Paper 380-2010

SAS® Enterprise Applications and IBM Virtual Environments

Alfredo Mendoza, IBM Corporation, Austin, TX
Frank Battaglia, IBM Corporation, Austin, TX
Darrell Barton, SAS Institute

ABSTRACT

Over the past several years, virtualization has been a hot topic of much discussion by IT personnel as a way to reduce cost, improve utilization and responsiveness, and increase flexibility. Although server consolidation still remains the primary reason for using virtualization, application isolation and continuity are also discussed as other uses. This paper presents three use cases for deploying SAS® Business Analytics Framework on IBM virtual environments when running on a POWER6 server. The information provided can serve as a guide for IT personnel in choosing the best virtual environment for their customer needs.

INTRODUCTION

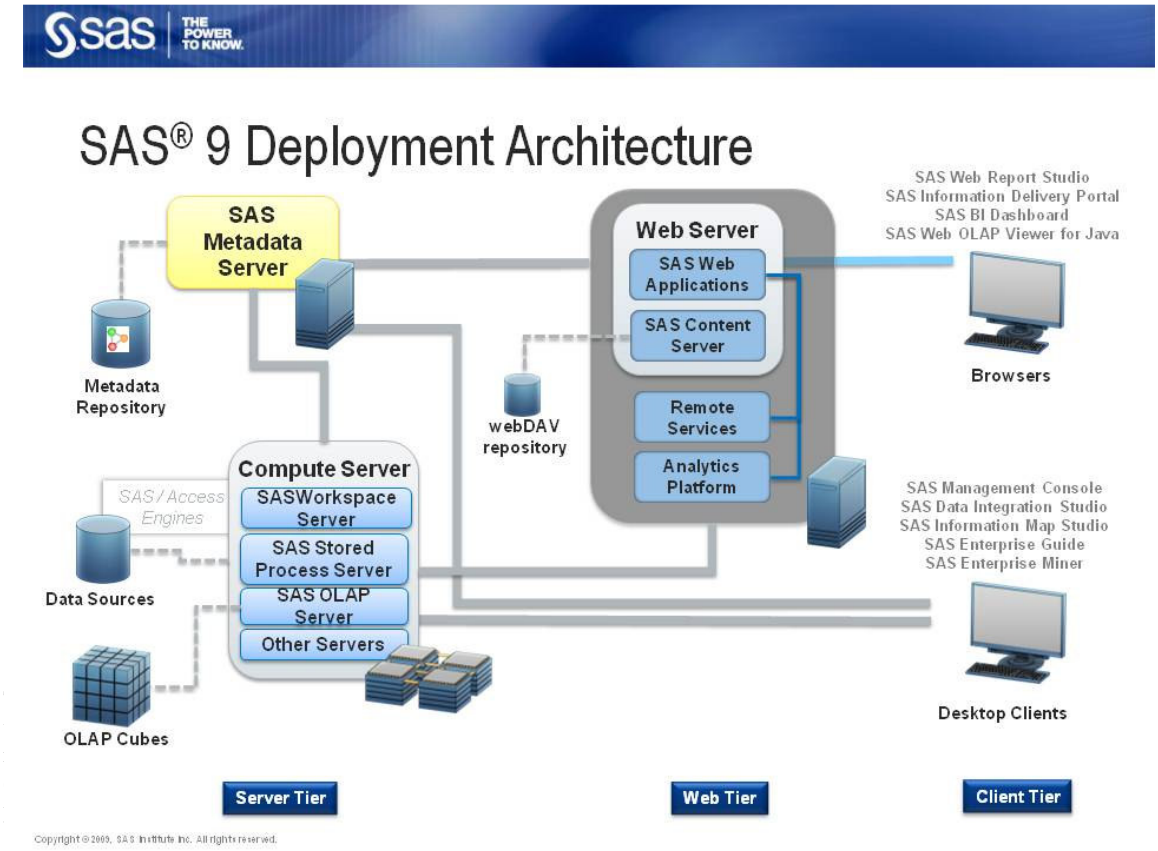
Transforming and analyzing data to provide proactive, evidence-based decision strategies have become increasingly important in today's highly competitive environment. Increased demand for business analytics that are designed to navigate through large amounts of data in the fastest time possible provides companies with a truly competitive advantage. The SAS Business Analytics Framework (Release 9.2) for IBM® AIX 6 on POWER™ processor-based systems provides such solutions. The unmatched expertise of IBM in hardware and software technology and services enables the SAS Business Analytics Framework to be deployed on an infrastructure that is designed to improve reliability, performance and scalability.

Although business decision makers deal with mission-critical issues, their IT organizations find ways to improve hardware resource utilization, responsiveness and flexibility while maintaining performance and scalability. One way that IT organizations are doing this is through the use of virtualization technologies. A 2008 Gartner study states that virtualization will be the highest impact trend in changing IT infrastructure and operations through 2012. Today, the desire to increase server utilization and lower total cost of ownership through server consolidation are the main driving forces behind the increasing use of virtualization. However, there are other reasons why an IT shop might choose to use more advanced virtualization technologies -- such as application isolation and business continuity, both of which are explained in this paper.

The paper presents three use cases that are common to a typical IT shop: the need to consolidate servers, isolate applications and provide business continuity. The use cases provide a way to demonstrate how IT personnel can use IBM PowerVM™ technology to improve resource utilization while maintaining service levels, performance and scalability of their SAS applications. Each use case shows the impact of virtualization on SAS applications and provides guidance on how to best use different virtualization features that PowerVM technology offers.

