



INX Data Center Virtualization Practice

V M w a r e v S p h e r e W h i t e p a p e r

vSphere™ 4 Enhancements – Kaplan/Jones

04/21/2009

Notices:

© 2009 INX Inc. All Rights Reserved. This document and its contents are the confidential and proprietary intellectual property of INX Inc. and may not be duplicated, redistributed or displayed to any third party without the express written consent of INX Inc.

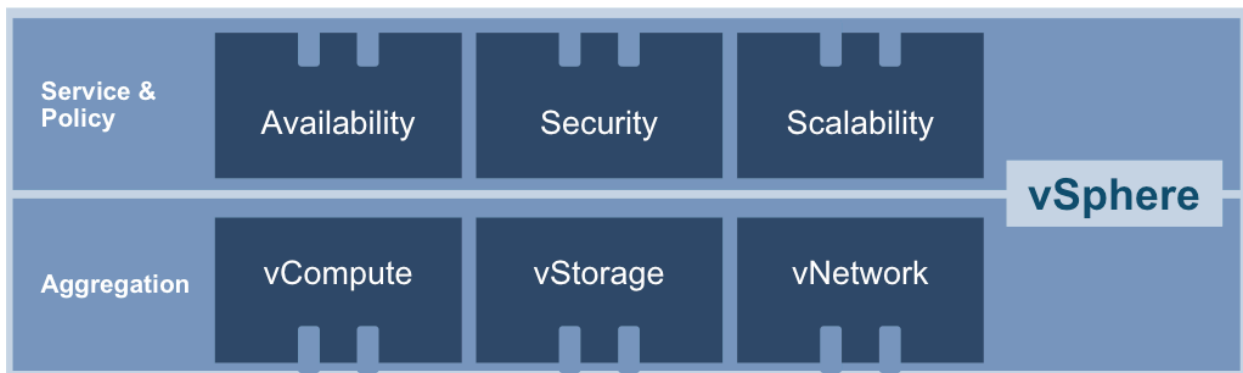
Other product and company names mentioned herein may be the trademarks of their respective owners.

Introduction

On April 21, 2009, VMware made the largest announcement in the history of the company with the introduction of vSphere 4, the rebranded new release of ESX 4. vSphere™ 4 (commonly referred to as “vSphere”) is a “Cloud Computing Operating System” whose domain encompasses an entire datacenter, including VMware’s latest version of their Virtual Data Center Operating System (vDC-OS). vSphere enables 100% data center virtualization by providing much higher levels of performance, reliability, security and management. vSphere is a foundation for running both internal clouds owned by the IT staff and external clouds owned by service providers, and uses federations and standards to facilitate seamless bridging between the two. vSphere renders Information Technology into a service, making technology more efficient, more flexible and much easier to operate.

As a vCloud OS, vSphere supports leading industry manufactures such as Cisco, NetApp, EMC, Symantec and many others that leverage its capabilities to provide increased functionality. Cisco’s Unified Computing System (UCS) and EMC’s next generation Symmetrix, V-Max, for example, are **highly** dependent upon on Cloud computing and thus vSphere as a means of transcending the normal boundaries of compute and storage platforms. Cisco’s Nexus 1000v Virtual Switch does, in fact, require vSphere to run at all.

vSphere™ 4 Architecture



VMware vSphere aggregates former IT silos of server, storage, network and policy into an automated service delivery model. Application service levels such as availability, security and scalability can be enabled simply and uniformly for any application running inside VMware virtual machines.

Availability

In addition to VMotion, Storage VMotion and HA, VMware vSphere introduces two new capabilities: VMware Fault Tolerance and VMware Data Recovery.

VMware Fault Tolerance enables a shadow VM to run on a second ESX host, receiving and executing each memory operation and each CPU instruction. If the primary host and protected VM fail, processing is seamlessly cut over to the shadow VM running on the alternate ESX host without requiring a reboot. A new shadow VM is



then automatically configured to run on another ESX host. Fault Tolerance is a simple, selectable feature that can be implemented for specific VM's deemed to be critical to the operation. Fault-Tolerance eliminates the need to purchase and manage expensive clustering software from Microsoft and other manufacturers.

