The Deployment of SAS Enterprise Business Intelligence Solution in a large IBM POWER5™ Environment

Edward Hayes-Hall
Brian Porter
Harry Seifert
Hsian-Fen Tsao
Introduction

A user of a SAS solution should not have to care about the underlying IT infrastructure. The user wants to use the solution to solve his or her unique business problem. The CFO of a retail company wants to understand how his business is doing. How many items are selling in his stores and which retail lines are making the most profit. The Business Analyst wants to develop models of how his business works. Neither of these personae are concerned with the underlying SAS architecture; they just want it to work.

However, “just wanting it to work” has some implied IT requirements for that solution and the environment in which it runs. The user wants the solution to be usable and to perform. If a CFO is used to seeing a Microsoft Excel spreadsheet that incorporates SAS data then he wants the SAS 9 Microsoft plug-in to work seamlessly and without perceived delays on his laptop or system on his desk in the corporate office.

User requirements have a profound effect on the IT deployment, management and administration of a SAS solution. SAS 9 is built around the cornerstones of Usability and Performance or scalability. In addition, SAS 9 strives to provide manageability and interoperability. These features work in concert to allow an IT Architect to deploy SAS 9 in a number of ways to fulfill the users requirements.

In recent years server/storage consolidation has proved to be a strategic direction in which Enterprises have gone. Server/Storage consolidation brings a number of benefits to the table such as ease of system management, lower IT costs and higher resource utilization.

This document describes the deployment of a large SAS Enterprise Business Intelligence solution in a server/storage consolidation scenario. While specific design criteria may be different for individual enterprise deployments, this example will describe a consolidated environment that successfully supported nearly 1500 concurrent batch and online SAS users. The methodology used to test the user count is documented in the SAS eBI Reference Architecture document (IBM, 2006)

SAS eBI Solution Components

SAS Enterprise BI Server provides a suite of business intelligence tools as well as an integrated business intelligence platform. SAS analytics and a suite of user interfaces provide users with self-service query and reporting capabilities. The SAS 9 eBI Solution is described more fully in “The SAS Enterprise Intelligence Platform: SAS Business Intelligence” (SAS, 2004)
SAS eBI builds directly upon the SAS Intelligence Value Chain and is the solid deployment of an Intelligence Model using quality ETL data to generate reliable information that is stored in fast Intelligence storage which can be accessed by higher functioning Analytic Intelligence.

SAS comprises a series of foundational components common across all SAS solutions. Individual solution components utilize the foundational components by means of a suite of common services and tools. SAS 9 is built on the concept of servers or “engines”. These engines are responsible for services such as Metadata, Stored Processes, ETL, analytics or grid computing.

SAS 9 also introduces the concept of personae. These fall into three basic groups

- **Informational Users**: Corporate executives and managers who run operational style reports and dashboards. These users are typically extracting data via pre-defined reports either in operational areas such as sales, marketing, finance, etc. or in specific areas such as technical support.
- **Power Users**: Business Analysts and consultants who typically generate and run ad-hoc reports
- **BI IT Support**: Data Modelers and IT administrators who are generating SQL programs, building generic reports for Informational Users, etc.

Each of these personas utilizes a different mechanism to access the eBI solution and is represented in the concurrent user count in different amounts relative to one another. These different groups will access the environment through different mechanisms.

- **SAS Information Delivery Portal**
- **SAS Web Report Studio**
- **SAS Enterprise Guide**
- **SAS ETL Studio**
- **SAS Management Console**

The SAS environment can be loosely defined as a 3-tier architecture comprising a web and application tier, a services middle tier and a computational back-end tier. These tiers are shown conceptually in Figure 1 with the individual access points for each of the persona that would use the solution.

**Tier 1**: The web/application tier comprises those components that deliver the solution to the end user. This tier is Java based and provides the Portal, Dashboard and web studio components along with query services and solution specific components. This tier is primarily accessed by the Informational and Business Analyst personae.
Tier 2: The “services” tier provides the Metadata, Remote services and job scheduling or solution launching specific components of the application. The Metadata Engine is a crucial component of a deployment. This tier will typically be accessed by the Analyst and administrative community rather than informational users.

Tier 3: The “computational” tier provides the raw processing capability to the solution. This tier provides components such as the Stored Process Engine, the OLAP server, Workspace engine and any Relational Database processing components.

Figure 1: Conceptual SAS 9 Solution Architecture

This conceptual deployment could be architected in several ways. Large Enterprise installations have typically deployed this type of scenario in a distributed environment. That is, each of the individual tiers running on separate servers. Those servers are not necessarily of the same type. Large Enterprise deployments could very well utilize a large zSeries on the back end to drive database and SAS foundational components with Intel systems deployed at the web tier for cost efficiency. A deployment of this type will be documented as a follow up to this paper.