THE TEST SUITE AND TEST ENVIRONMENT

In testing the IBM virtual environments, a SAS Business Analytics Framework was set up to run data integration, mixed workload analytics and enterprise business intelligence workloads. **Error! Reference source not found.** shows the high-level diagram of the test environment.



TEST SUITE DESCRIPTION

Three types of workloads were used in the test environment. They are:

1. I/O Intensive

The I/O intensive jobs simulated Data Integration (DI) jobs that perform a bulk load of data. A SAS data- step program inputs, transforms, validates, and loads three years of weekly data into SAS data sets. Each week of data is approximately 1 GB in size and SAS data sets approximately 1.4 GB in size are created as a result of the program. The script that starts the DI jobs creates 156 single- threaded processes to perform the work.

2. Compute Intensive

The compute-intensive jobs are more diverse. Several different SAS analytics procedures are used to evaluate the data in a variety of ways. The procedures include,

- The MIXED procedure fits a variety of mixed linear models to data and enables you to use these fitted models to make statistical inferences about the data. A mixed linear model is a generalization of the standard linear model used in the GLM procedure, the generalization being that the data are permitted to exhibit correlation and non-constant variability. Therefore, the mixed linear model provides you with the flexibility of modeling not only the means of your data (as in the standard linear model) but their variances and co-variances as well.
- The GLM procedure uses the method of least squares to fit general linear models. Among the statistical methods available in PROC GLM are regression, analysis of variance, analysis of covariance, multivariate analysis of variance, and partial correlation.
- The REG procedure is a general purpose procedure for regression.
- The LOGISTIC procedure fits linear logistic regression models for discrete response data by the method of maximum likelihood. It can also perform conditional logistic regression for binary response

data and exact conditional logistic regression for binary and nominal response data. The maximum likelihood estimation is carried out with either the Fisher scoring algorithm or the Newton-Raphson algorithm, and you can perform the bias-reducing penalized likelihood optimization (as discussed by Firth [1993] as well as Heinze and Schemper [2002]). You can specify starting values for the parameter estimates. The logit link function in the logistic regression models can be replaced by the probit function, the complementary log-log function, or the generalized logit function.

- The CORR procedure is a statistical procedure for numeric random variables that computes Pearson correlation coefficients, three nonparametric measures of association, and the probabilities that are associated with these statistics.
- The MEANS procedure provides data summarization tools to compute descriptive statistics for variables across all observations and within groups of observations. For example, PROC MEANS calculates descriptive statistics based on moments, estimates quantiles (including the median), calculates confidence limits for the mean, identifies extreme values, and performs a t test.
- The RISK procedure performs credit and risk analysis operations by creating base case market and position data, revaluing the data by using current prices, rates and other risk factors, and finally creating additional market scenarios by using stochastic processes.

The compute intensive script creates fewer processes compared to the I/O intensive script. The compute intensive script starts 19 processes then waits five minutes before starting the next 11 processes for a total of 30 processes. Each process may have between 1 and 4 threads in each process.

3. Web report generation

The goal of the Web report generation workload was to test client response while the SAS Business Analytics Framework middle-tier was being moved with Live Partition Mobility (LPM). LoadRunner software was used to simulate users who generate reports with SAS® Web Report Studio. LoadRunner is a performance and load testing software that can emulate hundreds or thousands of concurrent users.

It is used to put the application through the rigors of real-life user loads while collecting information from key infrastructure components (Web servers, database servers and others). The results can then be analyzed in detail, to explore the reasons for particular behavior.
The LoadRunner scenario mimics a typical user scenario of connecting to a web application,

1. Connect to SAS Web Report Studio web application
2. Log in using the sasdemo id
3. Select a report to execute
4. Select report parameters
5. Run the report

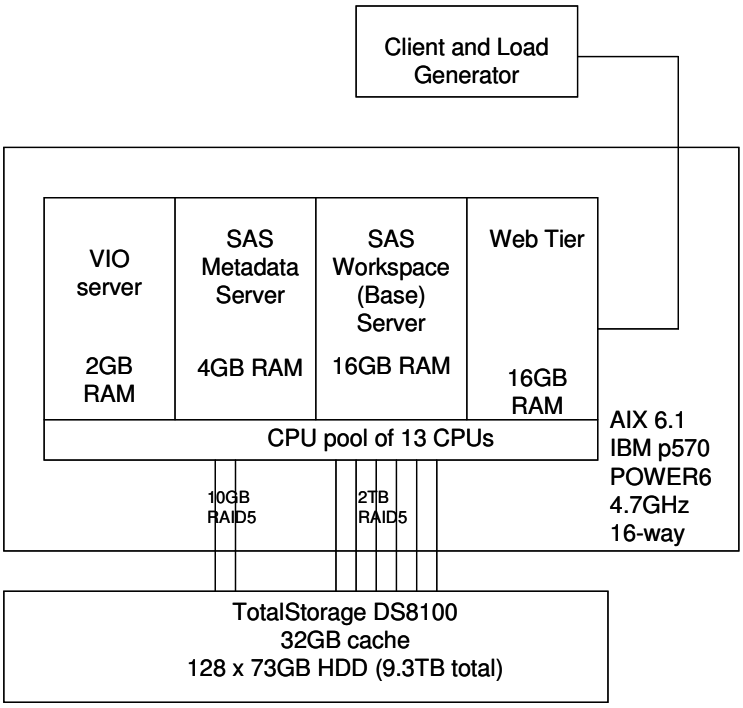
The scenario starts by running two users and then slowly adding two users every 15 seconds until the maximum of 25 users is reached. The 25 users run for 5 minutes. While all 25 users are running, the partition that hosts the SAS Business Analytics Framework middle tier is moved and the response of the users that run the scenario is monitored.

