The Adolescence of Virtualization

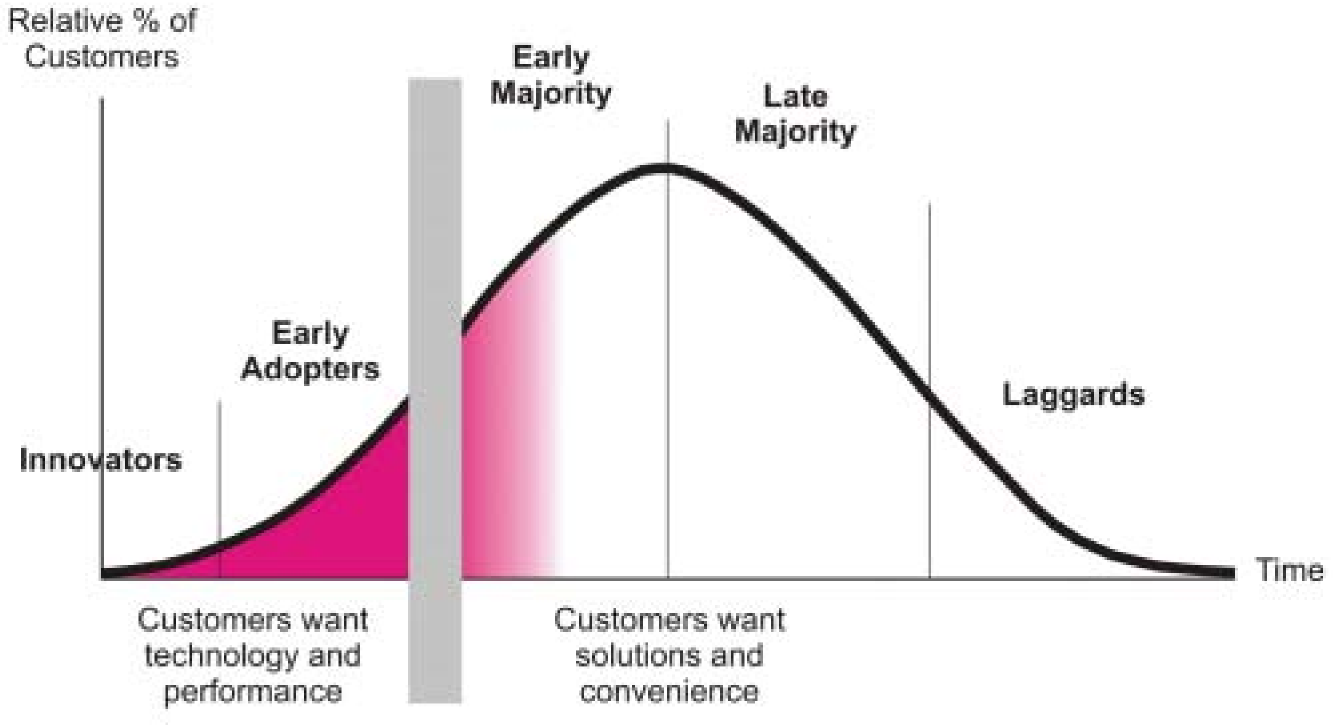
A status report on virtualization and Xen, with some observations on future challenges

June 1, 2006

Tom Christian
Principal Scientist, Virtualization Platforms
Hewlett-Packard Laboratories

Virtualization matures

Virtualization has entered the *early majority* phase of the technology adoption cycle, with customers looking for pragmatic solutions.



Xen has changed the playing field

- Xen is already considered a competitor, though it is still only available as a technology preview and XenSource has not yet shipped a product
- Three things have combined to force change:
  - The maturity and success of open source Linux
  - Xen’s existence as an open source virtualization technology
  - Industry interest in a standard virtualization platform

VMware and Microsoft have been forced to respond

So, if you are the “gorilla” in the marketplace, how do you respond to an open source threat?

Free Software!

Answer #1: Give stuff away

VMware Announces a Free Player

New Product Enables Anyone to Easily Run, Share or Evaluate Software in a Virtual Machine on a Windows or Linux PC

VMware, Palo Alto, California, October 24, 2005

VMware Introduces Free VMware Server

New Entry-level Virtualization Product to Accelerate Mainstream Adoption of Virtualization While Providing Path to Enterprise-class Virtual Infrastructure

VMware, Palo Alto, Calif., February 6, 2006

> *VMware President Diane Greene says her company is preparing for the low end of the software virtualization market to become commoditized, even free. Her product line, she says, increasingly stresses the sorts of higher-end software tools that allow companies to more easily manage their growing populations of virtual computers.*

Wall Street Journal, April 19, 2006

Answer #2: Give away $200,000

VMware Announces Virtual Appliance Challenge

Contest to Spur Developer Efforts to Innovate with Virtual Appliances

VMware, Palo Alto, California, February 27, 2006 VMware, Inc., the global leader in virtual infrastructure software for industry-standard systems, today announced the Ultimate Virtual Appliance Challenge contest with prizes totaling $200,000 to foster continued innovation in developing virtual appliances. Virtual appliances are pre-built, pre-configured and ready-to-run software applications, all packaged within virtual machines. They can be run using VMware virtualization products, including VMware Player and VMware Server which are both available for free download at www.vmware.com/download/.

