SAS® Grid Manager

Workload balancing. High availability. Faster processing.
All in a flexible, centrally managed grid computing environment.

SAS. THE POWER TO KNOW®

What does SAS® Grid Manager do?

SAS Grid Manager uses patented technology to deliver enterprise-class capabilities that enable many SAS solutions to automatically use a centrally managed grid computing infrastructure to provide workload balancing, high availability and parallel processing for analytics jobs and processes.

Why is SAS® Grid Manager important?

SAS Grid Manager makes it easy to accommodate compute-intensive applications and a growing number of users cost-effectively across your available hardware resources – while also ensuring continuous high availability for your analytics applications. With SAS Grid Manager, you can create a managed, shared environment for efficiently processing large volumes of SAS and open source jobs.

For whom is SAS® Grid Manager designed?

It is designed for CIOs, IT managers, data center managers and grid computing architects seeking to manage SAS programs in a shared environment. It also can be used by statisticians, business analysts and application developers to reduce processing times – and achieve faster results – for data integration, reporting and analytics jobs.

Analytical and data management workloads require a flexible infrastructure that can meet peak demands and support a growing and increasingly diverse set of user-driven jobs.

IT budgets are typically limited, which makes meeting these computing demands a constant challenge. Organizations need to find a way to scale cost-effectively, regardless of whether infrastructure is on premises or in the cloud. SAS Grid Manager is the answer.

SAS Grid Manager enables organizations to create a managed, shared grid computing environment for processing large volumes of data and analytics programs. It provides critical capabilities, including:

- Workload balancing.
- Job scheduling and prioritization.
- High availability.
- Parallel processing.
- Resource assignment and monitoring.

Benefits

- **Meet changing business demands with dynamic workload balancing.** SAS Grid Manager gives IT flexibility to meet service level commitments by easily reassigning computing resources to meet peak workloads or changing business demands. It provides a central point of control for administering policies, programs, queues and job prioritization across users and applications to achieve business goals under a given set of constraints.

- **Create a highly available SAS computing environment.** Having multiple servers in a grid computing environment enables jobs to run on the best available resource. If a server fails, its jobs can be transitioned seamlessly to another server. In addition, IT staff can perform maintenance on specific servers, as well as introduce additional computing resources, without interrupting analytics jobs or disrupting the business.

- **Get faster results from your existing IT infrastructure.** Multiprocessing capabilities let you divide individual jobs into subtasks that run in parallel, reducing processing time. This is particularly effective for

analytics programs with large data sets and long run times, as well as those with repetitive runs of independent tasks running against large data sets.

- **Grow computing resources incrementally and cost-effectively.** Take advantage of all available computing resources now and cost-effectively scale out as needed, adding capacity in single processing units with commodity hardware. There's no need to size today's environment for anticipated future needs.

- **Manage all your analytical workloads.** In an increasingly diverse analytics ecosystem, analysts are using a variety of programming languages. With SAS Grid Manager, you can manage all your jobs, including SAS and other languages, ensuring all your analytics run quickly.

- **Easily transition from a siloed server or multiple PC environment to a SAS grid environment.** Many SAS solutions are seamlessly integrated with SAS Grid Manager, enabling administrators to configure submission to the grid in SAS metadata. This means no change to the way analysts interact with SAS jobs that are directed to the grid.

Overview

SAS Grid Manager gives you faster results by balancing user and application workloads among available computing resources. IT can incrementally add computing resources – in the form of lower-cost commodity hardware – eliminating the need to size today's environment for tomorrow's demands.

Dynamic, resource-based workload balancing

SAS Grid Manager delivers enterprise-class dynamic workload balancing for multiple users and applications. It automates the management and optimization of SAS grids and provides resilience by efficiently distributing the processing of analytical programs, both SAS and open source, across multiple CPUs.

Centralized administration lets you enforce policies to determine job prioritization based on different users' needs, as well as deliver better service levels to business units. And it gives IT staff the flexibility to easily identify and allocate computing resources to match changing demands.

Shared environment with high availability

With SAS Grid Manager, there's no need to purchase a separate third-party tool for high availability because the solution includes high-availability capabilities for all services critical to a SAS environment. Failover to another node within the grid eliminates the need for a hot standby. And you can perform rolling maintenance with no interruption to users. Guaranteed execution of critical batch SAS jobs ensures that they complete successfully and within the expected time frame through the use of SAS checkpoint and automatic restart functionality.

Grid-enabled SAS®

Several solutions that generate SAS programs, including SAS Data Integration Studio (a component of SAS Enterprise Data Integration Server and SAS Data Integration Server) and SAS® Enterprise Miner™, are automatically tailored for parallel processing in a grid computing environment. To achieve maximum processing efficiency with minimum user intervention, these programs detect the grid environment at the time of execution. The grid-enabled logic that is produced can be saved as SAS Stored Processes for use by other SAS reporting clients to generate results for more users as cost-effectively as possible. As a result, you can integrate, cleanse and analyze larger volumes of data more quickly.