TEST ENVIRONMENT

The SAS test environment is configured as a multi-tier architecture as show in **Error! Reference source not found.**. The multi-tier configuration consists of a metadata server tier, a Web services middle tier and a computational back-end tier. The multi-tier configuration is used for a variety of tasks including data query, data analysis and interactive business visualization. Each tier is installed in a separate logical partition[1] (LPAR) instance. Typically, the multi-tier architecture is deployed in a distributed environment where the different SAS components are installed on separate physical servers. However, with IBM PowerVM virtualization technologies, the alternative approach of installing the SAS components in logical server partitions is recommended.

The LPARs for each tier component are configured to use PowerVM technology. **Error! Reference source not found.** shows the high-level configuration of the SAS environment installed on separate LPARs.

[1] An LPAR is a subset of a server's hardware resources, virtualized as a separate individual server with its own operating system.



lrives that
ıe DS8100 storage subsystem[2] are all configured as direct connects (that is, not using the VIO). The network connections for each LPAR are configured to use the Host Ethernet Adapter (HEA) that is available on the POWER6 server. A similar configuration was used in a recent SAS® 9.1.3 Enterprise Business Intelligence benchmark which showed how virtualization helped utilize server resources better to accommodate more users than was anticipated (500 users instead of the anticipated 350 users). To view the configuration used and the benchmark results refer to [1] in the resource section.

PowerVM increases resource utilization through dynamic load balancing and resource sharing. Traditional reference architectures look at each resource as a unique and separate entity. PowerVM changes that process so that it has a more holistic view of the environment.

The PowerVM features used in the study are as follows:

- **Micro-Partitioning™:** Allows processor resources to be assigned at a granularity of 1/100th of a core.
- **Shared processor pooling:** Allows for automatic non-disruptive balancing of processing power between partitions that are assigned to shared pools resulting in increased throughput.
- **Virtual I/O server:** Allows resource sharing between LPARs thus reducing costs by eliminating the need for dedicated network and disk adapters.
- **Live Partition Mobility:** Supports the movement of an active Linux and AIX partitions from one physical server to another compatible server without application downtime. This feature helps avoid application interruption for planned system maintenance, provisioning and workload management.
- **Workload Partitions (WPARs):** Allows virtualization of operating system environments within a single instance of the AIX operating system. WPARs secure and isolate the environment for the processes that the enterprise applications use.

Table 1 shows individual LPAR configuration for shared resources. The options to set desired, minimum and maximum values are needed when LPARs use the Micro-Partitioning feature of PowerVM.

System	Memory (in GB)	Processor-entitlement	Virtual processor	Weighting

[2] The DS8100 storage was configured as a RAID5. It has 16 arrays with eight disks each. Each array is seen by the AIX system as a logical disk drive. The logical volume configured on the disk is configured with a 64K stripe size.

	Online	Minimum	Maximum	Online	Minimum	Maximum	Online	Minimum	Maximum	Uncapped
lbmaix4 (metadata)	4	1	8	1	0.10	1.0	1	1	1	128
lbmaix5 (application)	16	1	32	6	0.10	8.0	6	1	8	128
lbmaix6 (middle tier)	16	1	32	4	0.10	4.0	4	1	4	128
VIO server	2	2	4	0.40	0.20	1.0	1	1	8	128

Table 1 Shared LPAR configuration table.

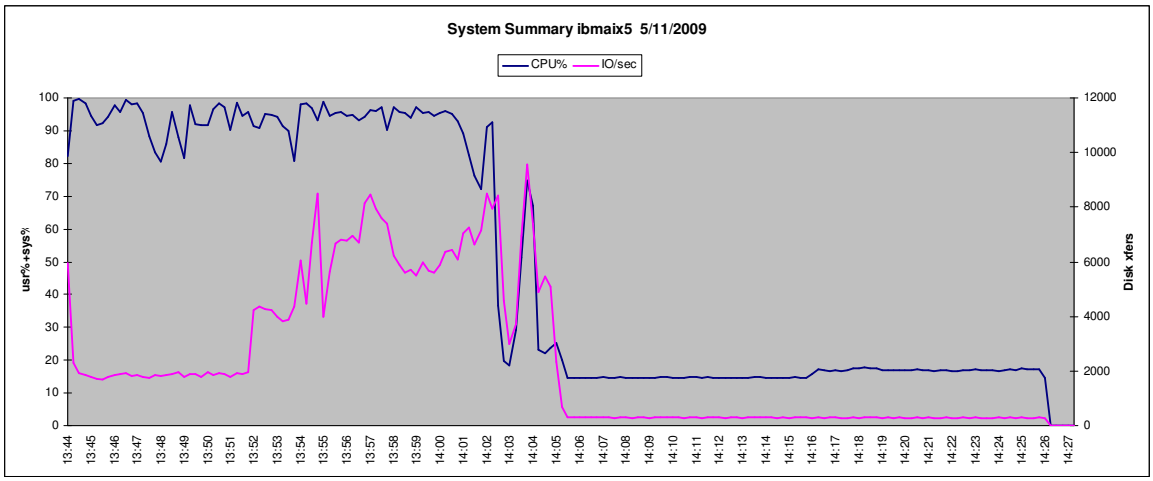
The PowerVM dynamic reconfiguration feature also allows for changing configuration without powering down the servers. This is useful during testing when some servers must be reconfigured to change their processor entitlements.