Answer #3: Focus on standards

VMware announces Virtual Desktop Infrastructure Alliance

New Alliance Formed to Speed Adoption and Deployment of Virtual Desktop Infrastructure

VMware, Palo Alto, Calif., April 24, 2006

VMware-friendly change likely for Linux

Linux programmers are moving toward a change that would put virtualization software from VMware on a more even footing with open-source rival Xen

CNET News.com, April 13, 2006

VMware introduces Open Virtual Machine Disk Format specification

Specification Enables Broad-based Usage for Virtualization Industry

VMware, Palo Alto, Calif., April 3, 2006

VMware proposes VMI as standard paravirtualization interface for Linux

The interface proposed by VMware takes into account the potential hurdles to development and maintainability of the operating system that [paravirtualization] can cause.

VMware, Palo Alto, Calif., March 13, 2006

VMware to add support for paravirtualized Linux

Market Leader Gives Customers Broadest Choice of Supported Operating Systems

VMware, Palo Alto, Calif., August 10 2005

What about Microsoft?

Microsoft Announces New Price, and Availability of Linux Support, for Virtual Server 2005 R2

Virtual Server 2005 R2: Now Available as a Free Download

Microsoft, April 3, 2006

Microsoft Adapts Windows Server System Licensing to Virtualization Scenarios

Licenses for the Datacenter Edition of the version of Windows Server, code-named “Longhorn,” will give customers the right to run an unlimited number of virtual instances on one physical server.

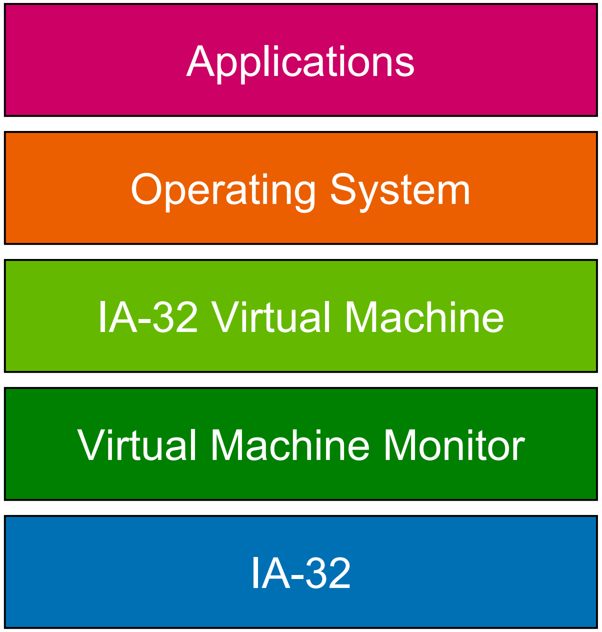
Microsoft, October 10, 2005

Microsoft is not part of the standards effort -- yet.

Classical Virtualization Definitions

Toward a common vocabulary…

VMs, VMMs and Hypervisors



A ***virtual machine*** (VM) provides a faithful implementation of a physical processor's hardware running in a protected and isolated environment.

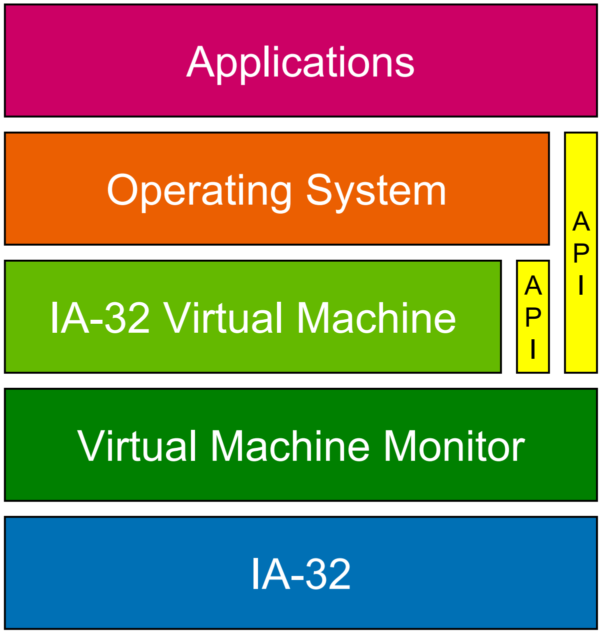
Virtual machines are created by a software layer called the ***virtual machine monitor*** (VMM) that runs as a privileged task on a physical processor.

