IBM GLOBAL SERVICES

G51

zSeries Logical Partitioning and Virtualization - Keeping the “z” in “Virtualization”

Romney White

zSeries Expo | Nov. 1 - 5, 2004

Miami, FL

[redacted]

RETURN TO INDEX

zSeries Logical Partitioning and Virtualization

Keeping the “z” in “Virtualization”

zSeries Technical Conference
November 1-5, 2004

Romney White
zSeries Virtualization

Agenda

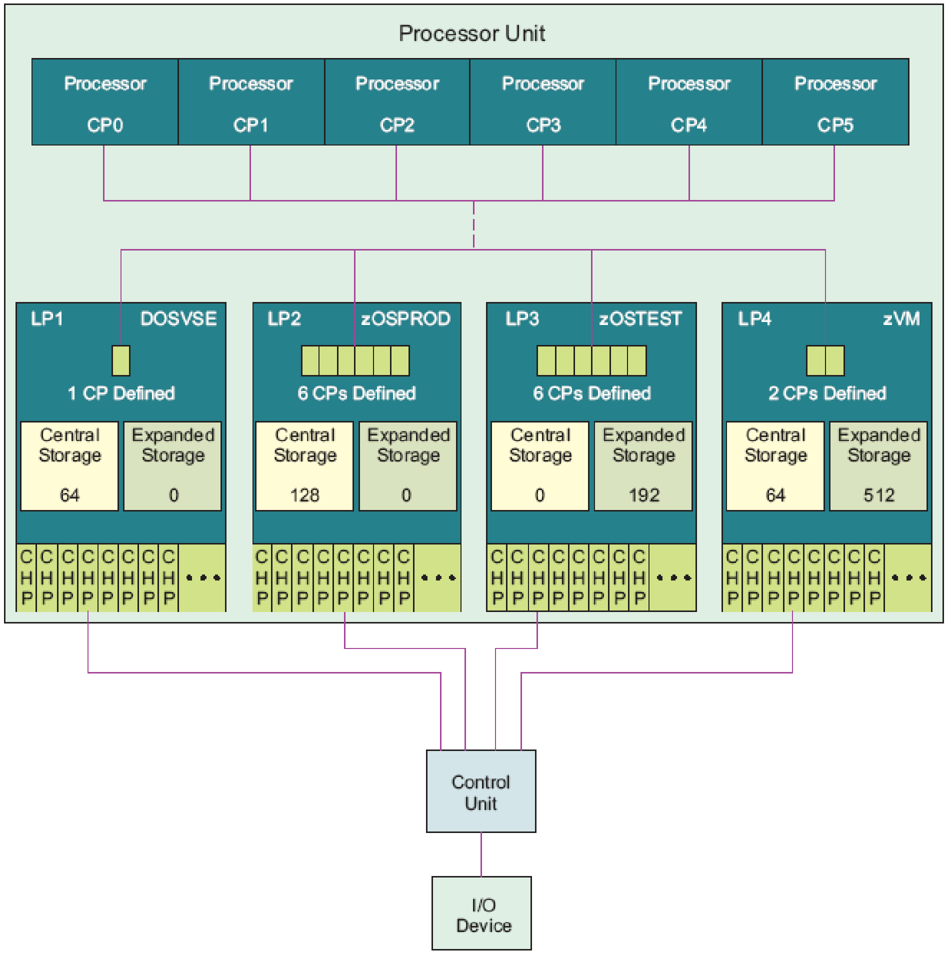
- Definitions
- Hypervisor Technologies
- zSeries Virtualization Evolution
- zSeries Virtualization Status
- Future Directions

System/360 Model 40: 1964

IBM eServer zSeries 890 (z890): 2004

Definitions

- **Partitioning**
  - **Server partitioning is the logical or physical division of a single server's resources into independent, isolated systems that can run independent operating systems and software**