Note: The memory to processor configuration ratio for the tests was sized specifically for the workload used in this test. From a general perspective SAS recommends a memory to processor configuration ratio of four gigabytes of memory to one processor.

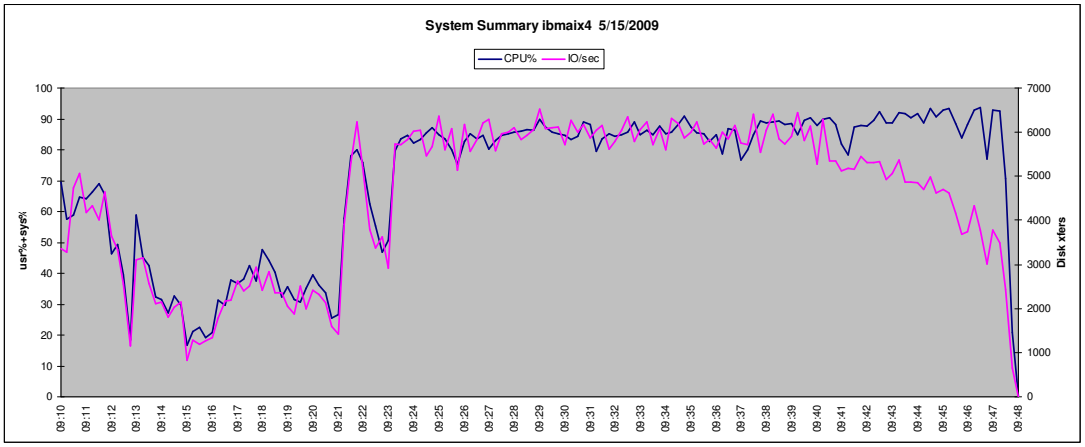
USE CASES

The main objective of this study is to show how SAS customers can use IBM's PowerVM technology to improve server utilization while maintaining service levels within their organizations. The study presents three use cases that were chosen to test different virtualization methods to see how they affect application run time and performance. Three types of workloads, discussed in the test suite section, are used for testing. The I/O- and compute-intensive workloads are used for the server consolidation and application-isolation use cases. The LoadRunner workload is used for the business continuity use case. Sizing and baseline measurements are taken before starting the test for each use case. The AIX 6.1 operating system was tuned based on recommendations from the *SAS for AIX 6.1: Tuning Guidelines* [2] document.

The compute-intensive workload is sized to run on an 8-processor system. Figure 3 shows the resource utilization of the compute-intensive workload. The workload uses close to 97 percent of the processors (represented by the blue line) during the first half of its execution. Disk transfers were in the range of 200 to 8000 disk transfers per second (represented by the red line) with the majority of disk transfers occurring in the 200 range.



red line) with the majority of disk transfers occurring in the 6500 range.



Fi

Fo
wo
is necessary to get baseline numbers for later comparison with shared- processor configuration. The WPAR reconfiguration is necessary to test the effects of WPAR overhead when both applications run on a single instance of the operating system. To limit the amount of data that will be discussed, only the run time numbers of the compute-intensive workload will be shown. I/O-intensive workloads will behave the same way as compute-intensive workloads as long as their I/O bandwidth remains unchanged during testing.

USE CASE 1: SERVER CONSOLIDATION

This use case demonstrates how IT personnel can consolidate multiple servers using PowerVM Micro-Partitioning and shared-processing pooling features. The IT personnel must guarantee a minimum amount of hardware resources to the various departments.

The use case used the ibmaix5 and ibmaix6 LPARs as shown in **Error! Reference source not found.**. Both are installed with AIX 6.1 running the SAS® foundation. The ibmaix5 LPAR with six fiber-channel ports run the compute-intensive workload while ibmaix6 with two fiber-channel ports run the I/O-intensive workload.

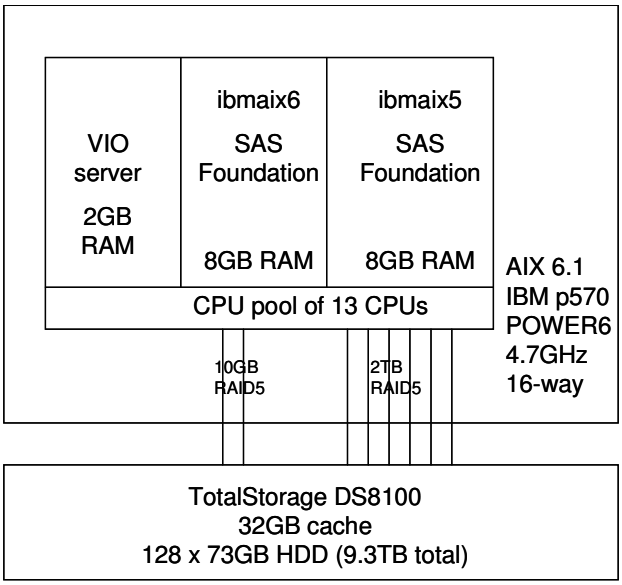


Figure 5 Consolidated servers using Micro-partitioning feature of PowerVM.

