z/VM Platform Update & Putting SMT to Work on z/VM

April 10, 2016 Version 4.18

Double Header!

Bill Bitner
z/VM Development Client Focus & Care
bitnerb@us.ibm.com

CAVMEN – May 5th, 2016

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

BladeCenter*	FICON*	OMEGAMON*	RACF*	System z9*	zSecure
DB2*	GDPS*	Performance Toolkit for VM	Storwize*	System z10*	z/VM*
DS6000*	HiperSockets	Power*	System Storage*	Tivoli*	z Systems*
DS8000*	HyperSwap	PowerVM	System x*	zEnterprise*	
ECKD	IBM z13*	PR/SM	System z*	z/OS*	

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.

Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

OpenStack is a trademark of OpenStack LLC. The OpenStack trademark policy is available on the OpenStack website.

TEALEAF is a registered trademark of Tealeaf, an IBM Company.

Windows Server and the Windows logo are trademarks of the Microsoft group of countries.

Worklight is a trademark or registered trademark of Worklight, an IBM Company.

UNIX is a registered trademark of The Open Group in the United States and other countries.

* Other product and service names might be trademarks of IBM or other companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

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

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

This information provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g., zIIPs, zAAPs, and IFLs) ("SEs"). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the "Authorized Use Table for IBM Machines" provided at www.ibm.com/systems/support/machine_warranties/machine_code/aut.html ("AUT"). No other workload processing is authorized for execution on an SE. IBM offers SE at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

Notice Regarding Specialty Engines (e.g., zIIPs, zAAPs and IFLs):

Any information contained in this document regarding Specialty Engines ("SEs") and SE eligible workloads provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g., zIIPs, zAAPs, and IFLs). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the "Authorized Use Table for IBM Machines" provided at
www.ibm.com/systems/support/machine_warranties/machine_code/aut.html ("AUT").

No other workload processing is authorized for execution on an SE.

IBM offers SEs at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

Acknowledgments

- Alan Altmark

- Bill Bitner

- Miguel Delapaz

- Glenda Ford

- John Franciscovich

- Les Geer

- Susan Greenlee

- Dan Griffith

- Brian Hugenbruch

- Emily Hugenbruch

- Arwed Tschoeke

- Romney White

Agenda

- Release Status and Information

- z/VM® Version 6 Release 3

  – 2014 Enhancements

  – 2015 Enhancements

  – 2016 Enhancements

- z/VM® Version 6 Release 4 Preview

- Futures and Statements of Direction

Release Status and Information

z/VM Release Status Summary

z/VM Level	GA	End of Service	End of Marktg.	Minimum Processor Level	Maximum Processor Level	Security Level
6.4	4Q2016			IBM System z196 & z114®	-	
6.3	7/2013	12/2017[4]		IBM System z10®	-	EAL 4+ OSPP-LS
6.2	12/2011	07/2017[5]	7/2013	IBM System z10®	z13 [3]	-
5.4	9/2008	12/2016[1]	3/2012	IBM eServer zSeries 800& 900	zEC12	-

[1] Or later (Announced August 6, 2014)
[2] Extended from original date (Announced February 4, 2014)
[3] Announced January 14, 2015
[4] Announced February 3, 2015
[5] Announced February 2, 2016

Marketed & Serviced

Serviced, but not Marketed

End of Service & Marketing

z/VM Version 6
Security Certification Results

- Common Criteria (ISO/IEC 15408)
  - ***new*** z/VM V6.3 has been certified: BSI-DSZ-CC-0903
  - z/VM V6.1 has been certified: BSI-DSZ-CC-0752
  - Evaluated to EAL 4+ for the Operating System Protection Profile (OSPP) with:
    - Virtualization extension (-VIRT)
    - Labeled Security extension (-LS)

- Federal Information Processing Standard (FIPS) 140-2
  - ***new*** z/VM V6.3 System SSL is FIPS 140-2 Validated$^{(TM)}$
    - http://csrc.nist.gov/groups/STM/cmvp/documents/140-1/1401val2014.htm#2139
  - z/VM V6.1 System SSL is FIPS 140-2 Validated$^{(TM)}$
    - http://csrc.nist.gov/groups/STM/cmvp/documents/140-1/1401val2012.htm#1735
  - Enablement requirements for certificate database and servers

- z/VM V6.2 is <u>designed to conform</u> to both Common Criteria and FIPS 140-2 evaluation requirements
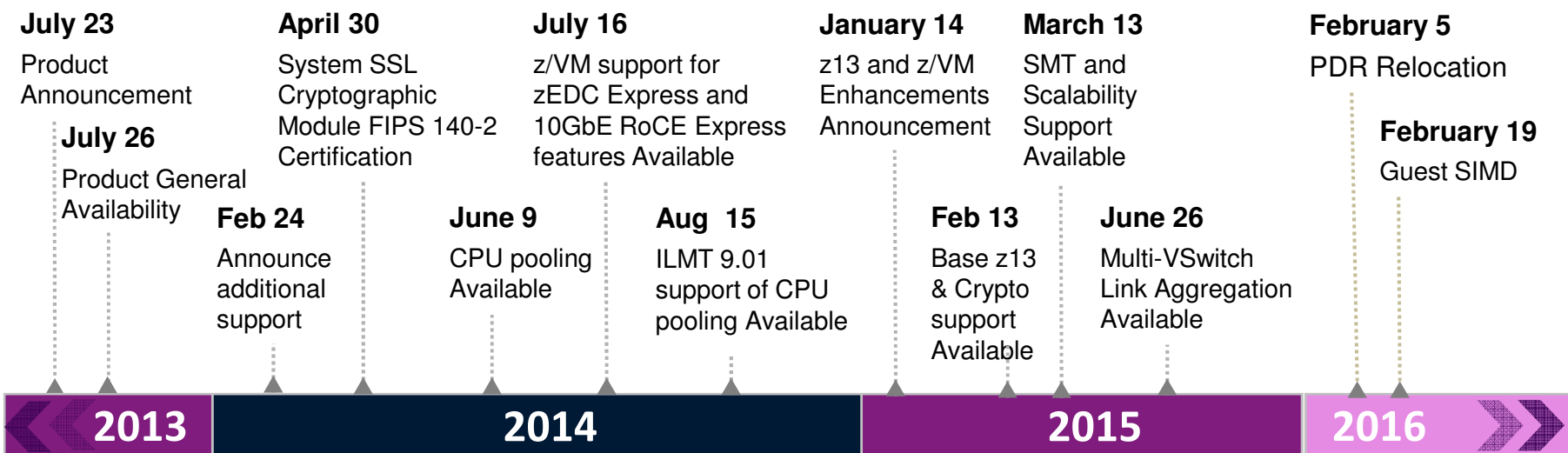
A Certification Mark of NIST, which does not imply product endorsement by NIST, the U.S. or Canadian Governments.
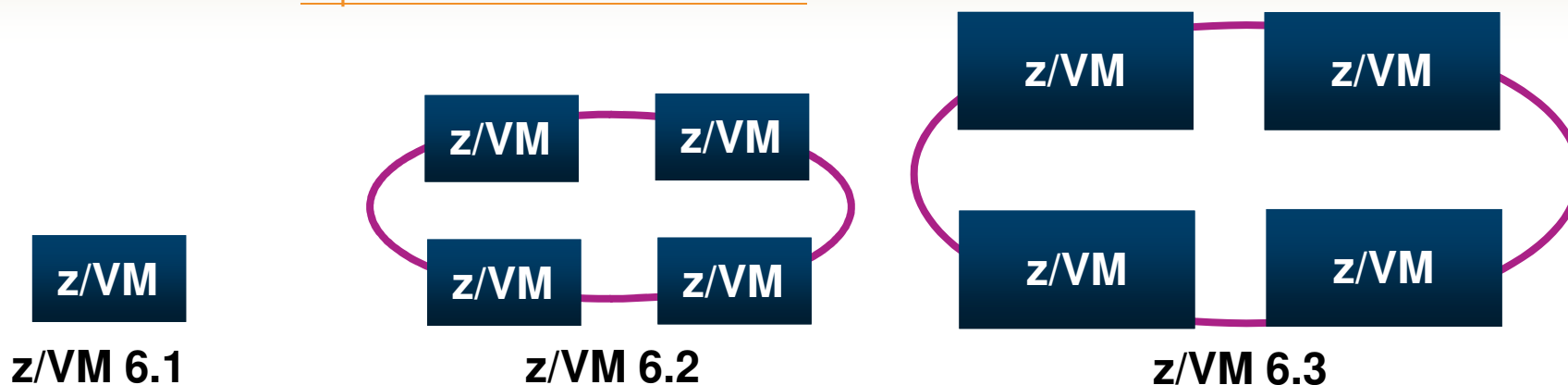
z/VM Version 6 Release 3

z/VM Version 6 Release 3
Making Room to Grow Your Business

July 23
Product Announcement

July 26
Product General Availability

April 30
System SSL Cryptographic Module FIPS 140-2 Certification

Feb 24
Announce additional support

July 16
z/VM support for zEDC Express and 10GbE RoCE Express features Available

June 9
CPU pooling Available

Aug 15
ILMT 9.01 support of CPU pooling Available

January 14
z13 and z/VM Enhancements Announcement

Feb 13
Base z13 & Crypto support Available

March 13
SMT and Scalability Support Available

June 26
Multi-VSwitch Link Aggregation Available

February 5
PDR Relocation

February 19
Guest SIMD

2013	2014	2015	2016

See http://www.vm.**ibm.com**/zvm630/

z/VM z/VM
z/VM z/VM

z/VM z/VM
z/VM z/VM

z/VM

z/VM 6.1

z/VM 6.2

z/VM 6.3

z/VM 6.3 Themes

- Improve Total Cost of Ownership
  - Expand z/VM systems constrained by memory or processor limitations with Large Memory Support, HiperDispatch, Scalability Enhancements, MSS for GDPS, etc.

- Enhance the Systems Management Experience
  - Enablement of OpenStack®, Upgrade in Place

- Continue Virtualization Leadership through Innovation
  - CPU Pooling, Simultaneous Multithreading (SMT), Efficiency Improvements, Virtual Switch Enhancements
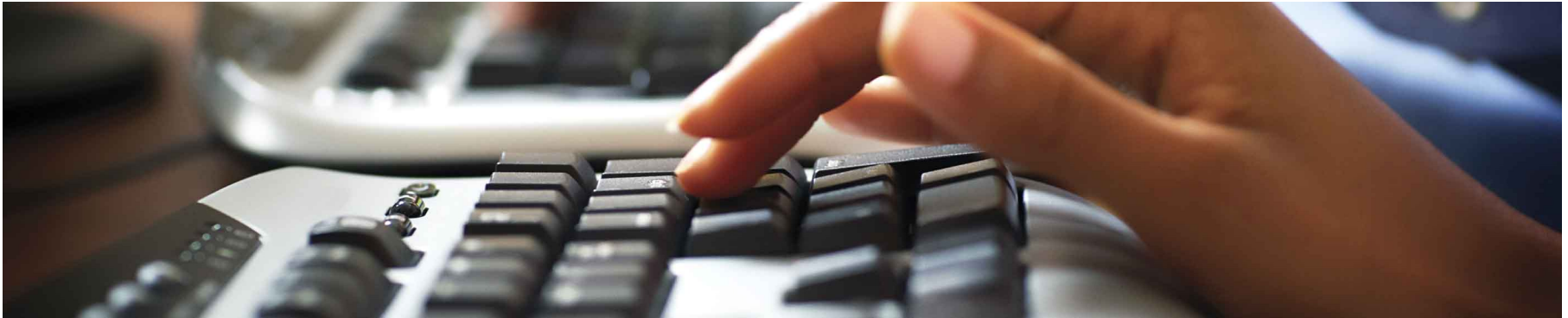
Large Memory Support

- Real memory limit raised from 256GB to **1 TB**
  - Proportionately increases total virtual memory based on tolerable overcommitment levels and workload dependencies

- Virtual machine memory limit remains unchanged at **1 TB**

- Paging DASD utilization and requirements change
  - Removed the need to double the paging space on DASD
  - Paging algorithm changes increase the need to have a properly configured paging subsystem

- Convert expanded storage to real storage
  - Expanded Storage continues to be supported with a limit of **128 GB,** but is no longer recommended.

Large Memory Support (continued)

- Reorder processing removed
  - Commands remain, but have no impact
  - Improves environment for running larger virtual machines

- Improved effectiveness of the CP SET RESERVE command
  - Stronger "glue" to hold reserved pages in memory
  - Support for reserving pages of NSS or DCSS
    - Example: Use with the Monitor Segment (MONDCSS)
  - Ability to limit the overall number of reserved pages for the system

Enhanced Dump Support

- Stand-alone Dump utility has been rewritten
  - Creates a CP hard abend format dump
  - Dump is written to ECKD™ or SCSI DASD

- Larger memory sizes supported, up to a maximum of 1 TB
  - Includes Stand-alone dump, hard abend dump, SNAPDUMP, DUMPLD2, and VM Dump Tool

- Performance improvements for hard abend dump
  - Reduces time to take a CP hard abend dump

HiperDispatch

- Improved processor efficiency
  - Better n-way curves
    - Supported processor limit of 32 remains unchanged
  - Better use of processor cache to take advantage of cache-rich system design of more recent machines

- Two components:
  - Dispatching affinity
  - Vertical CPU management

HiperDispatch – Dispatching Affinity

- Dispatcher is aware of the cache and memory topology
  - Dispatch virtual CPU near where its data may be in cache based on where the virtual CPU was last dispatched

- Better use of cache can reduce the execution time of a set of related instructions

- z/VM 6.2 and earlier uses "soft" affinity to dispatch virtual CPUs
  - No awareness of chip or book

HiperDispatch – Vertical CPU Management

- "Horizontal" management distributes the LPAR weight evenly across the logical processors of the z/VM LPAR

- "Vertical" management attempts to minimize the number of logical processors, allowing LPAR to similarly manage logical CPUs

Example:
- Ten Physical IFLs, seven logical IFLs, weight of 400 out of 1000
  - Each logical IFL (LPU) entitled to 57% of an IFL
- When CEC is constrained, the LPAR's entitlement is reduced to four IFLs, so seven is more than required
- z/VM and LPAR will cooperate
  - z/VM will concentrate the workload on a smaller number of logical processors
  - LPAR will redistribute the partition weight to give a greater portion to this smaller number of logical processors (~100% of four CPUs)
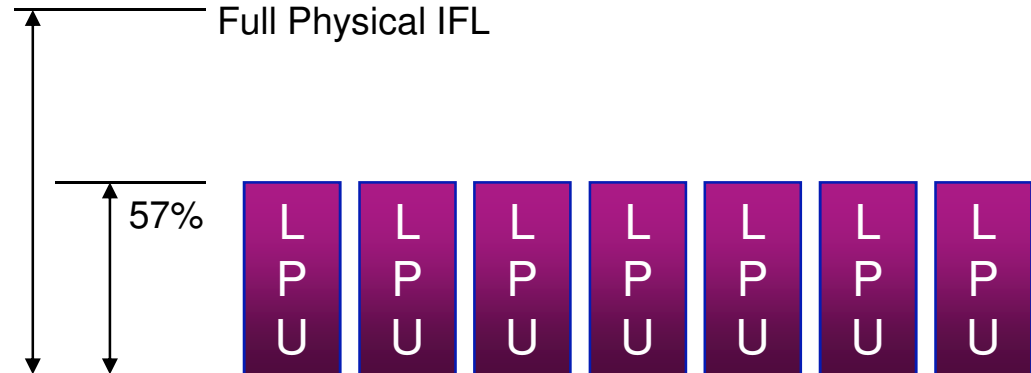
Horizontal vs. Vertical CPU Management

Horizontal:

- The logical processors are all created/treated equally.
- z/VM dispatches work evenly across the seven logical processors

Full Physical IFL

57%

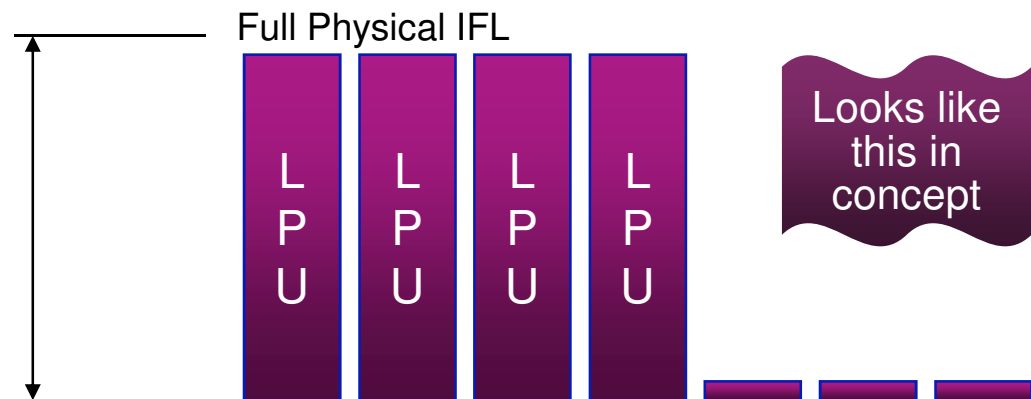| L P U | L P U | L P U | L P U | L P U | L P U | L P U |

Example:
- 10 Physical IFL Processors
- 7 Logical IFL Processors
- Entitlement of 4 IFL Processors

Vertical:

- The logical processors are skewed to where some get greater share of the weight.
- z/VM dispatches work accordingly to the heavier weighted workload.

