Not All Server Virtualization Solutions Are Created Equal

Participant Names:
Andre Metelo
metelo@us.ibm.com

IBM SWG Competitive Project Office

2010-08-13

Table of Contents

Executive Summary

Virtualization enables more efficient data centers through the use of fewer and more utilized servers. This paradigm changes the way servers and applications are deployed. It enables significant reductions in the Total Cost of Ownership (TCO) of servers through reduced requirements of floor space, power, cooling, administrative efforts, hardware acquisition, and several other factors.

Virtualization itself is not a new technology. In fact, IBM has been offering virtualization solutions for over 40 (forty) years. Power Systems have been taking advantage of virtualization for almost 15 (fifteen) years. However, what is commonly overlooked by many individuals in the IT industry is the fact that not all virtualization solutions provide the same capabilities. In other words, not all virtualization solutions are created equal.

Even if a solution fully complies with the basic requirements to be defined as a virtualization solution, it does not mean that it provides the robustness, scalability, flexibility and ease of management that is required by today's data centers.

This paper will show the unique features and implementation differences that make PowerVM the most robust, scalable and flexible virtualization solution available in its market space.

The paper will show that PowerVM is the only solution that has the scalability to provide to an enterprise the following attributes and functions that are required in the modern data centers:

- Robust foundation for a solid and secure virtualization solution,
- Ability to scale from the smallest virtual server all the way to the full system images using the same techniques and management methodology,
- Uncompromised flexibility to adapt to the ever changing requirements from applications and services provided by data centers.

This paper will make direct comparison to how the PowerVM provides a solution that is far more scalable and flexible than the virtualization solutions offered by Oracle, HP, and VMware.

An Unbeatable Foundation - The Importance of a Good Hypervisor

Defining server virtualization can be very tricky. However, KernelThread.com has been able to create a good definition:

> *"Virtualization is a framework or methodology of dividing the resources of a computer into multiple execution environments, by applying one or more concepts or technologies such as hardware and software partitioning, time-sharing, partial or complete machine simulation, emulation, quality of service, and many others."*

This is a very loose definition, and thus, the marketplace has created a multitude of different solutions that comply with it. Amongst these we can mention: complete machine emulations like Power Systems Logical Partitions, Oracle VM Server for SPARC and HP Hardware Partitions (npars) and VMware Virtual Machines; workload managers like IBM AIX System Workload Partitions or Oracle Solaris Containers; partial machine emulations like Mac-on-Linux, WINE, and PowerVM Lx86.

The common element present in all virtualization solutions examples listed is the masking of server resources – processors, memory, I/O adapters, disks, operating systems – from the server users. The server administrator divides a physical server into multiple virtual servers. These virtual servers are sometimes called partitions, but they are also known as virtual machines, guests, instances, containers or emulations depending on the technology used.

Although, there are several virtualization solutions available today, only a few of them make use of a Hypervisor[1]. The presence of a Hypervisor in a virtualization solution is fundamental to guarantee the isolation, security and flexibility in virtualized environments.

Technically speaking, a Hypervisor is a virtualization technique that allows multiple operating systems to run in a host computer at the same time. Hypervisors guarantee full isolation between the resources allocated to each of the operating systems. Without a Hypervisor, the resource isolation for processor, memory, I/O cards, and even Operating System Kernels is either non-existent or reduced. The virtualization environment will have shared resources that become contention points that impact performance and availability of the virtual servers. In some scenarios, this can cause system wide crashes that will bring all the workloads running in the server down.

Solutions like Solaris Containers, HP-UX Secure Resource Partitions (SRP) and even AIX's Workload partitions have their place in a data-center. However, the fact that they all share a single OS kernel for the workload virtualization makes them a riskier proposition on the performance and availability front for critical business applications. Without the complete isolation provided by a good Hypervisor, a problem in the OS kernel of the solutions mentioned above will bring down all the workloads. This may be acceptable for some environments, but not good enough for others. For applications that require the best performance and isolation, a Hypervisor is a must.

