Grid Computing in SAS® 9.4
Fourth Edition
# Contents

*What’s New in SAS Grid Manager 9.4* ........................................... vii

## PART 1  Introduction to SAS Grid Computing 1

*Chapter 1 • What Is SAS Grid Computing?* ........................................ 3
  SAS Grid Computing Basics ......................................................... 3
  SAS Grid Topology ........................................................................ 4
  What Types of Processing Does a Grid Support? .............................. 8
  What Business Problems Can a Grid Solve? .................................... 10

## PART 2  Grid Computing for SAS Using Platform Suite for SAS 13

*Chapter 2 • Planning and Configuring a Grid Environment* ................ 15
  Installation and Configuration Overview ........................................... 15
  Configuring the File Server ............................................................ 16
  Installing Platform Suite for SAS .................................................... 16
  Configuring the Grid Control Server ............................................... 16
  Configuring Platform Web Services ............................................... 21
  Configuring the Grid Nodes ........................................................... 23
  Configuring Client Applications ..................................................... 23
  Modifying Grid Monitoring Server Definitions ................................. 23
  Installing and Configuring SAS Environment Manager in a Grid Environment .................................................. 25

*Chapter 3 • Managing the Grid* .................................................... 27
  Overview of Grid Management ....................................................... 27
  Modifying Configuration Files with Platform RTM for SAS ................ 28
  Specifying Job Slots for Machines ................................................. 29
  Using Queues .............................................................................. 31
  Defining and Specifying Resources ................................................. 38

*Chapter 4 • Using Grid Management Applications* .......................... 41
  Using SAS Environment Manager ................................................... 41
  Using the SAS Grid Manager Agent Plug-in ..................................... 43
  Using the SAS Grid Manager Module ............................................. 45
  Using Platform RTM for SAS ......................................................... 55
  Using the Grid Manager Plug-in for SAS Management Console ........ 55

*Chapter 5 • High Availability* ...................................................... 63
  High Availability and SAS Grid Manager ........................................... 63
  Setting Up High Availability for Critical Applications ....................... 64
  Restarting Jobs ............................................................................. 67
### PART 5  SAS Grid Language Reference  

<table>
<thead>
<tr>
<th>Chapter 11 • SAS Functions for SAS Grid</th>
<th>137</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dictionary</td>
<td>137</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 12 • SASGSUB Command</th>
<th>153</th>
</tr>
</thead>
<tbody>
<tr>
<td>SASGSUB Overview</td>
<td>153</td>
</tr>
<tr>
<td>Disabling SASGSUB Actions</td>
<td>154</td>
</tr>
<tr>
<td>Dictionary</td>
<td>154</td>
</tr>
</tbody>
</table>

### PART 6  Appendix  

<table>
<thead>
<tr>
<th>Appendix 1 • Supported Job Options - Platform Suite for SAS</th>
<th>169</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Appendix 2 • Supported Job Options - SAS Grid Manager for Hadoop</th>
<th>171</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Recommended Reading</th>
<th>173</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glossary</td>
<td>175</td>
</tr>
<tr>
<td>Index</td>
<td>179</td>
</tr>
</tbody>
</table>
Contents
What’s New in SAS Grid Manager
9.4

Overview

SAS Grid Manager has the following new features and enhancements:

• Support for grid options sets has been added.
• Support for grid-launched workspace servers has been added.
• New options have been added to the SAS Grid Manager Client Utility, including support for grid options sets, support for interactive SAS sessions and command processing, and support for waiting for results from grid jobs.
• New job options for Platform Suite for SAS have been added.
• SAS Grid Manager for Hadoop adds support for SAS grid processing co-located on a Hadoop cluster.
• A SAS Grid Manager agent plug-in has been added to SAS Environment Manager for grids that use Platform Suite for SAS.
• A SAS Grid Manager module has been added to SAS Environment Manager for grids that use Platform Suite for SAS.
• The GRDSVC_HOSTLIST and GRDSVC_OPTSETS functions have been added.

Support for Grid Options Sets

A grid options set is a collection of grid options, SAS options, and required resources that are associated with a particular SAS client application. The purpose of a grid options set is to enable a SAS grid administrator to define a collection of options in SAS metadata. These options map to one or more SAS client applications and are automatically applied to workloads submitted to the grid based on the identity of the user accessing the client application. Some of the client applications that support grid options sets include the following:

• SAS Data Integration Studio
• SAS Grid Manager Client Utility
• SAS sessions that use the grdsvc_enable statement to submit jobs to the grid
• Grid-launched servers
Support for Grid-Launched Servers

With grid-launched servers, when an application requests a SAS workspace server from the object spawner, the spawner passes the request on to SAS Grid Manager. SAS Grid Manager checks the machines on the grid, using the grid’s policy management to determine which host is the least busy, and starts the server. Because the load-balanced servers are started by SAS Grid Manager, jobs on these servers appear as LSF jobs.

In the first maintenance release after SAS 9.4, stored process servers and pooled workspace servers can also be grid-launched.

New Options for the SAS Grid Manager Client Utility

The following new options have been added to the SAS Grid Manager Client Utility (SASGSUB):

-GRIDOPTSET
   specifies the name of a grid options set.

-GRIDWAITTIMEOUT
   specifies that the SAS Grid Manager Client Utility waits until either the job has completed running or a specified time-out value is reached.

-GRIDWAITRESULTS
   specifies that the SAS Grid Manager Client Utility waits until the job has completed running (either successfully or with an error) and then returns the results of the processing.

_PROMPT_
   when used as a value for the METAUSER or GRIDUSER parameters, specifies that the SAS Grid Manager Client Utility prompts the user for a user ID.

In the first maintenance release after SAS 9.4, support has been added to the SAS Grid Manager Client Utility to support interactive mode. These options enable you to start an interactive SAS session on the grid, using either SAS Display Manager mode or UNIX line mode, or to run an arbitrary command. It also enables you to monitor the output from SAS or a command running in batch mode.

-GRIDRUNSASLM
   specifies that the SASGSUB command runs a SAS session in line mode and that standard output and standard error are displayed by SASGSUB. SASGSUB also redirects standard input to the SAS session running on the grid. The grid must be on UNIX.

-GRIDRUNSASDMS
   specifies that the SASGSUB command runs a SAS session in SAS Display Manager mode. Standard output and standard error are displayed by SASGSUB, and the X Windows session is directed to the host and port specified on the -GRIDRUNSASDMS argument. The grid must be on UNIX.
GRIDRUNCMDINT
specifies that an arbitrary command is run on the grid in interactive mode. Standard output and standard error are displayed by SASGSUB. Standard input is directed to the program running on the grid.

GRIDWATCHOUTPUT
when SASGSUB is used for batch processing, specifies that the log output and the list output of the batch job are displayed on the user’s machine.

In the third maintenance release after SAS 9.4, options have been added to the SAS Grid Manager Client Utility to control whether the job directory, job log, and output listing are retained or deleted after processing.

GRIDWAITLOGLIST
specifies that, after a job has completed running, and after the job log and output listing are obtained, the job directory is deleted.

GRIDWAITNORESULTS
specifies that, after a job has completed running and the results of the processing are returned, the job log, output listing, and job directory are deleted.

GRIDWAITNORESULTSNOSLOG
specifies that, after a job has completed running and the results of the processing are returned, the job log, output listing, and job directory are deleted, and the log and list files are not copied to the current directory.

New Job Options

The following job options that are supported by Platform Suite for SAS have been added:

app
specifies the LSF application name to use.

maxpendingtime
specifies how long a job waits in the pending state before being terminated.

jobname
specifies a job name to be used for all jobs.

SAS Grid Manager for Hadoop

In the third maintenance release after SAS 9.4, SAS Grid Manager for Hadoop has been added. SAS Grid Manager for Hadoop provides workload management, accelerated processing, and scheduling of SAS analytics co-located on a Hadoop cluster. SAS Grid Manager for Hadoop leverages YARN to manage resources and distribute SAS analytics to a Hadoop cluster running multiple applications. It integrates with Oozie, which provides scheduling capability for SAS workflows. All of the existing SAS Grid syntax, submission modes, and integration with other SAS products and solutions is supported by SAS Grid Manager for Hadoop.
SAS Grid Manager Agent Plug-in for SAS Environment Manager

In the third maintenance release after SAS 9.4, a SAS Grid Manager agent plug-in has been added to SAS Environment Manager. The plug-in monitors your grid resources over time and generates events and alerts based on the monitored information.

The SAS Grid Manager agent plug-in provides metric data for the grid cluster and for individual grid hosts, and uses that data to perform these functions:

- display the current state of grid resources
- graph the data over time, providing a historical view and enabling you to see how the data changes
- create alerts that notify you whenever a selected measurement reaches a selected state

SAS Grid Manager Module for SAS Environment Manager

In the third maintenance release after SAS 9.4, a SAS Grid Manager module has been added to SAS Environment Manager. The module provides some of the same monitoring and management functions as Platform RTM for SAS, so you can monitor and manage your grid using the same application that you use to manage the rest of your SAS environment.

The SAS Grid Manager module enables you to configure and perform actions on grid resources and high availability applications, and provides a view of your grid at the current moment in time. The functions provided by the module are divided into these three areas:

**Monitoring**

- perform actions on grid resources (such as opening and closing hosts and activating and deactivating queues)
- view information for the cluster, grid hosts, queues, jobs, high-availability applications, and audit logs
- manage high availability applications

**LSF Configuration Manager**

modify the configuration information for the LSF cluster, including definitions for hosts, queues, users, administrators, resource limits, and cluster parameters

**HA Configuration Manager**

modify the configuration information for high availability applications

*Note:* The SAS Grid Manager module replaces the SAS Grid Manager server plug-in that was added to SAS Environment Manager in the second maintenance release after SAS 9.4.
New Functions

In the third maintenance release after SAS 9.4, these functions have been added:

**GRDSVC_HOSTLIST**
This function returns a list of hosts that are available to run grid jobs.

**GRDSVC_OPTSETS**
Returns a list of valid options sets for a specified combination of SAS Application Server, SAS client application, and user.
Part 1

Introduction to SAS Grid Computing

Chapter 1

Chapter 1
What Is SAS Grid Computing?

SAS Grid Computing Basics ................................................................. 3
SAS Grid Topology .............................................................................. 4
  SAS Grid Configuration .................................................................... 4
  Grid Architecture with Platform Suite for SAS ................................. 5
  Grid Architecture with Hadoop ......................................................... 6
What Types of Processing Does a Grid Support? ................................. 8
  Multi-User Workload Balancing ....................................................... 8
  Parallel Workload Balancing ............................................................ 8
  Distributed Enterprise Scheduling .................................................... 8
  SAS Applications That Support Grid Processing ............................... 9
What Business Problems Can a Grid Solve? ......................................... 10
  Many Users on Single Resource ..................................................... 10
  High Availability ............................................................................ 10
  Increased Data Growth .................................................................. 10
  Running Larger and More Complex Analysis ................................... 11
  Need for a Flexible IT Infrastructure ................................................. 11

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 cluster 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, prioritization, and a highly available analytic environment 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 grid provider with the SAS 4GL syntax and subsequently with several SAS products and solutions.

**SAS Grid Topology**

**SAS Grid Configuration**

A SAS Grid configuration consists of these main components:

- **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. You can also choose whether to configure the grid control server as a grid resource capable of receiving work. This machine must contain Base SAS and SAS/CONNECT. The grid control server might also configure a SAS workspace server so that SAS applications (SAS Data Integration Studio, SAS Enterprise Miner, SAS Enterprise Guide, and SAS Add-In for Microsoft Office) 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 are 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, and any applications and solutions needed to run grid-enabled jobs.

- **Shared file system**
  - this component is used to store data and shared directories for jobs that run on the grid. In order to simplify installation and ease maintenance, you can also install the SAS binaries on the shared file system if your grid runs on UNIX machines.

- **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 and to submit jobs to the grid through the Schedule Manager plug-in

- **SAS Grid Manager module**
  - a module for SAS Environment Manager that provides a management interface for the jobs and resources in the grid

- **SAS Grid Manager agent plug-in**
  - an agent plug-in for SAS Environment Manager that monitors grid resources over time and can generate events and alerts based on the collected metric data
Grid clients
clients submit jobs to the grid for processing, but they are not part of the grid resources that available to execute work. Examples of grid clients are:

- a SAS Data Integration Studio client, a SAS Enterprise Miner client, or a SAS Enterprise Guide client that uses a workspace server that can be spawned by the grid
- a SAS Management Console client that uses the Schedule Manager plug-in or another application to schedule SAS workflows
- a SAS Foundation install (SAS Display Manager) that is used to run a program that submits work to the grid. The submitted work can either be entire programs or programs broken into parallel chunks. This client must have Base SAS and SAS/CONNECT installed.
- a SAS Grid Manager Client Utility (SASGSUB). SAS is not required to be installed on this client.

Grid Architecture with Platform Suite for SAS

The following diagram is a conceptual view of the SAS Grid Architecture when using SAS Grid Manager with the Platform Suite for SAS. Platform Suite for SAS is a set of components that are provided by Platform Computing in order to provide efficient resource allocation, policy management, and load balancing of SAS workload requests.

**Figure 1.1  Grid Architecture Diagram — Platform Suite for SAS**

The Platform Suite for SAS includes these components:

Load Sharing Facility (LSF)
This facility 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. A server version of Platform LSF must be on the Grid Control Server and each Grid Node. A client version of Platform LSF must be installed with Platform PWS and with the SAS Display Manager and SASGSUB grid clients. Platform LSF is not required for any other grid clients.
Process Manager (PM)
This is the interface used by the SAS scheduling framework to control the submission of scheduled jobs to LSF and to manage any dependencies between the jobs. Process Manager includes two optional components, Calendar Editor and Flow Manager. Calendar Editor is a scheduling client for a Process Manager server. It enables you to create new calendar entries for time dependencies. Flow Manager provides a visual representation of flows that are created and scheduled through the Schedule Manager plug-in for SAS Management Console as well as reports scheduled through SAS Web Report Studio. Flow Manager enables you to view and update the status of jobs in a flow and rerun jobs. Platform PM is typically installed on the Grid Control Server.

Platform Web Services (PWS)
This is a REST-based web service that provides information to the SAS Grid modules and agent plug-in available in SAS Environment Manager. It requires that an LSF client be installed on the same machine.

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. GMS is typically installed on the Grid Control Server. This component is shown in parentheses because its use is being phased out beginning with the third maintenance release after SAS 9.4. Use the SAS Grid Manager modules for with SAS Environment Manager instead.

Enterprise Grid Orchestrator (EGO)
Provides support for critical service failover in a highly available (HA) environment. EGO provides functions for monitoring critical services, restarting services if they stop, and starting services on a failover host when needed. EGO is a component of LSF and not explicitly shown on the diagram above. However, LSF must be installed on any machine that is running a service for which EGO is to provide high availability.

Platform RTM for SAS
A web-based tool that enables you to graphically view the status of devices and services in a SAS grid environment as well as manage the policies and configuration. This application is not part of Platform Suite for SAS but can be downloaded separately from http://support.sas.com/downloads/package.htm?pid=669

Grid Architecture with Hadoop
The following diagram is a conceptual view of the SAS Grid Architecture when using SAS Grid Manager for Hadoop.
SAS Grid Manager for Hadoop does not include a Hadoop distribution or any of the Hadoop components. However, you must have one of the supported enterprise Hadoop distributions already installed and configured. SAS Grid Manager for Hadoop uses the following components and features:

- **YARN** provides the resource management for the SAS Grid jobs. YARN must be installed on the grid control server and on each of the grid nodes that are expected to run SAS Grid jobs. You must configure the grid control server on the same machine on which YARN’s Resource Manager is running.

- **Oozie** schedules SAS workflows across the cluster. The Oozie server must be installed on one of the nodes in the cluster and is shown on the Grid Control Server in the above diagram.

- **Kerberos** must be configured for the Hadoop cluster. This is required in order for YARN to be able to run the SAS Grid jobs with the identity of the user that submitted them.

- There are no Hadoop components required on the SAS web tier. All monitoring and management of SAS Grid jobs on the Hadoop cluster is performed through the usual Hadoop monitoring and management tools.

- the SAS clients require the Hadoop JAR and configuration files that are installed by the SAS Deployment Wizard and SAS Deployment Manager.

The need for a shared file system is greatly reduced since most data will be stored in HDFS. However, HDFS is not a POSIX compliant file system, so you will still need to use a shared file system rather than HDFS for the following:

- the shared directory required by SASGSUB

- installation and configuration of SAS. If you do not want to use a shared file system for installing and configuring SAS, you can use parcels to leverage Cloudera Manager or packages to leverage Ambari.

- shared project and data directories that are required by other SAS products running on the grid, such as SAS Enterprise Miner and SAS Forecast Server
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 acts as a gatekeeper for the jobs submitted to the grid. As jobs are submitted, SAS Grid Manager dispatches the job to grid nodes, preventing any one machine from being overloaded. If more jobs are submitted than can be run at once, SAS Grid Manager submits as many jobs as can be run. The rest of the jobs are held in a queue until resources are free, and then the jobs are dispatched to be run. SAS Grid Manager 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 in their processes 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 the SAS Grid Manager Client Utility.

**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. The jobs are then run on the grid using the resource and prioritization policies established by SAS Grid Manager.

You can also use SAS Grid Manager Client Utility to interface SAS Grid Manager with an existing enterprise scheduler so that scheduled jobs are processed by the grid without
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

<table>
<thead>
<tr>
<th>SAS Application</th>
<th>Multi-User Workload Balancing</th>
<th>Parallel Workload Balancing</th>
<th>Distributed Enterprise Scheduling</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any SAS program</td>
<td>yes</td>
<td>yes, with modifications</td>
<td>yes</td>
</tr>
<tr>
<td>SAS Grid Manager Client Utility (SASGSUB)</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>SAS Enterprise Guide</td>
<td>yes</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SAS Add-In for Microsoft Office</td>
<td>yes</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SAS Data Integration Studio</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>SAS Enterprise Miner</td>
<td>yes</td>
<td>yes</td>
<td></td>
</tr>
<tr>
<td>SAS Risk Dimensions</td>
<td>yes</td>
<td>yes, with modifications</td>
<td></td>
</tr>
<tr>
<td>SAS Web Report Studio</td>
<td></td>
<td></td>
<td>yes</td>
</tr>
<tr>
<td>SAS Marketing Automation</td>
<td></td>
<td></td>
<td>yes</td>
</tr>
<tr>
<td>SAS Marketing Optimization</td>
<td></td>
<td></td>
<td>yes</td>
</tr>
<tr>
<td>SAS JMP Genomics</td>
<td></td>
<td></td>
<td>yes</td>
</tr>
<tr>
<td>SAS Demand Forecasting for Retail</td>
<td></td>
<td></td>
<td>yes</td>
</tr>
<tr>
<td>SAS products or solutions that use workspace server load balancing</td>
<td>yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.

**High Availability**

Your organization might have services and long-running SAS programs that are critical to your operations. The services must be available at all times, even if the servers that are running them become unavailable. The SAS programs must complete in a timely manner, even if something happens to cause them to fail. For a SAS program that takes a long time to run, this means that the program cannot be required to restart from the beginning if it ends prematurely.

You can configure the critical services within your SAS grid environment to be highly available. SAS Grid Manager can monitor the critical services, detect if they fail or if the machine on which they are running fails, and automatically start the services on a failover host. Either a hardware load balancer or DNS name resolution is used to redirect clients to the service running on the failover host. This ensures that critical services remain available to clients without any manual intervention.

By using options on the SAS Grid Manager Client Utility, you can specify that SAS programs submitted to the grid are automatically restarted from the point where they stopped if they end before completion. The job restarts from the last completed procedure, DATA step, or labeled section. Jobs that take a long time to run do not have to start over at the beginning. You can also use the restart capability with queue options that automatically requeue jobs that end prematurely to provide a complete high-availability environment for SAS programs.

**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 that 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.
Part 2

Grid Computing for SAS Using Platform Suite for SAS

Chapter 2
Planning and Configuring a Grid Environment .......................... 15

Chapter 3
Managing the Grid .......................................................... 27

Chapter 4
Using Grid Management Applications ................................. 41

Chapter 5
High Availability ......................................................... 63

Chapter 6
Troubleshooting .......................................................... 71
Chapter 2
Planning and Configuring a Grid Environment

Installation and Configuration Overview

The process of configuring a Platform Suite for SAS grid consists of two main tasks:


2. Installing and configuring SAS products and metadata definitions on the grid. You can either install all SAS products on all machines in the grid or install different sets of SAS applications on sets of machines in the grid. However, Base SAS, SAS/CONNECT, and SAS Grid Manager must be installed on all grid machines. Using a grid plan file with the SAS Deployment Wizard guides you through the process of installing and configuring the SAS applications and metadata definitions on each machine in the grid. You must specify the same directory structure on all machines in the grid.

For information about performing a planned installation, see SAS Intelligence Platform: Installation and Configuration Guide.
Configuring the File Server

The central file server is a critical component of a grid environment. It is essential for each application on a grid node to be able to efficiently access data. Slowdowns caused by the performance of the file storage system could reduce the effectiveness and benefit of using a grid. The amount of storage required and the type of I/O transactions help determine the type of file storage system that best meets your needs. A shared file system is required if you want to use the restart functions of SAS Grid Manager Client Utility.

Assuming that the SAS jobs running on the grid perform an equal number of reads and writes, it is recommended that the file system be able to sustain 75–100 MB per second per core. This level can be adjusted up or down, depending on the level of I/O activity of your SAS jobs. For information about choosing and configuring a file system, see Best Practices for Data Sharing in a Grid Distributed SAS Environment, which is available at http://support.sas.com/rnd/scalability/grid/gridpapers.html.

Installing Platform Suite for SAS


Information for installing Platform Suite for SAS is available for both Windows and UNIX platforms.

The installation process for Platform Suite for SAS installs these components:

- Platform Process Manager
- Platform LSF
- Platform Grid Management Service

Configuring the Grid Control Server

After you install and configure Platform Suite for SAS, you can use the SAS Deployment Wizard to configure the grid control server. The SAS Deployment Wizard installs and configures these components:
Table 2.1  SAS Deployment Wizard Grid Control Server Components

<table>
<thead>
<tr>
<th>Installed SAS Software Components</th>
<th>Configured SAS Software Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>• SAS Foundation (including Base SAS and SAS/CONNECT)</td>
<td>• Platform Process Manager Server</td>
</tr>
<tr>
<td>• SAS Management Console</td>
<td>• Platform Web Services</td>
</tr>
<tr>
<td>• Grid Manager Plug-in for SAS Management Console</td>
<td>• SAS Environment Manager Server</td>
</tr>
<tr>
<td>• SAS Environment Manager</td>
<td>• SAS Grid Manager Agent Plug-in for SAS Environment Manager</td>
</tr>
<tr>
<td></td>
<td>• SAS Grid Manager module for SAS Environment Manager</td>
</tr>
<tr>
<td></td>
<td>• Grid Monitoring Server</td>
</tr>
<tr>
<td></td>
<td>• SAS Application Server (SAS Logical DATA Step Batch Server, SAS Logical Grid Server, SAS Logical Workspace Server)</td>
</tr>
<tr>
<td></td>
<td>• Object Spawner</td>
</tr>
<tr>
<td></td>
<td>• Grid script file</td>
</tr>
</tbody>
</table>

If you are installing Platform Suite for SAS on a UNIX machine, you must source the profile.lsf file before you start the SAS Deployment Wizard. The hostsetup command in the installation procedure for Platform LSF, starting with version 7, includes the ability to source the LSF profile from the default profile for all users. If this option was not used in the installation process or did not work correctly, you must use the following procedure. This procedure enables the SAS Deployment Wizard to find the addresource utility. To source the file, follow these steps:

1. Start the LSF daemons. The easiest method for doing this is to reboot the computer on which Platform Suite for SAS is installed.
2. Using the default profile for the machine, issue this command:

   `.LSF_TOP/conf/profile.lsf`

   Replace LSF_TOP with the directory in which Platform LSF is installed. Note that the command starts with a period.

The amount of user input that is required during the installation and configuration process depends on whether you choose an Express, Typical, or Custom install. For information about running the SAS Deployment Wizard, see SAS Deployment Wizard User's Guide.

An Express installation does not request any grid-specific information. Default values are used in all cases, so you must verify that these values match the values needed for your environment.

During an Express install, the Platform Process Manager information page enables you to specify the host name and port of the machine on which Platform Process Manager installed.
During the installation and configuration process for a Custom install, the SAS Deployment Wizard displays these pages that request grid-specific information:

1. The SAS Grid Control Server page enables you to specify the name of the Logical SAS Grid Server and the SAS Grid Server. Specify the grid control server machine.

**Figure 2.2  SAS Grid Control Server Page**
2. The Grid Control Server: Job Information page enables you to specify how jobs run on the grid. Specify the command used to start the server session on the grid, workload values, and additional options for the grid. The directory in the **Grid Shared Directory Path** field is used by grid programs (such as the SAS Grid Manager Client Utility) to store information. The location must be accessible by all grid nodes, and all grid users must have Read and Write access to the directory. For information about the values used in these fields, see “Modifying SAS Logical Grid Server Definitions” on page 99.

*Figure 2.3 Grid Control Server: Job Information Page*

3. The SAS Grid Monitoring Server page enables you to specify the name, machine, and port for the grid monitoring server.
4. The Platform Process Manager page enables you to specify the server on which you installed Platform Suite for SAS and the port that is used to connect to the server.

**Figure 2.4 SAS Grid Monitoring Server Page**

**Figure 2.5 Platform Process Manager Page for Custom Install**
**Configuring Platform Web Services**

After you have configured the grid control server, you can use the SAS Deployment Wizard to configure Platform Web Services (PWS). The SAS Deployment Wizard requests the following information:

1. **Use the Specify Software Location page to specify the location of your Platform LSF configuration file.** The correct path is displayed by default if you correctly sourced the LSF profile script before you started the SAS Deployment Wizard.

   ![Specify LSF Location Page](image)

2. **If you specified that SAS Deployment Wizard should display custom prompts, you are prompted to specify a name for a SAS Web Application Server for PWS.** This server must be dedicated to PWS. The default value of SASServer14 should not be changed.

3. **Use the Platform LSF Configuration page to specify the location of the Platform LSF configuration directory.** The correct location should be specified by default.
4. The Platform Web Services: Database Configuration page displays the Platform Web Services database and user ID, as specified in the Platform LSF configuration file. Specify and confirm the password for the user ID.

5. If you are installing on Windows, you must configure the SAS Web Application Server that runs PWS to run as the LSF Administrator.
At a Windows prompt, type `services.msc` to open the Services window.

6. Stop the service for the SASServer14 SAS Web Application Server.
7. Modify the service to run as the LSF Administrator.
8. Restart the service.

### Configuring the Grid Nodes

After you have installed and configured Platform Web Services, you can use the SAS Deployment Wizard to configure the grid nodes. The SAS Deployment Wizard installs and configures these components:

**Table 2.2 Required Software Components for Grid Nodes**

<table>
<thead>
<tr>
<th>Installed SAS Software Components</th>
<th>Configured SAS Software Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAS Foundation (Base SAS, SAS/CONNECT)</td>
<td>SAS Grid Node</td>
</tr>
<tr>
<td></td>
<td>Platform Suite for SAS</td>
</tr>
</tbody>
</table>

If more than one application server contains a logical grid server, you must choose which application server to use.

For information about the values required during a planned installation, see *SAS Intelligence Platform: Installation and Configuration Guide*.

*Note:* The configuration directory structure for each grid node must be the same as that of the grid control server.

### Configuring Client Applications

After the grid nodes have been installed and configured, you can install and configure the software required for the client applications that will use the grid. The software required depends on the type of client application. Applications such as SAS Data Integration Studio that can submit jobs through a workspace server do not need to install anything other than the client application. Applications such as Base SAS that submit jobs to the grid must also install Platform Suite for SAS in order to send jobs to the grid. When you install SAS Management Console, you must also install the SAS Grid Manager plug-in. SAS Management Console is used to monitor and control the grid.

### Modifying Grid Monitoring Server Definitions

The initial configuration of the grid monitoring server is performed by the SAS Deployment Wizard. However, a SAS grid administrator might need to modify the existing grid metadata or add new grid metadata definitions.
A SAS administrator performs these steps to specify or modify the required and optional properties as metadata for the Grid Monitoring Server:

1. In SAS Management Console, open the metadata repository that contains the metadata for the SAS Grid Server.

2. In the navigation tree, select **Server Manager**.

3. Find the metadata object for the Grid Monitoring Server.

4. Right-click the metadata object for the Grid Monitoring Server, and then select **Properties**.

5. In the Properties window for the Grid Monitoring Server, click the **Options** tab.

6. Here are the fields on the **Options** tab:

   **Provider**
   - the grid middleware provider. This value is used to communicate with the grid control server.

   **Module Name**
   - specifies the shared library name or the class name of the support plug-in for Platform Suite for SAS. Leave this field blank unless directed otherwise by SAS Technical Support.

   **Options**
   - the options needed by the grid monitoring server to connect to the grid server.

   **RTM Host Name**
   - specifies the URL for the Platform RTM for SAS host.

7. After you complete the field entries, click **OK** to save the changes and close the Grid Monitoring Server Properties window.

8. In the display area (right side) on SAS Management Console, right-click the **Connection** object for the Grid Monitoring Server, and then select **Properties**.

9. In the Properties window for the Grid Monitoring Server Connection, click the **Options** tab. Here are the fields on this tab:

   **Authentication Domain**
   - the authentication domain that is used for connections to the server. This value is the authentication domain of the machine that Grid Management Services (GMS) is running on.

   **Host Name**
   - the network address of the grid control server.

   **Port**
   - the port that is used to connect to the grid control server. The default value is 1976 for Platform Grid Management Service.

10. After you complete the entries, click **OK** to save the changes and close the Grid Monitoring Server Connection Properties window.
If you are installing SAS Environment Manager in a grid environment where more than one grid machine shares a configuration directory, you must perform additional configuration steps so that the SAS Environment Manager agents point to the correct machine. To run the configuration process, use the script deploy-ev-agents.sh, which you can download from http://support.sas.com/rnd/scalability/grid/HA/gridha.html.

If you are using the SAS Grid Manager agent plug-in or module for SAS Environment Manager, you must use the SAS Deployment Wizard to install and configure Platform Web Services on the middle-tier server. The SAS Deployment Wizard prompts you for the following information:

- Context root for Platform Web Services
- Platform LSF configuration directory
- The name of the Platform Web Services database, as well as a user ID and password for connections to the database (the database name and user ID are grayed out and cannot be changed)
Overview of Grid Management

Most organizations that use SAS consist of a variety of user categories. Each user
category has its own needs and expectations. For example, your organization might have
these users:

SAS Enterprise Guide and SAS Add-In for Microsoft Office users
These users are usually running interactive programs, and expect immediate results.

SAS Enterprise Miner users
These users might be using multiple machines to train models.

SAS Web Report Studio users
These users might be scheduling reports to run at a specified time.
SAS Risk Dimensions users

These users might be running jobs at night.

Some users in your environment might be running jobs that have a high priority. Other
users might be running jobs that require a large number of computing resources. A SAS
grid environment must be able to account for all of these different needs, priorities, and
workloads.

In order to manage this type of environment, you must be able to control when and
where jobs can run in the grid. You can manage grid resources using these strategies:

• Job slots. They enable you to control how many jobs can run concurrently on each
  machine in the grid. This enables you to tune the load that each machine in the grid
can accept. For example, you can assign a higher number of job slots to higher-
capacity machines, which specifies that those machines can process more jobs
concurrently.

• Queues. They enable you to control when jobs can run and what computing
  resources are available to the jobs that are submitted to the queue. You can create
  queues based on factors such as job size or priority. You can also define job dispatch
  windows and run windows for each queue. When you submit a job to a particular
  queue, the queue settings determine when the job runs and what priority the job has
  compared to other jobs that have been submitted to the grid. You can also specify the
  number of job slots across the grid that a queue can use at any one time. By
  combining the job-slot specification on the hosts and queues, you can specify how
  work is distributed across the grid.

