L14

Understanding the Technology Advantages of Running Linux on z/VM

Reed A. Mullen
mullenra@us.ibm.com
IBM Systems and Technology Group

IBM System z Expo
September 17-21, 2007
San Antonio, TX

IBM System z Expo – San Antonio, Texas

Session L14
Understanding the Technology Advantages of Running Linux on z/VM

Reed A. Mullen
mullenra@us.ibm.com
IBM Systems and Technology Group

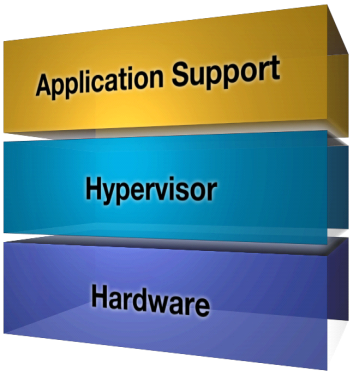
Trademarks

System z Virtualization: a Multidimensional Solution

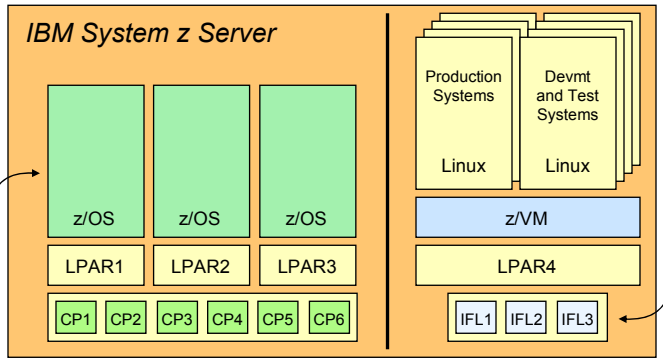
Virtualization Support is Built In, Not Added On

With coordinated investments in the virtualization technology stack

- **Application support layer**
  - Open, reliable operating system
  - Virtual server awareness infrastructure
  - Enterprise applications
- **Hypervisor layer (z/VM)**
  - Shared-memory based virtualization model
  - Highly granular resource sharing and simulation
  - Flexible virtual networking
  - Resource control and accounting
  - Server operation continuity (failover)
  - Server maintenance tools and utilities
- **Hardware layer**
  - Legendary reliability, scalability, availability, security
  - Logical partitioning (LPAR)
  - Processor and peripheral sharing
  - Interpartition communication
  - Virtualization support at the hardware instruction level



Sample z/VM IFL Configuration



z/VM and most Linux software fees are priced on real engine capacity...

IFL engines have no impact on z/OS license fees

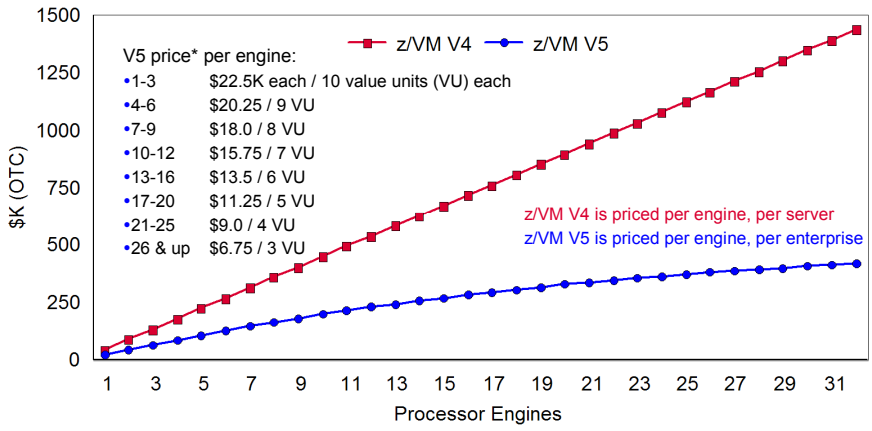
3-engine z/VM V5 license charges*

Year 1:	$84,390	OTC plus S&S
Year 2:	$16,890	S&S only
Year 3:	$16,890	S&S only
3-Year Total:	$118,170	

...another source of cost savings attributed to z/VM's ability to over-commit CPU capacity

*U.S. prices as of 1 Aug 2007

z/VM Version 5 Pricing



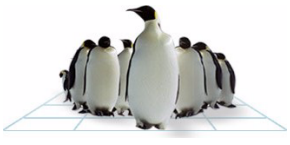
*U.S. prices as of 1 Aug 2007

Why Run Linux on z/VM?

- **Infrastructure Simplification**
  - Consolidate distributed, discrete servers and their networks
  - IBM mainframe qualities of service
  - Exploit built-in z/VM systems management
- **Speed to Market**
  - Deploy servers, networks, and solutions fast
  - React quickly to challenges and opportunities
  - Allocate server capacity when needed
- **Technology Exploitation**
  - Linux with z/VM offers more function than Linux alone
  - Linux exploits unique z/VM technology features
  - Build innovative on demand solutions

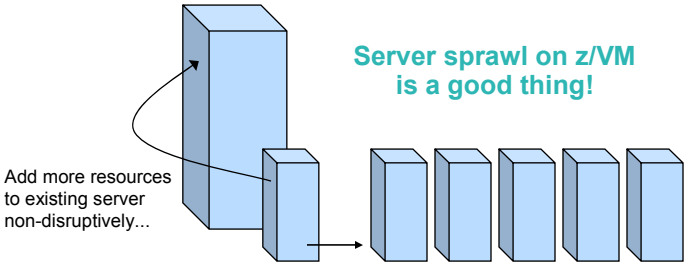
z/VM Technology Exploitation for Linux

- **Resource sharing and scalability**
- **CPU and memory**
- **Advanced disk support**
- **Virtual communications and network consolidation**
- **Systems management, provisioning, command and control**



