Virtualization: Clouds on the Horizon

Managing Virtualized Data Centers Now and in the Future

By Tom Joyce
CEO, Akorri, Inc.

August, 2008

Executive Summary

Virtualization technology is revolutionizing IT. It has taken root in large, medium and small enterprises throughout the world, and we are only just beginning to fully grasp the transformational potential of this technology. VMware is the clear leader, and is continuing to innovate and find ways to stay in front of the pack, while other server virtualization platforms such as Citrix XenServer and Microsoft Hyper-V are emerging as viable alternatives.

As we look into the future, virtualization technologies are set to push the envelope even further. Virtual Desktop Infrastructure is on the immediate agenda of many companies, and we believe that in the not distant future we will see server virtualization become the bedrock of a truly dynamic cloud computing model of IT services delivery. This "cloudy" vision of the future is getting serious investment from all of the major IT product and solutions companies, so it's coming, and fast.

But just like other revolutionary IT technologies that have come before it - open systems, Windows, and the Internet to name a few - VMware's rapid march into the data center has been accompanied by some "broken glass". Many of the companies that have gotten the most benefit from VMware have also found that wholesale changes were required in tools, processes, and talent to make this technology work in an enterprise-class fashion.

In the following we will look at the some of the most significant management challenges that are coming in the next wave of the virtual data center.

A Slight Return

First let's review what we have learned in the last few years. In our dealings with production users of VMware, in small, medium and large data centers, we have seen IT managers struggle with a number of common problems. Akorri provides technology that can help correct all of these issues, and customers that have done so have experienced great success with their VMware initiatives.

1) Performance Problems: Many customers experience acceptable performance initially because they virtualize non-critical applications that are not memory, CPU, or I/O intensive. The typical customer we work with has used one of the basic physical-to-virtual (P2V) planning tools from VMware, CiRBA, or Platespin. These tools help identify easy-to-virtualize non-critical applications, and all is well at start. However, as customers move to the next phase of deploying VMware as a corporate strategy, they move past the low hanging fruit, and start to go after applications that are important to the business. These applications require consistent performance. At this point in the roll-out of virtualization, basic P2V planning tools can no longer be trusted. They do not provide any insight into one of the most important aspects of performance: application demand for I/O. They also do not account for contention among VMs that are sharing server and storage resources. Not all VMs can live together on the same platform, and we regularly see environments where a single VM on a server is negatively impacting the performance of eight or ten others. Our analytics technology can help solve planning and performance problems such as these very easily.

2) Storage Problems: Most VMware users we encounter have not spent enough time planning their VMware-related storage requirements. Often we find that the storage department within IT has not been engaged in the planning. It cannot be assumed that what worked in the physical world will work the same way with a dynamic, highly consolidated virtual environment. Intuitively, it makes sense that if you take a physical server that is using twenty percent of it's capacity and load it with VMs so that it ramps up to seventy five percent utilization, the storage infrastructure that is

supporting it will need to be reconfigured. But this is often not fully understood, and the result can be storage-related performance bottlenecks and capacity issues that are difficult to root cause. Another contributing factor is so-called "VM sprawl", the phenomenon of rapid increase in the number of VMs. Without close monitoring of the impact of rapid growth, the storage infrastructure can get overwhelmed very quickly. Again, our BalancePoint software has features that help deal with these problems.

3) Increased Complexity: Many customers initially intend to reduce the number of physical servers they are running, but in our experience this typically does not happen. Usually what we see instead is the existing servers are eventually refreshed with multi core servers and blade infrastructure, and significant dollars are saved through this process, but the effective server capability does not decline significantly. Instead, it goes up. Most of the savings from a VMware project comes through increased flexibility, efficiency, and improvements in areas like disaster recovery, not from taking out server hardware. Along the way, the number of objects IT has to manage increases exponentially. If you start with one hundred servers and virtualize 10X over a one year period, you now have one thousand server objects to manage. If you factor in VM sprawl due to the ability to provision dozens of new virtual servers in minutes, and you can double your one thousand servers easily. Meanwhile, there is typically no additional headcount to manage all of these new entities. We often see scenarios where a small team that was managing tens of servers last year is now struggling to manage thousands. The only way to deal with this is intelligent automation. Our product is designed to help with this, but more often than not, customers do not plan for this in advance, and we are brought in to help fix things once they have reached a crisis stage.