• Resources. They enable you to specify where jobs are run on the grid by specifying
  resource names on hosts and using matching resource names on jobs. The resource
  names are specified on machines in the grid to indicate what type of job each
  machine should run. When you submit jobs to the grid, you can specify resource
  names to specify which machines should be used to process the job.

• Grid options sets. They enable you to create sets of SAS options, required resources,
  and grid options. Each options set can then be mapped to a specified combination of
  a SAS application and a user or user group. For example, you might want to define
  policies for different applications running on the grid (such as SAS Data Integration
  Studio and batch SAS programs) as well as for different business units that are using
  those applications on the grid. After defining the grid, you can define grid options
  sets that contain the options needed for the different type of processing. Then, you
  can map the appropriate options sets to each of the user and application combinations
  that you need.

Modifying Configuration Files with Platform RTM

for SAS

You can use Platform RTM for SAS to modify the configuration files that define queues
and resources on the grid. The Platform RTM for SAS download package contains
documentation about performing this task. However, if you use Platform RTM for SAS
to change any configuration files, you cannot make any further changes to the files
outside of Platform RTM for SAS. Changes in the configuration files are not
synchronized with Platform RTM for SAS.

pid=669
Specifying Job Slots for Machines

Overview

Platform LSF uses job slots to specify the number of processes that are allowed to run concurrently on a machine. A machine cannot run more concurrent processes than it has job slots. The default number of job slots for a machine is the same as the number of processor cores in the machine.

However, you can configure hosts with fast processors to have more jobs slots than the number of cores by setting the MXJ value for the given host to a fixed number of job slots. This enables the more powerful host to execute more jobs concurrently to take advantage of the processor’s speed.

Specifying Job Slots Using the SAS Grid Manager Module

To use the SAS Grid Manager module to specify the maximum number of job slots for all nodes, follow these steps:

1. From the SAS Grid Manager main window, select the LSF Configuration Manager tile.

2. Select Load to load the LSF configuration that is currently in use into the SAS Grid Manager module.

3. In the LSF Configurations table, select the entry for the configuration that you just loaded. A window appears with the tiles for editing the LSF configuration.

4. Select the Batch Hosts tile. The table of defined batch hosts appears.

5. Select the entry for the default host name.

6. Select the Maximum job slots field and select the User-defined value.

7. A text field appears below the Maximum job slots field. Specify a number that specifies the maximum number of job slots on each node, regardless of the number of cores. For example, a value of Auto-detect in this field on a machine with 16 cores results in 16 job slots, while a value of 2 on a machine with 16 cores results in just 2 job slots.

8. Click the Save icon to save the definition.

9. Click the back arrow in the SAS Grid Manager module to return to the table of LSF configurations.

10. Select the check box for the configuration that you edited and click Apply.
To use the SAS Grid Manager module to specify the maximum number of job slots for each node, follow these steps:

1. From the SAS Grid Manager main window, select the **LSF Configuration Manager** tile.
2. Select **Load** to load the LSF configuration that is currently in use into the SAS Grid Manager module.
3. In the **LSF Configurations** table, select the entry for the configuration that you just loaded. A window appears with the tiles for editing the LSF configuration.
4. Select the **Batch Hosts** tile. The table of defined batch hosts appears.
5. Select the **Add Batch Host** icon.
6. In the **Hostname** field, select one of the host machines on your grid.
7. In the **Maximum job slots** field, select **User-defined**.
8. In the text field below the **Maximum job slots** field, specify the maximum number of job slots for the machine.
9. Click the Save icon to save the definition.
10. Repeat the steps for each machine in your grid.
11. Click the back arrow in the SAS Grid Manager module to return to the table of LSF configurations.
12. Select the check box for the configuration that you edited and click **Apply**.

### Specifying Job Slots Manually

To change the number of job slots on a grid node by editing the **lsb.hosts** file, follow these steps:

1. Log on to the grid control server as the LSF Administrator (lsfadmin).
2. Open the file **lsb.hosts**, which is located in the directory `LSF-install-dir/conf/lsbatch/cluster-name/configdir`. This is the LSF batch configuration file. Locate the Host section of the file, which contains an entry for a **default** grid node.

```
Begin Host
HOST_NAME MXJ r1m pg ls tmp DISPATCH_WINDOW #Keywords
default ! () () () () () () #Example
End Host
```

3. Edit this file to specify the maximum number of job slots for all nodes or for each node. There are two ways you can do this.

- To specify the total number of job slots per node, edit the line for the **default** node. Here is an example:

```
Begin Host
HOST_NAME MXJ r1m pg ls tmp DISPATCH_WINDOW #Keywords
default ! () () () () () () #Example
End Host
```

The value ! represents one job slot per core for each node in the grid. You can replace this value with a number that specifies the maximum number of job slots on each node, regardless of the number of cores. For example, a value of ! on a
machine with 16 cores results in 16 job slots, while a value of 2 on a machine with 16 cores results in just 2 job slots.

- To specify the total number of jobs slots per node, add a line for each node in the grid. Here is an example:

<table>
<thead>
<tr>
<th>HOST_NAME</th>
<th>cores</th>
<th>rml</th>
<th>pg</th>
<th>ls</th>
<th>tmp</th>
<th>DISPATCH_WINDOW</th>
<th>#Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>default</td>
<td>16</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>#Example</td>
</tr>
<tr>
<td>D1234</td>
<td>16</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>#Example</td>
</tr>
<tr>
<td>D1235</td>
<td>16</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>#Example</td>
</tr>
<tr>
<td>D1236</td>
<td>16</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>#Example</td>
</tr>
<tr>
<td>D1237</td>
<td>16</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>#Example</td>
</tr>
<tr>
<td>D1238</td>
<td>16</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>#Example</td>
</tr>
</tbody>
</table>

Each line designates the concurrent execution of 16 jobs on each node.

4. Save and close the file.

5. Verify the LSF batch configuration file and reconfigure the LSF batch system by entering this command at the command prompt: `badmin reconfig`

6. For details about using this command, see *Platform LSF Reference*.

---

**Using Queues**

**Understanding Queues**

When a job is submitted for processing on the grid, it is placed in a queue and is held until resources are available for the job. LSF processes the jobs in the queues based on parameters in the queue definitions that establish criteria such as which jobs are processed first, what hosts can process a job, and when a job can be processed. All jobs submitted to the same queue share the same scheduling and control policy. By using multiple queues, you can control the workflow of jobs that are processed on the grid.

By default, SAS uses a queue named NORMAL. To use another queue that is already defined in the Grid Manager module or in the lsf.queues file, specify the queue using a `queue=queue_name` option. You can specify this option in the metadata for the SAS logical grid server (in the **Grid Options** field), in the job options macro variable referenced in the GRDSVC_ENABLE statement, or in the **Grid Options** field of a grid options set. For information about specifying a queue in the logical grid server metadata, see “Modifying SAS Logical Grid Server Definitions” on page 99. For information about specifying a queue in a GRDSVC_ENABLE statement, see “GRDSVC_ENABLE Function” on page 137.

---

**Configuring Queues Using the SAS Grid Manager Module**

The SAS Grid Manager module for SAS Environment Manager enables you to define queues. Follow these steps to define a new queue:

1. From the SAS Grid Manager main window, select the LSF Configuration Manager tile.
2. Select Load to load the LSF configuration that is currently in use into the SAS Grid Manager module.

3. In the LSF Configurations table, select the entry for the configuration that you just loaded. A window appears with the tiles for editing the LSF configuration.

4. Select Queues. The table of defined queues appears.

5. Click the Add Queue icon to define a new queue. The Add Queue window appears.

6. Specify values for the queue as described in the following sections.

7. When you finish defining the queue, click the Save icon to save the definition.

8. Click the back arrow in the SAS Grid Manager module to return to the table of LSF configurations.

9. Select the check box for the configuration that you edited and click Apply.

Configuring Queues Manually

Queues are defined in the lsb.queues file, which is located in the directory $LSF-install-dir/conf/lsbatch/cluster-name/configdir. The file contains an entry for each defined queue. Each entry names and describes the queue and contains parameters that specify the queue’s priority and the attributes associated with the queue. For a complete list of parameters allowed in the lsb.queues file, refer to Platform LSF Reference.
**Using the Normal Queue**

As installed, SAS Grid Manager uses a default queue called NORMAL. If you do not specify the use of a different queue, all jobs are routed to this queue and are processed with the same priority. Other queues enable you to use priorities to control the work on the queues.

In the SAS Grid Manager module, the values for the queue definition are the following:

- **Queue name** field
  - normal

- **Priority** field
  - 30

- **Description** field
  - default queue

The queue definition for the normal queue looks like the following:

```plaintext
Begin Queue
  QUEUE_NAME = normal
  PRIORITY = 30
  DESCRIPTION = default queue
End Queue
```

**Example: A High-Priority Queue**

This example shows the existing queue for high priority jobs. Any jobs in the high-priority queue are sent to the grid for execution before jobs in the normal queue. The relative priorities are set by specifying a higher value for the PRIORITY attribute on the high-priority queue.

In the SAS Grid Manager module, specify the following values:

- **Queue name** field
  - priority

- **Priority** field
  - 40

- **Description** field
  - 40

With the priority queue definition added to the normal queue definition, the manual queue definition file looks like the following:

```plaintext
Begin Queue
  QUEUE_NAME = normal
  PRIORITY = 30
  DESCRIPTION = default queue
End Queue

Begin Queue
  QUEUE_NAME = priority
  PRIORITY = 40
  DESCRIPTION = high priority users
End Queue
```
Example: A Night Queue

This example shows the existing queue for processing jobs (such as batch jobs) at night. The queue uses these features:

- The DISPATCH_WINDOW parameter specifies that jobs are sent to the grid for processing only between the hours of 6:00 PM and 7:30 AM.
- The RUN_WINDOW parameter specifies that jobs from this queue can run only between 6:00 PM and 8:00 AM. Any job that has not completed by 8:00 AM is suspended and resumed the next day at 6:00 PM.
- The HOSTS parameter specifies that all hosts on the grid except for host1 can run jobs from this queue. Because the queue uses the same priority as the normal queue, jobs from the high-priority queue are still dispatched first. Excluding host1 from the hosts that are available for the night queue leaves one host always available for processing jobs from other queues.

To define this queue in the SAS Grid Manager module, follow these steps:

1. Follow the general process for defining a queue. In the General tab, enter these values:

   - **Queue name** field
     night
   - **Priority** field
     30
   - **Description** field
     night time batch jobs

2. Select the Dispatch Window tab to display the list of defined dispatch windows.

3. Select the Add icon to define a new dispatch window entry. An entry appears, labeled 00:00 - 00:00.

4. Click on the 00:00 - 00:00 entry to display the Time Window Configuration window.

5. Click in the Start Time field and select a value of 18:00, and then click in the End Time field and select a value of 07:30. Click OK.
6. Select the Run Windows tab and follow the same procedure to specify a new run window with a start time of 18:00 and an end time of 08:00.

7. Select the Hosts tab and click the Add icon to add a host entry.

8. In the Add Host window, specify all in the Hostname field and click OK.

9. Repeat step 7 to add another host entry. Specify ~host1 in the Hostname field.

10. Save the configuration.

11. Return to the table of LSF configurations.

Select the check box for the configuration that you edited and click Apply.

With the night queue definition added, the manual queue definition file looks like the following:

Begin Queue
QUEUE_NAME = normal
PRIORIT Y = 30
DESCRIPTION = default queue
End Queue

Begin Queue
QUEUE_NAME = priority
PRIORIT Y = 40
DESCRIPTION = high priority users
End Queue

Begin Queue
QUEUE_NAME = night
Example: A Queue for Short Jobs

This example shows the existing queue for jobs that need to preempt longer-running jobs. The PREEMPTION parameter specifies which queues can be preempted as well as the queues that take precedence. Adding a value of PREEMPTABLE[short] to the normal queue specifies that jobs from the normal queue can be preempted by jobs from the short queue. Using a value of PREEMPTIVE[normal] to the short queue specifies that jobs from the short queue can preempt jobs from the normal queue. Using a value for PRIORITY on the short queue ensures that the jobs are dispatched before jobs from the normal queue. However, the jobs from the priority queue still take precedence.

To define these queues in the SAS Grid Manager module, follow these steps:

1. Edit the normal queue entry.
2. In the Preemption field, specify PREEMPTIBLE[short]. Save the queue entry.
3. Follow the general process for creating a new queue. Specify the following values:
   - Name field
     short
   - Priority field
     35
   - Preemption field
     PREEMPTIVE[normal]
   - Description field
     short duration jobs
4. Save the queue entry and apply the LSF configuration.

The manual queue definition file looks like the following:

```bash
Begin Queue
  QUEUE_NAME  = normal
  PRIORITY    = 30
  PREEMPTION  = PREEMPTABLE[short]
  DESCRIPTION = default queue
End Queue

Begin Queue
  QUEUE_NAME  = priority
  PRIORITY    = 40
  DESCRIPTION = high priority users
End Queue

Begin Queue
  QUEUE_NAME  = short
  PRIORITY    = 35
  PREEMPTION  = PREEMPTIVE[normal]
End Queue
```
Specifying Job Slot Limits on a Queue

A job slot is a position on a grid node that can accept a single unit of work or SAS process. Each host has a specified number of available job slots. By default, each host is configured with a single job slot for each core on the machine, so a multiple-core machine would have multiple job slots. For information about specifying job slots for a host, see Platform LSF Reference.

You can also use a queue definition to control the number of job slots on the grid or on an individual host that are used by the jobs from a queue. The QJOB_LIMIT parameter specifies the maximum number of job slots on the grid that can be used by jobs from the queue. The HJOB_LIMIT parameter specifies the maximum number of job slots on any one host that can be used by the queue. The following example sets a limit of 60 job slots across the grid that can be used concurrently by the normal queue and a limit of 2 job slots on any host that can be used.

To define this queue in SAS Environment Manager, follow these steps:

1. Edit the normal queue definition entry.
2. Select the Limits tab. Specify these values:
   
   **QJOB limit**
   60

   **HJOB limit**
   2

3. Save the queue entry and apply the LSF configuration.

The entry in the manual definition file looks like the following:

```
Begin Queue
QUEUE_NAME = normal
PRIORITY = 30
DESCRIPTION = default queue
QJOB_LIMIT = 60
HJOB_LIMIT = 2
End Queue
```

Working with Mismatches between Queues and Workspace Servers

Defining a queue that specifies fewer hosts than are available for processing can cause an error. For example, you might want to have jobs in a certain queue only be processed by the machine host2, so the queue definition would contain the line `HOSTS = host2`. You might also have a grid-launched logical workspace server definition that includes workspace servers for host1, host2, and host3. When the object spawner receives a request to start a workspace server, it then submits a job to Platform LSF to start the workspace server. Because the logical workspace server definition includes more hosts (host1, host2, host3) than the queue (host2), an error message results.

To accommodate this type of environment, add the line `ENABLE_HOST_INTERSECTION=Y` to the lsb.params file.
Defining and Specifying Resources

Overview

Defining resources enables you to specify where jobs are run on the grid. You can define resource names on grid nodes, and then specify those same resource names on jobs that are sent to the grid. The resource names that are specified on grid machines indicate a specific program or resource that the grid job might need (for example, a machine that has SAS/STAT installed on it or a machine that has access to a Teradata machine), so you can direct specific types of work to the nodes that are most appropriate for processing them.

By default, when a job is sent to the grid, the name of the SAS Application Server is sent as a resource name along with the job. You can further specify the type of machine used to run a job by specifying the WORKLOAD= parameter on the GRDSVC_ENABLE call.

For example, assume that you have installed and configured a grid that uses the application server name of SASApp. You now want to specify that SAS Data Integration Studio jobs should run on certain machines in the grid. To make this happen, follow these steps:

1. Create a resource name of DI for SAS Data Integration Studio jobs. (DI is only an example; you can use any user-defined string.)
2. Assign the resource names DI and SASApp to the machines that you want to use for processing SAS Data Integration Studio jobs.
3. Add the value DI to the Required Resources field for a grid options set definition that also specifies SAS Data Integration Studio in the SAS Application field.
4. When you send a job to the grid using SAS Data Integration Studio, the job is automatically sent to the DI workload, which sends the job to one of the machines with SASApp as a resource name and DI as a resource name. If there are no grid servers with resource names that match the value on the job, the job is not processed.

If you did not use a grid options set, you would have to manually choose the workload named DI in the Loop Properties window in SAS Data Integration Studio.

Defining Resource Names Using Addresource

SAS Grid Manager provides the addresource command to define hosts and resources. To use this command to specify resource names, follow these steps:

1. Log on to the grid control machine as the LSF administrator.
2. If you are on UNIX, source the LSF profile file and verify that the PATH is correct.
3. Issue the command addresource -r <resource_name> -m <machine_name>. If the machine_name contains spaces, you must change the spaces to underscores.
   For example, the command addresource -r DI -m D1234 assigns the resource name DI to the machine D1234.
4. Run the LSF commands to reconfigure the grid to recognize the new resources.
Specifying Resource Names Using GRDSVC_ENABLE

You can use the GRDSVC_ENABLE function to specify resource names for jobs that run on the grid. Use the SERVER= option to specify a resource to use if the specified grid server has specified a required resource. For more information, see “GRDSVC_ENABLE Function” on page 137.

Specifying Resource Names Using the SAS Grid Manager Client Utility

If you specify required resources in the metadata definition for a SAS Application Server, you can use the SAS Grid Manager Client Utility to specify that those resources are used. Specify the -GRIDAPPSERVER option to specify the SAS Application server to use. For more information, see “SASGSUB Syntax: Submitting a SAS Program in Batch Mode” on page 154.

Specifying Resource Names in SAS Data Integration Studio

In order to specify the resource name for SAS Data Integration Studio jobs, you must complete these tasks:

- Add the resource name as an allowed value for the logical grid server to which you send jobs.
- Specify the workload that corresponds to the resource name in the loop transformation properties.

To add the resource name to the logical grid server metadata's Workload values, see “Modifying SAS Logical Grid Server Definitions” on page 99.

To specify the workload value in SAS Data Integration Studio, follow these steps:

1. On the SAS Data Integration Studio menu bar, select Tools ➤ Options, and then select the SAS Server tab on the Options dialog box.
2. Select the SAS grid server in the Server field.
3. Select the workload to use for the submitted jobs in the Grid workload specification field.

You can also specify the resource name through a grid options set, either on GRDSVC_ENABLE or by using the GRIDOPTSET parameter on the SAS Grid Manager Client Utility.
# Chapter 4

## Using Grid Management Applications

## Using SAS Environment Manager

### Overview

If your grid uses Platform Suite for SAS, SAS Environment Manager provides an agent plug-in and a management module that enables you to monitor and manage a SAS grid cluster. These components provide some of the same functions as Platform RTM for SAS, so you can monitor and manage your grid using the same application that you use to monitor your SAS environment. The plug-in and module have different purposes, however:

- **Grid Manager Agent Plug-in**
  - Uses continuously collected metric data to monitor the performance of the grid and grid servers, graph changing metric data, and generate alerts

---

<table>
<thead>
<tr>
<th>Using SAS Environment Manager</th>
<th>41</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td>41</td>
</tr>
<tr>
<td>Configuring Platform Web Services</td>
<td>42</td>
</tr>
<tr>
<td>Removing the SAS Grid Manager HQU Plug-In</td>
<td>42</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Using the SAS Grid Manager Agent Plug-in</th>
<th>43</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td>43</td>
</tr>
<tr>
<td>Configuring the Plug-in</td>
<td>43</td>
</tr>
<tr>
<td>Monitoring Grid Resources</td>
<td>43</td>
</tr>
<tr>
<td>Creating Alerts</td>
<td>45</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Using the SAS Grid Manager Module</th>
<th>45</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td>45</td>
</tr>
<tr>
<td>Using the SAS Grid Manager Module as the LSF Administrator</td>
<td>46</td>
</tr>
<tr>
<td>Using the Monitoring Functions</td>
<td>47</td>
</tr>
<tr>
<td>Working with LSF Configurations</td>
<td>52</td>
</tr>
<tr>
<td>Working with High Availability Configurations</td>
<td>54</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Using Platform RTM for SAS</th>
<th>55</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using the Grid Manager Plug-in for SAS Management Console</td>
<td>55</td>
</tr>
<tr>
<td>Overview</td>
<td>55</td>
</tr>
<tr>
<td>Maintaining the Grid</td>
<td>57</td>
</tr>
</tbody>
</table>
Grid Manager Module
Configures and controls grid resources, views current performance data, and configures grid resources and high availability applications


Configuring Platform Web Services

You must install and configure Platform Web Services in order to use both the Grid Manager plug-in and module. Use the SAS Deployment Wizard to install Platform Web Services. See “Installing and Configuring SAS Environment Manager in a Grid Environment” on page 25 for more information.

Platform Web Services must run under the account of the LSF Administrator user. If the install account is not an administrator account, you can make the installer account an administrator account, or, if you are using Windows, you can change the account under which the service is running.

If you are running on UNIX or Windows, follow these steps to make the installer account an LSF Administrator account:

1. Edit the lsf.cluster.<cluster_name> file.
2. Locate the line Begin ClusterAdmins in the file.
3. In the next line, Administrators=admin1 admin2, add the ID of the installer account. Separate IDs with a space.
4. Save and close the file.
5. Run the command lsadmin reconfig.
6. Run the command badmin mbdrestart.

If you are running on Windows, follow these steps to change the account under which the service is running:

1. In the Windows Services window, click the entry for the service SASServer14_1 and select Stop from the context menu.
2. In the dialog box for the SASServer14_1 service, click the Log On tab.
3. Select the This account radio button and specify the user ID and password of the LSF Administrator user. Click OK to close the dialog box.
4. Click the entry for the service and select Start from the context menu.

Removing the SAS Grid Manager HQU Plug-In

In the second maintenance release after SAS 9.4, a SAS Grid Manager HQU plug-in was added to SAS Environment Manager. In the third maintenance release after SAS 9.4, this plug-in is replaced by the SAS Grid Manager module for SAS Environment Manager. Although leaving the old plug-in installed will not cause errors, it is recommended that it be removed.

To remove the SAS Grid Manager HQU plug-in, delete the folder <SASconfig>/<Levelroot>/Web/SASEnvironmentManager/server-5.8.0-EE/hq-engine/hq-server/webapps/ROOT/hqu/GridManagerPlugin
Using the SAS Grid Manager Agent Plug-in

Overview

The SAS Grid Manager agent plug-in provides metric data for the grid cluster and for individual grid hosts, and uses that data to perform these functions:

- display the current state of grid resources
- graph the data over time, providing a historical view and enabling you to see how the data changes
- create alerts that notify you whenever a selected measurement reaches a selected state

Configuring the Plug-in

Before the SAS Grid Manager agent plug-in can discover the grid resources, it must be configured. Follow these steps to configure the plug-in:

1. Log on to SAS Environment Manager with an administrative user ID.
2. Select Resources ⇒ Browse ⇒ Servers. The SAS Grid Manager entry in the table contains a gray question mark icon in the Availability column, which indicates that it has not been configured. Select the SAS Grid Manager entry in the table to display the Resource Details page.
3. Select Tools Menu ⇒ Configure Server to display the Configuration Properties page.
4. Specify the user ID and password for the LSF Administrator user and click OK.

Monitoring Grid Resources

After the plug-in is configured, SAS Environment Manager will automatically discover the resources associated with the grid. By default, the process to discover new resources runs every 24 hours. Whether the plug-in is discovering resources for the first time, or it is discovering new or changed nodes in the grid, up to 24 hours could pass before the resources are discovered.

If you want to discover resources sooner, you can restart the SAS Environment Manager agent, which will cause the discovery process to start immediately. If you want to change the interval at which new or changed resources are discovered, modify the file <SASConfig>/Lev1/Web/SASEnvironmentManager/agent-5.8.0-EE/conf(agent.properties. Modify the parameter autoinventory.runtimeScan.interval.millis=interval, where interval is the autodiscovery interval in milliseconds. The default value is 86400000. Restart the SAS Environment Manager agent after you change the value of the parameter.

After the resources are discovered, SAS Environment Manager starts collecting metric data that you can use to monitor the health and operation of your SAS grid. You can view data for the entire grid and for individual grid nodes.
To view availability and metrics for the grid as a whole, follow these steps:

1. In SAS Environment Manager, select **Resources → Browse** and select **Servers**. A table of server resources appears.

2. Locate and select the **SAS Grid Manager** entry in the table.

   The details page opens to the **Monitor** tab, and displays availability and metric data charts for the SAS Grid Manager cluster.

   **Note:** In a multi-resource environment, use the metric charts for individual resources rather than for the entire grid. The charts for the entire grid are an aggregate of the metrics for all the individual nodes, and can be misleading.

To view availability and metrics for individual nodes on the grid, select **Resources → Browse**, select **Services**, and search for a host name. You can also select **SAS Grid Manager Host** in the **All Service Types** field, and then select one of the hosts in the table. There is a separate entry in the table for each node in the grid. The details page opens to the **Monitor** tab, and displays availability and metric data charts for the selected grid node.

The details pages for both the grid and the separate grid nodes contain an availability bar, which enables you to check the availability of the grid or grid node at a glance. The bar displays a color-coded indicator that represents the availability during a time slice (the length of which depends on the display range that you select). The percentage of time that the resource was available is displayed at the end of the availability bar. The color codes are:

- **Green**
  - 100% availability

- **Yellow**
  - Partial availability (between 0% and 100%)

- **Red**
  - 0% availability

Click the percentage availability at the end of the availability bar to view a graph of the availability information.

Clicking on a dot in the availability bar highlights the time slice in the charts beneath the bar, which helps you diagnose availability problems.

To view detailed metric data, click **Metric Data**. The details page changes to display the metric data in table form. By default, the table displays only the available data. To see all of the collected metric data, click the arrow next to **Show All Metrics**. By default, the metric data refreshes every two minutes. To change the refresh interval, click a value next to **Metrics Refresh**.
If you have enabled the SAS Environment Manager Service Architecture, you can also view reports produced by metric data from the grid that is collected in the SAS Environment Manager Data Mart. To view these reports, select **Analyze ➔ Report Center**. In the Report Center, open the folders **Stored Processes ➔ Products ➔ SAS Environment Manager ➔ Kits ➔ SAS Grid Manager**. Use the prompts to produce reports that show the daily job completion totals per host and the job activity on the grid. See “Initializing and Enabling the Service Architecture” in *SAS Environment Manager: User's Guide* for more information.

### Creating Alerts

SAS Environment Manager enables you to create an alert for any of the metrics collected by the SAS Grid Manager agent plug-in. An alert is a user-defined type of event that indicates a critical condition in a selected resource. When an alert occurs, it must be acknowledged, and alerts are listed until they are marked as being fixed. You can define escalation schemes to identify the actions that happen if an alert is not fixed within a specified time.

For example, you can create an alert that occurs whenever availability of the grid cluster falls below a certain percentage, or when the number of available job slots falls below a certain threshold. See “Working with Resource Alerts” in *SAS Environment Manager: User's Guide* for more information about alerts.

The plug-in includes these pre-defined alerts for these conditions:

- Availability of the grid (SAS Grid Manager resource) falls below 100%
- CPU utilization of the grid (SAS Grid Manager resource) goes above 95%. This alert occurs once for every 12 times the condition is met within one hour.
- CPU utilization of each grid host (SAS Grid Manager Host resource) goes above 95%. This alert occurs once for every 12 times the condition is met within one hour.

To create an alert, follow these steps:

1. Select **Resources ➔ Browse** and either **Servers** (for the grid cluster) or **Services** (for the grid nodes). Click the alert icon (which is in the group of three icons to the left of the entry in the table) for the resource. The Alerts page for the resource appears.

2. Click **New** to display the New Alert page.

3. In the New Alert page, specify the metric for which you want to create an alert and the condition for the metric that causes the alert. The page also enables you to define other criteria for the alert, such as how often the alert can occur. See “Defining an Alert” in *SAS Environment Manager: User's Guide* for more information about defining an alert.

---

### Using the SAS Grid Manager Module

#### Overview

The SAS Grid Manager module enables you to configure and perform actions on grid resources and high availability applications, and also provides a snapshot of monitored data for grid resources. Start the SAS Grid Manager module by selecting **SAS Grid Manager** from the SAS Environment Manager Administration side menu.
The module displays these tiles:

**Monitoring**
- perform actions on grid resources (such as opening and closing hosts and activating and deactivating queues)
- view information for the cluster, grid hosts, queues, and jobs
- manage high availability applications
- track each transaction record

**LSF Configuration Manager**
modify the configuration information for the LSF cluster, including definitions for hosts, queues, users, administrators, resource limits, and cluster parameters

**HA Configuration Manager**
modify the configuration information for high availability applications

---

**Using the SAS Grid Manager Module as the LSF Administrator**

These SAS Grid Manager module functions are available only for the LSF Administrator user:

- **High Availability** tile under the **Monitoring** tile
- **Audit Log** tile under the **Monitoring** tile
- Ability to view job information for all users
- **LSF Configuration Manager** tile
- **HA Configuration Manager** tile

Users in the SAS Grid Manager module are mapped to users created in SAS metadata. In order for the LSF Administrator user identity to be available as a valid user in the SAS Grid Manager module, you must perform one of the following actions:

- If the LSF Administrator identity is defined in metadata but not in SAS Environment Manager, then add the identity to the group SAS Environment Manager Super Users
in SAS metadata and synchronize the users in SAS Environment Manager (select **Synchronize Users** under the **Manage** tab).

- If the LSF Administrator identity is defined in metadata and in SAS Environment Manager, select **Manage** ➔ **List Users**, select the user identity, then click **Add to List** to add the identity to the **Super User Role**.

- If identities have been defined in the SAS Environment Manager Super Users group, then you can specify one of those identities as an LSF Administrator by adding the identity to the LSF configuration file. You do not have to synchronize the SAS metadata and SAS Environment Manager user information if you use this approach.

Adding the LSF Administrator user to the SAS Environment Manager Super Users group enables the user to sign on to the SAS Grid Manager module using the LSF Administrator’s credentials and assigns the user to the Super User role in SAS Environment Manager. See *SAS Environment Manager: User's Guide* for information about user management in SAS Environment Manager.

### Using the Monitoring Functions

#### Viewing LSF Cluster Information

The SAS Grid Manager **Cluster** tile displays a table that contains the status of the LSF cluster. A cluster is a group of hosts, organized and managed by an LSF administrator. Clusters are the basis for job sharing in Platform LSF.

#### Managing Hosts

The SAS Grid Manager **Hosts** tile enables you to view configuration and usage information about hosts, as well as perform actions on the hosts.

To view information about a host, follow these steps:

1. Select the **Monitoring** tile, and then select the **Hosts** tile. The **Hosts** table appears, which contains summary information about the hosts in the grid.

2. Click on an entry in the **Host Name** column to view the **Host:hostname** tables for the host.

3. The **Information** tab contains configuration data for the host. The **Resource** tab contains details about the resources available for the host (such as CPU, memory, and disks). The **Load** tab contains a snapshot of the load on the host.

   **Note:** Little or no configuration, resource, or load information is displayed for hosts that are defined as grid client nodes.

To perform an action on a host in SAS Grid Manager, follow these steps:

1. Select the **Monitoring** tile, and then select the **Hosts** tile to display the **Hosts** table.

2. In the table, select the check box for the host that you want to close or open. Select multiple check boxes to perform an action on more than one host.

   **Note:** Actions have no effect on grid client nodes.

3. In the **Select Actions** menu, select the action that you want to perform. Choices for actions are:

   - **Close**
     
     Prevents jobs from processing on the machine. Closing a host is useful when you want to remove the host from the grid for maintenance.
Open

Enables jobs to be processed on the machine.

4. Confirm the action when prompted.

Click the refresh icon 🔄 to manually refresh the table.

Managing Queues

The SAS Grid Manager Queues tile enables you to manage queues and view detailed information about each queue. Use the LSF Configuration Manager in the SAS Grid Manager module to define queues.

To view information about a queue, follow these steps:

1. Select the Monitoring tile, and then select the Queues tile. The Queues table appears, which contains summary information for the queues in the grid.

2. Click on an entry in the Queue Name column to view the Queue: queue_name tables for the queue.
3. The **Information** tab contains configuration data for the queue. The **Load** tab contains a snapshot of the load on the queue. The **Attributes** tab contains any custom attributes that have been defined for the queue.