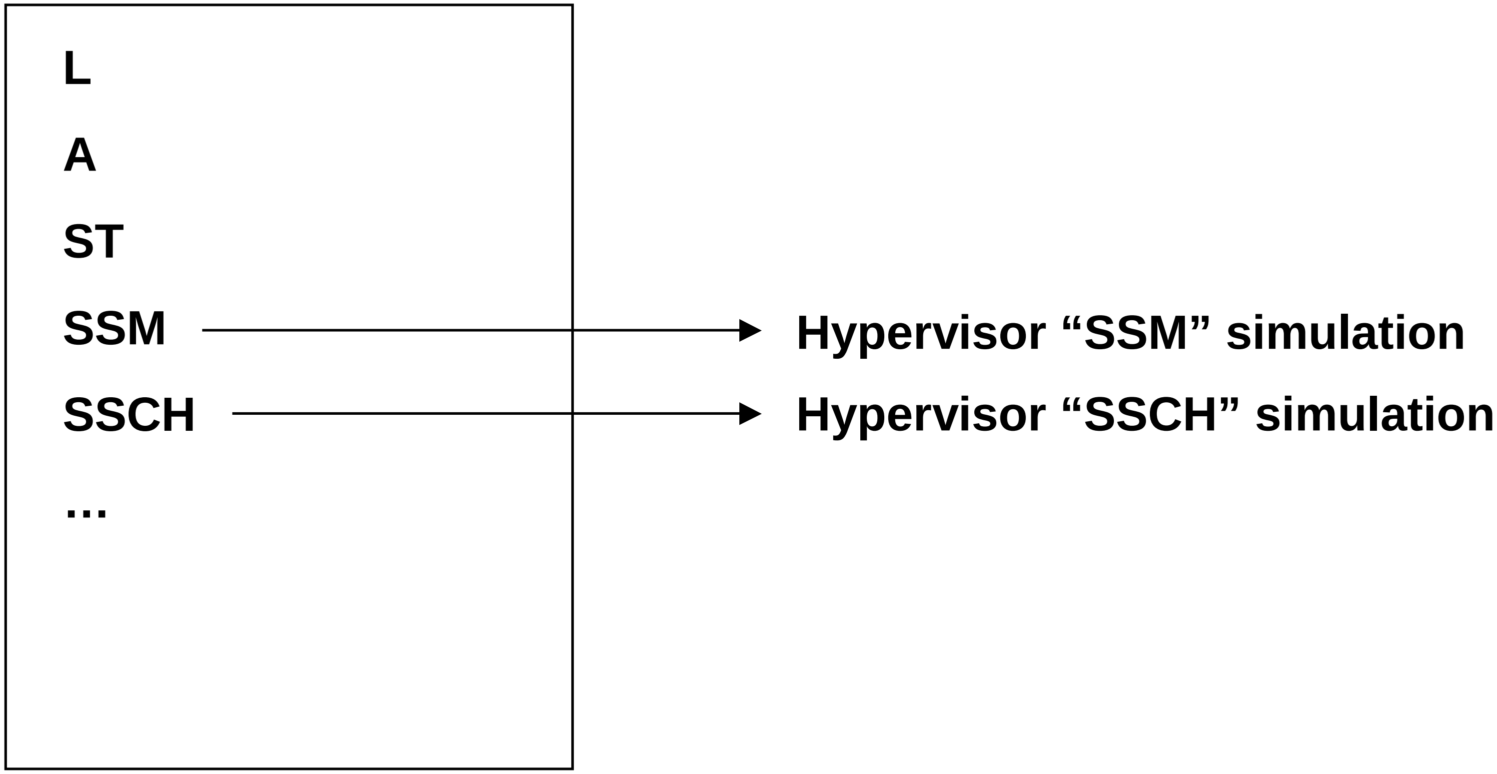
Definitions …

- **Virtualization**
  - **Virtualization is a method by which systems resources, which may be centralized or distributed, are aggregated and managed in shared resource pools and apportioned to users as virtual system resources**
  - **Virtualization separates the presentation of resources to users from the actual physical resources**
  - **Virtual resources correspond to all types of physical resources, including processors, memory, storage, SMP servers, clusters, and networks**

Hypervisor Technologies

- **"Trapping and Mapping" Method**
  - **Guest OS is run in user mode**
  - **Hypervisor runs in privileged mode**
  - **Privileged instructions trap to hypervisor**
  - **IA-32 complications**
    - **Instructions that behave differently in privileged and user modes (e.g, POPF treatment of interrupt enable flag)**
    - **User mode instructions that access privileged resources/state**
  - **Some guest kernel binary translation may be required**
  - **Originally used by CP/67 and VM/370**
  - **Used by VMware today**

