SAS® ENTERPRISE INTELLIGENCE PLATFORM REFERENCE CONFIGURATIONS

White Paper
May 2008

Table of Contents

Chapter 1
Introduction

The pursuit of information becomes more aggressive with every passing day, the volume of data continuously swells, and technology becomes more sophisticated. To succeed in today's complex and changing business environment, winning organizations must be able to transform raw information into useful knowledge. They need a comprehensive solution that allows them to gather, access, and use information, with enough strength to match that of their organization itself.

Architecting reliable, highly available, fast, scalable, cost-effective, and flexible business enterprise environments can be a time-consuming task. And time is something that is increasingly in short supply in today's globally competitive market. Information Technology (IT) managers need cost-effective, powerful solutions that can be quickly implemented to provide value and enable employees to turn data into a competitive edge for the business.

Leveraging vast engineering and field experience with benchmarking, architecting, and implementing SAS environments, Sun Microsystems and SAS have created a set of reference configurations for each of the four areas of the SAS® Enterprise Intelligence Platform (Data Integration, Business Intelligence, Analytics, and Intelligence Storage) to help IT managers quickly choose a configuration to meet their needs, thus reducing the risk and deployment times for Enterprise Intelligence (EI) solutions. Additionally, these reference configurations provide IT managers with information to help resolve common IT issues when deploying the SAS Enterprise Intelligence Platform.

Solution

Sun's offerings — software, servers, storage, and service — provide all of the technology requirements (such as performance, scalability, availability, security, systems management, etc.) for complete SAS Enterprise Intelligence Platform implementations.

Sun believes that IT matters. Sun can help IT solve the most pressing business problems by leveraging IT best practices gleaned from over two decades of experience working with top companies worldwide. Sun has vast experience in implementing SAS for EI environments and has developed a series of reference configurations to help reduce the risk and deployment time of implementing EI solutions. Sun can address every aspect of an EI environment in terms of infrastructure. However, the suggested reference configurations in this paper focus on business intelligence. SAS Business Intelligence provides a fully integrated and comprehensive suite of business intelligence software that addresses the needs of IT management and business users.

The entire set of reference configurations for SAS is divided into five papers for convenience and to highlight the specific needs of the different functional areas in the SAS Enterprise Intelligence Platform. Each paper defines reference configurations based on these needs.

- Data Integration and ETL
- Business Intelligence
- Analytics
- Intelligence Storage
- Putting it all Together — SAS Enterprise Intelligence Platform

---

Note – The reference configurations in this paper — for SAS Enterprise Intelligence Platform — as well as the other papers in the series, are suggested starting configurations that should be optimized based on specific requirements and preferences for technology. The configurations are based on generalized workloads. Final solutions need adjustment of CPUs, cores, memory, network interfaces, HBAs, RAID controllers, channels, and/or disk drives to fit the reference configuration to specific requirements. Contact a SAS Account Manager or Sun to customize the selected configuration to meet specific needs. Please note these reference configurations include adequate capacity to support reasonable growth. Additionally, these reference configurations are created with adequate resources to support test and development environments.

Sun and SAS

With a business relationship that is still going strong after 19 years, Sun and SAS have a long, proven track record of delivering open, scalable, and reliable technology solutions.

SAS is the leader in business intelligence and analytical software and services. Enterprises and organizations around the world use SAS software to improve performance through insight from data, resulting in faster, more accurate business decisions, more profitable relationships with customers and suppliers, compliance with governmental regulations, research breakthroughs, and better products and processes.

Sun is a relentless innovator, with a 25-year history of bringing innovative ideas to market with practical results for companies and organizations of all sizes. Sun integrates software, storage, systems, and services to provide complete network computing infrastructures. By investing in research and development, Sun creates products and services that address the complex issues that businesses face today, including increasing demands for network access, bandwidth, and storage — all driven by explosive growth in network participation and sharing. Sun innovates at all levels of the system and partners with market leaders such as SAS to provide value and choice.

Sun's commitment to choice is abundantly clear. With a full line of UltraSPARC®, SPARC64, AMD Opteron™, and Intel® Xeon® systems, all running the Solaris™ Operating System (Solaris OS), Sun can leverage the best technology to handle EI workloads.

With Solaris 10, Sun now guarantees source code compatibility between SPARC and x64 processors, ensuring applications can run across platforms with a simple recompile.

The Sun platform is an excellent match for the compute-intensive demands of advanced SAS Business Intelligence. Combined with SAS Enterprise Intelligence Platform, the Sun platform provides the capability to support hundreds of users building reports to quickly improve planning and decision making at all levels of business.

Together, Sun and SAS provide enterprises and organizations with the industry's leading enterprise intelligence solutions optimized to run on the world's most open and scalable hardware platform and Java™ technology-based environment. As a result, tightly integrated joint solutions enable companies to build closer, more productive relationships with customers, partners, suppliers, and employees, thus increasing revenue and market share while reducing costs.

Chapter 2

SAS® Enterprise Intelligence Platform

The SAS Enterprise Intelligence Platform, illustrated in Figure 1, optimally integrates the individual components of a business intelligence environment into a single, unified system. The result is secured access to data that delivers new insights to drive value within the business. All components and services, including SAS Data Integration, SAS Intelligence Storage, SAS Analytics, and SAS Business Intelligence, are managed from a single point. Data consistency is assured because metadata is stored in a single metadata repository and is shared across all SAS technologies and solutions.

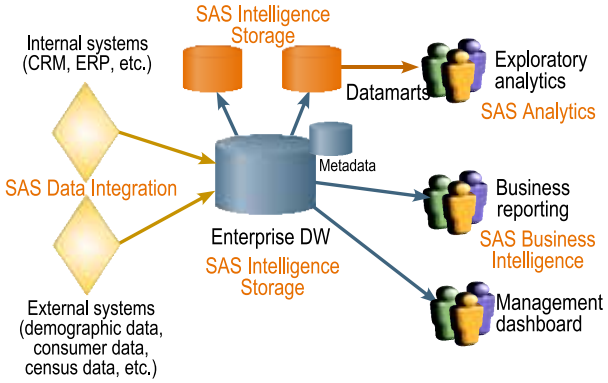

Figure 1. SAS Enterprise Intelligence Platform

SAS® Enterprise Data Integration Server

The SAS Data Integration Server, featuring DataFlux® technology, is a powerful, configurable, and comprehensive solution that can access virtually all data sources and extract, cleanse, transform, conform, aggregate, load and manage data. It supports data warehousing, migration, synchronization and federation initiatives, supports both batch-oriented and real-time master data management solutions, and creates real-time data integration services in support of service-oriented architectures. SAS Data Integration consists of SAS Data Integration Server, SAS Data Integration Studio, SAS Data Quality Solution, and SAS/ACCESS®. SAS Data Integration Server's grid-enabled, load balanced, multithreaded parallel processing architecture makes it an ideal fit for Sun's multi-core, multithreaded platforms.

SAS® Intelligence Storage

SAS Intelligence Storage is a collection of interoperable data stores, including SAS Scalable Performance Data Server® and SAS OLAP Server, designed for large business intelligence applications. As requests to the storage platform increase and the volume

of stored information continues to grow, SAS scales up by spawning multiple threads that balance the workload across processors and storage devices within a single server. SAS can also scale out by pooling multiple servers and storage devices into a single cluster, further distributing and balancing workloads across all available resources.

