Virtualization

Santa Clara Valley Chapter of the IEEE Communication Society
June 20, 2007
Scott Lurndal, 3Leaf Systems
Agenda

- Virtualization Overview
- Application Virtualization
- Operating System Virtualization
- Machine Virtualization
- I/O Virtualization
- 3Leaf Systems I/O Virtualization
Terminology

- Virtual Machine Monitor / Hypervisor
- Pure Virtualization
- Paravirtualization
- Emulation
- Host Operating System
- Guest Operating System
Virtualization History

- 1960’s  MIT’s CTSS
- 1970’s  IBM VM/370
- 1990’s  VMware
- 2000’s  Microsoft Virtual Server, Xen, KVM, TopSpin, PClexpress
- 2007   3Leaf Systems V-8000
- 2008/2009 Microsoft Viridian Hypervisor
Provides isolated, portable environment for application execution.

- Java Virtual Machine
  - Supports Java language
  - JVM executes ‘bytecode’ not native code.
- Microsoft Common Language Runtime
  - Supports C#, Managed C++, Python, .Net
- Microsoft Softricity Softgrid
  - SystemGuard layer isolates application from host Operating System.
  - Wraps and deploys existing Microsoft Windows applications

Simplifies deployment

- Java: Write once, run anywhere
Traditional Unix chroot(2) environment
   – Limits application access to filesystem subtree

BSD Jail
   – Adds process and network isolation to filesystem isolation

Solaris Containers
   – Creates virtual ‘systems’ where each container appears externally as a system with its own identity

Linux OpenVZ project
   – Similar to Solaris containers, adds the ability to create virtual private servers with individual identities to Linux

All containers/VPS’s execute the same operating system code
**VM/370**

- Allows multiple operating systems to simultaneously execute on System/370.
- Some hardware support for virtualization

**VMware Server (i386, x86_64)**

- Hosted virtualization
- Software platform emulation with mostly native code execution
  - Runs guest OS in non-privileged processor ring
  - Intercepts and emulates privileged operations
  - Emulates hardware devices
  - Rewrites some guest code dynamically
- Upcoming versions will leverage new processor virtualization features
  - Intel VT-X
  - AMD AMD-V
**Xen (Pre 3.0)**

- Paravirtualized; operating system leverages hypervisor API for certain performance intensive kernel operations
  - Page table updates
  - Input/Output
- Special I/O drivers are installed in guest
  - Front-end/Back-end architecture using ring buffers
- Runs special Linux guests (xenU architecture), doesn’t run Windows.
- 64-bit performance poor (due to processor architecture issues)

**Xen 3.0+**

- Leverages processor virtualization for better performance
- Paravirtual operations are optional for the guest
- Supports Windows and unmodified Linux guests
Microsoft Virtual Server
- Hosted Virtualization
- Runs Windows and Linux guests

Microsoft Viridian (Scheduled 2H2008)
- Hypervisor based virtualization
- Uses privileged domain to support guest I/O operations

Linux KVM
- New in-tree kernel virtual machine implementation leveraging processor virtualization support.
- Hosted implementation
- Supports Linux, Windows and other ia32 guests

QEMU, Bochs, AMD SimNow!
- Software platform emulations, including instruction set.
Input/Output Virtualization

- PCI Express Virtualization
  - I/O Subsystem Hardware Virtualization Support

- File Virtualization
  - Virtualization of Network Attached Storage

- Storage Virtualization
  - Virtualization of Storage Area Network
    - Fiberchannel
    - iSCSI with iSER RDMA extensions

- Network Virtualization
  - Virtual Local Area Networks (vlan)

- 3Leaf Systems Virtualization
PCI Express Virtualization

**Single Root Virtualization**
- Allows a single PCI express device to appear as multiple independent devices.
- Each device can be assigned directly to a Guest
- IOMMU ensures protection between guests

**Multi-root Virtualization**
- Allows multiple hosts to share a single-root enabled PCI express device
- Uses PCI express cabling and PCI express switch or blade backplane.

**IOMMU**
- A platform resident device to manage physical address translations between PCI devices and system memory controllers.
File Virtualization
- Virtualizes access to storage devices at the file level
- NAS Front-ends (CIFS, NFS) provide a global namespace
- Cluster Filesystems (CXFS(sgi), GPFS(ibm), OPFS(oracle))
- Acopia Networks, Attune Networks, IBRIX, Quantum’s StorNext

Storage Area Network Virtualization
- Hardware Raid subsystems
  - Virtualize access to physical disk. Raid controller manages physical storage.
- In-band Storage Virtualization
  - Maps virtual LUN’s to storage allocated from a pool of physical LUN’s.
  - Allows dynamic creation and resizing of LUNs
- Host-based Storage Virtualization
  - Solaris ZFS, Veritas Storage Foundation, Linux Device Mapper
3Leaf Systems Virtual Compute Environment

– Diskless Uni- or Multi-processor Compute Nodes + Infiniband HCA
  • 3Leaf driver provides virtual Host Bus Adapter for access to storage
  • 3Leaf driver provides virtual Network Interface Card for access to network
  • Supports RHEL 4, RHEL 5, SLES 10 and Windows operating systems.
  • Transports block requests and network packets to V-8000 I/O server
  • Devices can be added to and deleted from Compute Nodes dynamically

– V-8000 I/O Server
  • Infiniband HCA
  • 8 PCI express slots (2) 8x, (6) 4x with mix of GigE and 4Gbs Fiberchannel
  • Management Software
  • Serves up to 32 Compute Nodes
  • Multiple V-8000 provide redundant access to I/O resources
Introducing: Virtual Compute Environment (VCE)
3Leaf Company Overview

Company Facts
- Founded in 2004, headquartered in Santa Clara, CA.
- Approximately 80 employees worldwide.

Financial Backing
- Over $32M in funding raised to date.
- VC: Enterprise Partners, Alloy, Storm Ventures.
- Corporate: Intel Capital.

Market Focus
- X86 Server Market, Linux/Windows.
- Enterprise Datacenters.
- OEMs: Server Vendors.
- Partners: AMD, IBM, Storage Providers, Vmware.

Team Management
- Experience and proven track record in large system design, Operating Systems, ccNUMA, Storage and Network technologies.
- Priors: SGI, Unisys, IBM, Brocade, AMD, Transmeta, Verisign.
Conclusion

Thank You
Corporate Colors

– Use the default colors in the top row of the color table as much as possible for consistency.

- R: 122  G: 193  B: 66
- R: 0     G: 152  B: 74
- R: 203   G: 219  B: 42
- R: 255   G: 242  B: 0
- R: 255   G: 192  B: 14

– Press Ctrl-g to bring up the guides, and use them to align separate elements for a more organized appearance.

– To reset a page to the default, go to menu items Format > Slide Layout > then choose the style and click the Apply (or Reapply) button. If it doesn’t seem to correct things, try it one more time.