

Paper 390-2008

## Architecting a Finely Tuned SAS Grid Solution

Sandra Moulton, HP, Cupertino, CA

### ABSTRACT

SAS Grid Computing is a scale-out SAS solution that enables SAS applications to better utilize computing resources. When architecting a SAS Grid Computing solution it is important to understand the components required to ensure a scalable and high optimized solution. This paper details some of the components necessary to architect and tune a SAS Grid Computing solution.

### INTRODUCTION

SAS Grid Computing is an enterprise-class solution that enables SAS applications to take better advantage of computing resources. SAS Grid Manager manages the SAS workload dynamically across all nodes in the SAS Computing Grid. It provides dynamic load balancing, job prioritization and SAS job control. Part of the value-add in deploying SAS® for Grid Computing is that the solution can be architected for easy management and growth.

When architecting this solution it is important to understand the components required to ensure a scalable and highly optimized solution. SAS Grid Computing solutions are extremely *i/o* and compute intensive. They also require the use of a high performance clustered file system that allows all servers to access the same file systems. It is also important to consider management of the solution to ensure a smooth rollout and implementation and continued uptime. This paper will detail the important components of the SAS Grid Computing solution and how they scale as well as offering some tuning parameters.

In our testing we were able to demonstrate a linear solution for a SAS Grid Manager analytic environment that utilized HP ProLiant blade servers, StorageWorks storage arrays, and a highly available, scalable Clustered File system by PolyServe Matrix Server.

### COMPUTING REQUIREMENTS

SAS Grid Computing solutions allow for easy scalability as organizations grow. You can scale-up by adding processors into existing servers, or scale-out by adding new servers into the grid. The SAS Grid Manger solution provides dynamic resource-based load balancing and the ability to enforce policies based on time of day, workload, resource availability, etc.... This ability allows for greater utilization of diverse compute resources that allow for combination of slower speed compute engines with faster newer compute engines. Since the solution is presented as a single resource it is simple to add or remove compute resources without impact to the end users.

The speed of the processor will impact the speed of the throughput achieved. Compute intensive SAS jobs can take 100% of a single processor core for the duration of the job. Therefore it is important that the compute engine is extremely fast and capable of handling the compute needs of all SAS jobs. In our testing we found that optimal performance was achieved when splitting long running SAS jobs into multiple sub-tasks and matching the number of simultaneous processes to the number of processor cores available. It is also important to properly architect the SAS Grid Manager policies to meet the needs of the various SAS jobs. Testing the throughput of various compute servers will allow you to better understand the capabilities of the servers in your SAS Grid Manager solution and therefore properly configure the policies. Place high priority jobs on the faster compute engines and run longer lower priority jobs at off peak times.

### RECOMMENDATION

It is important that SAS Grid Manager is scalable and easy to manage. Given the compute intensive needs of a SAS Grid Manager solution it is important that the compute engine is extremely fast and capable of handling the compute needs of all SAS jobs. It is also important that SAS Grid Manager is scalable and easy to manage. Given these needs for SAS Grid Computing HP recommends HP ProLiant BL460c servers for SAS backend server functions.

HP ProLiant BL460c servers provide SAS Grid Manager customers with maximum compute power in a small form factor (up to 128 processor cores) in a single HP BladeSystem c7000 blade enclosure. Each HP ProLiant BL460c server supports up to two Quad-Core Intel Xeon processors (8 cores), 32 GB memory and 2 expansion slots to support your network and I/O needs.

With I/O throughput of significant importance to SAS® for Grid Computing customers the HP ProLiant BL460c blade servers help eliminate I/O system performance bottlenecks with support for up to two fibre channel (FC) host bus adapters (HBAs) providing redundant FC HBA support. These HBAs feature optimal performance utilizing PCI Express technology, 4/2/1 Gb/s auto-negotiating speeds and dual-ports for redundant path connections. The ProLiant BL460c servers also easily attach to SAN or NAS and can be easily managed from one location.

