Paper 216-31

Flying a Pilot BI Solution in a Virtualized UNIX Environment

Frank Bartucca
IBM eServer Solutions Enablement
January 2006

Table of contents

Abstract

The relentless drive of server vendors towards improved price/performance and reduced total cost of ownership has resulted in modern UNIX servers acquiring some of the virtualization features previously only available on mainframes. The prize sought after is to achieve the 90%+ average utilization rate typically enjoyed by mainframes instead of the typical 70% to 80% average idle time seen on traditional UNIX servers. Virtualized computing environments are certainly not new to SAS® users who run SAS on mainframes. However, virtualization is relatively new to those who run SAS in UNIX environments and somewhat foreign to those who exclusively use PCs. We tested our pilot BI implementation using the SAS Enterprise BI reference architecture test suite. One of our goals was to keep our midrange 8-CPU IBM® eServer pSeries® p570 server continuously "flying" with high utilization rates by taking advantage of its advanced features. In this paper we talk about all the things we learned as we tested our pilot BI implementation.

Introduction

We implemented our pilot deployment of the SAS Enterprise BI Server on a midrange system as the first step toward the target goal of scalability testing on an enterprise class system. The target enterprise class system was an IBM p590 with thirty-two 1.9-gigahertz POWER5™ CPUs, 128 gigabytes of main memory and an IBM TotalStorage™ DS8100 with sixty-four gigabytes of cache in front of ten gigabytes of storage. This configuration of the p590 had at least five times the capacity of our midrange p570.

We were not concerned about the capacity difference between the two systems because our main goal with the pilot system was to just get the deployment working. Our main concern was that our scalability test plan showed the p590 partitioned into eight servers. Four servers were used as Web Tier servers and one server each was used for the SAS Metadata Server, the SAS Workspace Server, the SAS OLAP Server and the IBM DB2 Server. Even though our p570 had eight CPUs there were not enough I/O adapters to make eight partitions and there were already other projects that needed partitions on the p570. The solution was to implement virtualization.

Steps toward virtualization

- Create Virtual I/O Server (VIOS) partitions
  - One real CPU per VIOS partition
  - Two gigabytes of main memory per VIOS partition
- Install the VIOS software that came with the p570
- Give ownership of all of the I/O adapters to the VIOS partitions
- Create virtual SCSI and Ethernet devices
- Create shared resource partitions as needed using real memory, virtual I/O and virtual CPUs
- Install an operating system (AIX, Redhat Linux, SUSE Linux or i5/OS) on the shared partitions

Virtualization concepts

Micro-partitioning

Mainframe functionality has been added to new POWER5 based IBM pSeries servers. In addition to the traditional dedicated partitions that have whole numbers of real CPUs allocated, shared partitions can now be created that have fractional amounts of real CPUs allocated in virtual CPUs. As an example, on an 8-way box, four partitions can be created that each have 8 virtual CPUs. Each virtual CPU looks like a real CPU to the OS running in the partition. The OS could be AIX 5.3, Redhat Linux, SUSE Linux or i5/OS. As an example if each virtual CPU is given a minimum CPU capacity of 0.2 real CPUs for each of four 8-way partitions then 6.4 real CPUs of processing power is the maximum *guaranteed* processing power the system has to meet. Individual partitions can also be *uncapped* meaning that they can use up to 1.0 real CPU per virtual CPU if it is available on the box. Also on a millisecond by millisecond basis AIX 5.3 (or higher), Linux or i5/OS can cede real CPU cycles to other partitions if they are not using the capacity. In our example above if three of the partitions are not busy and the fourth is uncapped then the fourth could have almost 8 real CPUs of processing power available to it.

Simultaneous Multi-threading

The POWER5 processors are capable of Simultaneous Multi-threading (SMT) and can have two hardware threads running on a single CPU core. Each thread looks to AIX 5.3 mostly like a real CPU so each thread is called a logical CPU. As an example, in a dedicated partition with four real CPUs with SMT enabled AIX will run on and schedule 8 logical CPUs as if they were real CPUs. As another example, in a shared partition with 8 virtual CPUs and SMT enabled AIX will run on and schedule 16 logical CPUs. So the hierarchy from lowest architectural level (least abstracted) to highest (most abstracted) is real, virtual and then logical CPU.

