



Red Hat Driving Virtualization Standards with Xen and Open Source Community

February 2006

PREPARED FOR

Red Hat

TABLE OF CONTENTS

Executive Summary 1

Introduction.....2

Virtualization and Consolidation2

Virtualization and Hardware Flexibility2

Virtualization and Workload Management.....3

The Emergence of Open Source Virtualization.....3

Paravirtualization Optimizes Guest Performance4

Xen vs. Intel VT and AMD “Pacifica”5

Production-Grade Xen Arriving in Red Hat Enterprise Linux 55

Red Hat Added Value6

Virtualization Management Strategy: Standardize Xen Interfaces7

Future Directions.....7

Red Hat Enterprise Linux and Xen: Value Proposition.....8

IDEAS Summary9

Executive Summary

As Linux matures and becomes more widely used in production environments, its users are becoming increasingly concerned with getting better utilization out of their existing servers, and minimizing the cost of migrating to new systems. When Red Hat introduces a production-grade implementation of Xen virtualization in Red Hat Enterprise Linux Version 5 later this year, Linux users will gain access to a safe, efficient, and integrated virtualization capability that meets these needs. The integration of Xen into Enterprise Linux will result in several important customer benefits, including better support for consolidation; more flexible lifecycle management for hardware and software; improved workload management; and enhanced systems availability.

Further, by making Xen a pervasive part of its Linux distribution, Red Hat will drive virtualization deeply into the mainstream. As virtualization technology such as Xen becomes embedded directly within operating systems, low-level resource management

RECOMMENDED USER GUIDELINES FOR DEPLOYING VIRTUALIZATION

The following are some recommended guidelines for users to prepare for the arrival of Xen and Enterprise Linux 5 later this year:

- » Identify workloads that are suitable for virtualization by understanding their load characteristics over time, including seasonal demands and differences between dependencies on CPU, memory, and I/O resources.
- » Begin experimenting with the version of Xen technology that is currently available in Red Hat Fedora.
- » Make sure virtualization management tools conform to the standard APIs and ABIs that Red Hat defines for controlling virtualized resources and management policies.
- » Assess the dependency of applications on the features of specific Linux distributions to determine if they can be rehosted on a paravirtualized version of Enterprise Linux, or if they require the use of full virtualization and virtualization hardware extensions.
- » Determine the suitability of workloads for virtualization in virtual machines versus other virtualization techniques, including virtual servers, hard partitions, and blade servers.
- » To prepare for live migrations, determine which storage network topology will be strategic in carrying virtual machine images.
- » When Enterprise Linux 5 becomes available, start hosting data on GFS in order to ease migration to a more complete, multisystem virtualized environment later.
- » Avoid overselling the capabilities of virtualization; gauge the overall risk averseness of the organization to determine its level of maturity for adopting virtualization in certain applications.

This document is copyrighted © by Ideas International, Inc. (IDEAS) and is protected by U.S. and international copyright laws and conventions. This document may not be copied, reproduced, stored in a retrieval system, transmitted in any form, posted on a public or private website or bulletin board, or sublicensed to a third party without the written consent of IDEAS. No copyright may be obscured or removed from the paper. All trademarks and registered marks of products and companies referred to in this paper are protected.

This document was developed on the basis of information and sources believed to be reliable. This document is to be used "as is." IDEAS makes no guarantees or representations regarding, and shall have no liability for the accuracy of, data, subject matter, quality, or timeliness of the content. The data contained in this document are subject to change. IDEAS accepts no responsibility to inform the reader of changes in the data. In addition, IDEAS may change its view of the products, services, and companies described in this document.

IDEAS accepts no responsibility for decisions made on the basis of information contained herein, nor from the reader's attempts to duplicate performance results or other outcomes. Nor can the paper be used to predict future values or performance levels. This document may not be used to create an endorsement for products and services discussed in the paper or for other products and services offered by the vendors discussed.

tools such as virtual machines and hypervisors will become commodities. As a result, the real competition will shift to virtualization management solutions. With few standards for virtualization management defined, some providers of these solutions will try to establish leverage by introducing proprietary interfaces at critical management layers. Red Hat's strategy to define documented, open source virtualization interfaces rather than compete with its own proprietary management tools will become a key added value for the industry. By establishing a level playing field for the development of virtualization tools, Red Hat will facilitate the emergence of virtualization management solutions that are based on industry standards rather than proprietary extensions.

