Server virtualization is rapidly entering the mainstream, as companies of all sizes adopt it to run everything from their development and test software to production level applications. The technology brings with it a powerful set of advantages, including greater resource utilization, improved IT flexibility, and reduced hardware and operating costs. Based on these compelling benefits, the server virtualization market is expected to continue to grow at a 40+% clip for each of the next several years.

But not all server virtualization architectures are created equal. While most server virtualization platforms are able to meet the consolidation needs of corporate workgroups and small to medium sized businesses (SMBs), very few are architecturally capable of addressing the requirements of the enterprise data center. To satisfy the more stringent demands of data center environments, hypervisors must possess some key architectural characteristics. Beyond meeting basic requirements such as fault isolation and hardware independence, “datacenter-ready” hypervisors must provide a high level of scalability and performance across all types of application workloads. In addition, these architectures must enable a high degree of security, availability and portability for workloads across a distributed virtualization environment.

Among the current providers of core server virtualization capabilities for commodity (x86 and emerging x64 based) hardware platforms, the VMware hypervisor architecture comes the closest to meeting the standards for data center readiness. We believe that the VMware hypervisor has some fundamental architectural advantages over competing virtualization platforms, which should enable it to remain a technology leader in the data center for some time to come. That said, we are confident that other virtualization companies will continue to work diligently to increase the readiness of their hypervisors for data center environments.

The Various Flavors of Virtualization

During the past ten years, vendors have taken a number of different approaches to virtualizing x86 (and more recently, x64) based server systems. Three of the most prominent methods are virtualization at the hardware (bare metal) level, application level, and operating system (OS) level. Hardware virtualization involves abstracting the hardware resources of an x86 or x64-based computer—including CPU, memory, disk and network controller—so that they can be shared by multiple virtual machines, each running its own OS and set of applications. To accomplish this, hardware virtualization vendors such as VMware, Citrix/XenSource and (later in 2008) Microsoft insert a thin layer of software directly on the computer hardware, or in the
hosted model, on a host operating system. This software layer contains a virtual machine monitor or “hypervisor” that is responsible for allocating hardware resources dynamically and transparently to virtual machines (see Figure 1 below).

In application virtualization, the abstraction of resources takes place at a higher layer in the software stack. The application itself is encapsulated in a run-time capsule, and the portions of the operating system that the application sees and uses are in effect virtualized. Though the application executes as if it is installed on a local client, in most cases it actually resides on a remote server and is streamed to each user device. The application believes it has direct access to the underlying operating system resources, although in reality it does not. Providers of application virtualization products include Availigent (Duration), Citrix (Presentation Server), DataSynapse (FabricServer), Microsoft (SoftGrid), Symantec (Altiris SVS), Trigence (AE) and VMware (Thinstall). These products enable centralized application distribution, installation and management. Application virtualization is complementary to hardware-based virtualization, and these two approaches may be used together on one or more physical servers.
Operating system (OS) virtualization, on the other hand, is a direct alternative to hardware virtualization, as illustrated in Figure 1. In this method, the underlying operating system resources are abstracted so that they can be shared by multiple applications, which typically run in isolated containers. Whereas each OS running on a (hypervisor-based) hardware virtualization platform believes it is on a dedicated physical system, each application running on an OS virtualization platform believes it is on a dedicated operating system. Examples of major OS virtualization vendors include Sun (Solaris Containers) and Parallels (Virtuozzo Containers).

OS virtualization enables extremely lightweight virtual environments, and as a result, can support up to hundreds of virtual environments (or containers) on a single physical server. This high level of density, which is not possible with hardware virtualization, is especially popular among hosting providers. Given that all containers are running on a single operating system, OS virtualization environments can also be simpler to manage. On the flip side, this approach provides a significantly lower degree of fault and security isolation than hardware virtualization, since the separation between virtual environments occurs at a higher layer in the stack. If either the host OS or a single container gets compromised, the entire platform is vulnerable. The sharing of a single OS kernel also reduces flexibility, in that it is not possible to patch a single virtual environment without patching all the others. The shared OS also prevents performance management and optimization at the granularity of a single virtual environment.

