FLEXIBLE COMPUTING ARCHITECTURES FOR SUPPORTING RAPID ALGORITHM DEVELOPMENT
TECHNOLOGY FOCUS: SERVER VIRTUALIZATION AND VIRTUAL APPLIANCES

Jason D. James
7b.4 National Oceanic and Atmospheric Administration

1. Introduction

Server Virtualization is a new technology concept that provides superior hardware utilization capabilities when compared to a traditional computing environment. This technology provides significant advancements in efficiency, productivity and it has also redefined the role and architecture of science computing.

Server Virtualization unlocks vast opportunities by "abstracting" the operating system from the physical hardware layer. The software to hardware link in today's traditional computing environment "locks" the software layer of a computer system into a fixed 1 to 1 operating system, and physical hardware match. Server Virtualization "unlocks" this inefficient concept as a part of the evolution of maximizing the use of available resources. The "unlocking" allows system administrators an unparalleled flexibility to create multiple operating systems inside one hardware platform.

The key to server virtualization is the software loads the operating systems into "files" that act as physical hard drives. This innovative approach allows all I/O components (network cards, hard drives, memory, CPU) to be made available to all the operating systems present in the physical hardware system.

As a result, when the system is started, all operating systems (Linux, Solaris, Windows) will each see and use all physical servers resources assigned to the virtual sever. Below is a pictorial representation of the virtualized versus the traditional server concepts.

Corresponding author address: Jason D. James
National Oceanic and Atmospheric Administration,
Environmental Satellite Processing System, 4231
Suitland, MD 20746
Email:jason.james@noaa.gov

2. How Does it Work?

Virtualization is based on a software component called a hypervisor. A hypervisor is the core component that manages and schedules access to

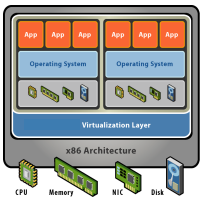

Figure 1

VS

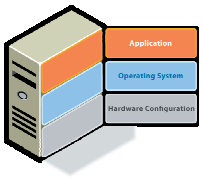

Figure 2

resources when request are sent from the virtual machines to server resources. In Figure-1 above is represented by the "virtual layer". While the function of the hypervisor is similar across all virtualization solutions, the difference is some hypervisors are proprietary and some are open source. The choice between the virtualization solutions depends on the tool set that has been built around each solution. Some solutions are more mature than others regarding the tools available to compliment the virtualization solution. That maturity is driven by industry acceptance and the community development of third party tools that add functionality to the product and solution. An additional consideration is the support provided with the virtualization solution. The support option, prepare the virtualization solution for enterprise usage in production environments.

3. Benefits of Virtualization

Why is server virtualization worth investigating? The technology provides the ability to squeeze value out of your current computing infrastructure, and it also creates a smart path to a more efficient and cost effective IT organization. Architecturally, virtual environments enable IT infrastructures to be flexible, manageable and most importantly agile. The ability to apply limited amounts of resources to a wide array of

projects provids the science community the flexibility needed to support a rapid development environment. When the server resources become a commodity in the IT environment, new doors open up regarding the options available in the virtual computing environment. These new doors come as operational improvements to IT management and customer response. In the science, research and government arenas, these operational improvements consists of:

Operational Benefits
Easy Hardware Refresh
System Approval and Templates
Streamlined Security Compliance

Science Benefits
Isolated and Standard Development Environment
Easy Provisioning of Development Resources
Managed Progression from Development to Production

The above is not a complete list, however it provides a glimpse into the possibilities for the virtualized data center concept in support of rapid development and IT operations management within the science community.

4. Modes of Virtualization

There are two modes of virtualization a server can be configured for.

1. Fully Virtualized
2. Para-Virtualized

A fully virtualized server does not know it is a virtual machine. It functions as though it were a physical server and all resources are static (non-adjustable) while the sever is operational. The Para-Virtualized server mode enables the operating system to be aware that it is a virtual machine. The awareness allows the management software for the para-virtualized server to adjust the resources on-the-fly. An administrator can add additional memory and CPU's if needed while the server is operational.

