Pete's All Things Sun: Comparing Solaris to RedHat Enterprise and AIX—Virtualization Features

PETER BAER GALVIN

Peter Baer Galvin is the chief technologist for Corporate Technologies, a premier systems integrator and VAR (www.cptech.com). Before that, Peter was the systems manager for Brown University's Computer Science Department. He has written articles and columns for many publications and is co-author of the *Operating Systems Concepts* and *Applied Operating Systems Concepts* textbooks. As a consultant and trainer, Peter teaches tutorials and gives talks on security and system administration worldwide.
Peter blogs at http://www.galvin.info and twitters as "PeterGalvin." pbg@cptech.com.

In our last episode, there was a battle royal comparing the general features of Solaris, RHEL, and AIX. In my view, there was no clear winner. Each had pros and cons, but they were all quite usable and feature-rich in the end. However, the battle does not end there. One area, perhaps the most important differentiator, was left for this, the final round.

Virtualization is extremely important to datacenter managers, providing a valuable tool to increase system use, reduce the number of systems, reduce costs, and increase manageability. It's also a complex topic, with technologies and features varying greatly, and few clear "best" approaches. Further, software vendors are increasing the importance of virtualization, in many cases limiting the licenses needed for an application to only the CPUs within a virtual machine in which the application runs. The Oracle Database, for example, running in a capped Solaris Container, needs fewer licenses than if the container were uncapped, potentially saving the datacenter quite a bit of money. Given its variations, complexity, and importance, it makes sense to give the virtualization battle a column of its own.

Comparison

For a full description of why these three operating systems are included and others are not, please see my December 2010 *;login:* column, which also provides an overview of each operating system and potential reasons to run each. Fundamentally, AIX 7.1, Solaris 10 9/2010, and Red Hat Enterprise Linux 5.5 are going to continue to be important operating systems, each having a potentially bright future (even if for different reasons). The purpose of this two-part column is to help readers sort through these operating system choices and provide a basis for comparison and consideration for use.

Virtualization is certainly many things to many people, but included here is hardware and operating system–provided virtualization. Certainly, Solaris and RHEL can also be virtualized by other technologies, such as VMware ESX and VirtualBox, and those are valid choices. However, those are not features of the selected operating systems and therefore are outside the scope of this column. Within the scope, though, is hardware virtualization, and even if Solaris only has that feature available on SPARC hardware, RHEL has no such feature, and AIX has that feature on all implementations (as it runs only on the Power CPU and Power has hardware virtualization). Given those differences, why is hardware virtualization "allowed" in this comparison? Fundamentally, most Solaris sites run Solaris on SPARC hardware, and many of those use the hardware virtualization features.

Likewise, AIX shops make extensive use of the Power hardware virtualization features. Hardware virtualization is an important tool available to administrators of those systems, and therefore must be considered when looking at the complete virtualization picture.

Virtualization

The gamut of virtualization technologies to consider is therefore hardware, multiple-OS software, and single-OS software. Those three are further explored in the next three sections.

Oracle/Sun Hardware Virtualization

In the Sun world, this is "Dynamic Domains" and "LDOMs" (now renamed by Oracle to "Oracle VM Server for SPARC," but called LDOMs here for brevity). This feature is implemented at the hypervisor level. By default, all hardware is seen in one chunk (for lack of a better word). The hypervisor allows the hardware, including CPUs, memory, I/O controllers, and system disk drives, to be divvied up into two or more chunks. Each chunk gets its own Solaris install and can run different OS releases and patch levels, different applications, and so on. Only Solaris 10 and Solaris 11 Express are supported on the M and T servers, however, so operating system choice is somewhat limited. For almost all intents, the given server acts like multiple servers that just share the same sheet metal.

LDOMs existing only on Sun "T" servers, and sharing the same motherboard, can have single points of failure. Dynamic Domains existing on M servers do not generally share single points of failure. Dynamic Domains and LDOMs are both somewhat dynamic, in that some resources can be moved between the chunks without downtime. The number of allowed Dynamic Domains varies based on the M-server model, from zero on the M3000 through two on the M4000 (a minimum of two CPUs per domain) to 24 on the M9000. The number of LDOMS is limited by the number of threads in the T-server box. A T3-1 has one socket containing 16 cores and 8 threads per core, for a total of 128 threads. LDOMs can be as small as one thread, for a possible total of 128 in the T3-1. Practically, one LDOM per core is more reasonable and common than one LDOM per thread.