A growing number of companies are choosing either hardware virtualization or OS virtualization platforms on which to run a portion of their production workloads. But the true test for any server virtualization platform is in the enterprise data center. An IT executive will not allow such a platform into his “inner sanctum” unless he has the utmost confidence in the quality, reliability and performance of the product. Thus, to meet the higher standard of “data center readiness”, a server virtualization platform must satisfy a demanding set of architectural requirements.

Requirements for Data Center Readiness

A virtualization architecture that meets the following core requirements will be equipped to handle the rigorous demands of a data center environment:

**Hardware virtualization.** A bare-metal hypervisor encapsulates an entire server environment, including the OS, representation of the hardware it runs on, configuration settings, application(s) and data, into a file. This is a key architectural foundation for fault and security isolation, availability and portability. The VM construct simplifies backup and recovery, as well as failover of one environment to another. It enables IT to move around workloads with the ease of copying a file.
Hardware independence. A data center typically includes a mix of server hardware that spans multiple vendors and generations of technology. To be most useful, a hypervisor must have the ability to run on any single-core or multi-core x86 or x64 hardware platform without virtualization-specific hardware features or added paravirtualization. This includes the ability to exploit hardware-based virtualization enhancements, such as cpu, memory and I/O assists, but also the ability to run well without them.

OS independence. The platform must support any guest operating system that runs on x86 and/or x64 systems, including major flavors of Windows, Linux and UNIX. This is in effect a pre-requisite to dynamic workload balancing and on-demand capacity capabilities, since it ensures that all the server nodes in a virtual infrastructure are available as needed to run any set of application workloads, including a range of different OSes or OS versions. OS independence also helps to reduce the attack surface of a virtualization platform. A purpose-built hypervisor architecture with a small footprint and minimal number of external interfaces is less vulnerable to security breaches and will likely suffer fewer code failures than an architecture that includes or is dependent upon a particular operating system.

Security. A high level of security must be designed into the hypervisor and the server virtualization architecture overall. Ideally, the hypervisor will be built specifically for its role, rather than derived from or dependent on an existing operating system. A minimal footprint is necessary to reduce security vulnerabilities. The architecture must ensure complete isolation of VMs, from the hardware layer on up. For example, a physical memory page in one VM must not be accessible from other VMs, and virtual disks must also be secured from unauthorized access. Virtualization platforms should be evaluated against industry security standards, such as the Common Criteria for IT Security Evaluation, to validate the security level of the design and implementation.

Platform for distributed virtualization. Though partitioning a single system delivers some benefits, the real value comes when virtualization spans multiple servers. The virtualization architecture must have the ability to create virtual hardware resource pools, including virtual computing, storage and networking, and then provide VMs with dynamic access to this pool of shared resources. A distributed virtualization architecture enables virtual server portability, availability and workload balancing. A distributed architecture also facilitates server management, maintenance and control.

Priority-driven resource allocation. The architecture must offer the ability to dynamically balance enterprise application workloads across available cpu, memory, storage and networking resources, and continuously optimize the allocation of these resources to support the highest priority applications. Resources should be allocated first to the most critical applications, to ensure that they run with optimal performance. This capability is key to ensuring efficient power management. The architecture should also support policy-based
automation to enable IT managers to meet Quality of Service (QoS) or Service Level Agreement (SLA) commitments.

**Scalability for a range of application workloads.** A server virtualization architecture must provide scalability and performance for typical data center applications, including ERP, enterprise messaging and CRM systems. Thus, a hypervisor must be able to run multiple types of virtualized applications, such as CPU-intensive, I/O-intensive and database-intensive workloads, with good performance. I/O performance should scale up as the number of virtual machines in the virtual infrastructure continues to grow. The hypervisor should ideally also support, though need not be optimized for, legacy application environments.

**On-demand capacity.** The architecture should enable non-disruptive scaling of physical and virtual environments, including the adaptability to redistribute existing VMs to run on newly added hardware and the flexibility to re-configure servers on the fly. For example, when new server hardware is added to a cluster in a virtual infrastructure, any number of VMs in the existing pool should have the ability to be migrated to take advantage of the new server capacity. Virtual machines should be re-balanced across physical servers in the cluster to take best advantage of newly available system resources.