SAS® Scalable Performance Data Server®

SAS Scalable Performance Data Server is designed to significantly speed up analytics and business intelligence applications and shorten ETL times, irrespective of the amount of data help within the enterprise. SAS Scalable Performance Data Server achieves its scalability through parallel processing and partitioning, hybrid bitmap and B-tree indexing, query optimizations, and eliminating transactional overhead.
To support the fastest access to large volumes of data, SAS Scalable Performance Data Server features a multithreaded I/O engine that launches multiple, independent lightweight processes to search for rows containing specific information in large tables.

SAS Scalable Performance Data Server speeds the processing of large amounts of data by partitioning the data across multiple disks and I/O channels. This enables parallelization of many SAS I/O functions over multiple data partitions. The maximum performance benefit with SAS Scalable Performance Data Server is gained when it is run on a system with multiple CPUs, multiple I/O channels, multiple disks, and a large amount of data that can be partitioned.

SAS® OLAP Server

SAS OLAP Server is a multidimensional data store designed to provide quick access to pre-summarized data generated from vast amounts of detailed data. The central component of SAS OLAP Server is a multithreaded query engine that ensures optimal use of hardware resources. Any query sent to the server is handled by an individual query thread, enabling the server to handle large user communities accessing the server in parallel. OLAP data sources can be spread across multiple file systems, enabling the multithreaded query engine to execute multiple queries in parallel by reading data from multiple locations at once.

SAS® Analytics

SAS has been developing analytics capabilities for nearly three decades and offers an integrated suite of analytics software that is delivered in a single environment.
SAS Analytics are built on an open, standards-based, multithreaded architecture designed to handle large volumes of data. SAS Analytics applications are frequently installed on the same system in the application tier of a multi-tiered architecture.

SAS Analytics solutions offer an integrated environment for predictive analytics and descriptive modeling, data mining, text mining, forecasting, optimization, simulation, experimental design, and more. SAS Analytics solutions provide a range of techniques

and processes to collect, classify, analyze, and interpret data to reveal patterns, anomalies, key variables, and relationships, leading ultimately to new insights for guided decision making.

SAS® Enterprise Miner™

SAS Enterprise Miner streamlines the entire data mining process from data access to model deployment by supporting all necessary tasks within a single, integrated solution. SAS Enterprise Miner 5.2 is designed with a Java client and SAS server architecture that separates the data mining computational server from the user interface. SAS Enterprise Miner provides advanced predictive and descriptive modeling tools and algorithms, including decision trees, neural networks, autoneural networks, memory based reasoning, linear and logistic regression, clustering, associations, time series, and more.

SAS® Forecast Server

SAS Forecast Server generates large quantities of statistically-based, high-quality forecasts quickly and automatically without the need of human intervention unless desired, allowing organizations to plan more effectively for the future. SAS Forecast Server predicts future activity based on time series data. It can detect underlying trends in data as well as adjust predictions for seasonality. For example, a typically requested forecast might be to predict the sales for the coming year or the sales for a coming holiday shopping season.

SAS® Business Intelligence

SAS Business Intelligence delivers a business intelligence platform that integrates data from across the enterprise to provide consistent, accurate information based on a comprehensive view of the business. The two key components of SAS Business Intelligence are SAS Enterprise BI Server and SAS Metadata Server.

SAS® Enterprise BI Server

SAS Enterprise BI Server provides a portal framework that can serve as a central place for surfacing business intelligence ranging from dashboards to full ad-hoc reporting. SAS Enterprise BI Server supports management with a high level view of business and more specialized users with a more in-depth view. It enables uses to discover hidden tidbits of information. It allows them to look at large volumes of data quickly from multiple angles, easily manipulate the data, add new data, modify the view of the data and interact with the data in many ways.

SAS' enterprise level BI application lends itself to a balanced configuration as major server groups such as the SAS server tier or the Web tier can be deployed on one or multiple software partitions or physical systems. Figure 2 shows how the various components of SAS Enterprise Intelligence Platform and SAS Enterprise BI Server can be deployed.

The middle tier provides an execution environment for business intelligence Web applications. I/O is usually not critical in the middle tier. The server tier executes SAS analytical and reporting processes for distributed clients. This server is typically accessed either by desktop clients or by Web applications running in the middle tier. SAS clients provide Web-based and desktop user interfaces to content, appropriate query and reporting interfaces, and business intelligence functionality. The software on the client tier includes Windows applications, Java™ applications, and a Web browser.

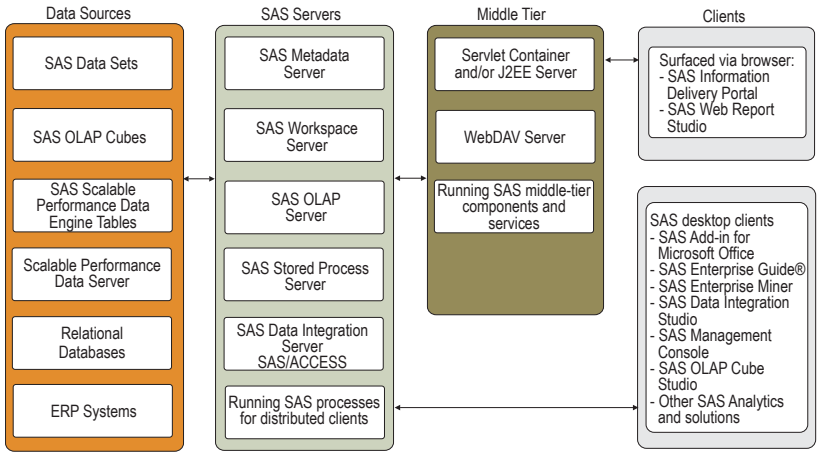

Figure 2. SAS Enterprise Intelligence Platform logical diagram

SAS® Metadata Server

SAS Metadata Server provides an open, central repository for all metadata that is created. The repository stores logical data representations of items such as users, security, information maps, and cubes, thus providing central control over the quality and consistency of data definitions and business rules. The repository also stores information about system resources such as servers, as well as authentication and authorization information. All of the data integration, ETL, and business intelligence tools read and use metadata from the repository and create new metadata as needed. By default, SAS Web applications authenticate users on the metadata server.

Chapter 3
Reference Configurations

These reference configurations for SAS Enterprise Intelligence Platform provide balanced performance for a variety of workloads derived from real-world customer environments. The configurations feature vertical scalability (within a server) options and horizontal scalability where appropriate (across multiple servers).

Balanced Configurations

Each of the reference configurations is designed to eliminate bottlenecks, providing a balance of processing power, throughput, and network capacity/throughput for several different types of workloads based on workload characteristics from all four of the SAS Enterprise Intelligence Platform components. The reference configurations are also designed to provide balanced, modular building blocks that can easily scale by adding CPUs and memory as processing needs increase. The Solaris 10 OS is scalable to over 100 CPUs and Sun systems and storage can be scaled up (vertically) or out (horizontally) without disturbing the business.

A SAS Enterprise Intelligence Platform environment needs to be resource balanced for good response. First, as with many enterprise-level systems, there are several system and user processes competing for resources. A good practice is to dictate that key processes, such as the SAS Metadata Server, be guaranteed resources by either physically or logically isolating the process and resources. Long running jobs that are resource intensive or less response sensitive can be managed so they can access resources that are not required by critical processes, and at the same time ensuring that they do not dominate the system. Proper resource management can yield better response time to business users while also increasing utilization rates of compute resources.

