



SAS® Grid Computing

Faster processing, increased resilience and improved hardware utilization

What is SAS Grid Computing

SAS Grid Computing delivers enterprise-class capabilities that enable SAS applications to automatically leverage grid computing, run faster and take optimal advantage of computing resources. With grid computing as an automatic capability, it is easier and more cost-effective to allocate compute-intensive applications appropriately across computing systems. SAS Grid Manager helps automate the management of SAS Computing Grids with dynamic load balancing, resource assignment and monitoring, and job priority and termination management.

Why is SAS Grid Computing important?

Customers can process high-volume SAS programs faster, improve hardware utilization and future proof computing infrastructures while increasing the resilience of SAS applications. Computing resources can be scaled out to cost-effectively add new users and meet fluctuating processing demands.

For whom is SAS Grid Computing designed?

It is designed for CIOs, IT managers, data center managers or grid computing architects seeking to improve performance of SAS jobs and optimize use of all appropriate resources. It also can be used by statisticians, analysts, data miners, application developers, DBAs and others seeking to reduce SAS processing time for faster data integration and analytic results.



**THE
POWER
TO KNOW®**

Meeting the computing demands of today's business environment with fixed and limited IT budgets is a challenge for every enterprise. Globalization, increasing competition and new technologies, such as RFID, are increasing the need to process more data in less time. Buying the latest and greatest servers (scaling up) to meet peak-demand computing loads is one way to meet those needs, but it can also be costly and inefficient.

Grid computing provides an innovative, strategic approach to building and managing a lower-cost IT infrastructure that can flex to meet rapidly changing and growing computing requirements. SAS Grid Computing offers a cost-effective solution for customers who want to accelerate the processing of SAS programs or increase the scale or scope (number of users, size of data sets, frequency of analysis) of a particular SAS application.

In addition to shortening time to intelligence through faster processing, SAS Grid Computing allows customers to maximize use of spare CPU cycles or add SAS applications to commodity computing grid infrastructures, resulting in significant improvements in average processing speed for SAS programs and a reduction in computing resource costs.

As query, reporting, OLAP and analytic capabilities grow in strategic importance and encompass thousands of workers and large quantities of data, the ability to cost-effectively scale a BI system to meet peak demand using grid computing becomes a competitive advantage.

Key benefits

- Increase compute hardware ROI and improve efficiency by fully exploiting existing computing resources.** SAS Grid Manager allows organizations to fully utilize all computing resources in the most efficient manner. With a comprehensive set of advanced load balancing and grid management functionality, it automates the management and optimization of application processing in a grid computing environment.
- Speed up applications using existing IT infrastructure.** Multiprocessing capabilities allow workflows to be processed in parallel on multiple machines. By using multiple CPUs, each running a part of a program with its own hardware resources, the time required to execute a SAS program is reduced substantially. SAS programs best suited for parallel processing are those with large data sets and long run times and those with replicate runs of independent tasks and independent tasks running against large data sets. Analytical and data integration processes will yield faster results, accelerating decision making across the enterprise.
- Future proof and increase flexibility of enterprise computing infrastructure by scaling out.** Using grid technology, SAS applications can be cost-effectively scaled out, adding capacity in single processing units for incremental IT spend. This reduces processing time as well as capital expenditures on ever-larger, stand-alone servers. Common repository enables the centralized storage, management and reuse of work based on user authorizations, reducing both development and maintenance time.

Product overview

SAS Grid Computing allows a variety of SAS applications to detect and leverage available grid environments. Compute-intensive programs can be allocated to run in pieces across grid environments, enabling organizations to optimize capacity and performance of hardware resources and produce faster results more cost-effectively. Users can submit SAS jobs to a shared pool of resources rather than an individual server in order to balance workload and better manage the SAS environment.

By combining proven industry-leading grid computing middleware from Platform Computing with the multiprocessing capabilities of SAS/CONNECT software, SAS Grid Computing pro-

vides automatic grid management and monitoring with the ability to spawn and distribute SAS programs across the various computing resources within the grid. SAS/CONNECT automatically builds and maintains the conduits between computers needed to move programs across CPUs.