**High level of availability.** The virtualization platform must provide capabilities that help to reduce the incidence and minimize the impact of planned and unplanned downtime. To address an unplanned outage, the virtualized infrastructure must be able to automatically detect failures and transparently fail over and restart workloads. To deal effectively with planned outages, such as the need for hardware maintenance or an OS upgrade, the architecture must offer the ability to take a cloned virtual machine offline, update/upgrade the OS or virtual hardware, and restore the VM without downtime.

**Manageability & extensibility.** The virtualization architecture should provide secure programmatic interfaces to enable remote management, and to allow the virtual infrastructure to be enhanced and extended so that it better fits in the data center environment. A command line interface (CLI) will enable remote provisioning and management of virtual machines, while a Common Information Model (CIM) or similar interface will facilitate management of server hardware. Additional interfaces should be present to permit remote management and control of the overall infrastructure, and to allow the virtual hardware platform to be enhanced (e.g. via the addition of new device drivers) and more tightly integrated with other tools or facilities in the data center.

Overall, a server virtualization platform in the data center should provide the reliability, security and performance of native hardware. To assess the fitness of each of the competing server virtualization architectures for deployment in the data center, we will evaluate them against this set of readiness criteria.
Assessing Architectural Fit

Given the wide variance in the design and architecture of the various server virtualization platforms, it is not surprising that some of the platforms are more architecturally capable than others of taking on the data center. So how do the various approaches and products stack up?

OS Virtualization
Operating system virtualization, and in particular the leading Parallels Virtuozzo Containers product, rates reasonably well against several of the readiness criteria. Virtuozzo Containers runs across all major x86 and x64 hardware platforms, and the architecture supports OS virtualization distributed across a collection of physical servers. The Virtuozzo architecture also supports dynamic resource allocation in both Windows and Linux environments, allowing fine-grained control of CPU, memory, disk and I/O resources. The biggest advantage of the OS virtualization architecture is in the high density and scalability of virtualized server workloads that can be achieved by leveraging a single, shared operating system kernel per physical server. The low overhead of this approach enables the consolidation of dozens or even hundreds of virtual workloads per server. The Virtuozzo architecture, with its support for industry-standard application programming interfaces (APIs) and software development kits (SDKs), additionally provides strong manageability and extensibility across servers running a uniform set of operating systems.

But on some of the other readiness criteria, OS virtualization, as embodied by the Virtuozzo architecture, falls short. The sharing of a single OS kernel, while enabling significant density and scalability, constrains flexibility and compromises both the security and availability of virtual application workloads. OS virtualization architectures such as that underlying Parallels Virtuozzo Containers do not support the mixing and matching of different operating systems on a single physical system, meaning that server resources across a typically heterogeneous data center can at best be aggregated into separate, distinct pools. Thus, a container running a Windows-based application workload can only take advantage of server resources on other compatible Windows servers. The common OS kernel and associated components have a much larger attack surface than an embedded bare-metal hypervisor, and are shared across multiple containers, making OS virtualization considerably more vulnerable to security breaches than a hardware virtualization architecture. The dependence on a shared underlying OS also limits availability, since an OS failure could impact multiple virtual server environments. To be fair, Parallels’ recently announced support for Windows Server Clustering and Red Hat Clustering helps to mitigate this issue. Finally, the migration of live, running virtualized workloads from one server to another is more challenging in an OS virtualization architecture. The underlying operating systems must be compatible, and even then, migrations today may require reboot.

Hardware Virtualization
Though hardware virtualization approaches tend to satisfy a significant percentage of the criteria for data center readiness, there is considerable variation among the competing
architectures in the market. VMware, which debuted as a hosted virtualization architecture in 1999, transitioned to a bare-metal hypervisor with the introduction of ESX Server two years later. The Xen architecture originated in an open source project at the University of Cambridge and first came to market in 2003. And Microsoft is expected to release the first version of the Hyper-V platform in the second half of 2008.