Protecting Investments

Enterprise intelligence environments require a long-term foundation that can encompass new technologies as they evolve without massively disrupting operations. Systems and software based on open standards can more readily absorb change and be redeployed than closed proprietary systems. Sun embodies this principle by maintaining binary compatibility between operating system releases and has been doing so for nearly a decade, enabling existing Solaris applications to run unmodified on Solaris 10. This means that Solaris applications developed ten years ago can run on Solaris 10 unchanged, taking advantage of new and advanced Solaris features. With Solaris 10 the same features and services run on SPARC, AMD Opteron, and Intel Xeon processor-based systems. In fact, Sun guarantees source code compatibility between SPARC and x64 processors, enabling applications to run across platforms with a simple recompile.

Sun's binary compatibility makes it easier to re-purpose systems to other tasks if necessary because the same applications run across the entire Sun product line. Furthermore, Sun storage maintains high levels of heterogeneous interoperability with leading platform and OS vendors, offering investment flexibility and protection now and in the future.

Scalability

Sun offers choice, supporting components of SAS Enterprise Intelligence Platform both vertical and horizontal scaling models with SPARC, AMD Opteron, and Intel Xeon running the Solaris 10 OS, Linux, or Windows.

SAS Enterprise Intelligence Platform middle tier and server tier are frequently installed on the same system in smaller environments. On the Solaris OS, applications can be isolated and guaranteed resources. Therefore, the reference configurations first detail a quantity of resources for a specific workload and then a suggest a Sun system capable of supporting those resources for each tier.

Vertical Scalability

Sun's systems portfolio includes a number of large servers that are ideal for running business intelligence on single servers. Scalability is accomplished by nondisruptively adding CPU and memory into the system, rather than adding entire nodes.

The benefits of implementing a vertically scaled environment on Sun servers are:

- Manageability — In general, fewer systems are easier to manage than many.
- Flexibility — Sun's unique Dynamic System Domains technology enables electrically fault isolated hard partitions that deliver the highest levels of availability and utilization. Combined with Solaris Containers, these capabilities help reduce costs by consolidating multiple servers and mission-critical applications onto a single midrange to high-end server. Each domain is completely protected from hardware or software faults originating in other domains. Each domain uses a separate boot disk with its own instance of the Solaris OS, as well as I/O interfaces to network and disk resources. Resources can be separately added and removed from running domains using Dynamic Reconfiguration, enabling optimal resource utilization and uptime.
- Dynamic capacity — The Dynamic Reconfiguration feature on larger Sun servers allows resources to be dynamically reallocated, or balanced, between domains by enabling a physical or logical restructuring of the hardware components while the system is running and the applications remain available. This high degree of resource flexibility allows the system to be easily scaled in order to provision the resources to meet changing workload demands.

- Consolidation — Consolidation collapses the functionality of multiple servers into a smaller number of systems, reducing costs and management complexity. Consolidation helps to better utilize resources and reduce administrative tasks, thus improving manageability and return on investment. Consolidation can also drive consistency for better data quality and compliance reporting.
- Capacity on demand (COD) — This is an innovative procurement model enabled by Dynamic System Domains. With COD, fully-configured systems are shipped with only a portion of their resources enabled, according to current needs. Additional processors and memory are already installed in the systems and can be quickly enabled with a licensing mechanism.

Horizontal (Grid)

The middle tier is horizontally scalable because of the software clustering features of the application server. An alternative approach to the single server model for the server tier of SAS Enterprise BI Server, SAS Enterprise Miner, and SAS Data Integration is a horizontally scalable, or grid, model. In this environment, scalability is accomplished by adding individual nodes to the grid. The benefits of grid computing are inherent system redundancy through multiple nodes and storage, as well as a performance growth path. As the number of servers increases, the risk and impact of a failed node decreases because there are more servers to take over the load of the failed node. In addition, by adding nodes, the computational power increases, thus increasing performance. The grid model can also be implemented on a single server that is partitioned using domains, VMware, or Solaris Containers.

SAS grid capabilities are enabled by SAS Grid Manager. One of the components of SAS Grid Manager is the Platform Suite for SAS, which includes:

- Process Manager — interfaces to the SAS scheduling capability
- Grid Management Services — provides run-time monitoring and management capability within SAS Management Console
- LSF for SAS — provides mapping and load balancing of SAS processing across the grid

SAS/CONNECT® provides the syntax to distribute SAS job components. When a flow contains a job with SAS/CONNECT syntax to identify parallel subtasks is triggered to run, the job is directed to a grid node by LSF for SAS. As the job executes, it creates parallel tasks to execute on the remaining grid nodes.

Storage Considerations

Business intelligence activities can be surprisingly I/O intensive. Performance of SAS Enterprise Intelligence Platform is closely tied to the performance of the underlying storage configuration. File systems, disk systems, controllers, host bus adapters (HBAs), and channels all need to be considered to eliminate potential I/O bottlenecks.

These issues are covered in the companion document to this paper: *SAS Intelligence Storage Reference Configurations*.

Every time a SAS program runs, it generates temporary work files, which can consume a lot of disk space, particularly for analytics processes. For users this space is defined in the environment variable SASWORK. SAS recommends allocating a minimum of two to three times the size of the source data load files for SASWORK storage per concurrent SAS process. For SAS Enterprise BI Server, SAS recommends allocating a minimum of 16 times the report size for SASWORK storage. SASWORK can be very I/O intensive with both sequential and random access competing for data, driving the need for a balanced storage configuration that provides performance for both types of I/O patterns.

It is important to allocate enough storage for SASUSER for SAS users. For SAS Enterprise Guide users that might require space, SAS recommends allocating 1 to 10 times the starting data set size for each named user. It is also necessary to allocate space for the report cache, which should be assigned a different volume than SASWORK. SAS recommends 1.5 times the report size for each concurrent user for report caching. In general, this storage is maintained on a central data store as described in the *SAS Intelligence Storage Reference Configurations* paper.

Sun's broad portfolio of modular and system storage with 3 Gb/sec. Serial-Adaptive-SCSI (SAS) and 4 Gb/sec. Fibre Channel (FC) architecture and drives enables optimal configurations of bandwidth, I/O rate, capacity, and cost for horizontal and vertical storage architectures. In general, SAS recommends using numerous small drives rather than fewer large drives in order to increase performance. In all cases, using 73 GB 15K RPM drives provides a solution with sufficient capacity, bandwidth, and I/O rate to meet the performance needs of the workloads in this paper. In the event that a system needs greater capacity for a given workload, 146 GB or 300 GB drives can be used instead of the 73 GB drives. For smaller, budget conscious IT shops, Sun's Serial-Adaptive-SCSI (SAS) arrays can be utilized to help keep costs down. In general, network attached storage (NAS) is not recommended for SAS environments.

Configuration Examples

Workloads

These reference configurations are designed to support small, medium, and large workloads observed by SAS and Sun engineers who work closely with many customers using Sun and SAS for business intelligence. The configuration examples utilize vertical scalability as well as different CPU architectures for optimum choice and flexibility. Contact SAS or a local Sun representative for information on other configurations and to customize these starting configurations to specific environments.