[1] The term Hypervisor originated in IBM's CP-370 reimplementation of CP-67 for the System/370, released in 1972 as VM/370.

Additionally, not all Hypervisors are created equal. Different implementations of a Hypervisor work in different ways. These differences are key to the performance, stability, and functionality provided by the Hypervisor.

In fact, Hypervisors can be divided into two categories:

- **Type 1 Hypervisor** (also known as native or bare metal Hypervisor) is similar to an operating system control program. Classic examples of Type 1 Hypervisor implementations are CP/CMS, z/VM, and VMware ESX Servers. Variations of this are Hypervisors that are embedded in the platform's firmware as IBM PowerVM and Oracle VM Server for SPARC.
- **Type 2 Hypervisor** (also known as hosted Hypervisor) is a piece of software that runs as an application within an operating system. In fact this type of implementation puts the virtual machine one level further away from the real physical hardware, adding additional management overhead when compared to a Type 1 Hypervisor. Current examples of Type 2 Hypervisors are VMware Workstation, Xen, Microsoft Virtual PC and HP Integrity Virtual Machine (IVM).

Figure1 shows a graphic view of the differences between a Type 1 Hypervisor (Bare Metal) when compared to a Type 2 Hypervisor (Hosted).

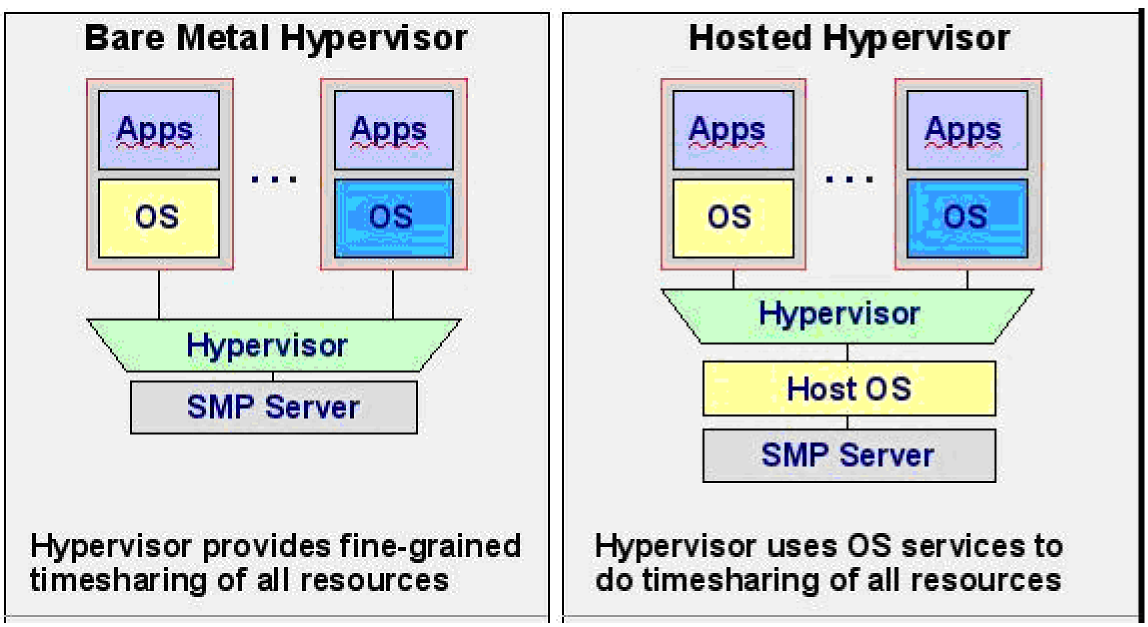

Figure 1 – Graphic differentiation between a Bare Metal and a Hosted Hypervisor

One key difference between Type 1 and Type 2 Hypervisor is the complexity and depth of the software stacks involved. Performance tests show that Type 1 Hypervisors provide better performance for the virtualized services and environments. Furthermore, when the Hypervisor is implemented in the platform firmware itself – see Figure 2 - these performance gains tend to be even more significant due to additional hardware integration and assists that are included during the architecture design phase[2].

