

Chapter 1

What Is SAS Grid Computing?

SAS Grid Computing Basics	3
SAS Grid Topology	4
What Types of Processing Does a Grid Support?	6
Multi-User Workload Balancing	6
Parallel Workload Balancing	6
Distributed Enterprise Scheduling	7
SAS Applications That Support Grid Processing	7
What Business Problems Can a Grid Solve?	8
Many Users on Single Resource	8
Increased Data Growth	8
Running Larger and More Complex Analysis	9
Need for a Flexible IT Infrastructure	9

SAS Grid Computing Basics

A SAS grid computing environment is one in which SAS computing tasks are distributed among multiple computers on a network, all under the control of SAS Grid Manager. In this environment, workloads are distributed across a grid of computers. This workload distribution enables the following functionality:

Workload balancing

enabling multiple users in a SAS environment to distribute workloads to a shared pool of resources.

Accelerated processing

allowing users to distribute subtasks of individual SAS jobs to a shared pool of resources. The grid enables the subtasks to run in parallel on different parts of the grid, which completes the job much faster.

Scheduling jobs

allowing users to schedule jobs, which are automatically routed to the shared resource pool at an appropriate time.

SAS Grid Manager provides load balancing, policy enforcement, efficient resource allocation, and prioritization for SAS products and solutions running in a shared grid environment. It also separates the SAS applications from the infrastructure used to execute the applications. This enables you to transparently add or remove hardware resources as needed and also provides tolerance of hardware failures within the grid infrastructure. SAS Grid Manager integrates the resource management and scheduling capabilities of the

Platform Suite for SAS with the SAS 4GL syntax and subsequently with several SAS products and solutions.

SAS Grid Manager includes these components, as illustrated in Figure 1.1 on page 5. :

Grid Manager plug-in

a plug-in for SAS Management Console that provides a monitoring and management interface for the jobs and resources in your grid

grid syntax

the SAS syntax necessary to grid-enable the SAS workload

Platform Suite for SAS

components provided by Platform Computing to provide efficient resource allocation, policy management, and load balancing of SAS workload requests

The Platform Suite for SAS includes these components:

Load Sharing Facility (LSF)

dispatches all jobs submitted to it, either by Process Manager or directly by SAS, and returns the status of each job. LSF also manages any resource requirements and performs load balancing across machines in a grid environment.

Process Manager (PM)

this is the interface used by the SAS scheduling framework to control the submission of scheduled jobs to LSF and manage any dependencies between the jobs.

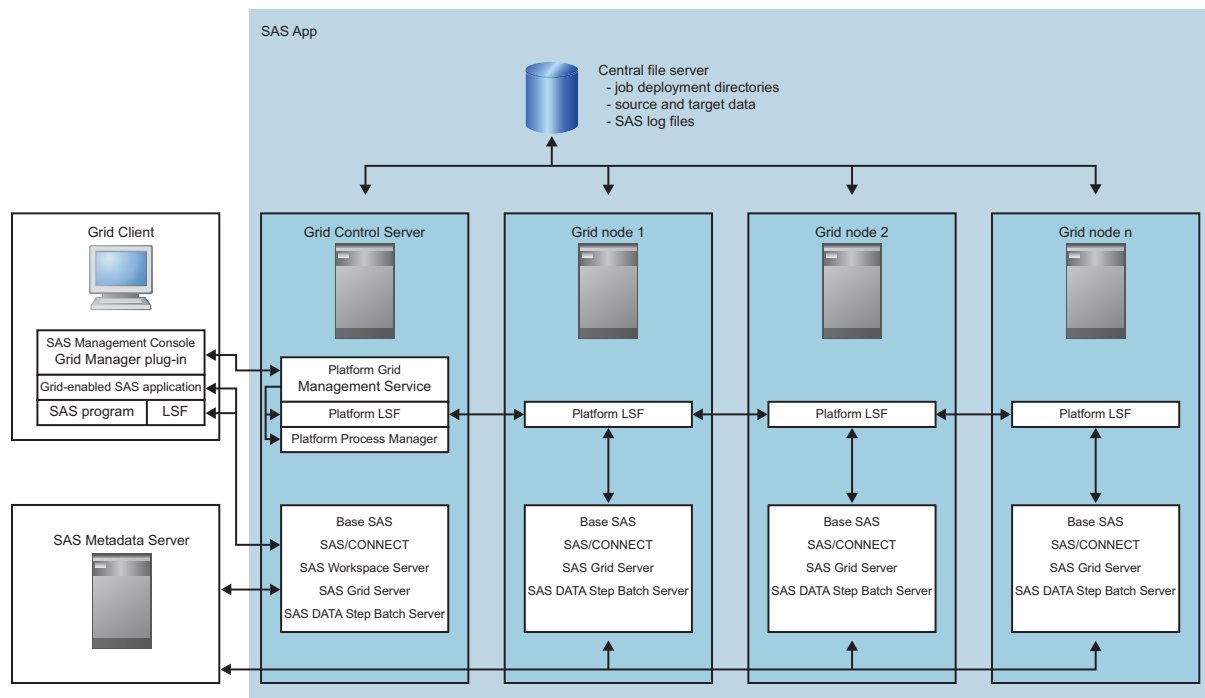
Grid Management Services (GMS)