Full Physical IFL

| L P U | L P U | L P U | L P U |

Looks like this in concept

Technology Exploitation

- Fibre Channel Protocol Data Router Support
  - FCP QEBSM support enhanced for guest support use of FCP Data Router

- FICON DS8000 Series New Functions
  - Storage Controller Health message
    - New attention message from hardware providing more details for conditions in past reflected as Equipment Check.
    - Intended to reduce the number of false HyperSwap® events.
  - Peer-to-Peer Remote Copy (PPRC) Summary Unit Check
    - Replaces a series of state change interrupts for individual DASD volumes with a single interrupt per LSS
    - Intended to avoid timeouts in GDPS environments that resulted from the time to process a large number of state change interrupts.
    - Satisfies a SOD from October 12, 2011

- Multiple Subchannel Set (MSS) support for mirrored DASD
  - Support to use MSS facility to allow use of an alternate subchannel set for Peer-to-Peer Remote Copy (PPRC) secondary volumes.
  - Satisfies a SOD from October 12, 2011

z/VM 6.3 and GDPS Support

- z/VM 6.3 alternate subchannel set support
  - GDPS V3.10 prereqs the PM71447 New Function: GDPS/PPRC XDR MSS1 Support APAR

- z/VM 6.3 FICON DS8000 Series new function (DS8K synergy initiative)
  - GDPS/PPRC V3.8, V3.9, & V3.10 and prereqs the PM44141 New Function: GDPS/PPRC XDR PPRCSUM and Storage Controller Health Message APAR, and DS8K R6.2 u-code.

- Cannot mix new MSS support in an SSI environment with older z/VM systems.

- See http://www-03.ibm.com/systems/z/advantages/gdps/whatsnew.html for details.

- See GDPS PSP buckets for required service (z/OS, Linux, and z/VM)
  - Remember to check for required service for systems that share the GDPS environment.

Environment	3.8	3.9	3.10
z/VM 6.3 w/ MSS 1	No	No	Yes[1]
z/VM 6.3 DS8K Synergy	Yes[1]	Yes[1]	Yes[1]
z/VM 6.3 SSI + LGR	No	No	Yes[1]

1 – with appropriate service – Check Bucket

Virtual Networking Improvements

- Live Guest Relocation support for port-based virtual switches built on existing support:
  – Allow relocation of port-based interface
  – Prevent relocation of an interface that will be unable to establish proper network connectivity
  – Adjust the destination virtual switch configuration, when possible, by inheriting virtual switch authorization from the origin

- MPROUTE server upgraded to z/OS V1.13 OMPROUTE functional equivalency

- Support for OSA-Express5S devices

- Virtual Switch recovery and stall prevention
  – New SET VSWITCH UPLINK SWITCHOVER command
  – Change from current device to one of the configured backup devices

Security Enhancements

- Crypto Express4S
  - Guest support for Crypto Express4S which is a feature available on zEC12 and zBC12
  - Can be configured in one of three ways:
    - IBM Common Cryptographic Architecture (CCA) Coprocessor mode
    - IBM CCA Accelerator mode
    - IBM Enterprise Public Key Cryptographic Standards (PKCS) #11 (EP11) coprocessor

- SSL Server Upgrade
  - System SSL update to z/OS V1.13 equivalency
  - Client certificate validation
  - Includes support for:
    - Transport Layer Security (TLS) protocol, Version 1.2
    - SHA2 certificate support
    - TLS Protocol Selection
  - IPv6 support for SSL-enabled Telnet, FTP, and SMTP

Installation Upgrade in Place Enhancement

- Upgrade an existing z/VM 6.2 system to z/VM 6.3 with minimal impact to the current running system.
    - Fewer manual steps such as directory merging and new virtual machine creation

- Upgrade Approach:
    - Install new release as temporary second level system
    - Move new level of z/VM to current system
    - For SSI Cluster, start with single member of the cluster on new level

- Provides a backup to support backing out in extreme cases

- Support for local modifications

Linux Disk Dump Utility can now include the NSS
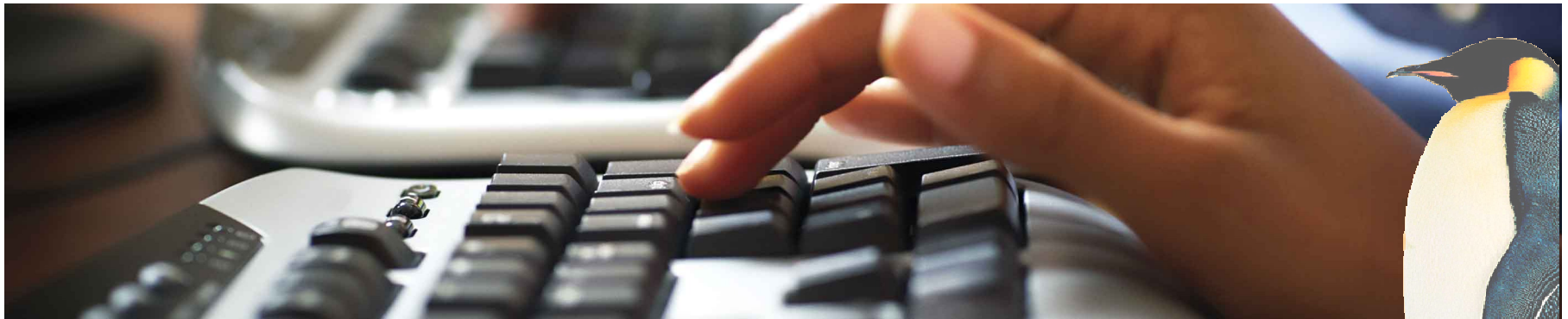
- The Linux Disk Dump utility is preferred over the CP VMDUMP command in most cases.

- Previously, the contents of an NSS could not be captured with Linux Disk Dump utility.

- Changes in IPL now allow the NSS to be included
  - New NSSDATA parameter

- For more background, see:
  - http://download.boulder.ibm.com/ibmdl/pub/software/dw/linux390/docu/l26ddt01.pdf for Linux Disk Dump utility information
  - http://www.vm.ibm.com/perf/tips/vmdump.html for information on differences between VMDUMP and Linux utility

# z/VM 6.3 Withdraws Cross System Extensions Support

- Satisfies a previous Statement of Direction

- The z/VM Single System Image (VMSSI) feature replaces the functions provided by CSE:
  - Logon once in the cluster, with exceptions
  - Cross-system MESSAGE and QUERY commands
  - Shared spool
  - Shared source directory

- VMSSI has additional value such as autonomic minidisk cache management and a single point of maintenance

- XLINK shared disk support is **not** affected. **Change from SoD**

z/VM 6.3 Withdraws support for TCP/IP Devices and Daemons

- Satisfies a previous Statement of Direction

- A220 HYPERchannel devices

- CLAW devices

- DHCP daemon

- LPSERVE (LPD)

    - RSCS LPD is provided at no charge

    - Does not affect LPR (client)
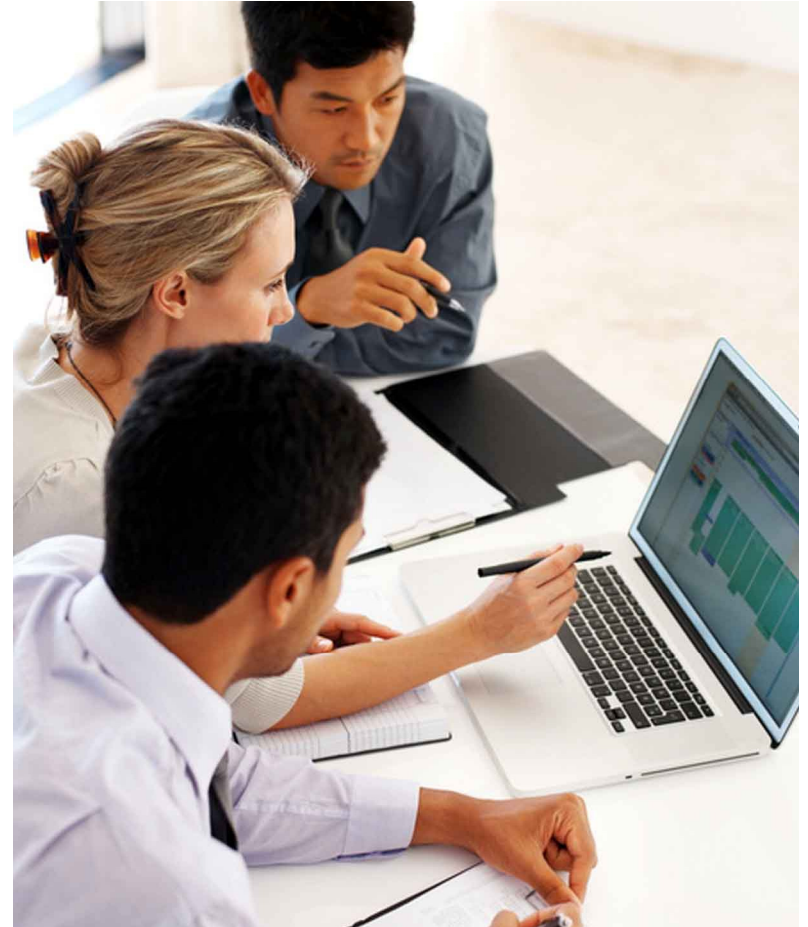
What IBM Wave is NOT:

Not just for novice users

Not just a Graphical User Interface

Not a z/VM system programmer replacement

Not excluded by in-house scripts

Not just for 100s of servers

Not just a cloning tool

ibm.com/linuxone

IBM Wave offers value in:

Advanced Visualization	Simplified Monitoring	Resource Management
▪ Shorten the learning curve needed to manage your Linux and z/V M environment	▪ Monitor z/VM system status through an innovative and interactive UI	▪ Manage your systems from a single workstation
▪ Organize and simplify administration of virtual Linux servers;; automate and simplify management steps	▪ Monitor performance of CPU, paging devices, spool disks and more;	▪ Assign and delegate administrative access using role based assignments
▪ View servers and storage utilization graphically; view resource status at a glance	▪ Use agentless and lightweight discovery for a current view of your environment	▪ Provision, clone, and activate virtual servers. Define and control virtual network and storage devices
▪ Use graphical or tabular displays with layered drill down; customize and filter views	▪ Use advanced filters, tagging, layout and layer selection to manage in a meaningful way	▪ Perform complex tasks such as live guest relocation using a few keystrokes
▪ Attach virtual notes to resources for additional policy-based management	▪ Complements IBM OMEGAMON® XE used for in-depth performance monitoring and historical views	▪ Execute complex scripts with a single mouse click
		▪ Report on resources with flexible resource reporting

IBM Infrastructure Suite for z/VM and Linux 1.1.0

- Announced and Available
  - Announced September 2, 2014
  - Available September 5, 2014
  - Announcement Letter ENUS214-350

- Includes following products:
  - IBM Tivoli® OMEGAMON® XE on z/VM and Linux V4.3
  - IBM Tivoli Storage Manager, part of IBM Spectrum Protect, Extended Edition V7.1
  - IBM Operations Manager for z/VM V1.5
  - IBM Backup and Restore Manager for z/VM V1.3
  - IBM Wave for z/VM V1.2

z/VM System Management – Related Products

- **Operations Manager for z/VM V1.5**
  - Facilitates automated operations
  - Monitor, view, and interact with consoles without logging on to service machines or Linux guests
  - Take actions based on service machine console messages and other system events
  - Schedule events for immediate execution or on a regular schedule

- **OMEGAMON® XE on z/VM and Linux V4.3**
  - Performance monitoring of z/VM and Linux guests
  - Part of the OMEGAMON and IBM Tivoli Monitoring infrastructure, including Tivoli Enterprise Portal
  - Uses IBM Performance Toolkit for VM as its data source

- **Backup and Restore Manager for z/VM V1.3**
  - Backup and restore file level data for CMS minidisks and Shared File System
  - Backup and restore images of Linux guests and/or z/VM volumes
    - Use Tivoli Storage Manager for file level backup and restore of Linux data

- **Tape Manager for z/VM V1.3**
  - Manage tapes: retention, access control, data security erase
  - Manage devices: share with other z/VM and non-z/VM systems
  - Manage mount requests for ATL, VTS, and manual mount devices
    - IBM TS7700: needs firmware update available as code level 8.21.0.165 (EC: M13120 / PN: 2727271 & 2727272 (DVD1&2.))
    - Oracle StorageTek automated tape libraries (ATL) and virtual tape libraries (VTL) - via either the STK VM Host Support Component or the STK VM Client
    - EMC Virtual Tape Libraries (VTL), such as the EMC DLm.

- **Archive Manager for z/VM V1.1**
  - Users and administrators manage disk space more efficiently and effectively
  - Archive infrequently used or large files to tape or other disk

- **zSecure™ Manager for RACF z/VM V1.11.1**
  - Automate complex, time consuming z/VM security management tasks
  - Quickly identify and prevent problems in RACF
  - Create comprehensive audit trails

All support z/VM 6.2 and 6.3!