However, with the increasing acceptance of logical partitioning, an alternative method of deployment presents itself. That deployment example includes the use of partitioning technology on IBM pSeries servers.
Server Partitioning

Server partitioning allows the centralization of IT resources into a single location. Centralization of resources reduces logistical complexity and redundancy of hardware, software and personnel. Enterprise can reduce complexity by partitioning larger servers and storage devices into fewer, more powerful systems running multiple workloads in separate partitions.

It is often apparent that multiple data repositories exist in such environments. Even if this state was known prior to moving to a partitioned environment, often the act of migrating will precipitate a uniting of all the data sources within the Enterprise into a single unified representation of that data to the applications running.

The move to a partitioned deployment is driven typically by three major forces:

- Many organizations have significantly more storage and server resources spread throughout the enterprise than they did only a couple of years earlier.

- Those increases resources typically span multiple server vendors and multiple storage vendors and irrespective of open standards it is still difficult to get all the components to work co-operatively together in a seamless fashion.

- The majority of these deployed, distributed resources are under utilized.

Partitioning technology is a key strategic approach to lowering TCO, improving error rates, improving scalability and performance. But as with any IT project, there's a right way and a wrong way to approach application deployment in such environments. The primary hurdles to success are not so much technical as they are organizational and political.

The deployment of Server and Storage partitioning and the adoption of a more centralized approach can bring significant benefits, including reduction in the system administration overhead, easier data management, and a simpler infrastructure. These benefits will differ from enterprise to enterprise, and are dependent on the demands of the business. IT operational costs have a direct impact on the bottom line of a business. The simplification of the systems infrastructure through partitioning can bring significant financial benefits, such as lower infrastructure overheads, with smaller and fewer data centers, reduced staff costs, lower licensing charges, and decreased maintenance costs.
IBM Server Technology

The IBM pSeries p5 product line, in conjunction with the AIX 5L operating system, provides unparalleled functionality and performance for implementing Business Intelligence solutions in large-scale enterprise-wide data warehouses. The IBM pSeries p5 product line encompasses Enterprise servers designed for a broad range of applications required by medium and large companies.

The performance of the 64-bit IBM POWER5™ platform, enabled by simultaneous multi-threading and the IBM AIX 5L™ operating system, is ideal for all Business Intelligence processing needs. IBM’s continued investment in the AIX 5L system and unique advanced virtualization technologies like the IBM Micro-Partitioning™ capability enable UNIX system users alike to do more with a single system.

The IBM eServer pSeries is designed for agility to enable you to quickly respond to changes in UNIX processing requirements, so you can meet anticipated and unanticipated business demands. pSeries servers provide the ability to dramatically increase system utilization in a consolidated server environment.

The IBM TotalStorage™ Difference

IBM provides organizations what they need to make the right storage decision for both now and the future.

IBM provides a focus on open standards, support for heterogeneous environments and continuous leverage of innovative systems technologies, including the IBM Power Architecture™ platform. This storage system approach provides Logical partitioning (LPAR) functionality on select combined with leading IBM POWER performance, permit new efficiencies of scale.

IBM TotalStorage SAN Volume Controller and TotalStorage SAN File System (SFS) software enable the virtualized management of heterogeneous storage resources.

SAS 9 eBI Large Deployment Example

Data support for this test suite was built from retail sales data. The data has several main fact tables which contain sales transaction related items. This data has been sanitized to protect actual sales and other competitive data including price, inventory, and sales volume data. This data includes more than 160 GB in relational data and 60 GB in cube data plus underlying cube detail data. The initial test defaults to SAS datasets, but can quickly be converted to utilize advanced storage products like SAS SPD Server or a third party relational database if desired for additional testing.

There are several user types and counts. The following table details user base is designed to simulate a typical field deployed system.
Storage is on striped, high performance disks sufficient for the task. The most common setting found in the field is RAID-5 for images and data; this storage will be largely read-only for the test. The most common setting found in the field is RAID-0 or RAID-1 for SASWork; this storage will be both heavily read and heavily write for this test. All physical machines, except for client machines, are on one common sub-net with 1Gbit/sec throughput.

**Hardware Deployment**

In order to support nearly 1500 concurrent users, the SAS eBI solutions was deployed as 8 physical partitions of varying sizes across a single IBM eServer p590 POWER5 system. The basic deployment of the hardware is shown in figure 2.
The web tier comprised four identical partitions. Each partition comprised two POWER5 physical CPUs with 8 GB RAM and four 36Gb hard disk drives. Each web tier partition was provided with a 1 Gbit Ethernet connection to an Ethernet switch providing network connectivity to a LAN comprising a number of xSeries Intel systems that provided the load generation of the workload. In addition, each web tier partition has a second 1Gbit Ethernet connection to a second subnet which is provided for data communication between the SAS components in the other tiers.

