PETER BAER GALVIN

Pete's all things Sun: the Sun virtualization guide

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 coauthor 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

BACK IN THE DAY (SAY, FOUR YEARS ago) there was only one choice available for a system administrator wanting to run multiple environments within a single Sun server—domains. Domains provide a valuable solution to some problems, but they leave many other needs unaddressed. Fast-forward to today, and there are literally five main virtualization options (and dozens more if you count all of the x86 virtual machine technologies). These technologies vary in features, functionality, maturity, performance overhead, and supported hardware. This month in PATS, I'll provide a brief overview of the main choices and a chart that can be used in deciding which virtualization solution to bring to bear given a set of criteria.

Virtualization Options

The reason for the great increase in Sun virtualization options—which for the purposes of this column includes technologies both from Sun and from other providers—are multifold. Sun now has SPARC and x86 systems, for example. Each has its own virtualization choices. And as technology advances and servers increase in capacity, there is a natural need for implementation of features that help administrators best utilize those systems. Sometimes maximum utilization can be had by running one big application, but more and more frequently many, many applications can run comfortably within a system's resources. This situation will increasingly occur over time as more and more cores and threads fit into smaller and smaller systems.

Sometimes resource management can allow those multiple applications to play well together within a system, but in other situations more segregation between applications is needed. With that segregation, be it virtual machines or similar functionality such as zones, comes a new set of abilities for system administrators. Consider that VMware ESX software allows the movement of processes between systems without interrupting the processes or their users. The use of these technologies can totally revolutionize how computing facilities are designed, implemented, and managed. Disaster recovery (DR), resource use, load balancing, availability, and flexibility can all benefit from vir-

tualization technologies. They are clearly worth using, but which should be used, and when?

Overview

A previous article in *;login:* provided details on many of the technologies [1]. To complete the picture, here is a brief overview of all of the options.

Dynamic System Domains [2] (or just "Domains") have existed for years in Sun Enterprise servers. They started as a method by which a larger server (such as an E10K) could be sliced into multiple, electronically isolated virtual systems. Each domain has its own operating system installation, and even a hardware fault in one domain typically does not affect any other domains. More recent Enterprise servers increase the ease with which domains can be dynamically reconfigured, in that CPU, memory, and I/O resources can be moved between them as needed. Typically this is done infrequently, to handle a temporary load or to change the system from interactive-oriented to batch-oriented (for example, between a daytime shift and an evening set of jobs).

Logical Domains [3] (LDoms) are the baby brother of Domains. They segregate a "coolthreads" (a.k.a. CMT or Niagara) system into virtual systems. Because these systems are motherboard-based, there is less fault isolation than on Enterprise servers. Some resources may be moved between LDoms, and "warm" migration is supported between systems. In this scenario, the two systems coordinate to move an LDom from the source to the destination, and the LDom is automatically suspended on the source system and resumed on the destination.

Zones/Containers [4] are a secure application environment, rather than a virtual machine implementation. There is one Solaris kernel running, and within that many, many zones can contain applications that run independently of other zones and can be resource-managed to a very fine degree. Fair-share scheduling can assure that each zone receives a "fair" amount of CPU, for example, while memory and CPU caps can limit the exact amount of each of those resources a zone can use.

xVM Server [5] is based on the Xen virtualization software project. It is a separate, open source product from Sun, in which Solaris is used as the software that creates and manages virtual machines. This is similar to the way ESX is the manager of virtual machines in a VMware environment.

VMware [6] and the like are virtual machine managers in which multiple operating systems, running multiple applications, can coexist within the same hardware.

The Guide

Choosing the right virtualization technology can be complicated and depends on many factors. Additionally, many of the available Sun virtualization technologies can coexist, adding functionality but also complexity. How is a system administrator to choose? Table 1 was designed to help guide such a decision. In this chart are summaries of all the major attributes of the technologies, including functionality and limitations. Use this as your initial guide, and then read the Explanation section to sort out the details.

Feature	Dynamic System Domains	LDoms	Zones / Containers	xVM Server	VMware ESX
Available on	USPARC	Niagara I and II	SPARC and x86	X86	X86
OS	≥Solaris 8	S10	S10	xVM Server host, Windows, Linux, Solaris, and other guests	VMware ESX host, Windows, Linux, Solaris, and other guests
Separate kernels/ packages/patches?	Yes	Yes	No	Yes	Yes
Fault isolation	Most	Some	Less	Some	Some
ISV support	None needed	None needed	Needed	Some needed	Some needed
Moveable between servers	Only via SAN boot, downtime	Yes, via warm migration	Yes, via attach/detach, downtime	Yes, live	Yes, via Vmotion, live
Resource management	Limited, at single CPU granularity	Yes, at single-thread granularity	Yes, at subthread granularity, very flexible	Yes, very flexible	Yes, very flexible
P to V effort	Using standard install tools	Using standard install tools	Solaris 8 Containers, Solaris 9 Containers, otherwise standard system tools	Physical to virtual tools included	Physical to virtual tools included
Efficiency	Higher	Higher	Higher	Lower	Lower
Speed to produce another instance	Slow (standard tools)	Slow (standard tools)	Fastest (cloning)	Fast (copying)	Fast (copying)
Other Features					Snapshots for version control
Maturity	Mature	Young	Middle-aged	Youngest	Middle-aged
Cost	Free	Free	Free	Free/$	$$