To perform actions on queues in the SAS Grid Manager module, follow these steps:

1. Select the **Monitoring** tile, and then select the **Queues** tile to display the **Queues** table.

2. In the table, select the check box for the queue that you want to perform an action on. Select multiple check boxes to perform an action on more than one queue.

3. In the **Select Actions** menu, select the action that you want to perform. Choices for actions are:

   - **Activate**
     Activate an inactivated queue. The queue can accept jobs if it is open, and the jobs in the queue are processed.

   - **Close**
     Closes a queue. A closed queue cannot accept any jobs that are sent to the grid. Closing a queue is useful when you need to make configuration changes.

   - **Inactivate**
     Makes a queue inactive. An inactive queue can still accept jobs if it is open, but none of the jobs in the queue can be processed.

   - **Open**
     Opens a closed queue. The queue can accept new jobs and processes the jobs in the queue.

   *Note:* Some queues might be defined to be inactive during certain time periods. If you attempt to open a queue during an inactive time period, the action will appear to be successful, but the queue remains inactive.
4. Confirm the action when prompted.

Click the refresh icon to manually refresh the table.

Managing Jobs
The SAS Grid Manager Jobs tile enables you to view the status of jobs, suspend or terminate running jobs, and resume or terminate suspended jobs. You can also view detailed information and the history for each job.

To view information about a job, follow these steps:
1. Select the Monitoring tile, and then select the Jobs tile. The Jobs table appears, which contains summary information for the jobs in the grid.
2. Click on an entry in the table to view the Job:jobnumber table for the job.
3. The Job Information tab contains information about the job, including where it is (or has been) running, the name of the queue used, the job command, and the start time. The Job History tab contains information about the execution history of the job, including summaries of the job information, the time information (such as the run time, pending time, and suspended time), and the job contents (the events in job processing, such as submission, job start and end, and any error messages).

To perform actions on jobs in the SAS Grid Manager module, follow these steps:
1. Select the Monitoring tile and select the Jobs tile to display the Jobs table.
2. In the table, select the check box for the job on which you want to perform an action. Select multiple check boxes to perform an action on more than one job.
3. In the Select Actions menu, select the action that you want to perform. Choices for actions are:
   - Resume
     Resume processing of a suspended job.
   - Suspend
     Pause the execution of the selected job.
   - Terminate
     Stop execution of the selected job.
4. Confirm the action when prompted.

Click the refresh icon to manually refresh the table.

If you log on to the SAS Grid Manager module using a user ID that is defined as an LSF Administrator ID, you can manage any jobs that have been submitted to the grid. Users who are not LSF administrators can manage only their own jobs. See “Using the SAS Grid Manager Module as the LSF Administrator” on page 46 for more information.

Note: Job names that use double-byte characters are not supported by the SAS Grid Manager module.

Managing High Availability Applications
The SAS Grid Manager module High Availability tile enables you to control and view information about high availability applications running on the grid. High availability applications are configured through the HA Configuration Manager tile in the SAS Grid Manager module, and are defined to have a primary host and a failover host. If the primary host fails, the application automatically starts on the failover host. The tile enables you to view the status of high availability applications running on hosts in the
grid, to stop applications, and to start applications on either the primary host or the failover host.

Note: The High Availability tile and its contents are visible only if you sign on to the SAS Grid Manager module using the credentials for an LSF Administrator. See “Using the SAS Grid Manager Module as the LSF Administrator” on page 46 for more information.

To perform actions on a HA application in the SAS Grid Manager module, follow these steps:

1. Select the Monitoring tile and select the High Availability tile, which displays the High Availability table.

2. In the table, select the check box for the application on which you want to perform an action. Select multiple check boxes to perform an action on more than one application.

3. In the Select Actions menu, select the action that you want to perform. Choices for actions are:
   - Migrate
     Start the selected application on the failover host. Because migrating an application might take a long time, you should migrate only a few applications at one time in order to avoid time-out errors. This option is useful if you are performing maintenance on the primary host.
   - Restart
     Restart the application on the original host if that host is running.
   - Start
     Start the selected application based on the application definition.
   - Stop
     Stop the selected application.

4. Confirm the action when prompted.

To manually refresh the table, click the refresh icon 🔄.

Viewing Audit Log Records
The Audit Logs tile enables you to view information about records in the LSF audit log database. Select the tile to view the Audit Logs table, which contains a summary of the audit log records.

Click on an entry in the table to view the ID: audit_log_ID table for the selected audit log record.

By default, audit logs are purged after 30 days. To change the amount of time that audit logs are kept, change the value of the AUDIT_LOG_KEEP parameter in the Platform Web Services database.

To manually refresh the table, click the refresh icon 🔄.

Note: The Audit Logs tile and its contents are visible only if you sign on to the SAS Grid Manager module using the credentials for an LSF Administrator. See “Using the SAS Grid Manager Module as the LSF Administrator” on page 46 for more information.
Working with LSF Configurations

Select the LSF Configuration Manager tile to manage sets of LSF parameters that are applied to all resources in the grid. An LSF configuration contains items such as a list of grid machines, definitions of queues that are used in the grid, and grid resource definitions. The initial LSF configuration is contained in the lsf.conf file, and it is created when you install Platform Suite for SAS.

To create or modify an LSF configuration, follow these steps:

1. In the SAS Grid Manager module, select the LSF Configuration Manager tile. The LSF Configuration table appears. The table lists all of the LSF configurations that you have created. A configuration that is created from the current settings is indicated by a check mark in the In Use column. When you use the module for the first time, this table will be blank, because there are no user-created configurations.

2. Click Load to use the settings for the currently active configuration to create a new configuration. The newly created configuration is marked with a check mark in the In Use column.

3. Click anywhere on the configuration’s row to edit the configuration settings. A set of tiles appears, where each one corresponds to a type of value in the LSF configuration. The number on the tile identifies the number of existing entries for the tile’s configuration area.

To change configuration values, select the tile for the type of value you want to change. A table appears containing any values that have already been defined.

**Cluster Hosts**

click the Cluster Hosts tile to change the hosts in the cluster and in the resource assignments for each host. The information for this tile is read from and written to the lsf.cluster.cluster_name file.
Batch Hosts
configure the host and job processing for the cluster. The information for this tile is read from and written to the Host section of the lsb.hosts file.

Host Groups
configure the host groups in the cluster. The information for this tile is read from and written to the HostGroup section of the lsb.hosts file.

Queues
define and edit the queues for the cluster. The information for this tile is read from and written to the lsb.queues file.

Resources
define and edit common resource definitions that are shared by all defined clusters. The information for this tile is read from and written to the lsf.shared file.

Users
define LSF users in the cluster. The information for this tile is read from and written to the User section of the lsb.users file.

User groups
define LSF user groups in the cluster. The information for this tile is read from and written to the UserGroup section of the lsb.users file.

Resource limits
define limits for how many resources must be available in order for different classes of jobs to start and which resource consumers the limits apply to. The information for this tile is read from and written to the Limits section of the lsb.resources file.

Cluster administrators
define users who can administer the cluster resources. The primary administrator is specified during installation and cannot be changed. The information for this tile is read from and written to the lsf.cluster.cluster_name file.

Cluster parametersedit the general LSF configuration. The information for this tile is read from and written to the lsf.conf file.

4. Select a tile to change the corresponding configuration information.

5. Click the name of an entry in the table to edit a definition. Click the Add icon in the table to add a new entry.

Note: Some configuration settings are time-based, which means that they can have different values during different time periods. By default, these settings are set for all times. As you add specific entries for time windows, the original setting still applies to all other times. A time-based parameter is identified by the Time-Based icon. An entry appears under the parameter labeled 00:00–00:00. Click the label to define the time period that the parameter applies to, and click the arrow beside the label to define the value for the time period. Click the Time-Based icon again to define further time periods, or click the arrow beside the label All Other Times to define a value for any time period not otherwise specified.

6. When you finish changing the configuration values, click the return arrow on the configuration settings page until you return to the configuration list.

7. Select the check box beside the configuration you just modified and click Apply. The SAS Grid Manager module writes the configuration settings to the configuration files. LSF then applies the settings in the files to the grid cluster.
8. To rename a configuration, select the check box beside the entry in the table and select **Rename**.

9. To delete a configuration, select the check box beside the entry in the table and select **Delete**. Click the header for the check box column to select all entries.

---

**Working with High Availability Configurations**

Select the **HA Configuration Manager** tile to manage configuration and operation of high availability applications in the grid. A HA configuration contains information to enable applications to automatically start on a failover server if the primary server goes down.

To create or modify a HA configuration, follow these steps:

1. In the SAS Grid Manager module, select the **HA Configuration Manager** tile. The **HA Configurations** table appears. The table lists all of the HA configurations that you have created. A configuration that is created from the current settings is indicated by a check mark in the **In Use** column. When you use the module for the first time, this table will be blank, because there are no user-created configurations.

2. Click **Load** to use the settings for the currently active configuration to create a new configuration. The newly created configuration is marked with a check mark in the **In Use** column.

3. Click anywhere in the configuration’s row to edit the configuration settings. The **High-Availability Applications** table appears. Each entry in the table represents a configuration to enable HA operation for an application.

4. Click the Add icon to create a new HA configuration for an application.

5. In the Add Application window, enter information in these locations:
   - **General tab**
     Specify a name, version, and description for the application.
   - **HA settings tab**
     Specify the HA mode, failover interval, and the hosts needed for HA operation. If you select active-passive mode, the application is started on the primary host and starts on the failover host only if the primary host fails. If you select active-active mode, the application is started on all selected hosts.
   - **Execution settings tab**
     Specify criteria for starting the application on the failover host, such as the commands used to start and stop the application.
   - **Dependencies**
     Specify applications that the HA application depends on to operate.

6. When you finish changing the values for the application, save the configuration to return to the application list. Continue adding applications as needed. When you have finished adding all applications, click the arrow icon to return to the **HA Configurations** table.

7. Select the check box beside the configuration you just modified and click **Apply**. The SAS Grid Manager module writes the configuration settings to the configuration files. LSF then applies the settings in the files to the grid cluster.
Using Platform RTM for SAS

Platform RTM for SAS is a web-based tool that lets you graphically view the status of devices and services within your SAS Grid environment as well as manage the policies and configuration of the grid. It is a visual tool to quickly track and diagnose issues before they affect service levels. Platform RTM for SAS provides these features:

• drill-down capabilities to view details of hosts, jobs, queues, and user activities
• instant alerts on job performance and grid efficiency to enable administrators to optimize usage and workloads
• customizable graphs to visually analyze resource usage, workload trends, and job behavior
• interfaces to allow administrators to update the policies and rules in the grid configuration as well as set up high availability for any critical grid services executing in the grid

Platform RTM for SAS helps system administrators improve decision-making, reduce costs, and increase service levels for SAS grid deployments. Refer to the documentation included with the Platform RTM for SAS installation package for instructions.

Platform RTM for SAS is supported on the following operating systems:

• Red Hat Linux 32/64 bit
• Windows Server 2008 R2 64 bit
• Windows 7 32/64 bit

If you have a UNIX grid, you must install the Linux version of Platform RTM for SAS on a Linux machine or a Linux virtual machine (VM). If you have a Windows grid, you must install the Windows version of Platform RTM for SAS on a Windows 7 or Windows Server 2008R2 machine or virtual machine (VM).


Using the Grid Manager Plug-in for SAS Management Console

Overview

The Grid Manager plug-in for SAS Management Console enables you to monitor SAS execution in a Platform Suite for SAS grid environment (the plug-in does not work with SAS Grid Manager for Hadoop. This plug-in enables you to manage workloads on the grid by providing dynamic information about the following:

• jobs that are running on the grid
• nodes that are configured in the grid
• job queues that are configured in the grid

Information is displayed in tabular or chart format. Here is an example of a job view:
Using the SAS Grid Manager, you can customize the view by selecting the columns of data to display and the order in which they should appear. In addition, you can filter, sort, and refresh the display of jobs.
Each grid that you define must have one grid monitoring server configured and running on a machine in the grid.

**Maintaining the Grid**

**Viewing Grid Information**

When you expand the **Grid Manager** node in the navigation tree, all of the grid monitoring servers that you have defined are listed under the name of the plug-in. Each one represents a unique grid. To view information about a specific grid, expand the server's node in the navigation tree. The information for a grid is grouped into three categories in the navigation tree:

- Job Information
- Host Information
- Queue Information

Select a category to display a table that contains information for the category. You can also display a graph of the job information. Click the column headings to select the information that is displayed in the table. Click **Options** to start the Filter wizard, which you can use to select which jobs to display.
After you have defined filters, you can select a filter and click **Filter Now** to filter the displayed information. You can also manage jobs, hosts, and queues from the tables.

Right-click the Grid Monitoring Server node in the navigation tree and select **Options** to specify that the information from the grid is automatically refreshed and how often it is refreshed.

**Managing Jobs**

Use the Grid Manager to terminate or suspend running jobs and terminate or resume suspended jobs.

To terminate a job, follow these steps:

1. In the selection tree, select the **Job Information** node.
2. In the table, locate the job that you want to cancel.
3. Right-click any column in the row for the job and select **Terminate Task** from the pop-up menu.

If you log on to SAS Management Console using a user ID that is defined as an LSF Administrator ID, you can terminate any jobs that have been submitted to the grid. Users can terminate only their own jobs. The LSF Administrator can terminate any job. If you are terminating a job on Windows, be sure to match the domain name exactly (including case).

To suspend a job (pause the job's execution), follow these steps:

1. In the selection tree, select the **Job Information** node.
2. In the table, locate the job that you want to suspend.
3. Right-click any column in the row for the job and select **Suspend Job** from the context menu.

To resume processing of a suspended job, follow these steps:
1. In the selection tree, select the **Job Information** node.
2. In the table, locate the job that you want to resume.
3. Right-click any column in the row for the job and select **Resume Job** from the context menu.

**Displaying Job Graphs**
You can use the Grid Manager to display GANTT charts for jobs running on the grid. To display a chart, follow these steps:
1. In the selection tree, select the **Job Information** node.
2. Right-click and select either **Create Graph by Host** or **Create Graph by Status** from the **Actions** menu, the context menu, or the toolbar.
3. Select **Create Graph by Host** to display a Gantt chart that shows the amount of time taken to process each job and identifies the machine on which the job ran.

**Figure 4.4** Display of Grid Jobs by Host

4. Select **Create Graph by Status** to display a Gantt chart that illustrates the amount of time that each submitted job spent in each job status (such as pending or running).
Closing and Reopening Hosts

You can use the Grid Manager to close or reopen hosts on the grid. A closed host cannot process any jobs that are sent to the grid. Closing a host is useful when you want to remove the host from the grid for maintenance. You can also close the grid control server to prevent it from receiving work.

Note: The status of a host does not change right away after it has been opened or closed. By default, the host status is polled every 60 seconds by the Grid Management Service. The polling time interval is specified by the GA_HOST_POLL_TIME property in the ga.conf file, which is located in the `<LSF_install_dir>/gms/conf` directory.

To close a host, follow these steps:

1. In the navigation area, open the node for the grid containing the host.
2. Select the Host Information node.
   The display area contains a table of the hosts in the grid.
3. In the table, right-click the host that you want to close and select Close from the context menu.

The host now cannot accept jobs that are sent to the grid.

To open a host that has been closed, follow these steps:

1. In the navigation area, open the node for the grid containing the host.
2. Select the Host Information node. The display area contains a table of the hosts in the grid.
3. In the table, right-click the host that you want to open and select **Open** from the context menu. The host can now accept jobs that are sent to the grid.

### Managing Queues

You can use the Grid Manager to close, open, activate, and inactivate queues. A closed queue cannot accept any jobs that are sent to the grid. An inactive queue can still accept jobs, but none of the jobs in the queue can be processed. Closing a queue is useful when you need to make configuration changes to the queue.

To close a queue, follow these steps:
1. In the navigation area, open the node for the grid containing the queue.
2. Select the **Queue Information** node.
   
   The display area contains a table of the queues in the grid.
3. In the table, right-click the queue that you want to close and select **Close** from the context menu.
   
   The queue is now prevented from accepting jobs that are sent to the grid.

To open a closed queue, follow these steps:
1. In the navigation area, open the node for the grid containing the queue.
2. Select the **Queue Information** node.
   
   The display area contains a table of the queues in the grid.
3. In the table, right-click the queue that you want to open and select **Open** from the context menu.
   
   The queue can now accept jobs that are sent to the grid.

To inactivate a queue, follow these steps:
1. In the navigation area, open the node for the grid containing the queue.
2. Select the **Queue Information** node.
   
   The display area contains a table of the queues in the grid.
3. In the table, right-click the active queue that you want to make inactive and select **Inactivate** from the context menu.

To activate a queue, follow these steps:
1. In the navigation area, open the node for the grid containing the queue.
2. Select the **Queue Information** node.
   
   The display area contains a table of the queues in the grid.
3. In the table, right-click the inactive queue that you want to make active and select **Activate** from the context menu.
Chapter 5
High Availability

High Availability and SAS Grid Manager

Your organization might have services and long-running SAS programs that are critical to your operations. The services must be available at all times, even if the servers that are running them become unavailable. The SAS programs must complete in a timely manner, even if something happens to cause them to fail. For a SAS program that takes a long time to run, this means that the program cannot be required to restart from the beginning if it ends prematurely.

SAS Grid Manager provides high availability through these capabilities:

- Multi-machine architecture. Because how a SAS grid is configured and operates, there is no single point of failure. Because jobs are processed on the available grid nodes, if a node becomes unavailable other nodes can take over the workload.

- Platform Suite for SAS. The default configuration of Platform Suite for SAS provides high availability for the grid operation. The LSF master daemon runs on a specified grid node (usually the grid control server), and a failover node is also identified. If the master daemon node fails, the failover node automatically takes over and broadcasts to the rest of the grid. The grid recognizes the new master daemon node and continues operation without interruption. Platform PM and GMS must be treated as critical services and configured for failover along with all other critical services.

- Critical service failover. There are certain services and processes that are critical to the operation of SAS applications on the grid and that must always be available (for example, the SAS Metadata Server). After providing a failover host for the service, you can use Platform Computing’s Enterprise Grid Orchestrator (EGO) to monitor the service, restart the service if it stops, and start the service on the failover host when needed. Once the service has started on the failover host, you can use either hardware (a load balancer) or software (EGO) to automatically direct clients to the
failover host. EGO is part of Platform Suite for SAS that is included with the SAS Grid Manager and installed as part of the LSF installation process.

- Automatic SAS program failover. If a long-running SAS job fails before completion, rerunning it from the beginning can cause a loss of productivity. You can use the SAS Grid Manager Client Utility to specify that the job is restartable. This means that a failed job restarts from the last successful procedure, DATA step, or labeled section. This capability uses the SAS checkpoint and restart functions to enable failed jobs to complete without causing delays. You can also use attributes on the queue definitions in the grid to automatically restart and requeue any job that ends with a specified return code or that terminates due to host failure. Using these options together ensures that critical SAS programs always run successfully and in a timely manner, even if they encounter problems.

All of these strategies are independent of one another, so you can implement the ones that provide the greatest benefit to your organization.

---

**Setting Up High Availability for Critical Applications**

On a grid, there are certain services that always need to be available and accessible to clients. These services are vital to the applications running on the grid and its ability to process SAS jobs. Examples include:

- SAS Metadata Server
- SAS object spawner
- Platform Process Manager
- Platform Grid Management Service
- Web application tier components

Configuring a grid that provides high availability for these services requires these components:

- providing failover hosts for machines that run critical applications. Using multiple machines for critical functions eliminates a single point of failure for the grid.
- providing a way to monitor the high-availability applications on the grid and to automatically restart a failed application on the same host or on a failover host if needed.
- providing a method to let the client know to connect to the failover host instead of the regular host. This can be done through software (DNS resolution) or hardware (the hardware load balancer), but only one is used.

In normal operations, the following sequence takes place:

1. The client determines that it needs to access a service on a machine in the grid.
2. The client sends a query to the corporate DNS server. The DNS server looks up the address for the machine and returns that information to the client.
3. The client uses the address to connect to the machine and use the application.
To provide business continuity for the application, a failover host must be provided for the critical services running in the grid environment. This provides an alternative location for running the critical services and ensures that it remains available to the applications on the grid. In addition, both the main and failover machines must have access to a shared file server. This ensures that the application has access to the data required for operation, regardless of which machine is running the service.

To provide business continuity for the application, the failover capability must also be automatic. EGO is configured to monitor any number of critical services running on the grid. If it detects that the application has failed or that the machine running it has gone down, it is configured to start the application on the failover server automatically, which enables applications to continue running on the grid.

However, once the application has started on the failover server, the client must have a way to know which server is running the application. There are two methods for accomplishing this:

- Using a hardware load balancer. The load balancer serves as an intermediary between the client and the services running on the grid, which decouples the grid operation from the physical structure of the grid. When the client wants to connect to the service, it connects to the load balancer, which then directs the request to the machine that is running the service. The load balancer knows the addresses of both the main and failover machines, so it passes the request on to whichever of the machines is running in the servers. During normal operation, the request goes to the main machine. When failover occurs, EGO starts services on the failover host, and the load balancer forwards connections to it (because it is not the host running the services).
Figure 5.2  Grid Failover with a Load Balancer

1. EGO detects failure of Machine 1, starts HA_App on Machine 2
2. EGO updates entry for HA_App in corporate DNS to point to Machine 2
5. Load balancer determines the address of the running host, connects to Machine 2
4. Client connects to load balancer

- DNS resolution. Once EGO starts the application on the failover server, it sends the address of the failover machine to the corporate DNS server. The entry for the application is updated in the server, so the next time a client requests a connection to the application, the DNS server returns the address of the failover machine.

Figure 5.3  Grid Failover with EGO

1. EGO detects failure of Machine 1, starts HA_App on Machine 2
2. EGO updates entry for HA_App in corporate DNS to point to Machine 2
5. Client connects to HA_App on Machine 2
4. DNS server returns address of Machine 2
3. Client queries DNS server for address of machine running HA_App

If you do not want EGO to directly update the corporate DNS, you can configure the DNS server to always point to EGO to provide the IP address for the machine. When EGO starts the application on the failover machine, it then points to the new machine.
The choice of whether to use a load balancer or a DNS solution depends on your organization’s policies. Using DNS resolution prevents you from having to purchase an addition piece of hardware (the load balancer). However, your organization’s policies might prohibit either the corporate DNS from being changed by an outside DNS (EGO) or DNS requests to be forwarded to an outside DNS. If this is the case, the hardware load balancer provides a high-availability solution.

## Restarting Jobs

### Overview

An essential component of a highly available grid is the ability to handle SAS jobs that fail or have to be restarted for some reason. If a long-running job fails, it can cause a significant loss of productivity. After the failure is noticed, you must manually resubmit the job and wait while the program starts over again from the beginning. For SAS programs that run for a considerable amount of time, this can cause unacceptable delays.

The SAS Grid Manager Client Utility, combined with LSF queue policies and the SAS checkpoint restart feature, provides support for these solutions to this problem:

- the capability to restart a job from the last successful job step
- the ability to set up a special queue to automatically send failed jobs to another host in the grid to continue execution

### Using SAS Checkpoint and Label Restart

The SAS Grid Manager Client Utility includes options that enable you to restart SAS programs from the last successful PROC or DATA step. When the program runs, it records information about the SAS procedures and DATA steps or labels in the program and tracks the ones that have been passed during execution.

If the program fails and has to be restarted, SAS first executes global statements and macros. Then, it reads the checkpoint or label library to determine which checkpoints or labels have been passed. When SAS determines where the program stopped, execution is resumed from that point. Program steps that have already successfully completed are not re-executed.

The restart capability is available on the grid only if you are using the SAS Grid Manager Client Utility or scheduling grid jobs. It is not available if you are using other application interfaces to submit work to the grid.

If you use the restart options, your SAS WORK library must be on shared storage. Using this capability adds some overhead to your SAS program, so it is not recommended for every SAS program that you run.

To set up the checkpoint or label restart capability, use the SAS Grid Manager Client Utility to submit the SAS program to the grid. Specify either the GRIDRESTARTOK argument (for checkpoints) or the GRIDLRESTARTOK argument (for labels). You cannot specify both arguments.

When you use the GRIDRESTARTOK argument, these options are automatically added to your SAS program:

- **STEPCHKPT**
  - enables checkpoint mode and causes SAS to record checkpoint-restart data.
STEPRESTART
enables restart mode, ensuring that execution resumes at the proper checkpoint.

When you use the GRIDLRESTARTOK argument, these options are automatically added to your SAS program:

LABELCHKPT
enables checkpoint mode for labeled code sections.

LABELRESTART
enables restart mode, ensuring that execution resumes at the proper labeled section.

Other options are automatically added to control restart mode. See “Checkpoint Mode and Restart Mode” in SAS Language Reference: Concepts for a list of options and their definitions as well as complete information about enabling checkpoint restart mode in your SAS programs.

If the host that is running the job becomes unresponsive, the program is automatically restarted at the last checkpoint.

Setting Up Automatic Job Requeuing

You can set up a queue that automatically requeues and redispatches any job that ends with a specified return code or terminates due to host failure. Using job requeuing enables you to handle situations where the host or the system fails while the job is running. Using the requeue capability ensures that any failed jobs are automatically dispatched to another node in the grid.

To use this functionality in a grid, you must use the SAS Grid Manager Client Utility and configure the SAS WORK library to run on shared storage.

To set up a queue for automatic restart, follow these steps:

1. Create a queue, including these two options in the queue definition:

   • `REQUEUE_EXIT_VALUES=return_code_a return_code_b ... return_code_n` option in the queue definition. The `return_code` values are the job exit codes that you want to filter. Any job that exits with one of the specified codes is restarted.

     Specifying `REQUEUE_EXIT_VALUES=all ~0 ~1` specifies that jobs that end with an exit code other than 0 (success) or 1 (warnings) are requeued.

     **Note:** If you specify a `return_code` value higher than 255, LSF uses the modulus of the value with 256. For example, if SAS returns an exit code of 999, LSF sees that value as (999 mod 256) or 231. Therefore, you must specify a value of 231 on `REQUEUE_EXIT_VALUES`.

   • `RERUNNABLE=YES`. This specifies that jobs sent from this queue can be rerun if the host running them fails.

2. Specify the queue that you created in step 1, either by modifying a grid server definition or by specifying the -GRIDJOBOPTS option.

   To create or modify a grid server definition, use the Server Manager plug-in in SAS Management Console. To specify the queue, specify `"queue=<name_of_requeue_queue>"` in the Additional Options field of the server definition.

   To use -GRIDJOBOPTS, submit the job using the -GRIDJOBOPTS `queue=name_of_requeue_queue` option.
3. Submit the job to the requeue queue on the grid. You must use the SAS Grid Manager Client Utility to specify the -GRIDRESTARTOK option. Send the job to the requeue queue by using the server that you specified in step 2.
Overview of the Troubleshooting Process

These topics provide the framework for a systematic, top-down approach to analyzing problems with a grid environment. By starting at the highest level (the network) and working downward to the job execution, many common problems can be eliminated.

For the troubleshooting information not contained here, go to http://support.sas.com/rnd/scalability/grid/gridinstall.html or contact SAS Technical Support.
Verifying the Network Setup

Overview

The first step in troubleshooting problems with a SAS grid is to verify that all computers in the grid can communicate with one another through the ports that are used by Platform Suite for SAS.

Host Addresses

Check the /etc/hosts file on each grid node to ensure that the machine name is not mapped to the 127.0.0.1 address. This mapping causes the sign-on connection to the grid node to fail or to hang. This happens because the SAS session that is being invoked on the grid node cannot determine the correct IP address of the machine on which it is running. A correct IP address must be returned to the client session in order to complete the connection. For example, delete the name "myserver" if the following line is present in the /etc/hosts file:

```
127.0.0.1 myserver localhost.localdomain localhost
```

Host Connectivity

You must verify that the network has been set up properly and that each machine knows the network address of all the other machines in the grid. Follow these steps to test the network setup:

1. Run the `hostname` command on every machine in the grid (including grid nodes, grid control servers, and SAS Foundation grid clients).

2. Run the `ping` command on all grid node machines and the grid control machine against every other machine in the grid (including grid client machines). When you ping a grid client machine, use the host name without the domain suffix.

3. Run the `ping` command on each grid client machine against every other machine in the grid (including itself). When a grid client machine pings itself using the value from the `hostname` command, verify that the returned IP address is the same IP address that is returned when the grid nodes ping the client. However, this might not occur on machines with multiple network adapters.

If the network tests indicate a problem, you must either correct the DNS server or add entries to each machine's hosts file. Contact your network administrator for the best way to fix the problem.

Platform LSF assumes that each host in the grid has a single name, that it can resolve the IP address from the name, and that it can resolve the official name from the IP address. If any of these conditions are not met, LSF needs its own `hosts` file, which is located in its configuration directory (LSF_ENVDIR/conf/hosts).
Host Ports

You must verify that the ports that SAS and LSF use for communication are accessible from other machines. The ports might not be accessible if a firewall is running on one or more machines. If firewalls are running, you must open ports so that communication works between the LSF daemons and the instances of SAS. Issue the `telnet <host><port>` command to determine whether a port is open on a specific host.

The default ports used in a grid include the following:

- **LSF:** 6878, 6881, 6882, 7869, 7870, 7871, 7872
- **Grid Monitoring Service:** 1976
- **Platform Process Manager:** 1966

If you need to change any port numbers, modify these files:

- **LSF ports:** LSF_ENVDIR/conf/lsf.conf and EGO_CONFDIR/ego.conf
- **Grid Monitoring Service port:** gms/conf/ga.conf
- **Platform Process Manager port:** pm/conf/js.conf

If you change the Grid Monitoring Service port, you must also change the metadata for the Grid Monitoring Server. If you change the Platform process Manager port, you must also change the metadata for the Job Scheduler Server.

Ports might be used by other programs. To check for ports that are in use, stop the LSF daemons and issue the command `netstat -an | <search-tool><port>`, where `search-tool` is `grep` (UNIX) or `findstr` (Windows). Check the output of the command for the LSF ports. If a port is in use, reassign the port or stop the program that is using the port.

SAS assigns random ports for connections, but you can restrict the range of ports SAS uses by using the `-tcpportfirst <first-port>` and `-tcpportlast <last-port>` options. You can specify these options in the SAS configuration file or on the SAS command line. For remote sessions, you must specify these options either in the grid command script (sasgrid.cmd on Windows or sasgrid on UNIX) or in the **Command** field in the logical grid server definition in metadata. For example, adding the following parameters to the SAS command line in the grid script restricts the ports that the remote session uses to between 5000 and 5005:

```
-tcpportfirst 5000 -tcpportlast 5005
```

---

Verifying the Platform Suite for SAS Environment

Verifying That LSF Is Running

After the installation and configuration process is complete, verify that all of the LSF daemons are running on each machine.

For Windows machines, log on to each machine in the grid and check the Services dialog box to verify that these services are running:

- Platform LIM
- Platform RES
- Platform SBD
For UNIX machines, log on to each machine in the grid and execute the **ps** command to check for processes that are running in a subdirectory of the $LSF_install_dir. An example command follows:

```
ps -ef|grep LSF_install_dir
```

The daemons create log files that can help you debug problems. The log files are located in the machine's $LSF_install_dir/logs directory (Windows) or the shared LSF_TOP/log directory (UNIX). If the daemon does not have access to the share on UNIX, the log files are located in the /tmp directory.