Hypervisor Technologies – Trapping and Mapping



L

A

ST

SSM → Hypervisor “SSM” simulation

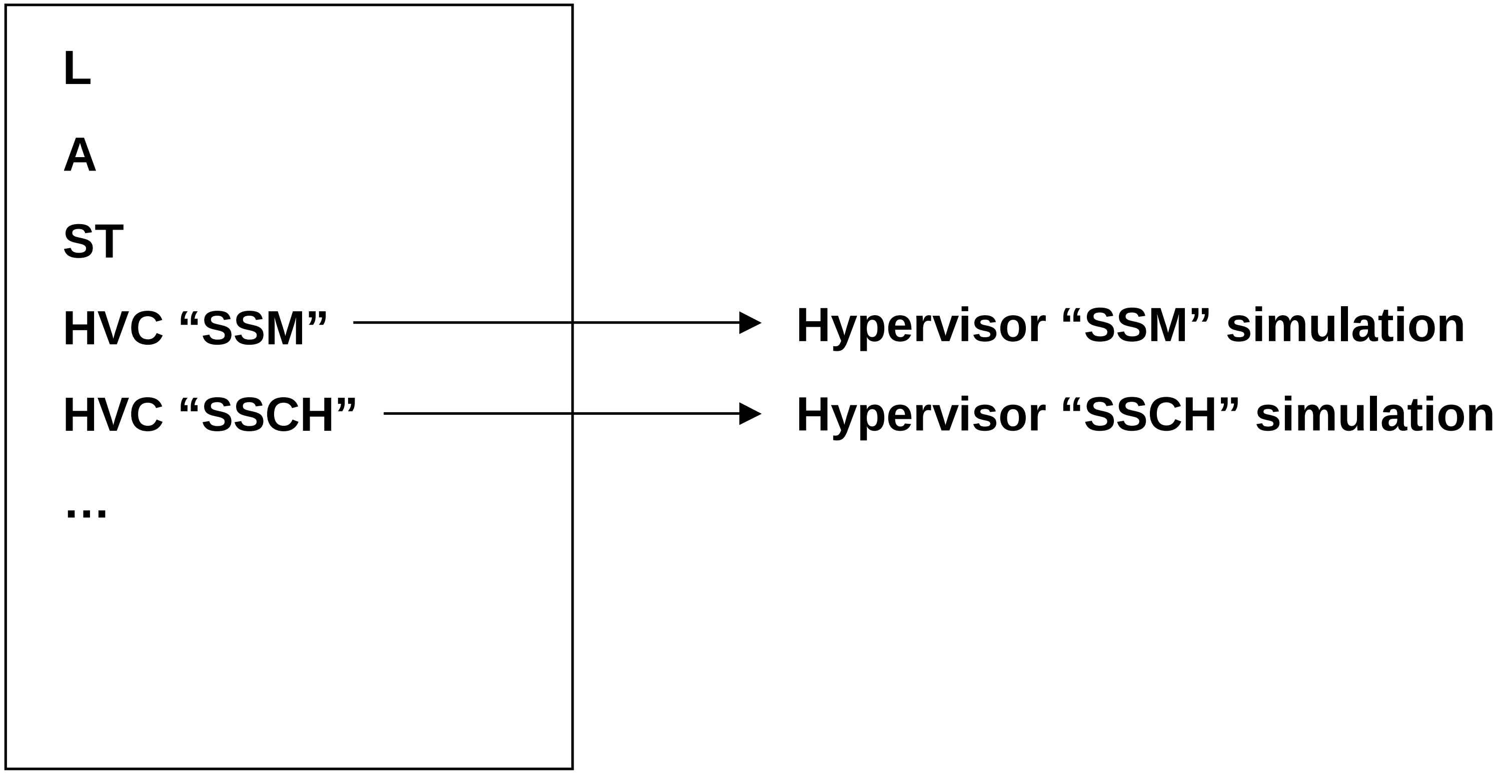
SSCH → Hypervisor “SSCH” simulation

…

Hypervisor Technologies

- **Hypervisor Call Method**
  - **Guest OS is run in privileged mode**
  - **Hypervisor runs in super-privileged mode**
  - **Guest OS kernel is modified to do hypervisor calls for I/O, memory management, yield rest of time slice, ...**
  - **Memory mapping architecture is used to isolate guests from each other and to protect hypervisor**
  - **Used by iSeries and pSeries today**