Performance tests run using threaded PROCs on mixed workloads by IBM at SAS showed an average performance improvement of 30% when SMT was enabled. This was possible because the workloads used in the tests were able to take advantage of the additional logical CPUs available to AIX when SMT was enabled. With SMT enabled it was like having 10.4 CPU cores of processing power (as normalized to the performance of 8 real CPU cores with SMT disabled).

Virtual I/O

The combination of specialized hardware and the firmware based Virtualization Engine on POWER5 servers allows I/O adapters to be virtualized. So far SCSI, Ethernet and terminals have been virtualized and in the future memory will be virtualized. Virtual adapters appear as real devices to the operating system running in the partition.

By using virtual Ethernet, a high speed network can be set up between partitions on the same server without using a physical Ethernet adapter. One or more partitions could also bridge the virtual network to the outside world using a physical adapter owned by the partition.

Specialized Virtual I/O Server (VIOS) partitions can be set up that own real Ethernet and SCSI adapters and present virtual Ethernet and SCSI to client partitions. In this way more virtual adapters can be defined on the server than there are physical adapters actually installed. Data transfers using virtual SCSI

adapters are just as efficient as using real SCSI adapters because the real SCSI adapter in the VIOS partition use DMA to transfer data directly to and from memory in the client partition.

Partition Load Manager

The Partition Load Manager (PLM) provides automated processor and memory distribution between dynamic LPAR and Micro-Partitioning-capable logical partitions running AIX. To improve the overall resource utilization of a partitioned system, PLM uses user-defined resource management policies to determine the additional resources, such as processors and memory, for each requesting partition. The PLM monitors for the occurrence of high threshold and low threshold events for Memory-pages-steal, Memory-usage and Processor-load-average.

SAS Enterprise BI Server reference architecture

The SAS Enterprise BI Reference Architecture Test Suite is a reference architecture for a 300 user enterprise work load based on scenarios of SAS customers addressing their data reporting needs. The simulated environment provided by the test suite uses real-world user volumes, reports and data. While the data used is retail based, the work load is representative of BI processing found in other business segments. SAS Enterprise BI Server reporting components included in the test suite are SAS Enterprise Guide, SAS Add-In for Microsoft Office®, SAS Information Delivery Portal and SAS Web Report Studio.

Types of users represented in this simulation

- Four SAS Enterprise Guide software client users
- Thirty users accessing SAS Stored Processes simulating use of SAS Add-In for Microsoft Office
- Thirty-six users building dynamic reports based on relational data via SAS Web Report Studio
- Thirty users viewing OLAP cubes via SAS Web Report Studio
- Two hundred users viewing static reports through SAS Information Delivery Portal, HTTP and SAS Web Report Studio

Types of reports used in this simulation

There were 23 dynamic reports and 40 static reports used in this simulation. These reports provide simulated users with the following information:

- Detection of low product inventory levels
- Detailed sales performance of product vs. location
- Comparison of item sales with and without marketing affects and effort
- Identification of high-value customers
- Cube views of sales performance by marketing effort, markdown and other causes and effects

Who developed the BI Reference Architecture test suite?

The SAS Enterprise Excellence Center (EEC) developed the SAS Enterprise BI Reference Architecture Test Suite. The EEC provides high-end performance metrics based on real-world customer scenarios. This enables customers to preview the performance of SAS technologies on their servers.

What we did

Implementing virtualization

First we made installable backups of the partitions we wanted to save. We then updated the system firmware and the Hardware Management Console (HMC) software to the latest levels plus the latest fixes. Two Virtual I/O Server (VIOS) partitions were created, named vios0 and vios1 and then installed using the HMC. The ownership of all real I/O adapters was split evenly between each VIOS and each VIOS was given one dedicated CPU and two gigabytes of memory. With 1 gigabyte reserved by the firmware this left us with 27 gigabytes of main memory and 6 real CPUs for creating the shared resource partitions.

Using a combination of VIOS and HMC commands we then created sixteen 36-gigabyte virtual SCSI disks and eight virtual Ethernet adapters. We then created eight shared resource partitions as shown in Table 1.