February 24, 2014 Announcements

Enhancing the Foundation for Virtualization

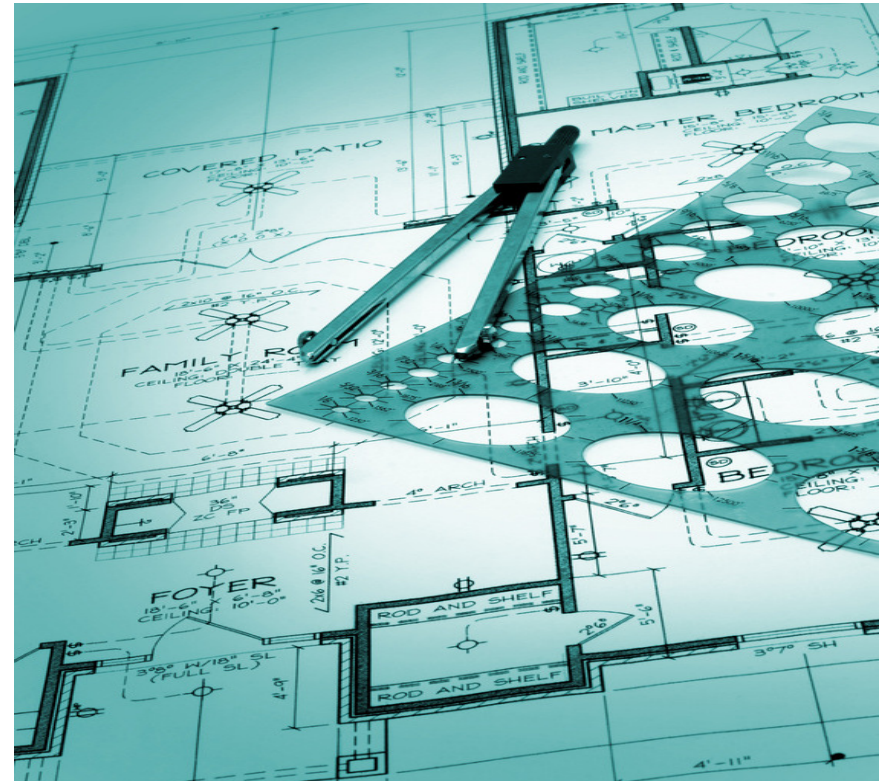- Release for Announcement – zBX and zEnterprise System Enhancements

  - February 24, 2014
  - http://www.vm.ibm.com/zvm630/apars.html

- Software Enhancements

  - CPU Pooling

  - Environment Information Interface

- Hardware Support

  - 10GbE RoCE Express Feature

  - zEDC Express Feature

CPU Pooling

- Fine-grained CPU limiting for a group of virtual machines

- Define one or more pools in which a limit of CPU resources is set.

- Two flavors of limits:
  - LIMITHARD - Percentage of system
  - CAPACITY – Number of CPUs

- Coexists with individual limit shares
  - More restrictive limit applies

- Support Details
  - z/VM 6.3 with APAR VM65418 – Available
    • Part of RSU 1501
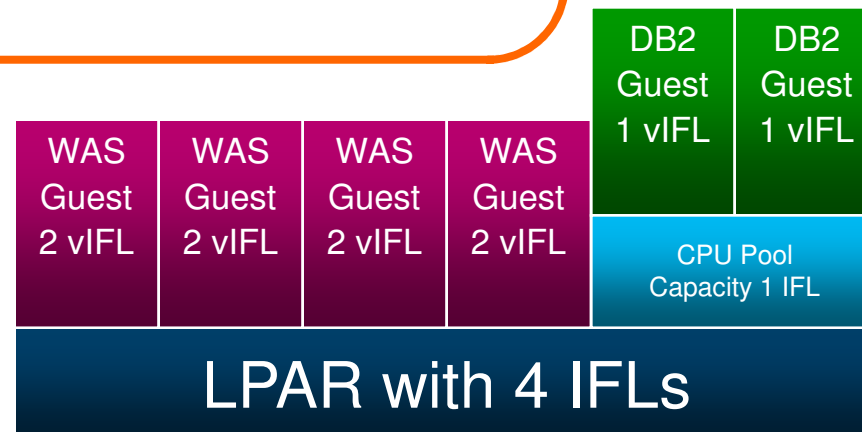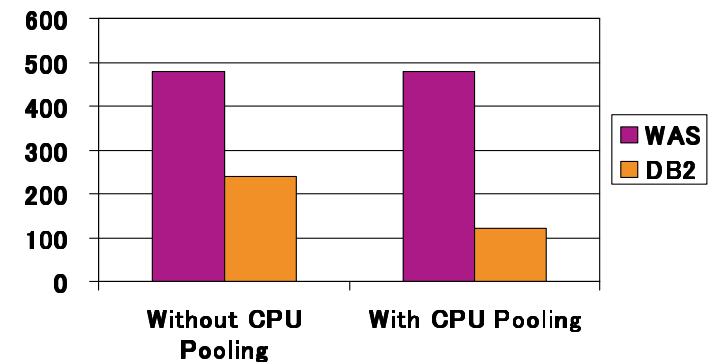
Environment Information Interface

- New interface allow guest to capture execution environment
  - Configuration and Capacity information
  - Various Levels:
    - Machine, logical partition, hypervisor, virtual machine

- New problem state instruction Store Hypervisor Information (STHYI)

- Includes support for CPU Pooling enhancement

- Foundation for future software licensing tools
  - IBM License Metric Tool 9.0.1updated August 2014- http://ibm.biz/cpupoolilmt
  - Greater flexibility for IBM Passport Advantage products

- Support details:
  - z/VM 6.3 with APAR VM65419 – Available
    - Part of RSU 1501

CPU Pooling Example

■4 WAS production guests
- Requires 4-engine WAS entitlement

■Create a 1-IFL pool

■Put the 2 DB2 production guests in pool
- Requires 1-engine DB2 entitlement (avoiding the need for 2-engine DB2 entitlement without CPU pooling)

PVU Entitlements

| | WAS |
| | DB2 |

Without CPU Pooling — With CPU Pooling

| DB2 Guest 1 vIFL	DB2 Guest 1 vIFL
WAS Guest 2 vIFL	WAS Guest 2 vIFL	WAS Guest 2 vIFL	WAS Guest 2 vIFL	CPU Pool Capacity 1 IFL

LPAR with 4 IFLs

■Allows new workloads and additional workload consolidation to be more cost effective

Note: All PVU Entitlement examples based on zEC12 (120 PVU per IFL) – will look proportionally the same on zBC12 (100 PVU per IFL)

10GbE RoCE Express Feature

- Support for RDMA over Converged Ethernet for guests

- Based on new hypervisor PCIe support

- Designed to support z/OS's Shared Memory Communications-Remote Direct Memory Access (SMC-R) in z/OS V2.1

- Support details:
  - IBM z13, zEC12, or zBC12 with appropriate updates – see support buckets
  - z/VM 6.3 with APAR VM65417 – Available – RSU 1501
    - System Config option – disabled by default.
    - You need to have required millicode fixes applied prior to enabling in system config
  - z/OS 1.12, z/OS 1.13, z/OS 2.1 with APAR OA43256
  - Fulfills 2013 Statement of Direction
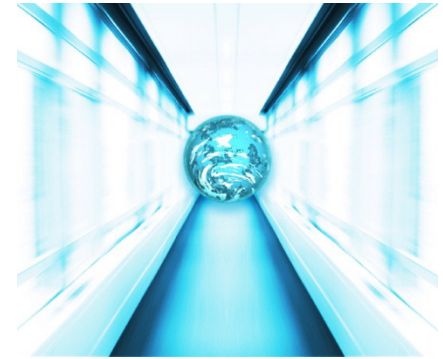
zEDC Express Feature

- Guest support for zEDC Express Feature

- High performance, low CPU consumption compression

- Possible disk utilization reduction

- Support details:
  - IBM zEC12 or zBC12 with appropriate updates – see support buckets
  - z/VM 6.3 with APAR VM65417 – Available – RSU 1501
    - System Config option – disabled by default.
    - You need to have required millicode fixes applied prior to enabling in system config
  - z/OS 1.12, z/OS 1.13, z/OS 2.1 with APAR OA43256
  - z/OS 1.12, z/OS 1.13, z/OS 2.1 with APAR OA44482
  - Fulfills 2013 Statement of Direction

January 14, 2015 Announcements

Expanding the Horizon of Virtualization

- Release for Announcement – The IBM z13™
  - January 14, 2015
  - Announcement Link

- z/VM Compatibility Support
  - PTFs available February 13, 2015
  - Also includes Crypto enhanced domain support
  - z/VM 6.2 and z/VM 6.3
  - No z/VM 5.4 support

- Enhancements and Exploitation Support - only on z/VM 6.3
  - IBM z13 Simultaneous Multithreading
  - Increased Processor Scalability
  - Multi-VSwitch Link Aggregation Support (Link Aggregation with Shared OSAs)

- Performance Report at http://www.vm.ibm.com/perf/reports/zvm/html/

z/VM Support for Crypto Express5S

- z/VM supports the z13 and Crypto Express5S feature
  - z/VM 6.2 and z/VM 6.3 only
  - APAR VM65577

- Expanded domain selection for dedicated domains
  - z/VM supports architected limits for CryptoExpress domains
  - CryptoExpress5S supports 85 domains per feature, with a maximum of 16 features

- Selection of APVIRT domains in System Configuration
  - Avoid collisions when reassigning domains in user directory
 - Minimize need for LPAR restart
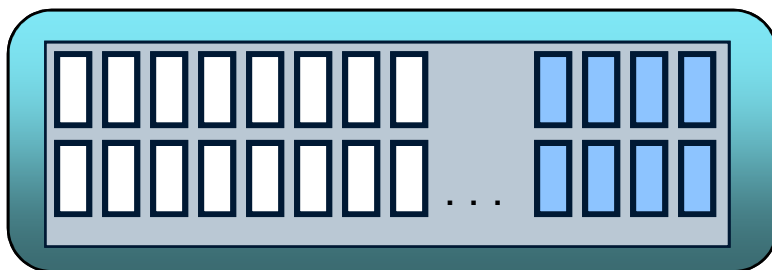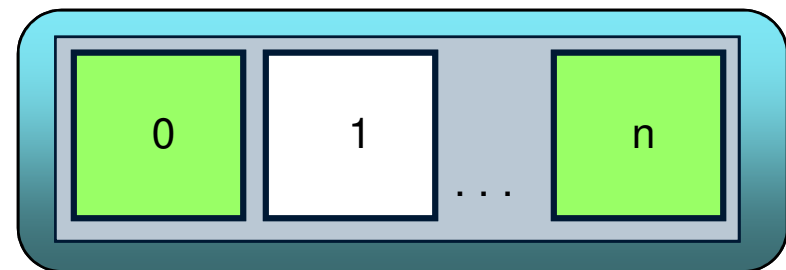
CEX5C 0   **CEX5A 1**

Simultaneous Multithreading (SMT)

- Objective is to improve capacity, not performance.

- Allows z/VM to dispatch work on up to two threads of a z13 IFL

- VM65586 for z/VM 6.3 **only**
  - PTFs available March 13, 2015

- At least z13 millicode bundle 11

- Transparent to virtual machine
  - Guest does not need to be SMT aware
  - SMT is not virtualized to the guest

- z13 SMT support limited to IFLs and zIIPs
  - z/VM support is only for IFLs

- SMT is disabled by default
  - Requires a System Configuration setting and re-IPL
  - When enabled, applies to the entire system

- Potential to increase the overall capacity of the system
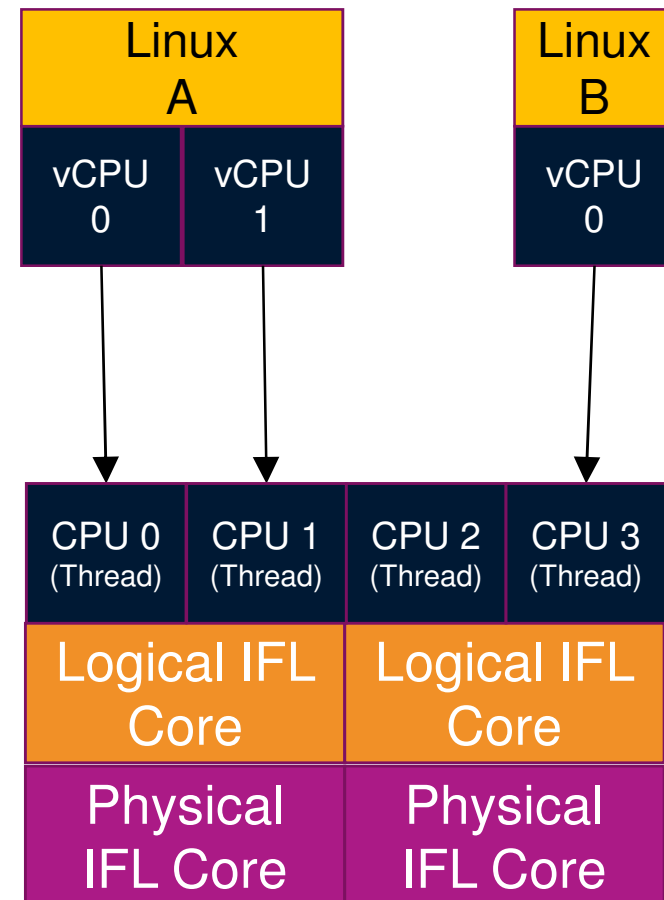 - Workload dependent
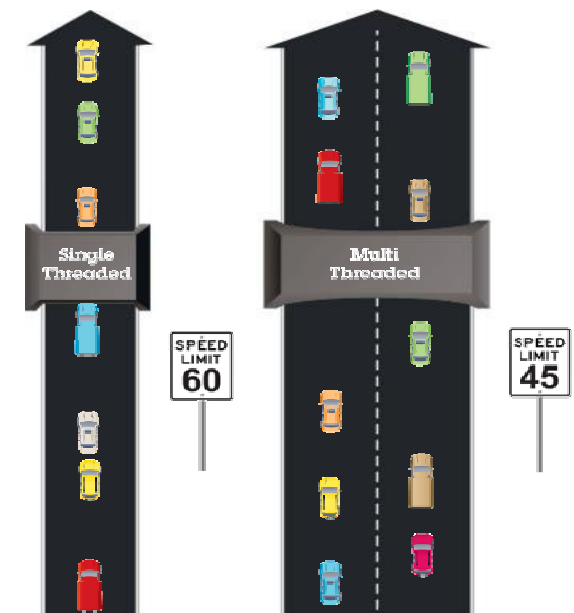
Which approach is designed for the higher volume of traffic? Which road is faster?

*Illustrative numbers only

SMT Usage

- Physical IFL Cores with SMT allow up to two threads to be used. You purchase these.

- Logical IFL Cores are presented to z/VM as in the past. You define these in the logical partition definition on HMC.

- z/VM creates a CPU or logical processor associated with each thread for it to use. Reflected in commands like QUERY PROCESSORS.

- The virtual CPUs of guests can then be dispatched on different threads intelligently, based on topology information.

Linux A		Linux B
vCPU 0	vCPU 1	vCPU 0

CPU 0 (Thread)	CPU 1 (Thread)	CPU 2 (Thread)	CPU 3 (Thread)
Logical IFL Core		Logical IFL Core	
Physical IFL Core		Physical IFL Core	

SMT Usage – Mixed Engine Environment

- In a mixed-engine environment, general purpose processors cannot do threading, but a second CPU address is consumed (CPU 1 in example).