Hypervisor Technologies – Hypervisor Call



L

A

ST

HVC “SSM” → Hypervisor “SSM” simulation

HVC “SSCH” → Hypervisor “SSCH” simulation

…

Hypervisor Technologies

- Direct Hardware Support Method
  - Guest OS is run in privileged mode
  - Guest OS can be run unmodified but may issue some hypervisor calls to improve performance or capability
    - I/O (z/VM)
    - Yield time slice (PR/SM, z/VM)
  - Extensive hardware assists for hypervisor (virtual processor dispatching, I/O pass-through, memory partitioning, ...)
  - Used by zSeries PR/SM and z/VM

Hypervisor Technologies – Direct Hardware Support

L

A

ST

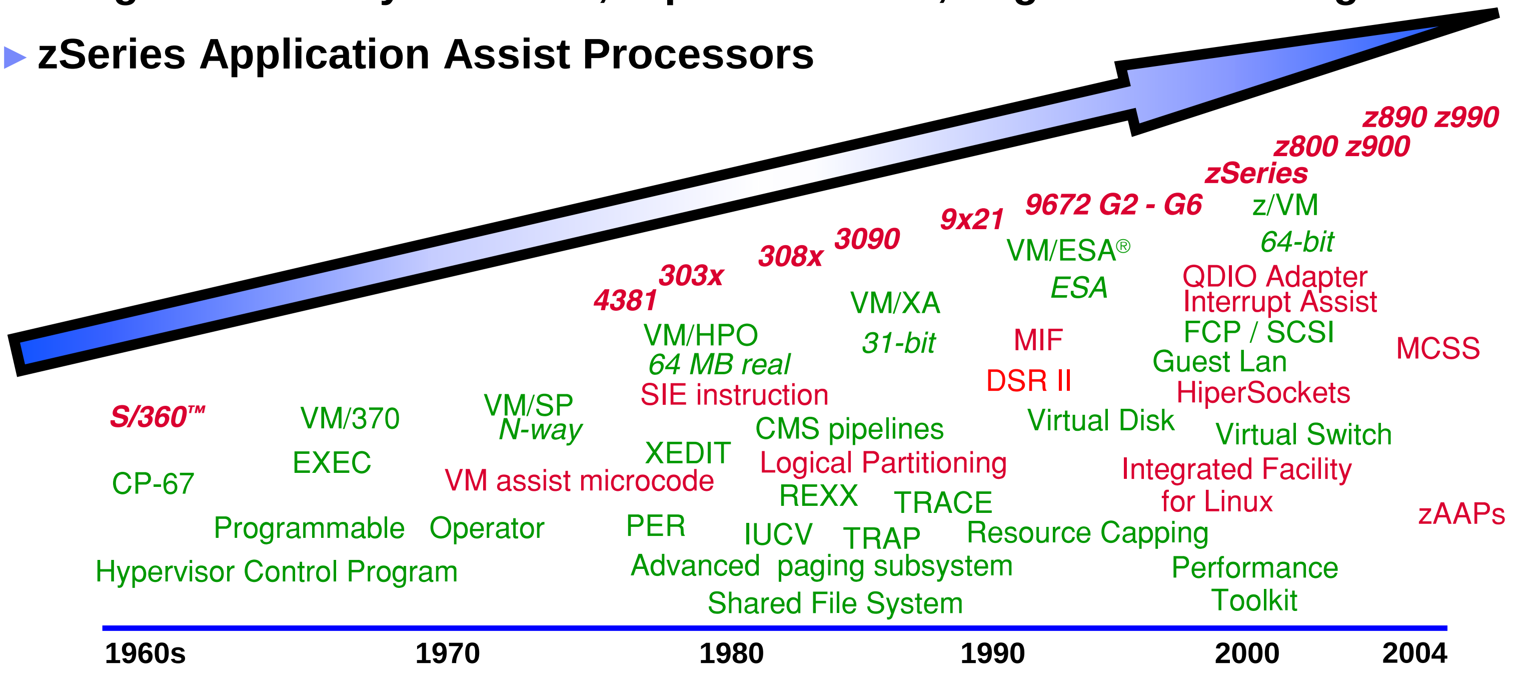
SSM

SSCH

…

IBM zSeries Virtualization Technology Evolution

- **Over 35 years of continuous innovation in virtualization**
  - **Refined to support modern business requirements**
  - **Exploit hardware technology for economical growth**
  - **Integrated Facility for Linux, HiperSockets™, Logical Partitioning**
  - **zSeries Application Assist Processors**