If the command fails, check the following:

- Verify that the path to the LSF programs is in the PATH environment variable. For LSF 7, the path is $LSF_install_dir/7.0/bin.
- On UNIX machines, you might have to source the $LSF_TOP/conf/profile.lsf file to set up the LSF environment.
- A machine might not be able to access the configuration files. Verify that the machine has access to the shared directory that contains the binary and configuration files, defined by the LSF_ENVDIR environment variable. If the file server that is sharing the drive starts after the grid machine that is trying to access the shared drive, the daemons on the machine might not start. Add the LSF_GETCONF_TIMES environment variable to the system environment and set the variable value to the number of times that you want the daemon to try accessing the share in each five-second interval before the daemon quits. For example, setting the variable to a value of 600 results in the node trying for 50 minutes ((600*5 seconds)/60 seconds per minute) before quitting.
- The license file might be invalid or missing. If LSF cannot find a license file, some daemons might not start or work correctly. Make sure that the license file exists, is properly referenced by the LSF_LICENSE_FILE parameter in the LSF_ENVDIR/conf/lsf.conf file, and is accessible by the daemons.
- All daemons might not be running. Restart the daemons on every machine in the grid using the **lsfrestart** command. If this command does not work, run the **init.d/lsf restart** command (UNIX) or use the Services Administration tool (Windows). Open Services Administration, stop the SBD, RES, and LIM services (in that order). Next, start the LIM, RES, and SBD services (in that order).
- A grid machine might not be able to connect to the SAS grid control machine. The grid control machine is the first machine listed in the lsf.cluster.<cluster_name> file. Make sure that the daemons are running on the master host, and verify that the machines can communicate with each other.

**Verifying LSF Setup**

You must verify that all grid machine names are specified correctly in the LSF_ENVDIR/conf/lsf.cluster.<cluster_name> file and the resource is specified in the lsf.shared file. Follow these steps to make sure the configuration is correct:

1. Log on as an LSF administrator on one of the machines in the grid, preferably the grid control server machine. The LSP Administrator ID is listed in the lsf.cluster.<cluster_name> file under the line

   `Administrators=username1 username2 ... usernameN`

2. Run the command **lsadmin ckconfig -v** to check the LSF configuration files for errors.
3. Run the command `badmin ckconfig -v` to check the batch configuration files for errors.

4. Run the command `lshosts` to list all the hosts in LSF, and to verify that all the hosts are listed with the proper resources.

5. Run the command `bhosts` to list all the hosts in LSF's batch system. Verify that all hosts are listed. Make sure that the Status for all hosts is set to `ok` and that the correct number of jobs slots have been defined for each host in the MAX column (the maximum number of jobs the host can process at the same time).

6. If you find any problems, correct the LSF configuration file and issue the commands `lsadmin reconfig` and `badmin reconfig` so that the daemons use the updated configuration files.

7. If you added or removed hosts from the grid, restart the master batch daemon by issuing the command `badmin mbdrestart`. To restart everything, issue the `lsfrestart` command.

**Verifying LSF Job Execution**

Some problems occur only when you run jobs on the grid. To minimize and isolate these problems, you can run debug jobs on specific machines in the grid.

To submit the debug job, run the command `bsub -I -m <host_name> set` from the grid client machine to each grid node. This command displays the environment for a job running on the remote machine and enables you to verify that a job runs on the machine.

If this job fails, run the `bhist -l <job_id>` command, where `job_id` is the ID of the test job. The output of the command includes the user name of the person submitting the job, the submitted command, and all the problems LSF encountered when executing the job. Some messages in the bhist output for common problems include the following:

- Failed to log on user with password
  - specifies that the password in the Windows `passwd.lsfuser` file is invalid.
  - Update the password using the `lspasswd` command.

- Unable to determine user account for execution
  - specifies that the user does not have an account on the destination machine. This condition can occur between a Windows grid client to a UNIX grid node because a domain has been prefixed to the Windows user name. Correct this problem by making sure that the user has an account on the UNIX machines. Also, add the line `LSF_USER_DOMAIN=` to the Windows lsf.conf file to strip the domain from the user name.

**Verifying the SAS Environment**

**Verifying SAS Grid Metadata**

SAS needs to retrieve metadata about the grid from a SAS Metadata Server in order to operate properly. Start SAS Management Console and use the Server Manager plug-in to verify the following:
Logical grid server
Under the SAS Application Server context (for example, SASApp), verify that a logical grid server has been defined.

Open the Properties window for the logical grid server. Verify that the properties contain the correct path to the script file or the correct command that is executed on the grid node. Verify that the path exists on every node in the grid and that the command is valid on every node in the grid.

Grid monitoring server
Verify that a grid monitoring server has been defined.

Open the connection properties for the server. Verify that the properties contain the name or address of the machine that is running the Grid Monitoring Server daemon (typically the SAS grid control machine). Verify that the port specified in the properties is the same as that specified in the Grid Monitoring Service configuration file (the default value is 1976).

Verifying Grid Monitoring
The Grid Manager plug-in for SAS Management Console displays information about the grid's jobs, hosts, and queues. After you define the Grid Monitoring Server and the Grid Management Service is running on the control server, grid information is displayed in the Grid Manager plug-in in SAS Management Console. Common error messages encountered in the Grid Manager plug-in include the following:

Connection timed out or Connection refused
The Grid Management Service is not running. Start the Grid Management Service on the grid control machine.

Your user ID or password is invalid. Please try again or contact your systems administrator

Either the user provided invalid credentials for the machine running the Grid Monitoring Service, or the user's credentials that are stored in the metadata do not include a password for the login that is associated with the authorization domain used by the Grid Monitoring Server connection. For example, "Grid 1 Monitoring Server" is defined in the metadata to use the "DefaultAuth" authorization domain. A login has been defined for "User1" in the User Manager for the "DefaultAuth" domain, but only the user ID has been specified and the password is blank.

There are three ways to correct the problem. First, provide complete credentials for the authorization domain for the user. Second, you can remove the login for the authorization domain. The third option is to use a different authorization domain for the Grid Monitoring Server connection. If you provide the correct credentials, the user is not prompted for a user ID and password. If you remove the login for that authorization domain or change the Grid Monitoring Server connection to use a different authorization domain without adding credentials for the user for that domain, the user is prompted for their user ID and password to connect to the machine where the Grid Monitoring Server is running.

Verifying SAS Job Execution
SAS provides a grid test program on the SAS support website that tests connectivity to all nodes in the grid. Run the program from a grid client. You can download the program from http://support.sas.com/rnd/scalability/grid/gridfunc.html#testprog. After you download the program, follow these steps:
1. Copy and paste the grid test program into a SAS Foundation Display Manager session.

2. If the application server associated with your logical grid server in your metadata is not named “SASMain”, change all occurrences of “SASMain” in the test program to the name of the application server that is associated with your logical grid server. For example, some SAS installations have named the application server “SASApp”, so all occurrences of SASMain should be replaced with “SASApp”.

3. Submit the code.

The program attempts to start one remote SAS session for every job slot available in the grid. The program might start more than one job on multi-processor machines, because LSF assigns one job slot for each core by default.

Here are some problems that you might encounter when running the grid test program:

**Grid Manager not licensed** message
Make sure that your SID contains a license for SAS Grid Manager.

**Grid Manager cannot be loaded** message
Make sure that Platform Suite for SAS has been installed and that the LSF and PATH environment variables are defined properly.

**Invalid resource requested** message
The application server name or workload value has not been defined in the lsf.shared file. Also, make sure that you associate the value with the hosts that you want to run SAS programs in the lsf.cluster.<\cluster_name> file.

The number of grid nodes is 0.
Possible reasons for this error include the following:

- The application server name was not defined as a resource name in the lsf.shared file.
- The application server name was not associated with any grid nodes in the lsf.cluster.<\cluster_name> file.
- The grid client where the job was submitted cannot communicate with the entire grid.

The number of grid nodes is not the same as the number of grid node machines.
As shipped, the number of grid nodes equals the number of job slots in the grid. By default, the number of job slots is equal to the number of cores, but the number of job slots for a grid node can be changed.

Another explanation is that the application server name has not been associated with all the grid nodes in the lsf.cluster.<\cluster_name> file.

Jobs fail to start.
Possible reasons for this problem include the following:

- The grid command defined in the logical grid server metadata is either not valid on grid nodes or does not bring up SAS on the grid node when the command is run. To verify the command, log on to a grid node and run the command defined in the logical grid server definition. The command should attempt to start a SAS session on the grid node. However, the SAS session might not run successfully because grid parameters have not been included. Platform Suite for SAS provides a return code of 127 if the command to be executed is not found and a return code of 128 return code if the command is found, but there is a problem executing the command.
• Incorrect version of SAS installed on grid nodes. SAS 9.1.3 Service Pack 3 is the minimum supported version. A return code of 231 might be associated with this problem.

• Unable to communicate between the grid client and grid nodes. Verify that the network is set up properly, using the information in “Verifying the Network Setup” on page 72.

Jobs run on machines that are supposed to be only grid clients. By default, all machines that are listed in the lsf.cluster.<cluster_name> file are part of the grid and can process jobs. If you want a machine to be able to submit jobs to the grid (a grid client) but not be a machine that can process the job (a grid node), set its maximum job slots to 0 or use the Grid Manager plug-in to close the host.

### Debugging Grid Jobs

When a grid-enabled SIGNON is used or a job is submitted by SASGSUB, the sasgrid script is used to start a SAS session on the grid node. To aid in debugging, the output from the script (but not the SAS log and SAS output) is written to a unique log file. The file is located in the GridServer/Logs directory of the machine where the job runs. If the job is successful, the log file is deleted when the job finishes running. If the job is unsuccessful, the log file is saved.

The log file is named using the format
SASGrid.userID.hostname.process_ID.YYYYMMDDhhmmss.log

The log file is not created if the job is run on a grid-enabled IOM server (workspace server, stored process server, or pooled workspace server), because these servers do not use the sasgrid script. These servers are started by the object spawner, so you must configure logging of the object spawner to debug grid-enabled IOM servers. For example, you can set the logger App.tk.tkegrid to TRACE and then review the object spawner’s log.

### Turning On Debugging in the SAS Environment Manager Agent Plug-In Loggers

To turn on debugging for the SAS Environment Manager agent plug-in loggers, add the following lines to the file <levelroot>/Web/SASEnvironmentManager/agent-5.8.0-EE/conf/agent.properties:

log4j.logger.com.sas.grid.hyperic=DEBUG
log4j.logger.com.sas.svcs.security.authentication.client=DEBUG

### Fixing Issues with the SAS Grid Manager Module for SAS Environment Manager

Here are some possible solutions if the SAS Grid Manager module fails to invoke in SAS Environment Manager:
You receive a “Service is temporarily not available” error
Verify that Platform Web Services is running. Restart the service if needed.

You receive a “Connecting to the server failed” error
Verify that SAS Environment Manager is running.

You receive an “Internal server error” message
Verify that Platform LSF is running.

You receive the message "You do not have permission to perform this action" when attempting to use the SAS Grid Manager module on Windows
This message indicates that Platform Web Services is not running under the LSF Administrator account. To change the account that the service is running under, follow these steps:

1. In the Windows Services window, click the entry for the service SAServer14_1 and select Stop from the context menu.
2. In the dialog box for the SAServer14_1 service, click the Log On tab.
3. Select the This account radio button and specify the user ID and password of the LSF Administrator user. Click OK to close the dialog box.
4. Click the entry for the service and select Start from the context menu.

If you would rather issue commands, open a Windows Command window and issue these commands:

- net stop "SAS [SASConfig-Lev] SAServer14_1 - WebAppServer"
- sc config "SAS [SASConfig-Lev] SAServer14_1 - WebAppServer" obj= domain \userid password= password
  
  Note: You must include a space between the equal sign and the value for the obj= and password= parameters.

Substitute your directory names for SASConfig-Lev.
You must update the entry for this service anytime the password for the LSF Administrator account changes.

Platform Web Services Errors
You receive permission errors or LSF library initialization errors
Verify that Platform Web Services is running under the LSF Administrator account. These errors occur if Platform Web Services runs under a local user account. To change the account that the service is running under, follow these steps:

1. In the Windows Services window, click on the entry for the service SAServer14_1 and select Stop from the context menu.
2. In the dialog box for the SAServer14_1 service, click the Log On tab.
3. Select the This account radio button and specify the user ID and password of the LSF Administrator user. Click OK to close the dialog box.
4. Click the entry for the service and select Start from the context menu.
Examining the debug log might help diagnose problems with the Grid Manager plug-in. To start logging, open the file server-log4j.xml (in the directory $SASEV/conf) and change the root-level value to DEBUG. The debugging information is written to the file server.log in the directory $SASEV/logs.
Part 3

Grid Computing for SAS Using SAS Grid Manager for Hadoop

Chapter 7
Using SAS Grid Manager for Hadoop
Chapter 7
Using SAS Grid Manager for Hadoop

Overview

SAS Grid Manager for Hadoop provides workload management, accelerated processing, and scheduling of SAS analytics co-located on your Hadoop cluster. SAS Grid Manager for Hadoop leverages YARN to manage resources and distribute SAS analytics to a Hadoop cluster running multiple applications. Integration with Oozie provides the scheduling capability for SAS workflows. All of the existing SAS Grid syntax, submission modes, and integration with other SAS products and solutions is supported by SAS Grid Manager for Hadoop. Therefore, end users of SAS applications that are submitting work to the SAS grid will not notice any difference in their operations. They will continue to submit jobs as they normally would, but those jobs will be processed on the Hadoop cluster. Grid monitoring and management is provided through the Hadoop monitoring tools, rather than through SAS Environment Manager or SAS Management Console.

Processing in the cluster is controlled by the queue allocations defined in the YARN scheduler and through the SAS Grid Manager for Hadoop's SASGRID policy file. The queues determine the allocation of resources between different groups. The SASGRID policy file enables you to specify which hosts in the Hadoop cluster are available for grid jobs. You can also define the application types (such as normal or priority) that are processed by the grid, specify which hosts can be used for processing those applications, and specify the computing resources and limits (such as memory and maximum processing time) that apply to each type. You can then specify application type to associate with a submitted job in a number of ways, including using a grid option on a grid options set. When a SAS user submits a job, the correct grid options set, including
the application type, is applied and the job is processed on the designated hosts in the Hadoop cluster.

Scheduling of SAS jobs on the Hadoop cluster is provided through Oozie. SAS uses Oozie to schedule DATA step batch servers to process SAS code in workflows. SAS users do not interface directly with Oozie. They deploy jobs for scheduling and schedule those jobs using an Oozie scheduling server (defined in SAS metadata), just as they would with any other type of scheduling. The Oozie scheduling server definition contains the information needed to send the scheduled jobs to Oozie on the Hadoop cluster.

High availability of the Resource Manager and the Name Node are supported by SAS Grid Manager for Hadoop. In addition, any failed SAS job is resubmitted.

---

**Configuration**

SAS Grid Manager for Hadoop supports these Hadoop distributions:

- Cloudera 5.2.x and higher
- Hortonworks 2.1.x and higher
- Hortonworks 2.2 (this version includes a REST API for job submission that results in better performance)
- MapR 4.0.x and higher
- MapR 4.1.x and higher (recommended, because this version adds support for crontab, which is required in order to schedule recurring flows)

The process of configuring SAS Grid Manager for Hadoop consists of these steps:

1. Install Hadoop services.
2. Enable Kerberos on the Hadoop cluster.
3. Enable SSL.
4. Update YARN parameters.
5. Set up HDFS directories.
6. Run the SAS Deployment Wizard to install and configure a SAS Grid Manager for Hadoop control server.

---

**Using SAS Deployment Manager to Deploy SAS Grid Manager for Hadoop**

1. Start the SAS Deployment Manager.

```bash
 cd /[SASHOME]/SASDeploymentManager/9.4
 ./sasdm.sh
```

The Select SAS Deployment Manager Task page opens.
2. Under **Hadoop Configuration**, select **Deploy SAS Grid Manager for Hadoop**. Click **Next** to continue.

   The Select Hadoop Distribution page opens.

3. From the drop-down menu, select the distribution of Hadoop that you are using.
Note: If your distribution is not listed, exit the SAS Deployment Manager and contact SAS Technical Support.

Click Next. The Hadoop Cluster Manager Information page opens.

4. Enter the host name and port number for your Hadoop cluster. For Cloudera, enter the location where Cloudera Manager is running. For Hortonworks, enter the location where the Ambari server is running. The port number is set to the appropriate default after Cloudera or Hortonworks is selected.

Note: The host name must be a fully qualified domain name. The port number must be valid, and the cluster manager must be listening.

Click Next. The Hadoop Cluster Manager Credentials page opens.
5. Enter the Cloudera Manager or Ambari administrator account name and password.

   *Note:* Using the credentials of the administrator account to query the Hadoop cluster and to find the Hive node eliminates guesswork and removes the chance of a configuration error. However, the account name does not have to be that of an administrator; it can be a read-only user.

   Click **Next**. If you are configuring a Cloudera cluster, the Specify SAS Product Deployment Media Version page opens
6. The version listed will be assigned to the media that will be used for deployment unless you enter a different version. The version number is calculated by the SAS Deployment Manager with the actual version of the installed product that you selected to deploy.

*Note:* You cannot deploy media that has the same version number as a deployment media that was previously deployed.

Click **Next**. The Hadoop Cluster Manager SSH credentials page opens.
7. Enter the root SSH account that has access to the cluster manager or enter a non-root SSH account if that account can execute sudo without entering a password. Click Next. The Specify the SAS Configuration and Deployment Directories page opens.

8. Enter the location of the SAS configuration and deployment directories:
Config Directory
Enter or navigate to the location of the `/standalone_installs` directory. This directory was created when your SAS Software Depot was created by the SAS Download Manager.

Deployment Directory
Enter or navigate to a working directory on the local server where the package or stack is placed.

Package Directory
Enter or navigate to a directory on the cluster manager host to which to send the package that the SAS Deployment Manager creates.

Click Next. The Checking System page opens and a check for locked files and Write permissions is performed.

9. If any files are shown in the text box after the system check, follow the instructions on the Checking System page to fix any problems.

Click Next. The Summary page opens.
10. Click **Start** to begin the configuration.

   *Note:* It takes several minutes to complete the configuration.

   If the configuration is successful, the page title changes to Deployment Complete.

   If warnings or errors occur, fix the issues and restart the configuration.

11. Click **Next** to close the SAS Deployment Manager.

---

### Creating the SASGRID Policy File

The SASGRID policy file specifies how SAS jobs use the resources with SAS Grid Manager for Hadoop. The policy file is an XML file that contains an entry for each type of processing that you want to perform on the grid (such as normal, priority, or night). Each type is identified as a GridApplicationType in the policy file. Within each application type entry, elements specify the details about how resources are allocated for those jobs. For example, an entry in the policy file for “normal” processing could specify the priority, memory, and number of cores used for jobs that should be run using normal processing.

Use the **appType** option on SAS jobs to specify the application type that a job should use. You can specify the **appType** in grid options. For example, including **appType=normal** in grid options specifies that jobs that use those grid options should use the parameters in the “normal” application type entry.

In addition, the SASGRID policy file contains **<HostGroup>** entries that enable you to specify which hosts in the Hadoop cluster are to be used for processing SAS grid jobs. If you want only a subset of hosts to be used for grid jobs, you can create a **<HostGroup>** entry that lists the available hosts, and then include that **<HostGroup>** entry in the application type entries. You can also create entries for groups of hosts that you want to
use for certain application types. For example, you can create a <HostGroup> entry for hosts that you want to use only for priority processing. Although you can also specify hosts individually in application type entries, using <HostGroup> enables you to specify groups that you can easily reuse in multiple entries.

If you do not create a SASGRID policy file, a job submitted using SAS Grid Manager for Hadoop can run on any host, and uses the maximum amount of memory and the minimum number of cores allowed by YARN (as defined by YARN properties in the yarn-site.xml file).

Use the policyFile grid option to specify the location of the policy file. You can specify a location on the file system of the grid client or a location in HDFS (prefix the path with hdfs://). If you do not specify a location using the policyFile option, SAS Grid Manager looks for a file named sasgrid-policy.xml in the directory that contains the cluster’s configuration files. If the application does not find the file in that location, it searches for the path /tmp/SASGrid/sasgrid-policy.xml in HDFS. See “Supported Job Options – SAS Grid Manager for Hadoop” on page 171 for more information.

<GridPolicy defaultAppType="name"> is the top-level element in the file. Use the defaultAppType attribute on the GridPolicy element to specify which application type should be used if a submitted job does not include an appType value.

Within the <GridPolicy> element are a series of <GridApplicationType> entries. Each entry contains some or all of these elements:

<GridApplicationType name="name">
  Specifies the name for the application type. The name must be in quotation marks.
  <jobname>
    Specifies the default job name if the application does not specify a name.
  </jobname>
  <priority>
    Specifies the priority for this job relative to other jobs. A higher number indicates a higher priority.
  </priority>
  <nice>
    Specifies the execution priority change, based on Linux “nice” values. Valid values are integers in the range of -20 to 19.
  </nice>
  <memory>
    The amount of physical memory (in megabytes) to reserve for the job.
  </memory>
  <vcores>
    The number of virtual machine cores to allocate to the job.
  </vcores>
  <runlimit>
    The maximum time (in minutes) that the job is allowed to run.
  </runlimit>
  <queue>
    Specifies the YARN scheduler queue that the job is submitted to. You create YARN queues when you set up YARN on your Hadoop cluster.
  </queue>
  <hosts>
    Specifies the hosts and host groups that the job can run on. Hosts are specified using the <host> element, and host groups are specified using the <hostGroup> element.
  </hosts>

A sample <GridApplicationType> entry looks like this:

  <GridApplicationType name="normal">
    <jobname>Normal Job</jobname>
    <priority>2</priority>
    <memory>1024</memory>
The `<GridPolicy>` element also includes `<HostGroup>` elements, which define host groups. Each `<HostGroup>` entry contains one or more `<host>` and `<HostGroup>` elements, which together specify which hosts are included in the group.

A sample `<HostGroup>` entry looks like this:

```xml
<hostGroup name="development">
  <host>myHost1.mydomain.com</host>
  <host>myHost2.mydomain.com</host>
</hostGroup>
```

A sample SASGRID policy file looks like this:

```xml
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<GridPolicy defaultAppType="normal">
  <GridApplicationType name="normal">
    <jobname>Normal Job</jobname>
    <priority>2</priority>
    <memory>1024</memory>
    <vcores>1</vcores>
    <runlimit>120</runlimit>
    <queue>default</queue>
    <hosts>
      <hostGroup>development</hostGroup>
      <host>myhost4.mydomain.com</host>
    </hosts>
  </GridApplicationType>
  <GridApplicationType name="priority">
    <jobname>High Priority</jobname>
    <priority>20</priority>
    <nice>-10</nice>
    <memory>1024</memory>
    <vcores>1</vcores>
    <queue>default</queue>
    <hosts>
      <hostGroup>development</hostGroup>
      <host>myHost4.mydomain.com</host>
    </hosts>
  </GridApplicationType>
  <HostGroup name="development">
    <host>myHost1.mydomain.com</host>
    <host>myHost2.mydomain.com</host>
  </HostGroup>
  <HostGroup name="test">
    <host>myHost3.mydomain.com</host>
    <hostGroup>development</hostGroup>
  </HostGroup>
</GridPolicy>
```
Sending Jobs to SAS Grid Manager for Hadoop

After you have created the SASGRID policy file, you can use application types that you defined to identify how SAS jobs are to be processed by the grid. Use any of the following methods to specify an application type to be associated with any jobs submitted to the SAS grid:

Grid server definition

Specify `AppType=application_type` in the Grid Options field of the YARN grid server definition (specify `YARN` in the Grid Provider field). This value is used if an AppType value is not otherwise specified when a job is submitted to the grid, so you might want to specify the application types used for normal processing. Specify `application_type` in the `jobOpts=` parameter of the `grdsvc_enable` function.

`grdsvc_enable` function

Specify `application_type` in the `jobOpts=` parameter of the `grdsvc_enable` function.

Grid options set

Grid options sets enable you to create sets of SAS options, required resources, and grid options. Each options set can then be mapped to a specified combination of a SAS application and a user or user group. When you create grid options sets for Grid Manager for Hadoop, go to the New Grid Options Set window and specify `AppType=application_type` in the Grid Options field. Specifying this value enables you to control the type of processing for jobs on the grid. See Chapter 10, “Working With Grid Options Sets,” on page 129 for more information.

Note: If you are using SAS Grid Manager for Hadoop and specify options in a grid options set that are specific to a Platform Suite for SAS grid, those options are ignored.

SASGSUB statements

Specify `AppType=application_type` as an option for the `-GRIDJOBOPTS` argument in a SASGSUB statement. You can also specify the application type in a grid options set and specify the grid options set in the `-GRIDOPTSET` argument. See “SASGSUB Syntax: Grid Job Arguments” on page 166.

If you are using SAS Grid Manager for Hadoop and the SAS Grid Manager Client Utility, you can restart grid jobs by using checkpoints and labels. See “Using SAS Checkpoint and Label Restart” on page 67 for more information.

Troubleshooting SAS Grid Manager for Hadoop

Debugging Grid Jobs

When a grid-enabled SIGNON is used or a job is submitted by SASGSUB, the sasgrid script is used to start a SAS session on the grid node. To aid in debugging, the output from the script (but not the SAS log and SAS output) is written to a unique log file. The file is located in the `GridServer/Logs` directory of the machine where the job runs. If the job is successful, the log file is deleted when the job finishes running. If the job is unsuccessful, the log file is saved.
The log file is named using the format
SASGrid.userID.hostname.process_ID.YYMMDDhhmmss.log

Debugging YARN

Log Sources
To debug SAS Grid Manager for Hadoop jobs in YARN, you can use the Hadoop Resource Manager application, and you can use information from these sources:

• SAS
• SAS Yarn AppMaster
• YARN server logs

YARN commands, and log entries. There are several types of logs that provide information about SAS Grid operation, and their locations differ depending on which grid provider you are using.

SAS Logging
Using the SAS logging facility, set the app.tk.tkegrid.yarn logger value to debug or trace so that logging messages are sent to the appender.

You can also use the SASGrid.*.log file, which is located at <config>/LevX/<SASAppServerContext>/GridServer/Logs.

SAS YARN AppMaster Logs
You can access the YARN AppMaster logs by using the Hadoop Resource Manager application or the command yarn logs -applicationID <applicationID>.

YARN Server Logs
The YARN Resource Manager log is contained in the file *-resourcemanager-<host>.log. The Node Manager log is contained in the file *-nodemanager-<host>.log. The location of these logs differs depending on the Hadoop distribution that you are using:

Hortonworks
    /var/log/hadoop-yarn

Cloudera
    /var/log/hadoop-yarn

MapR
    /opt/mapr/hadoop/hadoop-<ver>/logs

Scheduling with SAS Grid Manager for Hadoop

SAS Grid Manager for Hadoop uses Oozie for scheduling SAS workflows. To implement scheduling using Oozie, you define an Oozie scheduling server in SAS metadata and specify that server when deploying SAS DATA step jobs for scheduling. Only SAS DATA Step jobs are supported for scheduling. When a SAS DATA Step job is sent to the server, it generates these two XML files:
workflow file
contains a list of tasks that must be performed for the scheduled job

cooridnator file
contains the triggers that control when the job is to be run

Because of the functions supported by Oozie, the Oozie scheduling server provides support only for simple schedules, allowing only a single time event as a trigger, supporting only job events as flow dependencies, and only allowing AND conditions in a process flow diagram.

See *Scheduling in SAS* for complete information on scheduling with Grid Manager for Hadoop using an Oozie scheduling server.
Part 4

Common SAS Grid Manager Features

Chapter 8
  *Installing and Configuring Common Components* .......................... 99

Chapter 9
  *Enabling SAS Applications to Run on a Grid* ............................. 105

Chapter 10
  *Working With Grid Options Sets* ............................................. 129
Overview

Some aspects of configuring and using SAS Grid Manager are the same regardless of whether you are using Platform Suite for SAS or SAS Grid Manager for Hadoop as your grid provider. These parts of the SAS grid infrastructure are used in both environments and follow the same procedure for installation and configuration:

- logical grid server definitions
- SAS Grid Manager Client Utility

Likewise, the methods that end users of SAS applications use to submit jobs to SAS Grid Manager do not differ between grid providers. From the user’s perspective, they do not know (and do not need to know) whether the job that they submit is being processed on a Platform Suite for SAS grid or on a SAS Grid Manager for Hadoop grid.

Common Metadata Definitions

Modifying SAS Logical Grid Server Definitions

The initial configuration of the logical grid servers is performed by the SAS Deployment Wizard. However, a SAS grid administrator might need to modify the existing grid metadata or add new grid metadata definitions.

A SAS administrator performs these steps to specify or modify the required and optional properties as metadata for the SAS Grid Server:

1. In SAS Management Console, open the metadata repository that contains the metadata for the Logical Grid Server.
2. In the navigation tree, select Server Manager.
3. Expand the folders under Server Manager until you see the metadata objects for the SAS application server, such as SASApp, and its Logical Grid Server component.

4. Expand the Logical Grid Server component so that you see the metadata object for the Grid Server.

5. Right-click the metadata object for the Grid Server, and select Properties.

6. In the Grid Server Properties window, click the Options tab.

**Figure 8.1 Grid Server Properties**

![Grid Server Properties window]

7. The fields on the Options tab are:

**Provider**
the grid middleware provider. This value is used to communicate with the grid control server.

**Grid Command**
the script, application, or service that Platform Suite for SAS uses to start server sessions on the grid nodes. Any SAS options that are included in this command are passed to the grid jobs. You can specify any additional SAS options in the SAS Options field.

This value is the path to the sasgrid.cmd file (Windows) or sasgrid script file (UNIX). Because this same command is used to start the servers on all grid nodes, the path to the directory on each grid node must be the same. For example: `C:\SAS\Grid\Lev1\SASApp\GridServer\sasgrid`.

**SAS Options**
specifies the SAS options that are applied by default to jobs that are sent to the grid. Separate multiple options with a space. Any SAS options that are specified for a specific application mapping override this value.
Required Resources
specifies the name of the resources that are applied by default to jobs that are sent to the grid. Any resources that are specified for a specific application mapping are added to this value. The application server name is displayed in this field by default, but it can be removed if needed.

Grid Options
specifies the grid command options that are applied by default to jobs that are sent to the grid. Examples include the job priority, the job queue, or the user group that is associated with the job. Any grid options that are specified by a grid options set and mapped to a specific application override this value. Job options are specified as name/value pairs in this format. Multiple pairs are separated by semicolons:

    option-1=value-1;option-2="value-2 with spaces";
    ... option-n='value-n with spaces';

Here is an example of additional options that specify that all jobs that use this logical grid server go to the priority queue in the project “payroll”:

    queue=priority; project='payroll'

For a complete list of job options for Platform Suite for SAS, see Appendix 1, “Supported Job Options - Platform Suite for SAS,” on page 169.

For a complete list of job options for SAS Grid Manager for Hadoop, see “Supported Job Options – SAS Grid Manager for Hadoop” on page 171.

Grid Options Set Mappings
lists the grid server’s mappings of grid options sets to applications for a specified user or group. Each grid options set associates a set of SAS options, required resources, and grid options to a SAS application and user identities that use the grid. To change or remove an entry, click on the entry and select Edit or Delete from the pop-up menu. Click New to create a new application mapping. See Chapter 10, “Working With Grid Options Sets,” on page 129 for more information about grid options sets.

Advanced Options
enables you to specify the workload and module name. The workload specifies the optional resources that can be processed on the node. The module name specifies the shared library name or Java class name of the grid middleware provider's support plug-in.

8. After you complete the field entries, click OK to save the changes and close the Grid Server Properties window.

9. In the display area (right-hand side) on SAS Management Console, right-click the Connection object for the Grid Server, and then select Properties.

10. In the Properties window for the Grid Server Connection, click the Options tab. Here are the fields on this tab:

    Authentication Domain
    the authentication domain that is used for connections to the server. Set this value to <none> for Platform LSF.

    Grid Server Address
    the host name or network address of the grid control server.

    Grid Server Port
    the port that is used to connect to the grid control server. This value should always be set to 0 (zero) for Platform LSF.
Putting and Configuring SAS Grid Manager Client Utility

**Installation Overview**
The SAS Grid Manager Client Utility enables users to submit SAS programs to a grid for processing without requiring SAS to be installed on the machine that is performing the submission. Platform Suite for SAS must be installed on any machine on which the SAS Grid Manager Client Utility runs.

The SAS Grid Manager Client Utility is automatically installed and configured using the SAS Deployment Wizard if the utility is in the plan file.