#### **RAPID DEPLOYMENT AND SIMPLIFIED SCALABILITY**

With SAS Grid Computing technology and HP BladeSystem servers it is simple to increase your compute power as your environment grows. First add another blade into the existing blade enclosure. To simplify the installation of the new blade server HP provides ProLiant Essentials Rapid Deployment Pack (RDP) solution. RDP facilitates taking an image of an existing server and copying it to one or many new servers through either a GUI-based or web-based console, using either scripting or imaging technology. This will greatly reduce server configuration time and allow rapid high volume server deployments. On Linux implementations it is recommended to install the SAS software on the shared SAN storage, therefore when adding a new server it will not have to be reinstalled. So the second step is to configure storage arrays and the clustered file system to include the new server. Finally add the new server in the SAS Grid Manager configuration files and the system is ready to accept SAS jobs from the SAS Grid Manager. All configurations can be done remotely without interruption of the current running solution. The HP BladeSystem allows simple addition of up to 16 blades per c7000 enclosure allowing scalability of SAS Grid Computing.

#### **SIMPLIFIED MANAGEMENT**

HP ProLiant blade servers come with HP Integrated Lights-Out 2 (iLO 2) Management for unprecedented, high-speed management which includes a secure graphical remote console that allows access to the system KVM and allows control of Virtual Power and media from a single console. HP Insight Control Data Center Edition delivers foundation management for HP BladeSystem lifecycles, including hardware resource deployment, health monitoring, software version control, performance monitoring, vulnerability scanning and patch management.

#### **STORAGE REQUIREMENTS**

SAS Grid Manager solutions are highly dependant on the architecture of the storage solution. In our testing we found that if the i/o throughput was not sufficient the overall solution performance suffered severely. What affects i/o throughput the most is the storage systems architecture, the SAN infrastructure and the number of spindles (disks). It is important to understand your peak i/o throughput needs as well as the i/o capabilities of your storage solution. You should check with your SAN hardware manufacturer to ensure the architecture is sufficient to meet your current throughput requirements and is scalable to meet your future requirements. You should be aware of the increasing i/o needs as you scale your solution in the future and ensure that you address the growing storage throughput requirements while increasing your compute engines.

See “tuning recommendations and best practices” section below for more specific recommendations.

#### **RECOMMENDATION**

The scalability and performance of SAS® for Grid Computing is dependant on a high performance optimized SAN solution. Our testing required greater than 3.3 GB/s I/O throughput for which we utilized HP StorageWorks EVA6000 storage systems for storage of both SAS data and temporary space. An EVA6000 solution provides high-availability, unparalleled performance and the capacity necessary for a typical SAS Grid computing implementation. The StorageWorks storage systems also ease management through automation and an intuitive user interface which helps lower complexity. With HP’s variety of storage systems there is a cost-effective solution that would meet typical SAS implementations. The EVA6000 storage array is a high-performance, high-capacity, and high-availability virtual array storage solution. It has four 4 Gb/s host fibre channel connections and 4 GB cache per controller pair, allowing it to support the most data intensive applications. Expandable to 56 TB, it supports mixed RAID-1, RAID-0+1, RAID-5, and cross RAID configurations, allowing the flexibility and expansion needed for the most dynamic environment. The EVA6000 storage array is replaced by the EVA6100, which is expected to provide similar or greater I/O throughput.

#### **SHARED FILESYSTEM REQUIREMENTS**

SAS® for Grid Computing requires a high performance clustered file system for the SAS data. We recommend HP PolyServe Matrix Server which provides a cluster structure for managing a group of network servers and a Storage Area Network (SAN) as a single entity. HP PolyServe software consolidates servers and isolates pools of storage into manageable, available and scalable utilities for database and file serving. The File Servicing Utility provides

symmetrical and simultaneous data access to shared data required by SAS® for Grid Computing. Our testing showed that we were able to scale the I/O (to more than 3.3 GB/s) attained from multiple EVA storage arrays, with no negative performance impact from sharing the storage in a clustered file system.