Characteristics of SAS® Enterprise BI Server Workloads

The characteristics for the workloads used in these configurations are:

- Raw data — the total size of the raw data to be reported on.
- SAS Enterprise Guide users — the number of SAS Enterprise Guide users.
- OLAP cube users — the number of users accessing OLAP cubes through SAS Information Delivery Portal.
- Static report users — the number of users generating static reports through SAS Information Delivery Portal.
- Dynamic report users — the number of users generating dynamic reports through SAS Web Report Studio.
- Dashboard users — the number of SAS BI Dashboard users.
- Concurrent users — the total number of concurrent users.

SAS® Data Integration Workloads

The characteristics for the workloads used in these configurations are:

- Amount of data — the quantity of data to be processed in a specific time. For this paper, small volume is up to 5 GB, medium is up to 100 GB, and large is over 100 GB.
- Number of concurrent sessions — the parallel capability of the workflow.
- Complexity of transformations — ETL/DI flows are different for every environment and some transformations are more complex than others. For this reference configuration paper, ETL tasks are divided into simple, complex, and very complex.
  - Simple — such as index generation on one column or copy dataset.
  - Complex — data steps such as expression, hash lookup, and combined hash lookup and expressions.
  - Very complex — such as SCD (slowly changing dimension), DQMatch (SAS Data Quality Server tool for identifying potential matches so the data can be cleaned) on address, and sort on match code. These types of transformations require more CPU capacity.
- Frequency — the frequency of the data integration operation, e.g., daily, weekly, monthly.
- Time window — time constraint to complete. For example, a nightly job that must be completed in 2 hours. Service level agreements can dramatically affect the system more than any other workload data point.

SAS® Enterprise Miner™ Workloads

The characteristics for the workloads used in these configurations are:

- Data volume — the size of the data set to be analyzed.
- Number of variables — the number of columns (or different characteristics) to be examined.

- Number of concurrent users — the number of users working on the system at the same time.

SAS® Forecast Server Workloads

The characteristics for the workloads used in these configurations are:

Model selections per hour — the number of model selections that must be performed per hour to meet a specific SLA.

- Model fitting and forecasting per hour — the number of model fittings and forecastings that must be performed per hour to meet a specific SLA.
- Time series for model selection — the number of time series to be processed for model selection, fitting, and forecasting.
- Table size — the amount of workspace required to perform model selection, fitting, and forecasting. In general, model fitting and forecasting require twice as much workspace as model selection. The number of time series in the table is also important. The number of time series is the number of items to be forecasted multiplied by the number of BY groups, such as DEPARTMENT, STORE, DISTRICT, STATE, REGION, etc.

SAS® Scalable Performance Data Server® Workloads

The characteristics for the workloads used in these configurations are:

- Data load — amount of data to load into a star schema and create the index.
- Window — time in which the data load and index create must complete. This time varies widely depending on the complexity of the index and the star schema.

SAS® OLAP Server Workloads

The characteristics for the workloads used in these configurations are:

- Cube size in GB — size of the cube to be built. Remember to multiply this number by five in order to calculate required storage
- Window — time in which the cube build must complete
- Concurrent users — number of concurrent users querying the cube

The following configurations provide base requirements in terms of CPU cores, memory, and I/O. These requirements can then be used to size a single system, a grid of systems, or can be used to help size a single system that supports more than one SAS function.

Example One — Small

In this example, a manufacturing and retail company would like to use SAS Business Intelligence to analyze 25GB of raw data. The claims processing department would have 20 dynamic reporting and 10 static reporting users. They would also like to have 4 SAS Enterprise Guide users and 20 OLAP cube users work on understanding the root causes of claims in order to reduce their warranty cost.

The company's direct marketing division needs to analyze a 500 MB data volume with 100 variables to optimize direct marketing campaigns and create targeted customer relationship management projects such as predicting cancellations of cross-selling campaigns. The system needs to support three analysts working concurrently in SAS Enterprise Miner.

Their finance department is planning to use SAS Forecast Server to conduct time series studies of invoice collections, as well as forecast future collection behavior and results in order to increase the efficiency of their collections structure. They predict a workload of 100,000 model selections per hour, 1 million model fittings and forecasting per hour, 50,000 time series, and a table size of .5 GB.

On the retail side, this company needs a data integration system than can extract approximately 3 GB of data from their point-of-sale database per week, perform complex transformations on that data, and load it into a data warehouse in a 2 hour window during business hours. The data is only coming from one source and is well formatted, so it does not require any data quality transformations. The data is partitionable and can be divided into 10 concurrent sessions.

They also need to create a 10 GB OLAP cube in order to study statistics in a busy cities. The cube needs to be built in less than half an hour and the system needs to be capable of supporting five concurrent users querying the cube.

A good starting configuration for these workloads is 12 cores, 48 GB of memory, four HBA ports for data integration and access to the data warehouse, and four Ethernet ports to support network traffic to and from external systems, as well as provide load balancing and redundancy. A minimum of 750 GB storage is required to support SASWORK, data integration, the data warehouse, and OLAP cubes. Two CPU cores and 8 GB of memory should be isolated for the metadata server. The mid-tier components should also be isolated into a separate container so that it can be migrated to a Web farm if and when desired. This suggested configuration can also run a test and development environment in separate containers, and has adequate compute resources for future growth.

For this SAS Enterprise Intelligence Platform configuration, a single, powerful server such as the Sun SPARC Enterprise M5000 system provides the required capacity and throughput, with room to grow. The Sun SPARC Enterprise M5000 server supports up to

eight SPARC64 VI dual -core processors, 256 GB of memory, up to four dynamic domains, and up to 10 I/O slots. One domain should be configured in order to isolate and guarantee resources for SAS Metadata Server. Two Solaris Containers can be configured to provide failover within the domain for higher availability. A second domain can be configured to house the other SAS applications. Each application can be isolated in its own container. Solaris Resource Manager, a feature of Solaris Containers, can be used to dynamically allocate resources to applications such as SAS Data Integration, when other applications are not using their assigned resources. In this way, the system can be configured with fewer CPUs/memory than a configuration with a separate system for each application, thus greatly increasing resource utilization.

In addition, Dynamic Reconfiguration allows administrators to hot swap CPU/memory boards, dramatically improving availability and uptime. The system also supports Capacity on Demand in order to reserve hot spare system boards that can be activated when needed.

The Sun StorageTek 2540 array is suitable for this workload, supporting up to four host-side 4 Gb/sec. FC interfaces, dual active/active RAID controllers, and up to 36 Serial-Adaptive-SCSI (SAS) drives for total of up to 10.8 TB. The array also supports snapshots, storage domains, and up to 15 global hot spares, providing data protection, consolidation, and redundancy for SAS Scalable Performance Data Server and other SAS services. In this case, 750 GB configured with Solaris ZFS is recommended for performance and increased data protection. The entire configuration is depicted in Figure 3.