4) No Visibility: Most existing management software frameworks and tools were designed to deal with physical infrastructure. Furthermore, most of these tools were designed for only a part of the infrastructure, either servers, networks, or storage. It is very difficult to solve a VMware problem with a toolset that was designed for physical infrastructure. On the other hand, the new virtualization management tools such as VirtualCenter, CiRBA, Vkernel and DynamicOps only look at virtualized servers, not at non-VMware resources in the same environment such as non-virtualized servers databases, SAN fabrics and storage arrays. These VMware-only tools are completely blind to the outside world. For example, if a performance problem is due to a VMware server and a Solaris server sharing the same RAID group, tools like Vkernel or CiRBA will be unable to diagnose the problem due to lack of visibility into those resources. This simple example is an extraordinarily common one. Most companies we encounter are using multiple tools to try to make sense of what is going on in their environments. Our software can see and analyze the whole physical/virtual server/storage landscape. To put it succinctly, we help cure "blindness" in virtualized data centers.

What's Next?

We believe that we have solved many of the problems that have emerged in the first wave of production implementations of VMware. As we move into the next wave, it makes sense to be more proactive in our approach as an industry so that we don't have to address the new set of challenges after the fact. What do we have to look forward to over the next two years?

200:1 Consolidation: This might seem outlandish, as the conventional wisdom today is that 20:1 is about the best you can hope for. However in late 2006 one of our first customers was doing over 80 VMs on some VMware servers in a production environment. Clearly they were pushing the envelope. But within the next year massive single server consolidation will be enabled by a combination of technologies, including:

- 8-core blade servers and cheap 4-core servers with new memory designs
- High performance solid state disk (SSD) storage technologies
- New SAN transports such as FCOE, 10BbE and Infiniband

Buzzword Consolidation: Any technology that could reduce the number of buzzwords that litter the IT industry would be a great benefit to humanity. Consider the following IT industry buzzwords:

- "SOA": A computer systems architectural style for creating and using business processes, packaged as services.

- "Grid Computing": A form of distributed computing where a cluster of networked, loosely coupled computers forms a "virtual super computer" to perform large tasks.

- "Utility Computing: The packaging of IT resources including compute and storage as a metered service similar to a traditional utility.

- "Cloud Computing": Internet-based delivery of IT capabilities as a service, without requiring the user to have knowledge of, expertise with, or control over the technology infrastructure.

One of the most obvious things that these four computing models have in common is that virtualization is a key enabler of all of them. In fact, without virtualization in some form, none of them are possible. We believe that with the advent of production class virtual infrastructure, widespread use of these models will rapidly become a reality, and in fact they will all merge into one model for how the data center will look in the years to come. We see the infrastructure inside a corporate data center will look very similar to the "pay as you go" cloud computing model delivered by service providers. The benefits for IT and business will be significant. An ancillary, but no less welcome, benefit is that some of these buzzwords can go away. We hope.

Virtualization is not the only requirement however. Just as changes in management tools, processes and talent have been required in virtualization up to this point, some additional capabilities will be needed to manage the virtualized data center of the future. Here are a few examples of what will be required:

1) A "Systems Model" of the Infrastructure: In order to deliver consistent quality of service you must be able to model the entire environment in terms of its current and potential capability. In a virtualized world you have to be able to do this for each virtualized application that is running on the system. This means identifying the "slice" of resources that each app is using across the entire server/network/storage complex, and identifying the services levels, priority levels and policies associated with maintaining an appropriate level capacity for each requirement. This must take into account the fact that virtual servers and the virtual storage containers are frequently in motion, moving dynamically from host to host. Making this work this on pooled resources versus dedicated resources requires new analytical technology, and in our opinion the only way to do this is with mathematical analysis based upon queuing modeling. This is the key concept behind the modeling inside of Akorri BalancePoint.