You can set up other SAS solutions, including SAS® Enterprise Guide® and SAS Studio, to automatically submit SAS jobs to a grid of shared computing resources. All SAS programs can take
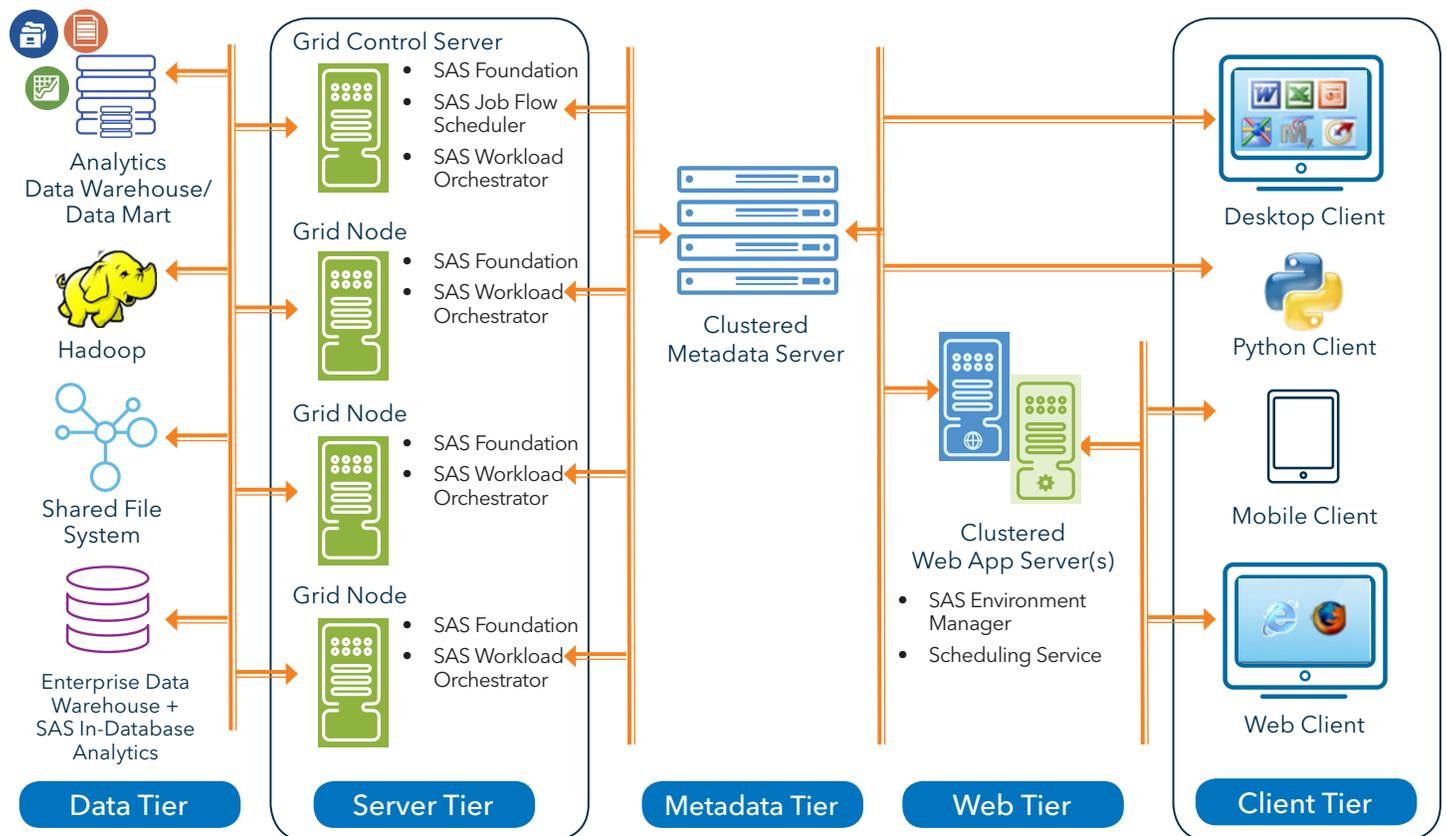

Figure 1: SAS Grid Computing architecture.

advantage of a grid computing environment with the addition of programming syntax and a structure that allows the submission of entire programs to the grid or the parallel execution of program steps (subtasks). The grid option of the SAS Code Analyzer in Base SAS software automates this process by generating a new version of a SAS program that has been restructured to run in a distributed environment.

Automated, web-based management and monitoring

SAS Grid Manager includes a web-based tool for monitoring and managing resources, users and jobs running in the grid. It also serves as an interface for configuring and managing high-availability services, as well as defining alerts to be sent when thresholds are exceeded.

Scalability for running more complex analytics faster

You can manage a wide variety of SAS and open source jobs across grid environments for optimal resource utilization and faster processing. Individual SAS jobs can be divided into subtasks that are then executed in parallel to accelerate processing and increase workload throughput. You can also schedule your production SAS workflows to run across grid resources using the schedule manager plug-in within the SAS Management Console.

To learn more about SAS Grid Manager and see other related material, please visit sas.com/grid.