VMware Data Recovery is fully integrated with vCenter Server. VMware Data Recovery is a virtual machine that provides simple agentless backup and recovery for virtual machines in smaller environments. It restores individual files or entire images, and uses a wizard to create and schedule backup jobs that work even as VMs are moved by HA, VMotion or DRS. Also included is software-based data de-duplication technology that saves a significant amount of space and more effectively leverages the backup window.

Security

VMware vSphere 's new VMsafe and vShield Zones enable efficient application of security policies.

VMware VMsafe enables, through published API's, 3rd party vendors to provide virtual appliances that automatically protect everything running within the vSphere virtual machines, thus no longer requiring the deployment of individual agents. In a physical world we have to over provision security appliances and run everything through them in anticipation of handling peak workloads. There is no such limitation or over-engineering in the VM world. With vSphere, we can assign virtual security specifically to individual VMs. VPN, Intrusion protection, virus protection and anti-root kit are all facilitated. VMsafe includes the ability for plug-ins from security manufactures such as Symantec, McAfee, Check Point, Trend Micro and Altor.

VMware vShield Zones is an application aware firewall technology that is built for virtualization and that uses the same VMsafe APIs. vShield Zones build a firewall "fence" around a VM cluster and ensures that the "fence" protects the VMs regardless of where they are running. vShield Zones allows an operation to build even higher security and separation into an already highly consolidated data center.

Scalability

DRS already facilitates unparalleled dynamic and seamless allocation of resources within the data center, and within VMware vSphere. vApp enables seamless application movement between clouds. VMware vSphere takes scalability to new levels by enabling plug-ins for third-party manufacturers. Some of the first manufacturers capitalizing on this scalability include Cisco, EMC and NetApp.

VMware vApp is a logical grouping of virtual machines comprising a multi-tiered application such as a Web server, application server and database back-end. VMware vApp creates a standard method that describes an applications operational policy along with specifying a mechanism for moving applications between internal or external clouds. In this way, resources such as bandwidth, storage and compute can be allocated as required whether internal or external to the organization.

Cisco's Unified Computing System (UCS) is based on the new Intel 5500 processor along with a patented memory controller that enables far more memory per server than afforded by traditional blade server architectures, helping to alleviate the biggest bottleneck for hosting large numbers of virtual machines. A single, fully populated UCS can host thousands of virtual machines. UCS utilizes role-based management to provision computer infrastructure with automatic coordination between server, network and storage domains. Policy-based management ensures allocation of resources that are in alignment with network, security and compliance policies.

Cisco's Nexus Family of Switches allows a Converged Network Adapter (CNA) to replace separate Ethernet and Fibre Channel HBAs in physical ESX hosts, thereby dramatically reducing cabling and network adapter requirements. Nexus enables storage to truly become a function of the network for fiber channel, iSCSI and NFS protocols –

leaving the decision to the architect as to what will provide best performance for a given environment (ref. [Unlocking network performance bottlenecks with Cisco's Nexus family of switches](#)).

EMC's V-Max is the next generation of Symmetrix – the Enterprise storage class that EMC has sold for years. Although V-Max does not **technically** require VMware vSphere, V-Max is highly dependent upon Cloud Computing and therefore, by extension, vSphere in order to fully realize V-Max' capabilities.

NetApp and Snap Manager for Virtual Infrastructure (SMVI) completely integrates into vCenter and allows the administrator to directly leverage NetApp's native WAFL filesystem for snapshot management. Snapshot creation is handed off to the ONTap OS and leverages the fact that WAFL "knows" where every block is located, thus ensuring that snapshots are created in the most efficient manner possible.