The SAS Metadata Server is deployed as a separate partition dedicated to Metadata functionality. This server has a both a dedicated 1Gbit Ethernet connection and four 1Gbit Fiberchannel connections to the DS8100 TotalStorage device. The Metadata log files are stored to locally attached disk.

Third tier computational processing is completed using three separate partitions. The SAS 9.1.3 base functionality (Workspace engine, etc.) are deployed in a dedicated 8-way partition with 32Gb RAM and sixteen 36Gb locally attached disk drives. The SAS OLAP cube is deployed in a separate partition and finally a small DB2 instance, along with an installation of Xythos are placed in a third partition. Locally attached drives to all these partitions contain the SAS binary files and various log files. The /saswork temporary file system and /data file system are allocated to logical volume groups deployed on the shared DS8100 storage device. Each of the computational partitions is deployed with four 1Gbit Fiber channel paths to the storage array.

The entire deployment uses a single IBM eServer p590 frame utilizing 28 physical POWER5 CPUs with 4Gb RAM/physical CPU (total of 112 Gb RAM) and 64 locally attached 36 Gb hard disk drives.

The performance and utilization of this environment is documented in the paper “Performance and Tuning of SAS workloads on IBM eServer pSeries POWER5 systems” (IBM, 2006)

**Software Deployment**

The software deployment can be best described by the following table.

<table>
<thead>
<tr>
<th>Server Name</th>
<th>Software</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata Server</td>
<td>AIX 5L</td>
<td>5.3 ML 02</td>
</tr>
<tr>
<td></td>
<td>SAS Metadata Server</td>
<td>9.1.3</td>
</tr>
<tr>
<td>DB2/Xythos (WebDav) Server</td>
<td>AIX 5L</td>
<td>5.3 ML 02</td>
</tr>
<tr>
<td></td>
<td>Xythos Webfile server</td>
<td></td>
</tr>
<tr>
<td></td>
<td>DB2 UDB</td>
<td>8.1</td>
</tr>
<tr>
<td>SAS Base Server</td>
<td>AIX 5L</td>
<td>5.3 ML 02</td>
</tr>
<tr>
<td></td>
<td>SAS Base</td>
<td>9.1.3</td>
</tr>
<tr>
<td></td>
<td>DB2 Client</td>
<td>8.1</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------</td>
<td>-----</td>
</tr>
<tr>
<td>SAS OLAP Server</td>
<td>AIX 5L</td>
<td>5.3 ML 02</td>
</tr>
<tr>
<td>SAS OLAP Server</td>
<td></td>
<td>9.13</td>
</tr>
<tr>
<td>Web Tier 1</td>
<td>AIX 5L</td>
<td>5.3 ML 02</td>
</tr>
<tr>
<td></td>
<td>Websphere MQ</td>
<td>5330.6</td>
</tr>
<tr>
<td></td>
<td>Websphere AS</td>
<td>5.0.2.2</td>
</tr>
<tr>
<td>Web Tier 2</td>
<td>AIX 5L</td>
<td>5.3 ML 02</td>
</tr>
<tr>
<td></td>
<td>Websphere MQ</td>
<td>5330.6</td>
</tr>
<tr>
<td></td>
<td>Websphere AS</td>
<td>5.0.2.2</td>
</tr>
<tr>
<td>Web Tier 3</td>
<td>AIX 5L</td>
<td>5.3 ML 02</td>
</tr>
<tr>
<td></td>
<td>Websphere MQ</td>
<td>5330.6</td>
</tr>
<tr>
<td></td>
<td>Websphere AS</td>
<td>5.0.2.2</td>
</tr>
<tr>
<td>Web Tier 4</td>
<td>AIX 5L</td>
<td>5.3 ML 02</td>
</tr>
<tr>
<td></td>
<td>Websphere MQ</td>
<td>5330.6</td>
</tr>
<tr>
<td></td>
<td>Websphere AS</td>
<td>5.0.2.2</td>
</tr>
</tbody>
</table>

Table 2: Server Software Matrix

Each partition is built with identical AIX 5L Version 5.3 operating systems. In a production environment it would be prudent to operate a NIM (Network Installation Manager) server with the latest fix pack versions of the operating system available. OS maintenance and upgrading of fix packs would take place on the NIM server production version and then would be pushed out to the individual production partitions. The Network Installation Manager is described in more detail in IBM publication SG24-6289 (Quintero, et al, 2004)

Conclusions

The successful completion of performance testing of the SAS Enterprise Business Intelligence reference architecture unequivocally showed that a IBM eServer pSeries p590 could easily sustain a concurrent steady state load of nearly 1500 users from across a number of varied personas. The environment was deployed in a consolidated fashion in eight partitions on a pSeries p590 using 28 physical CPUs and 4Gb RAM/physical CPU. The environment included a large Total Storage DS8100 storage array.

References