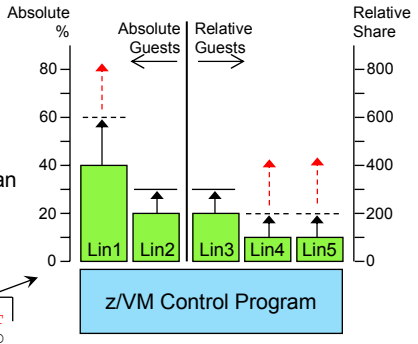
Resource Sharing and Scalability
Scale Up and Out with Linux on z/VM

- **With z/VM you can grow horizontally and vertically on the same System z server...dynamically**
- **Provision a virtual machine for peak utilization and allocate its resources to other servers during off-peak hours... automatically**



z/VM CPU Resource Controls
Highly Granular Sharing of System Resources

- Allocate system resources per guest image using SHARE command
  - This is a highly flexible and self-managed function of the z/VM Control Program
  - Reserve CPU capacity for peak usage
    - Use it when needed
    - Relinquish the processor cycles for other servers when not needed
  - "Absolute guests" receive top priority
  - The Virtual Machine Resource Manager can be used to monitor and adjust remaining capacity allocated to "Relative guests"



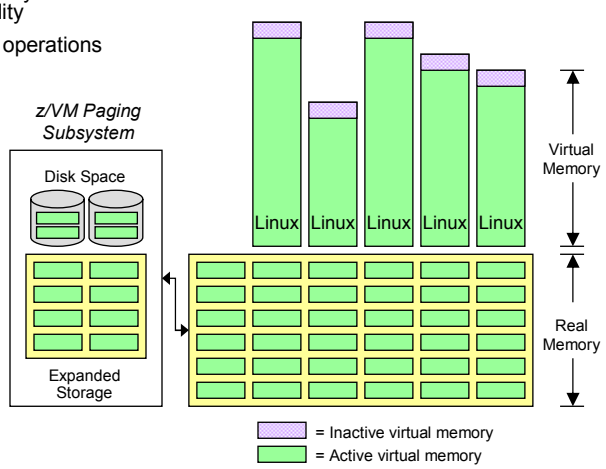
z/VM Directory Entries (or "on-the-fly" commands)

```
SHARE Lin1 ABSOLUTE 40% ABSOLUTE 60% LIMITSOFT
SHARE Lin2 ABSOLUTE 20% ABSOLUTE 30% LIMITHARD
SHARE Lin3 RELATIVE 200 RELATIVE 300 LIMITHARD
SHARE Lin4 RELATIVE 100 RELATIVE 200 LIMITSOFT
SHARE Lin5 RELATIVE 100 RELATIVE 200 LIMITSOFT
```

Notes:
----- = limit can be exceeded if unused capacity is available (LIMITSOFT)
——— = limit will not be exceeded (LIMITHARD)

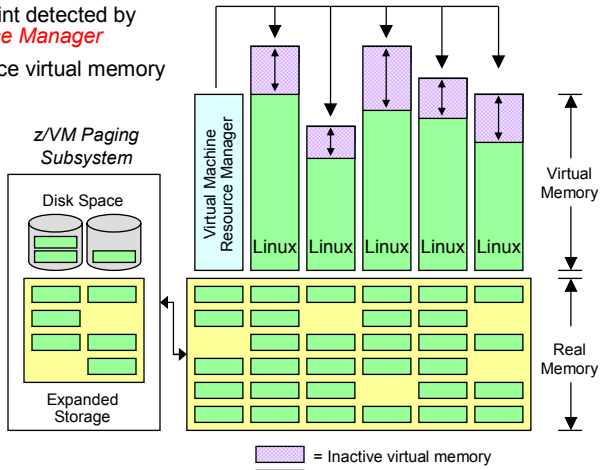
Linux and z/VM Technology Exploitation
Cooperative Memory Management (CMM)

- Problem scenario: virtual memory utilization far exceeds real memory availability
- z/VM Control Program paging operations become excessive
- Overall system performance and guest throughput suffers



Linux and z/VM Technology Exploitation
Cooperative Memory Management (CMM)

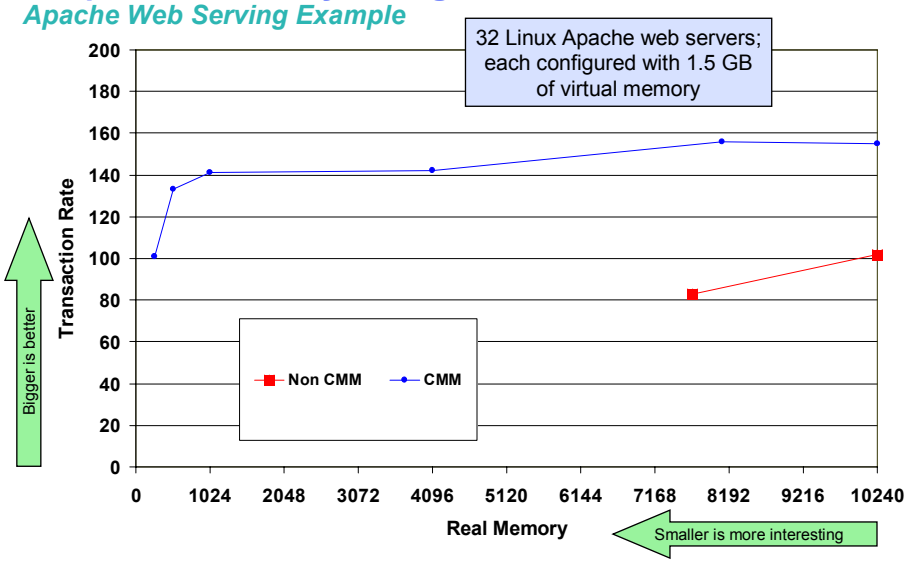
- Solution: real memory constraint detected by z/VM *Virtual Machine Resource Manager*
- Linux images signaled to reduce virtual memory consumption
- Linux memory pages are released
- Demand on real memory and z/VM paging subsystem is reduced
- Helps improve overall system performance and guest image throughput