The Micro-Partitioning configurations for both LPARs are shown in Table 2. Each partition is assigned memory, processor and virtual-processor entitlements. Each partition is marked as uncapped so that idle processing resources from each LPAR can be used by the others when needed.

System	Memory (in gigabytes)			Processor-entitlement			Virtual processor			Weighting
	Online	Minimum	Maximum	Online	Minimum	Maximum	Online	Minimum	Maximum	Uncapped
ibmaix5	8	1	8	8	0.10	12	12	1	12	128
ibmaix6	8	1	8	4	0.10	12	12	1	12	128
VIO server	2	2	4	0.40	0.20	1.0	1	1	8	128

Table 2 Micro-Partition configurations for ibmaix5 and ibmaix6.

Before running both jobs on the Micro-Partitioned LPARs, a baseline run-time measurement is taken of the compute-intensive jobs that are running on a dedicated-processor configuration. This dedicated baseline run-time number is used to compare the run-time performance numbers measured of the same compute-intensive jobs running on the Micro-Partitioned configuration. Only data from the processor- intensive LPAR are shown since both compute- and I/O-intensive jobs running on the LPARs showed the same run-time characteristics on both environments.

Configuration	Dedicated 8 processors 16GB memory (A)	Shared 8 processors from pool of 12. DI LPAR busy (B)	Shared 8 processor from pool of 12. DI LPAR idle (C)	Shared 6 processor from pool of 10. DI LPAR busy (D)	Shared 6 processor pool of 10. DI LPAR idle (E)	Shared 4 processor from pool of 8. DI LPAR idle (F)
Cumulative Time	3:00:14	3:05:17	2:55:59	3:16:20	3:01:19	2:59:14
Percentage Difference	--	-2.73 percent	2.42 percent	-8.20 percent	-0.60 percent	0.56 percent

Table 3 Run-time numbers for dedicated and various Micro-Partitioned configured LPARs.

The run-time numbers shown in Table 3 are the cumulative user times for each procedure in the compute- intensive workload. The columns that show the cumulative times are labeled A through F. The percent difference row represents the difference between the runtimes of the various Micro-Partition configured LPARs and the dedicated configured LPAR. A negative percentage means the cumulative run time for that configuration ran for a longer time period and a positive percentage means the cumulative run time means that the cumulative run time ran for shorter time period. Column B and C show the run-time numbers with the two LPARs sharing a pool of 12 processors – with the compute-intensive LPAR (ibmaix5) configured to have a guaranteed processor entitlement of 8 (8/12) and the I/O intensive LPAR (ibmaix6; also designated as DI LPAR in the table) configured with a guaranteed processor entitlement of 4 (4/12).

The compute-intensive workload was run two separate times to test the effects of processor sharing between the two LPARs. The first run shown in column B presents the compute-intensive number when the I/O- intensive workload was also running on the ibmaix6. There is a -2.37 percent increase in run time, as compared to the dedicated LPAR. This negative number represents the small overhead that is associated with processor sharing. The second run shown in column C presents the compute-intensive cumulative number when the I/O-intensive workload is *not* running on the ibmaix6. In this case, the cumulative run-time number shows the jobs finishing in a shorter time period. Column C shows a small but positive effect of processor sharing in a Micro-Partitioned environment.

Column D and E shows the same scenario, but this time the processor pool is reduced from 12 to 10. The compute-intensive LPAR is configured a guaranteed processor count of six instead of eight; a 25 percent reduction. The ibmaix6 LPAR guaranteed processor count is left at four. Column D shows the run-time number for the compute-intensive workload LPAR while the ibmaix6 LPAR was busy running the I/O-intensive workload. Column D shows an -8.2 percent change in run-time, meaning it runs slower, which was expected because of the reduced processor resource. Depending on workload priorities, an 8.2 percent increase in run time may be more favorable in order to free up 25 percent of processor resource for higher priority work. Column E shows the run-time number close to even with the baseline number when the ibmaix6 LPAR is not running the I/O-intensive workload. Column E shows that processor sharing works as it is intended to.

Column F shows a similar effect as column E when the pool is reduced further by 50 percent with the compute-intensive LPAR; now only guaranteed four processors out of eight. With ibmaix6 not running the I/O-intensive workload, the compute-intensive LPAR uses the resources from the idle ibmaix6 LPAR and performs as if it had eight

guaranteed processors to use.

Note: As a best practice, LPARs configured to use processors from a shared processor pool running critical SAS applications must be configured to have the recommended minimum number of processors to guarantee against application resource starvation during heavy server utilization periods.

USE CASE 2 : APPLICATION ISOLATION

The use case demonstrates how IT personnel can consolidate multiple applications running in a single instance of the AIX operating system using PowerVM WPARs. In this scenario, the IT personnel must manage peaks in resource utilization and still meet service level agreements (SLAs). The use case demonstrates how WPARs can help optimize resources, especially where applications have different resource requirements.