NetApp Cloning creates clones (both file and volume clones) significantly faster than can be performed through the vSphere software only solution. NetApp is integrated into vCenter via the Rapid Cloning Utility v2.0, a utility that significantly saves time and disk space. Controller data such as LUN mapping, de-duplication status, aggregate status, etc. is presented to the administrator along with available ESX hosts.

vCompute

	ESX 3.5	vSphere
CPU	5 vCPUs	8 vCPUs
Memory	64 GB per VM	256 GB per VM
Network	9 Gb/s	40 Gb/s
IOPS	100,000	200,000+

Last year, IBM demonstrated how it could double the number of Exchange mailboxes supported on a single server by breaking up Microsoft Exchange into smaller copies with a smaller number of CPUs assigned to each virtual machine. The aggregate performance was superior, enabling an ESX 3.5 host to double the record for number of Exchange mailboxes. Additionally, the other advantages of virtualization apply such as high-availability, the ability to continuously replicate it for disaster recovery and the ability to snap the server to facilitate upgrades, troubleshooting, patching, etc.

The inability for Exchange to efficiently scale up to the large number of cores available today is applicable to software in general. With servers rapidly accommodating ever more cores (Intel is about to go to 8 cores per CPU), this is becoming a bigger liability for physical servers. VMware vSphere now supports up to 8 virtual processors (vCPU) with 256 GB per virtual machine, and with support for Intel's Nehalem processor vSphere processor overhead is dramatically decreased. In an Oracle DB stress test, a single VM scaled up to 8 vCPUs and achieved a sustained I/O rate of 250Mbytes/second – all with less than 15% CPU. vSphere VM's on Intel 5500 CPUs can achieve a nearly linear vCPU scaling up to 8 vCPUs. vSphere enables record performance when running even the largest data bases as virtual machines.

VMware vSphere not only runs large workloads with superior performance, it also does it more efficiently from a power perspective. CPU throttling and other power management techniques are incorporated in order to minimize power requirements. During periods of low resource demand, such as overnight, VMs can be automatically be VMotioned off larger ESX hosts which are then be powered down, and repowered back up when required. CPU throttling and other power mgt techniques incorporated. VMs are moved off larger machines that can then be powered down automatically, and repowered when required. In this way workloads only use the power actually needed.

vStorage

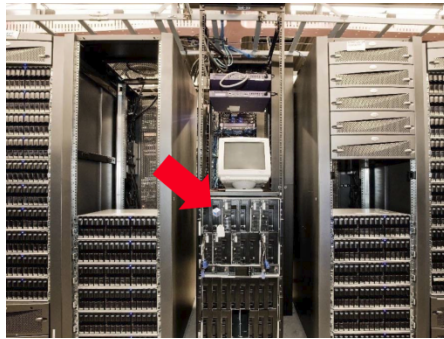


Figure 1: Oracle Stress Test

As demonstrated in an Oracle stress test shown, a single VMware vSphere VM running RedHat Enterprise Linux 5.1 can easily support over 200,000 IOPS, or 250MB/sec of disk I/O. The picture shows the 510 additional disk drives required before the IO rate was topped out.

VMware vStorage Thin Provisioning enables over subscription of storage. In this way, the number of allocated “disks” adds up to much more storage than what is actually owned. vStorage includes advanced management capabilities to ensure the over provisioning does not adversely affect the most important workloads, and takes advantage of technologies such as disk duplication. Most importantly, however, is the publication of an open API that allows vSphere to leverage the thin-provisioning offered by the storage vendors. Such hardware-based solutions are often higher performance solutions but historically required managing the storage via another interface. vSphere allows the storage vendors to write plug-ins that allow the administrator to manage all aspects of thin-provisioned storage right from the same, familiar vSphere console.

vNetwork

Today’s ESX virtual switch (vSwitch) is a rudimentary layer 2 switch with no routing or QOS capabilities, analogous to the Ethernet switches prevalent during the mid-1990’s. The virtual switch requires individual management as part of every ESX server, and all switches must be manually configured the same. The virtual machine’s policies must be manually reset as the virtual machine VMotions around the environment. To add insult to injury, the network administrator has no purview into the virtual machines or into the vSwitch. The network administrator can only see up to server level and can do nothing in terms of configuring and controlling the virtual network environment with security and network policies.