Learn more at:
ibm.com/servers/eserver/zseries/zvm/sysman/vmrm/vmrmcmm.html

Cooperative Memory Management with Linux on z/VM

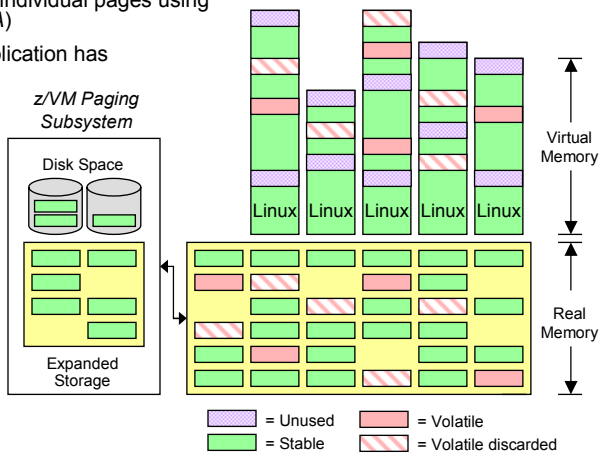
Apache Web Serving Example



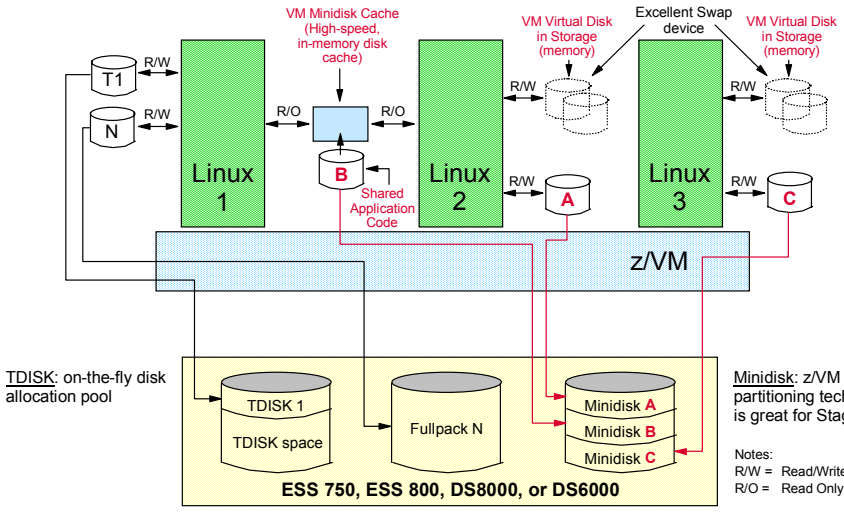
Linux and z/VM Technology Exploitation

Collaborative Memory Management Assist (CMMA)

- Extends coordination of memory and paging between Linux and z/VM to the level of individual pages using a new hardware assist (*CMMA*)
- z/VM knows when a Linux application has released a page of memory
- Host Page-Management Assist (*HPMA*), in conjunction with CMMA, further reduces z/VM processing needed to resolve page faults
- Can help z/VM host more virtual servers in the same amount of memory
- Supported by System z9 and z/VM V5.3
- IBM is working with its Linux distribution partners for exploitation support



z/VM Technology: Advanced Disk Support

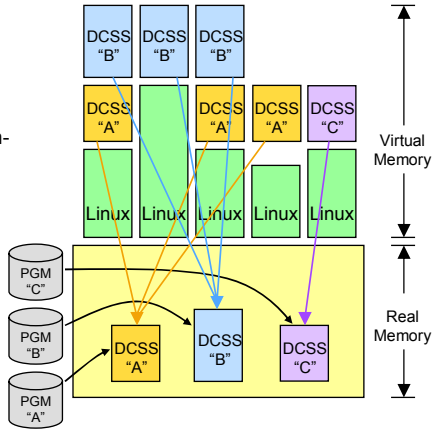


Linux and z/VM Technology Exploitation

Linux Exploitation of z/VM Discontiguous Saved Segments (DCSS)

- DCSS support is Data-in-Memory technology
  - Share a single, real memory location among multiple virtual machines
  - High-performance data access
  - Can reduce real memory utilization
- Linux exploitation: shared program executables
  - Program executables are stored in an execute-in-place file system, then loaded into a DCSS
  - DCSS memory locations can reside outside the defined virtual machine configuration
  - Access to file system is at memory speeds; executables are invoked directly out of the file system (no data movement required)
  - Avoids duplication of virtual memory and data stored on disks
  - Helps enhance overall system performance and scalability

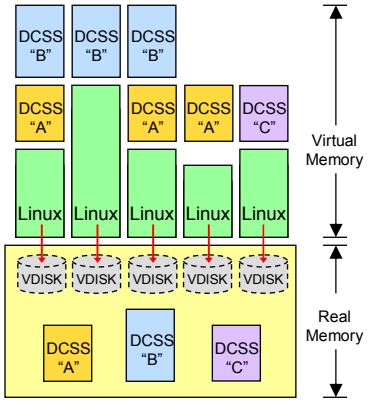
Learn more:
"Using DCSS/XIP with Oracle 10g on Linux for System z"
www.redbooks.ibm.com/redpieces/abstracts/sg247285.html



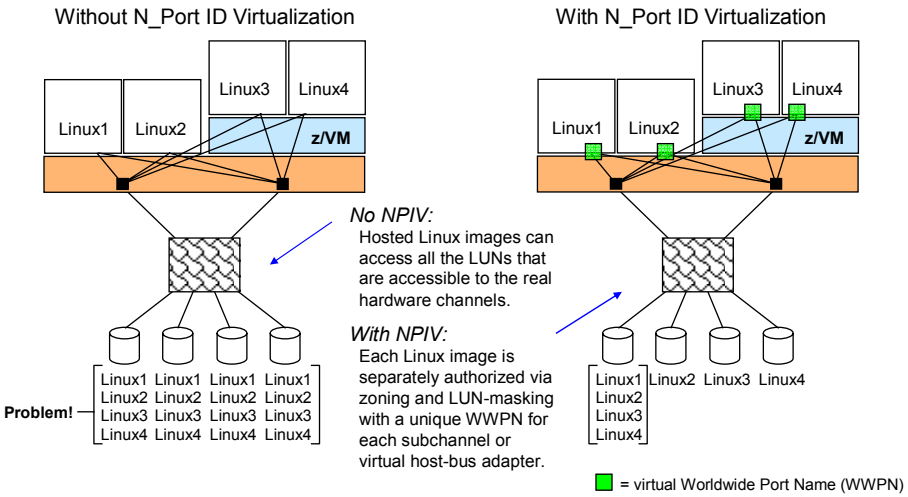
Linux and z/VM Technology Exploitation

Linux Exploitation of z/VM Virtual Disks in Storage (VDISK)