A WPAR is defined as a solution for partitioning one AIX operating system instance into multiple environments. Each environment, called a workload partition or WPAR, can host applications and isolate them from applications executing within other WPARs. A WPAR is purely a software partitioning solution that is provided by the AIX 6.1 operating system. Workload Manager (WLM), which is part of the AIX operating system, provides the underlying technology that supports WPAR resource controls. A WPAR has no dependencies on hardware features. WPARs can be configured on both dedicated and micro-partitioned LPARs. For this use case, the WPARs are configured on a dedicated LPAR to eliminate any overhead associated with virtualized processors using Micro-Partitioning. For more information on WPARs refer to the IBM Redbook [3] in the reference section.

Figure 6 is a high-level diagram of WPAR1 and WPAR2, each configured within a single operating system instance. Both WPARs are configured using ibmaix5 that has 6 fiber-channel ports connected to the DS8100 storage unit.

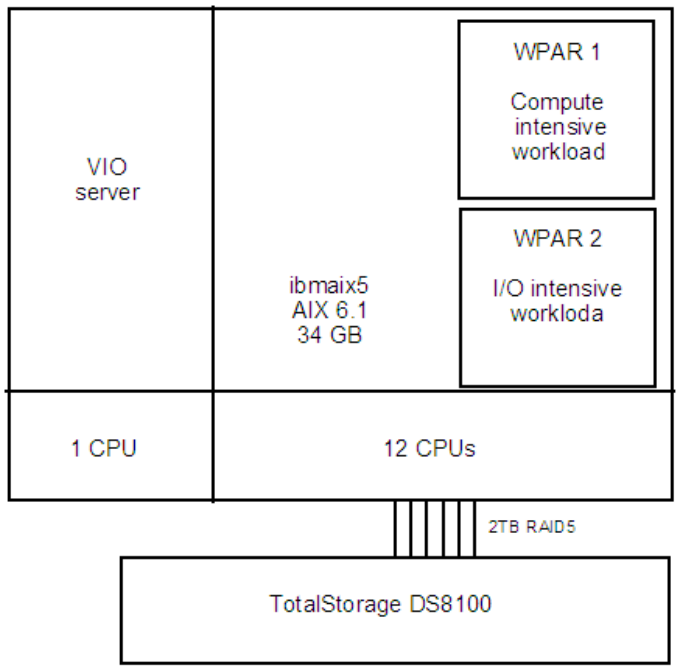


Figure 6 Configuration diagram of WPAR configuration for compute and I/O intensive workloads.

Table 4 shows the WPAR processor and memory configuration settings for each WPAR. The format, 50 to 66.7 and 100 percent (50-60,100 percent) refers to the processor limits configuration. The processor limits configuration shows the percentage of minimum to guaranteed processor resource allocation (50-66.7 percent) and the maximum resource allocation (100 percent) allowed for the WPAR. WPAR1 was configured two different ways for two test scenarios. The first configuration allocates a guaranteed 8 of the 12 processors (66.7 percent) with a maximum of 12 (100 percent) processors. And the second configuration allocated a guaranteed 8 of 12 processors with a maximum of 8 processors. WPAR2 was allocated with a guaranteed 4 of the 12 processors (33.33 percent) with a maximum of 12 processors (100 percent). Table 4 also shows which WPAR ran which type of workload. WPAR1 ran the compute- intensive workload

and WPAR2 ran the I/O-intensive workload.

	Processor Limits	Memory Limits
WPAR1 (Compute intensive)	50-66.7,100 percent 50-66.7.66.7 percent	0-50%,100 percent
WPAR2 (I/O intensive)	30-33.33,100 percent	0-50%,100 percent

Table 4 WPAR configuration for compute and I/O intensive workloads.

Table 5 shows the results of the WPAR tests. Before running the workloads on the WPARs, both workloads are run simultaneously in a dedicated 12-processor system. Column A shows the baseline number for the compute-intensive workload. This baseline number is taken when both compute- and I/O- intensive workloads are run on the same dedicated partition outside of the WPAR. This shows how the compute-intensive workload performs when the I/O-intensive workload is run at the same time and shares the resources of the same system.

Configuration	Dedicated 12- processors 16GB memory (A)	WPAR dedicated 12 processor CPU 50-66.7,100 % Memory 50%, 16GB, I/O workload running (B)	WPAR dedicated 12 processor CPU 50-66.7,66.7% Memory 50%, 16GB, I/O workload idle (C)	WPAR dedicated 12 processor CPU 50-66.7,100% Memory 50%, 16GB, I/O workload idle (D)
Cumulative Time	6:04:42	4:26:04	3:40:22	3:04:05
Percentage Difference	--	27.05 percent	39.58 percent	49.52 percent

Table 5 Run-time numbers of compute-intensive workload for WPAR configured system.

Column B shows the compute-intensive workload's cumulative run-time number when run on WPAR1 while the I/O-intensive workload runs on WPAR2. The number shows that separating the workloads to run on separate WPARs can help workloads run faster. In this case the compute-intensive workload runs 27.05 percent faster in a WPAR, separate from the I/O-intensive workload.

Why does the compute-intensive workload run faster when it runs on its own WPAR? As was mentioned in the first paragraph of this section, the AIX workload manager (WLM) is the underlying technology that WPARs are built on. The workload manager can prevent effective starvation of workloads with varying characteristics like high I/O or high processor utilization. WLM can separately manage how processes get percentages of processor and memory thus preventing applications from interfering with each other. By running each application in its own WPAR, applications are limited to using the resources allocated to them. In the case presented in column B, the compute intensive workload was given between 50-66.7 percent (equivalent of 8 processors) of the 12 available processor resources, thus effectively giving it more processors to run on. Contrast this to running the compute-intensive workload together with the I/O-intensive workload, in one system, and letting both applications contend for available resources.