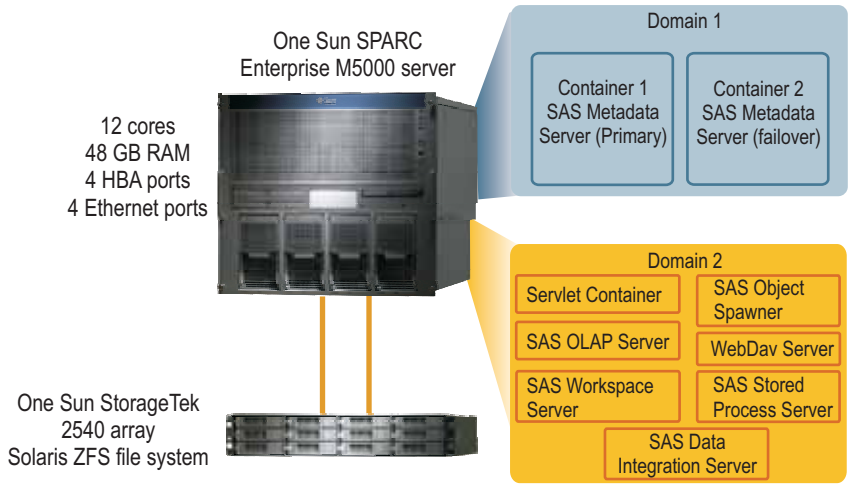

Figure 3. A sample configuration to support a small SAS Enterprise Intelligence Platform implementation.

Example Two — Medium

In this example, a health insurance company wants to understand risks and select the right customers accordingly. They have approximately 200 GB of raw data that needs to be processed. For analyzing various risk factors, there are 12 SAS Enterprise Guide users and 75 OLAP cube users. For analyzing the current customer accounts, there are 50 dynamic report users and 25 static report users.

They would like to mine data about customers' buying patterns and behavior in order to gain a greater understanding of customer motivations to help reduce fraud, anticipate resource demand, and curb customer attrition. Their current data volume is approximately 3 GB with 200 variables, and the system needs to support eight concurrent users. The company plans to add more analysts in the next several years, so the system must be easily scalable.

The company needs to add 50 GB of data to their data warehouse every week in a 4 hour window. The data comes from several sources including financial, marketing systems, and claims systems, and requires complex transformations before it can be loaded into the data warehouse or OLAP cubes. The ETL process can be partitioned into 25 sessions for better performance and load balancing, if it is running at night with dedicated resources.

With their current technology, it takes 10 hours to process a day's transactions of 250 GB from an external OLTP database into a state where the data can be queried with analytics tools. In order to increase their ability to fight fraud, the company needs to decrease processing time down to 2.5 hours. The company also needs to build a 50 GB cube on a weekly basis in three hours in order to manage and track the progress of a new marketing campaigns. The system needs to support queries from 40 concurrent users after the cube is created.

A good starting configuration for this SAS Enterprise Intelligence Platform with SAS Data Integration, SAS Business Intelligence, SAS OLAP Server, SAS Scalable Performance Data Server, SAS Enterprise Miner, and SAS Metadata Server is 24 CPU cores, 96 GB of memory, 8 HBA ports to connect to SASWORK and OLAP, and eight Ethernet ports to provide access to external servers, as well as provide load balancing and redundancy. Four CPU cores and 32 GB of memory should be isolated for the SAS Metadata Server.

This configuration example illustrates how the workload can be separated onto three servers to take advantage of x64 technology as well as the performance of the Sun Fire™ X4500 server for SAS Scalable Performance Data Server.

Both the middle and server tier for SAS Business Intelligence can be handled by a Sun Fire X4600 M2 server configured with four dual-core processors and 36 GB of memory. Powered by single- or dual-core AMD Opteron processors, and up to 256 GB of memory,

the 4- to 16-way Sun Fire X4600 M2 server provides cost-effective performance and scalability, and can also be upgraded to next-generation quad-core processor servers. The Sun Fire X4600 M2 server includes four Gigabit Ethernet ports for load balancing and redundancy when retrieving data from external data sources and communicating with the client, as well as eight PCI expansion slots and remote management features. The Sun Fire X4600 M2 supports virtualization technologies such as XEN, VMware, Solaris Containers, and Microsoft Virtualization, easily and effectively hosting and managing many virtual machines within the server. Allocating compute resources into these virtual machines can quickly and easily maximize utilization. The server should include six 4 Gb/sec. FC HBA ports to provide redundant connections and paths to SAN attached storage arrays that house SASWORK and OLAP cubes.

The Sun Fire X4500 server is a good fit for the SAS Scalable Performance Data Server workload for this example, providing a single platform for both applications and data, with enterprise server reliability features and extremely high data throughput rates. An integrated solution for applications and data, the Sun Fire X4500 server combines the functions of a high-performance, four-way x64 server, network fabric and switch, and storage into a single integrated system. This combination of multiple CPUs, I/O channels, disks, and support for up to 24 TB makes the Sun Fire X4500 server an ideal platform for SAS Scalable Performance Data Server. Powered by two dual-core AMD Opteron processors, 16 GB of memory, six eight-port SATA controllers, and Solaris ZFS, the server provides the processing power and I/O throughput to handle a load of 250 GB. For better performance, the system should be configured with as many small spindles as possible and Solaris ZFS should be configured to stripe across a single volume of mirrored disks. The system include four Gigabit Ethernet ports, which is more than adequate network bandwidth for this workload.

The remainder of the applications, SAS Data Integration, SAS Analytics, SAS OLAP Server, and SAS Metadata Server can be housed on a single server and separated through domains and containers. The Sun SPARC Enterprise M8000 server is an ideal server for the job. This suggested configuration can also run a test and development environment in separate containers, and has adequate compute resources for future growth.

Designed for demanding applications that require mission-critical, 24x7 performance, the high-end Sun SPARC Enterprise M8000 server delivers extreme reliability, availability, and serviceability, and is therefore ideal for this workload. Built on the advanced SPARC64 VI processor, the Sun SPARC Enterprise M8000 server is optimized for enterprise class applications such as SAS OLAP Server and SAS Analytics. Powered by up to 16 dual-core processors and up to 512 GB of memory, the system is also scalable up to 16 domains and 112 PCIe and PCI-X slots (with external I/O expansion unit), supported by a system I/O bus capable of providing 61 GB/sec. peak throughput. For this workload, the system is configured with 12 dual-core processors, 128 GB of

memory, four HBA ports, and a minimum of four Gigabit Ethernet ports, leaving room for growth in capacity and I/O throughput.

SAS Data Integration and SAS Analytics should be configured in the same domain in separate containers. Separate containers provide security and higher availability while also enabling SAS Data Integration to use the resources of SAS Analytics when those resources are not in use, such as for a few hours at night. This configuration obviates the need for separate resources for SAS Data Integration processes as long as SAS Analytics is not required on a 24 hour basis.

One domain should be configured in order to isolate and guarantee resources for SAS Metadata Server. Two Solaris Containers can be configured to provide failover within the domain for higher availability.

SAS OLAP Server can be configured in a third domain to isolate it from the other applications and to guarantee performance. The domain should contain eight CPU cores and 64 GB of memory. Four 4 Gb/sec. FC HBA ports can provide the throughput, redundancy, and load balancing required to access the OLAP storage system. Two Gigabit Ethernet ports are sufficient for interacting with other systems in the SAS environment, such as BI applications that need to query the OLAP cubes. A minimum of 2.5 TB is required to support a cube size of 500 GB.