- VDISK support is Data-in-Memory technology
  - Simulate a disk device using real memory
  - Achieve memory speeds on disk I/O operations
  - VDISKs can be shared among virtual machines
- Linux exploitation: high-speed swap device
  - Use VDISKs for Linux swap devices instead of real disk volumes
  - Reduces demand on I/O subsystem
  - Helps reduce the performance penalty normally associated with swapping operations
  - An excellent configuration tool that helps clients minimize the memory footprint required for virtual Linux servers
  - Helps improve the efficiency of sharing real resources among virtual machines



System z and N_Port ID Virtualization (NPIV)



z/VM Support for NPIV

- **FICON Express features on System z9 support FCP N_Port ID Virtualization**
- **NPIV enables zoning and LUN masking on a virtual machine basis**
- **Multiple operating system images can now concurrently access the same or different SAN-attached devices (LUNs) via a single, shared FCP channel**
  - Can increase channel utilization
  - Less hardware required
  - Helps reduce the complexity of physical I/O connectivity
- **Supported by z/VM V5.3 and V5.2; z/VM V5.1 support is available with the PTF for APAR VM63744**
  - Note: z/VM V5.1 cannot be installed from DVD to SCSI disks when NPIV is enabled

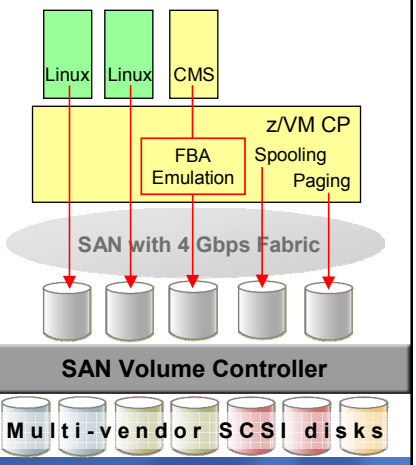
z/VM Support for Parallel Access Volumes

- **PAVs allow:**
  - Multiple concurrent I/Os to the same volume by one or more users or jobs
  - Automatic coordinated Read and Write I/O referential integrity when needed
- **z/VM V5.2 with PTF for APAR VM63952:**
  - Supports PAVs as minidisks for guest operating systems that exploit the PAV architecture (e.g., z/OS and Linux for System z)
  - Provides the potential benefit of PAVs for I/O issued to minidisks owned or shared by guests that do not support native exploitation of PAVs, such as z/VSE, z/TPF, CMS, or GCS
- **IBM System Storage DASD volumes must be defined to z/VM as:**
  - 3390 Model 2, 3, or 9 on a 3990 Model 3 or 6 Controller
  - Or…2105, 2107, or 1750 Storage Controller
  - Note: 3380 track-compatibility mode for the 3390 Model 2 or 3 is also supported.
- **Potential benefit:**
  - Designed to improve I/O response times by reducing device queuing delays

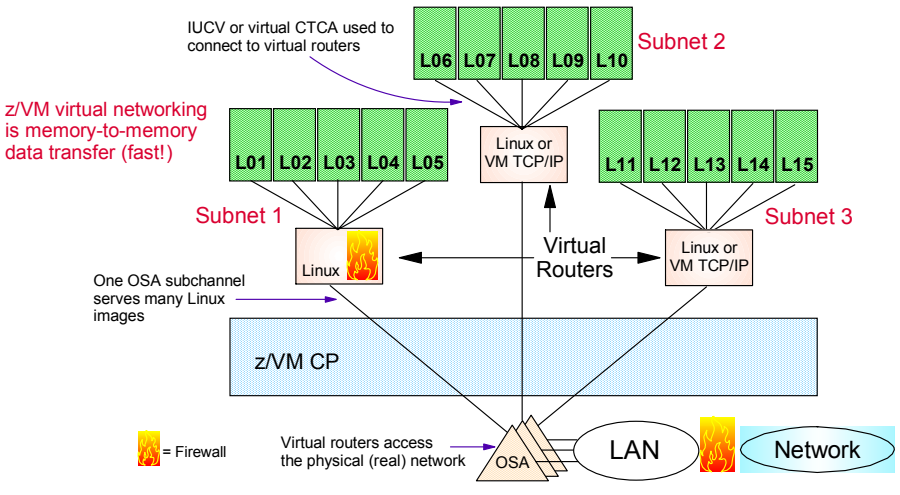
IBM System Storage SAN Volume Controller Software V4.2

- **z/VM and Linux for System z support SAN Volume Controller (SVC) V4.2**
- **SVC allows z/VM and Linux to access SCSI storage from multiple vendors as a single pool of disk capacity**
- **z/VM FBA emulation allows CMS users to access SVC-managed disk space**
- **New function in SVC V4.2:**
  - Multi-target FlashCopy support (up to 16 images)
  - Higher number of active FlashCopy relationships at the cluster level
  - Designed for improved cluster performance, especially when installed on IBM System Storage SVC 2145-8G4 storage engine
  - Support for additional OEM devices
- **Supported in z/VM V5.3 base product**
  - z/VM V5.2 support available with PTF for APAR VM64128