Column C shows the compute-intensive workload numbers when the I/O-intensive workload is *not* run on WPAR2 (designated as *I/O workload idle* in the column text). Column C shows the numbers when the WPAR1 is configured to have eight guaranteed processors and a maximum of eight processors (50-66.67, 66.7 percent). When comparing column C with column B, notice that the run-time number for the compute intensive workload in column C runs faster than the workload in column B. This shows the effect of shared I/O in a WPAR environment. Even if WPAR1 is only configured to use a maximum of no more than 8 processors, the workload still runs faster compared to column B. The reason the workload runs faster is due to the exclusive use of the fiber-channel ports for its own I/O (that is the I/O-

intensive workload in WPAR2 is not running). **This shows that careful consideration of I/O bandwidth needs to be planned when running applications on WPARs in a single server. Make sure that there is enough I/O bandwidth available to accommodate all applications to maintain service levels.**

Column D shows the compute-intensive workload when the I/O-intensive workload is not running on WPAR2. Column D shows the numbers when WPAR1 is configured with 8 guaranteed processors and a maximum of 12 processors (50-66.7, 100 percent). The main difference between columns D and C is the 12 processor maximum configuration for column D. The result is clearly a better run time than that of column C. **This can be attributed to the fact that the compute-intensive workload was able to use more than its guaranteed eight processors to run the workload.**

In this use case, the data clearly shows the benefits of WPAR from an application isolation point of view. WPARs can isolate workloads and give them the needed processor and memory resources they need to perform better.

USE CASE 3: PLANNED BUSINESS CONTINUITY

The use case demonstrates how IT personnel can use PowerVM LPM as part of their organization's business continuity strategy. This should include provisions to maintain normal IT operations when there are planned outages or sudden surges of resource utilization because of increased demand. For planned outages, such as a scheduled system hardware shutdown for maintenance, a critical application running on that system can be moved to another system to keep the application running. For sudden surges of resource utilization due to increase in demand, the critical application can be moved to a different system that can provide the needed resources.

Figure 7 shows us how a logical partition is configured for LPM. After it is configured all it takes is a few clicks on the hardware management console (HMC) to move the partition from one system to another. Table 6 shows some LoadRunner statistics taken for this use case.

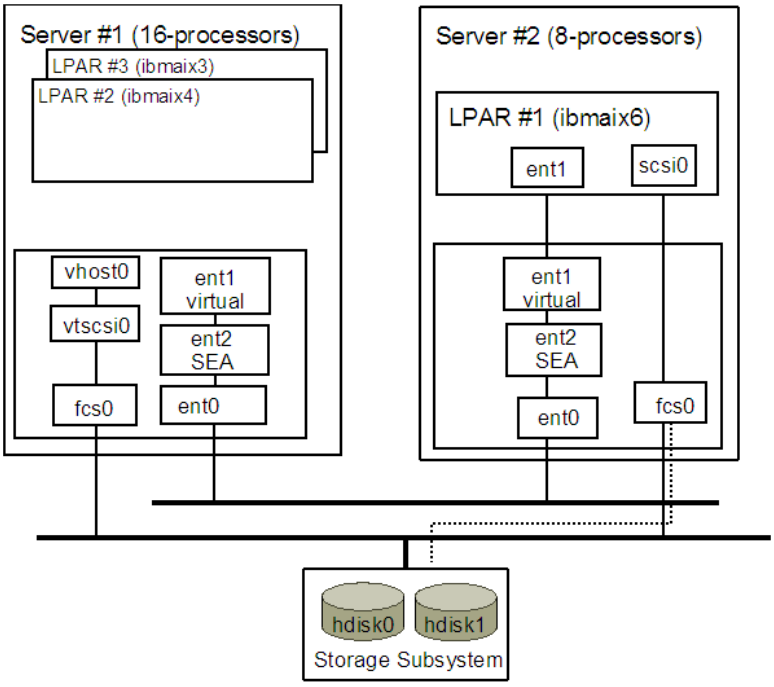


Figure 7 Starting configuration for LPM scenario.

For the purpose of this scenario the SAS Business Analytics Framework middle tier component is installed on an 8-processor system; the rest of the components (application server and SAS Metadata Server) are installed on a 16-processor system. The application and metadata server communicate through the internal bus of the 16-processor system using virtual Ethernet. The network communication between the application server and SAS Metadata Server is faster and more reliable than network communications through the local area network (LAN) which can be subject to degradation because of network traffic across the testing facility.

Column A shows the baseline numbers when running the LoadRunner test that is described in the *Describing the test*

suite section of this paper. Column B shows the numbers when the LPAR that runs the middle tier component is moved from the 8-processor system to the 16-processor system. Column C shows the percentage difference between the baseline and numbers during LPM operation. The negative number represents a decrease in throughput and hits-per-second which can be expected during LPM operations. After the middle tier LPAR is moved, the 8-processor system is ready for system maintenance.