A virtual machine may implement the instruction set for any processor. If it implements the instruction set for the physical processor on which it is running, it is called a ***hypervisor***.

Hypervisors are very efficient because they allow programs running in a virtual machine to execute instructions directly on the hardware processor whenever possible, interrupting the execution sequence only to emulate privileged operations.

All software running within a virtual machine runs *unprivileged*, but it may not be aware of that fact.

Paravirtualization



Paravirtualization is a technique in which the VMM is supplemented by an API that provides an assist for certain situations. The paravirtualization API is most often used by operating systems, but it may be used by applications for direct resource management.

The classical purpose of paravirtualization (circa 1974) was to improve performance by abstracting key functions at a higher level to enable them to be performed through a procedure call rather than as a sequence of privileged instructions.

Paravirtualization also allows you to avoid hard-to-virtualize processor instructions by replacing them with a procedure call that provides the functionality.

Paravirtualization simplifies tuning and provides a significant TTM advantage for VMM development. It can also be provided as an option to a pure VM.

The cost of paravirtualization is that the operating system has to be modified to take advantage of it.

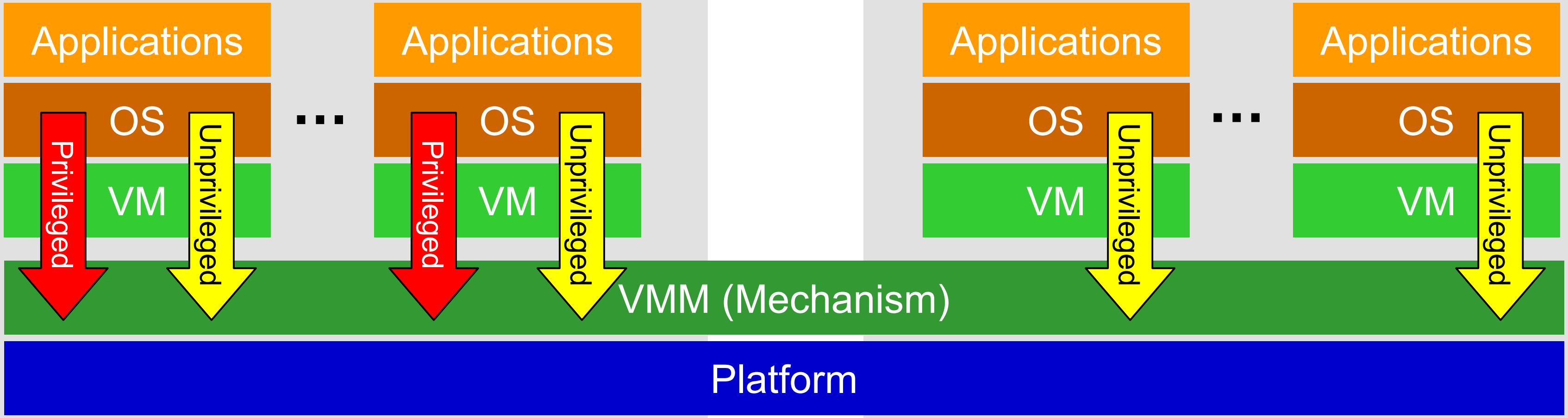
Virtual Server Architecture

Distinguished Domains (Policy)

- Run as unprivileged guests
- Have access to privileged VMM API
  - *Global resource management*
  - *System management*
  - *Performance and QoS monitoring*
  - *Network, storage, I/O throttling*
- Own and manage physical resources
- Network, storage and I/O virtualization
- *Probably* do not migrate

Guest Domains

- Run as unprivileged guests
- Have access only to unprivileged APIs
  - *Local resource management*
  - *Paravirtualization*
- Virtual servers are guest domains
- May migrate to another physical server



Hardware support for virtualization

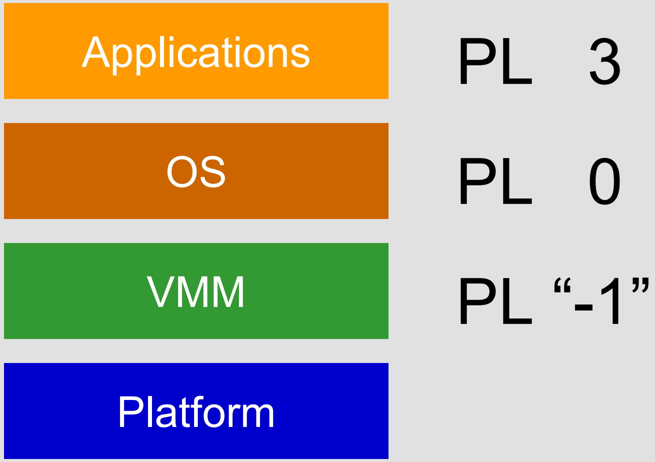
Intel delivers VT (Vanderpool)