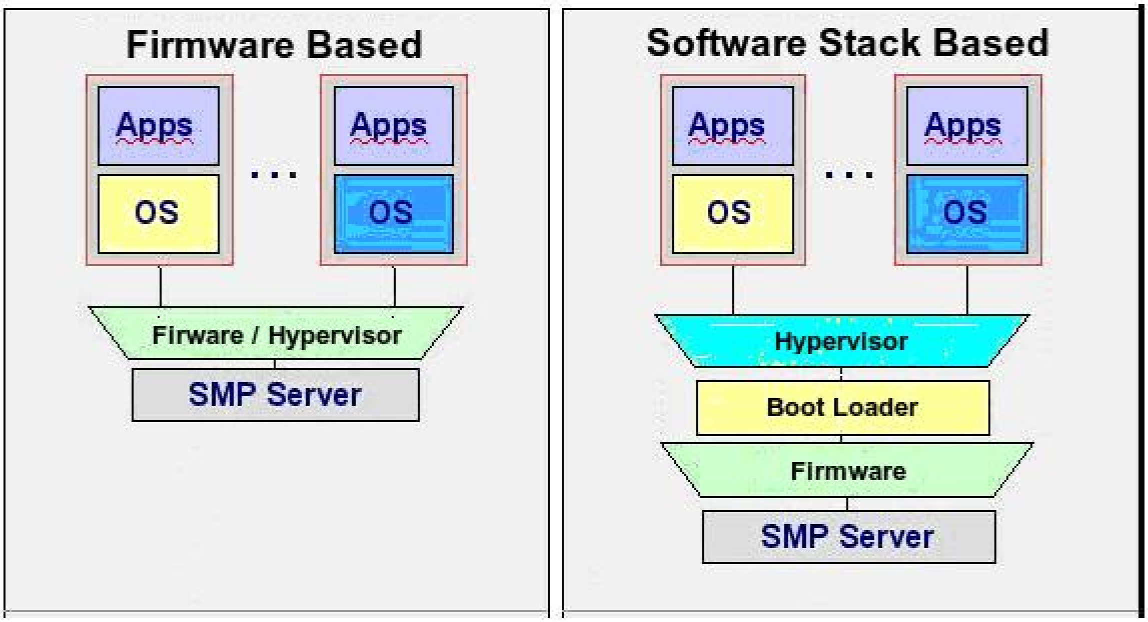

Figure 2 – Graphic differentiation between a Firmware and Software Stack based Hypervisor

To further compound the problem Type 2 Hypervisors present several security concerns[3] to IT staffs. Because this type of Hypervisor runs on top of an off-the-shelf operating system (OS), the OS itself has the potential to be the weakest link in the security chain. The techniques used to acquire unauthorized privileged access to the off-the-shelf OS can be applied against the host OS. When privileged access to the host OS of a Type 2 Hypervisor is obtained by unauthorized individuals, the Hypervisor will be under the control of said individuals. From that point on they would be able to control all virtualized services and environments hosted by that Hypervisor instance.

In fact, in 2009 multiple exploits were developed that allows the guest OS to compromise the host OS through Hypervisor related weaknesses. The first exploit exists in the Xen Hypervisor and through a virtual video buffer device code from a Guest OS, the Guest OS could access the Host OS. The second exploit used a similar technique to compromise a different Hypervisor - VMware Workstation. It demonstrates the ability for a Guest OS to execute an application at the Host OS.[4]

---

[2] http://www.webperformanceinc.com/library/reports/LoadTestingVirtualizationPerformance/ indicates that VMware running as a Type 2 Hypervisor can suffer a 43% performance penalty. Meanwhile IBM internal measurements of a POWER4 system running with the Hypervisor active incurs a penalty of around 3% in the over all system performance. Since the POWER5 all POWER Processors always have the Power Hypervisor active.
[3] For a more in depth discussion about security in virtualized environments please read the following white paper https://w3-03.ibm.com/sales/competition/compdlib.nsf/SearchView/ACF7306272DC153F0025748300707828?Opendocument
[4] Reference to the 1st exploit can be found at http://invisiblethingslab.com/pub/xenfb-adventures-10.pdf. Meanwhile a video of the second exploit can be found at http://www.immunityinc.com/documentation/cloudburst-vista.html.