Learn more at: **ibm.com**/storage/support/2145

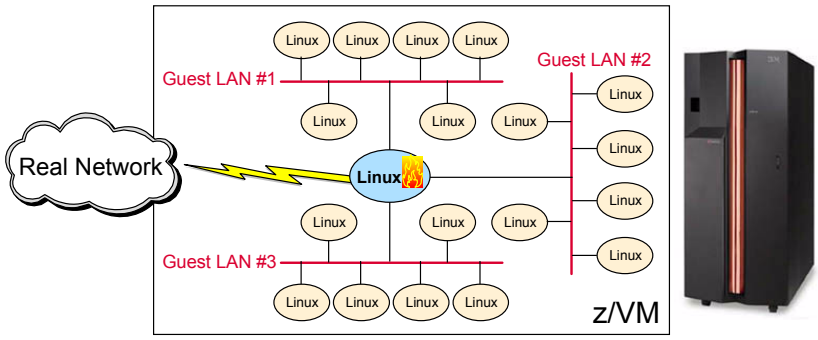


z/VM Virtual Networking
Point-to-Point Communications

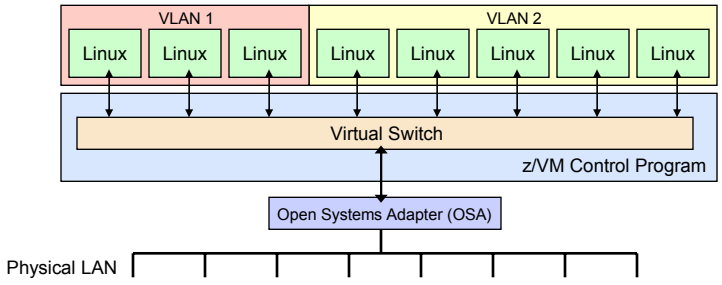


z/VM Virtual Networking
z/VM Guest LANs

- A Guest LAN is a "virtual" LAN created by the z/VM Control Program
- OSA Express (QDIO) and HiperSockets Guest LANs can be created
  - Point-to-point, Multicast, and Broadcast (QDIO) connections are supported
- Linux images can connect to one or more Guest LANs
  - And connect to real network adapters at the same time
  - This enables a Linux image to provide external routing and firewall services for other Linux images

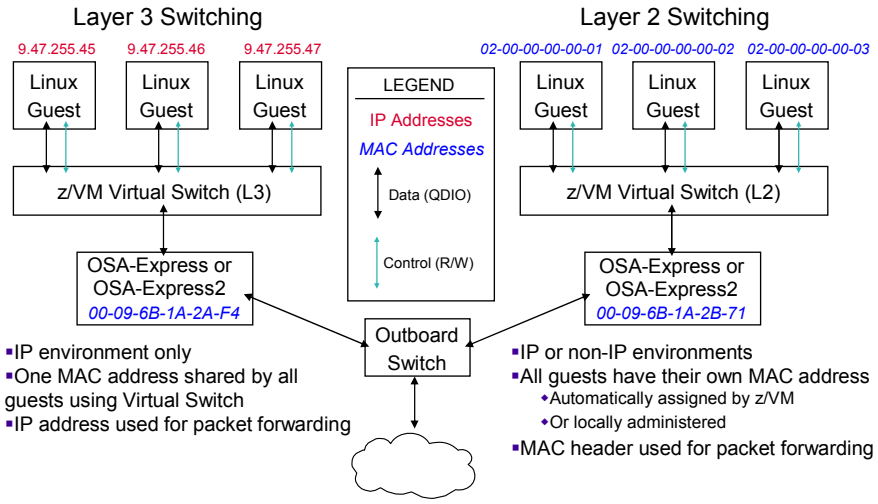


z/VM Virtual Networking
Using the z/VM Virtual Switch



- **Eliminates need for router to connect virtual servers to physical LAN segments**
  - May reduce overhead associated with router virtual machines
  - Allows virtual machines to be in the same subnet with the physical LAN segment
- **Supports Layer 2 (MAC) and Layer 3 (IP) switching**
  - Includes support for IEEE VLAN
  - Provides centralized network configuration and control
  - Easily grant and revoke access to the real network
  - Dynamic changes to VLAN topology can be made transparent to virtual servers

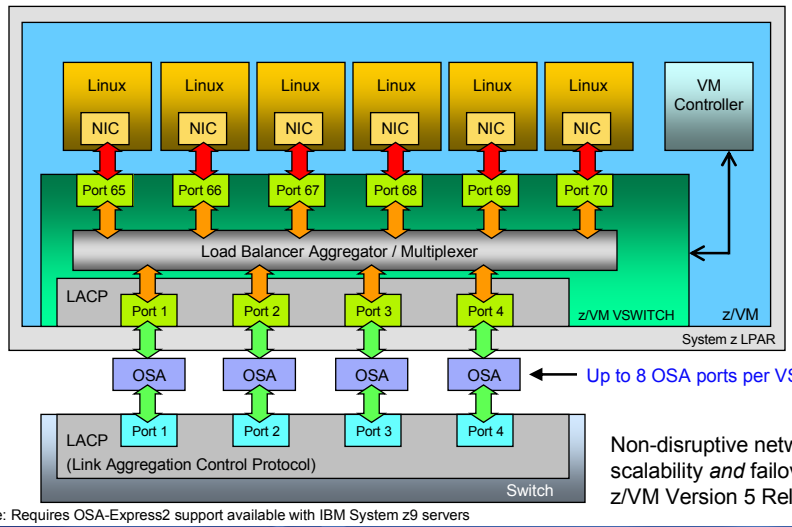
z/VM Virtual Switch Support
Layer 3 Compared to Layer 2 Switching



- IP environment only
- One MAC address shared by all guests using Virtual Switch
- IP address used for packet forwarding

- IP or non-IP environments
- All guests have their own MAC address
  - Automatically assigned by z/VM
  - Or locally administered
- MAC header used for packet forwarding