The Sun StorageTek™ 2540 array is a cost-effective choice for the data integration and SASWORK workloads, supporting up to four host-side 4 Gb/sec. FC interfaces, dual active/active RAID controllers, and up to 36 Serial-Adaptive-SCSI (SAS) drives for total of up to 10.8 TB. The array also supports snapshots, storage domains, and up to 15 global hot spares, providing data protection, consolidation, and redundancy for SAS Scalable Performance Data Server and other SAS services. In this case, a minimum of 250 GB configured with Solaris ZFS is recommended for performance and increased data protection.

Two Sun StorageTek 6140 arrays configured with Solaris ZFS provide the required throughput to handle the I/O intensive OLAP workload. The arrays support dual FC RAID controllers, up to 56 TB of raw capacity, up to eight switched 4 Gb/sec. host interfaces, up to 15 global hot spares, snapshot, and storage domains, providing performance and protection for large OLAP cubes. Each array should be configured with as many spindles of small disks as is economically feasible to provide optimal performance.

The entire configuration is shown in Figure 4.

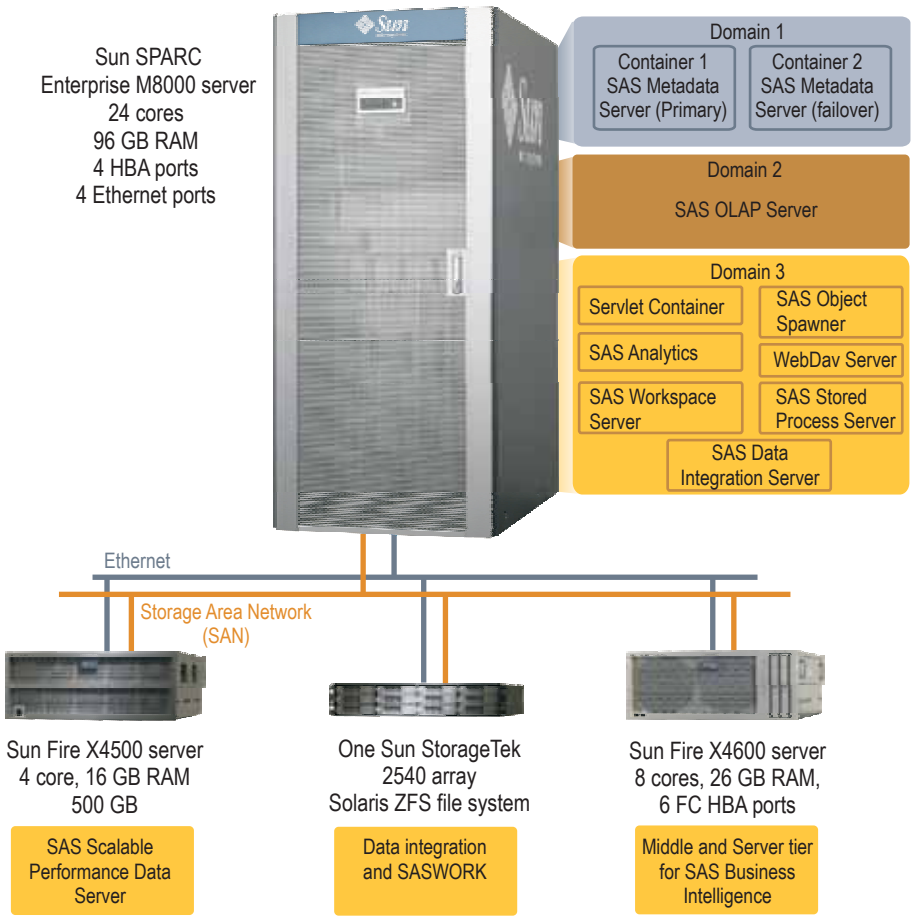

Figure 4. A sample configuration to support a medium SAS Enterprise Intelligence Platform implementation.

Example Three — Large

In this example, a large financial firm would like to grow their customer base by analyzing current customer behavior and trading activities. They have approximately 500 GB of raw data that can be processed. There are 18 SAS Enterprise Guide users and 140 OLAP cube users who can analyze this data. Finally, reports can be generated that show the name, portfolio, and trade history of current customers. They have approximately 140 dynamic report users and 100 static report users can generate these reports.

They are also interested in applying data mining techniques to manage the profitability of customer portfolios to stay ahead of the competition. They need a system that can handle a 50 GB data volume with 400 variables and 20 concurrent users. They also predict a workload of 750,000 model selections per hour, 2 million model fittings and forecastings per hour, 3 million time series, and a table size of 25 GB for forecasting.

They need to update their data warehouse with 1 TB of data every day in a 16 hour window. The data comes from a number of trading systems, and required very complex transformation including data quality transformations, which can be extremely CPU and memory intensive. The workflow can be divided and they would like to run 250 concurrent sessions during non-trading hours. The data load for the 16 hour requirement is 250 GB per hour.

Finally, they need to build a 120 GB cube on a daily basis in order to track daily trading activity for a business unit. The cube needs to be built in less than six hours and the resulting cube and system need to support 100 analysts.

A starting configuration for this very demanding workload is 72 CPU cores, and 432 GB of memory, 15 HBA ports, and six Ethernet ports for load balancing and redundancy. For workloads of this size, a second metadata server is usually configured to provide high availability. This suggested configuration can also run a test and development environment in separate containers, and has adequate compute resources for future growth.

The SAS Metadata Server contains metadata about everything and everyone in the SAS Enterprise Intelligence Platform, so performance is critical, especially for this demanding workload. The perfect system for this workload are the Sun SPARC Enterprise 5220 server, the World's first 64-thread, general purpose server powered by the ground-breaking UltraSPARC T2 *System on a Chip* processor. It integrates 8 cores, 64 threads, 8 Floating Point Units, cryptographic acceleration, dual 10Gb Ethernet ports, and PCI-E I/O directly onto the processor. With dramatically increased floating point capabilities on the chip, UltraSPARC T2 processor delivers up to 35x the performance of the first-generation UltraSPARC T1 processor when running floating point code. And, its on-chip dual 10 Gigabit Ethernet (10 GbE) technology delivers up to 10x more network bandwidth than the competition. Two systems, each with 64 GB of memory should be configured with Solaris Cluster software to provide high availability for this very critical application.

All of the applications for this example, except the metadata server, require over 10 CPU cores each, so it makes sense to house them all on a single server in one domain, thus through containers and resource management, each application is guaranteed resources but can also use the resources of the other applications when those resources are not in use.

By housing all of the applications on a single domain, that requirement can be decreased to 64 cores because there is some cross-over in users for SAS Business Intelligence and SAS Analytics, and because SAS Data Integration Server can use some resources from the other applications when they are under less pressure or idle.

The Sun SPARC Enterprise M9000-64 system can handle the requirement for 64 CPU cores with room to expand in the future. The Sun SPARC Enterprise M9000-64 server offers the highest availability, highest absolute performance, highest scalability, and the most sophisticated control of resources in Sun Microsystems' extensive server product line. It supports up to 64 dual core SPARC64 VI 64-bit Chip Multithreading (CMT) processors (up to 128 cores), 2 TB of memory, 128 hot-swappable PCI-Express (PCIe) I/O slots, up to 64 internal disks, support for external I/O, a new high performance interconnect capable of up to 304.2 GB/sec., support for up to 24 domains, and thousands of software applications for the Solaris Operating System. Exhibiting 7.5 times greater system bandwidth than previous generations and scalability in every dimension, the Sun SPARC Enterprise M9000-64 server sets a new standard for performance and configuration flexibility in large symmetric multiprocessing (SMP) platforms. To support this environment, the system should be configured with 96 CPU cores, 576 GB of memory, 28 HBA ports, and six Ethernet ports to support access to and from external servers and clients.

