



## SAS® Grid Manager

SAS demonstrates superior scalability with SAS® Grid Manager and Sun StorageTek QFS shared file system

### SAS® Grid Manager demonstrates massive scalability and sustained throughput across multiple grid nodes.

- During testing, sustained I/O throughput rate was 5.97 terabytes (TB) per hour.
- This is a 72-processor, 144-core, SAS Grid Manager implementation.
- A Solaris 10 and Sun StorageTek QFS shared file system is used.
- SAS Data Integration Server processes are used, with 210 running simultaneously.
- World-record SAS Data Integration Server test results are achieved.

In this SAS Data Integration Server world-record and benchmarking scenario, SAS Grid Manager and Sun's StorageTek QFS shared file system are used to demonstrate complex scalability across multiple grid nodes. A high-performance and scalable file system is critical to the success of any SAS grid solution. Sun servers and storage along with SAS Grid Manager are used to build a large grid environment to achieve the current data integration world record for throughput with complex transformations. Data integration tasks provide an excellent scenario to stress-test grid architecture because of their heavy CPU, memory and I/O requirements. The test's workload matches the large-scale of SAS customers' application requirements, especially those that use the SAS Enterprise Intelligence Platform.

### Scenario

In the SAS Data Integration Server benchmarking scenario, simultaneous processes are launched to build out a star schema data mart in SAS Scalable Performance Data Server®. These processes first build out dimensions across the grid. More than 200 simultaneous processes are spawned on the grid to build out the star schema's fact table. Input files are stored on the QFS file system, where they are visible to all SAS grid nodes. Each grid node processes a subset of the data and stores it in the shared file system inside. After loading is complete, all individual tables are snapped together into a single large fact table through SAS dynamic clustering. The ability to execute this scenario

on the grid is provided by the high-performance shared file system. Each server node had multiple fiber connections to storage, allowing for a shared view of all data, and processes could be launched on any and all servers.

### Grid architecture

In the scenario, three server domains with 24 UltraSPARC IV dual-core processors were used as the compute nodes managed by a SAS Grid Manager control machine. A smaller secondary server is used as the QFS manager, QFS Metadata store, SAS Metadata Server and SAS Grid Manager. Sun StorageTek 6140 arrays were attached to each compute node in the grid to house the shared QFS file system. Each compute node contains a copy of SAS Enterprise Intelligence Platform, SAS Grid Manager and QFS clients. These processes communicate with the SAS Grid Manager control machine, coordinating locking, data location and what to execute.

Three physical domains of a Sun E25K server were used for the world-record benchmarking scenario, but the configuration can easily be changed to an equivalent grid of smaller footprint systems. Figure 1 shows a high-level overview of the architecture used in the scenario. The focus of the architecture is always placed on the throughput capability of the storage network and the ability of the storage subsystem to use it



**THE  
POWER  
TO KNOW®**

## Reference architecture

### System Configuration

Sun Fire E25K with:

- Bulk load test
- Processors: 72 x 1.95GHz USIV+
- Memory: 288 GB
- I/O boards: 12 total, four per domain
- Three-domain configuration for SAS Grid Manager enabled test
- Sun StorageTek S1 internal boot disk

QFS storage configuration

- 20 Sun StorageTek 6140 arrays
- Four fiber connections per array to compute nodes
- 16x146 GB 15K rpm / array

- Four four-disk (3+1) RAID 5 LUNs surfaced per array 408 GB raw storage yielded
- 32 TB usable storage for SAS Scalable Performance Data Server, SAS Metadata server and QFS metadata storage

QFS metadata storage

- One Sun StorageTek 3510 array

### Software and operating system configuration

- Solaris 10, Update 3
- Sun StorageTek QFS 4.5
- SAS Grid Manager
- SAS Enterprise Data Integration Server
- SAS Enterprise Intelligence Platform 9.1.3 Service Pack 4

## Performance results

In the benchmarking scenario, SAS software sustained an overall throughput of 5.97 TB per hour. This overall throughput rate of more than 11.8 TB per hour is more than sufficient to handle the most demanding grid applications. This total throughput rate equates to 3.3 GB a second, which is well beyond the total I/O requirements of the largest SAS customers' multiuser and data warehousing applications.

Linear scalability across the grid nodes and file system is maintained with a simultaneous increase in compute resources (servers) combined with I/O throughput capability and storage arrays. When required, QFS along with SAS Grid Manager can easily be scaled up to higher throughput rates.

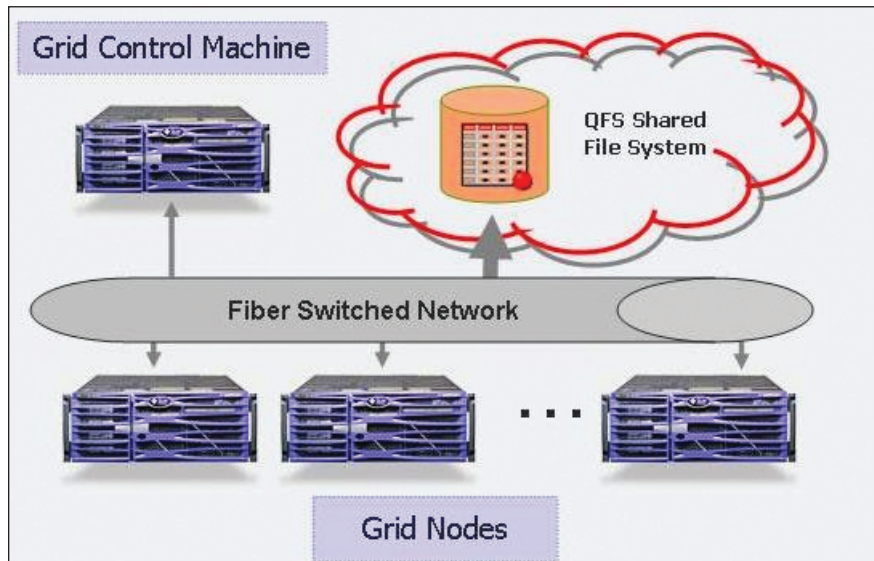


Figure 1 – High-level architecture diagram of design recommended for SAS® Grid Manager with Sun's StorageTek QFS shared file system.



THE  
POWER  
TO KNOW.

SAS Institute Inc. World Headquarters +1 919 677 8000

To contact your local SAS office, please visit: [www.sas.com/offices](http://www.sas.com/offices)

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. Copyright © 2007, SAS Institute Inc. All rights reserved. 103197\_460952.0907