## PRODUCT FEATURES

HP Clustered File System provides the following features:

- Fully distributed data-sharing environment. The PSFS filesystem enables all servers in the cluster to directly access shared data stored on a SAN. After a PSFS filesystem has been created on a SAN disk, all servers in the cluster can mount the filesystem and subsequently perform concurrent read and/or write operations to that filesystem. PSFS is a journaling filesystem and provides online crash recovery.
- Availability and reliability. Servers and SAN components (FibreChannel switches and RAID subsystems) can be added to a cluster with minimal impact, as long as the operation is supported by the underlying operating system. HP Clustered File System includes failover mechanisms that enable the cluster to continue operations without interruption when various types of failures occur. If network communications fail-between any or all servers in the cluster, HP Clustered File System maintains the coherency and integrity of all shared data in the cluster.
- Cluster-wide administration. The Management Console (a Java-based graphical user interface) and the corresponding command-line interface enable you to configure and manage the entire cluster either remotely or from any server in the cluster.

## TUNING RECOMMENDATIONS AND BEST PRACTICES

One of the most commonly misconfigured subsystems in a SAS Grid Computing architecture is the I/O. SAS Grid Computing solutions are extremely I/O intensive, but many times the storage isn't architected correctly to accommodate the high I/O demand. This can lead to severe performance problems. As a result, here are some I/O tuning tips:

- Put SASWORK on the SAN, it can be managed by PolyServe without issue, the main concern is that it is striped across as many physical SAN disks as possible
- Place all SAS and grid files on the HP PolyServe shared cluster file system so that it is available to all servers
- For ideal price / performance create LUNs using RAID 5 with default options on the EVA
- Group all disks in the HP StorageWorks arrays together in a disk group
- Create LUNs that use the primary path to the EVA rather than the failover path
- Create volumes in HP PolyServe that stripe across all EVAs in the SAN and use 32 KB stripe and block size. Select all other defaults in PolyServe Matrix Server for the LUNs
- Stripe LUNs across all disks in an EVA
- Use 15k rpm disks

Our testing showed the following Linux settings helped. They may or may not help your performance but are listed here to assist with performance tuning options to try.

- Modify the device queue depth to 8 (echo 8 > /sys/block/sdx/device/queue\_depth), make sure you add this change to your startup file to ensure the change is permanent
- Change the block device size to 8k (blockdev --setra 8192 /dev/psv/psv\*)
- Modify your virtual memory settings to increase performance, some suggestions include:
  - Decrease how much the kernel favors swap over RAM: /sbin/sysctl -w vm.swappiness=20
  - Decrease how long a disk buffer can remain in RAM in a dirty state expressed in 100<sup>th</sup> of a second: /sin/sysctl -w vm.dirty\_expire\_centisecs=100
  - Set the poll interval between iterations of any pdflush daemons, expressed in 100<sup>th</sup> of a second: /sbin/sysctl -w vm.dirty\_writeback\_centisecs=500
  - Increase the amount of free memory that the kernel keeps available (in other words, not used in any kernel caches): /sbin/sysctl -w vm.min\_free\_kbytes=512000
  - Increase the amount of synchronizing of dirty pages back to the filesystem: /sbin/sysctl -w vm.dirty\_background\_ratio=20
  - Increase the limit at which processors which are generating dirty buffers will begin to synchronously write out data to disk, rather than relying on pdflush daemons to do it, expressed as a percentage of total system memory: /sbin/sysctl -w vm.dirty\_ratio=80
  - Change the default I/O scheduler to run "deadline" (add elevator=deadline in /boot/grub/menu.lst, for example, Kernel /vmlinuz-2.6.9-42.EL\_3\_5\_1\_0418PolyServe\_smp to

```
root=/dev/VolGroup00/LogVol00 elevator=deadline rhgb quiet)
```

SAS® Grid Computing performance can vary greatly depending on the number of jobs that are allowed per processor core. Our testing found that on HP ProLiant blade servers setting the number of job slots in SAS Grid Manager for each server should be equal to the number of cores on that system to deliver the best overall throughput.