AMD delivers SVM (Pacifica)

Hardware Support for Virtualization

Intel and AMD virtualization decoded

- Intel: VT (*Vanderpool)*
  - VTx (IA32/EM64T), VTi (IA-64)
  - Hardware support for virtualization
- Intel: LT (*LaGrande)*
  - Secure start-up and I/O security
- AMD: SVM (*Pacifica)*
  - Hardware support for virtualization
  - Secure start-up and I/O security
- From 10,000 feet, Pacifica ≈ VT + LT



Applications PL 3

OS PL 0

VMM PL “-1”

Platform

Key Features of Hardware Virtualization

- New operating modes:
  - Intel: VMX Root, VMX non-root
  - AMD: Host Mode, Guest Mode
- Effectively implements a privilege level “-1”
  - Privilege levels 0-4 available to guests
  - VMM runs in root/host mode at PL “-1”
  - Guests run in non-root/guest mode at PL 0-3
  - Privileged instructions *always trap* in guest mode
  - Problematic instructions also trap in guest mode
  - Guests cannot control virtual machine operation
- Secure start-up: Validated software configuration
  - Trusted Platform Module provides root of trust
  - Sequential module validation provides chain of trust
- Secure I/O
  - Constrain I/O to domain boundaries
  - Hardware support required for (efficient) secure DMA
  - Guests can have direct access to physical devices

Are VT and SVM magical?

Hardware Virtualization FAQ

Q: What is the principal purpose of VT and SVM?

A: The principal purpose is to correct architectural defects that make IA–32 and IA–64 difficult to virtualize. In addition, AMD and Intel have added features to enhance performance and security.

Q: Assuming mature technology solutions, will processor support for virtualization enable unmodified guests to run as fast as paravirtualized guests?

A: No. Paravirtualization offers the opportunity to abstract privileged services at a higher level than individual privileged instructions, resulting in fewer context switches and better performance.

Q: How will VT and SVM benefit VMware?

A: VMware uses binary translation to modify guest operating systems to make it possible to run them in a virtual environment. VT and SVM may improve performance and/or reliability and eliminate the need for some of these translations.

Elements of a Legacy

- *Disaggregation*
- *Standardization*
- *Trust*

If Xen gets these three things right,
it will drive virtualization for the industry.