Each Dynamic Domain has its own I/O controllers and is independent of other Dynamic Domains to perform its I/O. LDOMs need to have one or two "I/O Domains" which include all of the I/O controllers and have all I/O routed through them. If there are two I/O Domains, then they can be configured to provide redundant paths to the devices for high availability.

IBM Power Hardware Virtualization

IBM Power servers have LPARs (Logical PARtitions). While Dynamic Domains and LDOMs are included with the cost of the Oracle server, IBM frequently has an additional cost for features such as virtualization. LPARs are more flexible than LDOMs in that they can be configured to the single core increment, up to a maximum of the number of cores in the system, as well as in micro-partitions. Most Power servers have a "micro-partitions" feature in which an LPAR can be split into up to 10 sub-chunks. There are some fixed limits: for example, currently the IBM Power 790 can only have 256 partitions (a combined total of LPARs and micro-partitions). That is still a large number, though. And even the micro-partitions are full hardware instances in that they have their own copy of AIX installed.

AIX versions 5.2 and beyond are supported on most Power servers, as are RHEL and SUSE, so a wide variety of operating systems can be running within the same system.

LPARs can act like Dynamic Domains and have dedicated I/O, or have a "Virtual I/O Server" partition which manages I/O and which other LPARs and micro-partitions talk to to get I/O performed. It can also mix these two options. There does not seem to be a way to make the Virtual I/O Server highly available, but a Virtual I/O Server is running AIX and can be configured for multi-pathing within the LPAR. As with Dynamic Domains, resources may be moved between LPARs. LPARs with that feature enabled are known as Dynamic LPARs or DLPARs.

There are some features provided by IBM Power hardware virtualization that are not found within Oracle's Sun products. The most impressive and useful of these is "Partition Mobility," which can move an LPAR or micro-partition between Power servers without disrupting service, taking a few seconds. Partition Mobility works for all operating systems, including Linux. There is also an "inactive" version of Partition Mobility that makes it easy to move a halted partition between servers. Of course, all storage used by the partition (the boot disk as well as data disks) must be able to be mounted on the target server, via FC SAN or iSCSI connectivity. Also, all I/O must occur via the Virtual I/O Server.

Active Memory Sharing is another interesting feature, in which a pool of memory is configured and that pool is made available to multiple partitions. This can be simpler and more flexible than moving memory between partitions.

PowerVM Lx86 provides a supported solution for running Linux x86 binaries within a Power partition. This solution involves a binary translator that converts and then caches the x86 instructions into Power instructions. No modification of the original Linux x86 binary is needed. Also translated are system calls from the x86 versions to ones compatible with the native Linux operating system running on the Power CPU within the partition. There is obviously a performance penalty incurred by translation, but the fast Power CPU probably helps decrease the impact of that translation on the performance of the x86 Linux application.

RHEL Hardware Virtualization

This section might be surprised to find itself here, but at least a bit needs to be said about the possibility of using VMWare ESXi to "hardware virtualize" RHEL. This solution to having hypervisor support running virtualized RHEL (and many other OSes) is legitimate and functional. In fact, it provides the "vMotion" feature, which enables the transfer of a running virtual machine from one VMware server to another without service interruption. ESXi itself is free, but there is a maintenance cost if desired, and features such as vMotion are only available for a price. Perhaps the biggest difference between the ESXi solution and the others discussed above is that ESXi comes from a third party and adds support complexity because a problem resolution may require calls to two support desks rather than just one. So, rather than crossing "hardware virtualization" off of the RHEL feature list, consider this option.

Of course, this is one of many third-party virtualization options, which include software solutions such as VMware ESX, Xen, and VirtualBox, and even hosting solutions such as Amazon EC2 and Joyent's Cloud. Those are well worth understanding and evaluating as part of performing virtualization planning due diligence.