Five Sun StorageTek 6140 arrays configured with Solaris ZFS provides a good fit for SAS Scalable Performance Data Server. The arrays supports dual FC RAID controllers, up to 56 TB of raw capacity, up to eight switched 4 Gb/sec. host interfaces, up to 15 global hot spares, snapshot, and storage domains, providing both the performance and data protection necessary for business-critical data. For better performance, the system should be configured with as many small spindles as possible and Solaris ZFS should be configured to stripe across a single volume of mirrored disks, for a total of 6 TB of storage.

Five additional Sun StorageTek 6140 arrays configured with Solaris ZFS provide the required throughput to handle this I/O intensive OLAP workload. Two more Sun StorageTek 6140 arrays provide storage for data integration on (one array) as well as SASWORK for SAS Analytics and SAS Business Intelligence (the second array).

The entire configuration is illustrated in Figure 5.

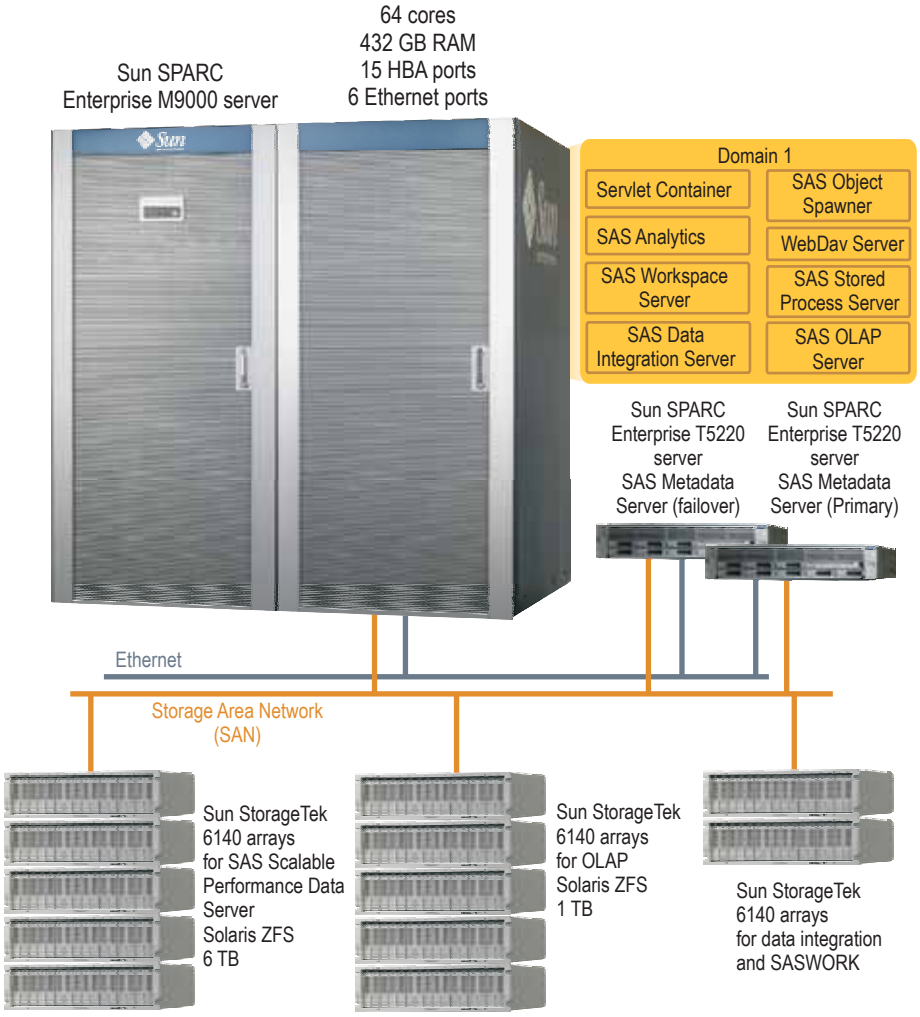

Figure 5. A sample configuration to support a large SAS Enterprise Intelligence Platform implementation.

Other Things to Consider

The configuration examples above are provided as a starting point to help IT architects and managers size data integration requirements. No two environments are exactly the same and there are a number of variables that need to be considered in order to provided optimal data integration performance for individual organizations.

For example, the source data in many companies is scattered through the enterprise and stored in inconsistent formats. The ability of the data source (DB2/UDB, MS/SQL, Oracle, Sybase) to support parallel data extract processes can influence the number of CPUs that the data integration server can effectively utilize. The type of storage

(SAN, direct attach, NAS) on the source data system can cause I/O bottlenecks that affect performance if not properly configured to support SLAs. The data refresh rate (nightly, weekly, monthly, etc.), volume of data, and anticipated annual growth rate of the source data need to be considered when sizing systems. Finally, the complexity of the transactions can greatly impact performance as this is a very subjective descriptor that can only be measured through actual testing.

Consolidating with Solaris Containers

Solaris Containers are a unique, multi-core aware feature in the Solaris 10 OS that enable consolidation and better asset utilization, which leads to reduced complexity in the data center. A Solaris Container is a complete runtime environment for applications that creates very little administrative overhead and has miniscule impact on system performance, unlike other virtualization technologies. One instance of the Solaris OS enables nearly limitless partitioning, with a theoretical limit of 8000 containers per instance. Solaris 10 Resource Manager and Solaris Zones software partitioning technology are both parts of the container. Zones allow application components to be isolated from one another even though the zones share a single instance of the Solaris Operating System. The Solaris Resource Manager provides the capability to dynamically allocate the quantity of CPU and memory resources that a workload in a zone receives.

This is particularly important for data integration processes, because it is often assumed that these processes run at night using resources that are otherwise dedicated to other functions during the day. The Solaris Resource Manager can be configured to help ensure that data integration processes are guaranteed the resources required to finish processing within specified time limits.

Containers can be used to consolidate all of the SAS tiers and/or environments (including test, development, and production) onto a single server or servers, while isolating the applications and providing guaranteed resources for each application. For example, the SAS Data Integration Server can run in a container on the same server as SAS Business Intelligence and/or SAS Analytics. Resource Management can be configured to enable the SAS Data Integration Server to use processing capacity if the other containers are not using it, which might be for a few hours in the middle of the night. In addition, containers could be used to implement grid on a single system.