- Virtual IFL CPUs would get dispatched to the logical IFLs and virtual CP CPUs would get dispatched on the logical CPs

TPF A	Linux A		Linux B
vCPU 0 CP	vCPU 0 IFL	vCPU 1 IFL	vCPU 0 IFL

CPU 0 (Thread)	CPU 2 (Thread)	CPU 3 (Thread)	CPU 4 (Thread)	CPU 5 (Thread)
Logical CP Core	Logical IFL Core		Logical IFL Core	
Physical CP Core	Physical IFL Core		Physical IFL Core	
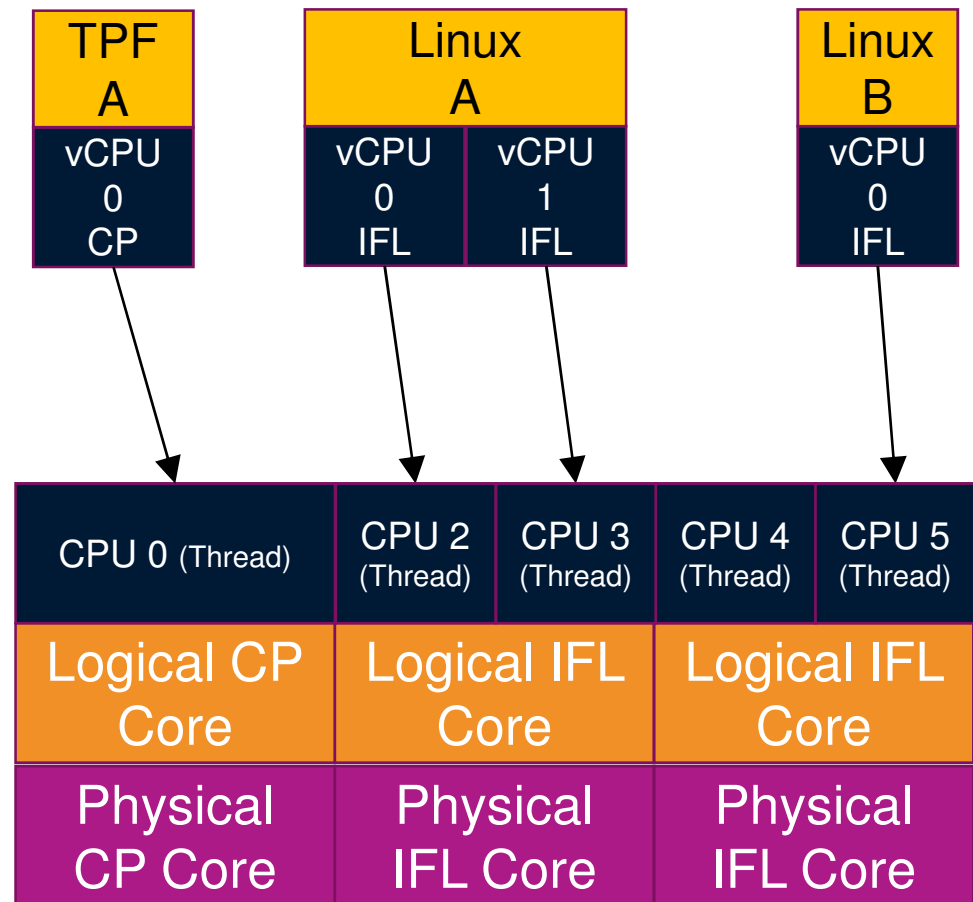
Increased CPU Scalability

- Various improvements to allow z/VM systems to be larger in terms of processors and more efficient, improving the n-way curve

- APARs VM65586 & VM65696 for z/VM 6.3 **only**
  - PTFs available March 13, 2015

- For z13
  - With SMT disabled, increases logical processors supported from 32 to 64
  - With SMT enabled, the limit is 32 cores (64 threads)

- For processors prior to z13
  - Limit remains at 32 cores
  - May still benefit from improved n-way curves

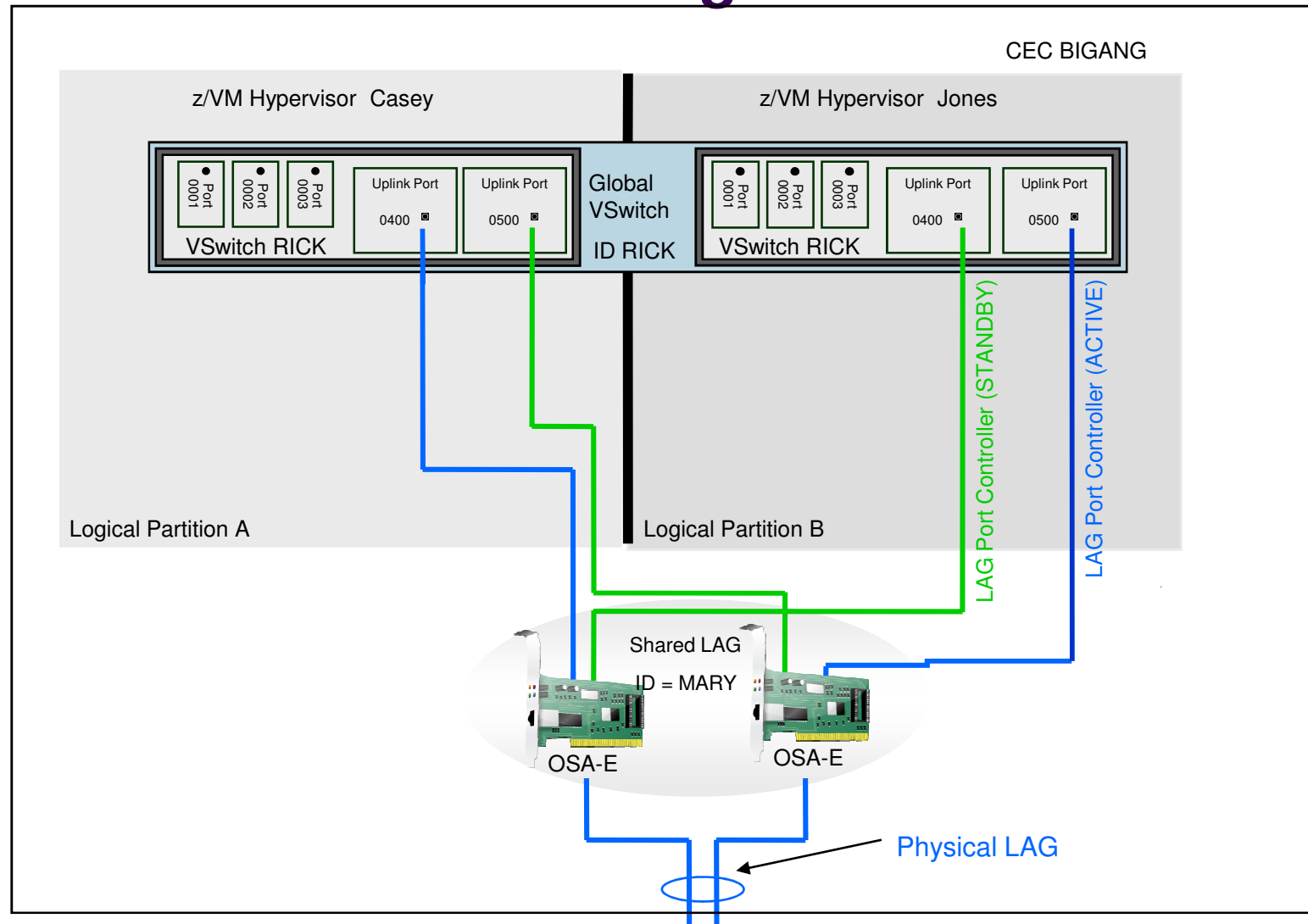# Areas Improved with Scalability Enhancements

- z/VM Scheduler Lock
  - Management of internal stacked work
  - Guests going into a wait state

- Locking for Memory Management
  - Most benefit during system initialization and when very constrained with memory

- Serialization and processing of VDisk I/Os

- Batching and processor-local queues for VSWITCH buffers

Multi-VSwitch Link Aggregation

- Link Aggregation is ability to combine or aggregate up to 8 OSAs to increase the bandwidth available to a VSwitch

- This enhancement makes it possible to do Link Aggregation with VSwitches with shared OSAs rather than previous the requirement for dedicated OSAs

- Allows a port group of OSA-Express features to span VSwitches within a single or multiple z/VM systems.
  - Cannot be shared with non-z/VM logical partitions

- APARs VM65583 (CP), PI21053 (TCP/IP), VM65528 (Performance Toolkit), and VM65670 (SMAPI) for z/VM 6.3 **only**
  - PTFs planned to be available June 26, 2015

- Available only on the z13
  - Requires OSA enhancements introduced with the z13

- Allows better consolidation and availability while improving TCO

Multi-VSwitch LAG Configuration

CEC BIGANG

z/VM Hypervisor Casey

z/VM Hypervisor Jones

VSwitch RICK

Port 0001 · | Port 0002 · | Port 0003 ·

Uplink Port 0400 ▣ | Uplink Port 0500 ▣

Global VSwitch ID RICK

VSwitch RICK

Port 0001 · | Port 0002 · | Port 0003 ·

Uplink Port 0400 ▣ | Uplink Port 0500 ▣

Logical Partition A

Logical Partition B

LAG Port Controller (STANDBY)

LAG Port Controller (ACTIVE)

Shared LAG

ID = MARY

OSA-E

OSA-E

Physical LAG

September 2015 Updates

Securing the Path to Virtualization

- Updates to RACF for z/VM V6.3 – **APAR VM65719**
  - PTFs available 14 September, 2015
  - Security enhancements include:
    - Password Encryption upgrade
    - Helpdesk support
    - Special character support
    - Minimum password change intervals

- Updates to the z/VM TLS/SSL Server – **APARs PI40702, VM65717, VM65718**
  - PTFs available 14 September, 2015
  - Maintains FIPS 140-2 and NIST SP 800-131a compliance
  - Function includes:
    - System SSL V2.1 Equivalency (V6.3 only)
    - AES Galois/Counter Mode Encryption (V6.3 only)
  - Changes to default cipher suites and protocols (all releases)

- More information at http://www.vm.ibm.com/security/

RACF Password Encryption Upgrade

- Enables of a stronger encryption mechanism of passwords and/or passphrases in a RACF database
  - *Matches support delivered by z/OS APAR OA43999*
  - *Strengthen RACF database against offline attacks*
  - Mitigate compliance issues of older encryption algorithms

- Migration to KDFAES is for an entire RACF database
  - May cause problems if sharing this RACF database with another system!
  - Utilities available to convert databases and clean password histories

- Some restrictions may apply
  - Support is for z/VM 6.3 only
  - RACF template has been updated; run RACFCONV accordingly
  - CPACF (Feature 3863) must be enabled

SMT Prorated Core Time Support

- APAR VM65680 available for z/VM 6.3 on September 2, 2015

- Applies only to z/VM systems where SMT has been enabled

- This support enforces capacity limits using core time rather than thread time so that a CPU Pool will not be limited prematurely.

- Following interfaces have been updated:
  - Commands: DEFINE CPUPOOL, QUERY CPUPOOL, SET CPUPOOL , SET SHARE
  - Accounting records
  - Monitor records
  - Store hypervisor information (STHYI) instruction

- Also resolves SRN005 Abends in previous APAR VM 65613

- CPU Pooling and ILMT can now be used without the need to potentially adjust the pool values to be equivalent to non-SMT environment.

- Use QUERY CPUPOOL to determine if APAR is applied (shows core instead of cpu)

February & March 2016 Updates

Dynamically Migrate the SSI PDR Volume

- Enhancement to be able to relocate the Single System Image (SSI) Persistent Data Record (PDR) volume without a planned outage

- Avoid the need for a cluster-wide shutdown in order to move PDR volume to a new device or new storage server
  - Facilitates moving to a new storage server
  - Does not addressed unplanned outage of the PDR volume

- New option on the CP **SET SSI** command

- VM65712 – Closed February 5, 2016, PTF UM34736
  - z/VM 6.3 Only

SIMD Guest Exploitation Support

- Allows guests with appropriate support to use the Vector Facility for z/Architecture, aka Single-Instruction Mulitple-Data (SIMD) functions introduced with the z13 servers.
  - Access to 32, 128-bit registers.
  - Potential performance improvements for exploiting software

- Supported by Live Guest Relocation

- VM65733 – Closed February 19, 2016, PTF UM34752
  - z/VM 6.3 Only
  - Requires z13 or z13s and guest with support

A Different Better vs. A Standard Good
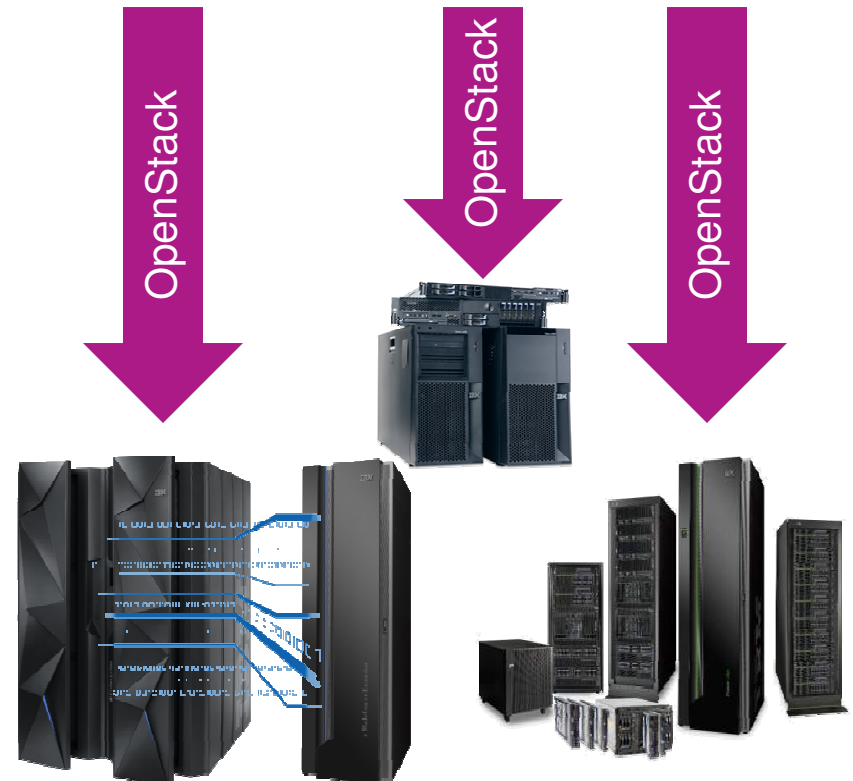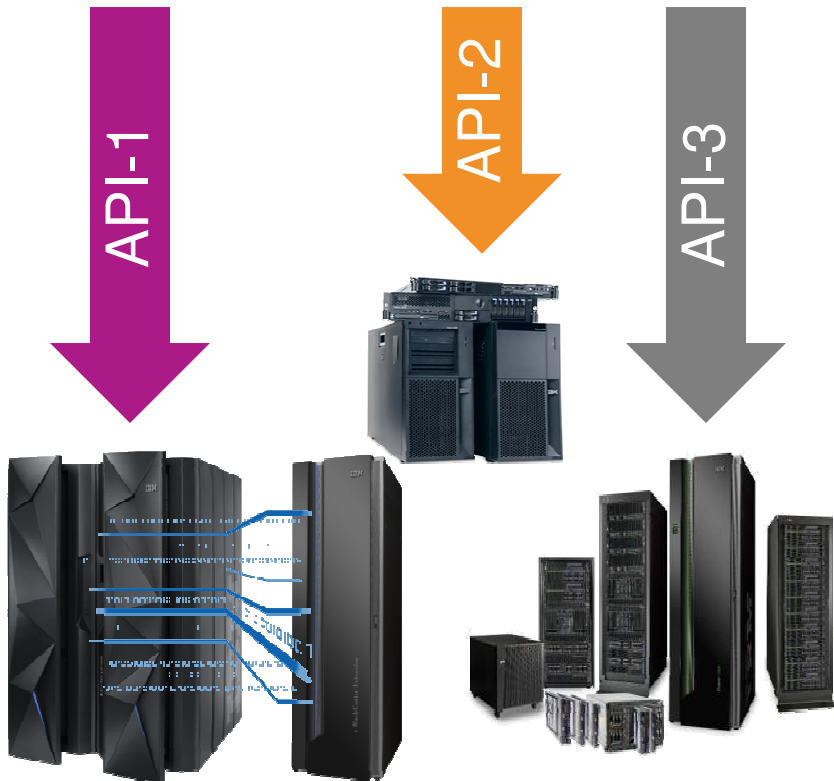
The OpenStack Food Chain

IBM Cloud Technology Products

openstack
CLOUD SOFTWARE

Communicates with z/VM

xCAT Appliance

SMAPI Services

Directory Product

IBM
zEnterprise

- **Top Half of the Solution:**
  - An IBM Cloud Technology product or other vendor product will include the OpenStack support.
  - Portions of that OpenStack support will know z/VM (i.e. code that connects and understands how to talk to z/VM).