Despite their different origins, the Xen and Hyper-V architectures have a number of common design elements (see Figure 2). Both contain a set of separate partitions (or domains) running side by side on a thin hypervisor layer. Running inside each partition is a guest operating system, which the hypervisor manages and schedules across the physical CPUs. The first partition, known as “domain zero” in Xen terminology and the “root partition” in Hyper-V, is effectively a “parent” partition that runs a guest OS with special privileges, including automatic booting along with the hypervisor, direct access to the physical hardware, and management capabilities. In the Xen architecture, the first partition (domain zero) can run modified versions of Linux, NetBSD, or Solaris; on Hyper-V, the initial or root partition must run Windows Server 2008. Additional Windows, Linux or UNIX guest OSes in child partitions are started and managed through the parent partition.

![Figure 2. Basic Xen / Hyper V Architecture](image-url)
Xen and Hyper-V are both hardware virtualization architectures, and provide pooling of virtual CPU, memory, networking and storage resources across multiple physical systems. These distributed resources can then be dynamically allocated to VMs based on priority. In Xen, these resources are managed through a Global Resource Pool, and administrators may adjust allocations using VM-level controls. A virtual switch in Hyper-V enables Windows Network Load Balancing (NLB) across physical servers. The architectures also meet basic readiness requirements for data center scaling and performance, though Hyper-V has yet to be proven in production environments. The availability story in each architecture is based around clustering, including support for both host and guest OS clustering in Hyper-V. Microsoft supports live backups through Windows Server 2008 Volume Shadow Copy Services, and minimally disruptive migration of VMs between systems through the use of its Quick Migration facility. XenSource provides hot NIC bonding for failover, and XenMotion for migration of live workloads.

Both platforms also offer a rich set of tools and interfaces for manageability and extensibility. XenCenter can manage both physical and virtual environments in the Platinum edition of XenServer, and XenSource provides a Command Line Interface (CLI) and XenAPI for integration with other management scripts and tools. Microsoft Windows System Center will provide proven management capabilities across virtual and physical environments, along with Windows Management Instrumentation (WMI) interfaces and Hyper-V APIs for custom development and integration. Finally, the two companies’ engineering teams have been working together to deliver support and management of the other’s VMs on their respective hypervisor platforms.

But while the Xen and Hyper-V architectures meet a majority of the readiness criteria for data center deployment, both suffer from some significant shortcomings. For one thing, both architectures take an OS-centric approach. The primary or “parent” guest OS running in the initial domain or partition plays a key role, providing direct hardware access as well as management and control functions for the guest OSes running in the “child” partitions. While this approach enables the architecture to take advantage of advanced management capabilities, such as those provided by Windows System Center in Hyper-V, the platform-wide dependence on the parent-level guest OS can compromise reliability and security. For example, a fault or security breach in the parent-level guest OS can affect software running in the other VMs, which depend on the parent-level guest for management and other services. This makes the primary guest a potential weak link in the chain of processing on the virtual platform.

In general, the larger the footprint of the primary guest OS, the greater the exposure to faults and security issues. For this reason, it is highly advisable to run a more tailored OS in the primary partition, such as a Windows Server Core installation on Hyper-V and a relatively compact Linux distribution on the Xen platform. Even so, these OS builds can be quite large - in the range of 100’s of megabytes for Linux to more than a gigabyte for Windows Server Core - relative to the size of the hypervisor itself. This level of exposure is not present in the
virtualization-centric VMware architecture, in which hardware access is provided by the hypervisor, and management and other advanced capabilities are provided in higher-level layers of the virtual infrastructure. Virtual machines in the VMware architecture are completely independent of each other, ensuring a high level of fault and security isolation.

A second shortcoming of Xen and Hyper-V for data center deployments is their lack of hardware dependence: both architectures require the latest Intel or AMD hardware assists in order to support a full range of Windows and Linux guest operating systems. Specifically, each hypervisor will only run on systems containing Intel VT or AMD-V (aka “Pacifica”) hardware acceleration extensions, which are present in the latest Xeon and Opteron chips. The extensions are needed to run Linux guests in Hyper-V, and to run Windows guests in Xen. The two architectures also require modified OS kernels and special paravirtualized device drivers to be run in the VM environment to facilitate enhanced I/O and networking performance. VMware does not have these restrictions: though the VMware ESX Server 3i hypervisor fully exploits hardware assists, it can run without them. Similarly, paravirtualization is not a prerequisite for any VMware guest operating systems. VMware is thus able to run effectively on any x86 or x64 based system in a data center, including servers that have been in place for several years, and to virtualize legacy Windows and Linux based workloads without the need for OS modifications.