However, just like the difference between a Type 1 and Type 2 Hypervisor, not all Type 1 Hypervisors are implemented the same away. In fact, Hypervisors implemented as a software stack added on top of the system hardware in place of an Operating System can be less secure and suffer from performance problems.

For example, on the security front, it is possible to install a modified fileset on the hard disk where the Type 1 Hypervisor resides. Then, the modified fileset can be used to grant access to the administration stack of a software-based Hypervisor. As soon as the modified code is executed, the Type 1 Hypervisor would be compromised.[5] This is where the advantage of Type 1 Hypervisors implemented at hardware or the firmware level provides additional security advantages. It is virtually impossible to remotely install a modified fileset into the EPROMs (Erasable Programmable Ready Only Memory) of systems like IBM Power Systems.

A firmware-based Type 1 Hypervisor provides further advantages in performance, workload isolation, scalability, flexibility and security. Although the benefits are many, today there are very few Hypervisor implementations for UNIX and distributed servers that are implemented as firmware-based Type 1 Hypervisor. Currently, IBM PowerVM, and Oracle VM Server for SPARC are most prominent firmware based Type 1 Hypervisors in this market segment.

[5] This is a common technique used to invade Systems that date back from the early days of networked computers. An invader tries to install a piece of software that will be executed in super-user mode granting him or her privileged access to the system. This is similar to a Trojan horse attack in the PC environment.

Only PowerVM Scales for Any Workload, Large or Small

When deploying new applications and IT related services, it is very common for enterprises to "bump" into capacity limits. To further complicate things, in the modern business environments there are different workloads that demand significantly different amounts of resources to execute properly. In order to respond to such requirements, the underlying infrastructure has to be able to scale from tiny images all the way to the full system capacity.

This ability of a system, a network or a process to handle such variations on the resources needed to manage workloads in a graceful manner has become a key requirement. Scalability, as a property of systems, is generally difficult to define[6] and in any particular case it is necessary to define the specific requirements for scalability on the dimensions which are deemed important. Nonetheless, more then ever, scalability is a highly significant issue in data centers. It affects the technical specification and requirements of servers, databases, routers, networking and many other IT applications.

In the classical data center environments, before virtualization became a common practice, enterprises typically handled the scalability requirements through two paths:

- **Scale vertically (scale-up)** means to add resources to a single node in a system, typically involving the addition of CPUs or memory to a single computer.
- **To scale horizontally (scale-out)** means to add more nodes to a system, such as adding a new computer to a distributed software application. An example can be the use of hundreds of small computers configured in a cluster to obtain aggregate computing power that exceeds single traditional scientific computers.

The problem with these models is that they tend to be too wasteful. Typically the use of either technique causes waste of precious physical resources - processing capacity, power and real state. To further complicate things, a scale-out solution tends to impose a much heavier workload on system administrators as well as require more complex application models.

With the rise of virtualization techniques, these models are becoming obsolete, and are going through a migration process to enable significant improvements in data center resource utilization as well as administration productivity gains.

As more virtualized solutions are deployed by enterprises of diverse sizes, statistics have shown that vertical scaling systems are able to more effectively use this new technology to deliver higher resource utilization levels and reduced system administration efforts. This happens because scale-up solutions naturally provide more resources for the Hypervisor to share amongst the Operating System and Application modules that need to be deployed.

---

[6] See Mark D. Hill, 'What is scalability?' in ACM SIGARCH Computer Architecture News, December 1990, Volume 18 Issue 4, pages 18-21, (ISSN 0163-5964) and Leticia Duboc, David S. Rosenblum, Tony Wicks, 'Doctoral symposium: presentations: A framework for modeling and analysis of software systems scalability' in Proceeding of the 28th international conference on Software engineering ICSE '06, May 2006. ISBN 1-59593-375-1, pages 949 - 952

As a consequence, the scalability of a virtualization solution can be measured by its ability to be deployed on systems that can smoothly and easily be set up to handle very little virtual servers, and allow the growth of it to take advantage of the full capacity of the largest physical servers in the product line. However, not many virtualization solutions available today are able to achieve this requirement. Going one step further, not all vendors are able to provide a solution that applies to its complete product line and takes advantage of their Hypervisor-based virtualization offering.