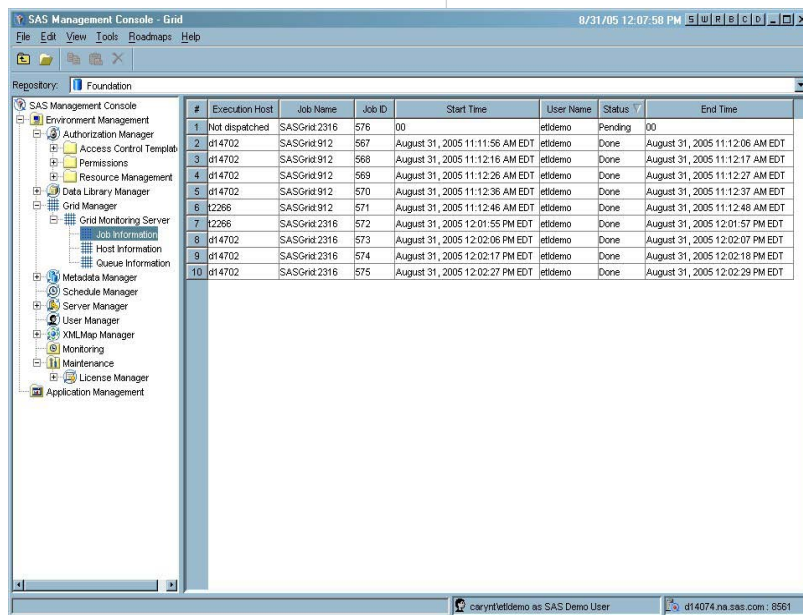
Grid-enabled SAS

SAS Data Integration Studio 3.3 and SAS Enterprise Miner 5.2 are the first SAS products that are automatically tailored for parallel processing in a grid computing environment. To achieve maximum processing efficiency with minimum user intervention, these programs detect the grid environment at the time of execution. Not only does

this speed results for these particular applications, the grid-enabled logic that is produced can be saved as a SAS Stored Process for use by other SAS BI clients to produce faster results for more users as cost-effectively as possible. Other SAS applications will add automatic grid-enablement in future product releases. In addition, all SAS programs can take advantage of a grid computing environment with the addition of programming syntax and structure that allows program steps (subtasks) to be executed in parallel.

Automatic grid management and monitoring

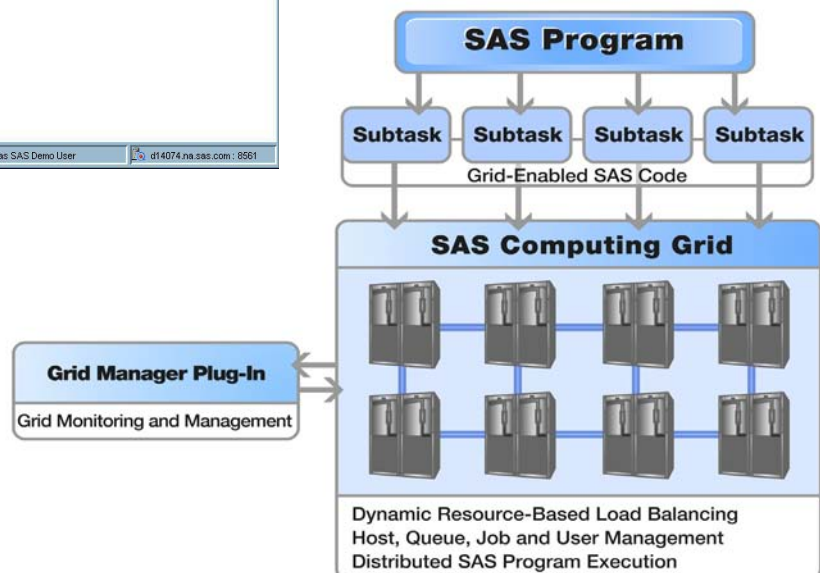
SAS Grid Manager includes the Grid Manager plug-in, which is integrated into the SAS Management Console. This plug-in provides a visual interface for monitoring and managing hosts, jobs, users and queues in the SAS Computing Grid. Resource and activity information can be filtered allowing users to focus on specific aspects of grid operations. Additionally, SAS Grid Manager provides job priority management and termination. SAS Grid Manager monitors grid comput-



#	Execution Host	Job Name	Job ID	Start Time	User Name	Status	End Time
1	Not dispatched	SASGrid.2316	576	00	etldemo	Pending	00
2	d14702	SASGrid.912	567	August 31, 2005 11:11:56 AM EDT	etldemo	Done	August 31, 2005 11:12:06 AM EDT
3	d14702	SASGrid.912	568	August 31, 2005 11:12:16 AM EDT	etldemo	Done	August 31, 2005 11:12:17 AM EDT
4	d14702	SASGrid.912	569	August 31, 2005 11:12:26 AM EDT	etldemo	Done	August 31, 2005 11:12:27 AM EDT
5	d14702	SASGrid.912	570	August 31, 2005 11:12:36 AM EDT	etldemo	Done	August 31, 2005 11:12:37 AM EDT
6	k2266	SASGrid.912	571	August 31, 2005 11:12:46 AM EDT	etldemo	Done	August 31, 2005 11:12:48 AM EDT
7	k2266	SASGrid.2316	572	August 31, 2005 12:01:55 PM EDT	etldemo	Done	August 31, 2005 12:01:57 PM EDT
8	d14702	SASGrid.2316	573	August 31, 2005 12:02:06 PM EDT	etldemo	Done	August 31, 2005 12:02:07 PM EDT
9	d14702	SASGrid.2316	574	August 31, 2005 12:02:17 PM EDT	etldemo	Done	August 31, 2005 12:02:18 PM EDT
10	d14702	SASGrid.2316	575	August 31, 2005 12:02:27 PM EDT	etldemo	Done	August 31, 2005 12:02:29 PM EDT

(Above) The SAS Management Console and the Grid Manager plug-in provide an easy-to-use interface for job monitoring and management.