Introduction

Virtualization is generating genuine excitement in the IT community. Over the long term, the use of virtualization will cause a fundamental transformation of IT operations. Viewed broadly, virtualization involves decoupling workloads and data from the functional details of the physical platforms on which they are hosted. This decoupling increases the flexibility with which the workloads and data can be matched with physical resources, enabling administrators to develop business-driven policies for delivering resources that are appropriate given specific time, cost, and service-level requirements. Over the long term, virtualization will enable IT operations to be performed with far better economies of scale, maximizing the utilization of existing resources by allowing infrastructures to be managed efficiently, even as they undergo high rates of growth. In the short term, virtualization has the ability to solve some specific problems faced by IT managers today.

Virtualization and Consolidation

The most straightforward application of virtualization is consolidation. Many organizations are finding they need to consolidate a profusion of small servers (i.e., with one or two processors) that are extraordinarily underutilized, often running at 5% of their capacity or less. At this level of utilization, users can dramatically lower their hardware costs and administrative burden if the workloads of 10 to 20 systems can be rehosted on a single processor. Virtualization tools such as virtual machines help to solve this problem by allowing multiple workloads to be hosted simultaneously on larger servers, whereby each application operates under the illusion that it is running on its own computer. Server consolidation through virtualization allows administrators to reduce the number of physical machines that they have to manage, lowering maintenance costs and consumption of cooling and power resources.

Virtualization and Hardware Flexibility

Virtualization can also help to simplify the task of reconfiguring hardware. Because virtualization isolates workloads from details about the servers on which they are hosted, it becomes possible to move a workload from one machine to another without disturbing its application environment. For example, virtual machine implementations typically allow the entire state of a running virtual machine to be captured in a standard file, so that it can be transported across a network. Workloads can be migrated from one physical server to another simply by copying this file across the network. Applications running in the virtual machine continue to

operate in the exact same environment (including operating system and middleware) without any knowledge of the actual physical server they are running on. This capability has a variety of benefits, such as the ability to swap a new server in place of an old one without reinstalling the operating system or application software. This facility also allows administrators to continue supporting older operating systems for which support is no longer available on current hardware.

Virtualization and Workload Management

Virtualization can help administrators increase uptime and balance resource consumption in response to changing load conditions. On single systems, administrators can assign resources to virtual machines at virtually any granularity, from fractions of processors to multiple processors on SMP-enabled hosts. The ability to migrate virtual machines across a network simply by copying files dramatically simplifies the implementation of IT techniques such as:

- » *server load balancing*, whereby workloads are migrated from heavily loaded servers to less busy machines;
- » *high availability (HA) clustering*, in which backup servers stand by to recover from the failure of individual servers; and
- » *disaster recovery (DR)*, in which workloads are copied over geographically significant distances to recover from the failure of entire datacenters or sites.

The Emergence of Open Source Virtualization

In practice, there are a variety of ways to implement virtualization. One of the most basic virtualization functions is the ability to run multiple operating systems simultaneously on a single server. By hosting multiple operating systems on one server, users gain the ability to deploy many workloads on the server in a way that allows the workloads to continue running in their native OS environments. The most flexible way to accomplish this task is to use virtual machines, which create entire computer systems in software that can be operated and controlled as if they were applications. Virtual machines have long been available on mainframes, and have more recently become available for industry-standard X86 systems in commercial products like VMware and Microsoft Virtual Server.

Now, a new virtualization technology has emerged called Xen that provides similar functionality using open source software. Like virtual machine solutions such as VMware and Virtual Server, Xen allows multiple instances of an operating system to run simultaneously on a single server. However, unlike pure virtual machine implementations such as VMware GSX Server and Microsoft Virtual Server, in which virtualized operating system instances are hosted as processes or threads on a standard operating system, Xen has a hypervisor-based design similar to that of VMware's ESX Server and IBM's p5 server Micro-Partitions. As with these systems, the Xen hypervisor runs directly on hardware (i.e., in "Ring 0" of the processor), and dynamically partitions the server into "domains," each of which hosts an instance of a guest operating system. For each domain, the hypervisor guarantees a certain amount of CPU and I/O resources, including network I/O or block I/O for disk storage. The hypervisor itself simply acts as a domain switcher

and thus has a very low profile, providing just enough abstraction of the server hardware to allow effective isolation and resource management between domains. Indeed, early implementations of the Xen hypervisor consumed no more than 42,000 lines of code.