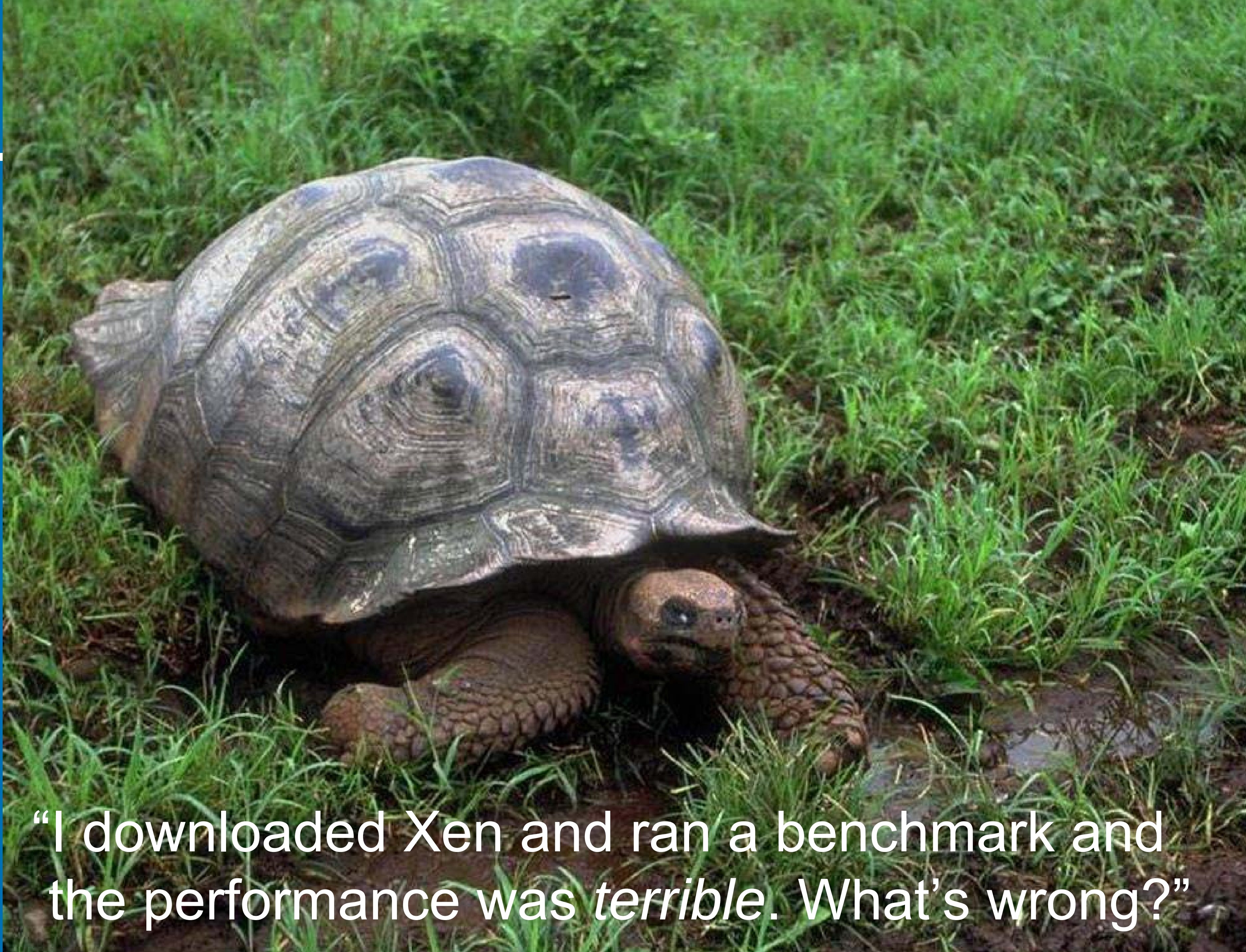
Xen continues to gain support…

…but there have been a few tactical errors.

Disaggregation

Decomposition into the right component parts

- The VMM should be open source and managed by a neutral person or organization
  - Eliminates a potential business control point
  - Enables and encourages broad support
  - Prevents manipulating the technology roadmap to favor one company
- The VMM should not be part of any OS
  - OS integration creates a business control point
  - May have licensing or distribution implications
  - Legal trend is to mandate component substitutability
    - E.g., Windows Media Player in Europe



"I downloaded Xen and ran a benchmark and the performance was *terrible*. What's wrong?"

There is not just one Xen

Some Xen versions perform better than others:

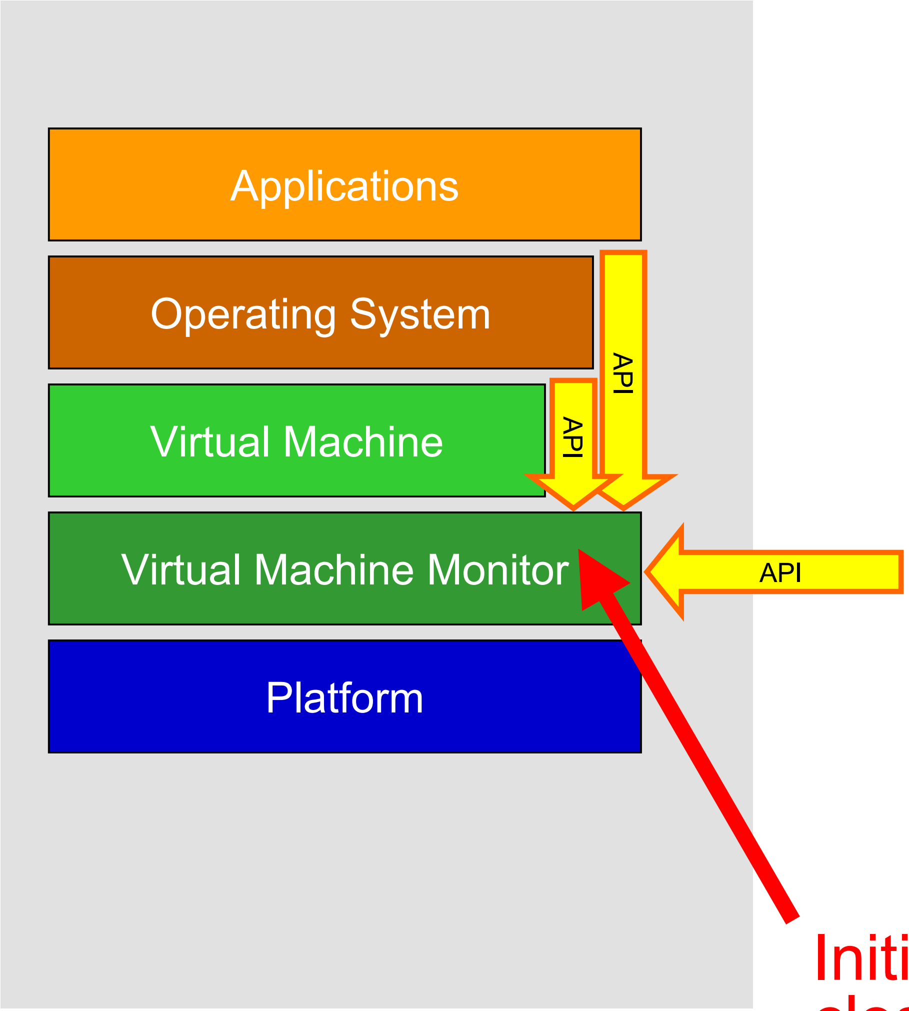
- XenSource has retained some key Xen technology that improves performance
- This technology will be offered *as a product* by XenSource
- Some of this technology may be made available to the open source code base in the future
- XenSource's principal objective is to deliver a Windows virtualization product and compete head-to-head with VMware