VMware vSphere’s vNetwork Distributed Switch (vDS) aggregates all networking capabilities and creates a vSwitch whose components and policies span all ESX hosts. Administrators can then apply policies such as

availability, security and maximum latency to a virtual machine. That policy then follows the virtual machines as VMotion moves the VM around. VMware vDS also provides an open API, thus allowing third-party manufacturers to provide their own switch plug-ins.

Cisco's Nexus 1000v virtual switch plugs directly into vSphere and in essence takes over and extends the functions of vDS. Nexus 1000v finally places back into the hands of the Cisco Network Administrators the management of the virtual switches with the familiar command line tools that are used in the day-to-day management of physical Cisco switches. For a detailed comparison between vDS and the Nexus 1000v, see our companion white paper, [The Nexus 1000v Virtual Switch](#).

VMDirectPath for Virtual Machines offloads I/O processing from the hypervisor by allowing virtual machines to directly access the underlying hardware devices.

Virtual Data Center Management

vCenter Suite is a collection of management tools that allow service and policy implementations to facilitate availability, scalability and automation. VMware vCenter 4 “speaks” to the ESX hosts that control virtual machines and connects through both thick and thin clients. vCenter 4 includes many new capabilities designed to simplify and enhance management.

vCenter Server Heartbeat enables monitoring, replication and rollback for instances of vCenter Server, especially those servers running on a physical server. Each instance of vCenter Server enables management of 3,000 VMs and 300 hosts.

vCenter Linked Mode is a long awaited new capability designed for very large enterprises. vCenter Linked Mode allows up to 10 vCenter instances (or Datacenters) to be aggregated into a single vSphere client and includes an advanced Search capability designed to assist in finding VMs throughout the linked vCenter instances. As with the other components of vSphere, partners can enhance vCenter with SDK and toolkits.

VMware vCenter AppSpeed enables the monitoring and managing of individual VMs. AppSpeed provides more fine-grained DRS controls, performance statistics, disk usage reports and more wizards.

VMware vCenter Host Profiles enables more fully automated host provisioning through the use of host “templating”. Host Profiles facilitates efficiently and safely adding new hardware to the data center and monitoring of the new hosts to ensure they’re still in compliance.

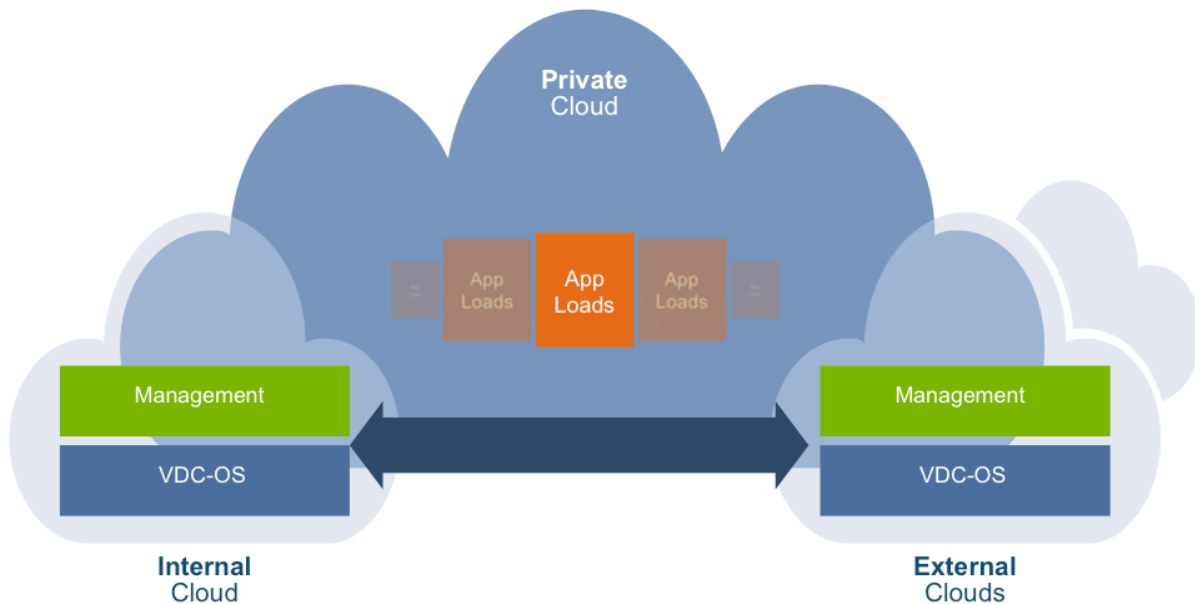
The VMware vCenter vCloud Plug-in allows management of, and authentication to, both internal and external clouds. the vCloud plugin uses a variety of techniques to bridge together and offer private cloud using resources from both services.