## BENCHMARK TEST RESULTS

The SAS Grid Manager analytic benchmark scenario, developed by SAS and HP, replicated a real customer SAS Grid Manager implementations. The scenario generated a workload similar to a large multi-user ad-hoc analytics environment. This test suite was designed to provide an understanding of expected performance, scalability and system resources required to support SAS Grid Computing analytic activities.

### SCENARIO

This SAS Grid Manager analytic benchmark scenario utilizes real-world data volumes and structures of a typical SAS customer.

The SAS Grid Manager analytic benchmark scenario simulates the types of jobs received from various SAS clients such as display manager, batch, SAS® Data Integration Studio, SAS® Enterprise Miner™, SAS® Microsoft® Add-In and SAS® Enterprise Guide®. Many customer environments have large numbers of ad-hoc SAS users or jobs that utilize analytics in support of their company's day to day business activities. This benchmark scenario represents simultaneous users that replicate a mix of light to heavy users.

The SAS Grid Manager analytic benchmark scenario consisted of typical analytic jobs designed to replicate a heavy workload. These jobs were launched via a script which included time delays to simulate scheduled jobs and interactive users launching at different times.

### SAS JOB CHARACTERISTICS

The SAS® Grid Manager workload used for this test had the following characteristics:

- 50% CPU intensive jobs and 50% I/O intensive jobs
- Utilized SAS procedures including:  
SAS DATA step, PROC RISK, PROC LOGISTIC, PROC GLM (general linear model), PROC REQ, PROC SQL, PROC MEANS, PROC SUMMARY, PROC FREQ and PROC SORT
- SAS program input sizes up to 50 GB per job\*
- Input data types are text, SAS dataset and SAS transport files
- Memory use per job is up to 1 GB
- Each SAS job uses approximately 25 MB/s of I/O throughput
- Job runtimes were varied (short and long-running tasks)

\* File sizes were chosen to eliminate the caching effect of SAN storage and system RAM, larger file sizes would not have effected run results.

### TEST METHODOLOGY

The workload simulated a SAS Grid Manager analytic workload of 6,276 SAS jobs/day sustaining more than 3.3 GB/s I/O throughput evenly distributed across the SAS Grid Computing nodes. This distribution optimized throughput for each SAS job.

### TEST TOPOLOGY

The application environment under test included several different software components including:

- SAS® Foundation 9.1.3 with Service Pack 4
- SAS Grid Manager
- Red Hat Enterprise Linux ES (v.4 x86\_64)
- HP PolyServe Matrix Server version 3.5.1