- **Bottom Half of the Solution:**
  - Rest APIs are used to communicate with the OpenStack code from the top half.
  - The xCAT (Extreme Cloud Administration Toolkit) Appliance utilizes new and existing Systems Management APIs (SMAPI) to interact with the z/VM system
  - SMAPI can interact with additional products or features (e.g. a directory manager).

Product with OpenStack Support

z/VM 6.3 Product

Additional Product or Feature

OpenStack Enablement Changes

- Support of Liberty Release of OpenStack

- Integration of xCAT function into the z/VM Cloud Manager Appliance (CMA)
  - Running a fully functional z/VM OpenStack solution in a single virtual machine per z/VM system
  - z/VM Cloud Manager Appliance 1.2.0

- Support for provisioning Red Hat RHEL 7 and SUSE SLES 12 servers

- Uses the following:
  - OpenStack authentication (Keystone) V3 API.
  - RabbitMQ as the OpenStack message queue implementation.
  - OpenStack Dashboard (Horizon) as the self service portal.

- Available March 24, 2016 via APAR VM65780

z/VM Cloud Management Appliance (CMA)
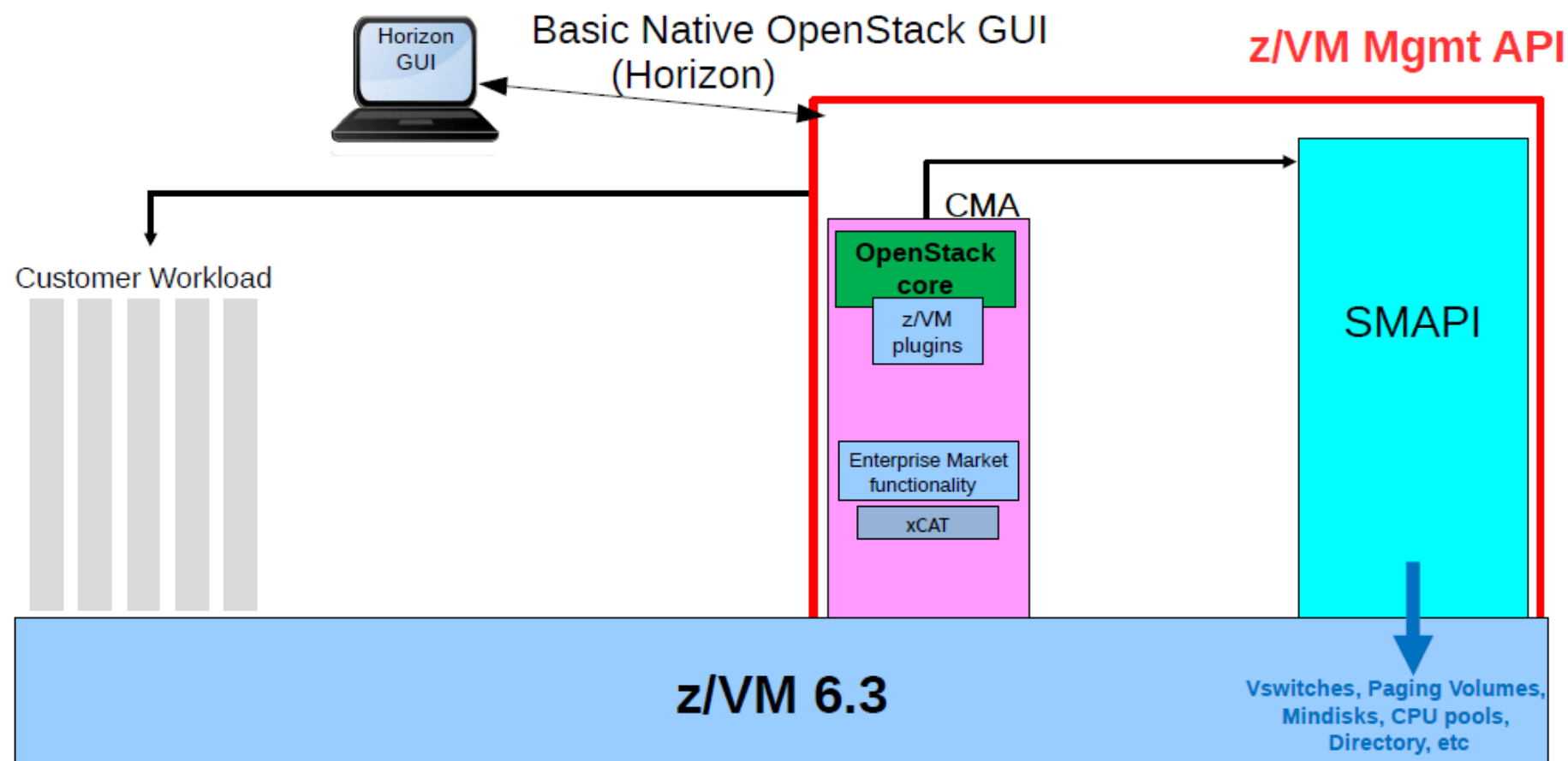
February 16, 2016 Announcement

Preview IBM z/VM 6.4

- Preview announcement 216-009, dated February 16, 2016
  – http://www.vm.ibm.com/zvm640/index.html

- Planned availability date Fourth Quarter 2016

- A release born from customer feedback

- Key components:
  – Enhanced technology for improved scaling and total cost of ownership
  – Increased system programmer and management capabilities

- New Architecture Level Set (ALS) of z196 and higher

Improved Scalability and TCO

- z/VM Paging enhancements
  - Use of HyperPAV when available to increase bandwidth for paging
  - Increases number of paging I/Os that can be in-flight at once
  - Exploitation for Paging, Spooling, z/VM user directory, and minidisk pools that are mapped to z/VM data spaces.

- Guest large page support
  - Enhanced DAT facility for guest use
  - 1 MB pages
  - Decreases memory needed for DAT structures by guest with Enhanced DAT support
  - z/VM maps to 4KB pages at the host level.

- Guest Transactional Execution support
  - Potential efficiency and scaling improvements for guests and guest software that exploits
  - Alternative for serializing a set of operations.

Improved Scalability and TCO

- Memory scalability improvements
  - Enhanced algorithms to further improve the efficiency of memory management
  - Provide a foundation for future enhancements in scaling and efficiceny

- FlashSystems support for FCP-attached SCSI disks.
  - Removes requirement of a San Volume Controller (SVC) to use FlashSystems for z/VM system volumes and EDEVs

System Programmer & Management Capability

- QUERY SHUTDOWN command
  - Allows better understanding of state of the system
  - Allows for increased programmatical management of the system

- CP environment variables
  - New framework to allow information to be set and queried for automatic processing
  - Example: Indicate system is being started for Production or DR Test or Actual DR

- New management queries for SCSI environment.
  - Allows SCSI detailed information to be gathered for emulated devices (EDEVs)

System Programmer & Management Capability

- CMS Pipelines enhancements
  - Pipelines is a powerful programming construct available in the CMS environment
  - Objective is to make available, with the product, many of the advances made to Pipelines since it was last updated in the product
  - Allows use of various tools and programming without the need to download additional code

- DirMaint to RACF Connector
  - Modernizes the Connector with a collection of functional enhancements
  - Brings processing in line with modern z/VM practices
  - Allows better passing of directory information to RACF
  - Facilitates proper security policy in environment managed by IBM Wave for z/VM or OpenStack

System Programmer & Management Capability

- Upgrade In Place migration enhancements

  - Upgrade In Place migration was introduced in z/VM 6.3

  - Enhanced to allow migration to z/VM 6.4 from
    - z/VM 6.2 or z/VM 6.3 (but not both at same time in cluster)
    - Supports migration for clustered or non-clustered systems

z/VM 6.4 Supported Hardware

- Following z Systems servers:
  - z13
  - z13s
  - LinuxONE Emperor
  - LinuxONE Rockhopper
  - IBM zEnterprise EC12
  - IBM zEnterprise BC12
  - IBM zEnterprise 196
  - IBM zEnterprise 114

- Electronic and DVD install
  - No tapes

Middleware Highlights

WebSphere Application Server Liberty Profile

For rapid development and light-weight production deployment with high scalability

- Web Profile certified (Liberty Core Edition)

- Small footprint (< **54MB**), quick startup (< **3 sec**)

- Developer-first design of simple, shareable XML configuration

- Dynamic runtime and configuration

- Unzip install and deploy

- Fidelity to WebSphere Application Server Full Profile

- Monitoring and mgmt through Admin Center or scripting

- Install new features from repository with no server restart

- Lightweight collective management scales to 10,000 servers

- Friendlier to a virtualized environment.

August 17, 2015 Announcements

IBM LinuxONE Portfolio ™
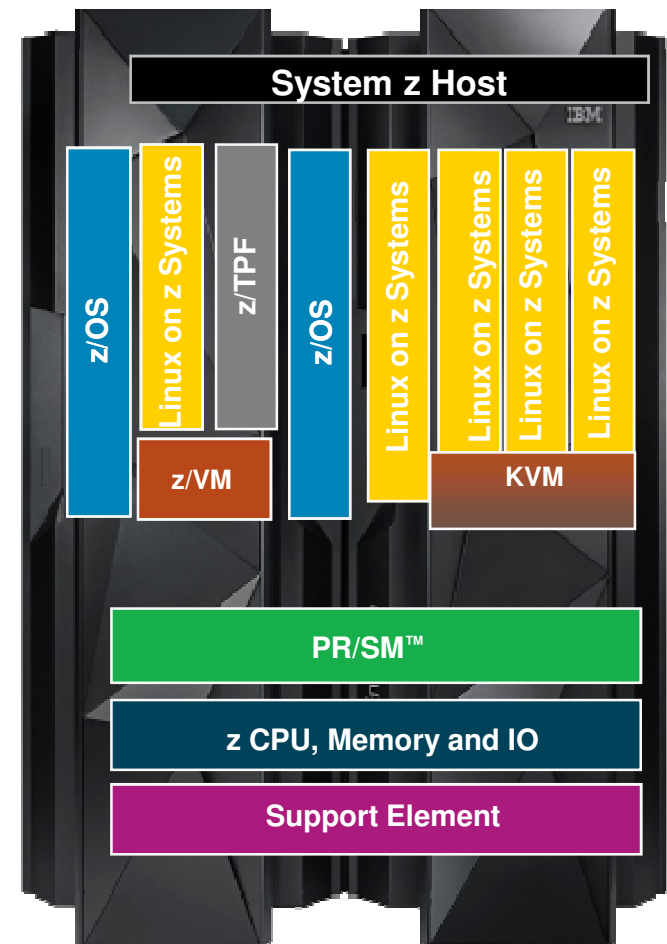
Linux Your Way

Linux without Limits

Linux without Risk

KVM for IBM z Systems™

- Announced: August 17, 2015

- General Availability: September 18, 2015

- An additional hypervisor for z Systems, joining z/VM & PR/SM

- Meant for customers with rich KVM skills on other platforms who want to leverage z Systems hardware and infrastructure.

- Not meant as a replacement for z/VM

KVM for IBM z Systems Requirements

KVM for IBM z v1.1.0 Pre-reqs	
Servers	IBM z13™ IBM zEnterprise® zEC12 IBM zEnterprise zBC12 IBM LinuxONE Rockhopper™ IBM LinuxONE Emperor™
Operating systems supported	SUSE Linux Enterprise Server (SLES 12 SP1)
Networking features supported	IBM OSA-Express5S IBM OSA-Express4S IBM OSA-Express3S (zEC12 and zBC12 only) -- OSA-Express3S in KVM for IBM z does not support using VLANs or flat networks together with OpenvSwitch

Storage devices are supported	ECKD™ DASD:
	   ▪ DS8000® (FICON®-attached)
	FCP SCSI disks:
	   ▪ XIV®
	   ▪ Storwize® V7000
	   ▪ FlashSystems™
	   ▪ SAN Volume Controller
	   ▪ DS8000 (FCP-attached)

Note: Refer to the KVM for IBM z Systems: Planning and Installation Guide (SC27-8236) for the most current information.

IBM LinuxONE Portfolio ™

IBM LinuxONE Rockhopper ™   IBM LinuxONE Emperor™

The most trusted, efficient and performant Linux platform for emerging high volume business critical applications

	IBM LinuxONE Rockhopper ™	IBM LinuxONE Emperor™
Positioning	*An **entry point** into the LinuxONE Systems family offering all the same great capabilities, innovation and value of LinuxONE with the flexibility of a smaller package*	*The **ultimate** flexibility, scalability, performance and trust for business critical Linux applications. With a **huge capacity** range you can grow with virtually limitless scale to handles the most demanding workloads in a single system*
Hardware	Better environmentals, lowest cost, less memory, fewer cores, no SMT	SMT, large memory, more cores, more I/O, enhanced crypto performance, more Logical Partitions, zAware
Cores	2 to 13 LinuxONE Cores	6 to 141 LinuxONE Cores Up to 10 TB
Hypervisor	z/VM and/or KVM for IBM z	

Linux Your Way – Pricing Your Way

- **Pay-for-Use IBM Hardware AND Software**
  - Fixed monthly payments
  - Costs can scale up or down based on usage
  - "Use" in terms of cores on machine, not usage in terms of core time consumed.

- **Per-Core Pricing Model on IBM Software stack for Linux**
  - Software licenses only on designated core
  - Order what you need, when you need it
  - Decrease licenses with 30 days notice
  - Cancellation with 30 days notice

- **Rental model with No up front payment required**
  - 36 month term: title remains with IBM
  - Choice at contract end – return, buy, replace
  - Right to return hardware after only 1 year

- Note: Clients must run ILMT to produce reports for SW compliance

Hardware Support

Support for IBM z13

- **Updates for z/VM 6.2 and 6.3**
  - http://www.vm.ibm.com/service/vmreqz13.html
  - Many components affected
  - Note: Directory space requirements increased slightly.

- No z/VM 5.4 Support

- No z/VM 6.1 Support even if you have extended support contract.