vCenter Orchestrator is a workflow engine that automates tasks for VMware vSphere and enables orchestration between multiple solutions, and to integrate the vCenter Server operations into automated processes. Orchestrator also allows integration with other management and administration solutions with its open plug-in architecture.

Record and Replay Virtual Machine Execution — ESX/ESXi 4.0 provides the ability to record and replay the execution of a virtual machine for forensic or debugging purposes. APIs enable third parties to control this functionality.

VMware vSphere™ 4 Features Summary

Why is vSphere yet another major evolutionary jump in VMware arsenal? Because vSphere, coupled with vCloud, brings organizational infrastructure closer to the realization of a truly seamless and connected computational resource whose boundaries are, virtually, limitless. Even the connectedness of the much touted (but extremely proprietary) High Performance Computing clusters pales when compared to the cooperation, transparency, and resiliency presented to us in the form of vSphere and vCloud.



The following is a table of new vSphere features, segregated into high-level functional categories. The feature list, while not exhaustive, represents the most compelling technology behind vSphere.

New Feature	Discussion	Differences from VI3	Why we Care
Architecture			
64 bit VMkernel	The VMkernel is the core component of the ESX/ESXi 4.0 Hypervisor and it is now a 64bit operating system. The 64bit architecture provides much greater host physical memory capacity, enhanced performance and more seamless hardware support over earlier releases.	VI3 ran a 32bit VMkernel	The improved architecture translates into much better overall performance and scalability. In addition, the 64bit architecture allows the VMkernel to provide much greater capacity to each of the Guest OS': support for up to 64 logical processors, 512gb of ram, 512vCPU's/host, 8vCPU's and 256gb/Guest.
64-bit Service Console (COS)	The Service Console for ESX 4.0 remains but as a 64bit version of Red Hat Enterprise Linux (RHEL) 5.	VI3 runs a COS based upon RHEL3, a 32bit OS.	Many customers run enhancement processes inside the COS (e.g., HP Insight client). RHEL3 has been, for all intents and purposes, a dead OS for many years. RHEL4 was not up to par with RedHat's tradition of well developed OS' and was soon supplanted by RHEL5. Adoption of RHEL5 by VMware, and leveraging the 64bit kernel, is an excellent decision.
CPU Power Management Aware	ESX/ESXi 4.0 supports Enhanced Intel Speedstep and AMD PowerNow power management capabilities. With Dynamic Voltage and Frequency Scaling (DVFS), ESX/ESXi 4.0 can reduce power consumption on hosts that are running at reduced loading. Power will also be tracked via the vSphere client and will be considered to be a performance metric.	VI3 did not support either technology.	The management of power usage down to the CPU level is critical when businesses look for technology that more finely controls the expenditure of their OpEx dollars. Being able to gradually "turn-down" consumption provides an organization with the tools needed to better track, predict, and manage the behavior of the infrastructure.