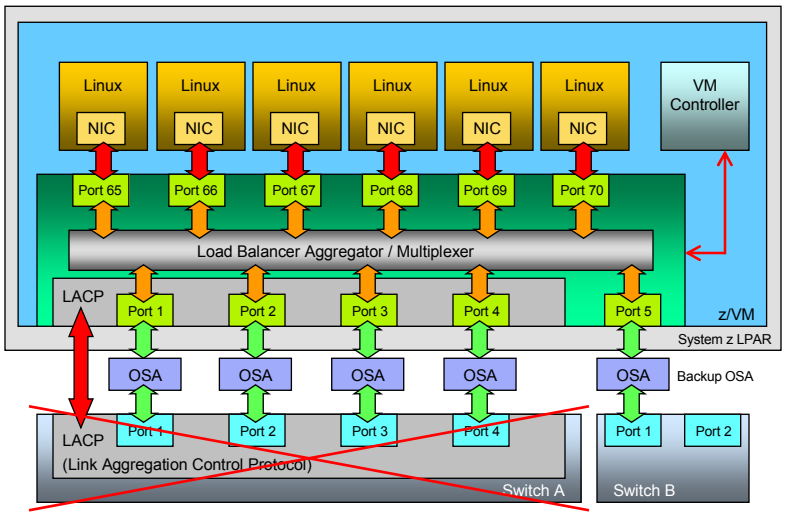
z/VM Virtual Switch Link Aggregation Support
Enhanced Networking Bandwidth and Business Continuance



Non-disruptive networking scalability *and* failover with z/VM Version 5 Release 3

Note: Requires OSA-Express2 support available with IBM System z9 servers

Recovery of a Failed Switch

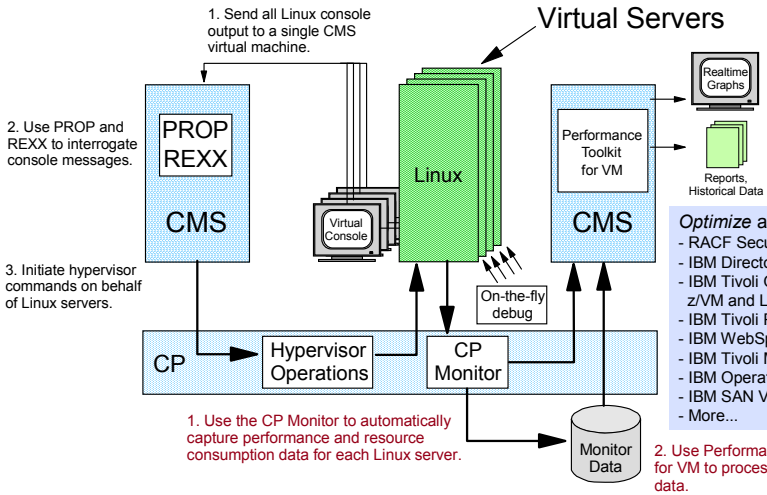


z/VM Command and Control Infrastructure

- **Built-in z/VM facilities enable cost-effective command and control**
  - Performance data collection and reporting for every Linux image
  - Log accounting records for charge-back
  - Automate system operations with CMS, REXX, Pipelines, virtual console interrogation using PROP (VM programmable operator)
  - Dynamic I/O reconfiguration (e.g., dynamically add more disks)
  - Run EREP on z/VM for system-level hardware error reporting
  - Priced z/VM V5.3 features:
    - DirMaint – simplifies task of adding/modifying/deleting users
    - Performance Toolkit for VM – performance recording and reporting
    - RACF Security Server for z/VM – security services (including LDAP)
    - RSCS – provides NJE connectivity support for Linux systems
- **Samples, examples, downloads available**
  - IBM Redbooks
  - z/VM web site (www.vm.ibm.com/download)
- **Extensive suite of solutions available from ISVs**
  - Visit ibm.com/systems/z/os/linux/apps/all.html

z/VM Technology – Command and Control Infrastructure

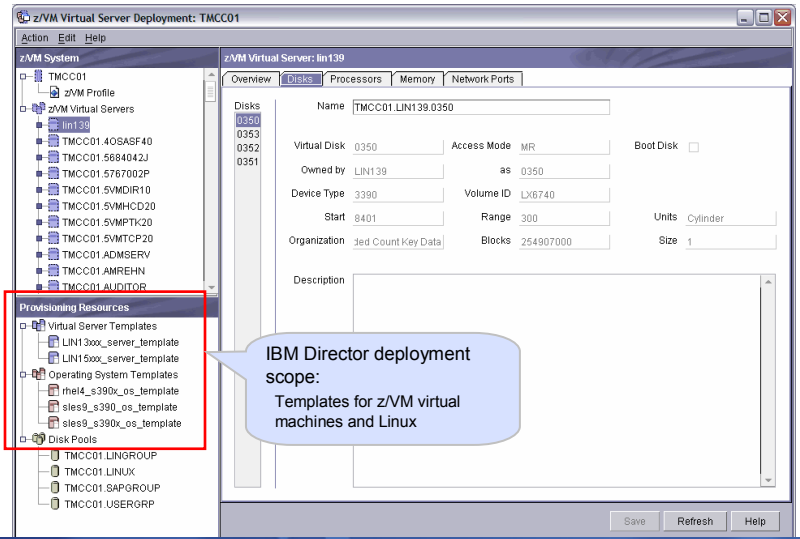
Leveraging the IBM Software Portfolio



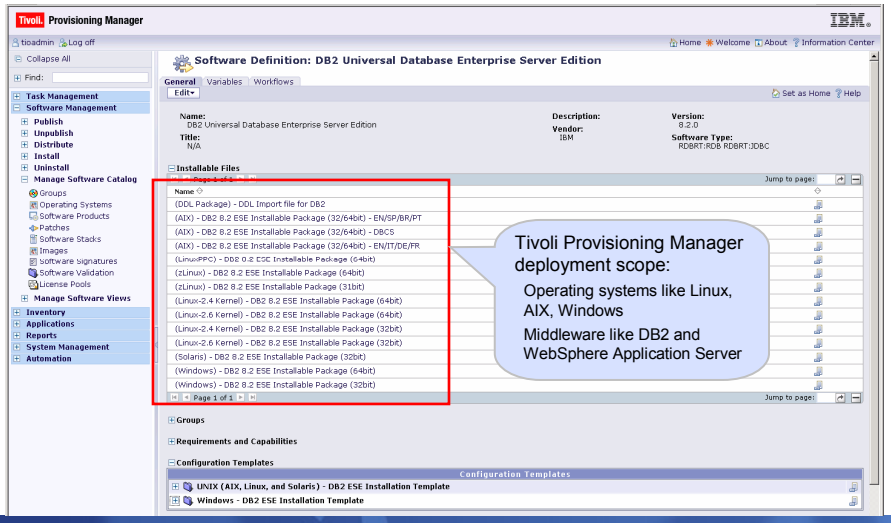
Optimize and *Integrate* with:
- RACF Security Server for z/VM
- IBM Director (z/VM Center)
- IBM Tivoli OMEGAMON XE for z/VM and Linux
- IBM Tivoli Provisioning Manager
- IBM WebSphere solutions
- IBM Tivoli Monitoring
- IBM Operations Manager for z/VM
- IBM SAN Volume Controller
- More...