- **PSP Bucket**
  - Upgrade   **2964DEVICE**
  - Subset   **2964/ZVM**

- **If running Linux, please also check for required updates prior to migration.** **WARNING!**

Tested Linux Platforms

http://www.ibm.com/systems/z/os/linux/resources/testedplatforms.html

Distribution	LinuxONE Emperor LinuxONE Rockhopper z13s and z13	zEnterprise - zBC12 and zEC12	zEnterprise - z114 and z196	System z10 and System z9
RHEL 7	✔ (1)	✔ (3)	✔ (3)	✗
RHEL 6	✔ (1)	✔ (4)	✔	✔
RHEL 5	✔ (1)	✔ (5)	✔	✔
RHEL 4 (*)	✗	✗	✔ (8)	✔
SLES 12	✔ (2)	✔	✔	✗
SLES 11	✔ (2)	✔ (6)	✔	✔
SLES 10 (*)	✗	✔ (7)	✔	✔
SLES 9 (*)	✗	✗	✔ (9)	✔

z/VM z13 GA2 / z13s Support

- z/VM 6.2 and 6.3 will be supported on z13 (Driver 27) and z13s. No specific PTFs beyond GA1 support are required for z/VM 6.2

- GA1 PTFs are required to be applied before applying GA2 Support

- z/VM support includes:
  - Support for the Mid-Range machine (6.2 and 6.3)
  - Guest Exploitation Support for GA2 Crypto (6.2 and 6.3)
  - Guest Exploitation for the GA1 Vector Facility (SIMD) (Improved performance for analytics) (6.3)
  - LPAR Group Absolute Capacity Capping
    - Set an absolute capacity cap by CPU type on a group of LPARs
    - Allows each of the partitions to consume capacity up to their individual limits as long as the group's aggregate consumption does not exceed the group absolute capacity limit
    - Includes updated SysEvent QVS support (used by vendors who implement SW pricing)
  - Guest Exploitation of SMC-D; Virtual PCI Devices type ISM (6.3)
  - Dynamic Memory Management (6.3)
    - Improves efficiencies when a dynamic memory upgrade uses only a portion of the reserved main storage for the partition; initialize and clear just the amount of storage requested

z/VM z13 GA2 / z13s Support

- z/VM 6.3 GA2 APARs (February 2016):
  - VM65716: CP Support
  - VM65698: Performance Toolkit support
  - VM65729: VMHCD support
  - VM64844: VMHCM support
  - VM65704: EREP support
  - VM65736: IOCP support
  - HLASM does not require an APAR for support

- Additional z/VM 6.3 GA1 Support
  - VM65680: SMT-2 Prorated Core Time Support (September 2015)
  - VM67533: Vector Facility (SIMD) support (March 2016)

- Updates will be available on the z/VM web page at announce:
  http:/www.ibm.com/vm

- Hardware PSP bucket z/VM subset should also be reviewed
  - Upgrade **2965DEVICE**
  - Subset **2965/ZVM**

Statements of Direction
July 23, 2013
January 14, 2015
February 16, 2015

- Subset of IBM Statements of General Direction that are most important to the z/VM environment. See announcement materials for additional statements.
- Subject to change or withdrawal without notice, representing IBM goals and objectives only.

Completed Statements of Direction

Statement of Direction	From Announce Letter
z/VM Support for Single Instruction Multiple Data (SIMD)	January 2015
Enhanced RACF® password encryption algorithm for z/VM	January 2015
KVM Offering for z Systems	January 2015
GDPS/PPRC Multiplatform Resiliency Capability	January 2015
Security Evaluation of z/VM 6.3	July 2013
FIPS 140-2 Validation of z/VM 6.3	July 2013
Support of 10 GbE RoCE Express Feature	July 2013
Support of zEDC Express Feature	July 2013
Stabilization of z/VM 5.4 Support	July 2013

- Requires support from hardware and/or guests operating systems as appropriate

- Refer to www.vm.ibm.com or www.vm.ibm.com/security for more information

Withdrawal of Support for Expanded Storage

July 23, 2013

z/VM 6.3 will be the last release to support expanded storage (XSTOR) as part of the paging configuration. With the enhanced memory management support added in z/VM V6.3, expanded storage is no longer recommended as part of the paging configuration. z/VM can run efficiently in a configuration using only central storage

- In z/VM 6.3, it is recommended to configure all processor memory as central storage.
  - Support remains to use expanded storage in z/VM 6.3, but is suggested for use only in special cases.

Look for update with z/VM 6.4.

Removal of Support for Expanded Storage
January 14, 2015

z/VM V6.3 is the last z/VM release that will support Expanded Storage (XSTORE) for either host or guest usage. The IBM z13 server family will be the last z Systems server to support Expanded Storage (XSTORE).

- The previous SoD spoke of removal of paging to expanded storage, but there is more.

- All z/VM support for expanded storage will be removed in future release
  - Attaching to guests
  - Minidisk Cache
  - Paging
  - etc.

- This SoD also goes on to speak to hardware support being removed as well, after the z13 server family.

Look for update with z/VM 6.4.

Removal of ESA/390 Architecture Mode
January 14, 2015

The IBM z13 will be the last z Systems server to support running an operating system in ESA/390 architecture mode; all future systems will only support operating systems running in z/Architecture mode. This applies to operating systems running native on PR/SM as well as operating systems running as second level guests. IBM operating systems that run in ESA/390 mode are either no longer in service or only currently available with extended service contracts, and they will not be usable on systems beyond IBM z13. However, all 24-bit and 31-bit problem-state application programs originally written to run on the ESA/390 architecture will be unaffected by this change.

- While a hardware statement, there are potentially changes required for z/VM.

- Note implication of older operating systems.

Stabilization of z/VM 6.2 Support
January 14, 2015

The IBM z13 server family is planned to be the last z Systems server supported by z/VM V6.2 and the last z systems server that will be supported where z/VM V6.2 is running as a guest (second level). This is in conjunction with the statement of direction that the IBM z13 server family will be the last to support ESA/390 architecture mode, which z/VM V6.2 requires. z/VM V6.2 will continue to be supported until December 31, 2016, as announced in Withdrawal Announcement 914-012, dated February 04, 2014.

- While z/VM 6.2 will be supported until the end of 2016, there will **not** be support for the next server family.

- Similar to the statement of direction with z/VM 5.4 not supported on z13.

Product Delivery of z/VM on DVD/Electronic Only
January 14, 2015

Product Delivery of z/VM on DVD/Electronic only: z/VM V6.3 will be the last release of z/VM that will be available on tape. Subsequent releases will be available on DVD or electronically.

- No more tapes for z/VM product delivery for future z/VM releases.

- Allows testing resources to be spent else where.

Look for update with z/VM 6.4.

Dynamically Managed Thread Activation Levels
February 16, 2016

Dynamically managed thread activation levels. IBM intends to provide support in a future z/VM deliverable that will allow clients to dynamically manage the number of activated threads per configured core that can be enabled for simultaneous multithreading (SMT) without requiring an IPL of the z/VM system.

- Effectively, switch between 1 or 2 threads per IFL core via command

Stabilization of z/VM Support for the z10 Server Family
February 16, 2016

Stabilization of z/VM support for the IBM System z10 server family.
z/VM V6.3 is the last z/VM release planned to support the IBM System z10 server family of servers. Either an IBM zEnterprise 196 (z196) or an IBM zEnterprise 114 (z114) is the required minimum level of server for z/VM V6.4.

- New Architecture Level Set (ALS) for z/VM 6.4 of z196 and higher.

Putting SMT to Work in a z/VM Environment

Version 4

Bill Bitner
z/VM Client Focus and Care

bitnerb@us.ibm.com

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

BladeCenter*	FICON*	OMEGAMON*	RACF*	System z9*	zSecure
DB2*	GDPS*	Performance Toolkit for VM	Storwize*	System z10*	z/VM*
DS6000*	HiperSockets	Power*	System Storage*	Tivoli*	z Systems*
DS8000*	HyperSwap	PowerVM	System x*	zEnterprise*	
ECKD	IBM z13*	PR/SM	System z*	z/OS*	

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.
Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.
ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.
Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and
Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
OpenStack is a trademark of OpenStack LLC. The OpenStack trademark policy is available on the OpenStack website.
TEALEAF is a registered trademark of Tealeaf, an IBM Company.
Windows Server and the Windows logo are trademarks of the Microsoft group of countries.
Worklight is a trademark or registered trademark of Worklight, an IBM Company.
UNIX is a registered trademark of The Open Group in the United States and other countries.

* Other product and service names might be trademarks of IBM or other companies.

Notes:

Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.

This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.

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

Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

This information provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g., zIIPs, zAAPs, and IFLs) ("SEs"). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the "Authorized Use Table for IBM Machines" provided at www.ibm.com/systems/support/machine_warranties/machine_code/aut.html ("AUT"). No other workload processing is authorized for execution on an SE. IBM offers SE at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

Agenda

- **Key Concepts**

- **Terminology and Basic Concepts**

- **Determining Application Performance**

- **Variability Factors**

- **Metrics**

- **Performance Measurement Results**

- **Summary**

Notice Regarding Specialty Engines (e.g., zIIPs, zAAPs and IFLs):

Any information contained in this document regarding Specialty Engines ("SEs") and SE eligible workloads provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g., zIIPs, zAAPs, and IFLs). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the "Authorized Use Table for IBM Machines" provided at
www.ibm.com/systems/support/machine_warranties/machine_code/aut.html ("AUT").

No other workload processing is authorized for execution on an SE.

IBM offers SEs at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

Acknowledgements

- **The following people contributed charts or information to this presentation:**
  - **Bill Bitner**
  - **John Franciscovich**
  - **Emily Hugenbruch**
  - **Damian Osisek**
  - **Dan Rosa**
  - **Xenia Tkatschow**
  - **Brian Wade**
  - **Charles Webb**
  - **Romney White**
  - **Don Wilton**

Terminology and Basic Concepts

Why Simultaneous Multithreading?

- Other architectures are already doing it.

- We're reaching the physical limits of the machine, we can't just keep making chips smaller and faster.

- We need now to look at ways to use the chip resources more efficiently.

Work per Virtual CPU-second

z9   z10   z196   zEC12   z13 non-SMT

Processor Frequency

6000
5000
4000
3000
2000
1000
0

System z
POWER Systems
x86

2002   2004   2006   2008   2010   2012
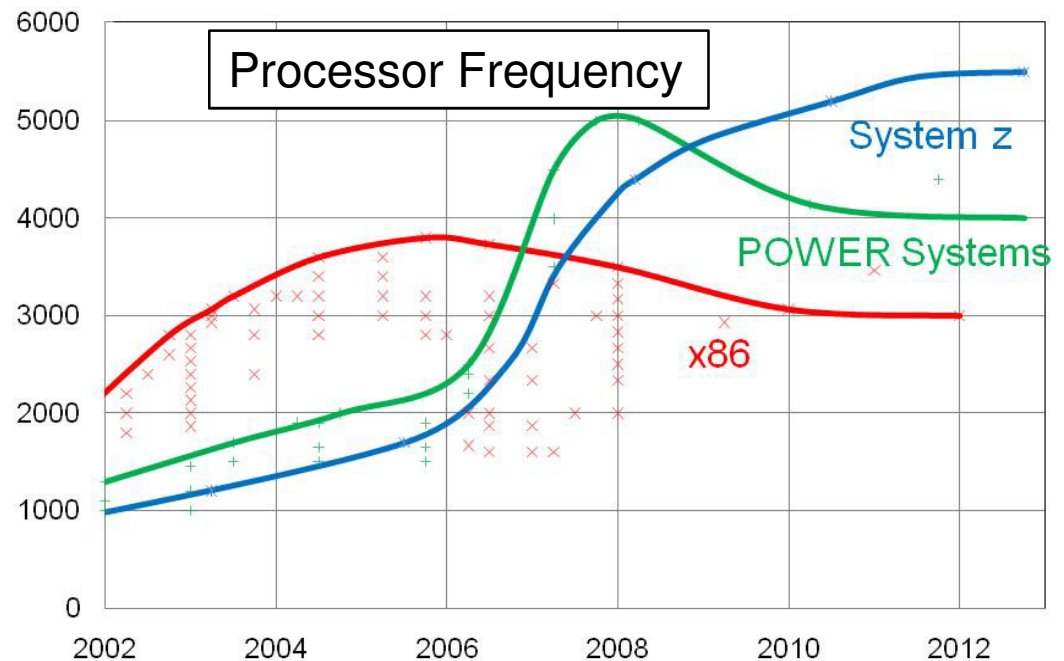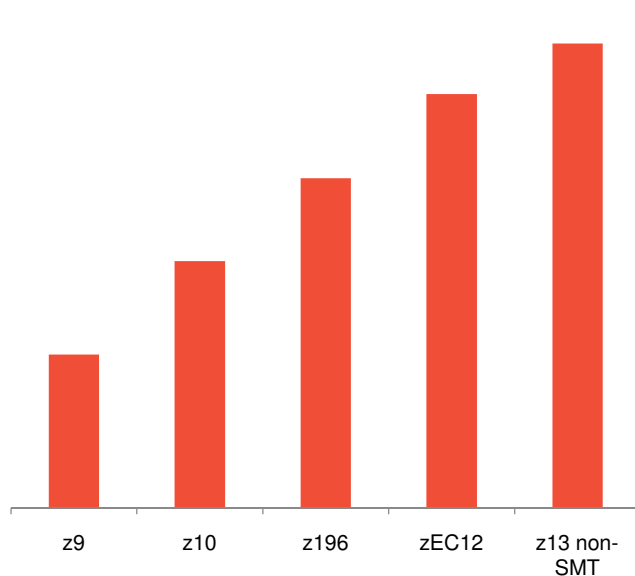
Not to Worry – The Future is Coming

QUBIT
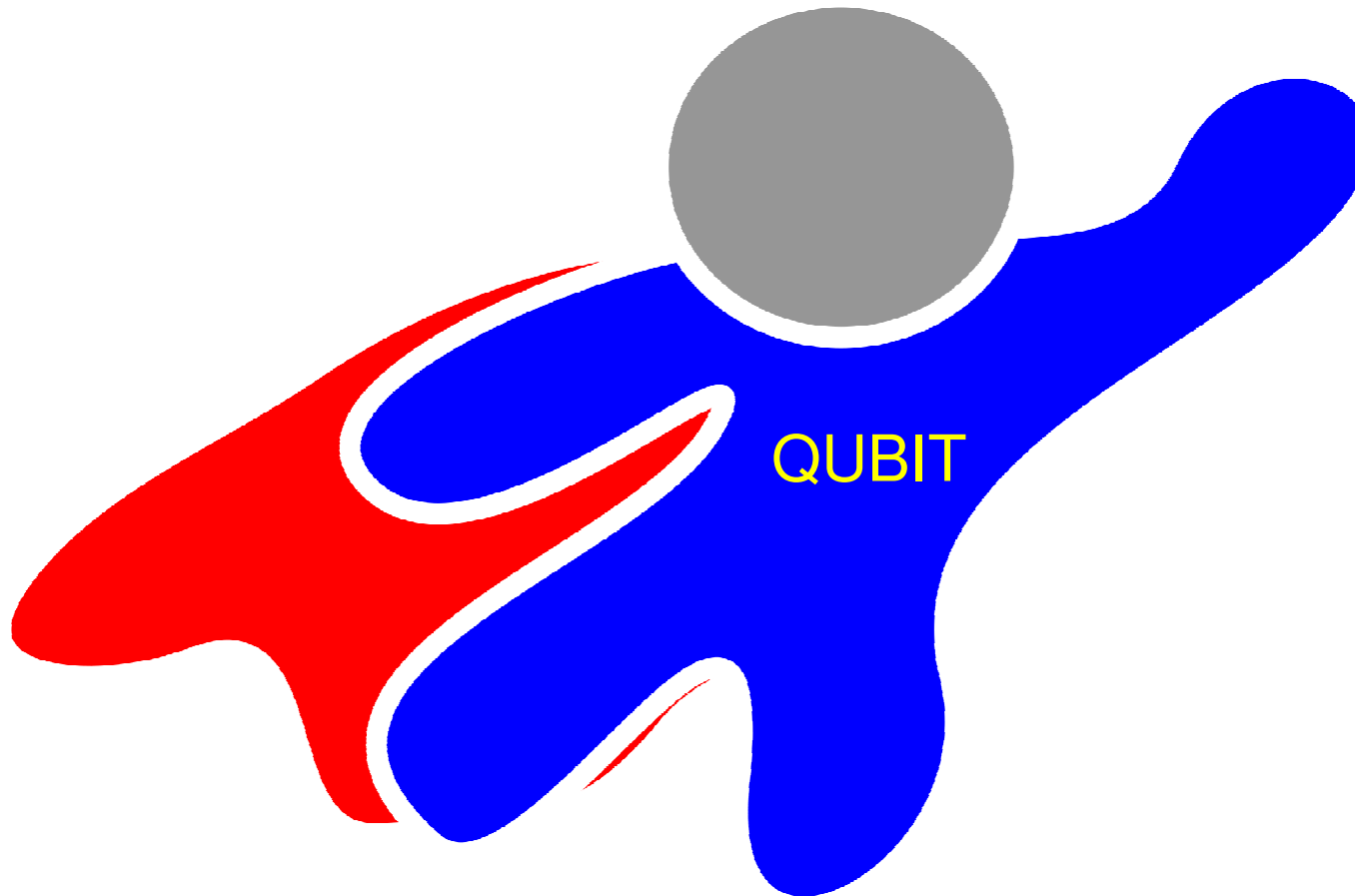
Quantum Computing – search ibm quantum computing cloud