New Feature	Discussion	Differences from VI3	Why we Care
Installation			
Kickstart Replacement	The ESX scripted installation is similar to RHEL scripted installations, however, the two are incompatible.	VI3 leveraged scripted installs as well.	Automated installations are more powerful and flexible. Fully automated ESX server bring-ups can address configuration control and compliance issues in very large deployments (as found in VDI).
USB Support	The Kickstart configuration file and the ESX 4.0 installation files can reside on a USB device. NOTE: Do not confuse this USB support with full USB support for Guest OS'.	VI3 had no native support for USB devices. Some system manufacturers could use a USB thumb-drive for local booting.	USB support provides an alternative solution over pulling installation resources from across the network. Scenarios where this can be critical would be remote locations/offices.
Host Profiles	The task is to maintain consistency across all ESX 4 Hosts, regardless of when they were built. Host Profiles is designed to address this need for those who do not wish to delve into the arcane art of Kickstart scripting. Not a replacement for scripting, Profiles are meant to provide a template-oriented mechanism for ensuring compliance with the standard.	VI3 supported scripting but not template-based builds	While scripts are nice, they should only be used for those configuration steps that require different parameters from host to host. Host Profiles smells an awful lot like Microsoft Group Policies – a tool that is horribly underutilized in the MS world but one which could save organizations mounds of money in managing system configurations (not to mention doing away with third party config apps that are simply not needed).
Power Management			
DPM	DPM now fully supports Out-Of-Band IPMI and iLO management interface Remote Power On features.	VI3 only supported Wake-On-LAN to trigger a power-on event. DPM was also experimental.	DPM is no longer experimental and also supports Out-Of-Band Host power management signaling in addition to In-Band Wake-On-LAN. The issue has been an historical un-reliability of Wake-on-LAN implementations in the chipsets. Adding support for the OOB management interfaces adds just that many more options to the mix – and the OOB systems are built as truly independent of the motherboard operation.

New Feature	Discussion	Differences from VI3	Why we Care
Storage			
vStorage	Storage plug-ins from various storage vendors will provide direct access to such features as backend “snapshot” technology and storage multi-pathing.	Few vendors provided integration	Leveraging native storage functionality results in vastly superior performance when compared to the VMware native capabilities.
Storage VMotion - GUI	Storage VMotion has been around but relied on either third party GUI plug-ins or the VMware CLI. Storage VMotion comes into its own with an integrated GUI.	VI3 relied upon third party GUI's	Storage VMotion has actually proven its worth when organizations found themselves having to deal with moving from Direct Attached Storage (DAS) to Shared Storage (NAS, iSCSI, FC) or in having to deal with the shortcomings of VMFS, extents, and LUN expansion. Third party GUI's were not always reliable and the CLI was rather cryptic.
Dynamic VMFS expansion	VMFS will now have the capability of being expanded after a storage LUN has been expanded.	VI3 VMFS could only be expanded by adding VMFS extents – vastly different.	Adding VMFS extents, while easy, pose an element of risk that many admins deem unreasonable. Only the Master extent contains the meta data that describes where all of the blocks are located – loose the Master and all is lost.
Thin Provisioning	Just like the big storage vendors, vSphere will be able to create VMFS stores that do not pre-allocate all of their storage. This feature is very reminiscent of the current VMDK creation capabilities in Workstation and Fusion.	VI3 does not have this feature.	<p>We often see a tendency to over-commit or over-estimate on storage requirements, and with good reason -- we want our applications to have adequate storage. So, we over-estimate and in the process insure that our server has storage when it needs it. After all, running out of storage on a critical server is a sure ticket to being shown the door.</p> <p>Problem: over-commits result in wasting the storage we have (because the over-commit results in nearly 40% unused space), thus more quickly using up our shared storage, thus causing us to have to go out and buy more. Then the cycle starts all over again.</p> <p>Thin-provisioning was designed to arrest the cycle by allowing real storage use to grow with the demands of the OS. The OS thinks it has 100gb but in reality only 30gb of space is actually in use.</p>



New Feature	Discussion	Differences from VI3	Why we Care
			Caution: Thin-provisioning introduces and entirely new set of challenges – actively monitoring real storage and anticipating real storage growth. One can still be shown the door if a critical server really runs out of space!