5. Performance Concerns

There is a 3 to 6 percent performance overhead associated with virtualizing servers. The overhead is associated with the hypervisor scheduling request from the virtual machines. This concern has been resolved by configuring virtual machines in the para-virtualization mode. The para-virtualization mode enables the kernel and the hypervisor to communicate faster and thus minimizes the overhead associated with the fully virtualized mode. Some linux distributions enable para-virtualization inherently when the operating system is loaded in a virtual environment. These Linux distributions consists of Redhat and Ubuntu Linux.

6. Easy Information Sharing

Server virtualization enables developers to share the entire systems. How is this done? Developers today are creating virtual appliances. A virtual appliance is an application integrated and packaged with just enough operating system (JeOS) *--pronounced juice* -- that can easily be in a virtual container. One of the advantages of virtual appliances is IT consumers are not required to integrate and configure application and operating systems into a working solution themselves rpath(2008). These virtual appliances can be saved and downloaded by anyone who wishes to use them. How does this affect the science community? The science community can begin to pre-configure complex installations of science algorithms with an operating system of their choice and share the configuration throughout the community. Any scientist who wishes to add-to or improve a processing system can do so by downloading the server or appliance, test the software and make improvements to the system and repost the server or appliance for download in the community. This type of collaboration is possible because the virtual servers are a set of files that represent the hard drive of a physical system. Like all other files, the virtual server files can be posted on the web for download. The community can then download those

files and run a pre-configured appliance to test data sets. This is just one scenariomany where servers can be thought of as a set of files in need of some resources to consume to perform a job. Visit http://vmware.com/appliance or http://rpath.org to learn more about virtual appliances and the community supporting them.

Quick Bullets: Easy Information Sharing

- Systems are easily downloadable
- Processing Systems are a group of Files
- Systems can be pre-configured

7. Virtual Appliance and Its Effects on Development

The development of virtual appliances gives the ability for the developer to create a lean operating system that is built and configured specifically to support the algorithm that will be executed on the appliance. Once the appliance system has been built, the user only needs to have a virtual platform to run the appliance and the data inputs required for the algorithm. The user can now if needed make algorithm adjustments or system adjustments and consequently share the system back out to the community. This all can be done with open source software and accessible inexpensive hardware.

8. What Problems Are Solved by Virtual Server Environments?

Operationally virtual server environments streamline system support and development. In a virtual server environment the server operators are able to build servers once and then create a server template that can be cloned or copied as many times as needed. This functionality enables Data Center Operators to maintain standard system installations and keep security standards intact across multiple system installs. The virtual server environment also allows for a quicker response to server request from customers, based on the template methodology described above. Together, these benefits look to reduce much of the repetitive duties of the support staff and create a better environment to manage critical systems and facilitate compliance.

In a virtual server environment, operators are only bound by the amount of physical resources available for consumption by the virtual servers created in the environment. One hundred percent physical server resource utilization can be an indicator of needed growth, which can provide justification for additional resource investment to increase computing capacity. The software that provides the virtualization layer collects data on virtual server resources utilization which can help determine when servers are busy and when they are not. This data can help system planners and application owners understand their system needs to enable better resource planning in the future. Further more, that utilization and performance data can help the system operators create resource groups within the system to support application needs. These resource groups can be adjusted to provide support where needed across the enterprise. Simply, server virtualization enables system owners to have the resource flexibility and mobility to solve and manage many problems throughout an rapid enterprise computing environment.

9. Environmental Benefits

Virtual server environments provide many benefits for conserving power in data centers. The power savings come in many different forms. Normally, the idea is that a single physical server can support the jobs of many virtual servers thus improving the utilization of the physical server and reducing the amount of physical servers needed in a Data Center to get multiple jobs done. The impact of less servers in a data center affects:

1. Power Consumption to cool the Data Center.

2. Electrical cost to power a smaller number of servers.

3. Ability to operate a lean, more efficient support system.

4. Rapid Development Environment Support.

10. Conclusion

Virtual server technology opens up many opportunities within a science computing environment. The software especially offers the science community improved methods to share complex systems that will enable others with limited resources to learn about advanced technologies with little technical computer knowledge. Sharing, developing and executing science algorithms can become an easier process than what is today.

11. References

Rpath (2008). Best Practices for Building Virtual Appliances.Retrieved December 12, 2008, from http://rpath.com/corp/images/stories/white_papers/WP_BestPractices.pdf)