Name	Virtual CPUs	Dedicated CPUs	Entitled real CPU capacity	Capped?	Memory	Virtual disks	Real disks	Ethernet
db2	2		0.2	no	3	2		virtual
meta	2		0.2	no	3	2		virtual
olap	2		0.2	no	3	2		virtual
workspace	4		0.4	no	6	2		virtual
web0	2		0.2	no	3	2		virtual
web1	2		0.2	no	3	2		virtual
web2	2		0.2	no	3	2		virtual
web3	2		0.2	no	3	2		virtual
vios0		1	1.0	yes	2		24	real
vios1		1	1.0	yes	2		24	real

Table 1: Partition configurations

We gave the VIOS partitions dedicated CPUs because of the anticipated heavy I/O load. The other partitions received virtual CPUs with uncapped capacity because we did not know what the workload would be like. The *entitled capacity* for virtual CPUs guaranteed that if the virtual CPU had work to do

then, no matter what the demands were on the rest of the system, the virtual CPU would always get its entitled capacity of real CPU backing it up.

For efficiency if a virtual CPU has nothing to do then it can potentially use near zero real CPU capacity, instead of spinning in a wait loop. This *ceding* of CPU capacity back to the free CPU pool is an operating system enhancement. An OS could actually choose to just spin on an idle CPU.

A *shared partition*, which is a partition that uses virtual CPUs, can have the attribute of being capped or uncapped. If a partition is *capped* then its virtual CPUs can never use more real CPU capacity than the partition's entitled capacity. If a partition is *uncapped* then each virtual CPU defined in a partition can use up to a full real CPU worth of capacity if it is available from the free pool.

Implementing some redundancy

One of the concerns we had with using virtual I/O was that if a VIOS crashed or was brought down for maintenance then the virtual I/O devices it served to partitions would disappear and cause multiple partitions to go down. We addressed this problem for the virtual disks by using volume group mirroring. Each of the VIOS servers contributed a virtual disk to make 2-way mirrored root volume group on each partition. Therefore, one VIOS could go down and the shared partitions would still have functioning, although no longer redundant, volume groups. We did not try it but in theory when the second VIOS comes back up its stale virtual disks would be resynchronized and we would again have a 2-way mirror. Using a 2-way mirror cut our disk capacity in half but we still had more than enough disk space.

Creating redundancy for the virtual Ethernet was not as much of a problem. The virtual Ethernet devices are implemented in firmware and can be created for partition to partition communications without the involvement of a VIOS. However, to have connectivity to the outside world a physical Ethernet adapter must be used to bridge between the internal LAN and the external LAN. We used a physical Ethernet adapter in each VIOS and had vios0 and vios1 backing each other up incase an adapter failed or one of the VIOS partitions went down.

Installing the software

We decided to install all of the software on one partition, get it all working and then move components of the BI Server to the other partitions.

IBM software

This was the easy part. IBM WebSphere Applications Server V6, IBM HTTP Server V2.0 (DAV/2), IBM DB2 UDB ESE V8.2, and the mozilla and firefox web browsers had already been installed for us. All we had to do was apply fix packs.

SAS Software

Installing the SAS Enterprise BI Server took a bit longer but it was relatively easy using the SAS Software Navigator. We selected the choice: `Enterprise BI Server, one machine`. After installation there were still a few manual configuration steps to go through but the file `/home/sas/SAS/EntBIServer/instructions.html` produced by the install step was a great guide

through the rest of the setup. The various shell scripts listed in the file saved a lot of error prone manual typing.

Running the software

After completing the setup we ran various examples and samples to make sure everything was working on a single partition. We then moved the Metadata Server to the meta partition, DB2 to the db2 partition and so on.

Crashing our pilot BI solution

After manually moving various pieces of IBM and SAS software to the other partitions we were never able to correctly reconfigure all of the software pieces to get the BI solution working well enough to run our tests. We had trashed an implementation that was working on a single machine and had effectively crashed our pilot BI solution.

Server setup summary

For the demonstration featured in this paper, the following hardware and software configuration was established.