Provisioning Linux Virtual Machines on System z

Using IBM Director for Linux on System z with z/VM Center



Provisioning Software in System z Virtual Linux Servers
Using IBM Tivoli Provisioning Manager



Monitoring System z Virtual Linux Servers
Using IBM Tivoli OMEGAMON XE for z/VM and Linux

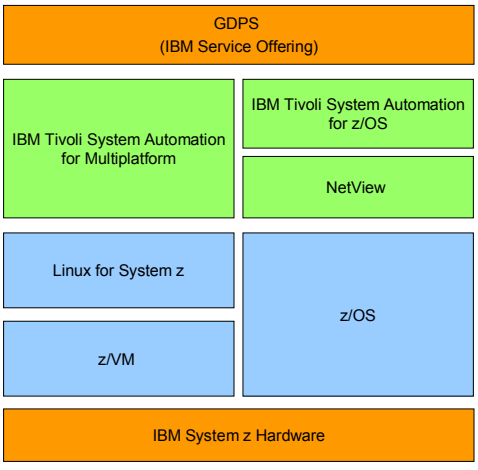
- **Combined product offering that monitors z/VM and Linux for System z**
- **Provides work spaces that display:**
  - Overall system health
  - Workload metrics for logged-in users
  - Individual device metrics
  - LPAR Data
- **Provides composite views of Linux running on z/VM**

z/VM Systems Management Products from IBM

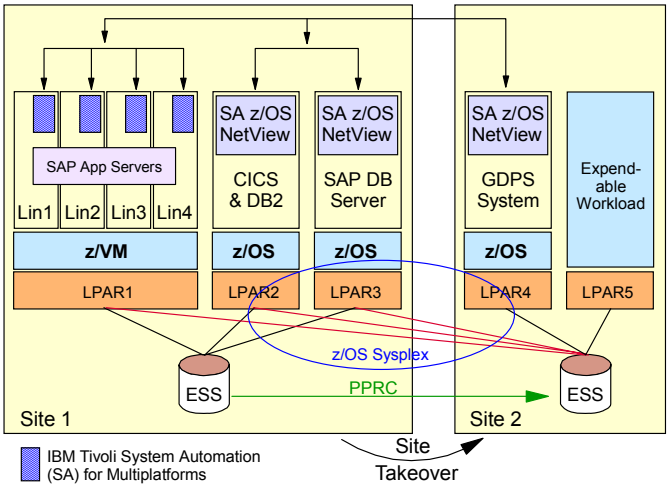
- **IBM Backup and Restore Manager for z/VM V1.2**
  - Provides z/VM system administrators and operators the ability to efficiently and effectively backup and restore files and data on z/VM systems
  - Can also backup and restore images of non-z/VM guest systems such as Linux
- **IBM Tape Manager for z/VM V1.2**
  - Manages and monitors tape resources, helping increase data availability and improve operator efficiency
  - Automates common daily tape operations and helps eliminate tedious, often error-prone, manual tasks
- **IBM Archive Manager for z/VM V1.1**
  - Addresses storage and data management concerns by allowing users to archive historical or other infrequently used data to increase data availability
  - Helps companies comply with data storage requirements mandated by fiscal or legal regulations and policies
- **IBM Operations Manager for z/VM V1.2**
  - Helps improve the monitoring and management of z/VM virtual machines by automating routine maintenance tasks
  - Enables users to automatically respond to predictable situations that require intervention

GDPS/PPRC Multiplatform Resiliency for System z

- Business Resiliency Solution for z/OS and Linux on z/VM
- Based on IBM Geographically Dispersed Parallel Sysplex (GDPS) Service Offering
- Leverages existing proven solutions
  - GDPS
  - Tivoli System Automation for z/OS
  - Tivoli System Automation for Multiplatforms
- Provides coordinated cross platform business resiliency for operating systems running on System z hardware
- Integration point for z/OS and Linux for System z



GDPS/PPRC Multiplatform Resiliency for System z



- Designed for customers with distributed applications
- SAP application server running on Linux for System z
- SAP DB server running on z/OS
- Coordinated near-continuous availability and DR solution for z/OS, Linux guests, and z/VM
- Uses z/VM HyperSwap function to switch to secondary disks
- Sysplex support allows for site recovery

z/VM Virtualization Leadership Support

- **High levels of RAS built into the hardware**
- **Non-disruptive On/Off Capacity on Demand capability**
- **Linux and z/OS application integration**
- **Highly granular allocation of hardware assets**
  - Add "small" server images to existing configuration with minimal impact to other server images expected
- **Large-scale server hosting**
  - Potentially hundreds of server images
- **Resource consumption recording / reporting**
  - Capture data at hypervisor level (CP Monitor)
  - Useful for charge-back, capacity planning, problem determination, and fix verification
- **Hot stand-by without the hardware expense**
  - Idle backup images ready to run (or be booted) if primary servers fail
- **Autonomic, non-disruptive disk failover to secondary storage subsystem capability**
- **Architecture simulation**
  - Help satisfy configuration requirements without necessarily suffering expense of real hardware
- **In-memory application sharing**
  - Share program executables among multiple server images
- **Server-memory-cached disk I/O**
  - High-speed read access to files on disk
- **Virtual Disks in Storage**
  - High-speed read and write access to files in memory (excellent swap devices for Linux)
- **Built-in console message routing**
  - Route messages from all virtual servers to a single virtual machine (system automation)
- **Virtual Machine Resource Manager**
- **"Hands free" auto-logon of server images**
  - Using z/VM "Autolog" support
- **Initiate operating system shutdown from "outside" the server image**
  - Without requiring agent running on guest operating system
- **Up to 256 Linux servers can share a single System z cryptographic card using z/VM**
- **Clone, patch, and "go live" with easy rollback**