**Installation Prerequisites**
The configuration for the SAS Grid Manager Client Utility assumes that all of the following actions have been performed:

- The grid control server has already been installed. The configuration must retrieve the logical grid server definition from metadata.
- The user name under which jobs are submitted is defined in metadata. If not, jobs submitted to the grid fail.

**Configuring the SAS Grid Manager Client Utility**
The amount of user input that is required during the installation and configuration process depends on whether you choose an Express, Typical, or Custom install. For information about running the SAS Deployment Wizard, see *SAS Deployment Wizard User's Guide*.

1. The SAS Grid Manager Client Utility: Options page enables you to specify the user credentials that are used to connect to the SAS Metadata Server, the method for transferring files to and from the grid (either through a shared file system or remote copy), and the path to a SAS license file that contains a SAS Grid Manager license (only shown during a custom installation). By default, the metadata is searched for the SAS license file.
2. If you choose to use remote copy (also known as staging) to transfer files to and from the grid, the SAS Grid Manager Client Utility: Staged File Options page is displayed. This page enables you to specify the path to the directory used to stage files moving into and out of the grid. If you are performing a custom installation, you can also specify the staging host and the path to the staging directory as seen by the staging host.

**Figure 8.3** SAS Grid Manager Client Utility: Staged File Options Page
If you choose to use a shared directory to copy files to and from the grid, the SAS Grid Manager: Shared Directory Options page appears. This page enables you to specify the grid-shared directory on the grid control server.

Figure 8.4 SAS Grid Manager Client Utility: Shared Directory Options Page

Using the SASGSUB Configuration File

Most of the options that are used by the SAS Grid Manager Client Utility are contained in the sasgsub.cfg file. This file is automatically created by the SAS Deployment Wizard. These options specify the information that the SAS Grid Manager Client Utility uses every time it runs. The sasgsub.cfg file is located in the Applications/SASGridManagerClientUtility/<version> directory of the configuration directory. The following information from the SAS Deployment Wizard is collected in the sasgsub.cfg file:

- information to connect to the SAS Metadata Server (SAS Metadata Server name, port, user ID, and password). By default, the metadata password value is set to _PROMPT_, and the user is prompted for a password.

- the path that is used to store files used by the grid. If you are using a shared file system, then this is the path to the shared file system. If you are staging files, then this is the location where grid clients store files that are retrieved by the grid.

- the name of the SAS Application Server that contains the logical grid server definition.
Chapter 9
Enabling SAS Applications to Run on a Grid

Overview of Grid Enabling .................................................. 106
Submitting SAS Programs to a Grid Interactively .................. 106
  Overview ............................................................................. 106
  Submitting Jobs from the Program Editor to the Grid ............. 107
  Viewing LOG and OUTPUT Lines from Grid Jobs .................. 108
  Using the SAS Explorer Window to Browse Libraries .......... 108
  Running Interactive SAS Sessions Using the SAS Grid Manager Client Utility . 108
  Running Interactive Commands Using the SAS Grid Manager Client Utility . 109

Submitting Batch SAS Jobs to the Grid ............................... 109
  Overview ............................................................................. 109
  Grid Manager Client Utility File Handling ......................... 109
  Submitting Jobs in Batch Using the SAS Grid Manager Client Utility . 110
  Running Commands in Batch Using the SAS Grid Manager Client Utility . 110
  Viewing Job Status Using the SAS Grid Manager Client Utility . 111
  Ending Jobs Using the SAS Grid Manager Client Utility .......... 111
  Retrieving Job Output Using the SAS Grid Manager Client Utility . 111
  Retrieving a SAS Grid Manager Client Utility Log ................. 112
  Monitoring Batch Processing Using the SAS Grid Manager Client Utility . 112
  Using a Grid without a Shared Directory ......................... 112

Scheduling Jobs on a Grid ................................................. 113
Comparing Grid Submission Methods .................................. 114
Enabling Distributed Parallel Execution of SAS Jobs ............. 114
Using SAS Studio with a SAS Grid ...................................... 115
Using SAS Enterprise Guide and SAS Add-In for Microsoft
  Office with a SAS Grid .................................................. 116
    Types of Grid Enablement .............................................. 116
    Parallel Execution and Grid Enablement ......................... 117
    Using Extended Attributes to Specify Grid Usage ............. 118
    Assigning Libraries in a Grid ........................................ 118
    Developing SAS Programs Interactively Using a Grid ....... 118
Using SAS Stored Processes with a SAS Grid ....................... 119
Using SAS Data Integration Studio with a SAS Grid .............. 120
  Scheduling SAS Data Integration Studio Jobs on a Grid ....... 120
  Multi-User Workload Balancing with SAS Data Integration Studio . 120
  Parallel Workload Balancing with SAS Data Integration Studio . 121
  Specifying Workload for the Loop Transformation .............. 123
Overview of Grid Enabling

After you have configured your grid, you can configure your SAS applications and programs to take advantage of the grid capabilities. Some SAS applications require you to change only an option to take advantage of the grid; other applications require more extensive changes.

Submitting SAS Programs to a Grid Interactively

Overview

You can use SAS Display Manager as a client to submit SAS programs to the grid for execution. The results of the execution are returned to the local workstation. When you submit a SAS program from a SAS Display Manager client to execute on a grid, the program runs on a grid machine in a separate session with its own unique work library. The SAS log and output of the grid execution are returned to the local workstation. You might need to perform additional actions in order to view data from the SAS Display Manager session that was created or modified by the program that ran on the grid. For example, modifications might be required in order to use the Explorer to browse SAS libraries that are modified by grid execution.

Starting with the first maintenance release after SAS 9.4, you can also use the SAS Grid Manager Client Utility to start an interactive session on the grid or watch the output from a batch session. An interactive session enables you to perform functions in interactive mode or batch monitor mode:

Interactive mode
- Run SAS in line mode.
- Run SAS in Display Manager mode.
- Run and interact with a command (watch the output and provide input).

Batch monitor mode
- Run SAS in batch mode and watch the output of the session.
- Run a command in batch mode and watch the output.

Using SAS line mode or SAS Display Manager mode requires that the grid nodes are UNIX machines. Also, SAS Display Manager mode requires that you have X Server on the client machine.

If you terminate the SAS Grid Manager Client Utility session while performing any interactive mode function, the remote job is also terminated. Terminating the session while running a batch monitor mode function does not terminate the batch job.
Submitting Jobs from the Program Editor to the Grid

The first step in integrating SAS processes with the grid is to get your SAS programs running on the grid.

In order to submit a SAS program to the grid, you must add a set of grid statements to the program. For programs submitted through the SAS Program Editor, you can save the statements to an external file and then specify a key definition that issues the statements. Submit the contents of the SAS Program Editor window to the grid, rather than to the local workstation.

Some of the examples in this topic use SAS/CONNECT statements (such as signon, rsubmit, and signoff). For detailed information about these statements, see *SAS/CONNECT User’s Guide*.

**Note:** This procedure does not work if the Explorer window is open in your SAS session.

To add grid statements to a program and submit the program to the grid, follow these steps:

1. Save these statements to an external file, referred to as *grid-statement-file* (for example, c:\gpre.sas):

   ```sas
   %global count;
   %macro gencount;
   %if %bquote(&count) eq %then %do; %let count=1;%end;%else %let count=%eval(&count+1);
   %mend;
   %gencount;
   options metaserver='metadata-server-address';
   options metaport=metadata-server-port;
   options metauser=username;
   options metapass="password";
   %let rc=%sysfunc(grdsvc_enable(grid&count, server=SASApp));
   signon grid&count;
   ``

   *metadata-server-address* is the machine name of the SAS Metadata Server, and *metadata-server-port* is the port used to communicate with the metadata server.

2. Open the Keys window and specify the following for an available key (for example, F12):

   ```sas
   gsubmit "%include 'grid-statement_file';";
   rsubmit grid&count wait=no persist=no;
   ``

   *grid-statement-file* is the path and filename of the file (for example, c:\gpre.sas) containing the grid statements.

3. Type or include a SAS program in the Program Editor window, and then press the key assigned to the grid statements. The program is automatically submitted to the grid for processing. Your local machine is busy only until the program is submitted to the grid.

Using the same key to submit multiple jobs causes multiple jobs to be executed in parallel on the grid. Anything specified in the WORK library is discarded when the grid session ends and is not available to later steps in the code.
Viewing LOG and OUTPUT Lines from Grid Jobs

The example in “Submitting Jobs from the Program Editor to the Grid” on page 107 uses asynchronous rsubmits. This causes the results of the execution to be returned to the local log and output windows only after the entire program finishes execution on the grid. To cause the log and output lines to be displayed while the program is executing, delete the options noconnectwait; line in the program.

The rsubmit executes synchronously, and the returned log and output lines are displayed while the job is executing. This also results in the Client SAS session being busy until the entire grid job has completed. You cannot submit more code until the job completes.

Using the SAS Explorer Window to Browse Libraries

The Client SAS session and the grid SAS session are two separate instances of SAS. Any code or products needed to access data must be submitted and available on both the client machine as well as the grid nodes. Use the following steps to browse libraries from the SAS Explorer Window that are accessed and modified by jobs executing in the grid:

1. Define all of your SAS libraries within SAS metadata under your server context (for example, under SASApp).

2. Ensure that the following option is in the SAS invocation in the sasgrid script file used to start SAS on the grid nodes. This option should have been added by the SAS Deployment Wizard.

   metaautoresources SASApp

   SASApp is the name of your application server context.

3. Include this option on the Client SAS session invocation on the workstation.

   metaautoresources SASApp

   SASApp is the name of your application server context.

   Note: If you are accessing data through any SAS/ACCESS product, you must license the SAS/ACCESS products on the SAS Client machine in order to be able to browse those libraries from the SAS Explorer. The SAS/ACCESS products must also be licensed on the grid nodes in order to enable the job to access data during execution.

   Each SAS session executing on the grid is a unique session with a unique WORK library. In order to view the work libraries that are created on each of the grid nodes, you must add the following line after the signon statement in the code provided in “Submitting Jobs from the Program Editor to the Grid” on page 107:

   libname workgrid slibref=work server=grid&count;

   grid&count is the label used as the remote session ID in the signon statement.

Running Interactive SAS Sessions Using the SAS Grid Manager

Client Utility

To start an interactive SAS session on a UNIX grid using the SAS Grid Manager Client Utility and SAS Display Manager, issue the following command from an operating system command line:
To start an interactive SAS session in line mode on a UNIX grid using the SAS Grid Manager Client Utility, issue the following command from an operating system command line:

<path>/SASGSUB -GRIDRUNSASDM

The path option specifies the path for the SASGSUB program. By default, the location is <configuration_directory>/Applications/SASGridManagerClientUtility/<version>.

The -GRIDRUNSASDM or -GRIDRUNSASLM option starts the interactive SAS session on the grid. Once the session has started, you can submit SAS code or programs that will run on the machine in the grid. There are other arguments, some required and some optional, that you can include in the SASGSUB statement. These arguments are passed to the grid when the SAS session is started and include parameters such as the metadata connection information, grid application server name, and workload resource names. See “SASGSUB Syntax: Starting SAS in Interactive Mode” on page 155 for complete syntax information.

Running Interactive Commands Using the SAS Grid Manager Client Utility

To interactively run a command on a grid using the SAS Grid Manager Client Utility, issue the following command from a SAS command line:

<path>/SASGSUB -GRIDRUNCMDINT command

Standard input, standard output, and standard error are directed to your command prompt window.

Submitting Batch SAS Jobs to the Grid

Overview

The SAS Grid Manager Client Utility enables you to run SAS jobs on the grid in batch. You can also use the utility to check job status, end a job, and retrieve job output. Most of the options that are used by the SAS Grid Manager Client Utility are contained in the sasgsub.cfg file. This file is automatically created by the SAS Deployment Wizard. These options specify the information that the SAS Grid Manager Client Utility uses every time it runs.

The SAS Grid Manager Client Utility and Platform LSF must be installed on any machine on which the SAS Grid Manager Client Utility runs.

Grid Manager Client Utility File Handling

This is how files are handled by the SAS Grid Manager Client Utility when processing a job on the grid in batch mode:

1. SASGSUB creates a job directory in the GRIDWORK directory under the directory of the user who is submitting the job. For example, if GRIDWORK is /grid/share and the submitting user is sasuser1, then a job directory is created in /grid/share/sasuser1 for the files.
2. SASGSUB copies the SAS program and any files specified by GRIDFILESIN into the new directory.

3. SASGSUB submits a job to the grid that includes information about the location of the job directory. It uses either GRIDWORK or GRIDWORKREM to specify the location of the job information to the grid. If you are staging files, SASGSUB also passes the stage file command specified by the GRIDSTAGECMD option to the grid.

4. If the grid job is using staging when the job starts, the grid copies the files in the job directory under GRIDWORK to a temporary job directory. The temporary directory is in the grid's shared directory location that is specified during the SAS Deployment Wizard installation process.

5. The grid runs the SAS program from the job directory and places the LOG and LST file back into the same job directory. For a shared file system, this directory is the one specified by the GRIDWORK option. This is also the directory that SASGSUB copied files into. If you are staging files, this directory is the job directory that is in the grid shared directory.

6. If you are staging files, after the job is complete, the files in the job directory in the grid shared location are copied to the job directory that is specified by the GRIDWORK option.

7. At this point in processing, the job directory in GRIDWORK contains all of the files that are required and produced by SAS batch processing. You can then retrieve the files using the GRIDGETRESULTS command.

**Submitting Jobs in Batch Using the SAS Grid Manager Client Utility**

To submit a SAS job in batch mode to a grid using the SAS Grid Manager Client Utility, issue the following command from an operating system command line:

```
<path>/SASGSUB -GRIDSUBMITPGM sas-program-file
```

The `path` option specifies the path for the SASGSUB program. By default, the location is `<configuration_directory>/Applications/SASGridManagerClientUtility/<version>`.

The `-GRIDSUBMITPGM` option specifies the name and path of the SAS program that you want to submit to the grid.

In addition, you can specify other options that are passed to the grid or used when processing the job, including workload resource names. For a complete list of options, see “SASGSUB Syntax: Submitting a SAS Program in Batch Mode” on page 154.

Specifying the `-GRIDWATCHOUTPUT` argument displays the standard output and standard error of the submitted batch job on your machine.

**Running Commands in Batch Using the SAS Grid Manager Client Utility**

To submit a command to a grid in batch mode using the SAS Grid Manager Client Utility, issue the following command from an operating system command line:

```
<path>/SASGSUB -GRIDRUNCMD command
```
Viewing Job Status Using the SAS Grid Manager Client Utility

After you submit a job to the grid, you might want to check the status of the job. To check the status of a job, issue the following command from a command line:

```bash
<path>/SASGSUB -GRIDGETSTATUS [job-ID | ALL]
```

- `GRIDGETSTATUS` specifies the ID of the job that you want to check, or ALL to check the status of all jobs submitted by your user ID. For a complete list of options, see “SASGSUB Syntax: Viewing Job Status” on page 158.

The following is an example of the output produced by the SASGSUB -GRIDGETSTATUS command.

Output 9.1  Output Produced by SASGSUB -GRIDGETSTATUS Command

<table>
<thead>
<tr>
<th>Current Job Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Job information in directory U:\pp\GridSub\GridWork\user1\SASGSUB-2008-11-24_13.17.17.327_testPgm is invalid.</td>
</tr>
<tr>
<td>Job 1925 (testPgm) is Submitted: Submitted: 08Dec2008:10:28:57</td>
</tr>
</tbody>
</table>

Ending Jobs Using the SAS Grid Manager Client Utility

If a job that has been submitted to the grid is causing problems or otherwise needs to be terminated, use the SAS Grid Manager Client Utility to end the job. Issue the following command from a command line:

```bash
<path>/SASGSUB -GRIDKILLJOB [job-ID | ALL]
```

- `GRIDKILLJOB` specifies the ID of the job that you want to end, or ALL to end all jobs submitted by your user ID. For a complete list of options, see “SASGSUB Syntax: Ending a Job” on page 158.

Retrieving Job Output Using the SAS Grid Manager Client Utility

After a submitted job is complete, use the SAS Grid Manager Client Utility to retrieve the output produced by the job. Issue the following command from a command line:

```bash
<path>/SASGSUB -GRIDGETRESULTS [job-ID | ALL] -GRIDRESULTSDIR
```

- `GRIDGETRESULTS` specifies the ID of the job whose results you want to retrieve, or you can specify ALL to retrieve the results from all jobs submitted by your user ID.
- `GRIDRESULTSDIR` specifies the directory in which the jobs results should be moved. When the results are retrieved, they are removed from the GRIDWORK directory, which keeps this directory from filling up with completed jobs. If you do not specify this parameter, the results are copied to a job subdirectory in the current directory.

A file named job.info is created along with the job output. This file contains information about the execution of the job, including the submit time, start time, end time, the machine on which the job ran, the job ID, and the return code from the SAS program.
The following is an example of the output produced by the SASGSUB -GRIDGETRESULTS command.

Output 9.2  Output Produced by SASGSUB -GRIDGETRESULTS Command

```
Current Job Information
    Moved job information to .\SASGSUB-2008-11-21_21.52.57.130_testPgm


    Moved job information to .\SASGSUB-2008-11-24_13.16.06.060_testPgm

Job 1925 (testPgm) is Submitted: Submitted: 08Dec2008:10:53:34
```

Retrieving a SAS Grid Manager Client Utility Log

After a submitted job is complete, you can find the SAS program log file for the job in this location: `GRIDWORK/user id/SASGSUB-YYYY-MM-DD_HH:MN_SS_mmm_job_name/program_name.log`

The SAS Grid Manager Client Utility uses the standard SAS logging facility. Output from the SAS Grid Manager Client Utility is directed to the console unless you use the SAS logging facility to create a log. See the -LOGCONFIGLOC option in “SASGSUB Syntax: Submitting a SAS Program in Batch Mode” on page 154 for a list of the supported logging keys.

Monitoring Batch Processing Using the SAS Grid Manager Client Utility

When you use SASGSUB to submit a program or a command to the grid in batch mode, you can use the -GRIDWATCHOUTPUT argument to interactively monitor the processing on the grid. The option specifies that the output of what was submitted by the SASGSUB command is displayed on your machine. If you use this argument when submitting a SAS program using -GRIDSUBMITPGM, the SAS log and output are displayed. If you use the argument when submitting a command using -GRIDRUNCMD, the command’s standard output and standard error are displayed. While the output is being displayed, entering the command prompt does not affect the processing in the grid.

If you terminate the SASGSUB session while in interactive monitoring mode, the batch job continues to run and does not terminate.

Using a Grid without a Shared Directory

If your grid configuration does not permit a directory structure to be shared between the grid client machines and the grid nodes, you can specify that the grid job move files into the grid before processing and move files out of the grid when the job is complete. The file movement (called file staging) is performed by the grid job using a remote copy program such as rcp, scp, or lsrcp. When using file staging, files are moved into and out of the grid using the GRIDWORK directory. The SAS Grid Manager Client Utility
passes information to the grid that indicates which files need to be sent to the grid and where the files are located. After the grid processes the job, the results are copied back to the GRIDWORK directory. If the user is offline, the results are held in the shared file system until they are retrieved.

During the installation process, the SAS Deployment Wizard enables you to specify whether you will use a shared directory or if you will be staging files. If you specify that you will be staging files, you must also specify the staging command that you want to use to move the files (rcp, lsrcp, scp, pscp, or smbclient). You can also specify the host that you will use to stage files to and from the grid, if you are not using the current host.

To submit jobs to a grid without a shared file system, follow these steps:

1. Use the GRIDSTAGECMD parameter on the SASGSUB command to specify the transfer method to use for moving the files from the staging directory to the grid.

2. If the machine that stages the files is not the current host, use the GRIDSTAGEHOST parameter on the SASGSUB command to specify the host that is used to stage the files. For example, use this parameter if you are using a laptop to submit jobs to the grid and then disconnecting or shutting down the laptop before the jobs are completed or submitted. The laptop must have a GRIDWORK directory on a file server that is always available to the grid. Use the GRIDSTAGEHOST command to specify the file server host name.

Scheduling Jobs on a Grid

Using the scheduling capabilities, you can specify that jobs are submitted to the grid when a certain time has been reached or after a specified file or job event has occurred (such as a specified file being created).

To schedule a job to run on a grid, follow these steps:

1. Deploy the job for scheduling.

   Some SAS applications, such as SAS Data Integration Studio, include an option to deploy jobs for scheduling. If you want to schedule an existing SAS job, use the Deploy SAS DATA Step Program option in the Schedule Manager plug-in of SAS Management Console.

2. Use the Schedule Manager plug-in in SAS Management Console to add the job to a flow.

   A flow contains one or more deployed jobs as well as the schedule information and time, file, or job events that determine when the job runs.

3. Assign the flow to a scheduling server and submit the flow for scheduling.

   You must assign the flow to a Platform Process Manager scheduling server in order for the scheduled job to run on the grid.

For detailed information about scheduling, see Scheduling in SAS.
Comparing Grid Submission Methods

You can use the SAS Grid Manager Client Utility, the Schedule Manager plug-in to SAS Management Console, and SAS language statements to submit jobs to the grid. The following table compares the methods.

Table 9.1 Comparison of Grid Submission Methods

<table>
<thead>
<tr>
<th>Feature</th>
<th>SAS Grid Manager Client Utility</th>
<th>Schedule Manager Plug-in</th>
<th>SAS Language Statements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interface</td>
<td>Command line</td>
<td>SAS Management Console interface</td>
<td>SAS language syntax</td>
</tr>
<tr>
<td>Duration of client connection</td>
<td>Duration of the submission</td>
<td>Duration of the submission</td>
<td>Duration of the execution</td>
</tr>
<tr>
<td>Minimum client installation requirements</td>
<td>SAS Grid Manager Client Utility and Platform LSF</td>
<td>SAS Management Console</td>
<td>Base SAS, SAS/CONNECT, Platform LSF</td>
</tr>
<tr>
<td>Support for checkpoint restart</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Support for SAS options, grid options, and policies</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Support for event-triggered workflow execution</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Enabling Distributed Parallel Execution of SAS Jobs

Some SAS programs contain multiple independent subtasks that can be distributed across the grid and executed in parallel. This approach enables the application to run faster. To enable a SAS program to use distributed parallel processing, add RSUBMIT and ENDRSUBMIT statements around each subtask and add the GRDSVC_ENABLE function call. SAS Grid Manager automatically assigns each identified subtask to a grid node.

You can use the SAS Code Analyzer to automatically create a grid-enabled SAS job. To use the SAS Code Analyzer, add PROC SCAPROC statements to your SAS program, specifying the GRID parameter. When you run the program with the PROC SCAPROC statements, the grid-enabled job is saved to a file. You can then run the saved SAS job on the grid, and SAS Grid Manager automatically assigns the identified subtasks to a grid node.
This is an example of the syntax for the SAS Code Analyzer:

```sas
proc scaproc;
   record '1.txt' grid '1.grid':
run;
remainder of SAS program...
```

For complete information and syntax for the PROC SCAPROC statement, see *Base SAS Procedures Guide*.

This is an example of the syntax used for enabling distributed parallel processing:

```sas
%let rc=%sysfunc(grdsvc_enable(_all_,server=SASApp));
options autosignon;
rsubmit task1 wait=no;
   /* code for parallel task #1 */
endrsubmit;
rsubmit task2 wait=no;
   /* code for parallel task #2 */
endrsubmit;
   ...
rsubmit taskn wait=no;
   /* code for parallel task #n */
endrsubmit;
waitfor _all_ task1 task2 . . . taskn;
signoff _all_
```

For more information, see “GRDSVC_ENABLE Function” on page 137.

For detailed syntax information, see *SAS/CONNECT User's Guide*.

---

### Using SAS Studio with a SAS Grid

SAS Studio can take advantage of the processing capabilities of a SAS Grid. The approach to take to add grid support depends on the edition of SAS Studio that you are using.

With all editions of SAS Studio, you can use the SAS/CONNECT GRDSVC_ENABLE, SIGNON, and RSUBMIT functions in your code to send the code to the grid. Use the GRDSVC_ENABLE function to identify the SAS/CONNECT session that should be sent to the grid, specify that session name in the SIGNON statement, and then submit the code to the session in the RSUBMIT statement.

For example, you could use the following code to specify processing on the grid:

```sas
%let rc=%sysfunc(grdsvc_enable(gridsess,server=SASApp));
options autosignon;
rsubmit gridsess wait=no;
   /* code to be processed on the grid */
endrsubmit;
waitfor gridsess;
signoff gridsess;
```


If you are using SAS Studio Enterprise Edition together with SAS Grid Manager, you can convert the SAS Studio workspace server to use grid load balancing. During the conversion process, you can choose to launch the workspace server either by the grid or
by the object spawner on the least-used machine in the grid. After you convert the workspace server, the grid load-balanced server processes the SAS statements in SAS Studio. See “Using SAS Grid Manager for Server Load Balancing” on page 125 for more information.

In order to run code from any edition of SAS Studio on the grid, you must have SAS Grid Manager licensed and Platform Computing’s LSF installed on the same machine as the SAS Studio workspace server.

If you are using version 3.1 of SAS Studio, modify the SAS Studio property metadata to specify that the application is grid capable. Identifying SAS Studio as grid capable enables you to associate grid options sets with the application. Grid options sets are collections of grid options, SAS options, and required resources that you can map to users and applications (in this case, SAS Studio). For more information, see Chapter 10, “Working With Grid Options Sets,” on page 129.

To specify that SAS Studio is grid capable, use SAS Management Console to add the keyword `isGridCapable:*SAS Studio Mid-Tier*` to the metadata for the SAS Studio folder. See “Specifying That an Application Is Grid Capable” on page 133 for complete instructions. If you are using version 3.2 or higher of SAS Studio, you do not need to modify the metadata.

---

Using SAS Enterprise Guide and SAS Add-In for Microsoft Office with a SAS Grid

**Types of Grid Enablement**

Jobs generated by SAS Enterprise Guide and SAS Add-In for Microsoft Office can take advantage of a SAS grid using one or a combination of these approaches:

- **Using server-side load balancing through a grid-launched workspace server.** When SAS Enterprise Guide or SAS Add-In for Microsoft Office request the use of a workspace server to run a job, the object spawner launches a new workspace server in the grid on a node that meets the criteria of the queue to which the job was submitted. The job then runs on the workspace server. See “Using SAS Grid Manager for Server Load Balancing” on page 125 for more information about specifying grid load balancing on a workspace server.

- **Using server-side load balancing through a grid load-balanced workspace server.** When SAS Enterprise Guide or SAS Add-In for Microsoft Office request the use of a workspace server to run a job, the object spawner uses SAS Grid Manager to automatically launch the server on the least busy node in the grid. The code from SAS Enterprise Guide or SAS Add-In for Microsoft Office is then run on the workspace server. See “Using SAS Grid Manager for Server Load Balancing” on page 125 for more information about specifying grid load balancing on a workspace server.

- **Using SAS Enterprise Guide or SAS Add-In for Microsoft Office’s ability to enable projects or tasks to run on the grid.** If you are using SAS Enterprise Guide or SAS Add-In for Microsoft Office 5.1 or later, the code for the project or the task is wrapped in RSUBMIT/ENDRSUBMIT statements and automatically sent to the SAS grid for processing. To specify this type of processing, perform one of the following actions:
  - Specify **Use grid if available** in the Project Properties window or the Task Properties window.
• Specify the **Initialize Grid (if available)** option when connecting to a workspace in the Options dialog box under SAS Programs.

• Set the EGGridPolicy or the AMOGridPolicy extended attribute on the logical grid server to a value of *Force*. See “Using Extended Attributes to Specify Grid Usage” on page 118 for more information about these attributes.

**Note:** These options are not needed if you use grid-launched workspace servers.

If you are using grid-launched workspace servers and do not want the **Use grid if available** option to be available, set the EGGridPolicy or the AMOGridPolicy to a value of *Ignore*. The *Ignore* value is available in the first maintenance release after version 6.1 of SAS Enterprise Guide and SAS Add-In for Microsoft Office.

• Using SAS Enterprise Guide or SAS Add-In for Microsoft Office’s ability to insert custom code before and after tasks and SAS programs to route the processing to the grid. If you are using SAS Enterprise Guide 4.3 or earlier, you can wrap the code for the project or the task in RSUBMIT/ENDRSUBMIT statements through the use of custom code options that are inserted before and after the code that is submitted with the program. You must also modify the application servers, the SAS Enterprise Guide configuration files, and the SAS Add-In for Microsoft Office options. Download the code and instructions from [http://support.sas.com/rnd/scalability/grid/download.html](http://support.sas.com/rnd/scalability/grid/download.html).

**Note:** This option is not needed if you use grid-launched workspace servers.

If you use grid-launched workspace servers, you should not specify any of the other grid-related options. If you are using the first maintenance release of version 6.1 of SAS Enterprise Guide or SAS Add-In for Microsoft Office, you can prevent users from using the grid options by specifying a value of *Ignore* for the EGGridPolicy or the AMOGridPolicy extended attribute on the logical grid server. See “Using Extended Attributes to Specify Grid Usage” on page 118 for more information about these attributes.

If you are not using grid-launched workspace servers, you must specify grid load balancing options, and you might have to wrap your code in RSUBMIT/ENDRSUBMIT statements. The statements are required if you are using SAS Enterprise Guide or SAS Add-In for Microsoft Office 4.3 or earlier. If you are using SAS Enterprise Guide or SAS Add-In for Microsoft Office 5.1 or later, the statements are not required.

**Parallel Execution and Grid Enablement**

The option **Allow parallel execution on the same server** is not an option to control processing on the grid, but it can affect how grid-enabled jobs are processed. The option causes SAS Enterprise Guide to create a workspace server for each parallel process in a project, regardless of whether the project is grid-enabled or not. The following list explains how this option interacts with the **Use grid if available** option:

Neither option is selected

All code runs in a single workspace server.

Only **Use grid if available** option selected

A workspace server starts, performs a grid-enabled SIGNON, and submits code to the grid.

Only **Allow parallel execution** option selected

Multiple workspace servers are started, one for each parallel code path in a project.

Both options selected

Multiple workspace servers are started. Each server signs on to the grid. Each parallel code path in a project is submitted to the grid.
Selecting both options might cause problems because the GRIDWORK and RMTWORK libraries are duplicated in each workspace server. This causes problems for SAS Enterprise Guide when it tries to read from one of those libraries. Selecting the **Allow parallel execution** option might cause problems because one execution stream might need to use or write to a data set that is in use by another stream.

If you are using grid-launched workspace servers, you do not need to enable the **Use Grid if available** option because the workspace server for SAS Enterprise Guide is already running on the grid. The **Allow parallel execution** option can improve run-time efficiency, but it is more complex because it can run multiple and separate SAS sessions. When possible, perform debug and testing operations with the **Allow parallel execution** option disabled.

**Using Extended Attributes to Specify Grid Usage**

You can use the EGGridPolicy and AMOGGridPolicy attributes on the logical grid server definition to control whether SAS Enterprise Guide or SAS Add-In for Microsoft Office version 5.1 or later uses the grid. These attributes override any other settings in the applications. The following are the possible values for the attributes (values are not case-sensitive):

- **Noforce**
  - jobs are sent to the grid depending on the setting of the option **Use grid if available**.

- **Force**
  - jobs are always sent to the grid, regardless of the value of the option **Use grid if available**.

- **Ignore**
  - jobs are never sent to the grid, regardless of the value of the option **Use grid if available**. This value is supported for the first maintenance release after version 6.1 of SAS Enterprise Guide and SAS Add-In for Microsoft Office.

**Assigning Libraries in a Grid**

In SAS 9.2 and later versions, SAS sessions on the grid use the METAAUTORESOURCES option by default. This option causes SAS libraries that are defined in metadata and identified as “pre-assigned” to automatically be assigned when the SAS session is started. Using pre-assigned libraries with the METAAUTORESOURCES option ensures that the libraries used in the code generated by SAS Enterprise Guide and SAS Add-In for Microsoft Office are available to the SAS sessions on the grid.

However, if your programs use a large number of libraries, you might not want to make all of these libraries pre-assigned. Automatically assigning a large number of libraries could cause performance problems, and not all libraries are likely to be used for all programs. To minimize the performance overhead, define the libraries in SAS metadata but do not identify them as pre-assigned. When you need to refer to the library, you can then use a LIBNAME statement using the META LIBNAME engine.