2) Dynamic Tiered Resource Management: One of the biggest challenges will be assuring that the more expensive, higher performing systems are allocated to the more important and more resource intensive applications, and that the less critical apps are run on the less expensive systems. Tiered

resource allocation must take into account business cycles and seasonality. For instance, applications that are critical at month-end and are essentially dormant at other times, it will be immensely valuable to be able to allocate resource automatically based on policy. In the securities industry it would be valuable to allocate more resources to key trading apps while the market is open, or to make changes dynamically in response to unexpected changes in demand. This technology exists, but typically not in a way that extends across the whole infrastructure. There is QOS for the server layer, and QOS for the storage layer, and today they do not connect. Coordinating tiered server and tiered storage allocation into a single uniformly responsive system is the ultimate goal. Akorri BalancePoint's cross-domain management capabilities will be critical to enabling tiered resource management across both the server and storage layers.

3) End-to-End Business Service Alignment: How do you know if you are doing well? How do you relate the performance of the IT infrastructure to the requirements of the business? Are service levels clearly understood by both the service provider and the customer? Whether in an internal IT environment or a hosted environment, these are the questions that are ultimately important at the CIO level. Answering them requires the ability to measure performance and utilization in ways that are relevant to the business. Akorri BalancePoint delivers business-critical key performance indicators that can form the basis of performance and utilization related service levels.

Next let's look at some of the core capabilities of Akorri BalancePoint for enabling the next generation of data center virtualization.

BalancePoint Capabilities

BalancePoint is a virtual appliance that collects data agentlessly across the server and storage infrastructure, processing it in near real-time to identify potential performance issues and configuration problems. BalancePoint identifies performance and capacity resource dependencies for each application running on the virtualized infrastructure, automatically defining a "fingerprint" for each. It next creates an end-to-end topology map for each application, showing every element in the topology with the potential to affect application performance, and indicates existing and potential points of conflict. This model is dynamic and is automatically updated. In its infrastructure view, BalancePoint examines the I/O infrastructure to determine which parts are shared by applications, flagging existing and potential points of contention, bottlenecks, hotspots, etc.

BalancePoint Unique KPIs: BalancePoint's most important key performance indicators are Infrastructure Response Time (IRT) and Performance Index (PI). IRT shows the performance delivered to an application by the total IT resources assigned to it, and is the main service delivery metric for managers of IT infrastructure, enabling BalancePoint to baseline and predict future infrastructure service, and to send alerts when there are service deviations. Performance Index (PI) scores the balance between application requirements and the infrastructure's capability to deliver. PI directly measures if IT infrastructure is under- or over-utilized, or is fully optimized.

Both IRT and PI are dynamically calculated as both application workloads change. These metrics are stored historically and used to project and predict future states. The mathematical complexity of all the underlying non-linear "queuing" analysis is simplified for the user by these simple scores and a "performance alerting" interface that clearly presents system states as good, bad or critical. The BalancePoint GUI visualizations of cross-domain topologies requires neither domain-specific knowledge nor any particular technical expertise to easily identify performance issues.

Key BalancePoint Features:

- **Dynamically mapped server/storage data path topologies**, mapped and modeled back through virtualization layers to datacenter application volumes.
- **Automatic appliance-based analysis of performance** in that mapped infrastructure (Red/yellow/green status on topology views, alerts with textual explanations of issues).
- **Key Performance Indicators** - Akorri IRT and PI help identify performance issues, show where resource pools are over or under-utilized and enable service level management and reporting.

ABOUT AKORRI

Akorri BalancePoint cross-domain performance management software with advanced analytics provides automated visibility and analysis across IT silos to help enterprises fix problems, optimize utilization, and improve performance in the virtualized data center. For more information, visit www.akorri.com.

CORPORATE HEADQUARTERS

305 Foster Street
Littleton, MA 01460
Phone: 978.431.1200
Fax: 978.428.0091
General information: info@akorri.com
Sales: sales@akorri.com