New Feature	Discussion	Differences from VI3	Why we Care
Networking			
Network Distributed Switch	vDS provides the ability to define a vSwitch that is distributed amongst multiple ESX hosts, and those portgroups and uplinks that make up the distributed switch. The vDS configuration is then propagated to all members of the vDS.	VI3 only provided the “vSwitch”	In large installations the configuration of vSwitches and portgroups can be cumbersome at best. vDS renders this activity very simple.
Nexus 1000v	<p>The Nexus 1000v is the ultimate in virtualized networking. Think of the 1000v (Virtual Ethernet Module or VEM) as akin to a Catalyst 6509 line card. The VEM leverages the vDS API and thus is intimately aware of what is happening within the vSphere HOWEVER the VEM also operates as a fully functional piece of native Cisco gear. Add the Virtual Supervisor Module (VSM) and one has a “Switch” comprised of VEM line cards and those line cards are spread among all of the ESX servers (up to 64) in the vSphere.</p> <p>Nexus 1000v, once virtualized, gives the Network Admin (not the Server Admin) the ability to create policy-based VM connectivity (through Port Profiles) and transparently move those port profiles as the dynamics of the vSphere change.</p>	VI3 had no such construct.	<p>VI3 required that the server admins configure the vSwitch using either the VMware CLI or vCenter – thus eliminating the separation of duties that is often found in large organizations.</p> <p>Nexus 1000v moves network operations back into the hands of the Network Admin and leverages the time-tested (and well understood) Cisco CLI.</p>
Management			
vCenter for Linux	vCenter will now run under Linux as well as Windows (NOTE: vCenter for Linux will remain in Technology Preview until further notice from VMware. ref vCenter Server - Linux)	VI3 vCenter only ran under Windows.	For organizations that do not “do” Windows, this is a revelation. Many Linux/Unix organizations were required to install and support a single copy of Windows just so they could manage their Virtual Infrastructure. No longer.
VMware Fault Tolerance	VMware’s answer to complex, application targeted cloning. vSphere FT provides for a clone of a running VM to exactly track the changes within the primary VM. Failure of the primary	VI3’s closest relative to FT is HA, but HA does not provide	FT is the “Universal Clustering” that organizations have been chasing for years. FT provides non-stop operation and is Application agnostic, meaning an organization can have a non-stop cluster



New Feature	Discussion	Differences from VI3	Why we Care
	VM causes an uninterrupted failover to the tracking VM running on a completely different ESX 4.0 host.	uninterrupted service.	without having to invest in proprietary clustering technology for every single application.
vApp	Akin to the “VM Teaming” feature in VMware Workstation, vApp allows us to designate a team of VM’s that operate as a system. Rules may be applied that govern how the “system” is started and in what order. vApp ultimately defines	VI3 had no such offering	vApp ultimately defines that elusive IT goal of providing rule-based dependencies between independently hosted applications. In fact, Business Continuity practices nearly always dictate that it is the *System” uptime that determines whether one is meeting uptime goals or not – yet there has been little success in defining policy that ensures the “System” is operational. vApp does just that – it allows the admin to define the policy that actually governs the viability of the “System”.

Authors:

Steve Kaplan is VP, Data Center Virtualization for INX. Steve can be reached at steve.kaplan@inxi.com, or followed on Twitter at <http://twitter.com/roidude>

Steve Jones is Managing Consultant, Data Center Virtualization Practice, Federal Division for INX. Steve can be reached at steve.jones@inxi.com