Paper SAS072-2014

SAS in the Enterprise
A Primer on SAS® Architecture for IT

Gary Spakes, SAS Institute

ABSTRACT

How does the SAS® server architecture fit within your IT infrastructure? What functional aspects does the architecture support? This session helps attendees understand the logical server topology of the SAS technology stack:

- resource and process management
- in-memory architecture
- in-database processing

The session also discusses process flow from data acquisition through analytical information to visual insight. IT architects, data administrators, and IT managers from all industries should leave with an understanding of how SAS has evolved to better fit into the IT enterprise and to help IT's internal customers make better decisions.

INTRODUCTION

When it comes to IT architecture and the data center, SAS traditionally "does not compute." But it should. A solid SAS enterprise strategy will provide consumers with consistency, dependability, reliability, availability, and functionality while allowing IT to service the consumer base and administer the system. With the SAS platform, IT can address scalability, high availability, maintainability, usability, and productivity, which are the typical core competencies of the data center.

This paper covers three main topic areas—security, performance, and server topology—while touching on typical "-ability" words.

SECURITY

SAS security is predicated on four tenets:

- Identification – Who are you?
- Authentication – Prove it.
- Authorization – Now that you have proved it, what are you allowed to see or do?
- Auditing – What are you doing? What did you do?

IDENTIFICATION

If users are not properly identified, then the computer system is potentially vulnerable to access by unauthorized users. Commonly called a "user ID," a person's computer system "identity" is used as a non-confidential, auditable representation. In the SAS world, identities for people and processes are stored in the SAS Metadata Server. A person cannot leverage SAS in the enterprise without a metadata identity. A standard SAS enterprise deployment uses "host-based identification," which means that SAS leverages the computer system's internal identification mechanisms. SAS can also leverage common identification protocols such as LDAP and Microsoft Active Directory, but this approach requires some configuration.

AUTHENTICATION

Authentication involves the verification of an identity. In other words, "prove who you say you are." The typical mechanism is a password or PIN. As previously mentioned,SAS leverages "host-based authentication," LDAP, and Active Directory for verifying the user ID and password combination. SAS can also use web authentication, single-signon, and Kerberos tickets.

AUTHORIZATION

Enterprise SAS provides an authorization layer that augments the security mechanisms configured in the host environment and other systems. The SAS authorization mechanisms are used to manage access to almost any system object (for example, reports, servers, applications, database objects, Hadoop data, and SAS libraries).

Permissions can be set at the following levels of granularity:

- Baseline controls. Called repository-level controls, these are managed from the default access control template (ACT). All enterprise SAS users are restricted by these controls.
- Object-level controls. These controls manage access to a specific object such as a report, a table, a column, or a folder. These controls can be defined individually (as explicit settings) or by creating an ACT.
- Fine-grained controls affect access to subsets of data within a resource. To establish fine-grained controls, constraints called permission conditions are added.

AUDITING

Enterprise SAS uses a standard logging mechanism for all of its components. The logs support problem diagnosis and resolution, performance and capacity management, and auditing and regulatory compliance. Logged events can be categorized in a hierarchical naming system.

PERFORMANCE

There are many facets of good computer performance. The following are some examples (but certainly not an exhaustive list) of performance-related questions: How fast is the processor? How much memory does the computer have? How fast does it read and write data? But the most relevant question is: How fast will I get my answer? Enterprise SAS focuses on the most relevant question. The enterprise SAS architecture processes work in three complementary ways: by managing resources and workload; by dealing with all the data, quickly; and by operationalizing SAS with the data repositories.

RESOURCE AND WORKLOAD MANAGEMENT

The Enterprise SAS component called SAS Grid Computing enables the creation of a managed, shared environment in which analytic programs can quickly process large volumes of data by using dynamic, resource-based load balancing. SAS jobs can be split into smaller units of work, and each piece can run in parallel across multiple computers using shared physical storage. Using these capabilities, IT can build and manage a lower-cost, flexible infrastructure that can scale to meet rapidly changing computing requirements.

Central administration is used to monitor and manage multiple users and applications under a given set of constraints. Service-level demands can be met by reassigning computing resources to manage peak workloads and changing business needs.

Multiple servers in a grid environment enable jobs to run on the best available resource. If a server fails, jobs transition to other servers—providing a highly available business analytics environment.

Multiprocessing capabilities divide individual jobs into subtasks that are run in parallel on the best-available hardware resources in a grid environment. Faster processing of data integration, reporting, and analytical jobs speeds decision making across the enterprise.

DEALING WITH ALL THE DATA, QUICKLY

The slowest part of a computer process occurs when data is read from and written to disk. SAS changes the game by loading all relevant data into memory. Instead of “going to the well” every time data is needed, computer processes execute faster by simply accessing the data in memory. While each computer has its own memory, SAS In-Memory processes “see” memory on many computers as one big block.