Figure 1. Architectural diagram of test bed

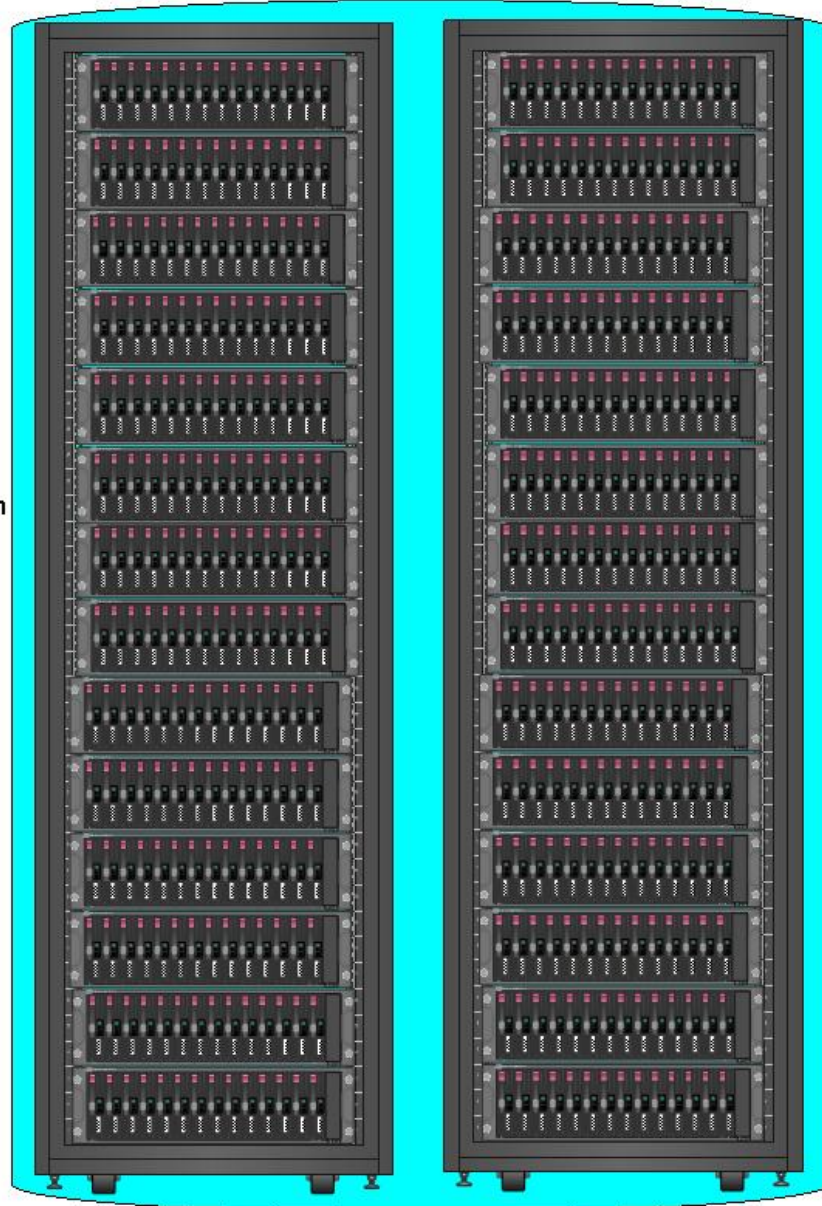
## SAS Grid Manager analytic benchmark

Fourteen HP ProLiant BL460c Blade Servers  
each with two quad-core 2.66 GHz cpu, 32 GB  
Linux Red Hat ES 4



HP PolyServe Matrix Server  
providing Shared Storage,  
Symmetrical data access  
Clustered File System

Eight HP StorageWorks  
EVA6000 Storage Arrays each  
with fifty-six 72 GB 15k rpm  
disks