Paravirtualization Optimizes Guest Performance

Xen performs traditional virtualization as well as a technique called “paravirtualization” to manage guest operating systems, which means that guest operating systems require specific modifications in order to run in Xen domains. By contrast, traditional virtual machine tools such as VMware and Microsoft Virtual Server provide “full virtualization,” which means that they theoretically allow any operating system to run unmodified in a virtual machine (however, actual support from vendors may vary, depending on which systems they decide to invest in with testing and certification). The newest version of Xen technology is designed to work with Linux, whereby Linux can control the hypervisor and run as a paravirtualized guest. However, other operating systems can potentially be adapted to support Xen as well. For example, developers have started adapting OpenSolaris, the open source version of Sun’s Solaris operating system, with Xen’s paravirtualization functions.

While Xen’s paravirtualization design constrains the choice of operating systems that can be fully controlled to those that have been appropriately adapted, in return it offers considerable performance advantages over the full virtualization applied in traditional virtual machine approaches. By allowing guest operating systems to cooperate directly with the hypervisor, resource management becomes much more efficient. For example, using the paravirtualization mechanism, Xen can dynamically request that guest operating systems reduce their memory footprint, if necessary. Each paravirtualized guest operating system is required to support a “balloon” memory management driver that the Xen hypervisor can use to dynamically adjust the guest’s memory usage. By inflating or deflating the balloon, the guest OS is forced to use its own memory management algorithms to swap pages out to the virtual disk if the hypervisor needs to reallocate memory to another guest OS. Similarly, deflating the balloon decreases the memory “pressure,” allowing the guest to dynamically reclaim memory. This flexibility can be extraordinarily useful for responding to changing workload conditions, in which one virtual machine may have to suddenly swap in the resources from another machine that is hosting a less-critical task. Such techniques are much harder to apply with full virtualization, because the hypervisor has no direct way to influence the behavior of guest operating systems.

Paravirtualization also helps to solve other problems associated with full virtualization that may seem trivial, but can in fact become quite tedious in practice, such as maintaining the correct time of day in guest operating systems. Some virtual machine environments have a side effect whereby the system clocks in guest environments start to drift because they do not have direct access to the physical hardware that maintains the correct time. Although this problem can be overcome by installing special system tools, paravirtualization does not require such add-ons, since the hypervisor can communicate the time of day directly to paravirtualized guest operating systems.

Xen vs. Intel VT and AMD “Pacifica”

While developers continue to refine the capabilities of virtual machine software, the leading suppliers of processors are also introducing virtualization capabilities at the hardware level. Both Intel and AMD will be shipping production-grade virtualization functions in their respective X86 processors later this year, including Intel's Virtualization Technology (VT, formerly code-named “Vanderpool”) and AMD's technology code-named “Pacifica.” These hardware-based virtualization functions make the process of hosting fully virtualized operating systems more efficient. Without them, traditional operating systems running in virtual machines incur a performance overhead resulting from the need to perform a translation or emulation procedure every time a guest operating system attempts to execute a “privileged” instruction (i.e., a low-level instruction that, in order to maintain consistency, only the host OS has the right to perform). These privileged instructions are frequently invoked during I/O operations, which can become crucial bottlenecks in commercial applications that depend heavily on disk and network access. Hardware-based virtualization functions help to overcome these bottlenecks by managing the execution of privileged instructions of multiple virtual machines. Intel VT and AMD Pacifica thus complement Xen by enabling it to host unmodified (i.e., fully virtualized) guest operating systems together with paravirtualized guest operating systems within the same infrastructure.