XenSource has to become profitable, but differentiating at the platform level puts standardization at risk

Standardization

- The VMM is becoming *infrastructure technology* with more value as a standard than as a medium for delivering innovation.
  - Much work – including research – remains to be done on virtualization, but it's best done as part of an effort to create a standard platform.
- Future value-added solutions will (and *should*) compete principally in the policy layers above the VMM.

Standard APIs



Control and Management

- Create / destroy virtual servers
- Allocate / de-allocate resources
- Virtual server migration
- Application resource management
- Image management / update

Security

- Create and manage trust relationships

Performance

- Resource utilization
- QoS monitoring
- Dynamic resource management

Paravirtualization

- Mitigate virtualizability problems
- Optional performance tuning

Initial focus is on OS/VMM API, closely followed by management

OS/VMM API Standardization

Xen may not define the standard VMM API for Linux

- Xen’s APIs have been evolving
  - The functionality has been changing as Xen matures
- VMware has been participating in Xen API development
  - VMware’s objective is to define a standard VMM interface for Linux
- VMware submitted its VMI API to Linux on March 13, 2006
  - Xen missed an opportunity to set the standard
    - “VMware programmers suggested a documented, stable interface…” – Andrew Morton
    - “I've heard precious little from the Xen team on the topic…” – Andrew Morton
  - Xen submitted a competing API three days later
- Andrew Morton prefers a neutral interface that works with any VMM
  - “I'd say [the API is] a long way from making it into mainline.“

A neutral interface is the *best outcome* for customers

Trust

Virtualization will never be accepted by customers unless there is implicit trust in the solution.

Virtual servers must provide the same level of trust as a well-managed physical server, including secure boot and establishment of a chain of trust to the application level.

Trust – HP Involvement

- HP is actively involved in developing technologies to secure the Next Generation Data Center
  - Chairmanship of Trusted Computing Group (TCG) Technical, Server Specific and Hard Copy Work Groups
  - HP's Mark Schiller is the current President and CEO of TCG
  - Focus on extending the TCG concepts of Chain of Trust and Root of Trust to broader infrastructure components, including virtualization
- HP leads several EU collaborative research projects:
  - Open Trusted Computing
  - Trust and Security for Next Generation Grids
  - Global Network for Secure Communications based on Quantum Cryptography

Secure Platform Architecture

Using Itanium’s security features to create a secure OS

> ***“Unless security is designed into the system from the bottom up, we’re constantly going to be fighting a holding action.”***
>
> – *Bruce Schneier*
> *“Secrets & Lies: Digital Security in a Networked World”*

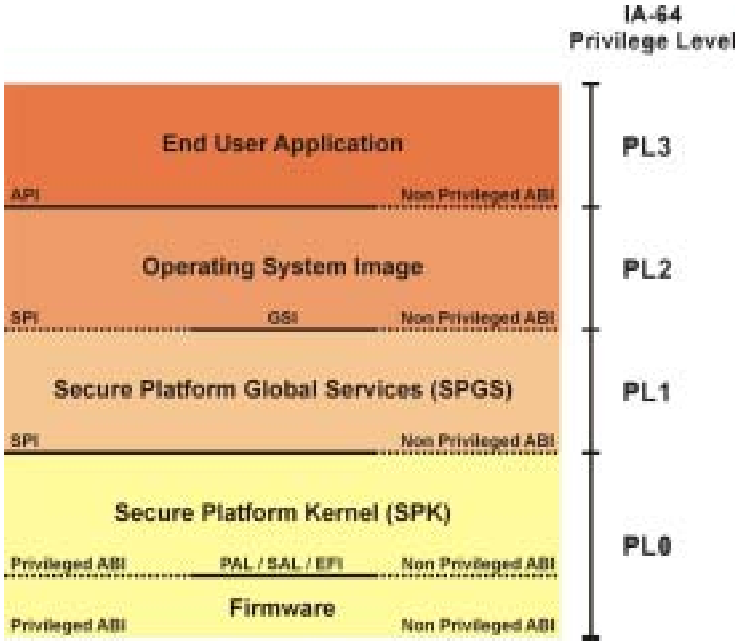
SPA Architecture – April 2001

Ring 0: The SPK abstracts privileged instructions, physical resources and interruptions, replacing them with an API to request and manage privileged objects and services. The SPK provides privileged *mechanisms* and is the only SPA layer that runs in privileged mode.