**Developing SAS Programs Interactively Using a Grid**

**Maintaining a Connection to the Grid**

By default, when you start SAS Enterprise Guide or SAS Add-In for Microsoft Office, it connects to a single workspace server and keeps that connection active for the length of
the session. If you interactively develop programs in SAS Enterprise Guide or SAS Add-In for Microsoft Office by highlighting and submitting lines of code, the codes use items such as libraries, WORK files, and SAS global statements on the workspace server.

If you are submitting grid-enabled code from SAS Enterprise Guide or SAS Add-In for Microsoft Office in a grid environment, then items such as libraries and SAS global statements must be accessed through the grid rather than through a single workspace server. To maintain access to these items, you must maintain a connection to the grid while you are developing programs interactively. If you are using grid-launched workspace servers, this is not an issue.

**Managing Interactive Workload**

When SAS Enterprise Guide or SAS Add-In for Microsoft Office is used for interactive program development, the workload is likely to consist of short bursts of work interspersed with varying periods of inactivity while the user considers their next action. The SAS grid configuration can best support this scenario with these configuration settings:

- Increase the number of job slots for each machine.
  
  Increasing the number of job slots increases the number of simultaneous SAS sessions on each grid node. Because the jobs that are run on the grid are not I/O or compute intensive like large batch jobs, more jobs can be run on each machine.

- Implement CPU utilization thresholds for each machine.
  
  If all users submit CPU-intensive work at the same time, SAS Grid Manager can suspend some jobs and resume the suspended jobs when resources are available. This capability prevents resources from being overloaded.

The following example shows a sample LSB.HOSTS file that is configured with job slots set to 32 and CPU utilization thresholds set to 80%. The settings needed for a specific site depend on the number of users and the size of the grid nodes.

<table>
<thead>
<tr>
<th>HOST_NAME</th>
<th>MXJ</th>
<th>ut</th>
<th>rlm</th>
<th>pg</th>
<th>ls</th>
<th>tmp</th>
<th>DISPATCH_WINDOW</th>
<th>#Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>#default</td>
<td>1</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>#Example</td>
</tr>
<tr>
<td>host01</td>
<td>32</td>
<td>0.7/0/8</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td># host01</td>
</tr>
<tr>
<td>host02</td>
<td>32</td>
<td>0.7/0/8</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td># host02</td>
</tr>
<tr>
<td>host03</td>
<td>32</td>
<td>0.7/0/8</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td># host03</td>
</tr>
<tr>
<td>host04</td>
<td>32</td>
<td>0.7/0/8</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td># host04</td>
</tr>
<tr>
<td>host05</td>
<td>32</td>
<td>0.7/0/8</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td>()</td>
<td># host05</td>
</tr>
</tbody>
</table>

**Using SAS Stored Processes with a SAS Grid**

As with SAS Enterprise Guide, SAS stored processes can take advantage of a SAS grid using either one of these approaches or a combination of both:

- Using server-side load balancing through a stored process server. You can convert a stored process server to use load balancing, and then specify grid as the balancing algorithm. When you send jobs from a SAS stored process to this server, the server automatically sends the job to the least busy node in the grid. See “Using SAS Grid Manager for Server Load Balancing” on page 125 for more information about specifying grid load balancing on a stored process server.
Modifying the stored process code in order to enable the code to run on the grid. You must also modify options associated with the stored process. Download the code that you must add to the stored process, along with complete information about other steps that you must take, from http://support.sas.com/rnd/scalability/grid/download.html.

Using SAS Data Integration Studio with a SAS Grid

Scheduling SAS Data Integration Studio Jobs on a Grid

You can schedule jobs from within SAS Data Integration Studio and have those jobs run on the grid. You deploy the job for scheduling in SAS Data Integration Studio, and then use the Schedule Manager plug-in in SAS Management Console to specify the schedule and the scheduling server. For more information, see “Scheduling Jobs on a Grid” on page 113. Also see Scheduling in SAS.

Multi-User Workload Balancing with SAS Data Integration Studio

SAS Data Integration Studio enables users to directly submit jobs to a grid or to a workspace server that is launched by the grid. This capability allows the submitted jobs to take advantage of load balancing and job prioritization that you have specified in your grid. SAS Data Integration Studio also enables you to specify the workload that submitted jobs should use. This capability enables users to submit jobs to the correct grid partition for their work.

To submit a job to the grid, select the SAS Grid Server component in the Server menu on the Job Editor toolbar. Click Submit in the toolbar to submit the job to the grid.

This step is not needed if you use grid-launched workspace servers.

Figure 9.1 Submitting a Job to the Grid

To specify a workload value for the server, follow these steps:

1. On the SAS Data Integration Studio menu bar, select Tools ➪ Options, and then select the SAS Server tab on the Options dialog box.

2. Select the SAS grid server in the Server field.
3. Select the workload to use for the submitted jobs in the **Workload specification** field.

**Figure 9.2 Selecting the Workload**

4. You can also specify the workload in a grid options set and specify the options set name in the **Grid options set specification** field. Only grid options sets that are mapped to SAS Data Integration Studio are listed in this field. For more information, see “Modifying SAS Logical Grid Server Definitions” on page 99.

SAS Grid Manager uses the workload value to send the submitted job to the appropriate grid partition. For more information about the other steps required, see “Defining and Specifying Resources” on page 38.

**Parallel Workload Balancing with SAS Data Integration Studio**

A common workflow in applications created by SAS Data Integration Studio is to repeatedly execute the same analysis against different subsets of data. Rather than running the process against each table in sequence, use a SAS grid environment to run the same process in parallel against each source table, with the processes distributed among the grid nodes. For this workflow, the Loop and Loop-End transformation nodes can be used in SAS Data Integration Studio to automatically generate a SAS application that spawns each iteration of the loop to a SAS grid via SAS Grid Manager.
To specify options for loop processing, open the Loop Properties window and select the **Loop Options** tab. You can specify the workload for the job, and you can specify how many processes can be active at once.

For more information, see *SAS Data Integration Studio: User's Guide*. 
Specifying Workload for the Loop Transformation

A SAS Data Integration Studio user performs these steps to specify an LSF resource in the properties for a Loop Transformation in a SAS Data Integration Studio job. When the job is submitted for execution, it is submitted to one or more grid nodes that are associated with the resource.

It is assumed that the default SAS Application Server for SAS Data Integration Studio has a Logical SAS Grid Server component, which was updated in the metadata repository. For more information, see “Defining and Specifying Resources” on page 38.

1. In SAS Data Integration Studio, open the job that contains the Loop Transformation to be updated.
2. In the Process Designer window, right-click the metadata object for the Loop Transformation and select Properties.
3. In the Properties window, click the Loop Options tab.
4. On the Loop Options tab, in the Grid options set specification text box, select the name of the desired grid options set. The entry is case sensitive.
5. Click OK to save your changes, and close the Properties window.

Using SAS Enterprise Miner with a SAS Grid

There are three cases where SAS Enterprise Miner uses a SAS grid:

• during model training, for parallel execution of nodes within a model training flow
• during model training, for load balancing of multiple flows from multiple data modelers
• during model scoring, for parallel batch scoring

The workflow for SAS Enterprise Miner during the model training phase consists of executing a series of different models against a common set of data. Model training is CPU-intensive and I/O-intensive. The process flow diagram design of SAS Enterprise Miner lends itself to processing on a SAS grid because each model is independent of the other models. SAS Enterprise Miner generates the SAS program to execute the user-created flow and also automatically inserts the syntax needed to run each model on the grid. Because the models can execute in parallel on the grid, the entire process is accelerated.

In addition, SAS Enterprise Miner is typically used by multiple users who are simultaneously performing model training. Using a SAS grid can provide multi-user load balancing of the flows that are submitted by these users, regardless of whether the flows contain parallel subtasks.

The output from training a model is usually Base SAS code that is known as scoring code. The scoring code is a model, and there are usually many models that need to be scored. You can use SAS Grid Manager to score these models in parallel. This action accelerates the scoring process. You can use any of these methods to perform parallel scoring:

• Use the SAS Grid Manager Client Utility to submit each model independently to the SAS grid.
- Use the Schedule Manager plug-in to create a flow that contains multiple models and schedule the flow to the SAS grid. Because each model is independent, the models are distributed across the grid when the flow runs.

- Use SAS Data Integration Studio to create a flow to loop multiple models, which spawns each model to the grid.

Starting with SAS Enterprise Miner 6.2, you can specify that projects are sent to the grid by default.

In SAS Enterprise Miner 12.3, specify default grid processing by changing the attributes on the logical workspace server. Follow these steps:

1. In SAS Management Console, expand the Server Manager plug-in and then expand the SAS Application Server (typically named SASApp).
2. Select the Logical Workspace Server and select `File ➔ Properties`.
3. On the `Extended Attributes` tab, click `New`.
4. In the `Field Name` column, enter `EM_USE_GRID_IF_AVAILABLE`.
5. In the `Value` column, enter `Y`.
6. Click `OK` to save the property changes.

In SAS Enterprise Miner 6.2 to 12.1, specify default grid processing by changing a configuration property. To set this property, locate the file `app.config`. A typical location for this file is `C:\SAS\Config\AnalyticsPlatform\apps\EnterpriseMiner\app.config`. Locate the property `em.enablegrid`. To turn on default grid processing, specify a value of `Y`. If you specify a value of `N`, you must send projects individually to the grid.

*Figure 9.5  Grid Processing with SAS Enterprise Miner*
Using SAS Risk Dimensions with a SAS Grid

The iterative workflow in SAS Risk Dimensions is similar to that in SAS Data Integration Studio. Both execute the same analysis over different subsets of data. In SAS Risk Dimensions, the data is subsetted based on market states or by instruments. Each iteration of the analysis can be submitted to the grid using SAS Grid Manager to provide load balancing and efficient resource allocation.

Because every implementation is different, an implementation of SAS Risk Dimensions in a grid environment must be customized to your specific business and data requirements.

Using SAS Grid Manager for Server Load Balancing

Overview

SAS servers, such as SAS workspace servers and SAS OLAP servers, are capable of performing load balancing across multiple machines. These servers can be configured to use one of the default algorithms to provide load balancing. However, with SAS Grid Manager installed you can configure workspace servers and other SAS servers to use SAS Grid Manager to provide load balancing. Because SAS Grid Manager accounts for all of the resource consumption on the machine, SAS Grid Manager can make better decisions about which machine is the best candidate for a server session.

SAS Grid Manager can provide load-balancing capabilities for these types of servers:

- workspace servers
- pooled workspace servers
- stored process servers
- OLAP servers

Any SAS product or solution that uses workspace servers, pooled workspace servers, stored process servers, or OLAP servers would benefit from using SAS Grid Manager to provide load balancing. This includes (but is not limited to) SAS Enterprise Guide, SAS Studio Enterprise Edition, SAS Data Integration Studio, SAS Enterprise Miner, SAS Web Report Studio, and SAS Marketing Automation. However, using the grid to provide load balancing also increases overhead, so your jobs might take longer to run.

You can use two methods to provide workspace server load balancing through SAS Grid Manager. You can have SAS Grid Manager start the workspace servers, or you can have the object spawner start the servers.

Load Balancing Methods

Grid-Launched Servers

With grid-launched workspace servers, the spawner passes the request on to SAS Grid Manager when an application requests a SAS workspace server from the object spawner.
SAS Grid Manager submits the job to the grid provider, which determines the best server to run the job based on the policy of the queue. The grid provider uses factors such as CPU utilization, network and disk input and output, disk space, available memory, and queue limits in determining which machine to use.

Because the grid-launched workspace servers are started by SAS Grid Manager, jobs on these servers appear as LSF jobs.

In the first maintenance release after SAS 9.4, pooled workspace servers and stored process servers can also be grid-launched.

An error will result if you use a grid-launched workspace server that specifies more hosts than the queue definition that you are using. For example, using a queue definition that specifies only host2 together with a logical workspace server definition that specifies host1, host2, and host3 causes an error. To accommodate this type of environment, add the line `ENABLE_HOST_INTERSECTION=Y` to the lsb.params file.

**Servers Started by the Object Spawner**

When an application or process (for example, SAS Enterprise Guide) requests a SAS server from the object spawner, the spawner determines which hosts can run the new servers. The spawner sends the list of hosts to SAS Grid Manager, which then determines which host is open and is the least busy. The object spawner then directs the client connection to a server on the least busy node on the grid. If there are multiple servers already running on the machine, then the spawner directs the client to the server with the fewest number of connections.

When using the grid load balancing algorithm, jobs are not started by LSF. As a result, jobs that run on the load-balanced servers do not appear as LSF jobs.

For more information, see “Understanding Server Load Balancing” in *SAS Intelligence Platform: Application Server Administration Guide*.

**SAS OLAP Servers**

Because the SAS object spawner is not used to start SAS OLAP servers, the request is not processed through the object spawner if the client requests a SAS OLAP server. However, the same load-sharing logic is used to determine which OLAP server is used for the client connection. As a result, the SAS client connections are spread around to all of the servers in the grid.

**Converting Servers to Use Grid Load Balancing**

To use SAS Grid Manager for load balancing for a workspace server or an OLAP server, follow these steps:

1. In the Server Manager plug-in in SAS Management Console, select the server that will use load balancing.

2. Select **Convert To Load Balancing** from the **Actions** menu or the context menu for the logical server. The Load Balancing Options dialog box appears.
3. Specify the following values:
   
   **Balancing algorithm**
   Select **Grid**.

   **Launch Servers via Grid**
   Select this check box to specify that SAS Grid Manager will launch the new servers. Deselect this check box if you want the object spawner to start the servers. This option is not available for OLAP servers.

   **Grid server**
   Select a grid server that was defined during installation and configuration. This field is active only if you did not select **Launch Servers via Grid**.

   **Grid server credentials**
   Select the credentials that the object spawner uses to authenticate to the grid server. This field is active only if you did not select **Launch Servers via Grid**.

4. Click **OK** to save your changes to the server metadata.

   To use SAS Grid Manager for load balancing for a stored process server or a pooled workspace server, follow these steps:

1. In the Server Manager plug-in in SAS Management Console, select a logical server that will use load balancing.

2. Select **Properties** from the **Actions** menu or the context menu. The Properties dialog box appears.

3. Select the **Load Balancing** tab.
4. Specify the following values:

Balancing algorithm
Select Grid.

Launch Servers via Grid
Select this check box to specify that SAS Grid Manager will launch the new servers. Deselect this check box if you want the object spawner to start the servers.

Grid server
Select a grid server that was defined during installation and configuration. This field is active only if you did not select Launch Servers via Grid.

Grid server credentials
Select the credentials that the object spawner uses to authenticate to the grid server. This field is active only if you did not select Launch Servers via Grid.

In addition to modifying these server definitions, you must also change server configuration files. For information about the changes that you need to make, see “Understanding Server Load Balancing” in SAS Intelligence Platform: Application Server Administration Guide.
Chapter 10
Working With Grid Options Sets

Understanding Grid Options Sets

A grid options set is a collection of grid options, SAS options, and required resources. Grid options sets are then mapped to a particular SAS client application and one or more specific metadata identities. The purpose of a grid options set is to enable a SAS grid administrator to define a collection of options in SAS metadata that map to one or more SAS client applications. Those options and are automatically applied to the workload that is submitted to the grid based on the identity of the user accessing the client application. SAS client applications such as the SAS Grid Manager Client Utility have been enhanced to support grid options sets.

For example, consider a scenario where you have many SAS Data Integration Studio users. The users are classified as either “power users” or “ad hoc users.” The power users should always submit their jobs to the POWERUSER queue and use the SAS options MEMSIZE=2048M and SORTSIZE=512M. The ad hoc users should always submit their jobs to the DIS queue and use the default SAS invocation options.

To support this scenario, define the workspace servers to be launched using the grid. You could define a grid options set called DISPOWERUSER that contains the grid option QUEUE=POWERUSER and SAS options MEMSIZE=2048M and SORTSIZE=512M. You would then map this grid options set to the SAS Data Integration Studio application and to the metadata identities that you identify as power users. You could then define a second grid options set called DISADHOCUSER that contains only the grid option QUEUE=DIS. You would also map this grid options set to the SAS Data Integration Studio application, but to the SASUSERS group. When a SAS Data Integration Studio user submits a job to the grid, the grid options set mapping identifies which grid options set to use, and applies the proper options for the job. This process is completely transparent to the user.
Merging Option Values

When you use a grid options set, its values are merged with the default option values for the logical grid server. The way the values are merged and evaluated differs according to the type of option:

SAS Options
values in the grid options set are appended to the values from the SAS Options field on the logical grid server’s Properties window. Because the options are evaluated from left to right, the value from the grid options set is evaluated last and therefore overrides any similar default SAS option.

For example, you have a logical grid server with a SAS option set to “-memsize 256” and you create a grid options set with a SAS option value of “-memsize 0”. When the job is executed on the grid, the SAS command line has the options “-memsize 256 -memsize 0” at the end of the command line. The option “-memsize 0” is evaluated last, so that is the value used when processing the job.

Required resources
values in the grid options set are added to the values from the Required Resources field on the logical grid server’s Properties window. When the job runs on the grid, all specified resources must be available.

For example, you might have a logical grid server with “SASApp” specified in the Required Resources field and a grid options set with a required resource specified as “MySQLDB”. When the grid provider selects a host on which to run the job, it must select a host that has both “SASApp” and “MySQLDB” as resources associated with the host.

Grid options
values in the grid options set are appended to the values from the Grid Options field on the logical grid server’s Properties window. Because the options are evaluated from left to right, the value from the grid options set is evaluated last and therefore overrides any similar default grid option.

For example, if you are using Platform Suite for SAS, you might have a logical grid server with a Grid Options value of “queue=normal” and you create a grid options set with a value of “queue=priority”. When the job is passed to the grid provider module, the option string that is passed include the options “queue=normal queue=priority”. Because the option “queue=priority” appears after “queue=normal”, the priority queue is used for the job.

Creating Grid Options Sets

Grid options sets are created in SAS Management Console as part of a Grid Server definition. You must be a member of the SAS Administrators group in order to create a grid options set. To create a grid options set, follow these steps:

1. In the Server Manager of SAS Management Console, select a Grid Logical Server and then the Grid Server component.

2. In the Properties window, select the Options tab. The Grid Options Set Mappings table lists any existing mappings.
3. Click **New** to display the Grid Options Set Mapping Wizard dialog box.

The **New**, **Edit**, and **Delete** buttons are available only if you are a member of the SAS Administrators group.
4. In this dialog box, specify these options:

**SAS Application**
Specify the SAS application that should be associated with this grid options set. The applications that are listed in this field are ones that have been identified as grid capable either during deployment or by specifying the isGridCapable keyword in the application’s properties. If you select an application that launches a workspace server and you have specified that load balanced servers should be launched by the grid, rather than by the spawner, then the options in the grid options set can be used when the workspace server starts. The application name must also be matched with the isGridCapable keyword. See “Specifying That an Application Is Grid Capable” on page 133 for more information.

**Grid Options Set**
Specifies a named set of grid options, required resources, and SAS options. The grid options set mapping associates this group of options and resources with a SAS application and identities. Click **New** to define a new grid options set.

**Identities**
Select identities (users or groups) to associate with the grid options set. You must select at least one identity to associate with the mapping. If no identity is selected, the mapping is not saved.
Specifying That an Application Is Grid Capable

The only applications that you can associate with a grid options set are ones that have been identified as being grid capable. Applications are identified as grid capable either by the application’s deployment process or by specifying the isGridCapable keyword through the application’s property values.

To verify which applications have been identified as grid capable during deployment, you can define a grid options set mapping before you have set the isGridCapable keyword on any applications through the application properties. The applications that are listed as available for mapping are the ones that were identified as grid capable during deployment.

For all other applications that you want to map to a grid options set, you must manually identify each of them as grid capable. In order to apply the isGridCapable keyword to the correct application when running a job, SAS Grid Manager must be able to match the application name as identified on the application folder with the name that is sent by the application when the object spawner starts the workspace server. If they do not match, then you can use wildcards to help match the names. This situation might occur if the application sends a different name from one invocation to the next or if you have multiple versions of the same application that you want to map to the same grid options set.

These examples illustrate several scenarios:

• If the folder for an application is specified as “SAS Application A” and the application sends the name “SAS Application A”, then the names match exactly and you can set the keyword to isGridCapable.

• If the folder for an application is specified as “SAS Application A” and the application sends the name “SAS Application A 4.6” some times and “SAS Application A 4.7” at other times, then you should specify the keyword in the form isGridCapable:SAS Application A*.

• If the folder for an application is specified as “SAS Application A” and the application sends the name “SAS Application A” some times and “Web Infrastructure - SAS Application A” at other times, then you should specify the keyword in the form isGridCapable:*SAS Application A.

• If the folder for an application is specified as “SAS Application A” and the application sends the name “Web Infrastructure - SAS Application A” some times and “SAS Application A 4.6” at other times, then you should specify the keyword in the form isGridCapable:*SAS Application A*.

For example, to make SAS Studio a grid-enabled application, you might specify the keyword as isGridCapable:*SAS Studio Mid-Tier*.

• The isGridCapable portion of the keyword enables SAS Studio Mid-Tier to be included in the list of grid-enabled applications.

• The *SAS Studio Mid-Tier* portion of the keyword enables the IOM application name as sent by SAS Studio to be matched to the SAS Studio Mid-Tier application folder so that grid options sets can be applied. The wildcards enable SAS Grid Manager to recognize the IOM name (Web Infra Platform Services 9.4 - SAS Studio Mid-Tier 3.1) and match it with the SAS Studio Mid-Tier options sets.

To set the isGridCapable keyword for an application, follow these steps:
1. In SAS Management Console, select the **Folders** tab.
2. Open the **System ➔ Applications** folder.
3. Select the folder for the application whose property you want to change. Then, select **File ➔ Properties.**
4. In the Properties window, click **Add** next to the **Keyword** area.
5. In the Add Keyword dialog box, type `isGridCapable` or `isGridCapable: <application_name>` with wildcards as needed in the **Name** field.

6. Click **OK** to close the Add Keyword dialog box and then the Properties window.

After applications are identified as being grid capable and grid options sets are mapped to those applications, the options in the grid options set are used to start any grid-launched servers. See “Grid-Launched Servers” on page 125 for more information.
Part 5

SAS Grid Language Reference

Chapter 11
SAS Functions for SAS Grid ........................................ 137

Chapter 12
SASGSUB Command ................................................. 153
Chapter 11
SAS Functions for SAS Grid

Dictionary

GRDSVC_ENABLE Function
Enables or disables one or all SAS sessions on a grid.

Valid in: %SYSFUNC or %QSYSFUNC Macro, DATA step
Category: Grid

Syntax

grdsvc_enable(identifier <,option-1; ... option-n> )
grdsvc_enable(identifier,"" | "")

Required Argument

identifier
specifies one or all server sessions to be enabled or disabled for grid execution. The identifier is specified as follows:

server-ID
specifies the name of a SAS/CONNECT server session to be enabled or disabled for grid execution.

You use this server-ID when you sign on to a server session using the SIGNON or the RSUBMIT statement. For information about ways to specify the server ID, see SAS/CONNECT User's Guide.
Requirement
If the function is used in a DATA step, enclose server-ID in double or single quotation marks. A server-ID cannot exceed eight characters.

_ALL_
specifies that all SAS sessions are enabled or disabled for grid execution.

See
SIGNON statement and RSUBMIT statement in *SAS/CONNECT User's Guide*

Example
```sas
%let rc=%sysfunc(grdsvc_enable(grdnode1,server=SASApp));
%let rc=%sysfunc(grdsvc_enable(_all_,server=SASApp));
%let rc=%sysfunc(grdsvc_enable(notgrid1,**));
```

### Optional Arguments

**SASAPPSERVER=server-value**
specifies the name of a SAS Application Server that has been defined in the SAS Metadata Repository. The SAS Application Server contains the definition for the logical grid server that defines the grid environment.

**Alias**
SERVER=, RESOURCE=

**Restriction**
Although a SAS Application Server is configured as a required grid resource in most environments, some grids are not partitioned by resource names. In these environments, passing the SAS Application Server name as a required resource causes the job to fail. To find out whether the SAS Application Server is designated as a required resource value or not in the SAS Metadata Repository, use the GRDSVC_GETINFO function call.

**Interaction**
The name of the SAS Application Server is passed to Platform Suite for SAS as a resource value. When the job is executed, the grid provider selects a grid node that meets the requirements that are specified by this value. If SAS-application-server contains one or more spaces, the spaces are converted to underscores before the name is passed to Platform Suite for SAS as a resource value.

**Tip**
For Platform Suite for SAS, this server-value corresponds with the value of a resource that the LSF administrator has configured in the lsf.cluster.cluster-name file and the lsf.shared file on the grid-control server.

**See**
“GRDSVC_GETINFO Function” on page 143 to find out whether the SAS Application Server is designated as a required resource value in the SAS Metadata Repository. To remove the SAS Application Server name as a required resource, see “Modifying SAS Logical Grid Server Definitions” on page 99.

**Example**
```sas
%let rc=%sysfunc(grdsvc_enable(_all_, server=SASApp));
```

**GRIDOPTSET="grid_options_set_name"**
specifies the grid options set to be used when running jobs on the grid. A grid options set is a collection of grid options, SAS options, and required resources that are associated with a particular SAS client application. The purpose of a grid options set is to enable and disable grid options for the user or to change the default options that are used when running jobs on the grid.

**Alias**
GRIDOPTSET=

**Restriction**
GRIDOPTSET must be used in a DATA step.

**Interaction**
The grid options set is passed to Platform Suite for SAS as a resource value. When the job is executed, the grid provider selects a grid node that meets the requirements that are specified by this value. If grid_options_set contains one or more spaces, the spaces are converted to underscores before the name is passed to Platform Suite for SAS as a resource value.

**Tip**
For Platform Suite for SAS, this grid_options_set corresponds with the value of a resource that the LSF administrator has configured in the lsf.cluster.cluster-name file and the lsf.shared file on the grid-control server.

**See**
“GRDSVC_GETINFO Function” on page 143 to find out whether the SAS Application Server is designated as a required resource value in the SAS Metadata Repository. To remove the SAS Application Server name as a required resource, see “Modifying SAS Logical Grid Server Definitions” on page 99.

**Example**
```sas
%let rc=%sysfunc(grdsvc_enable(_all_, server=SASApp));
```

**Note**
If the function is used in a DATA step, enclose server-ID in double or single quotation marks. A server-ID cannot exceed eight characters.
set is to enable a SAS grid administrator to define a collection of options in SAS metadata. These options will map to one or more SAS client applications and will be applied to workload submitted to the grid.

**Requirements**

A grid options set must be defined in SAS Management Console before it can be used in the **GRDSVC_ENABLE** function.

| The grid options set name must be enclosed in quotation marks. |

**WORKLOAD=workload-value**

Identifies the resource for the job to be executed on the grid. This value specifies an additional resource requirement for which Platform Suite for SAS selects the appropriate grid nodes.

The specified workload value should match one of the workload values that are defined in the SAS Application Server in the SAS Metadata Repository.

**Requirement**

Workload values are case sensitive.

**Interaction**

If *workload-value* contains one or more spaces, the spaces are converted to underscores before the value is passed to the grid provider. If *workload-value* is not located in the SAS Application Server definition and no other errors occur, a 0 result code is returned, and this note is displayed:

```
NOTE: Workload value "gridResource" does not exist in the SAS Metadata Repository.
```

**Tip**

For Platform Suite for SAS, this *workload-value* corresponds with the resource that the LSF administrator has configured in the lsf.cluster.*cluster-name* file and the lsf.shared file on the grid-control computer.

**Example**

```sas
%let rc=%sysfunc(grdsvc_enable(grdnode1, server=SASApp; workload=EM));
```

The workload value EM specifies the resource name. EM must be assigned to a grid node in order to process this job. An example is assigning EM to machines that can process SAS Enterprise Miner jobs.

**JOBNAME=job-name-macro-variable**

Specifies the macro variable that contains the name that is assigned to the job that is executed on the grid.

**Example**

```sas
%let hrjob=MyJobName;
%let rc=%sysfunc(grdsvc_enable(grdnode1, server=SASApp; jobname=hrjob));
signon grdnode1;
```

In this example, hrjob is the name of the macro variable to which the job name is assigned. The actual job name is MyJobName. The status of the job can be tracked using the SAS Grid Manager Plug-in for SAS Management Console. In this example, you track the status of the job named MyJobName.
**JOBOPTS=**_job-opts-macro-variable_

specifies the macro variable that contains the job options. The job option name/value pairs are assigned to _job-opts-macro-variable_.

The job options are used by the grid job to control when and where a job runs. Job options are specified as name/value pairs in this format:

```
option-1=value-1;option-2="value-2 with spaces"; ... 
option-n='value-n with spaces';
```

For a list of the job options that you can specify, see Appendix 1, “Supported Job Options - Platform Suite for SAS,” on page 169.

**Requirement**  
Use a semicolon to separate job option and value pairs. For multiple values, use a macro quoting function for the semicolon or use single or double quotation marks to enclose all job options. If the value contains one or more spaces, tabs, semicolons, or quotation marks, enclose the value in single or double quotation marks.

**Example**

```
%let rc=%sysfunc(grdsvc_enable(all, server=SASApp; jobopts=hrqueue)); 
%let hrqueue=queue=priority%str(;)project="HR Monthly"; 
signon grdnode1; 
%let hrqueue='queue=priority;project="HR Yearly"'; 
signon grdnode2 
```

Both jobs are sent to the priority queue. The first job is associated with the project named “HR Monthly” and the second job is associated with the project named “HR Yearly.”

**Disables**

```
""
```

disables grid execution for the specified server ID or all server sessions.

This value is intended to be used when you have specified _ALL_ in a previous call but you want to disable it for a small number of exceptions.

**Requirement**  
Double or single quotation marks can be used. Do not insert a space between the double or single quotation marks.

**Interaction**  
When quotation marks are used with _ALL_, it clears all previous grid settings that were specified using the GRDSVC_ENABLE function.

**Example**

```
%let rc=%sysfunc(grdsvc_enable(grdnode1,"")); 
%let rc=%sysfunc(grdsvc_enable(_all_,""));
```

**Details**

The GRDSVC_ENABLE function is used to enable and disable a grid execution. Grid execution can be enabled for a specified SAS session or for all SAS grid sessions. If a grid environment is not configured or is unavailable, the job is started as a symmetric multi-processor (SMP) process instead.

The GRDSVC_ENABLE function does not resolve to a specific grid node, and it does not cause grid execution. The server ID is mapped to a specific grid node. The server session starts on the grid node when requested by subsequent SAS statements (for example, when the SIGNON statement or the RSUBMIT statement is executed).
In order to restrict the use of specific grid nodes to be used by server sessions, the name of the SAS Application Server and the workload resource value are passed as required resources to Platform Suite for SAS.

*Note:* An exception to this behavior is when the SAS Application Server is disabled as a required resource for the grid server. For details, see the restriction for the SASAPPSERVER= option.

The grid can be partitioned according to resource or security requirements. If grid nodes do not have the required resources, then SAS requests fail. If grid nodes have the required resources but are busy, SAS requests are queued until grid resources become available. For information, see “Defining and Specifying Resources” on page 38.

Some SAS applications are appropriate for execution in a grid environment, but not in an SMP environment. Such applications should contain a macro that checks the return code from the GRDSVC_ENABLE function to ensure that a grid node, rather than an SMP process, is used.

Here are the result codes:

<table>
<thead>
<tr>
<th>Result Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Reports that one or all server sessions were disabled from grid execution.</td>
</tr>
</tbody>
</table>
| 1           | Reports that a grid environment is unavailable due to one or more of these conditions:  
• A connection to the SAS Metadata Server is unavailable.  
• A logical grid server has not been defined in the SAS Metadata Repository.  
• The current user identity does not have authorization to use the specified logical grid server.  
• SAS Grid Manager has not been licensed.  
  Instead, server sessions execute on the multi-processor (SMP) computers as a SASCMD sign-on. One of these commands, in order of precedence, is used to start the server session:  
• the value of the SASCMD system option  
• !sascmd -noobjectserver |
| 0           | Reports that the specified session was enabled. |
| -1          | Reports a syntax error in the function call. An example is the omission of the server ID. |
| -2          | Reports a parsing error in the function call. An example is an invalid option. |
| -3          | Reports an invalid server ID in the function call. |
| -5          | Reports an out-of-memory condition while the function is executing. |
### Result Code Explanation

-6  Reports that the function cannot connect to the SAS Metadata Server or cannot access the grid metadata information. This condition frequently occurs when the user is not explicitly defined in metadata. By default, users without a definition in metadata are assigned to the PUBLIC group, which is not granted the ReadMetadata permission.