Production-Grade Xen Arriving in Red Hat Enterprise Linux 5

Red Hat plans to introduce its production-grade implementation of Xen virtualization in the next release of its fully supported Linux distribution, Red Hat Enterprise Linux Version 5, which is due by the end of this year. A preliminary version of Xen is already available now in Fedora Core (FC), Red Hat's community-driven Linux distribution in which new features can be tested by a broad and diverse base of users before being formally packaged and supported in the commercial Enterprise Linux release. Throughout this year, the FC community will have the opportunity to help test and debug Xen before it goes into production. During that time, Red Hat will continue to prepare its ISVs, partners, and management infrastructure providers for supporting Xen virtualization in their products and services.

Enterprise Linux 5 will support the Xen hypervisor, as well as the paravirtualization functions needed to run instances of different Red Hat operating systems as guests within Xen domains (see Table 1). At minimum, the first release will allow multiple instances of Enterprise Linux 5 to be deployed on a single server, whereby the Xen hypervisor can precisely control the resources assigned to each instance. At the same time that Enterprise Linux 5 ships, or shortly thereafter, Red Hat also plans to issue an update to the current release of Red Hat Enterprise Linux 4 with Xen paravirtualization functions. This update will allow Enterprise Linux 4 to be deployed in Xen domains as well, with full fine-grained control by the Xen hypervisor.

Further, Enterprise Linux 5 will support virtualization hardware extensions from both Intel and AMD. This means that even without the paravirtualized version of Enterprise Linux 4, multiple instances of Enterprise Linux 4 can be consolidated on a single Enterprise Linux 5 system using Intel VT extensions or AMD Pacifica

extensions. In this case, Enterprise Linux 4 would be running fully virtualized, meaning that Xen will not be able to manage the resources of Enterprise Linux 4 guests with the same level of control as it could for Enterprise Linux 5 guests. However, the use of virtualization hardware would still minimize the performance overhead of consolidating fully virtualized Enterprise Linux 4 workloads onto a single server. Red Hat may even extend support for hosting fully virtualized guests to older versions of Red Hat Enterprise Linux, but it has not yet determined how far back that support will stretch.

Table 1. Guest OS Support with Xen in Red Hat Enterprise Linux 5

OS Release	Standard HW	Intel VT / AMD Pacifica HW
Red Hat Enterprise Linux 5	Paravirtualized	Fully Virtualized
Red Hat Enterprise Linux 4 (Current)	-	Fully Virtualized
Red Hat Enterprise Linux 4 (Update)	Paravirtualized	Fully Virtualized

Red Hat Added Value

As open source technology that is optimized for Linux, Xen will most likely be supported in many Linux distributions across the industry. Red Hat has the opportunity to differentiate its Xen implementation from others by adding value in several ways. First, Red Hat will optimize its storage products for use in Xen environments, including the Global File System (GFS) and Logical Volume Manager (LVM). For example, Red Hat plans to adapt GFS so that it supports single-node configurations, allowing users to host data on GFS file systems even before adopting multisystem configurations such as clusters. With this plan, users will not have to undergo a migration to move their data off traditional, single-node-oriented file systems when they start embracing multisystem virtual environments. Xen itself does not depend on file-system implementation details. However, with virtualization, a datacenter can easily triple or quadruple the number of operating system instances in a very short period of time. The environment also becomes much more dynamic. Instances may reside quiescently on a standby machine and suddenly attach to the network. Optimizations are therefore required in GFS to accommodate conditions in which very large numbers of systems attempt to use shared storage simultaneously so that it can efficiently deal with the resulting coherency and contention issues.

Red Hat will also make sure that the necessary expertise about Xen is introduced in its training classes for Red Hat Certified Engineers (RHCNs) so that they can establish credibility for being able to use and support this technology. Red Hat will also work with its OEMs to make sure all of their hardware is supported with Red Hat's Xen implementation. Finally, Red Hat will extend Red Hat Network (RHN) so that it works properly with virtualized systems. Red Hat will introduce the notion of virtualized instances to the RHN schema, so that just as RHN can currently monitor, provision, and update physical machines, it will also be able to do the same thing with virtual machines. As a result, users who have invested in RHN will not have to adopt a new tool to manage their virtual environments.