Solaris Software Virtualization

Solaris includes "Containers," also called "Zones," as a core virtualization feature. Given that Containers have been discussed at length elsewhere and in previous issues of this column, only the basics are discussed here. Containers provide a secure environment in which applications run. Those applications have no knowledge of other applications in other Containers. Applications in one Container can only communicate with other Containers over network protocols. Most importantly, Containers do not run their own kernel. Rather, one kernel is used to create all Containers and all applications. There are Solaris 8 and Solaris 9 Containers, allowing a Solaris 8 or Solaris 9 system to be converted from a physical instance to a virtual instance (P to V) and run within a Container (at added cost). Containers may not reside on NFS, but can reside on block storage such as DAS, SAN, and iSCSI. Finally, Containers can be detached, and then attached to another server that sees the same storage, and can be duplicated, but they cannot be moved without service interruption.

AIX Software Virtualization

AIX has Workload Partitions, or "WPARs," which are very, very similar to Containers. Again, there are no separate kernels, applications are protected from each other, resources can be managed, and so forth. It is an AIX feature, unlike LPARs, which are a Power CPU feature. Unlike Containers, WPARs can reside on NFS storage. Also unlike Containers, WPARs can be moved without service interruption via the AIX Live Application Mobility feature. Such a move is much like VMware's vMotion.

RHEL Software Virtualization

Unlike AIX WPARs and Solaris Containers, RHEL KVM is very new, having been released in 2009, and seemingly not very heavily used in production environments. KVM replaces Xen, which was the chosen virtualization technology for RHEL 5. KVM has some impressive features, including supporting 64 virtual CPUs per guest with low overhead, as well as live migration, but the only supported guests are RHEL versions 3 through 6 and Windows 2003, 2008, XP, and 7.

KVM is included in RHEL, and also in a separate product from Red Hat named "Enterprise Virtualization." RHEV is a virtualization platform, in some ways similar to VMware ESX. It supports RHEL and Windows guests and provides virtualization features such as live migration, high availability, power management, and system scheduler (for scheduling activities). Oddly, the RHEV manager only runs on Windows. Apparently, RHEL and RHEV are in a transition period and only time will tell how enterprise-worthy these components are. As is the case with Linux in many areas, and one of its strongest benefits, there are many free choices surrounding virtualization and virtualization management. Determining which, if any, of them are production-ready can be a challenge, though.

Conclusions

If just the native, included-with-the-OS virtualization features are considered, then Solaris has a distinct advantage over the other contenders via its Domains, LDOMs, and Containers. However, expanding the possibilities to features available for purchase from the vendors, IBM puts on an impressive show with features that match all of Sun's, and exceeds them with the ability to move partitions between servers without service interruption. RHEL can only match these features when adding third-party solutions such as VMware ESXi for hardware virtualization. That is a valid choice, as long as the challenge of needing to involve another entity in support issues is balanced against the benefits of the added feature set.

Finally, RHEL 6.0 was just announced and promises to add new features to the already sound offering. For more details on what is included see http://press.redhat.com/2010/11/10/red-hat-enterprise-linux-6-a-technical-look-at-red-hats-defining-new-operating-platform/.

Tidbits

Solaris 11 Express has finally been born, after a long pregnancy and even more prolonged labor (of love?). Certainly the waiting was the hardest part, and was possibly extended by the purchase of Sun by Oracle. S11 Express has many of the features of OpenSolaris, except that there is no source code available for it. It does have advanced features such as ZFS encryption and a vastly better network functionality code named "crossbow." According to the FAQ (http://www.oracle.com/technetwork/server-storage/solaris11/overview/faqs-oraclesolaris11express-185609.pdf), S11E can be used for non-production use for free, but use in production requires a support contract. It is available for download from http://www.oracle.com/us/products/servers-storage/solaris/solaris-11-express-185123.html.

References

AIX 7: ftp://public.dhe.ibm.com/common/ssi/ecm/en/pod03054usen/POD03054USEN.PDF.

IBM Power server features: https://www-304.ibm.com/businesscenter/fileserve?contentid=205264.

AIX virtualization: http://www.ibm.com/developerworks/wikis/display/virtualization/Virtual+IO+Server.

IBM Live Partition Mobility: http://www.redbooks.ibm.com/redbooks/pdfs/sg247460.pdf.

RHEV overview: http://www.redhat.com/rhev/.

RHEL KVM: http://www.redhat.com/f/pdf/rhev/DOC-KVM.pdf.