### See Also

- *SAS/CONNECT User's Guide*
- *SAS/CONNECT User's Guide*
- *SAS Language Reference: Dictionary*
- *SAS Macro Language: Reference*

### GRDSVC_GETADDR Function

Reports the IP address of the grid node on which the SAS session was chosen to execute.

**Valid in:**  
%SYSFUNC or %QSYSFUNC Macro, DATA step  

**Category:**  
Grid

### Syntax

```
grdsvc_getaddr(identifier)
```

### Without Arguments

### Required Argument

**identifier**  
identifies the server session that is executing on the grid. The identifier can be specified as follows:

```
""| "
```

is an empty string that is used to refer to the computer on which the function is executing.

**server-ID**  
specifies the server session that is executing on a grid.

You use the same server-ID that was used to sign on to a server session using the RSUBMIT statement or the SIGNON statement. Each server ID is associated with a fully qualified domain name (FQDN). The name resolution system that is part of the TCP/IP protocol is responsible for associating the IP address with the FQDN. The output is one or more IP addresses that are associated with the server. IP addresses are represented in IPv4 and IPv6 format, as appropriate.

**Requirement**  
Double or single quotation marks can be used. Do not insert a space between the double or single quotation marks.
If the function is used in a DATA step, enclose server-ID in double or single quotation marks.

Example

```bash
/*-----------------------------------*/
/* The following sets the macro variable 'myip' to the IP address */
/* of the grid node associated with the server session 'task1' */
/*-----------------------------------*/
%let
myip=%sysfunc(grdsvc_getaddr(task1));
```

See Also

RSUBMIT statement

• SAS/CONNECT User's Guide

SIGNON statement

• SAS/CONNECT User's Guide

DATA step

• SAS Language Reference: Dictionary

%SYSFUNC or %QSYSFUNC

• SAS Macro Language: Reference

GRDSVC_GETINFO Function

Reports information about the grid environment.

Valid in: %SYSFUNC or %QSYSFUNC Macro, DATA step

Category: Grid

Syntax

```
grdsvc_getinfo(identifier)
```

Required Argument

*identifier*

specifies the server session or the SAS Application Server whose details you want to have reported to the SAS log.

The *identifier* is specified as follows:

*server-ID*

reports details about the specified server ID. The details that are returned by the GRDSRV_INFO function reflect the arguments that are specified in the
GRDSVC_ENABLE function. You can request details about a server-ID that you have used to create a server session or that you will use to create a server session on the grid.

Requirement A server-ID cannot exceed eight characters.

_ALL_ reports details about all server IDs to the SAS log. The details that are returned by the GRDSRV_INFO function reflect the arguments that are specified in the GRDSVC_ENABLE function.

SASAPPSERVER=SAS-application-server reports information about the specified SAS Application Server to the SAS log.

Alias SERVER=, RESOURCE=

_SHOWID_ lists each server session and its status: enabled for grid execution, enabled for SMP execution, or disabled.

Interaction If the GRDSVC_GETINFO function is used in a DATA step, enclose the identifier in single or double quotation marks. The identifier can be specified as server-ID, _ALL_, SASAPPSERVER=SAS-application-server, or _SHOWID_. If no grid processes were enabled using the GRDSVC_ENABLE function or if all grid processes were disabled using the GRDSVC_ENABLE function with _ALL_ option, this message is displayed:

NOTE: No remote session ID enabled/disabled for the grid service.

Tip You do not have to be signed on to a specific server session in order to get information about it.

Example This log message reports that the SAS Application Server is a required resource.

```sas
%put %sysfunc(grdsvc_getinfo(server=SASApp));
NOTE: SAS Application Server Name= SASAPP
    Grid Provider= Platform
    Grid Workload= gridwrk
    Grid SAS Command= gridsasgrid
    Grid Options= gridopts
    Grid Server Addr= d15003.na.sas.com
    Grid Server Port= 123
    Grid Module= gridmod
Server name is a required grid resource value.
```

If the SAS Application Server is a disabled required resource, this message is displayed:

Server name is not a required grid resource value.

Optional Arguments

CLIENTAPP="client_application"

returns information about the grid options set associated with the specified application. The information returned includes the grid options, SAS options, and required resources that are specified in the grid options set and are associated with the specified application.
GRIDOPTSET="grid_options_set_name"
returns information about the specified grid options set. A grid options set is a
collection of grid options, SAS options, and required resources that are associated
with a particular SAS client application. The purpose of a grid options set is to
to enable a SAS grid administrator to define a collection of options in SAS metadata.
These options will map to one or more SAS client applications and will be applied to
workload submitted to the grid.

**Requirement**
The grid options set name must be enclosed in quotation marks.

**Example**
This example returns information about the grid options set “OptionsSet1.”

```sas
%put %sysfunc(grdsvc_getinfo(server=CNTWin;gridoptset="OptionsSet1"));
NOTE: SAS Application Server Name= CNTWin
      Grid Provider= Platform
      Grid Workload= idbvm1, idbvm2, idbvm3
      Grid SAS Command= gridsas
      Grid Server Addr= rlsgrid.mycorp.com
      Grid Required Resources= <none>
      Grid SAS options= -set GAOvar OptionsSet1
```

### Details
Here are the result codes:

<table>
<thead>
<tr>
<th>Result Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Reports that the specified server ID is not enabled for grid execution.</td>
</tr>
<tr>
<td>1</td>
<td>Reports that the specified server ID is enabled for SMP execution.</td>
</tr>
<tr>
<td>0</td>
<td>Reports that the specified server ID is enabled for a grid execution or that no error occurred.</td>
</tr>
<tr>
<td>-1</td>
<td>Reports a syntax error in the function call. An example is that an empty string is specified for the server ID.</td>
</tr>
<tr>
<td>-2</td>
<td>Reports a parsing error in the function call. An example is the failure to specify the SAS Application Server using the SASAPPSERVER= option.</td>
</tr>
<tr>
<td>-3</td>
<td>Reports an invalid server ID in the function call.</td>
</tr>
<tr>
<td>-5</td>
<td>Reports an out-of-memory condition while the function is executing.</td>
</tr>
<tr>
<td>-6</td>
<td>Reports that an error occurred when the SAS Metadata Server was accessed or when the information was returned from the SAS Metadata Server.</td>
</tr>
</tbody>
</table>

### Example

```sas
/*------------------------------------------------------------------------*/
/* Show grid logical server definition for SAS Application Server 'SASApp'*/
```
See Also

RSUBMIT statement
- SAS/CONNECT User's Guide

SIGNON statement
- SAS/CONNECT User's Guide

DATA step
- SAS Language Reference: Dictionary

%SYSFUNC or %QSYSFUNC
- SAS Macro Language: Reference

GRDSVC_GETNAME Function
Reports the name of the grid node on which the SAS grid server session was chosen to execute.

Valid in: %SYSFUNC or %QSYSFUNC Macro, DATA step
Category: Grid

Syntax
grdsvc_getname(identifier)

Required Argument
identifier
identifies the server session that is executing on the grid. The identifier can be specified as follows:
is an empty string that is used to refer to the computer at which the statement is executed.

`server-ID` specifies the server session that is executing on a grid.

You use the same `server-ID` that you used to sign on to a server session using the RSUBMIT statement or the SIGNON statement.

If the function is used in a DATA step, enclose `server-ID` in double or single quotation marks.

**Example**

```/* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * */ */
/* The following sets the macro variable 'mynodea' to the name of */
/* the grid node associated with the server ID 'task1'. */
/* * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * * */ */
%let
mynodea=%sysfunc(grdsvc_getname(task1));
```

**See Also**

- **RSUBMIT statement**
  - *SAS/CONNECT User's Guide*
- **SIGNON statement**
  - *SAS/CONNECT User's Guide*
- **DATA step**
  - *SAS Language Reference: Dictionary*
- **%SYSFUNC or %QSYSFUNC**
  - *SAS Macro Language: Reference*

---

**GRDSVC_HOSTLIST Function**

Returns a list of hosts that are available to run grid jobs. The list is ordered according to the priority set by the grid provider.

- **Valid in:** %SYSFUNC or %QSYSFUNC Macro, DATA step
- **Category:** Grid
Syntax

`grdsvc_hostlist('SERVER=SAS-application-server; HOSTLIST=grid_host_machine,[grid_host_machine] | WORKLOAD=workload-value || CLIENTAPP=client_application | GRIDOPTSET=grid_options_set_name;')`

Required Argument

**SASAPPSERVER=SAS-application-server**

specifies the name of the SAS Application Server that has been defined in the SAS Metadata Repository. The SAS Application Server contains the definition for the logical grid server that is used to access the grid environment. The name of the SAS Application Server is passed to the grid provider as a required resource.

Optional Arguments

**HOSTLIST=grid_host_machine,[grid_host_machine]**

specifies a comma-separated list of host machines that are in the grid and might be available to process grid jobs. If the list includes machines that are not included in the grid definition or are not available to process jobs, they are not included in the output of the function. If none of the machines in the list are included in the grid, the function returns an error message.

**WORKLOAD=workload-value**

identifies the resource for the job to be executed on the grid. This value specifies an additional resource requirement for which Platform Suite for SAS selects the appropriate grid nodes.

The specified workload value should match one of the workload values that are defined in the SAS Application Server in the SAS Metadata Repository.

Requirement

Workload values are case sensitive.

Interaction

If `workload-value` contains one or more spaces, the spaces are converted to underscores before the value is passed to the grid provider. If `workload-value` is not located in the SAS Application Server definition and no other errors occur, a 0 result code is returned, and this note is displayed:

```
NOTE: Workload value "gridResource" does not exist in the SAS Metadata Repository.
```

Tip

For Platform Suite for SAS, this `workload-value` corresponds with the resource that the LSF administrator has configured in the `lsf.cluster.cluster-name` file and the `lsf.shared` file on the grid-control computer.

Example

```sas
%let rc=%sysfunc(grdsvc_enable(grdnode1, server=SASApp; workload=EM));
```

The workload value EM specifies the resource name. EM must be assigned to a grid node in order to process this job. An example is assigning EM to machines that can process SAS Enterprise Miner jobs.
CLIENTAPP="client_application"
returns information about the hosts that are available to run jobs from the specified application.

GRIDOPTSET="grid_options_set_name"
returns information about the hosts associated with the specified grid options set. A grid options set is a collection of grid options, SAS options, and required resources that are associated with a particular SAS client application. The purpose of a grid options set is to enable a SAS grid administrator to define a collection of options in SAS metadata. These options will map to one or more SAS client applications and will be applied to workload submitted to the grid.

**Requirement**
The grid options set name must be enclosed in quotation marks.

**Example**
The code
```
%put
%sysfunc(grdsvc_hostlist('server=SASApp;hostlist=node1,node2,node3;'));
```
returns the output
```
node1, node3
```
This output indicates that node1 and node3 machines are available to process grid jobs, and that node1 is first priority to process jobs.

---

**GRDSVC_NNODES Function**

Reports the total number of job slots that are available for use on a grid.

**Valid in:** %SYSFUNC or %QSYSFUNC Macro, DATA step

**Category:** Grid

**Syntax**
```
grdsvc_nnodes(argument;option)
```

**Required Argument**

**SASAPPSERVER=SAS-application-server**
specifies the name of the SAS Application Server that has been defined in the SAS Metadata Repository. The SAS Application Server contains the definition for the logical grid server that is used to access the grid environment. The name of the SAS Application Server is passed to Platform Suite for SAS as a required resource. The grid provider selects the grid nodes that meet the requirements for the specified SAS Application Server and returns the total number of job slots in the grid.

An exception to this behavior is when the SAS Application Server is disabled as a required resource for the grid server. For details, see the SASAPPSERVER= option for the GRDSVC_ENABLE function on page 137.

**Alias**
SERVER=, RESOURCE=
Interaction
If `SAS-application-server` contains one or more spaces, the spaces are converted to underscores before the name is passed to Platform Suite for SAS.

Example
```%let numofnodes=%sysfunc(grdsvc_nnodes(server=SASApp));```

**Optional Argument**

`WORKLOAD=workload-value`

identifies the resource for the type of job to be executed on the grid. This value specifies the workload requirements for which Platform Suite for SAS selects the grid nodes that contain these resources.

The specified workload value should match one of the workload values that are defined in the SAS Application Server in the SAS Metadata Repository.

Requirement
If you specify `WORKLOAD=`, you must also specify the `SASAPPSERVER=` option. Workload values are case sensitive.

Interaction
If `workload-value` contains one or more spaces, the spaces are converted to underscores before the value is passed to Platform Suite for SAS. If `workload-value` is not located in the SAS Application Server definition and no other errors occur, a 0 result code is returned. A 0 result code means that no grid nodes contain the requested resources. Also, this note is displayed:

```
NOTE: Workload value "gridResource" does not exist in the SAS Metadata Repository.
```

If `workload-value` is undefined to Platform Suite for SAS, the `GRDSVC_NNODES` function returns the result code 0.

Tip
For Platform Suite for SAS, this `workload-value` corresponds with the resource that the LSF administrator has configured in the `lsf.cluster.cluster-name` file and the `lsf.shared` file on the grid-control computer.

Example
```%let numofnodes=%sysfunc(grdsvc_nnodes(server=SASApp; workload=em));```

The workload value, EM, specifies the resource name. EM must be assigned to a grid node in order to process this job. An example is assigning EM to machines that can process SAS Enterprise Miner jobs.

**Details**

When a grid environment is available, the `GRDSVC_NNODES` function returns the total number of job slots (busy and idle) that are available for job execution. This value is resolved when the function is called. Because of this, the value might vary over time, according to whether job slots have been added or removed from the grid. It can also vary based on the user, the queue that is being used, or other slot limits that are defined in the LSF configuration.

Note: If you are using SAS Grid Manager for Hadoop, the value returned by the `GRDSVC_NNODES` function might not match the number of jobs that can actually be executed at the same time. In a SAS Grid Manager for Hadoop environment, jobs contain resources for the app master as well as resources for the actual job. Because
the app master and the job will sometimes run on the same machine and will
sometimes run on different machines, an exact count of the number of executing jobs
cannot be determined.

Here are the result codes:

<table>
<thead>
<tr>
<th>Result Code</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>nnn</td>
<td>If a grid environment is available, reports the total number of job slots (idle and busy) that have been configured in a grid environment. The grid contains the resources that are specified by the SASAPPSERVER= argument and the WORKLOAD= option. If a grid environment is not available, assumes a multi-processor (SMP) environment, and reports the value of the CPUCOUNT system option. In this case, the lowest value that can be reported is 1.</td>
</tr>
<tr>
<td>1</td>
<td>If a grid environment is not available, assumes a multi-processor (SMP) environment, and reports the value of the CPUCOUNT system option. In this case, the lowest value that can be reported is 1.</td>
</tr>
<tr>
<td>0</td>
<td>Reports that no grid nodes contain the requested resources.</td>
</tr>
<tr>
<td>-1</td>
<td>Reports a syntax error in the function call. For example, a syntax error would result from supplying no value, or an empty string, to the SASAPPSERVER= option.</td>
</tr>
</tbody>
</table>

Example

```plaintext
/*---------------------------------------------*/
/* Get the number of grid nodes that have 'SASApp' as a resource */
/*---------------------------------------------*/
%let NumNodes=%sysfunc(grdsvc_nnodes(server=SASApp));
/*---------------------------------------------*/
/* Get the number of grid nodes that have 'SASApp' 'EM' as resources */
/*---------------------------------------------*/
%let numofnodes=%sysfunc(grdsvc_nnodes(server=SASApp;workload=EM));
```

See Also

RSUBMIT statement
- *SAS/CONNECT User's Guide*

SIGNON statement
- *SAS/CONNECT User's Guide*

DATA step
- *SAS Language Reference: Dictionary*
CPUCOUNT= system option

- SAS Language Reference: Dictionary

GRDSVC_OPTSETS Function

Returns a list of valid options sets for a specified combination of SAS Application Server, SAS client application, and user. If the user is not specified, the function returns a list of all valid grid options sets that apply to the specified SAS client application.

Valid in: %SYSFUNC or %QSYSFUNC Macro, DATA step

Category: Grid

Syntax

grdsvc_optsets("SERVER=SAS-application-server;CLIENTAPP='SAS-client-application'; USER='user-ID';")

Required Arguments

SERVER=SAS-application-server

specifies the name of the SAS Application Server that has been defined in the SAS Metadata Repository. The SAS Application Server contains the definition for the logical grid server that is used to access the grid environment. The name of the SAS Application Server is passed to the grid provider as a required resource.

CLIENTAPP='SAS-client-application'

specifies the name of a SAS client application for which a grid options set has been created.

Optional Argument

USER='user-ID'

specifies the ID of a user defined in metadata who might be subject to a grid options set for the specified SAS client application. The grid options set can apply to the user explicitly or as a member of a user group.

Example

The code

```sas
%put
%sysfunc(grdsvc_optsets("server=GOSWin;clientapp='SAS Grid Manager Client Utility';user='sasuser1';"));
```

returns the output

"Sasuser1_optset_SASSSUB","Group1_optset_SASGSUB"

which are the two grid options sets that apply to user sasuser1 for the SAS Grid Manager Client Utility. The options set Sasuser_optset_SASSSUB is explicitly defined for user sasuser1. The options set Group1_optset_SASGSUB applies because the user belongs to group Group1.
Chapter 12
SASGSUB Command

SASGSUB Overview

SAS Grid Manager Client Utility is a command-line utility that enables users to submit SAS programs, operating system commands, or command files to a grid for processing and to start interactive sessions on grid machines. This utility allows a grid client to submit SAS programs to a grid without requiring that SAS be installed on the machine performing the submission. It also enables jobs to be processed on the grid without requiring that the client remain active.

You can use the SAS Grid Manager Client Utility's SASGSUB command to submit jobs to the grid, start interactive sessions, view job status, retrieve results, and terminate jobs. The SAS Grid Manager Client Utility options can be specified in a configuration file so that they do not have to be entered manually. SASGSUB uses the sasgsub.cfg configuration file, which contains the required options. This file is automatically created by the SAS Deployment Wizard during installation. It stores the file in <config_dir>/Applications/SASGridManagerClientUtility/<version>.

If you are using Platform Suite for SAS as your grid provider, Platform LSF must be installed on any machine where the SAS Grid Manager Client Utility runs.
Disabling SASGSUB Actions

You might want to prevent users from using the SAS Grid Manager Client Utility to perform certain actions, such as terminating jobs. To disable these actions, you can specify these SasgsubDisable extended attributes on the grid server metadata definitions:

- SasgsubDisableGRIDSUBMITPGM
  prevents SAS jobs from being submitted
- SasgsubDisableGRIDKILLJOB
  prevents jobs from being terminated
- SasgsubDisableGRIDRUNCMD
  prevents arbitrary commands from running in batch
- SasgsubDisableGRIDRUNCMDINT
  prevents arbitrary commands from running in an interactive session
- SasgsubDisableGRIDRUNSASLM
  prevents an interactive SAS line mode session from starting
- SasgsubDisableGRIDRUNSASDMS
  prevents an interactive SAS Display Manager session from starting

You do not need to specify a value for these extended attributes.

Dictionary

SASGSUB Syntax: Submitting a SAS Program in Batch Mode

The following is the complete syntax for submitting a SAS program to a grid in batch mode. Enter the command on a Windows or UNIX command line.

Syntax

SASGSUB
   -GRIDSUBMITPGM sas-program-file
   common_arguments batch_arguments
   <GRIDRESTARTOK | GRIDRESTARTOK>
   <GRIDSASOPTS grid-sas-options>
   <grid_job_arguments>

Required Arguments

-GRIDSUBMITPGM sas-program-file
  specifies the path and filename of the SAS program that you want to run on the grid.

common_arguments
  specifies any of the arguments that are common to all uses of the SASGSUB command. Some arguments are required and some arguments are optional. See
“SASGSUB Syntax: Common Arguments” on page 161 for the syntax for each argument.

**Batch Arguments**

Specifies any of the arguments that are common to all SASGSUB commands that run in batch mode. Some arguments are required and some are optional. See “SASGSUB Syntax: Batch Options” on page 164 for the syntax for each argument.

**Optional Arguments**

- **GRIDRESTARTOK | -GRIDLRESTARTOK**
  Specifies that the job can be restarted at a checkpoint or a label.

- **GRIDSASOPTS grid-sas-options**
  Specifies any SAS options that are applied to the SAS session started on the grid. Multiple options can be enclosed in quotation marks. If you are specifying options that contain quotation marks, you can enclose all of the options in parentheses before enclosing them in quotation marks. Each of these forms is valid:

  - GRIDSASOPTS option
  - GRIDSASOPTS “option1 option2”
  - GRIDSASOPTS “(option1 option2 option3)”
  - GRIDSASOPTS “(option1 option2 ‘option3’)”

  When SASGSUB runs, the outer set of quotation marks is removed.

**Grid Job Arguments**

Specifies any of the arguments valid when submitting jobs or commands to the grid. See “SASGSUB Syntax: Grid Job Arguments” on page 166 for the syntax for each of these arguments.

---

**Example: Submitting a SAS Program to the Grid in Batch Mode**

The following is an example of a SASGSUB statement used to submit the SAS program Lab_report.sas in the directory C:\SAS_programs to the grid in batch mode. The job has been enabled for restarting, and it needs to run in the overnight queue. The grid uses a shared file system.

```
SASGSUB -GRIDSUBMITPGM C:\SAS_programs\Lab_report.sas -GRIDRESTARTOK
-GRIDJOBOPTS queue=overnight
```

---

**SASGSUB Syntax: Starting SAS in Interactive Mode**

The following is the complete syntax for starting SAS in interactive mode. You can use either line mode or SAS Display Manager mode to run SAS in a UNIX grid, with input coming from and output going to the client machine. SAS Display Manager mode also requires that an X server be present on the client machine. If you terminate SASGSUB in interactive mode, you also terminate the SAS session running in the grid.
Syntax

SASGSUB
  -GRIDRUNSASLM | -GRIDRUNSASDMS
    <hostname:displaynumber<.screennumber>
    common_arguments
    <GRIDWAITTIMEOUT seconds>
    <grid_job_arguments>

Required Arguments

-GRIDRUNSASLM
  specifies that the SASGSUB command runs a SAS session in line mode. Standard output and standard error are displayed by SASGSUB. SASGSUB also redirects standard input to the SAS session running on the grid.

-GRIDRUNSASDMS <hostname:displaynumber<.screennumber>
  specifies that the SASGSUB command runs in SAS Display Manager mode. Standard output and standard error are displayed by SASGSUB. The X Windows session is directed to the host and port specified by the hostname:displaynumber:screennumber options.

common_arguments
  specifies any of the arguments that are common to all uses of the SASGSUB command. Some arguments are required and some are optional. See “SASGSUB Syntax: Common Arguments” on page 161 for the syntax for each argument.

Optional Arguments

-GRIDWAITTIMEOUT timeout-value
  specifies that the interactive SAS session is terminated when the timeout-value (in seconds) is reached.

grid_job_arguments
  specifies any of the arguments that are valid when submitting jobs or commands to the grid. See “SASGSUB Syntax: Grid Job Arguments” on page 166 for the syntax for each of these arguments.

SASGSUB Syntax: Running a Command in Batch Mode

The following is the complete syntax for submitting a grid job to run a command in batch mode.

Syntax

SASGSUB
  -GRIDRUNCMD command common_arguments batch_arguments
    <grid_job_arguments>

Required Arguments

-GRIDRUNCMD command
  specifies a command that you want to run on the grid.
**common_arguments**

specifies any of the arguments that are common to all uses of the SASGSUB command. Some arguments are required and some are optional. See “SASGSUB Syntax: Common Arguments” on page 161 for the syntax for each argument.

**batch_arguments**

specifies any of the arguments that are common to all SASGSUB commands that run in batch mode. Some arguments are required and some are optional. See “SASGSUB Syntax: Batch Options” on page 164 for the syntax for each argument.

**Optional Argument**

**grid_job_arguments**

specifies any of the arguments valid when submitting jobs or commands to the grid. See “SASGSUB Syntax: Grid Job Arguments” on page 166 for the syntax for each of these arguments.

---

**Example: Submitting a Command in Batch Mode**

The following is an example of a SASGSUB batch mode statement used to submit a command to the grid to copy a set of files from the `prod` directory to the `backup` directory:

```
SASGSUB -GRIDRUNCMD "cp /prod/file1.* /backup"
```

---

**SASGSUB Syntax: Running a Command in Interactive Mode**

The following is the complete syntax for submitting a grid job to run a command interactively. If you terminate SASGSUB in interactive mode, you also terminate the SAS session running in the grid.

**Syntax**

```
SASGSUB
  -GRIDRUNCMDINT command common_arguments
  <!-GRIDWAITTIMEOUT>timeout-value
  <!grid_job_arguments>
```

**Required Arguments**

**-GRIDRUNCMDINT command**

specifies that an arbitrary command is run on the grid in interactive mode. Standard output and standard error is displayed on the user’s machine. Standard input is redirected from the client to the grid.

**common_arguments**

specifies any of the arguments that are common to all uses of the SASGSUB command. Some arguments are required and some are optional. See “SASGSUB Syntax: Common Arguments” on page 161 for the syntax for each argument.
Optional Arguments

-GRIDWAITTIMEOUT timeout-value
  specifies that the SAS Grid Manager Client Utility waits until either the job has completed running (either successfully or with an error) or timeout-value (in seconds) is reached. If you use this option with -GRIDWAIT or -GRIDWAITRESULTS, it enables you to specify how long to wait for results.

grid_job_arguments
  specifies any of the arguments valid when submitting jobs or commands to the grid. See “SASGSUB Syntax: Grid Job Arguments” on page 166 for the syntax for each of these arguments.

SASGSUB Syntax: Ending a Job

The following is the complete syntax for ending a job on a SAS grid. Enter the command on a Windows or UNIX command line.

Syntax

SASGSUB
  -GRIDKILLJOB job-id | ALL common_arguments

Required Arguments

-GRIDKILLJOB job-id | ALL
  terminates the job specified by job-id. If you specify ALL, all jobs are terminated.

common_arguments
  specifies any of the arguments that are common to all uses of the SASGSUB command. Some arguments are required and some are optional. See “SASGSUB Syntax: Common Arguments” on page 161 for the syntax for each argument.

Example: Ending a Job

The following is an example of a SASGSUB statement used to end the job 61361 that is running on the grid:

SASGSUB -GRIDKILLJOB 61361

SASGSUB Syntax: Viewing Job Status

The following is the syntax for using SASGSUB to view the status of a job on a SAS grid. Enter the command on a Windows or UNIX command line.
Syntax

SASGSUB
   -GRIDGETSTATUS job-id | ALL
   common_arguments batch_arguments

Required Arguments

-GRIDGETSTATUS job-id | ALL
   displays the status of the job specified by job-id. If you specify ALL, the status of all
   jobs for the current user is displayed.

common_arguments
   specifies any of the arguments that are common to all uses of the SASGSUB
   command. Some arguments are required and some are optional. See “SASGSUB
   Syntax: Common Arguments” on page 161 for the syntax for each argument.

batch_arguments
   specifies any of the arguments that are common to all SASGSUB commands that run
   in batch mode. Some arguments are required and some are optional. See “SASGSUB
   Syntax: Batch Options” on page 164 for the syntax for each argument.

Examples

Example 1: Viewing the Status of a Single Job
The following is an example of a SASGSUB statement used to view the status of job
61361 that is running on the grid:

   SASGSUB -GRIDGETSTATUS 61361

The output from the command looks like this:

Output 12.1 Status of a Single Job

   Current Job Information
   Job 61361 (testPgm) is Finished: Submitted: 06Jan2011:10:28:57, Started:

Example 2: Viewing the Status of All Jobs
The following is an example of a SASGSUB statement used to view the status of all jobs
running on the grid:

   SASGSUB -GRIDGETSTATUS ALL

The output from the command looks like this:
Output 12.2 Status of All Jobs

Current Job Information

Job information in directory U:\jobs\GridSub\GridWork\user1\SASGSUB-2011-05-11_13.17.17.327_testPgm is invalid.
Job 1925 (Pgm_04) is Submitted: Submitted: 08May2011:10:28:57

SASGSUB Syntax: Retrieving Job Output

The following is the syntax for using SASGSUB to retrieve the output of a job that has completed processing on a SAS grid. Enter the command on a Windows or UNIX command line.

Syntax

SASGSUB

-GRIDRESULTS job-id | ALL
  
  common_arguments batch_arguments
  
  <-GRIDRESULTSDIR directory>
  
  <-GRIDFORCECLEAN> <-VERBOSE>

Required Arguments

-GRIDRESULTS job-id | ALL

Copies the job information from the work directory to the directory specified by -GRIDRESULTSDIR for the specified job-id or for all jobs. If you do not specify a value for -GRIDRESULTSDIR, the job information is copied to the current directory. An error results if the current directory cannot be written to.

common_arguments

specifies any of the arguments that are common to all uses of the SASGSUB command. Some arguments are required and some are optional. See “SASGSUB Syntax: Common Arguments” on page 161 for the syntax for each argument.

batch_arguments

specifies any of the arguments that are common to all SASGSUB commands that run in batch mode. Some arguments are required and some are optional. See “SASGSUB Syntax: Batch Options” on page 164 for the syntax for each argument.

Optional Arguments

-GRIDRESULTSDIR directory

specifies the directory to which the job results are moved. If you do not specify a value for this argument, the default value is used, which is the current directory.

-GRIDFORCECLEAN

specifies that the job directory on the grid is deleted, regardless of whether the job was successful or not.
Examples

**Example 1: Retrieving the Output of a Job**
The following is an example of a SASGSUB statement used to view the output of job 61361 that is running on the grid:

```
SASGSUB -GRIDGETRESULTS 61361
```

The output from the command looks like this:

**Output 12.3  Output of a Single Job**

```
Current Job Information
Moved job information to \SASGSUB-2011-01-06_21.52.57.130_testPgm
```

**Example 2: Retrieving the Output of All Jobs**
The following is an example of a SASGSUB statement used to view the output of all jobs on the grid:

```
SASGSUB -GRIDGETRESULTS ALL
```

The output from the command looks like this:

**Output 12.4  Output of All Jobs**

```
Current Job Information
Moved job information to \SASGSUB-2011-05-06_21.52.57.130_Pgm1

Moved job information to \SASGSUB-2011-05-06_13.13.39.167_Pgm2

Moved job information to \SASGSUB-2011-05-06_13.16.06.060_Pgm3

Job information in directory U:\jobs\GridSub\GridWork\user1\SASGSUB-2011-05-06_13.17.17.327_Pgm4 is invalid.
Moved job information to \SASGSUB-2011-05-06_13.17.17.327_Pgm4

Job 1925 (Pgm4) is Submitted: Submitted: 08May2011:10:53:34
```

**SASGSUB Syntax: Common Arguments**

The following arguments are valid for all SASGSUB commands. Argument values that contain spaces must be contained in quotation marks.
Syntax

-`METASERVER server`
-`METAUSER user-ID`
-`METAPORT port`
-`METAPASS password`
-`METAPROFILE profile-name`
-`METACONNECT connection-name`
-`XMLCONFIGFILE pathname`
-`GRIDAPPSERVER sas-application-server`

\[
<\text{GRIDCONFIG grid-option-file}>
\]

\[
<\text{GRIDLICENSEFILE grid-enabled-license-file}>
\]

\[
<\text{LOGCONFIGLOC logging-option-file}>
\]

\[
<\text{GRIDLIBPATH path}>
\]

\[
<\text{VERBOSE}>
\]

\[
<\text{GRIDUSER grid-logon-username} | \_PROMPT_>
\]

\[
<\text{GRIDPASSWORD grid-logon-password} | \_PROMPT_>
\]

\[
<\text{GRIDPLUGINPATH grid-jar-file-path}>
\]

\[
<\text{JREOPTIONS java-runtime-options}>
\]

Required Arguments

-`GRIDAPPSERVER sas-application-server`
  specifies the name of the SAS Application Server that contains the grid's logical grid server definition. This option is stored in the configuration file that is automatically created by the SAS Deployment Wizard.