Virtualization Management Strategy: Standardize Xen Interfaces

As virtualization becomes increasingly embedded directly within operating systems, low-level resource management tools such as virtual machines and hypervisors will soon become commodities. Instead, the value will shift to the tools for managing virtualized resources. From the user standpoint, the question of which standards to apply for managing those resources will become increasingly critical, and competition between various providers of proprietary virtualization management frameworks will heat up considerably as they try to control those standards.

Rather than try to establish its own alternative for managing virtualized resources, Red Hat's strategy is to deliver a stable set of interfaces and scripting tools that can be used to interact with the Xen system at a low level. Because there is no accepted standard and stable interface for managing Xen-based systems, Red Hat is taking the initiative of defining the Application Program Interfaces (APIs) and Application Binary Interfaces (ABIs) needed to control the Xen hypervisor. Because Red Hat is oriented around open source, it will make sure that all of the Xen interfaces are fully exposed and documented in order to establish a level playing field and ensure that any third party can take advantage of the interfaces to add manageability to their products. As part of this effort, Red Hat recently created a code library called `libvirt` with Python and C language bindings that applications can use to monitor and configure Xen.

Red Hat has also started developing a set of interfaces to define policies for the use of virtualization. The interfaces deal with such functions as resource monitoring; monitoring the resources that are consumed by different virtual machines; controlling the level of resources that are allocated to virtual machines; and performing allocation changes between virtual machines. These interfaces will allow users to implement rudimentary policies for automating resource assignment (i.e., limiting the amount of CPU capacity a virtual machine can consume by gracefully shutting it down if it goes over a certain level).

Red Hat will apply these interfaces to adapt its existing tools – including the Anaconda, KickStart, and Red Hat Network installation programs– for use in virtualized environments. Red Hat feels that the new Xen capabilities should plug into existing infrastructure management tools as much as possible, rather than demand the adoption of an entirely new management and provisioning framework. Not all systems will be virtualized, and Red Hat wants to make sure that the same tools and scripts can be used to provision “bare metal” physical servers and virtual machines, so that users will not be forced to adopt new tools and new problems.

Future Directions

In the future, Red Hat will apply its concept for “stateless Linux” to virtualized environments. With stateless Linux, large numbers of networked systems are set up as replaceable “appliances” that do not maintain any significant state locally. In the scheme, an administrator would configure hundreds of systems as clones of a master system, and the stateless Linux system would make sure that all of them are kept synchronized whenever the master system is updated. In virtualized environments, the problem of maintaining different file system layouts on each node is significantly compounded by the large number of systems that can be instantiated with virtual

machine technologies such as Xen. For example, when Xen is used to consolidate 5 to 10 virtual machines on a physical server, the utilization of that server may improve, but its management could become more difficult as well. In this case, stateless Linux would be used to deploy a single file system image that is employed by many virtual machine guests at the same time. Indeed, Red Hat envisions the use of a single, unified, read-only file system, with one instance per hundreds of virtual machines. It would be implemented as an identical read-only file system that can be used by more than one virtual instance. The first incarnation of that system will use an in-memory writable file system for caching temporary changes.

Red Hat also plans to strengthen the integration between virtualization and its clustering software, Red Hat Cluster Suite. As users start to embrace multisystem virtualization, in which virtualized workloads migrate between physical systems across the network, they will increasingly take on procedures that are familiar in the HA clustering domain, such as monitoring the uptime of a virtual machine; generating alerts when a system goes down; relocating a virtual machine; and defining distributed policies for responding to failure events. Red Hat Cluster Suite already implements many of these procedures for physical servers, although its policies are currently generally defined in the context of applications. Red Hat is planning to converge Cluster Suite capabilities with the policy management for virtual machines using the new virtualization management interfaces discussed earlier. Ultimately, Red Hat envisions Cluster Suite evolving into a single layer of software that can implement HA clustering on physical servers and also manage distributed virtual machines.