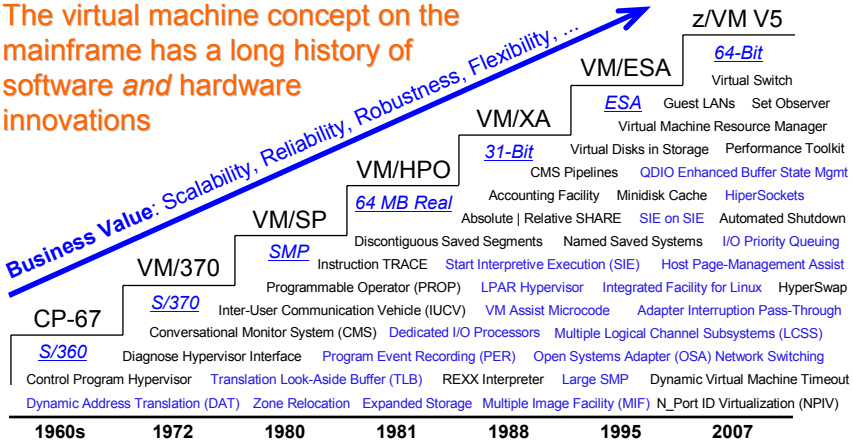
Linux and z/VM Resources

- **IBM Learning Services Classes**
  - Linux Implementation for System z (ZL100)
  - Installing, Configuring, and Servicing z/VM for Linux Guests (ZV062)
  - Linux Basics – A System z perspective (ZL120)
  - Advanced Solutions for Linux on System z (ZL150)
  - Deploying WebSphere and Advanced e-business Applications on Linux for System z (LINX7)
  - Deploying WebSphere Centric Products on Linux for System z (LINX6)
  - DB2 Universal Database Administration Workshop for Linux (CF201)
  - z/VM and Linux Connectivity and Management (ZV100)
  - z/VM RACF and DirMaint Implementation (ZV200)
  - Start Your Engines – IBM Virtualization Engine – for Enterprise Release 1 (VE011)
  - Find more info at: ibm.com/servers/eserver/zseries/os/linux/ed.html
- **Linux for System z and S/390 listserver**
  - www.marist.edu/htbin/wlvindex?linux-390
- **IBM z/VM web site**
  - www.vm.ibm.com
- **Linux for System z support documentation (including Redbooks)**
  - www.ibm.com/systems/z/os/linux/support_documentation.html

Reference Material

IBM System z Virtualization Genetics

The virtual machine concept on the mainframe has a long history of software *and* hardware innovations



System z virtualization starts on the chip; an integration of hardware, firmware, and software functionality

Reference Material

System z Interpretive Execution
Advanced Technology for Virtual Server Hosting

- **Start Interpretive Execution (SIE) instruction**
  - Operand is a state descriptor for an LPAR or virtual machine
  - Accommodates fixed-storage and pageable guests
  - Interception controls allow hypervisor intervention
  - Reduces context switch time
- **System z implements two levels of SIE**
  - No performance penalty for running z/VM in an LPAR
  - No shadow page tables required for DAT-on guests
  - Considerable architectural and hardware investment required
    - Potential instruction behavioral differences at each level
    - Multiple control register sets

Reference Material

Additional Mainframe Virtualization Facilities

- **Zone Relocation**
  - SIE capability that provides multiple zero-origin storage regions (logical partitions) on one system
  - Enables I/O subsystem to access partition memory directly, without requiring hypervisor intervention
- **Translation Lookaside Buffers (TLBs)**
  - Large allocation of microprocessor space for TLBs directly benefits virtual server scalability
  - z9 and z990 provides a TLB arrangement which advantageously uses two buffers
    - Second-level TLB feeds address translation information to the first-level TLB when the desired virtual address is not contained in the first-level TLB
- **Multiple Image Facility (MIF)**
  - Enables channel sharing among multiple LPARs
  - I/O devices on shared channel paths can be accessed simultaneously by sharing LPARs (or restricted to a subset of sharing LPARs)
- **Logical Channel Subsystem (LCSS) support**
  - Allows a z9 and z990 to be configured with up to 1024 channels (512 channels for z890)
  - 256 channels can be configured for each LPAR, with selected channel sharing among LPARs possible

Reference Material

Additional Mainframe Virtualization Facilities

- **I/O Priority Queuing**
  - Allows high-priority workloads to receive preferential access to I/O subsystem
  - Supported by Intelligent Resource Director and virtualized by z/VM
- **HiperSockets**
  - High-speed, security-rich TCP/IP connectivity among LPARs
  - Memory speed communications
- **Adapter Interruption Pass-Through**
  - OSA-Express (Ethernet) and FCP (SCSI) virtual machine I/O can be performed while z/VM guest image is running in SIE mode
  - "Thin" interrupt passed to z/VM Control Program when I/O operation belongs to an idle guest system
- **QDIO Enhanced Buffer-State Management (QEBSM)**
  - Two new machine instructions designed to help eliminate overhead of hypervisor interception
- **Host Page-Management Assist (HPMA)**
  - Interface to z/VM paging and storage management
  - Designed to allow hardware to assign, lock, and unlock page frames without hypervisor assistance
- **Layer 2 (MAC) and Layer 3 (IP) network switching**
  - OSA and z/VM support enables virtual IP and MAC network switching without requiring a hosting partition

Thank you

For more information, please contact

Reed A. Mullen
mullenra@us.ibm.com
+1 607 429 3824

ibm.com/systems/z