Table 1, shows the CPU allocation limits for the Hypervisors offers from the main virtualization providers on the UNIX and distributed server market today. It also shows how these offerings scale within these vendors server product line.

	IBM	**Oracle**	**HP**	**VMware**
Hypervisor	PowerVM	Oracle VM Server for SPARC	Integrity Virtual Machines	vSphere4.1
Leading Operating System	AIX	Solaris	HP-UX	Windows/Linux
Processor Architecture	POWER	SPARC[7]	ITANIUM	X86
Max Number of Cores in Product Line	256	256 (CMT & SPARC64)	128	64
Smallest CPU Allocation for a Virtual Server	0.1 cores	1 thread (1/8th of a core)	0.05 cores	1 VP (Reserve CPU Cycles in MHz)
Largest CPU Allocation for a Virtual Server	256 cores	32 cores (CMT only)	8 cores	8 cores
Maximum Number of Virtual Servers	1000	128	256	320

Table 1 – CPU scalability of different leading virtualization solutions in the UNIX and distributed server market

The data from the table clearly shows that only PowerVM is able to provide a virtualization environment that can be setup from the smallest requirements all the way to a single large virtual server that takes over the capacity of the largest SMP system offered for that platform. Meanwhile, the solutions from all other vendors reach a point where they have to either change the virtualization solution or completely abandon it in order to continue to scale-up their deployments within the same platform. This is a serious limitation in a virtualization solution, as it makes the system administration more complex, and the administrators are forced to learn different deployment, performance tuning, management and end of life techniques to deal with the whole deployment scope of the solution.

However, it is not only about the number of cores in a system. After all, not all cores have the same performance characteristics. When taking into consideration the processor performance, it becomes

[7] The SPARC processors currently are forked into 2 processor families that are based on the same ISA, but have a completely different set of capabilities and performance characteristics – SPARC64 and CMT. For this comparison purposes we are looking at the maximum capacity of the combined SPARC product line and how many of them can be leveraged for Oracle's Hypervisor based virtualization solution.

clear that a 64-core POWER7 server is able to handle a much larger workload than an equivalent server with the same number of cores from the competition. Figure 3 shows how the platforms that were discussed in Table 1 compare to each other in terms of full server performance when deployed in systems with different number of processor cores from their respective product lines.