S/360™
CP-67
Hypervisor Control Program
Programmable
VM/370
EXEC
Operator
VM/SP
N-way
VM assist microcode
4381
303x
VM/HPO
64 MB real
SIE instruction
XEDIT
PER
Advanced
308x
CMS pipelines
Logical Partitioning
REXX
IUCV
TRAP
paging subsystem
Shared File System
3090
VM/XA
31-bit
TRACE
9x21
9672 G2 - G6
VM/ESA®
ESA
MIF
DSR II
Virtual Disk
Resource Capping
zSeries
z800 z900
z890 z990
z/VM
64-bit
QDIO Adapter
Interrupt Assist
FCP / SCSI
Guest Lan
HiperSockets
Virtual Switch
Integrated Facility
for Linux
Performance
Toolkit
MCSS
zAAPs

1960s **1970** **1980** **1990** **2000** **2004**

zSeries – comprehensive and sophisticated suite of virtual function

LPAR and z/VM – World-Class Server Virtualization

- **LPAR has grown up as a hardware feature supporting *virtual servers* in high-performance partitions by logically partitioning physical resources**

- **VM has grown to support 1000s of *virtual servers* by truly virtualizing resources such as storage and I/O**

- **Both employ great hardware and firmware innovations developed over the years that make virtualization part of the *basic componentry* of the zSeries platform**

Interpretive Execution

- SIE (Start Interpretive Execution) instruction
  - Operand is a state descriptor for a logical partition or virtual machine
  - Accommodates fixed-storage and pageable guests
  - Interception controls allow hypervisor intervention
- zSeries implements two levels of SIE
  - No performance penalty for running z/VM in a logical partition
    - Exception: preferred guests not supported
  - 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

Zone Relocation

- **SIE capability to provide multiple zero-origin storage regions (i.e., logical partitions) on one system**
  - **Enables I/O subsystem to access partition memory directly, without hypervisor intervention**

Multiple Image Facility

- **I/O subsystem channel path resources can be manifested in multiple logical partitions and shared among them**
- **I/O devices on shared channel paths can be accessed simultaneously by sharing logical partitions**
- **I/O devices on shared channel paths can be restricted to use by a subset of the sharing logical partitions**

Multiple Channel Subsystems

- **Channel subsystem limited to 256 channel paths**
- **With multiple channel subsystems**
  - **Architecture is preserved for logical partitions and guests**
  - **Additional I/O resources can be configured for a single hardware system**

Hipersockets

- **Very high speed, secure, memory-based communication mechanism**
- **zSeries hardware provides internal Queued Direct I/O channel paths for inter-LPAR communication**
- **z/VM provides virtual internal Queued Direct I/O subchannels for inter-virtual machine communication**

Adapter Interruption Pass-Through

- **QDIO devices (FCP, OSA Express) induce overhead due to high interruption rates**
  - **z/VM Control Program has to mediate between hardware interruptions and guests**
  - **As interruption rates go up, this overhead increases**
- **New hardware facility designed to address this problem**
  - **Allows interruptions to be presented directly by hardware for active guest**
  - **Delivers "thin" signal to CP when interruption is for idle guest**