Ring 1: The SPGS layer supports multiple secure OS partitions, each with its own OS image, and implements resource management *policy* for SPA.

Ring 2: Operating System images run *as unprivileged tasks* under SPA.

Ring 3: Application Programs



SPA uses the designed-in security features of Itanium to provide a platform for trustworthy systems. It has been prototyped on Linux, showing less than 2% impact on performance while maintaining application compatibility.

Virtualization Futures

A glimpse at HP’s research plans

What's next?

- Many technologies will emerge or gain importance as virtualization solutions mature:
  - Delivery of pre-installed applications
  - Simplification of the system test matrix
  - Low-cost high-availability solutions
  - Patch and upgrade management
  - Service automation and thin clients
  - Performance and security isolation
  - User-based performance management

We will explore one:

Scalable, automated system management

Future System Attributes

- Resource allocation and utilization will shift from physical servers to virtual servers.
- Virtual server reliability and security will meet or exceed the levels possible with physical servers.
- Eventually, servers will be dissociated from any particular physical existence and be viewed as a fungible resource allocated as needed to perform a specific task.
- Operating system and application vendors will target virtual servers as supported platforms.

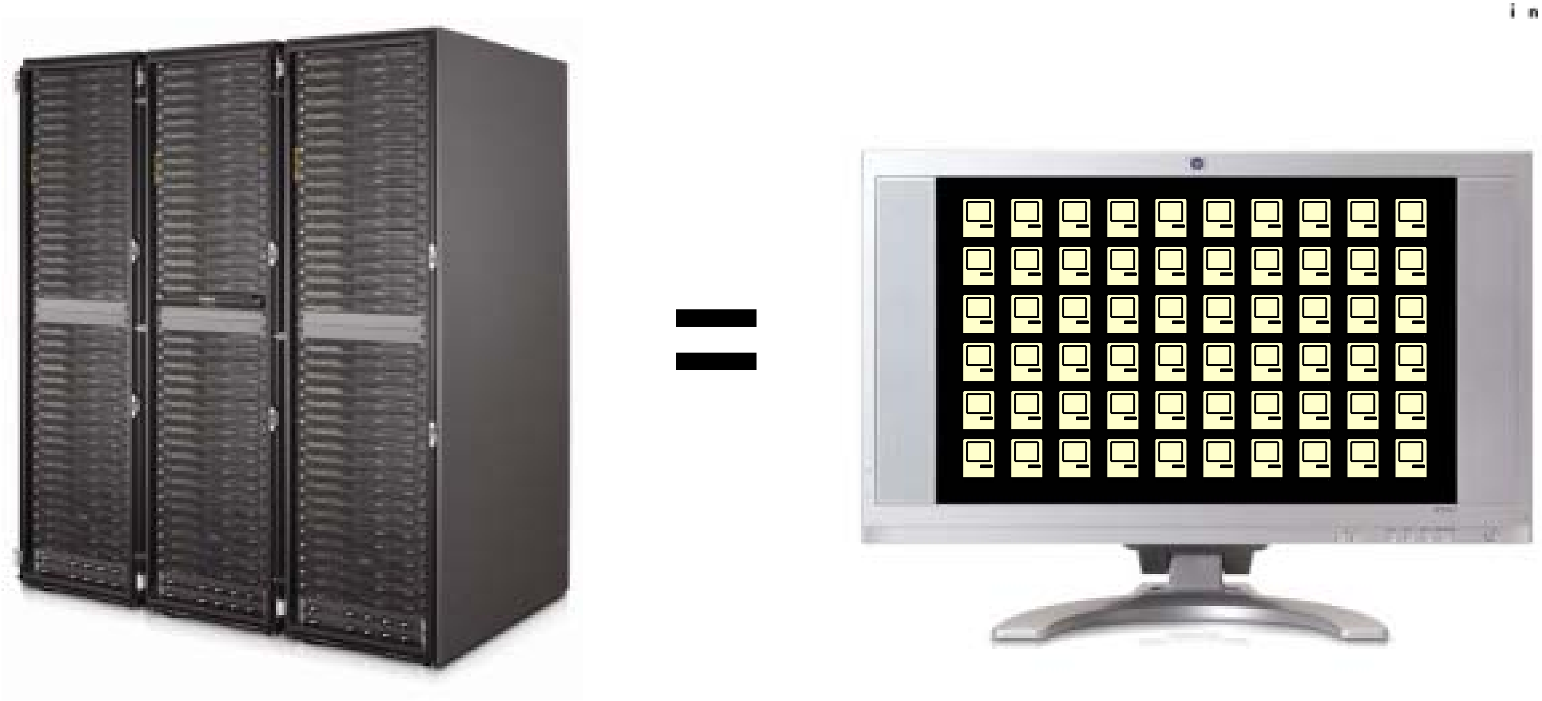
Ubiquitous Virtualization

The virtualization layer will:

- Become a standard layer on servers and workstations, eventually becoming part of the firmware
- Provide full virtualization that will run any operating system without modification and optional paravirtualization for added performance
- Support different operating systems in separate virtual machines on a single physical server
- Export a standard API for integration of management and control applications for heterogeneous data centers

System Management Goals

- A principal objective of HP's virtualization strategy is the significant reduction of system management costs.
- The ultimate goal is zero human management overhead in a steady-state environment.
- To achieve this goal we must take work out of the system, not just move the work somewhere else or transform its nature.

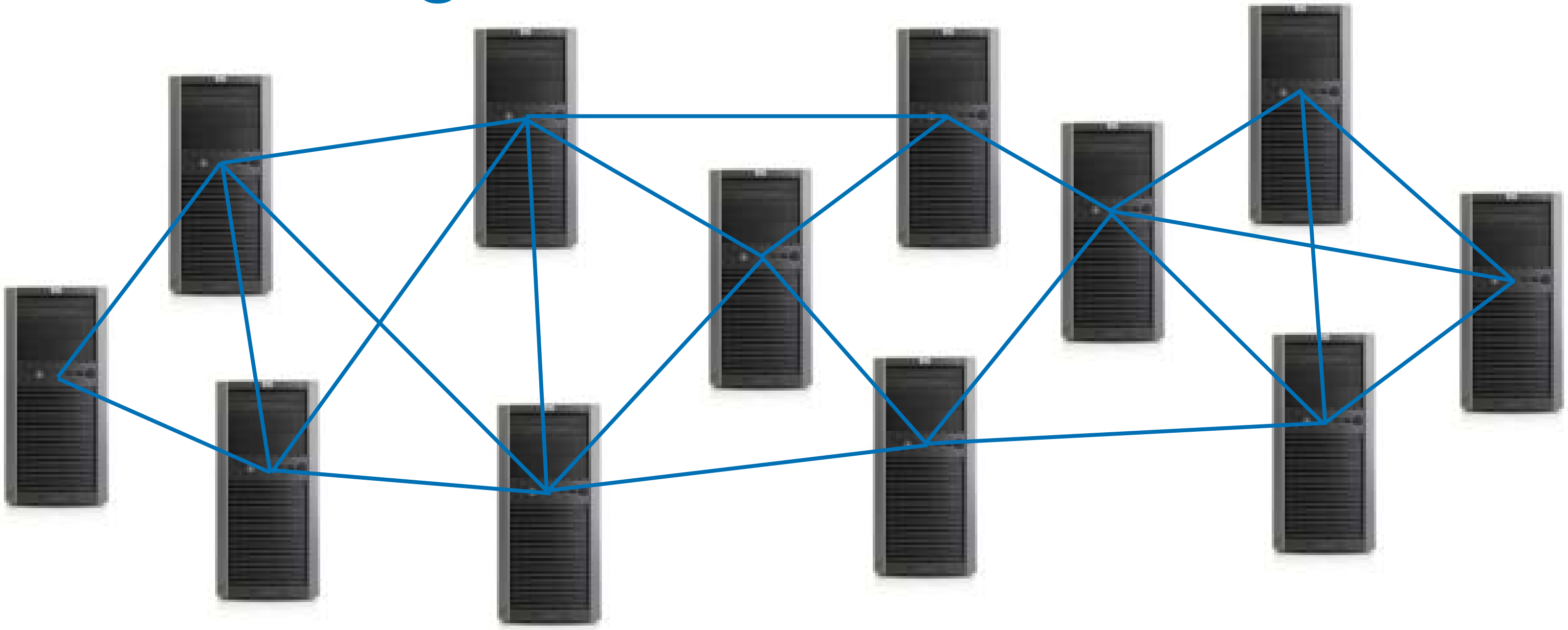


Managing individual systems remotely through a management application *moves* and *transforms* the work, but it does not eliminate a significant quantity of work.

Scalable System Management

- System and resource management for physical servers will be handled by management software running in privileged partitions.
- System management will be automated and distributed with no dedicated management servers and no single points of failure.
- Each system will participate in the management process through local, low-overhead applications that have minimal impact on system performance.
- Complexity scales linearly with the number of systems.

Self-management Research



Each system within a group is given limited knowledge and simple goals:

- Information about system and environmental conditions
- Knowledge about several “neighbors” within the collection
- A goal to pursue optimum utilization of local resources
- Rules for guest domain migration when local resources can’t meet needs

Balance and control become an *emergent behavior* invent