this is the interface to the Grid Manager plug-in in SAS Management Console. It provides the run-time information about jobs, hosts and queues for display in SAS Management Console.

SAS Grid Topology

As illustrated below, a grid configuration consists of these main components:

Figure 1.1 Grid Topology



Grid control server

this machine controls distribution of jobs to the grid. Any machine in the grid can be designated as the grid control server. Also, you can choose whether to configure the grid control server as a grid resource capable of receiving work. This machine must contain the grid middleware software (such as Platform Suite for SAS). The grid control server also configures a SAS workspace server so that SAS Data Integration Studio and SAS Enterprise Miner can run programs that take advantage of the grid.

Grid node

these machines are grid computing resources that are capable of receiving the work that is being distributed to the grid. The number of nodes in a grid depends on the size, complexity, and volume of the jobs that will be run by the grid. You can add or remove nodes as specified by your business needs. Each grid node must contain Base SAS, SAS/CONNECT, Platform LSF (or other grid middleware software), and any applications and solutions needed to run grid-enabled jobs.

Central file server

this machine is used to store data for jobs that run on the grid. In order to simplify installation and ease maintenance, you can also install the SAS binaries on the central file server.

Metadata server

this machine contains the metadata repository that stores the metadata definitions needed by SAS Grid Manager and other SAS applications and solutions that are running on the grid. Although it is recommended that the SAS Metadata Server be on a dedicated machine, it can be run on the grid control server.

SAS Management Console

this application is used to manage the definitions in the metadata repository, to submit jobs to the grid through the Schedule Manager plug-in, and to monitor and manage the grid through the Grid Manager plug-in.

Grid clients

submits jobs to the grid for processing, but is not part of the grid resources available to execute work.

Examples of grid clients are:

- a SAS Data Integration Studio client. Platform LSF is not required on this client machine.
- a SAS Enterprise Miner client. Platform LSF is not required on this client machine.
- a SAS Management Console client, that uses the Schedule Manager plug-in or another application to schedule SAS workflows. Platform LSF is not required on this client machine.
- a SAS foundation install that is used to run a program that submits work to the grid. The submitted work can be entire programs or programs broken into parallel chunks. This client must have Base SAS, SAS/CONNECT, and Platform LSF installed. Platform LSF is required to submit the SAS workload to the grid.
- a SAS Grid Manager Client Utility. SAS is not required to be installed on this client, but Platform LSF is required to submit the SAS workload to the grid.

What Types of Processing Does a Grid Support?

Multi-User Workload Balancing

Most organizations have many SAS users performing a variety of query, reporting, and modeling tasks and competing for the same resources. SAS Grid Manager can help bring order to this environment by providing capabilities such as the following:

- specifying which jobs get priority
- deciding the share of computing resources used by each job
- controlling the number of jobs that are executing at any one time

In practice, SAS Grid Manager submits work to the grid middleware, which acts as a gatekeeper for the jobs submitted to servers. As jobs are submitted, the middleware (such as Platform LSF) does them out to grid nodes, preventing any one machine from being overloaded. If more jobs are submitted than can be run at once, the grid middleware submits as many jobs as can be run. The middleware then holds the rest in a queue until resources are free. The grid middleware can also use job priority to determine whether a job is run immediately or held in a queue.

The application user notices little or no difference when working with a grid. For example, users can define a key sequence to submit a job to a grid rather than running it on their local workstation. Batch jobs can be run using wrapper code that adds the commands needed to run the job in the grid. SAS Enterprise Guide applications can be set up to automatically insert the code needed to submit the job to the grid.

Parallel Workload Balancing

Some SAS programs consist of subtasks that are independent units of work and can be distributed across a grid and executed in parallel. You can use SAS syntax to identify the parallel units of work in these programs, and then use SAS Grid Manager to distribute the

programs across the grid. Using parallel workload balancing can substantially accelerate the entire application.