So, given the mixed report card for Xen and Hyper-V, how does VMware stack up against the readiness criteria for data center deployment?

**VMware: Ready for the Data Center**

The VMware bare-metal architecture provides a set of fundamental advantages that make it particularly well suited for use in the data center:

- Virtualization-centric design; purpose-built hypervisor architecture. The ultra-compact ESX Server 3i kernel is designed to do one thing very well: run virtual workloads.

- Complete independence of VMs from a general-purpose OS and from one another, helping to ensure strong fault and security isolation.

- Lack of dependency on hardware acceleration features or paravirtualization, enabling the platform to run across all x86 and x64 systems in the data center.

- Advanced design features that enable scalability and performance for a range of virtualized application workloads:
  - Memory: unique memory over-commit feature, which enables greater efficiency through transparent page sharing and memory ballooning. The ESX Server
hypervisor saves memory capacity by automatically recognizing all pages that are identical between VMs in read-only mode, and very carefully allowing multiple VMs to share the same physical memory pages. With ballooning, the hypervisor and guest OS use their joint intelligence to determine with pages should get forced out of memory. The net effect of these two VMware-unique features is the ability to achieve greater consolidation ratios with Virtual Infrastructure.

- **I/O:** direct driver model, in which drivers are optimized for VMs and I/O performance is designed to scale up as new VMs are added. This feature, which VMware added to the most recent release of the ESX Server 3 and 3i hypervisors, is unique to the VMware architecture. In this model, the I/O stack is embedded in the hypervisor, allowing I/O workloads to run on the same processors that are running the VMs. Context switching is reduced because I/O control flows from the VM to the same context in the hypervisor, and more processor cores are utilized for I/O. This enables users to more efficiently scale their VMs across all cores available in the underlying physical server platform.

- **Storage:** As an extra virtualization layer in the storage architecture, the VMFS clustered file system enables users to manage the storage behind their VMs as a single large LUN. This approach, which is another source of differentiation between VMware and other hypervisor architectures, provides greater independence and portability of storage, accelerates provisioning, and increases administrative simplicity and flexibility.

  - The plug-and-play capabilities of ESX Server 3i, which can power on new servers, incorporate them into a DRS (Dynamic Resource Scheduling) cluster, and automatically re-balance VMs in the cluster to most effectively utilize the newly expanded pool of hardware resources.

  - Proven and time-tested VMM (virtual machine monitor) architecture, currently in its sixth generation and in use by millions of customers; along with the bare-metal hypervisor architecture of ESX Server, which is continuing to build on its 3.x release.

Overall, the VMware architecture satisfies the essential criteria for data center readiness. VMware’s hardware virtualization platform outdoes Parallel’s OS virtualization offering in a head-to-head comparison, as shown below in Figure 3, and VMware also scores better against the readiness requirements than competing hardware virtualization architectures. Of all the hypervisor architectures currently in the market, VMware comes the closest to providing the reliability, security and performance of native hardware, which is a true indicator of its fitness for enterprise deployments.
Where Do We Go From Here?

Though core virtualization suppliers have made great strides in building and outfitting hypervisors for the data center, much work remains to be done. The hypervisor is far from becoming a commodity, despite arguments to the contrary. The major platform vendors should continue to enhance hypervisor architectures to provide capabilities such as the following:

**Continuous availability of virtualized applications.** Virtualization platforms must afford data center managers the ability to keep critical applications up and running through all types of outages, planned and unplanned. Vendors must improve their platforms and infrastructure-level applications to provide automated failover of transaction-oriented ERP and other database-driven applications with no loss of data or interruption of service. Virtualized server backups should enable rapid recovery of both data and applications in a consistent state, without impacting processing. The platforms must also allow non-disruptive administration and maintenance of guest OSes and applications.