-`METASERVER server`
  specifies the name or IP address of the SAS Metadata Server. You must specify either -`METASERVER`, -`METAPORT`, -`METAUSER`, and -`METAPASS`; or -`METAPROFILE` and -`METACONNECT`; or -`XMLCONFIGFILE`. You cannot specify more than one group of options.
  This option is stored in the configuration file that is automatically created by the SAS Deployment Wizard.

-`METAPORT port`
  specifies the port to use to connect to the SAS Metadata Server specified by the `-METASERVER` argument. This option is stored in the configuration file that is automatically created by the SAS Deployment Wizard.

-`METAUSER user-ID`
  specifies the user ID to use to connect to the SAS Metadata Server specified by the `-METASERVER` argument. This option is stored in the configuration file that is automatically created by the SAS Deployment Wizard. Specify _PROMPT_ for `user-ID` to specify that the SAS Grid Manager Client Utility prompts the user for a user ID.

-`METAPASS password | _PROMPT_`
  specifies the password of the user specified in the `-METAUSER` argument. If the value of the argument is set to _PROMPT_, the user is prompted for a password. This option is stored in the configuration file that is automatically created by the SAS Deployment Wizard.

-`METAPROFILE profile_pathname`
  specifies the pathname of the connection profile for the SAS Metadata Server.
You must specify either -METASERVER, -METAPORT, -METAUSER, and 
-METAPASS; or -METAPROFILE and -METACONNECT; or -XMLCONFIGFILE.
You cannot specify more than one group of options.

This option is stored in the configuration file that is automatically created by the 
SAS Deployment Wizard.

If you are using a clustered metadata server, you can use -METAPROFILE to specify 
the clustered metadata server definition XML file. In this scenario, you do not need 
to specify the -METACONNECT option.

-METACONNECT connection-name
specifies the name of the connection to use when connecting to the SAS Metadata 
Server. The connection must be defined in the metadata profile specified in the 
-METAPROFILE argument. This option is stored in the configuration file that is 
automatically created by the SAS Deployment Wizard.

-XMLCONFIGFILE pathname
specifies the path to an XML file that contains the metadata profile connection 
information.

You must specify either -METASERVER, -METAPORT, -METAUSER, and 
-METAPASS; or -METAPROFILE and -METACONNECT; or -XMLCONFIGFILE.
You cannot specify more than one group of options.

Optional Arguments

-GRIDLICENSEFILE grid-enabled-license-file
specifies the path and filename of a SAS license file that contains the SAS Grid 
Manager license. The default value is to retrieve the license file information from 
metadata. If specified, the location must point to a valid SID file that contains a SAS 
Grid Manager license. Do not use this option unless instructed by SAS Technical 
Support.

-GRIDCONFIG grid-option-file
specifies the path and filename of a file containing other SASGSUB options. The 
default value is sasgsub.cfg.

-LOGCONFIGLOC logging-option-file
specifies the path and name of a file containing any options for the SAS logging 
facility. SASGSUB uses the App.Grid logger name with these keys:

App.Grid.JobID specifies the job ID as returned by the grid 
provider.
App.Grid.JobName specifies the job name.
App.Grid.JobStatus specifies the job status. Possible values are 
Submitted, Running, or Finished.
App.Grid.JobDir specifies the job directory name.
App.Grid.JobDirPath specifies the full path of job directory.
App.Grid.JobSubmitTime specifies the time at which the job was submitted.
App.Grid.JobStartTime specifies the time at which the job started running.
App.Grid.JobEndTime specifies the time at which the job completed.
App.Grid.JobHost specifies the host that ran the job.
GRIDLIBPATH path
the path to the shared libraries used by the utility. This value is set in the
configuration file and should not be altered. The path cannot contain spaces.

VERBOSE
specifies that extra debugging information is printed. If this argument is not
specified, only warning and error messages are printed.

GRIDUSER grid-logon-password | _PROMPT_
specifies the user name to be used to log on to the grid, if required by the grid
provider. This option is not required if the grid uses Platform Suite for SAS. Specify
_PROMPT_ for grid-logon-username to specify that the SAS Grid Manager Client
Utility prompts the user for a user ID.

GRIDPASSWORD grid-logon-password | _PROMPT_
specifies the password to log on to the grid, if required by the grid provider. This
option is not required if the grid uses Platform Suite for SAS. Specify _PROMPT_
for grid-logon-password to specify that the SAS Grid Manager Client Utility
prompts the user for a password.

GRIDPLUGINPATH grid-jar-file-path1 ... grid-jar-file-pathN
specifies a list of paths to search for additional grid provider JAR files. Paths are
separated by semicolons and cannot contain spaces. This option is not required if the
grid uses Platform Suite for SAS.

JREOPTIONS java-runtime-options
specifies any Java run-time options that are passed to the Java Virtual Machine. This
argument is required if you are using a grid provider other than Platform Suite for
SAS. This option is stored in the configuration file that is automatically created by
the SAS Deployment Wizard.

SASGSUB Syntax: Batch Options
The following options are used when running SASGSUB in batch mode. Argument values that contain
spaces must be contained in quotation marks.

Syntax

-GRIDWORK work-directory
  -GRIDWORKREM shared-file-system-path
  -GRIDFILESIN grid-file-list
  -GRIDSTAGECMD command
  -GRIDSTAGEHOST hostname
  -GRIDWAIT
  -GRIDWAITTIMEOUT timeout-value
  -GRIDWAITRESULTS
  -GRIDWAITLOGLIST
  -GRIDWAITNORESULTS
  -GRIDWAITRESULTSNLOG
  -GRIDWATCHOUTPUT

Required Argument

-GRIDWORK work-directory
specifies the path for the shared directory that the job uses to store the program,
output, and job information. On Windows, the path can contain spaces, but must be
in quotation marks. On UNIX, the path cannot contain spaces. This option is stored
in the configuration file that is automatically created by the SAS Deployment
Wizard.

Optional Arguments

-GRIDWORKREM shared-file-system-path
  when using a shared file system, specifies the pathname of the GRIDWORK
directory in the shared file system relative to a grid node. Use this argument when
the machine used to submit the job is on a different platform than the grid. The path
cannot contain spaces.

  When you are using file staging, this argument specifies the location of the
GRIDWORK directory as passed to the GRIDSTAGECMD parameter. For example,
Windows clients could have GRIDWORK set to \myServer\myShare\SharedPath with the grid using LSRCP to stage files with
GRIDSTAGEFILEHOST set to myServer.mydomain.com. If you set Windows
to map C:myShare on myServer to the Windows share \myServer\myShare,
you would set GRIDWORKREM to C:\myShare\SharedPath.

-GRIDFILESIN grid-file-list
  specifies a comma-separated list of files that need to be moved to the grid work
directory before the job starts running.

-GRIDSTAGECMD command
  specifies the remote copy command used to stage files to the grid. Valid values are
LSRCP, RCP, SCP, or SMBCLIENT. This option is used when the grid client
machines and the grid machines do not share a common directory structure.

-GRIDSTAGEHOST hostname
  specifies the name of the host that stores files that are staged into the grid. This
option is used when a machine submits files to the grid and then shuts down.

-GRIDWAIT
  specifies that the SAS Grid Manager Client Utility waits until the job has completed
running, either successfully or with an error. If the job does not complete, it must be
ended manually. The results of the job are not returned, and the job directory is
retained on the GRIDWORK location.

-GRIDWAITTIMEOUT timeout-value
  specifies that the SAS Grid Manager Client Utility waits until either the job has
completed running (either successfully or with an error) or timeout-value (in
seconds) is reached. If you use this option with -GRIDWAIT or -
GRIDWAITRESULTS, it enables you to specify how long to wait for results.

-GRIDWAITRESULTS
  specifies that the SAS Grid Manager Client Utility waits until the job has completed
running (either successfully or with an error). It then returns the results of the
processing, including the job directory and the job log and output listing. If the job
does not complete, it must be ended manually.

-GRIDWAITLOGLIST
  specifies that the SAS Grid Manager Client Utility waits until the job has completed
running (either successfully or with an error). It then returns the results of the
processing. After obtaining the job log and output listing, the job directory is deleted.
If the job does not complete, it must be ended manually.

-GRIDWAITNORESULTS
  specifies that the SAS Grid Manager Client Utility waits until the job has completed
running (either successfully or with an error). It then returns the results of the
processing and deletes the job log, the output listing, and the job directory. If the job
does not complete, it must be ended manually.

-GRIDWAITRESULTSNOLOG
specifies that, after a job has completed running and the results of the processing are
returned, the job log, output listing, and job directory are deleted, and the log and list
files are not copied to the current directory.

-GRIDWATCHOUTPUT
specifies that the output of what was submitted by the SASGSUB command is
displayed on the user’s machine. If this argument is used with -GRIDSUBMITPGM,
the SAS log and output are displayed. If the argument is used with
-GRIDRUNCMD, the command’s standard output and standard error are displayed.
While the output is being displayed, entering the command prompt does not affect
the processing in the grid.

SASGSUB Syntax: Grid Job Arguments
The following arguments are used when submitting a SAS job or command to the grid. Argument values
that contain spaces must be contained in quotation marks.

Syntax

<GRIDJOBNAME grid-program-name>
<GRIDJOBOPTS grid-provider-options>
<GRIDOPTSET options-set-name>
<GRIDWORKLOAD grid-resource-names>

Optional Arguments

-GRIDJOBNAME grid-program-name
specifies the name of the grid job as it appears on the grid. If this argument is not
specified, the SAS program name is used.

-GRIDJOBOPTS grid-provider-options
specifies any options that are passed to the grid provider when the job is submitted.
You can separate multiple options with spaces or with semicolons. If you use spaces,
you must enclose the option string in quotation marks. For a grid that uses Platform
Suite for SAS, see Appendix 1, “Supported Job Options - Platform Suite for SAS,”
on page 169. For a grid that uses SAS Grid Manager for Hadoop, see Appendix 2,
“Supported Job Options - SAS Grid Manager for Hadoop,” on page 171.

-GRIDOPTSET options-set-name
specifies the name of a grid options set to be used in job processing. This option is
needed only if there is more than one grid options set associated with the current
SAS Grid Manager Client Utility user. If there is only one grid options set for the
user, it is used by default. A grid options set is a collection of grid options, SAS
options, and required resources that are associated with a particular SAS client
application and SAS user or group.

-GRIDWORKLOAD grid-resource-name
specifies a resource name to use when submitting the job to the grid.
Appendix

Appendix 1
Supported Job Options - Platform Suite for SAS ............... 169

Appendix 2
Supported Job Options - SAS Grid Manager for Hadoop ........ 171
Appendix 1

Supported Job Options - Platform Suite for SAS

The following table lists the job options that are supported by Platform Suite for SAS. You can specify these options in these locations:

- the JOBOPTS= option of the GRDSVC_ENABLE function
- the Additional Options field in the metadata definition for the SAS Logical Grid Server
- the Grid Options field in the grid server definition and in grid options set definitions

Options specified in metadata override those specified in a GRDSVC_ENABLE statement.

Table A1.1  Platform Suite for SAS Job Option Name/Value Pairs

<table>
<thead>
<tr>
<th>Job Option Name/Value Pairs</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>app=grid_application_name</td>
<td>specifies the LSF application name to use. This option is valid only for LSF version 7 and higher.</td>
</tr>
<tr>
<td>exclusive=0</td>
<td>1</td>
</tr>
<tr>
<td>ignoreFull=value</td>
<td>if value is anything other than 0, specifies that during object spawner load balancing, SAS Grid Manager ignores the closed status of a host if the host is closed because it is full of jobs. If the value is 0, a host that is closed because it is full is not used for the next server.</td>
</tr>
<tr>
<td>jobgroup=job-group</td>
<td>specifies the name of the job group to associate with the job.</td>
</tr>
<tr>
<td>jobname=job_name</td>
<td>specifies a job name to be used for all jobs. The name specified in this option overrides any job name specified by an application.</td>
</tr>
<tr>
<td>jobSlots=number_of_slots</td>
<td>specifies the number of available job slots for the grid. Use this option in configurations where the grid is controlled by Platform Computing’s Enterprise Grid Orchestrator (EGO). EGO always shows grid hosts as being closed. However, specifying a high value for this option shows the grid as being open and enables jobs to be submitted to the grid.</td>
</tr>
<tr>
<td>Job Option Name/Value Pairs</td>
<td>Explanation</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>maxpendingtime=\textit{time}</td>
<td>specifies the number of seconds a job can wait in the pending state before it is terminated.</td>
</tr>
<tr>
<td>priority=\textit{job-priority}</td>
<td>specifies the user-assigned job priority. This is a value between 1 and MAX_USER_PRIORITY, as defined in the \texttt{lsb.params} file.</td>
</tr>
<tr>
<td>project=\textit{project}</td>
<td>specifies the name of the project to associate with the job.</td>
</tr>
<tr>
<td>queue=\textit{queue}</td>
<td>specifies the name of the queue to put the job in. The default queue name is \texttt{normal}.</td>
</tr>
<tr>
<td>reqres=&quot;\textit{requested-resources}&quot;</td>
<td>specifies additional resource requirements.</td>
</tr>
<tr>
<td>runlimit=\textit{time-in-seconds}</td>
<td>specifies the maximum amount of time that a job is allowed to run. This value is used as an absolute limit or as part of an SLA job.</td>
</tr>
<tr>
<td>sla=\textit{service-level-agreement}</td>
<td>specifies the name of the service-level agreement to associate with the job.</td>
</tr>
<tr>
<td>usergroup=\textit{user-group}</td>
<td>specifies the name of the user group.</td>
</tr>
</tbody>
</table>

For complete information about job options, see \textit{Platform LSF Reference}. 
Appendix 2

Supported Job Options - SAS Grid Manager for Hadoop

The following table lists the job options that are supported by SAS Grid Manager for Hadoop. You can specify these options in these locations:

- the JOBOPTS= option of the GRDSVC_ENABLE function
- the Additional Options field in the metadata definition for the SAS Logical Grid Server
- the Grid Options field in the grid server definition and in grid options set definitions

Options specified in metadata override those specified in a GRDSVC_ENABLE statement.

<table>
<thead>
<tr>
<th>Job Option</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>appMasterMem=memory</td>
<td>specifies amount of memory in megabytes needed for the application master to run.</td>
</tr>
<tr>
<td>appType=applicationTypeName</td>
<td>specifies the application type to use when running the job. applicationTypeName must be defined in the SASGRID policy file. If applicationTypeName does not match an entry in the policy file, the application type that is specified by the defaultAppType attribute is used. See “Creating the SASGRID Policy File” on page 91.</td>
</tr>
<tr>
<td>cfgPath=path</td>
<td>specifies the absolute path to the Hadoop configuration files.</td>
</tr>
<tr>
<td>jaasCfgFile=path</td>
<td>specifies the absolute path to the JAAS configuration files.</td>
</tr>
<tr>
<td>jarFile=path</td>
<td>specifies the path to the sas.grid.provider.yarn.jar file.</td>
</tr>
<tr>
<td>jarPath=path</td>
<td>specifies the path to the Hadoop JAR files.</td>
</tr>
<tr>
<td>javaOpts=options</td>
<td>specifies the Java options to add to the Java command for Hadoop clusters.</td>
</tr>
<tr>
<td>krb5ccFile=path</td>
<td>specifies the absolute path to the Kerberos 5 credential cache for the user.</td>
</tr>
<tr>
<td>Job Option</td>
<td>Explanation</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>krb5CfgFile=\path</td>
<td>specifies the absolute path to the Kerberos 5 configuration file.</td>
</tr>
<tr>
<td>noHDFS</td>
<td>specifies that HDFS is not used.</td>
</tr>
<tr>
<td>noMapRLogin</td>
<td>specifies that <code>maprlogin kerberos</code> is not automatically issued when submitting jobs to MapR clusters.</td>
</tr>
<tr>
<td>noSecurity</td>
<td>specifies that security is not enabled.</td>
</tr>
<tr>
<td>policyFile=\path</td>
<td>specifies the absolute path to the SAS Grid policy XML file.</td>
</tr>
<tr>
<td>testHost=\hostname</td>
<td>specifies a host to use for grid validation. The job is sent to the specified host using minimum resource requirements.</td>
</tr>
<tr>
<td>useJava</td>
<td>specifies that Java should be used, even if the Hadoop cluster supports the YARN REST API.</td>
</tr>
<tr>
<td>useLog4j</td>
<td>specifies that the file <code>hdfs://tmp/SASGrid/log4j.properties</code> file should be used for logging. You must copy the file to this location manually.</td>
</tr>
<tr>
<td>useSSL</td>
<td>specifies that the job should use an SSL connection.</td>
</tr>
<tr>
<td>vjrPath</td>
<td>specifies the absolute path for the versioned JAR repository (VJR).</td>
</tr>
</tbody>
</table>
Recommended Reading

Here is the recommended reading list for this title:

- *SAS/CONNECT User's Guide*
- *SAS Environment Manager: User's Guide*
- *SAS Intelligence Platform: Installation and Configuration Guide*
- *SAS Statements: Reference*
- *SAS Macro Language: Reference*
- *Scheduling in SAS*

- SAS offers instructor-led training and self-paced e-learning courses to help you get started with the SAS add-in, learn how the SAS add-in works with the other products in the SAS Enterprise Intelligence Platform, and learn how to run stored processes in the SAS add-in. For more information about the courses available, see sas.com/training.

For a complete list of SAS publications, go to sas.com/store/books. If you have questions about which titles you need, please contact a SAS Representative:

SAS Books
SAS Campus Drive
Cary, NC 27513-2414
Phone: 1-800-727-0025
Fax: 1-919-677-4444
Email: sasbook@sas.com
Web address: sas.com/store/books
application server
a server that is used for storing applications. Users can access and use these server applications instead of loading the applications on their client machines. The application that the client runs is stored on the client. Requests are sent to the server for processing, and the results are returned to the client. In this way, little information is processed by the client, and nearly everything is done by the server.

authentication
the process of verifying the identity of a person or process within the guidelines of a specific authorization policy.

authentication domain
a SAS internal category that pairs logins with the servers for which they are valid. For example, an Oracle server and the SAS copies of Oracle credentials might all be classified as belonging to an OracleAuth authentication domain.

grid
a collection of networked computers that are coordinated to provide load balancing of multiple SAS jobs, scheduling of SAS workflows, and accelerated processing of parallel jobs.

grid computing
a type of computing in which large computing tasks are distributed among multiple computers on a network.

grid control server
the machine on a grid that distributes SAS programs or jobs to the grid nodes. The grid control server can also execute programs or jobs that are sent to the grid.

grid monitoring server
a metadata object that stores the information necessary for the Grid Manager plug-in in SAS Management Console to connect with the Platform Suite for SAS to allow monitoring and management of the grid.

grid node
a machine that is capable of receiving and executing work that is distributed to a grid.

identity
See metadata identity.
job
a metadata object that specifies processes that create output.

load balancing
for IOM bridge connections, a program that runs in the object spawner and that uses an algorithm to distribute work across object server processes on the same or separate machines in a cluster.

logical grid server
a metadata object that stores the command that is used by a grid-enabled SAS program to start a SAS session on a grid.

logical server
in the SAS Metadata Server, the second-level object in the metadata for SAS servers. A logical server specifies one or more of a particular type of server component, such as one or more SAS Workspace Servers.

login
a SAS copy of information about an external account. Each login includes a user ID and belongs to one SAS user or group. Most logins do not include a password.

metadata identity
a metadata object that represents an individual user or a group of users in a SAS metadata environment. Each individual and group that accesses secured resources on a SAS Metadata Server should have a unique metadata identity within that server.

metadata repository
a collection of related metadata objects, such as the metadata for a set of tables and columns that are maintained by an application. A SAS Metadata Repository is an example.

metadata server
a server that provides metadata management services to one or more client applications. A SAS Metadata Server is an example.

plug-in
a file that modifies, enhances, or extends the capabilities of an application program. The application program must be designed to accept plug-ins, and the plug-ins must meet design criteria specified by the developers of the application program. In SAS Management Console, a plug-in is a JAR file that is installed in the SAS Management Console directory to provide a specific administrative function. The plug-ins enable users to customize SAS Management Console to include only the functions that are needed.

SAS Management Console
a Java application that provides a single user interface for performing SAS administrative tasks.

SAS Metadata Repository
one or more files that store metadata about application elements. Users connect to a SAS Metadata Server and use the SAS Open Metadata Interface to read metadata from or write metadata to one or more SAS Metadata Repositories. The metadata types in a SAS Metadata Repository are defined by the SAS Metadata Model.
SAS Metadata Server
a multi-user server that enables users to read metadata from or write metadata to one or more SAS Metadata Repositories. The SAS Metadata Server uses the Integrated Object Model (IOM), which is provided with SAS Integration Technologies, to communicate with clients and with other servers.

SAS Workspace Server
a SAS IOM server that is launched in order to fulfill client requests for IOM workspaces. See also IOM server and workspace.
## Index

**A**  
accelerated processing 3  
address resource utility 17  
addresses  
  grid server address 101  
  host address 72  
  IP address of grid nodes 142  
analysis on data 11  
aplications  
  configuring client applications 23  
  grid enabling 106  
  SAS applications supporting grid processing 9  
appType option 91, 94  
asynchronous rsubmits 108  
authentication domain 24, 101  

**B**  
batch jobs 8  
  submitting to grid 109  
business problems 10  
  high availability 10  
  increased data growth 10  
  many users on single resource 10  
  need for flexible IT infrastructure 11  
  running larger and more complex analysis 11  

**C**  
central file server  
  configuring 16  
check point restart 67  
clients  
  configuring client applications 23  
  grid clients 5  
  running jobs and 78  
Cloudera 84, 86  
complexity of analysis 11  
computer failure 11  
configuration  
  central file server 16  
  client applications 23  

**D**  
data analysis 11  
  flexible IT infrastructure and 11  
data growth 10  
data volume 10  
debug jobs 75  
distributed enterprise scheduling 8  
distributed parallel execution of jobs 114  
DNS resolution 66  

**E**  
ending jobs 58, 111, 158  
environment  
  information about grid environment 143  
  verifying Platform Suite for SAS environment 73  
  verifying SAS environment 75  

**F**  
failover
of critical services 63
of SAS programs 64
functions 137

G
Gantt charts 59
graphs
displaying job graphs 59
GRDSVC_ENABLE function 137
result codes 141
specifying resource names 39
GRDSVC_GETADDR function 142
GRDSVC_GETINFO function 143
result codes 145
GRDSVC_GETNAME function 146
GRDSVC_NNODES function 149
result codes 151
grid
comparing submission methods 114
configuring 4, 15
enabling or disabling grid execution 137
information about 57, 143
installing Platform Suite for SAS 16
maintaining connection to 118
grid architecture
Platform Suite for SAS 5
SAS Grid Manager for Hadoop 6
grid clients 5
jobs running on 78
Grid Command property 100
grid computing 3
business problems solved by 10
processing types 8
grid control machine 74
grid control server 4
configuring 16
grid enabling 106
distributed parallel execution of jobs 114
SAS Add-In for Microsoft Office and 116
SAS Data Integration Studio and 120
SAS Display Manager and 106
SAS Enterprise Guide and 116
SAS Enterprise Miner and 123
SAS Grid Manager for workspace server load balancing 125
SAS Risk Dimensions and 125
scheduling jobs on grid 113
grid environment
information about 143
planning and configuring 15
grid maintenance 57
closing and reopening hosts 60
displaying job graphs 59
managing jobs 58
managing queues 61
viewing grid information 57
grid management 27
defining resources 38
grid options sets 129
overview 27
queues 31
specifying job slots 29
Grid Manager agent plug-in 4
Grid Manager module 4
Grid Manager plug-in 55
closing and reopening hosts 60
displaying job graphs 59
job views 55
maintaining the grid 57
managing jobs 58
managing queues 61
viewing grid information 57
grid metadata 99
verifying 75
grid monitoring
verifying 76
grid monitoring server
modifying definitions 23
verifying SAS grid metadata 76
viewing grid information 57
grid nodes 4
configuring 23
IP address of 142
name of 146
number of grid nodes is zero 77
required software components for 23
testing connectivity to 76
grid options sets 28, 129
creating 130
mapping 132
merging option values 130
use with SAS Grid Manager for Hadoop 94
grid processing
processing types supported 8
SAS applications supporting 9
grid server
address 101
port 101
grid topology 4
grid-launched workspace servers 125
GRIDOPTSET= option
GRDSVC_ENABLE function 138, 145

H
Hadoop 83
Hadoop distributions 84
hardware load balancer 65
high availability
and critical applications 64
overview 10
high-priority queues 33
Hortonworks 84, 86
host name 24
hosts
addresses 72
closing and reopening 60
connectivity 72
information about 57
ports 73

I
installation 15
Platform Suite for SAS 16
Platform Suite for SAS, on UNIX 17
SAS Grid Manager Client Utility 102
SAS products and metadata definitions 15
interactively developing SAS programs 118
invalid resource request 77
invalid userid or password 76
IP address of grid nodes 142
IT infrastructure 11
iterative processing 8

J
job
requeuing 68
job graphs 59
job options 169
job scheduling
See scheduling jobs
job slots 28
increasing 119
specifying 29
specifying limits on a queue 37
total number available on grid 149
job views 55
JOBNAME= option
GRDSVC_ENABLE function 139
JOBOPTS= option
GRDSVC_ENABLE function 140
jobs
See also scheduling jobs
comparing grid submission methods 114
debug jobs 75
distributed parallel processing 114
ending 58, 111, 158
failure to start 77
information about 57
machines that are only grid clients 78
managing 58
prioritizing 120
queue for short jobs 36
restarting 67
resuming 58
retrieving output 111, 160
status of 111, 158
submitting batch jobs to grid 109
submitting from Program Editor to grid 107
submitting with SASGSUB command 154, 155
suspending 58
terminating 58, 111, 158
verifying LSF job execution 75
verifying SAS job execution 76
viewing log and output lines from 108

L
label restart 67
LIBNAME statement
assigning SAS Add-In for Microsoft
Office libraries 118
assigning SAS Enterprise Guide
libraries 118
libraries
assigning SAS Add-In for Microsoft
Office libraries 118
assigning SAS Enterprise Guide
libraries 118
browsing with SAS Explorer Window 108
license
for SAS Grid Manager 77
license file 74
load balancer 65
load balancing
multi-user workload balancing with
SAS Data Integration Studio 120
SAS Grid Manager for workspace
server load balancing 125
Load Sharing Facility (LSF)
specifying job slots 29
specifying workload for Loop
Transformation 123
terminating LSF jobs 50, 58
verifying job execution 75
verifying LSF is running 73
verifying setup 74
log lines
viewing from grid jobs 108
logical grid server
modifying definitions 99
verifying SAS grid metadata 76
Loop Transformation
    specifying workload for 123

M
maintenance issues 11
many users on single resource 10
MapR 84
merging option values 130
METAAUTORESOURCES option
    assigning SAS Add-In for Microsoft Office libraries 118
    assigning SAS Enterprise Guide libraries 118
metadata
    modifying logical grid server definitions 99
    verifying SAS grid metadata 75
metadata definitions
    installing and configuring 15
metadata server 4
    verifying grid metadata 75
model scoring 123
model training 123
Module Name property 24
monitoring
    verifying 76
multi-user workload balancing 8, 120

N
names
    grid nodes 146
    host name 24, 130
    Module Name property 24
    specifying resource names 39
network setup verification 72
    host addresses 72
    host connectivity 72
    host ports 73
night queues 34
nodes
    See grid nodes
NORMAL queue 31, 33

O
OLAP servers
    converting to grid load balancing 126
Oozie 95
option values
    merging with grid options sets 130
Options property 24
output
    retrieving 111, 160
viewing output lines from grid jobs 108

P
parallel execution, distributed 114
parallel scoring 123
parallel workload balancing 8, 121
partitions 28
password invalid 76
performance
    assigning SAS Add-In for Microsoft Office libraries 118
    assigning SAS Enterprise Guide libraries 118
ping command 72
planning grid environment 15
Platform RTM for SAS
    host name 24
Platform Suite for SAS
    installing and configuring 16
    installing on UNIX 17
    supported job options 169
    verifying environment 73
    verifying LSF job execution 75
    verifying LSF setup 74
    verifying that LSF is running 73
policy file 91
ports 24
    grid server port 101
    verifying host ports 73
pre-assigned libraries
    assigning SAS Add-In for Microsoft Office libraries 118
    assigning SAS Enterprise Guide libraries 118
prioritizing jobs 120
processing types 8
    distributed enterprise scheduling 8
    multi-user workload balancing 8
    parallel workload balancing 8
processing, iterative 8
profile.lsf file 17
Program Editor
    submitting jobs to grid from 107
programs
    developing interactively with SAS Add-In for Microsoft Office 118
    developing interactively with SAS Enterprise Guide 118
Provider property 24, 100

Q
queues 28, 31
    activating 61
    closing 61
configuring 32
for short jobs 36
high-priority 33
inactivating 61
information about 57
managing 61
mismatch with workspace servers 37
night queue 34
NORMAL queue 31, 33
opening 61
server specification 37
specifying 31
specifying job slot limits on 37

R
refused connection 76
requeuing 68
required resources 130
resource names
specifying in grid options sets 130
specifying in SAS Data Integration Studio 39
specifying with GRDSVC_ENABLE function 39
specifying with SAS Grid Manager Client Utility 39
resources
many users on single resource 10
specifying LSF resources for Loop Transformation 123
specifying names 38
restart
checkpoint and label 67
restarting jobs 67
resuming jobs 58
rsubmits, asynchronous 108

S
SAS Add-In for Microsoft Office
assigning libraries 118
developing SAS programs interactively 118
grid enabling and 116
SAS applications
grid enabling 106
specifying as grid capable 133
supporting grid processing 9
SAS Code Analyzer 114
SAS Data Integration Studio
grid enabling and 120
multi-user workload balancing with 120
parallel workload balancing with 121
scheduling jobs 120
specifying resource names 39
specifying workload for Loop Transformation 123
SAS Deployment Wizard
configuring grid control server 16
configuring grid nodes 23
SAS Display Manager 106
browsing libraries with SAS Explorer Window 108
submitting jobs from Program Editor to grid 107
viewing log and output lines from grid jobs 108
SAS Enterprise Guide
assigning libraries 118
developing SAS programs interactively 118
grid enabling and 116
maintaining connection to the grid 118
setting workload values 119
users 27
SAS Enterprise Miner
grid enabling and 123
users 27
SAS environment
verifying 75
verifying grid metadata 75
verifying grid monitoring 76
verifying SAS job execution 76
SAS Explorer Window
browsing libraries 108
SAS Grid Manager 3
components 4
for workspace server load balancing 125
grid enabling and 125
license for 77
loading 77
multi-user workload balancing 8
parallel workload balancing 8
SAS Grid Manager Client Utility 109
ending jobs 111
ending jobs with SASGSUB command 158
installing and configuring 102
retrieving job output 111, 160
SASGSUB command 153
specifying resource names 39
submitting batch jobs to grid 110
submitting jobs 114, 154, 155
use with SAS Grid Manager for Hadoop 94
viewing job status 111, 158
SAS Grid Manager for Hadoop
deployment 84
overview 83
specifying for Loop Transformation 123
workload balancing 3
multi-user 8, 120
parallel 8, 121
workload values
setting for SAS Enterprise Guide 119
WORKLOAD= option
GRDSVC_ENABLE function 139
workspace servers
converting to grid load balancing 126
grid-launched 125
SAS Grid Manager for load balancing 125
started by object spawner 126
Gain Greater Insight into Your SAS® Software with SAS Books.

Discover all that you need on your journey to knowledge and empowerment.

support.sas.com/bookstore for additional books and resources.