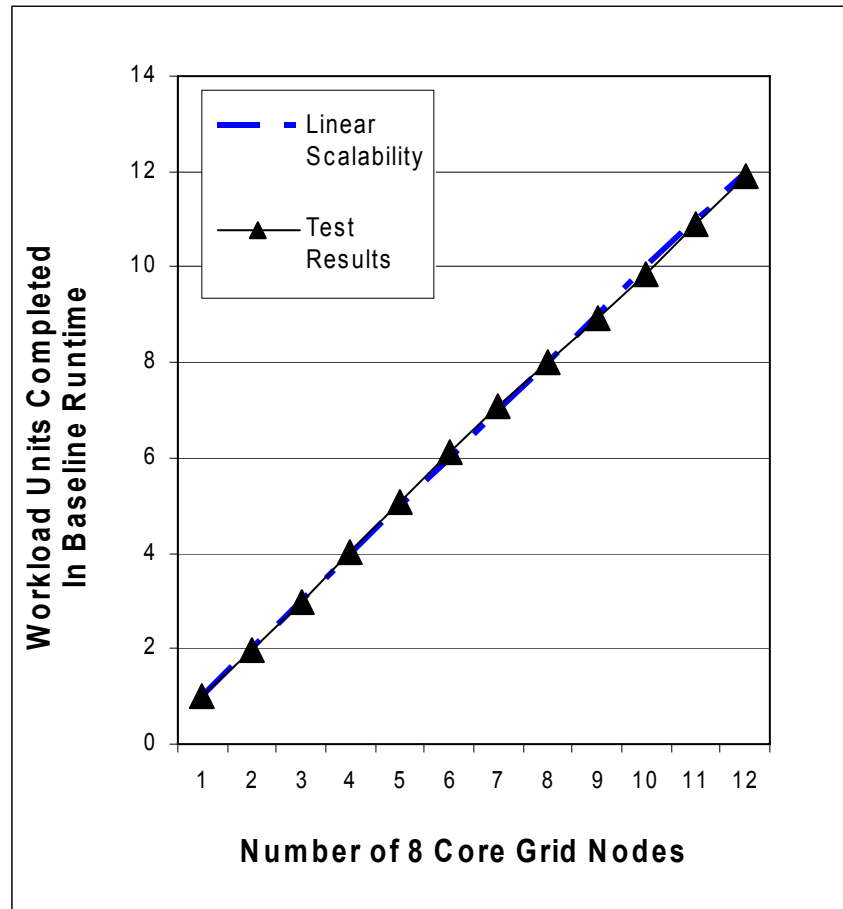
### TEST RESULTS

Testing results showed that during the 12 server tests we maintained a rate of over 6,276 jobs for an 8 hour period

with I/O throughput rates over 3.3 gigabytes/second. This workload simulated a mixed workload of users accessing the grid nodes.

We performed further tests to identify the scalability of SAS Grid Manager running on the HP solution described above. We found that the solution scaled linearly when the amount of resources was sufficient to support the user load. Below is a graph showing the scalability of SAS Grid Manager running on HP hardware.

Figure 2. Test scalability of SAS Grid Manager running on HP blade servers, HP StorageWorks EVA and a Clustered File system by PolyServe Matrix Server



#### TEST ANALYSIS SUMMARY

The execution of the SAS Grid Manager analytic workload demonstrated a high-performance, highly scalable HP BladeSystem with HP StorageWorks and a Clustered File system by PolyServe Matrix Server that is ideal for enterprise SAS Grid Computing.

The SAS Grid Manager analytic benchmark load clearly showed how HP BladeSystem servers are able to support the server demands of a typical solution for SAS Grid Computing. SAS Grid Manager jobs can be extremely I/O intensive. The HP StorageWorks EVA6000 storage arrays and Clustered File Systems by HP PolyServe Matrix Server is able to handle the I/O load and manage the storage resources required for an enterprise solution for SAS Grid Manager.

#### CONCLUSION

In our testing we found that given a properly architected HP solution for SAS Grid Manager it can scale up linearly. Important considerations include IO architecture and processor speeds as well as the clustered file system solution. HP and SAS Grid Manager can provide a highly efficient utilization of existing resources, as well as increased time to yield SAS results and cost-effective scaling possibilities for today and future needs.

**RECOMMENDED READING**

There are several HP papers that detail SAS Grid Computing solution benchmarks and reference architectures as well as more detailed solution information found below:

HP and SAS: [www.hp.com/go/sas](http://www.hp.com/go/sas)

HP Complimentary Customized Sizing: [sastech@hp.com](mailto:sastech@hp.com)

HP BladeSystem c-Class Server Blades: <http://www.hp.com/go/proliant>

HP Storage Array Systems: <http://h18006.www1.hp.com/storage/array systems.html>

HP PolyServe Shared Data Clustering for Linux and Windows®:  
<http://h71028.www7.hp.com/ActiveAnswers/cache/70994-0-0-225-121.html>

SAS Partner Directory: <http://www.sas.com/partners/directory/hp>

SAS Grid Computing: <http://support.sas.com/rnd/scalability/grid/>

**CONTACT INFORMATION**

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

Sandra Moulton  
Hewlett-Packard  
19055 Pruneridge Ave  
Cupertino, CA, 95014  
Work Phone: 415-762-0136  
Fax: 415-762-0136  
E-mail: [sandra.moulton@hp.com](mailto:sandra.moulton@hp.com)

**TRADEMARK INFORMATION**

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.

© 2007 Hewlett-Packard Development Company, L.P. The information contained herein is subject to change without notice. The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

Microsoft and Windows are U.S. registered trademarks of Microsoft Corporation. Intel and Xeon are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries. 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.

4AA1-6058ENW, October 2007