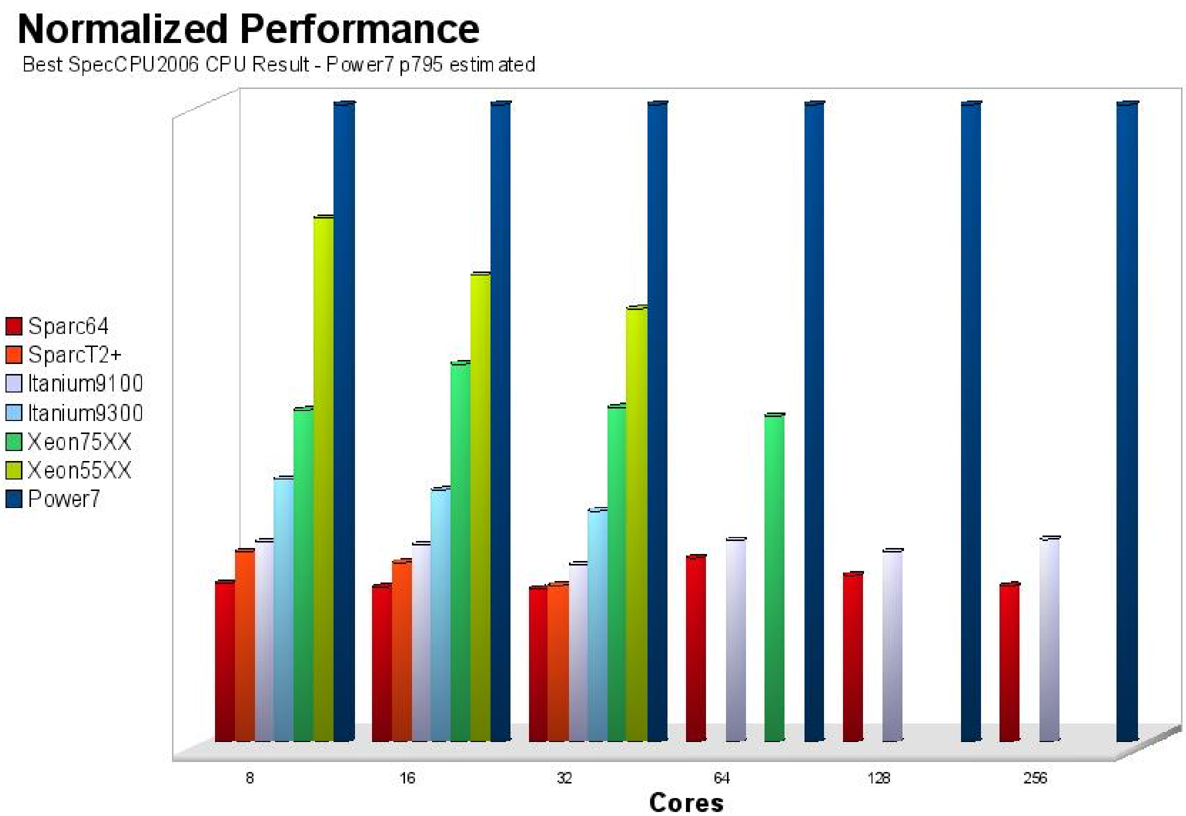

Figure 3– Normalized system performance comparison for UNIX and distributed servers available in the market

Additionally, the scalability problems from competing server virtualization solutions are not limited to the processor capacity and the number of cores of the virtual server deployments. Memory scalability is a problem as well. There are severe limitations on the solutions from the vendors compared. PowerVM is the only solution that allows for a virtual server to scale from a meager amount of memory – 128MB currently – all the way to the full capacity of the system – 8TB on the largest servers available today. Meanwhile the competition has a considerable lower limit on the maximum amount of memory that can be allocated to a virtual server – HP IVM only allows 64 GB; Oracle VM Server for SPARC only allows 128GB; VMware vSphere4 .1 only allows 255GB. Again, the solutions from the competition have a scalability problem.

A similar situation happens with I/O resources as well. Once again the virtualization solutions from the competition have severe limitations on how an I/O resource can be allocated to the different virtual server. The solutions from other vendors are limited by a low count of virtual adapters per virtual servers and the inability to dedicate a physical adapter to a specific virtual server. PowerVM is the only solution that allows a system administrator full control of how to deploy the I/O resources on a system. From fully virtualized, to fully dedicated and literally, anywhere in between.[8]

In the end, only PowerVM is able to deliver the full scalability capabilities needed to deal with applications that require a very small amount of resources at the same time that other applications require a large portion of the server resources. The solutions offered by the competition come up short in their ability to scale. They force the system administrators to learn different procedures, performance tuning techniques and daily maintenance techniques in order to deploy workloads that have such a wide range of requirements. In a data center that is struggling to optimize its resources to the maximum, this is unacceptable.

[8] Although Oracle VM Server for SPARC is able to dedicate some adapters to specific Domains, currently only four domains ara allowed to have control of the physical adapter present in the system

Flexible, Adaptable, Dynamic: PowerVM Is the Leading Virtualization Solution

In the context of a dynamic and modern data center, if the scalability of a virtualization solution is considered one side of a coin, then the other side of this coin is the flexibility features of the solution.