Key Features

A managed, shared environment
- Improves efficiency of program distribution and CPU utilization through dynamic, resource-based load balancing.
- Makes computing resources available to multiple users and multiple applications for running larger or more complex analysis.
- Provides job, queue, host and user management across your enterprise.
- Enables job prioritization by rules-based job queues to govern the use of computing resources.
- Provides automatic identification, allocation, management and optimization of computing resources and program flows.
- Allows administrators to easily create a set of metadata-defined grid and SAS options to be applied automatically to workloads submitted to the grid based on the user's identity and the application being used to access the grid.
- Simplifies administration of SAS environments through centralized policies.

High availability
- Provides high-availability capabilities for critical SAS services such as the SAS Metadata Server.
- Uses a grid node as a hot-standby machine for failover and enables nondisruptive, rolling maintenance.
- Detects hardware and software failures in the grid and recovers appropriately.
- Ensures that SAS and open source jobs will be completed optimally.
- Restarts SAS jobs automatically from the last successful checkpoint.

Grid-enabled SAS®
- Speeds up processing of analytics jobs and provides more efficient computing resource utilization.
- Provides load balancing for all SAS workloads to improve throughput and response time of all SAS clients.
- Enables many SAS solutions and user-written programs to be configured easily for submission to a grid of shared resources.
- Uses the SAS Code Analyzer to analyze job dependencies in SAS programs and generate grid-ready code.
- Provides automated session spawning and distributed processing across a set of diverse computing resources.
- Enables scheduling of production analytical workflows to be executed across grid resources:
  - Provides a process flow diagram to create SAS flows of one or more SAS jobs that can be simple or complex to meet your needs.
  - Uses all of the policies and resources of the grid.
- Integrates with all SAS analytics applications by storing grid-enabled code as SAS Stored Processes.
- Provides greater resilience for mission-critical applications and high availability for the SAS environment.
- Includes a command-line utility to automate batch submissions.
- Allows you to submit and forget, and reconnect later to retrieve results.
- Enables integration with other standard enterprise schedulers.
- Enables batch submission to leverage checkpoint information and automatically restart jobs.
- Applies grid policies to SAS workspace servers when they are launched through the grid.

Key Features (continued)

Real-time monitoring and administration
- Provides a web-based tool for monitoring and administering multiple SAS grids, including:
  - Visuals for tracking resource usage, users and jobs running on the grid.
  - A GUI for modifying grid configurations and defining alerts when critical thresholds are crossed.
  - A GUI for configuring and managing critical services for high availability.
  - Support for filtering and role definitions for customizing displays and activities that can be performed.

Flexible infrastructure
- Allows you to add computing resources incrementally to cost-effectively accommodate a growing number of users, as well as meet increased business needs.
- Decouples the computing infrastructure from the SAS applications to allow business users to focus on their processes.
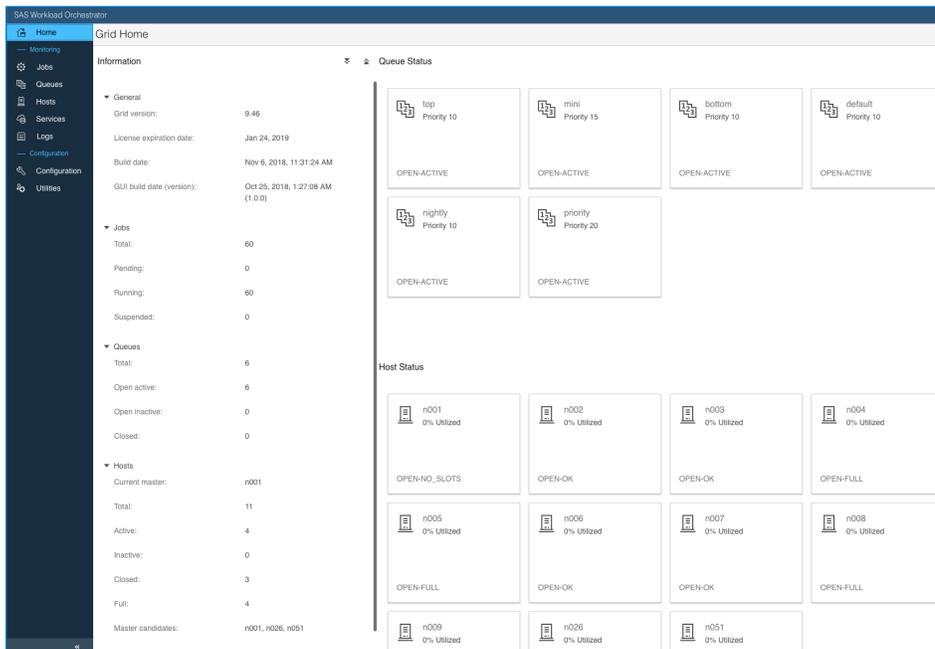- Creates a shared environment for easily and dynamically allocating resources.

Figure 2: The SAS Workload Orchestrator Home area provides an overview of the status of your SAS grid on a day-to-day basis.