Reliability

To make data integration processes more available and reliable, Sun systems are equipped with redundant, hot-swappable components. The Solaris 10 OS includes other technologies such as self-healing, automatic reconfiguration around failed components, automatic restart of services, multipathing to disks and networks, and failover mechanisms:

- Solaris Predictive Self Healing — Solaris 10 can detect failing system components — including CPUs, memory, and I/O — and take them offline without requiring the application to go offline.
- Multipathing to disks — Implementing higher levels of reliability and availability requires redundant host connectivity to storage devices. The Solaris FC and Storage Multipathing software, incorporated into the Solaris 10 OS, manages the failure of storage paths while maintaining host I/O connectivity through available secondary paths.
- IP Multipathing (IPMP) — A feature in the Solaris Operating System that enables IP fail-over and IP link aggregation. It helps balance network workloads and provides higher network availability by managing IP link failures on the Sun servers.
- Service Management Facility — A Solaris 10 features that provides a mechanism for failed services to restart automatically in dependency order. Service Management Facility can be configured to automatically restart SAS applications such as SAS Data Integration on the same node if they should fail. SAS does not retain state across failures, so jobs need to be restarted after the application is restarted.
- Solaris Dynamic Tracing (DTrace) — SAS includes a lot of integrated products and the ability to access almost any data source, which can make it difficult to track down the source of performance bottlenecks. DTrace provides dynamic instrumentation and tracing for both application and kernel activities — even allowing tracing of application components running in a Java Virtual Machine (JVM).
  DTrace enables developers and administrators to explore the entire system to understand how it works, track down performance problems across many layers of software, or locate the cause of aberrant behavior. Tracing is accomplished by dynamically modifying the operating system kernel to record additional data at locations of interest. DTrace has no impact on system performance when not in use, making it particularly effective for monitoring and analyzing production systems.

High Availability

Sun systems are highly reliable, however, unforeseen outages might still occur.
SAS Grid Manager provides higher availability through redundant nodes. However, the grid software (SAS/CONNECT, Platform, SAS Grid Manager) can present a single point of failure. Highly available clusters provide nearly continuous access to data and applications by keeping the cluster running through failures that would normally bring down a single server system. No single failure — hardware, software, or network — can cause a cluster to fail.

Solaris Cluster software's high-availability framework detects a node failure quickly and migrates the application or service to another node that runs in an identical environment.

Solaris Cluster offers built-in support for the broadest range of commercial and open source applications in the industry and Web tier solutions. An easy-to-use toolkit is provided to guide developers or administrators through creating custom agents.

In single system environments, a Solaris Cluster can be created to provide SAS application failover to another node in the cluster, if the application cannot be successfully restarted on its primary node. Again, SAS does not retain state across failures, so jobs need to be restarted after the application restarts on a different node.

For greater availability across a greater distance, cluster nodes can be separated by up to 400 kilometers using Solaris Cluster software and DWDM (dense wave division multiplexing) technology to provide application service continuity in the event of a catastrophic failure.

Solaris Cluster Geographic Edition software provides the highest level of availability for data and application services in the event of a large disaster such as the United States' East Coast power failure of 2003. If a primary cluster fails, Solaris Cluster Geographic Edition software enables IT operators to start up services with replicated data on the secondary cluster. It can also be used when companies need their data and applications to follow the sun for global operations.

Protecting Data

The demands of a global economy mean backup and recovery times are very limited, requiring technologies that minimize backup windows and optimize data availability. Technologies such as snapshots, mirroring, and replication can help. Snapshots take a point-in-time snapshot of data that only keeps references to data after it has changed on the source. Snapshots take less space than a complete copy and can be used to back up and restore data quickly. Mirroring makes a full copy of data. The mirror can then be separated and used for other purposes or for backup. Replication transmits changes as they occur to another storage system for disaster recovery purposes. The replicated data can also be used to create remote tape backups. Finally, storing data on near-line tape enables faster access for compliance and can still allow access to the data for queries when necessary. Sun offers all of these solutions and can help implement them in data integration environments.

Chapter 4
Conclusion

In today's global economy IT managers need the ability to quickly implement robust, scalable data integration systems that have the power to rapidly integrate data from many sources to help employees make smarter decisions to stay competitive. Leveraging vast engineering and field experience architecting the SAS Enterprise Intelligence Platform, Sun and SAS have created a set of reference configurations for small, medium, and large environments to reduce the risk and deployment times for data integration solutions. Table 1 summarizes the reference configurations.

Table 1. SAS Enterprise Intelligence Platform reference configuration summary.

Workload	Servers	Storage
Small	• Sun SPARC Enterprise M5000 server, 12 cores, 48 GB RAM, 4 HBA ports, 4 Enet ports	• Sun StorageTek 2540 array
Medium	• BI — Sun Fire X4600 M2 server, 8 cores, 36 GB RAM, 6 HBA ports, 4 Enet ports • Sun SPARC Enterprise M8000 server, 24 cores, 128 GB RAM, 4 HBA ports, 4 Enet ports • Warehouse — Sun Fire X4500 server	• Sun StorageTek 2540 array for SASWORK and data integration • 2x Sun StorageTek 6140 arrays for OLAP
Large	• Metadata — 2x Sun SPARC Enterprise T5220 servers, 64 GB RAM, 2 HBA ports each • Sun SPARC Enterprise M9000-64 server, 96cores, 576 GB RAM, 28 HBA ports, 6Enet ports	• 12x Sun StorageTek 6140 arrays

To start, use the reference configurations to get an idea of the systems and storage required for the amount of data and workload to be supported in the business intelligence environments. To integrate the solution into an entire business intelligence solution, Sun and SAS can help with Sun Client Solutions for Business Intelligence, Sun[SM] Solution Centers, and the SAS Enterprise Excellence Center.

Sun realizes that architecting an end-to-end infrastructure is quite a challenge and believes the people who intimately understand the technology are the best people to determine specific needs and design the right solution. Toward that end, the consultants in Sun Client Solutions provide a business-focused, architecture driven approach that uniquely satisfies business intelligence deployment challenges. Sun's business intelligence solutions include requirements gathering, architecture design, migration, and performance tuning, along with a wide variety of applications for the Solaris OS.

Sun Solution Centers can help take the guesswork out of implementing Enterprise Intelligence solutions. IT staff can try out concepts, size systems, test performance and scalability, and see working prototypes (proof-of-concept) on Sun's equipment — all at no charge. They can leverage the combined expertise of Sun business intelligence specialists and third-party vendors.

References

Sun Microsystems posts product information in the form of data sheets, specifications, and white papers on its Web site at *www.sun.com*.

For information on SAS and Sun and to access the other reference configurations in this series, visit *www.sun.com/sas*.

To contact the Sun Solution Center for SAS Competency, select the Contact tab from the following Web site: *www.sun.com/third-party/global/sas/sas-cc.html.*

For more information on the SAS Enterprise Intelligence Platform see: *www.sas.com*.

Sun Microsystems, Inc. 4150 Network Circle, Santa Clara, CA 95054 USA **Phone** 1-650-960-1300 or 1-800-555-9SUN (9786) **Web** sun.com

© 2008 Sun Microsystems, Inc. All rights reserved. © 2006-2007 Sun Microsystems, Inc. Sun, Sun Microsystems, the Sun logo, Java, Solaris, Sun Fire, and Sun StorageTek are trademarks or registered trademarks of Sun Microsystems, Inc. in the U.S. and other countries. All SPARC trademarks are used under license and are trademarks or registered trademarks of SPARC International, Inc. in the U.S. and other countries. Products bearing SPARC trademarks are based upon architecture developed by Sun Microsystems, Inc. AMD Opteron and the AMD Opteron logo are a trademarks or registered trademarks of Advanced Micro Devices, Inc. Intel(R) Xeon(R) is a trademark or registered trademark of Intel Corporation or its subsidiaries in the United States and other countries. Information subject to change without notice. Printed in USA 05/08