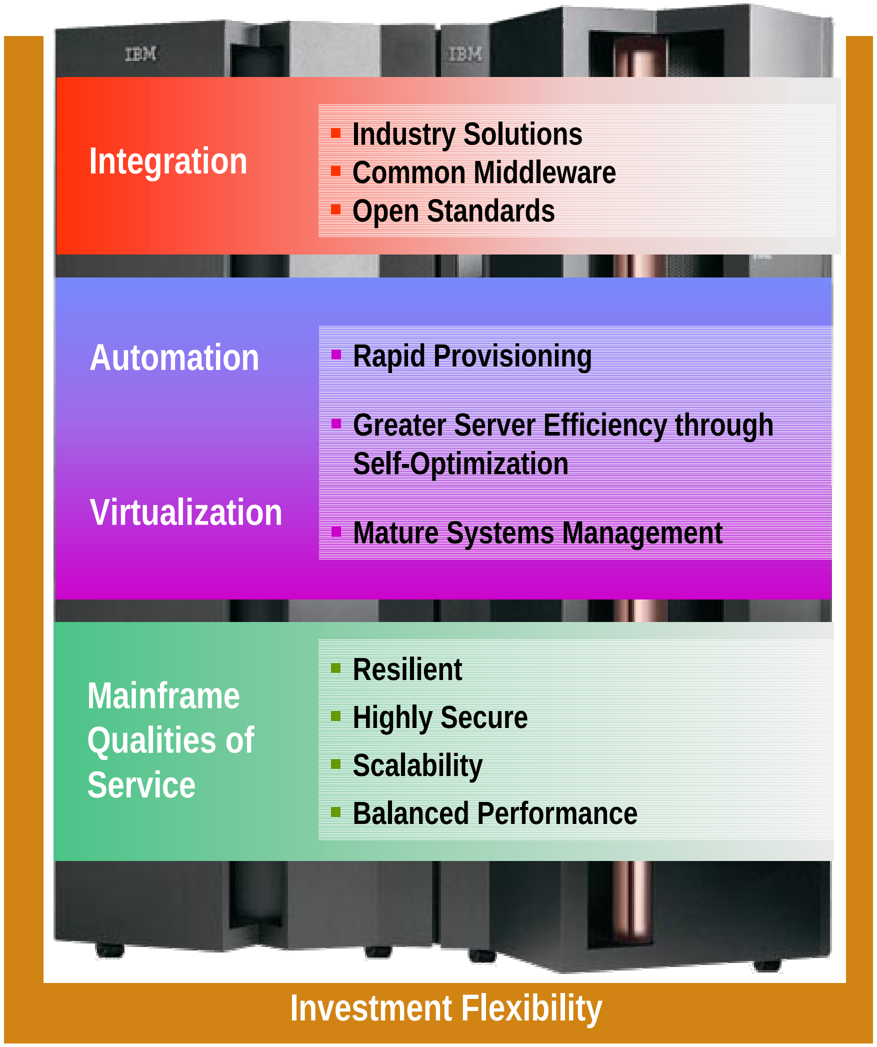
LPAR and z/VM – World-Class Server Virtualization

- **Together on zSeries, LPAR and z/VM technologies provide**
  - **A modest but growing number of high-performance “on the metal” virtual servers for larger, performance-critical workloads**
  - **The ability to provision 1000s of additional virtual servers flexibly and on demand**
- **How many Virtual Servers can you do on zSeries?**
  - **How many do you need?**
  - **Yeah, we can do that**

Environments and Uses – Reasons Remain Today

- **Hardware Consolidation**
- **Software Migration**
- **Development, Test, and Maintenance**
- **Diverse Workloads**
- **Constrained Systems**
- **Backup and Recovery**
- **Workload Isolation**
- **Coupling and Parallel Sysplex**
- **LPAR Clusters**

Future Direction - Continue Core Value Investments



Integration

- Industry Solutions
- Common Middleware
- Open Standards

Automation

Virtualization

- Rapid Provisioning
- Greater Server Efficiency through Self-Optimization
- Mature Systems Management

Mainframe Qualities of Service

- Resilient
- Highly Secure
- Scalability
- Balanced Performance

Investment Flexibility

- **Scalability**
  - Higher performance and capacity processors
  - More processors per server
  - Increased Connectivity – more connections, higher speed
- **Virtualization**
  - Increased number of LPARs
  - Increased number of LCSSs
- **Systems Management**
  - eWLM
  - Tivoli Provisioning Manager
  - Tivoli Intelligent Orchestrator
- **Security**
  - Trusted Computing
  - WebServices security
- **Resiliency**
  - Extended GDPS capabilities for Linux

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

RETURN TO INDEX