Hardware

The hardware configuration of the IBM p570 server used was:

- Four Dual Chip Modules (DCMs) with:
  - One 32-megabyte L3 cache chip per DCM
  - One 1.65 GHz POWER5™ chip per DCM with:
    - 1.9-megabyte L2 shared cache per chip
    - Two CPU cores per chip with:
      - 32-kilobyte L1 D-cache per core
      - 64-kilobyte L1 I-cache per core
- 32-gigabyte main memory
- Seven PCI-X ULTRA320 SCSI adapters
- Forty-eight 36.4-gigabyte ULTRA320 15K RPM SCSI disks
- Twelve 1-gigabyte/s Ethernet adapters

Software

The following releases of IBM and SAS software were installed:

- IBM
  - DB2 UDB ESE V8.2
  - IBM HTTP Server V2.0.47.1
  - WebSphere Application Server (WAS) V6.0.2.5
- SAS Enterprise BI Server components:
  - SAS Version 9.1.3 + SP3
  - SAS Metadata Server
  - SAS Information Delivery Portal
  - SAS Stored Process Server
  - SAS Web Report Studio

Summary

Manually moving and reconfiguring the various pieces of IBM and SAS software onto the other partitions proved to be more complicated and error prone than we had anticipated. We were no longer protected from the complexities taken care of by the SAS Software Navigator and by the DB2 and WebSphere experts who had already done the base installs for us.

In retrospect we should have:

- brought back our DB2 and WAS experts to distribute DB2, WAS and HTTP to three of the partitions
- used SAS Software Navigator and selected the install choice `Enterprise BI Server, three machines`

We plan to complete this work and present the results at SUGI31.

Resources and References

SAS Enterprise BI Server: *Reference architecture for a 300 user enterprise work load* (SAS Enterprise Excellence Center, Performance Brief)

These Web sites provide useful materials to supplement the information contained within this document.

- IBM Techdocs (the Technical Sales Library)
  ibm.com/support/techdocs/
  - Hints and Tips for Running SAS on an IBM eServer™ pSeries Server and AIX 5L (TD102517)
  - Tune AIX 5L for the SAS 9 System (TD102515)
- AIX 5L Performance guides:
  publib.boulder.ibm.com/infocenter/pseries/index.jsp
  Navigate **AIX documentation** > **AIX PDFs** > then, under **Performance**
  - AIX 5L V5.3 Performance Management Guide
    in the right pane, click **Performance Management Guide**
  - AIX 5L V5.3 Performance Tools Guide and Reference
    in the right pane, click **Performance Tools Guide and Reference**
  - Performance Toolbox Version 2 and 3 Guide and Reference
    in the right pane, click **Performance Toolbox Version 2 and 3 Guide**
- Microcode survey information
  https://techsupport.services.ibm.com/server/aix.invscoutMDS
- AIX commands and tools information:
  publib.boulder.ibm.com/infocenter/pseries/index.jsp
  Navigate **AIX documentation** > **Commands reference** > **Alphabetical list of commands** >
- SAS Home page
  sas.com
- SAS and IBM white papers
  sas.com/partners/directory/ibm/papers.html
  - Configuration Options using IBM SSA Storage for SAS
  - Maximizing Performance on IBM Enterprise Storage Servers®
  - High Availability Storage on IBM Enterprise Storage Servers
  - Managing SAS Storage on IBM Enterprise Storage Servers
- A Practical Approach to Solving Performance Problems with the SAS System (October 2001)
  support.sas.com/rnd/papers/sugi27/SolvingPerformance.pdf
- IBM eServer p5 Information Center
  publib.boulder.ibm.com/infocenter/pseries/index.jsp
- IBM Publications Center
  www.elink.ibmlink.ibm.com/public/applications/publications/cgibin/pbi.cgi?CTY=US
- IBM Redbooks™
  www.redbooks.ibm.com/
  - AIX 5L Practical Performance Tools and Tuning Guide (SG24-6478)

About the author

Frank Bartucca

IBM Systems and Technology Group

Raleigh, North Carolina

Mr. Bartucca is a senior engineer and scientist with IBM eServer Solutions Enablement in the IBM Systems and Technology Group. He has been working with AIX and IBM RISC technologies since 1984. In his current capacity, as a pSeries technical consultant, he provides technical support to software developers and solution providers who are developing software products to run on the AIX operating system.

Trademarks and special notices

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

The following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both: IBM, the IBM logo, eServer, pSeries, AIX, AIX 5L, Redbooks, Enterprise Storage Server, and ibm.com.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Other company, product or service names may be trademarks or service marks of others.

Information is provided "AS IS" without warranty of any kind.

All customer examples described are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Information concerning non-IBM products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capability of non-IBM products should be addressed to the supplier of those products.

Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.