**Greater scalability and performance of I/O-intensive workloads.** Server I/O poses a significant challenge in virtualized server environments. Virtualized servers demand greater network bandwidth than traditional physical servers, and they need connections to more networks and storage to meet the requirements of the multiple applications they host. One solution is to dedicate connectivity to each VM, but this is not always possible or cost effective. Excessive sharing of physical host bus adapters (HBAs) and network interface cards (NICs) by multiple VMs can make I/O the resource bottleneck, severely degrading application...
performance. New approaches are required to allow I/O performance to scale up efficiently with compute performance.

**Enhanced security of hypervisor architectures.** As the hypervisor increasingly occupies the strategic layer between hardware and applications, vendors must take additional steps to fortify and protect it against security threats of all types. This includes hypervisor hardening steps such as further shrinking the footprint to minimize the attack surface, and protecting or regulating access to any external interfaces. Vendors should also focus on further strengthening the isolation of VMs and their associated data; encrypting sharing and communications between VMs and between each VM and the hypervisor; and authenticating user and administrative access to VMs. To accelerate innovation at the virtual infrastructure level, platform suppliers would be well advised to enable select third-party vendors to build security solutions for their virtualized environments, as VMware is planning to do with its VMsafe API sharing program.

**Improved virtual storage capabilities.** While fibre channel-based Storage Area Networks (SANs) continue to be the primary shared storage platform in virtualized environments, vendors should work to provide improved support for other popular storage approaches and protocols, including iSCSI-based SANs and Network Attached Storage (NAS) products. In addition, platform suppliers should support storage virtualization enhancements such as thin provisioning for increased storage utilization, and N-Port ID Virtualization (NPIV) for greater visibility from VMs to physical storage. Finally, the major virtualization platform vendors need to enable dynamic and non-disruptive migration of virtual machine disk files across storage arrays, such as VMware has done with Storage VMotion.

**Interoperability of different VM formats.** To ease VM portability and encourage the virtualization of more applications, platform vendors must work diligently to implement the Open Virtual Machine Format (OVF), the multi-vendor packaging and distribution format for virtual machines that has already been accepted by the Distribution Management Task Force (DMTF). The OVF standard, once fully implemented, will allow users to routinely import and convert a given VM, regardless of origin, to run on any major server hardware and hypervisor platform. This will motivate third party software and service providers to adapt their products and services to run in VMs, thereby creating a large and uniform market for all manner of virtual appliances.

**Unified management of physical and virtual environments.** A consistent and compatible set of tools and interfaces is required to enable users to manage different flavors of VMs across major hypervisor architectures, and also to manage applications across both physical and virtual environments.
Full automation of VM monitoring, maintenance and lifecycle management functions. As VMs proliferate in enterprise environments, policy-based tools are needed to automate the monitoring and management of virtual machines across complex environments, and to ensure that maintenance functions such as guest OS upgrades happen reliably and without human intervention.

Enhancements such as these will make virtualization architectures even more capable of addressing the rigorous demands of data center environments.

Taneja Group Opinion

As virtualization products begin in earnest to make their way into the data center, hypervisor architectures will matter more than ever. Server virtualization provides an ideal platform on which to run data center workloads of the future: network (or “cloud”) based component software and services that run across multi-core systems. By pooling underlying hardware resources across multiple servers, virtualization enables applications to fully utilize available compute, memory and I/O capacity.

But to securely and reliably run and manage data center workloads, a hypervisor must go beyond basic virtualization capabilities. In particular, a hypervisor architecture must satisfy ten core readiness requirements before it can be deemed fit for data center deployment. Of all the OS and hardware based virtualization platforms that are currently competing for end user attention and investment, the VMware hypervisor architecture comes the closest to meeting the standards for data center readiness. VMware ESX Server 3i aspires to deliver the reliability, security and performance of native hardware, which makes it a compelling choice for enterprise use.

While VMware and other virtualization platform providers have made significant progress in designing and preparing their hypervisors for data center deployment, a lot still remains to be done. We believe that the large and growing market opportunity, coupled with intensifying competitive pressures, will drive considerable innovation in hypervisor architectures and features over the next few years. However, those suppliers that don’t make the grade in the data center will likely be excluded from consideration. That said, we are hopeful that other x86/x64 based virtualization vendors will soon join VMware in satisfying the rigorous demands of data center users.