For any data center, scalability is a key characteristic of a virtualized solution. But, when taking into consideration the need of a virtualized deployment to handle the changes and adjustments that are required by applications and services, then scalability is not the only element that has to be taken into account. Typically these applications and services require more than just the ability to scale while deployed in a modern data center.

In order to easily change the deployment performance parameters of the infrastructure where applications and services run, data centers need a solution that supports online change of the deployment specifications. In other words, they need a solution that allows changes to the virtual servers' specifications easily, without any downtime and that does not cause any performance problems.

Let us call this characteristic of a virtualization solution *Flexibility*. Flexibility gives a virtualization solution the ability to respond to the workloads fluctuations that are expected in modern applications and services. In the end, the solution's flexibility to dynamically change the running environment is fundamental for enterprises to fully become a dynamic and modern data center.

Without a rich set of flexibility related features in a virtualization solution, any data center deployment will cause extra work to its system administrators. Without built-in flexibility all a data centers applications and services will suffer through outages in order to allow the operation teams to fully optimize the resources utilization in response to the multiple workload fluctuations.

Unlike the scalability considerations, flexibility features are not a simple and direct feature and number comparison. Flexibility is not about executing code faster, loading data in the shortest amount of time, completing transactions with the minimum amount of resource consumption or executing backups with the least amount of data transferred.

Flexibility is about enabling the system administrators to control how, where and when they assign resources – CPU, memory, I/O adapters – to different workloads with little or no effort. Most importantly this needs to be done with zero downtime and no performance impact for the users.

However, just like with the scalability comparison, when a deeper analysis of the flexibility features is done, similar results appear. Once more the Hypervisors compared do not provide capabilities similar to PowerVM. Many of them miss key features that are a "must have" for the dynamic infrastructure that is required by a modern data center.

Table 2, shows key flexibility related capabilities that are important for a solution to support. Without support for many of these features, a virtualization solution will lack the ability to fully optimize the resource allocations by the system administrator.

	IBM	Oracle	HP	VMware
Hypervisor	PowerVM	Oracle VM Server for SPARC	Integrity Virtual Machines	vSphere4.1
Host OS Required to Implement Virtualization	No	Yes (Solaris becomes control domain)	Yes (HP-UX is the host for the Hypervisor)	No
Add / Remove CPU without reboot	Yes	Yes	Yes	Add Only
Dedicate Processor	Yes	Yes	No	Limited[9]
Processor Sharing / Over Commitment	Yes	Limited[10]	Yes	Limited [11]
Add / Remove Memory without reboot	Yes	No	Yes	Only Add
Memory Over commitment	Yes	No	No	Yes
Memory Compression	Yes	No	No	Yes
Add / Remove Dedicated I/O without reboot	Yes	Limited [12]	Not Supported[13]	Not Supported[14]
Add / Remove Shared I/O without reboot	Yes	Yes	Yes	Yes
Live Migration	Yes	Limited[15]	Limited[16]	Yes
Combine full OS Partitions with Workload Management	Yes (AIX WPARs)	Yes (Containers)	Yes (SRP)	No
Cross Platform Code Execution	Yes (Linux x86 Binaries)	No	No	No

Table 2 – Flexibility related capabilities of the leading virtualization solutions in the UNIX and distributed server market

In the end, just like when comparing the scalability capabilities, not all vendors are able to provide a solution with all the features needed in a data center to fully enable resource optimization with zero downtime.

[9] When reserving CPU capacity utilization will be lower due to the scheduling design

[10] Requires a complex scripting language controlled by the administrator in order to do the sharing of processors. Over-commitment is sill impossible. You can only allocate the limit of threads.

[11] Expect VM processes to behave and not hog CPU while the core is not in reserved mode

[12] Not all adapters can be dynamically changed

[13] AVIO (Accelerated Virtual I/O) is not the same as dedicating an adapter to a VM. It is just a shortcut to reduce the amount of layers that a packet has to cross in order to get from the Application to the Physical adapter.