TABLE 1: COMPARISON OF THE SERVER VIRTUALIZATION TECHNOLOGIES AVAILABLE WITH SOLARIS 10

Explanation

This column includes enterprise virtualization features, not desktop options. For desktops there are also many choices, including the new kid on the block, VirtualBox. It is now owned by Sun, open source, and feature-rich. It's available at http://www.virtualbox.org/.

Standard install tools include install from DVD, Solaris Jumpstart, or Flash Archive and other network installation tools. Each of these has functions and limits that need to be planned for. For example, Flash Archive can only capture and build like-architecture systems (e.g., you can't capture a SPARC system and build an x86 system from that image).

There are several important limits with respect to Domains. For example, the system must be preconfigured to allow domains to change dynamically.

LDoms are limited to only the "coolthreads" CMT servers from Sun. For warm migration, the LDom's boot disk must be stored on an NFS server or SAN (and SAN functionality was limited at the time of this writing).

Some applications are not supported with Zones, so it is important to check the vendor support matrix for every application.

"Solaris 8 Containers" and "Solaris 9 Containers" are a commercial offering from Sun allowing physical to virtual (P to V) transfer of existing S8 or S9 (SPARC) systems into containers within S10 (SPARC).

DTrace may be used within Zones, but it may not probe kernel resources. Only the "global zone" (Solaris, not within a Zone) may do that. Further, DTrace can only go so far when used in conjunction with other operating systems. With xVM server, DTrace will be usable within the host but likely not provide any value looking within the guests (unless they are Solaris too).

xVM Server has not yet shipped (at the time of this writing) so the expected features may or may not be included or work as expected.

Each of these technologies has features to manage the resources used by the guests. Those technologies vary wildly in their abilities, ease of use, and ease of monitoring. Generally, domains are the least granular and zones are the most granular. If you have specific resource management needs, then a thorough read of the documents and an on-line search for best practices are your next steps.

The creation of a new "clone" virtual machine from an existing virtual machine is a very useful feature. Although some tricks can be played to speed up the process, I only address the feature set included with the virtualization technology when describing how quickly a clone can be made. For example, if SAN or NAS boot is used for Domains or LDoms, the SAN or NAS capabilities can be used to quickly clone new instances.

Certainly, the efficiency of all of these solutions will vary depending on many aspects of the deployment. However, I assert that the closer to the raw hardware the applications are, the faster they will be. For example, there is essentially no difference in the path that the application code takes when run within a Domain or on a Sun server without virtualization. Similarly, Zones are very efficient because only one kernel is running. Hundreds of zones can run within one system. Of course, with VMware and ESX Server, if there is only one kind of guest operating system running then the system likewise can be highly efficient. As with all performance analysis, consider the details of your specific deployment to determine how efficient one virtualization technology would be compared to the others.

More so with virtualization than many other technologies, the details, features, and functionality are changing all the time. I plan to make this article available on-line and keep the chart up-to-date, so if you are interested in updates keep an eye on my blog at http://www.galvin.info.

Conclusion

The many virtualization choices on Sun platforms vary in all degrees. Fundamentally, each has pros and cons, as well as uses and abuses. This guide should help sort through all of those aspects to help you determine the best choice(s) given your environment and needs. Aside from the features, be sure not to lose track of the complexity added with each of these technolo-

gies. Management, monitoring, debugging, and performance analysis and tuning are all major aspects of a virtualized environment. That is where tools such as VMware Virtual Center and the new Sun xVM Ops Center can ease the burden and control virtual machine sprawl. Consider the tools to manage the environment as well as the environment itself.

Random Tidbits

Sun LDoms 1.1 shipped in January and added new features (as described in this column).

Also, if you have any interest in using OpenSolaris, check out the recently published book *The OpenSolaris Bible* [7].

There is a new Sun blueprint that goes into great detail on configuring Sun 7000 storage for running Oracle databases [8].

REFERENCES

[1] http://www.usenix.org/publications/login/2008-10/openpdfs/hu.pdf.

[2] http://docs.sun.com/source/819-3601-14/Domains.html.

[3] http://www.sun.com/servers/coolthreads/ldoms/index.jsp.

[4] http://en.wikipedia.org/wiki/Solaris_Containers.

[5] http://xvmserver.org/.

[6] http://www.vmware.com/.

[7] N.A. Solter, J. Jelinek, and D. Miner, *The OpenSolaris Bible* (New York: Wiley, 2009).

[8] http://wikis.sun.com/display/BluePrints/Configuring+Sun+Storage+7000 +Systems+for+Oracle+Databases.