Red Hat Enterprise Linux and Xen: Value Proposition

The integration of Xen virtualization into Enterprise Linux offers the following customer benefits:

- » *Consolidation* – As described earlier, consolidation is one of the most fundamental benefits enabled by virtualization. For workloads that can be hosted on Enterprise Linux 5, paravirtualization enables computing resources to be assigned to workloads with far greater precision than before, so that systems can respond to changing business cycles by executing particular tasks at a performance level and cost that is appropriate for their priority. Customers with virtualization-enabled hardware will be able to consolidate multiple existing Enterprise Linux 4 systems into one new server. With the paravirtualized update to Enterprise Linux 4, these systems will also be able to run on older hardware (i.e., without VT or Pacifica).
- » *Flexible lifecycle management* – Virtualization enables users to migrate legacy applications to new hardware without disturbing their environment. For example, users can migrate an entire Enterprise Linux 4 operating environment, along with all of the applications that may have been running for some time, onto a Enterprise Linux 5-based hypervisor without having to requalify the software.
- » *Horizontal scaling* – Users can deploy scale-out infrastructures within a single larger server. For example, an application that does not scale well on an SMP system might be able to improve its throughput by replicating its processing across multiple servers. However, that course of action may not be possible on a single OS instance if the software was not designed to run multiple copies on the same machine (i.e., if it is using System V semaphores that have the same

name). In this case, multiple instances of the application could be hosted in different virtual machines on the same physical system.

- » *Instrumentation* – Although the Xen hypervisor is designed to have limited function and very low overhead, it can still be combined with instrumentation tools such as SystemTap and Frysk. This capability allows users to run a fully virtualized system, but perform testing, tuning, and monitoring of workloads within the hypervisor.
- » *Live migrations* – The ability to migrate a workload across the network by saving its virtual machine to a file is a key benefit of virtualization, and it will be implemented in the first release of Xen in Enterprise Linux 5. That will allow users to create a checkpoint, copy the checkpoint to another machine, restart it, and delete it from the first machine. Moreover, unlike with traditional HA clustering software, Xen can preserve network connections during a migration because the entire memory of the operating system is transported, keeping all of the TCP/IP socket data intact. The only requirement that Enterprise Linux 5 will have is a storage network topology that can support it, which could be some form of Storage Area Network (SAN), iSCSI, or Network File System (NFS). Otherwise, all of the necessary controls and capabilities needed to migrate both live and suspended virtual machines will be part of the base Enterprise Linux 5 platform.

Americas

Ideas International, Inc.
800 Westchester Avenue
Suite S620
Rye Brook, NY 10573-1330
USA
Tel + 1 914 937 4302
Fax +1 914 937 2485

Asia/Pacific and Worldwide Headquarters

Ideas International Limited
Level 3
20 George Street
Hornsby, NSW, 2077
Australia
Tel +61 2 9472 7777
Fax +61 2 9472 7788

Europe, Middle East, Africa

Ideas International Europe
1 Deanes Close
Steventon
Oxon OX13 6SZ
United Kingdom
Tel +44 (0) 1235 437 850
Fax +44 (0) 1235 437 851

www.ideasinternational.com



IDEAS Summary

Xen has emerged as the leading open source virtualization technology, and it has generated considerable interest across the industry, from both end users and systems vendors. As a key open source player, Red Hat will drive Xen deeply into the mainstream by making it a pervasive part of its Linux distribution. Red Hat's market presence will help to put Xen on the short list of virtualization platforms, along with VMware and Microsoft Virtual Server.

However, as basic resource management tools like virtual machines and hypervisors become commoditized, the real competition will shift to virtualization management solutions, including offerings from vendors such as BMC; VMware with VirtualCenter and VMotion; XenSource with XenOptimizer; IBM Tivoli; HP OpenView; and others. With few standards for virtualization management defined, some of these providers will try to establish leverage by introducing proprietary interfaces at critical management layers.

Thus, Red Hat's approach of focusing on the definition of open and documented virtualization interfaces, rather than competing with proprietary management tools, will provide a key added value for the industry. Red Hat's virtualization APIs and ABIs will create a level playing field for developers of virtualization management tools that are based on industry standards. Further, the approach will allow Red Hat Linux users to take advantage of virtualization with many of their existing tools, including scripts and provisioning frameworks.

The industry is at a critical phase in the adoption of virtualization, and the decisions that users make now could have a long-lasting impact. The use of open source software has been shown to minimize the risk of vendor dependency in many software domains. With its exclusive focus on open source, Red Hat is now poised to extend that benefit to virtualization.