[14] Although vSphere4.1 allows for some level of direct access for physical adapter when using systems that provide support for Intel's VT implementation, the allocation/deallocation issues are still present

[15] Requires domain reduction to 1 thread, and reboot to re-enable CPU relocation

[16] Limitations to the OS supported and I/O devices must be disconnected before migration

Only PowerVM is able to deliver a full set of features that enables an enterprise to respond to the ever more common variable workloads that can quickly change full order of magnitudes within minutes while allowing the administrators the freedom to choose where to run the workloads without adding the burn of too many manual interventions and/or downtime of the applications and services managed by them.

Conclusion

Virtualization is being extensively used by enterprises to modernize their IT infrastructure. In fact McKinsey quarterly has indicated that since 2006, 86% of the CIOS are looking at consolidating their datacenters. And virtualization is the key technology to enable an extremely high level of consolidation and resource utilization.

IT departments have to understand and properly plan for virtualization deployment. They have to be able to guarantee that the infrastructure can scale and respond quickly to the changes in workloads deployed in order to avoid the collapse of their applications and services. An improperly capable, planned or implemented virtualized environment will be a huge liability.

A solution based on an architecture that is designed to be robust, scalable and flexible is a must. IBM's PowerVM is the only solution that combines these attributes and delivers the best virtualization solution in the UNIX and distributed server market segment.

This paper has shown that only PowerVM:

- Provides a robust firmware based Type 1 Hypervisor
- Has a simple and unified solution that scales from 0.1 core and 128MB of memory all the way to 256 cores and a 4TB of memory
- Has the ability to freely add and remove resources – CPU, Memory, I/O adapters – from a virtual server through a simple, unrestricted and unified process
- Allows the freedom to choose where to run the applications and services deployed in the data center.

The bottom line is: Virtualization requires a solid architecture as a foundation where proper policies and best practices are built upon by an enterprise in order to maximize its investment in information technology. There is not a more solid, scalable, and flexible foundation in the UNIX and distributed server market than the IBM Power Systems and when you combine that with the uncompromised capabilities of IBM PowerVM you get an unbeatable combination.

References

Shoumack, Tony, Beginners Guide to LDOMS: Understanding and Deploying Logical Domains – for Logical Domains 1.0 Release, Sun BluePrints, 2007

Logical Domains 1.2 Release Notes, Sun MicroSystems, 2009

Logical Domains 1.3 Release Notes, Sun MicroSystems, 2010

HP, Overview: HP Integrity Virtual Machine s (Integrity VM), 2010, http://h18000.www1.hp.com/products/quickspecs/12715_div/12715_div.HTML

HP Integrity Virtual Machines Release Notes Version A.03.50, HP, 2008

HP Integrity Virtual Machines Release Notes Version A.04.20, HP, 2010

Configuration Maximums – Vmware vSphere 4.0 and vSphere 4.0 update 1, VMware, 2009

Performance Best Practices and Benchmarking Guidelines, VMware, 2009

IBM, Advertisement, 2010, http://www-03.ibm.com/systems/power/

IBM System p Advanced POWER Virtualization (PowerVM) Best Practices, IBM Redbooks, 2009

PowerVM Virtualization on IBM System p: Introduction and Configuration Fourth Edition, IBM Redbooks, 2010

Gerry Kovan, CPO Lunch and Learn - Oracle-Sun Attack: Oracle's Virtualization Technology - Nothing New Under the Sun, IBM CPO, 2010

About The Author

Andre Metelo is a Certified IT Power System Specialist in the IBM Software Group Power System Competitive Project Office in the United States.

Andre has two and a half years of experience in Competitive strategy as well as ten years of experience in the Power Systems field and three and a half years in the networking field where he developed, planned and implemented multiple complex solutions.

Andre has worked at IBM for 15 years. His areas of expertise include Power System hardware, Linux, AIX, TCP/IP networking, and ATM. He holds a degree in Industrial Engineering from the Universidade Federal do Rio de Janeiro. He has published the Redbook Tuning Linux OS on System p The POWER of Innovation (SG24-7338-00), articles in the CPU magazine in Brazil, as well as several IBM internal and externals Whitepapers.

Special thanks to the following people who helped in the development of this white paper:

Randy Kuseske

Copyright