HiperDispatch – Dispatching Affinity

- Processor cache structures become increasingly complex and critical to performance

- Goal is to re-dispatch work close (in terms of topology) to where it last ran

SMT Technology

- **Multiple programs (software threads) run on same processor core**

- **Active threads share core resources**
  - **In space: e.g, data and instruction caches, TLBs, branch history tables**
  - **In time: e.g., pipeline slots, execution units, address translator**

- **Increases overall throughput per core when SMT active**
  - **Amount of increase varies widely with workload**
  - **Each thread runs more slowly than a single-thread core**

Load/Store (L1 Cache)

instructions

Cache

Execution Units (FXU/FPU)

Use of Pipeline Stages in SMT2

A	Thread-A		Both threads
B	Thread-B		Stage idle

SMT in z/VM

- **Physical IFL Cores (you purchase these) with SMT allow up to two threads to be used**

- **Logical IFL Cores are presented to z/VM as in the past (you define these in the logical partition profile on the HMC)**

- **z/VM creates a CPU or logical processor associated with each thread (reflected in commands like QUERY PROCESSORS)**

- **The virtual CPUs of guests can then be dispatched on different threads intelligently, based on topology information**

Linux A		Linux B
vCPU 0	vCPU 1	vCPU 0

CPU 0 (Thread)	CPU 1 (Thread)	CPU 2 (Thread)	CPU 3 (Thread)
Logical IFL Core		Logical IFL Core	
Physical IFL Core		Physical IFL Core	
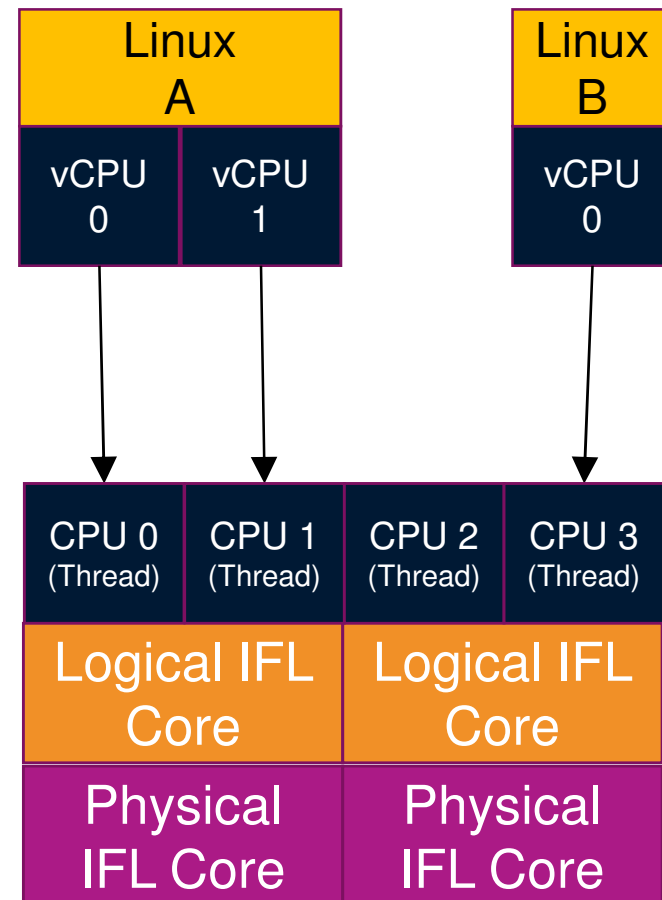
SMT in z/VM – Mixed Engine Environment

- **In a mixed-engine environment, general purpose processors are not enabled for threading, but a second CPU address is consumed (CPU 1 in this example) when SMT is enabled for the partition**

- **Virtual IFL CPUs are dispatched on the logical IFLs and virtual CP CPUs are dispatched on the logical CPs**

TPF A	Linux A		Linux B
vCPU 0 CP	vCPU 0 IFL	vCPU 1 IFL	vCPU 0 IFL

CPU 0 (Thread)	CPU 2 (Thread)	CPU 3 (Thread)	CPU 4 (Thread)	CPU 5 (Thread)
Logical CP Core	Logical IFL Core		Logical IFL Core	
Physical CP Core	Physical IFL Core		Physical IFL Core	
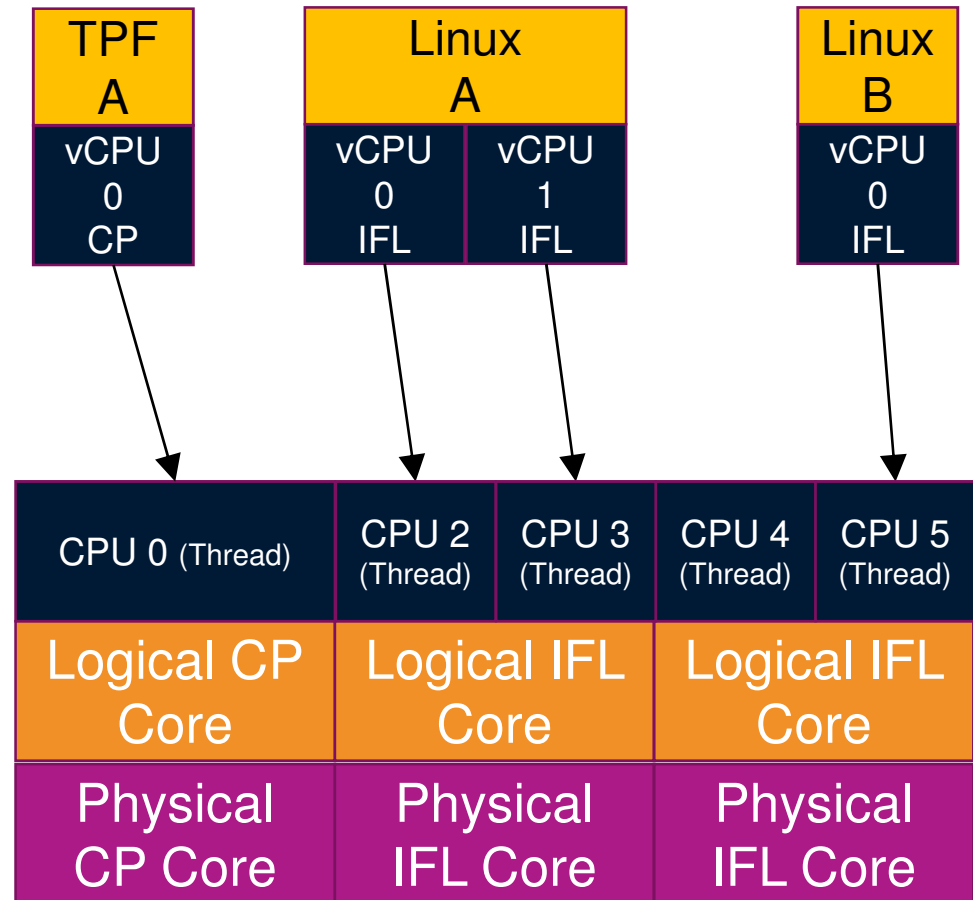
Determining Application Performance

Throughput

- Given that SMT's objective is to increase system throughput, it makes sense to measure it

- Applications can span virtual machines, z/VM systems, and even include components outside of z Systems

- An accurate view of throughput is available only from the application itself or externally

- Some poor alternatives from z/VM data can be used
  - Virtual I/O rate: assumes for a given transaction the virtual I/O will remain constant (could be true)
  - Network Input or Output rate: for workloads driven by network requests

- These measures can be for particular virtual machines or groups, depending on the structure of the application

Response Time

- Multiple threads that are slower than a single thread may potentially degrade response time

- An accurate view of response time is available only from the application itself or externally

- Unlike throughput, there are no alternatives from z/VM data that can map to response time, making it more important to have application or external data for this analysis

- z/VM State Sampling information can be helpful from a different perspective
  - z/VM does virtual CPU state sampling by default every two seconds
  - Captures 'state' of virtual CPU (e.g., Running, CPU wait, Page wait)
  - Delays due to waiting for (as opposed to using) a resource can be a major influence on response time
  - All systems wait at the same speed (wall clock)

Variability Factors

Work and Processor Time

- Old Myth: "Virtual CPU time should be constant for a given workload"
  - True on much older machines where
    - Resources were not shared with other instruction streams
    - More sophisticated optimizations (e.g., pipelining) were not used
  - True when competition or demand for shared resources was constant
  - True when fewer instruction streams were competing for those resources
  - True when memory was closer and uniform, in relative terms

- The variability has been there for a long time; it is just more noticeable today

- Key to recognize
  - Performance of a system, or a single part of a system, needs to be measured in that system to get an accurate picture
  - Some things vary that you can control
    - E.g., number of guests, number of logical partitions

Changes to z/VM when SMT is Enabled

- **Thread Affinity** – An effort is made to run the virtual CPU on the same thread as long as the virtual CPU remains on the core's dispatch vector (avoids TLB penalty)

- **Preemption Disabled** – To give the current CPU more time on the thread, benefit from build up of cache L1, L2 and TLB

- **Minor Time Slice Increased** – Allow virtual processor to benefit from population of L1/L2 cache and TLB

- **Time Slice Early** – If the virtual CPU loads a wait PSW, and certain conditions are true, CP ends the virtual CPU minor time slice early (helps assure virtual CPU not holding guest spin-lock at time slice end)

More details
http://www.vm.ibm.com/perf/reports/zvm/html/1q5smt.html

Metrics

Need to Rethink Metrics

CAPACITY ≠ UTILIZATION ≠ PRODUCTIVITY

Processor Time Reporting

- **Raw time** (the old way, but with new implications)
  - Amount of time each virtual CPU is run on a thread
  - This is the only kind of time measurement available when SMT is disabled
  - Used to compute dispatcher time slice and scheduler priority

- **MT-1 equivalent time** (new)
  - Used when SMT is enabled
  - Approximates what the raw time would have been if the virtual CPU had run on the core all by itself
    - Adjusted downward (decreased) from raw time
  - Intended to be used for chargeback

- **Pro-rated core time** (new with VM65680)
  - Used when SMT is enabled
  - "Discounts" raw time proportionally when core is shared between active threads
    - Full time charged while a virtual CPU runs alongside an idle thread
    - Half time charged while vCPU is dispatched beside another active thread
  - Suitable for core-based software license metrics

A Word About Metrics

- Estimate vs. Measure
  - Some metrics are actually measured by z/VM or firmware, others are estimated by either z/VM or the firmware

- Full wall clock vs. Core dispatched interval
  - Some metrics cover all of time, i.e. Wall Clock
  - Other metrics are based on the span of time in which z/VM's logical core was dispatched

What are all those numbers?

SMT Parameters	Definition	Range of Value
Core Busy	Percent of time at least one thread of the core was busy (aka core utilization)	0 – 100 %
Thread Density	average number of threads running on the core during the times the core was running at least one thread	1.00 – 2.00
Productivity	the ratio of the amount of work the core accomplished to the estimated amount of work the core would have accomplished if all of its busy time had been spent with all threads busy	0 – 100 %
MT Utilization	Percentage of estimated core capacity in use	0 – 100 %
Capacity Factor	core's work rate compared to its work rate when running with one thread busy (the "SMT benefit")	100 – 200% Sometimes CF < 100%; implies workload not MT - friendly
Maximum Capacity Factor	core's work rate when running with all threads busy compared to its work rate when running with one thread busy	100 – 200% Sometimes MCF < 100%; implies workload not MT- friendly

Measures: Core Busy and Thread Density

Two Threads on a Core

| Thread 0 | Idle | | | Busy | |

| Thread 1 | Busy | | Idle | Busy | Idle |

- Core Busy = $4 \div 5 = 80\%$

Wall Clock

Measure

- Thread Density = Average(1,1,2,1) = 1.25

Dispatch Interval

Measure

Measures: Productivity

Two Threads on a Core

Thread 0	**Idle**			**7**	**9**

Thread 1	**9**	**9**	**Idle**	**7**	**Idle**

Numbers indicate "work" completed metric (illustrative purposes here)

Compared to keeping both threads busy whenever core is in use

Thread 0	**7**	**7**	**Idle**	**7**	**7**

Thread 1	**7**	**7**	**Idle**	**7**	**7**

- Productivity = Estimate of "work" completed compared to potential work that could be completed if threads kept busy
- Productivity = $41 \div 56 = 73\%$

Measures: MT Utilization

Two Threads on a Core

Thread 0	**Idle**			7	9

Thread 1	9	9	**Idle**	7	**Idle**

Numbers indicate "work" completed metric (illustrative purposes here)

Compared to keeping both threads busy (TD 2) all the time (100% core busy)

Thread 0	7	7	7	7	7

Thread 1	7	7	7	7	7

- MT Utilization = Percentage of estimated core capacity in use
  - A view of how close the workload is to saturating the core
- MT Utilization = 41 ÷ 70 = 59%
  - Observation: MT Utilization ~= Core Productivity * Core Busy

Measures: Capacity Factor & Max Capacity Factor

Estimate

Dispatched Interval

Two Threads on a Core

| Thread 0 | Idle | | | 7 | 9 |

| Thread 1 | 9 | 9 | Idle | 7 | Idle |

- Capacity Factor = TotalWorkRate / SingleThreadWorkRate =
  - $(41 \div 4) \div (27 \div 3) = 10.25 \div 9 = 1.13 = 113\%$

- Max Capacity Factor: Projected upper bound
  - TwoThreadWorkRate / SingleThreadWorkRate =
  - $(14 \div 1) / (27 \div 3) = 14 \div 9 = 1.56 = 156\%$

Measures: Putting It All Together

Two Threads on a Core

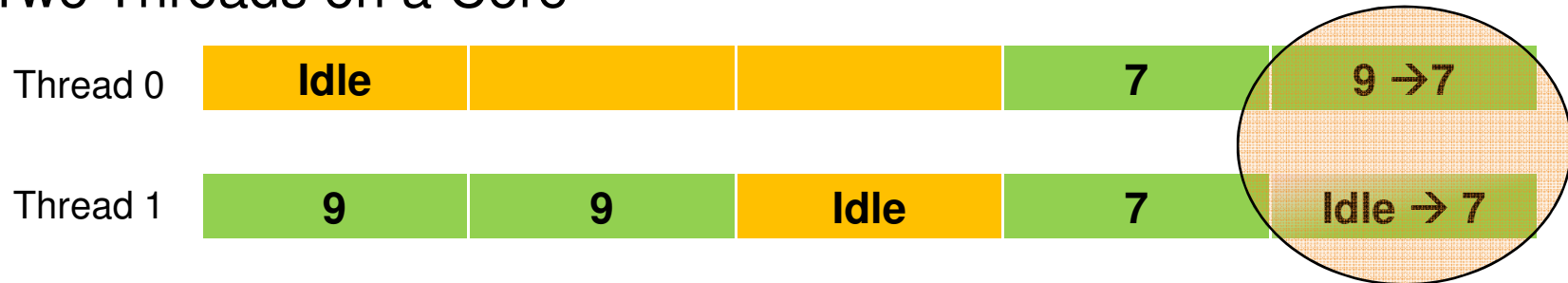| Thread 0 | Idle | | | 7 | 9 |

| Thread 1 | 9 | 9 | Idle | 7 | Idle |

Metrics	Example
Core Busy	80%
Thread Density	1.25
Productivity	73%
MT Utilization	59%
Capacity Factor	113%
Max Capacity Factor	156%

Relationship: As Thread Density increases

Two Threads on a Core

Start to use 2nd thread in last time segment

Thread 0	**Idle**			**7**	**9 →7**

Thread 1	**9**	**9**	**Idle**	**7**	**Idle → 7**

Metrics	Example	w/ Increased TD	
Core Busy	80%	80%	No effect
Thread Density	1.25	1.5	1.25 → 1.5
Productivity	73%	82%	Increases
MT Utilization	59%	66%	Increases
Capacity Factor	113%	127%	Increases
Max Capacity Factor	156%	156%	No effect

Relationship Two: More efficient threading

Two Threads on a Core

Go from 7 to 8 when both threads active

Thread 0	Idle			7 → 8	9
Thread 1	9	9	Idle	7 → 8	Idle

Metrics	Example	w/ more Efficient Threading	
Core Busy	80%	80%	No effect
Thread Density	1.25	1.25	No effect
Productivity	73%	67%	Decreases
MT Utilization	59%	54%	Decreases
Capacity Factor	113%	119%	Increases
Max Capacity Factor	156%	178%	Increases

Thread Density Effect on Capacity Factor

Maximum Capacity Factor

Capacity Factor

Other things being equal, as thread density increases, the Capacity Factor approaches the Maximum Capacity Factor.

