High-Performance Analytics performance. These resource limits and parameters are highly customizable. Protecting RAM priority for SAS Visual Analytics and SAS High-Performance Analytics Server reduces performance degradation when other applications are sharing the same physical resources.

NETWORK ATTACHED STORAGE (NAS) SHARED STORAGE VERSUS SAN

Using network-based file systems like NFS are typically not as fast as fiber attached storage systems. However, they can be easy to administer and deploy. When using NFS or other NAS devices design the storage subsystem and network to supply the required I/O throughput across the virtual environment. Ensure the system can deliver sustained I/O across all of the virtual machines and file systems during peak system usage. Network connections can limit to I/O performance per file system. It is important to understand how applications use each file system and what the maximum I/O requirements are.

ADDITIONAL TOPICS

NETWORKING

In the current architecture the test environment uses a 1Gbit switched Ethernet to support the vMotion, vMware administration and boot disk traffic. In an enterprise class deployment more than one 10Gbit Ethernet is recommended to support the virtual environment to increase the speed of administrative functions. Isolate virtual management traffic from other activities to improve performance for the virtual machines and administration functions.

MIGRATION

In the test environment, SAS engineers replicated various components and entire multi-machine servers to create additional virtual server environments. Environments are cloned using virtualization cloning capability, operating system utilities, and SAS utilities. This feature is valuable for administrators who need to bring up additional environments in support development, testing, and migration to production.

NON-UNIFORM MEMORY ACCESS (NUMA)

Machines used for virtual hosts typically have multiple NUMA nodes (multiple self-contained CPU and RAM components). SAS systems that use large amounts of memory for processing are sensitive to improper memory management related to NUMA. VMware has several features that allow virtual machines to be isolated to NUMA nodes. This feature helps guarantee consistent memory access performance. SAS testing in other projects showed that using this capability when building out virtual environments can improve performance for applications with critical memory access requirements. Memory availability is critical for SAS High-Performance Analytics Server and SAS Visual Analytics. This can also be true for virtual machines that use a significant I/O (storage) because memory is used heavily by the operating system file cache.

PHYSICAL MEMORY ON VMWARE VIRTUAL HOSTS

The systems used in testing had 48GB of RAM with 12 cores each. For enterprise class environments, more than 16GB of RAM per CPU core is recommended. Memory is the most important resource in a virtual and SAS High-Performance Analytics environment.

AUTOMATED VERSUS MANUAL LOAD BALANCING

Most virtual software products like VMWare are able to move live virtual machines from one physical server to another. This involves copying active physical memory between machines and making the copied instance the active server. This allows for failover and the ability to manually move resources to physical systems that are not as busy. The virtual server products can be configured to automatically move virtual machines to help balance resource usage. This is useful in environments with multiple solutions and constantly changing resource usage. Automatic resource balancing increases overall physical system usage that helps eliminate system bottlenecks that can occur on physically dedicated servers.

CONCLUSION

Testing has shown that virtualization of resource demanding applications like SAS High-Performance Analytics are possible. Virtualization features such as resource management, cloning, and migration enhance the ease of deployment, help reduce cost and increase flexibility and scalability. The test results significantly emphasize that resource management is a critical best practice in order to help achieve predictable performance by protecting priority virtual machines. Based on this effort, virtualization is a viable option for all SAS applications when best practices and resource requirements are considered during planning.

ACKNOWLEDGMENTS

Rachel Wynne – SAS

Donna Bennett - SAS

Margaret Crevar - SAS

VMware, Inc.

EMC Corporation

Red Hat, Inc.

RECOMMENDED READING

- *Moving SAS Applications from a Physical to a Virtual VMware Environment*
- *VMware ESX5 Performance Best Practices*

CONTACT INFORMATION

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

Tom Keefer
Cary, NC 27513

Tom.Keefer@sas.com

Rich Pletcher
Cary, NC 27513
Rich.Pletcher@sas.com

Dan Zuniga
Woodbine, MD 21797
Daniel.Zuniga@sas.com