(Right) SAS Grid Computing takes SAS programs and divides them into subtasks for parallel processing. The programs are then distributed in the computing grid using Platform Computing middleware to provide dynamic load balancing. Hosts, users and jobs are monitored and tracked within the SAS Grid Manager Plug-in to the SAS Management Console.



ing resources and allows for detection of individual equipment failures to help ensure that mission-critical programs are completed.

Dynamic resource-based load balancing

By incorporating proven industry-leading grid computing middleware from Platform Computing, SAS Grid Computing delivers enterprise-class dynamic load balancing. This automates the management and optimization of SAS grids and provides execution resilience by efficiently distributing the processing of SAS programs across multiple CPUs.

It also facilitates job, queue, host and user management. The Grid Manager plug-in provides a visual interface for centrally managing intelligent grid computing functionality, including the automatic identification, allocation, management and optimization of computing resources and program flows. SAS Grid Computing enables automatic program scheduling across two or more host machines.

Scale out to run more, faster analyses and take advantage of all computing resources

By making grid computing an automatic capability within multiple intelligence applications, SAS helps organizations reduce data processing times so larger volumes of data can be integrated, cleansed and analyzed in less time. SAS Grid Computing can schedule a wide variety of SAS jobs across grid environments for optimal resource utilization and faster processing. In today's international organizations, nightly batch processing windows no longer exist. Someone, somewhere needs information 24/7, and the faster, more efficiently data can be loaded and analyzed, the more it can be utilized for competitive advantage.

Key Features

Grid-enabled SAS

- SAS Data Integration Studio 3.3 and SAS Enterprise Miner 5.2 are automatically tailored for parallel processing in a grid environment.
- Automates session spawning and distributed processing of SAS programs across a set of diverse computing resources.
- Provides greater resilience for mission-critical applications.
- Speeds up processing of applicable SAS programs and applications and provides more efficient computing resource utilization.
- Automatic selection of SMP or grid environment at run time.
- Grid-enabled SAS logic can be saved as a SAS Stored Process, which can then be used by all SAS BI clients and analytic applications.

Automatic grid monitoring and management

- Integrated with SAS@9 Enterprise Intelligence Platform via the SAS Management Console to provide enterprise-class grid computing infrastructure administration and management.
- SAS Grid Manager plug-in provides a visual interface to monitor and manage hosts, jobs, users and queues in the SAS computing grid.
- Provides centralized monitoring and management of grid status, program attributes, program priority, program termination, host queues and multiple grids.
- Provides real-time monitoring of programs and computing resources.
- Manages multiple grids.
- Resource and activity information can be filtered allowing users to focus on specific aspects of grid operations.
- Metadata driven.

Dynamic resource-based load balancing

- Improves efficiency of program distribution and CPU utilization through dynamic, resource-based load balancing.
- Leverages industry-leading grid computing middleware from Platform Computing.
- Automates management and optimization of SAS Computing Grids.
- Flexibly allows the addition of peak load computing capacity one unit (CPU) at a time.
- Provides execution resilience by distributing the processing of SAS programs across multiple computers.
- Enables virtualization of resources.
- Provides job, queue, host and user management.
- Enables job prioritization by rules-based job queues.
- Comprehensive centrally managed intelligent grid computing functionality: automatic identification, allocation, management and optimization of computing resources and program flows.
- Ability to automate program scheduling across two or more host machines.

Scaling out

- Builds on the proven multiprocessing functionality of SAS 8.2.
- Cost-effectively increase the scale or scope of compute-intensive programs, add more users, and run more analyses against larger volumes of data with shorter processing times.
- Run new analyses because more time and resources are available.
- Improve application recovery time – if a job runs faster and it fails, there is more time for recovery and restart.
- Schedule a wide variety of SAS jobs across grid environments for optimal resource utilization and faster processing.

SAS® Grid Computing

Grid computing environment

SAS Servers, including Base SAS and SAS Metadata Server, can be installed on one or more hardware systems in a multi-tier configuration.

- AIX (64-bit): Release 5.1, 5.2, 5.3 on POWER
- HP-UX PA-RISC: Release 11i Version 1, 2 and 3 *
- HP-UX Itanium: Release 11i Version 1, 2 and 3
- Linux for Intel (x86-32): Red Hat Linux 8.0, RHAS 2.1, RHEL 3.0 and 4.0, SuSE 8 and 9
- Linux for Itanium (64-bit): Red Hat RHEL 3.0
- Solaris on SPARC: Version 8, 9, 10 on SPARC
- Solaris on x64: Version 10
- Windows (x86-32): Windows 2000 Professional, Windows XP Professional, Windows 2000 Server, Windows Server 2003, Windows NT 4 Server *

* Note: SAS Metadata Server must be installed on an alternate supported platform.

Required/optional software

- Base SAS
- SAS/CONNECT
- Any application-specific software required on the servers
- SAS Grid Manager



**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**

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. 102295_443887.0507