Applications such as SAS Data Integration Studio, SAS Risk Dimensions, and SAS Enterprise Miner are often used for iterative processing. In this type of processing, the same analysis is applied to different subsets of data or different analysis is applied to a single subset of data. Using SAS Grid Manager can improve the efficiency of these processes, because the iterations can be assigned to different grid nodes. Because the jobs run in parallel, the analysis completes more quickly and with less strain on computing resources.

Distributed Enterprise Scheduling

The Schedule Manager plug-in for SAS Management Console provides the ability to schedule user-written SAS programs as well as jobs from numerous SAS applications. You can schedule the jobs and programs to run when specified time or file events occur. If jobs are scheduled using the scheduling capabilities provided by Platform Suite for SAS, the jobs can be processed on a grid without any change to the scheduling process. This capability provides further control over use of computing resources, because you can use the scheduling capability to control when a job runs and the SAS Grid Manager capability to determine which computing resource processes the job.

SAS Applications That Support Grid Processing

The following table lists the SAS applications that currently support grid processing and the type of processing that each supports.

Table 1.1 *Grid Support in SAS Applications*

SAS Application	Multi-User Workload Balancing	Parallel Workload Balancing	Distributed Enterprise Scheduling
Any SAS program	yes	yes, with modifications	yes
SAS Enterprise Guide	yes, with modifications		
SAS Data Integration Studio	yes	yes	yes
SAS Enterprise Miner	yes	yes	
SAS Risk Dimensions	yes	yes, with modifications	
SAS Web Report Studio			yes
SAS Marketing Automation			yes
SAS Marketing Optimization			yes

SAS Application	Multi-User Workload Balancing	Parallel Workload Balancing	Distributed Enterprise Scheduling
SAS JMP/Genomics		yes	
SAS Demand Forecasting for Retail		yes	
SAS products or solutions that use workspace server load balancing	yes		
SAS stored processes	yes, with limitations	yes, with limitations	

For a current list of SAS applications that support grid processing, see <http://support.sas.com/rnd/scalability/grid/index.html>.

What Business Problems Can a Grid Solve?

Many Users on Single Resource

An organization might have multiple users submitting jobs to run on one server. When the environment is first configured, the server might have been sufficient to handle the number of users and jobs. However, as the number of users submitting jobs grows, the load on the server grows. The increased load might lead to slower processing times and system crashes. In a SAS grid environment, jobs are automatically routed to any one of the servers on the grid. This spreads the computing load over multiple servers, and diminishes the chances of a server becoming overloaded. If the number of jobs exceeds the resources available, the jobs are queued until resources become available. If the number of users continues to increase, you can increase capacity by adding servers to the grid.

Increased Data Growth

Your organization might have a process running to analyze a certain volume of data. Although the server that is processing the job is sufficient to handle the current volume of data, the situation might change if the volume of data increases. As the amount of data increases, the load on the server increases, which can lead to longer processing times or other problems. Changing to a larger-capacity server can involve considerable expense and service interruption.

A SAS grid environment can grow to meet increases in the amount of data processed. If the volume of data exceeds the capacity of a server on the grid, the processing load can be shared by other grid servers. If the volume continues to increase, you can add servers to the grid without having to make configuration changes to your processes. Adding servers to the grid is also more cost-effective than replacing a single large server, because you can add smaller servers to handle incremental increases in data volume.

Running Larger and More Complex Analysis

Your organization might have a process running to perform a certain level of analysis on data. If you want to increase the complexity of the analysis being performed, the increased workload puts a greater strain on the processing server. Changing the computing power of the server involves considerable expense and interrupts network availability.

Using a SAS grid environment enables you to add computing power by adding additional computers on the grid. The analysis job can be divided up among the grid nodes, which enables you to perform more complex analysis without increasing the load on any single machine.

Need for a Flexible IT Infrastructure

Your organization's ability to perform the data analysis you need depends on a flexible computing infrastructure. You must be able to add needed resources quickly and in a cost-effective manner as the load increases. You must also be able to handle maintenance issues (such as adding or replacing resources) without disrupting your work. A SAS grid environment enables you to maintain a flexible infrastructure without disrupting your operations.

As your data-processing needs grow, you can incrementally add computing resources to your grid by adding smaller, less-expensive servers as new server nodes. This ability prevents you from having to make large additions to your environment by adding large and expensive servers.

When you need to perform maintenance on machines in the grid, the grid can still operate without disruption. When you take the servers offline for maintenance or upgrades, SAS Grid Manager routes to work to the machines that are still online. Users who send work to the grid for processing do not have to change their way of working. Work that is sent to the grid is processed just as before.

Likewise, the SAS grid environment adapts if a computer fails on the grid. Because SAS Grid Manager automatically avoids sending work to the failed machine, the rest of the grid is still available for processing and users do not see any disruption.