	Baseline 8-way (A)	**Move App LPAR 8- to 16- way (B)**	**percentage difference (C)**	**Baseline 16-way (D)**	**Move App LPAR 16 to 8 way (E)**	**% diff (F)**
Maximum Vusers	25	25		25	25	
Total throughput	1057128085	981089132	-7.19 percent	1269504448	1242739881	-2.11 percent
Avg throughput (bytes per sec)	1594462	1479772	-7.19 percent	1914788	1877251	-1.96 percent
Total hits	74553	69147	-7.25 percent	89658	87750	-2.13 percent
Average hits per second	112.448	104.294	-7.25 percent	135.231	132.553	-1.98 percent

Table 6 LoadRunner measurements before and after moving the application LPAR with LPM.

Now that the middle tier LPAR is running on the 16-processor system, another baseline is captured and the numbers are shown in column D. Notice that the baseline numbers on the 16-processor system is 16.73 percent better than the baseline numbers when running on the 8-processor system. Although the processor and memory configuration of the middle tier LPAR was not changed, the network communications between the middle tier and the other LPARs are now going through the internal bus, as opposed to the wired LAN – the reason why the baseline numbers are significantly better. As mentioned earlier in this section, network communications through the internal bus runs faster than communications through the external local network. Additionally, if more processors and memory are needed by the middle tier LPAR, the 16-processor system has extra resources that can be dynamically assigned to the middle tier LPAR.

Column D shows the numbers when the LPAR is moved from the 16-processor system back to the 8-processor system. The difference between the numbers in columns E and D show the overhead that is associated with the LPM move. This use case shows how LPM can be used to move critical business application component from one system to another for planned system maintenance. It also shows how an LPAR that needs more resources can be moved to a different system that has spare resources that can be dynamically added to the LPAR.

SUMMARY

The different features on IBM PowerVM offer SAS customers several options to reduce costs, improve utilization and responsiveness, and increase flexibility. Although server consolidation remains the number one reason to use virtualization, other reasons such as application isolation and planned business continuity are not far behind. SAS customers need to evaluate their application profile and IT operational needs to find the right virtualization feature they need to use.

For applications that need total isolation from other applications, a dedicated LPAR or a shared-processor LPAR are two options to consider. Both LPAR solutions have hardware- and firmware-based technology that offer the best system isolation from a processor, memory and I/O point of view. The one difference that a shared-processor LPAR has is the added ability to dynamically share fractional processor capacity with other LPARs in the same system. The shared-processor LPAR configuration offers improved utilization and responsiveness as compared to the dedicated LPAR configuration for a very small overhead penalty (around 2.73 based on the application profiles used in the testing).

For SAS customers who run several workloads simultaneously in a single instance of AIX 6.1, WPARs can help isolate the applications from other applications. WPAR is a software-based technology that allows SAS customers to allocate processor and memory resources to individual applications. With WPARs, SAS customers can easily allocate more system resources to the application that have higher priority to meet service level agreements (SLAs).

Live Partition Mobility allows SAS customers to move LPARs from one system to another with minimum impact to business operations. LPARs can be moved from a system that is scheduled for hardware maintenance. Or a resource-starved LPAR can be moved to a system that has spare resources to help with the LPAR's resource requirements.
All of the features tested in this paper are applicable to the majority of software applications that run on AIX. **Careful planning and consideration must be done to get the full benefit of these features. As shown in the use-case sections, some small overhead can be expected in exchange for reducing cost, improving utilization and increasing flexibility.**

REFERENCES

[1] Mendoza, A., Pu, J., 2009, Deploying *SAS® Enterprise Business Intelligence in an AIX virtual environment,* Austin, TX: IBM Corporation

[2] Mendoza, A., 2009, *SAS for AIX 6.1: Tuning guidelines,* Austin, TX: IBM Corporation, http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP101529

[3] Blanchard, B., et al., *Introduction to Workload Partition Management in IBM AIX Version 6.1,* IBM Redbook: IBM Corporation, http://www.redbooks.ibm.com/abstracts/sg247431.html

[4} Bailey, J., et al., *IBM PowerVM Live Partition Mobility*, IBM Redbook: IBM Corporation, http://www.redbooks.ibm.com/abstracts/sg247460.html?Open

ACKNOWLEDGEMENTS

Thanks go to the following IBM people who reviewed this paper for technical accuracy: Frank Battaglia and Harry Seifert of IBM, Darrell Barton of SAS. Thanks to Ralph Nissler of IBM and Rachel Wynn of SAS for editorial help.

SPECIAL NOTICE

The information contained in this document has not been submitted to any formal IBM test and is distributed "AS IS". While each item may have been reviewed by IBM for accuracy in a specific situation, there is no guarantee that the same or similar results will be obtained elsewhere. The use of this information or the implementation of any techniques described herein is a customer responsibility and depends on the customer's ability to evaluate and integrate them into the customer's operational environment. Customers attempting to adapt these techniques to their own environments do so at their own risk.

IBM is not responsible for printing errors in this publication that result in pricing or information inaccuracies.

There is no guarantee these measurements will be the same on generally available systems. Actual results may vary. Users of this paper should verify the applicable data for their specific environment.

CONTACT INFORMATION

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

Author: Alfredo Mendoza
E-mail : mendoza1@us.ibm.com

The following terms are trademarks of International Business Machines Corporation in the United States, other countries, or both: IBM, AIX, AIX 5L, AIX 6, Micro-Partitioning, PowerVM. A full list of U.S. trademarks owned by IBM may be found at: http://www.ibm.com/legal/copytrade.shtml.

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.

Other brand and product names are trademarks of their respective companies.