A rudimentary statistical example is the calculation of the mean (average) versus the median (midpoint). To calculate the mean, each computer can look at its own data, get a sum and a count, sum all of the sums, sum all of the counts, and finally divide the total sum by the total count. To calculate the median, the computer must sort and count all of the data and then find the middle number. The median is not easy to calculate if the data is on many computers. But with the data in-memory, and all memory “seen” as one big block, SAS turns this task into a trivial process.

SAS divides data and analytic processes into manageable pieces with computations distributed in parallel across dedicated computers. SAS leverages big data and sophisticated analytics to tackle complex problems quickly and to solve dedicated, industry-specific business challenges faster than ever before.

SAS provides concurrent, in-memory, and multiuse access to structured and unstructured data, no matter how big or small. As long as the dedicated compute environment has enough memory, SAS can hold the data. Since the software is optimized for distributed, in-memory processing, new scenarios can be run and complex analytical computations can be completed very quickly. Data can be explored, visualized, and analyzed, all to help solve problems that could never before be considered due to computing constraints.

OPERATIONALIZING SAS WITH THE DATA REPOSITORIES

Organizations have made large investments in data storage. SAS In-Database technologies leverage these massively parallel processing (MPP) database architectures for faster execution of key data management, analytic development, and deployment tasks. SAS analytics and other relevant tasks have been moved closer to the data, and computations are run inside the database to avoid time-consuming data movement and conversion.

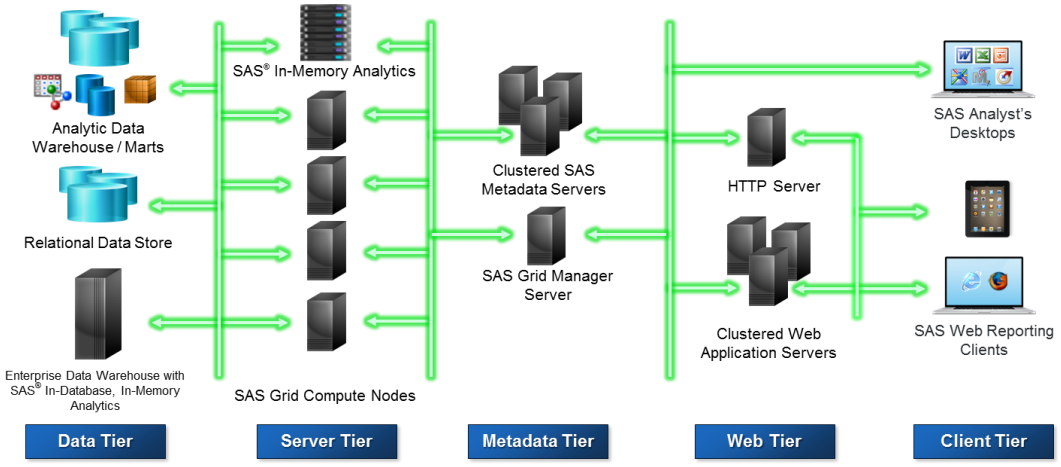
Data governance is enhanced by the reduced need to replicate or move large amounts of data between a data warehouse and the analytical environment or data marts.

Increased efficiencies and decreased costs are realized through the use of existing infrastructure and resources. This leverages existing investments and increases operational efficiency, yielding a faster time to value and reducing cost of ownership.

SAS In-Database technologies enable organizations to realize true insight more quickly by eliminating the need to move data between the modeling environment and the database for analytic scoring.

SERVER TOPOLOGY

Enterprise SAS leverages several “tiers” within an IT infrastructure, from data to insight delivery. SAS can be embedded on the data tier, SAS processes are executed on the server tier, security is managed on the metadata tier, and SAS user interfaces are surfaced on the web and client tiers.



All SAS functionality—from data acquisition through analytic calculation, insight delivery, and visualization—leverages this architecture.

For high availability, a complete Enterprise SAS environment has failover and redundancy on the web tier and the metadata tier. In addition, it has multiple, managed computing servers and dedicated in-memory servers.

CONCLUSION

Any analytics application follows this flow:

1. Operational data flows through a data acquisition and data credibility component to determine what data is relevant to solving the problem.
2. The relevant data is then analyzed and turned into useful information.
3. The useful information is then visualized and consumed to provide insight.
4. Insight is acted upon for decision making.

The process flow can be represented as follows:

Here is the same flow, seen from an IT process perspective:

SAS enterprise architecture provides reliable infrastructure, security, performance, and software versatility to help your organization transform the way you do business.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Gary Spakes
100 SAS Campus Drive
Cary, NC 27513
SAS Institute Inc.
gary.spakes@sas.com
http://www.sas.com