1.0   1.1   1.2   1.3   1.4   1.5   1.6   1.7   1.8   1.9   2.0

Thread Density

Thread Density Effect on Capacity Factor

Maximum Capacity Factor

Capacity Factor

Other things being equal, as thread density increases, the Capacity Factor approaches the Maximum Capacity Factor.
(Damian's corollary: … but other things are rarely equal.)

Thread Density

1.0 1.1 1.2 1.3 1.4 1.5 1.6 1.7 1.8 1.9 2.0

New Metrics on INDICATE command & monitor

- **Indicate Load** will still show information by processor, which means by individual thread on multithreaded cores.
  - The percent-busy is thread-busy aka logical CPU-busy

- A new command, **INDicate MULTITHread (MT)** will show you the per type information, giving you an idea of how much capacity you have left for each type. The utilization shown is an average of the utilization of the cores of that type.

```
indicate multith
Multithreading is enabled.
Statistics from the interval 12:00:53 - 12:01:23
Core Type CP   Busy   8%   TD  1.00 of  1   Prod 100%   Util   8%
   CF  100%   MaxCF  100%
Core Type IFL  Busy   1%   TD  1.50 of  2   Prod  90%   Util   1%
   CF  113%   MaxCF  125%
Core Type ZIIP Busy   0%   TD  1.00 of  1   Prod 100%   Util   0%
   CF  100%   MaxCF  100%
Ready;
```

SMTMET Display Tool

- An EXEC that extracts MT metrics from Domain 0 Record 2.

- Available on z/VM download library:
  http://www.vm.ibm.com/download/packages/

- SMTMET Documentation: http://www.vm.ibm.com/perf/tips/smtmet.html

The process for reducing the SMTMET counters is the following:

1. Start with a MONWRITE file that contains D0 R2 records.

2. Command Syntax from CMS prompt:
   SMTMET *filename* MONDATA *filemode*

3. Resultant file from CMS prompt:
   filename $SMTMET *filemode*

SMTMET Resultant File Sample: Per-Core-type Report

```
D0R2 Per-Core-type Report for file: AMPDGLD1 MONDATA
```

Interval __Ended_	Core Type	___Secs___	Sampled ___Cores__	Pct Core Prodctvity	Pct Cap __Factor__	Pct Max _Cap Fct__	Pct MT Utilztion_	Average Thread Den
>>Mean>>	IFL	120.0	4.0	93.6	156.4	167.1	86.0	1.83
21:32:02	IFL	120.0	4.0	93.6	159.2	170.1	74.5	1.84
21:32:32	IFL	120.0	4.0	93.6	158.7	169.6	89.7	1.84
21:33:02	IFL	120.0	4.0	93.2	157.7	169.2	89.3	1.83
21:33:32	IFL	119.6	4.0	93.4	158.9	170.1	89.3	1.84
21:34:02	IFL	120.0	4.0	93.3	159.2	170.5	89.4	1.84
21:34:32	IFL	120.0	4.0	93.5	158.9	169.9	89.8	1.84
21:35:02	IFL	120.0	4.0	94.1	161.1	171.1	91.2	1.86
21:35:32	IFL	120.0	4.0	93.4	159.0	170.1	89.8	1.84
21:36:02	IFL	120.0	4.0	93.8	159.5	170.0	90.5	1.85
21:36:32	IFL	120.0	4.0	93.0	158.5	170.4	88.7	1.83
21:37:02	IFL	120.0	4.0	93.4	159.1	170.3	89.7	1.84
21:37:32	IFL	120.0	4.0	93.7	159.4	170.1	90.2	1.85
21:38:02	IFL	120.0	4.0	93.7	159.0	169.6	90.4	1.85

SMTMET Resultant File Sample: Per-Core Report

D0R2 Per-Core Report for file: AMPDGLD1 MONDATA

Interval __Ended_	Core _ID_	Core Type	___Secs___	Pct Core Prodctvity	Pct MT Utilztion_	Average Thread Den	Pct Core ___Busy___
>>Mean>>	00	IFL	30.0	93.6	86.0	1.83	92.06
>>Mean>>	01	IFL	30.0	93.5	86.0	1.83	91.92
>>Mean>>	02	IFL	30.0	93.7	86.3	1.83	92.20
>>Mean>>	03	IFL	30.0	93.6	85.9	1.84	91.86
21:32:02	00	IFL	30.0	93.4	74.0	1.84	79.26
21:32:02	01	IFL	30.0	93.1	74.4	1.83	79.91
21:32:02	02	IFL	30.0	93.8	74.2	1.85	79.11
21:32:02	03	IFL	30.0	93.8	75.4	1.85	80.39
21:32:32	00	IFL	30.0	94.1	91.2	1.86	96.86
21:32:32	01	IFL	30.0	93.3	88.8	1.84	95.16
21:32:32	02	IFL	30.0	92.7	88.3	1.82	95.28
21:32:32	03	IFL	30.0	94.0	90.7	1.86	96.42

SMTMET Resultant File Sample: Per-Core Report

D0R2 Per-Core Report for file: IDLESYS MONDATA

Interval __Ended_	Core _ID_	Core Type	___Secs___	Pct Core Prodctvity	Pct MT Utilztion_	Average Thread Den	Pct Core ___Busy___
>>Mean>>	00	IFL	30.0	71.7	0.1	1.20	0.05
>>Mean>>	01	IFL	30.0	78.8	0.1	1.39	0.11
>>Mean>>	02	IFL	30.0	0.0	0.0	1.38	0.01
>>Mean>>	03	IFL	30.0	0.0	0.0	1.40	0.01
13:52:42	00	IFL	30.0	....	....	1.24	0.02
13:52:42	01	IFL	30.0	80.8	0.1	1.55	0.12
13:52:42	02	IFL	29.9	....	....	1.40	0.01
13:52:42	03	IFL	30.0	....	....	1.40	0.01
13:53:12	00	IFL	30.0	....	....	1.25	0.02
13:53:12	01	IFL	30.0	80.3	0.1	1.55	0.12
13:53:12	02	IFL	30.1	....	....	1.39	0.01
13:53:12	03	IFL	30.0	....	....	1.40	0.01

- **At very high and very low core utilizations, z/VM may not be able to calculate Productivity and Utilization values**
- **Represented as '..........' in report and not included in the >>Mean>> calculation**

SMTMET Resultant File Sample: Per-Core Type

```
D0R2 Per-Core-type Report for file: IDLESYS MONDATA

Interval Core                Sampled   Pct Core   Pct Cap    Pct Max    Pct MT     Average
__Ended_ Type ___Secs___ ___Cores__ Prodctvity __Factor__ _Cap Fct__ Utilztion_ Thread Den
>>Mean>> IFL      90.0        3.0       76.7       124.1      172.1       0.0        1.36

13:51:42 IFL     119.8        4.0       80.6       137.3      175.7       0.0        1.49
13:52:12 IFL     120.1        4.0       80.9       136.1      173.4       0.0        1.49
13:52:42 IFL     120.0        4.0       80.8       135.8      173.7       0.0        1.49
13:53:12 IFL     120.1        4.0       80.3       137.2      177.3       0.0        1.50
13:53:42 IFL     120.0        4.0       82.1       132.1      164.1       0.0        1.48
13:54:12 IFL      60.0        2.0       81.2       134.0      171.9       0.1        1.49
13:54:42 IFL      60.0        2.0       70.0       106.5      167.1       0.1        1.18
13:55:12 IFL      60.0        2.0       59.2       103.2      200.0       0.1        1.12
13:55:42 IFL      60.0        2.0       86.0       101.0      118.2       0.1        1.09
13:56:12 IFL      60.0        2.0       66.0       118.0      200.0       0.1        1.26
```

When looking at this data, first look at MT Utilization; the cores are practically idle.

CPUMF Display Tool

- An exec that extracts z System CPU Measurement Facility records; internal performance experience; metrics as instructions completed, clock cycles used, and cache misses

- Available on z/VM download library:
  http://www.vm.ibm.com/download/packages/

- CPUMF Documentation

  http://www.vm.ibm.com/perf/tips/cpumf.html

The process for reducing the CPUMF counters is the following:
1. Start with a MONWRITE file that contains Domain 5 Record 13 records.
2. Command Syntax
   EXEC CPUMFINT *filename* MONDATA *filemode*
3. Resultant file:
   filename CPUMFINT *filemode*   ← *Interim file*
4. Command Syntax
   EXEC CPUMFLOG *filename* CPUMFINT *filemode*
5. Resultant file:
   filename $CPUMFLG *filemode* ← Final report file

Sample $CPUMFLG Output

IntEnd	LPU	Typ	___L1MP___	___L2P____	___L3P____	___L4LP__
>>Mean>>	0	IFL	2.05	87.22	12.71	0.0
>>Mean>>	1	IFL	2.01	87.27	12.66	0.0
>>Mean>>	2	IFL	2.02	87.13	12.80	0.0
>>Mean>>	3	IFL	2.04	87.06	12.86	0.0
>>Mean>>	4	IFL	2.01	87.25	12.68	0.0
>>Mean>>	5	IFL	2.01	87.21	12.72	0.0
>>MofM>>			2.02	87.19	12.74	0.0
>>AllP>>						
00:46:02	0	IFL	1.99	87.00	12.93	0.0
00:46:02	1	IFL	1.99	87.04	12.91	0.0
00:46:02	2	IFL	1.96	87.01	12.93	0.0
00:46:02	3	IFL	1.96	86.93	13.01	0.0
00:46:02	4	IFL	1.97	86.95	12.98	0.0
00:46:02	5	IFL	1.99	86.96	12.96	0.0

Memory Footprint within the Cache

- 2% of the instructions incur an L1 cache miss. (L1MP)

- 87% of the L1 misses are sourced from the L2 cache (L2P)

- 13% of the L1 misses are sourced from the L3 cache (L3P)

z/OS RMF Workload Activity Report (MT=2)

```
                    W O R K L O A D   A C T I V I T Y
z/OS V2R1 SYSPLEX PATPLX29 DATE 01/21/2015 INTERVAL 10.14.053
        RPT VERSION V2R1 RMF          TIME 21.46.55
REPORT BY: POLICY=PATPLEX   WORKLOAD=WASWKLD   SERVICE CLASS=WASTRANS
        RESOURCE GROUP=*NONE          CRITICAL=CPU
-TRANSACTIONS-    ...  SERVICE TIME   ---APPL %---
AVG        12.69      CPU 3545.743  CP      252.44
MPL        12.69      SRB    0.000  AAPCP     0.00
ENDED   5502452      RCT    0.000  IIPCP     0.43
END/S   8960.87      IIT    0.000
#SWAPS        0      HST    0.000  AAP        N/A
EXCTD         0      AAP      N/A  IIP     220.69
AVG ENC    12.69     IIP 1995.640
```
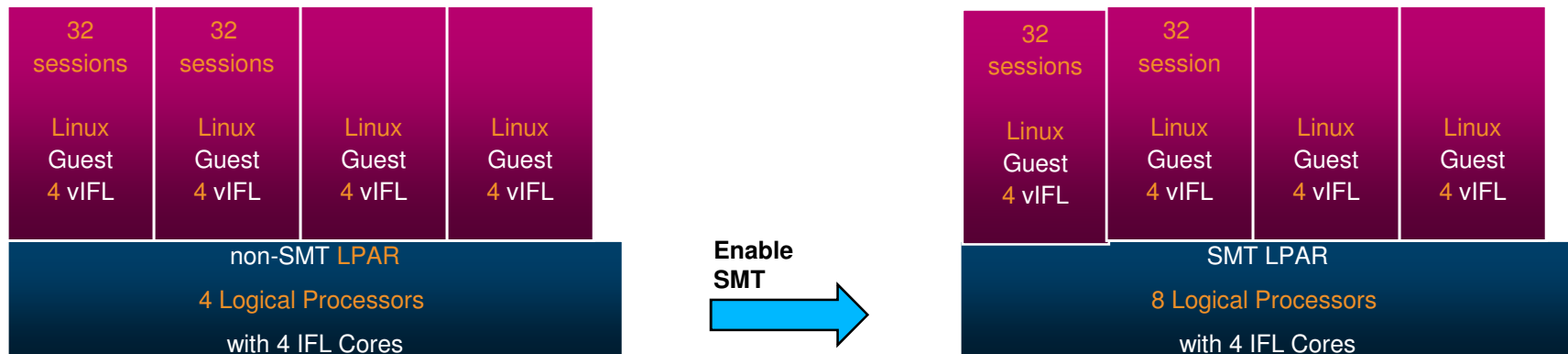
- **Service Times in MT=1 Equivalent Time units**
- **APPL % is % of core relative to its maximum Capacity Factor**
- **IIP APPL % = 1995.64 x 100 ÷ (614 x 1.473) = 220.69 %**
- **CP APPL % = (3545.743 - 1995.640) x 100 ÷ (614 * 1.0) = 252.44 %**

Performance Measurement Results

Performance Measurements

- **SMT2 Ideal Application**

- Maximum Storage Configuration

- Maximum Logical Processor Configuration

- **Linux-only mode with Single Processor serialization Application**

  - **Mitigation 1: Increasing virtual processors**

  - **Mitigation 2: Increasing servers in workload**

- Linux-only mode with Master Processor Serialization Application

- z/VM-mode with Master Processor Serialization Application

- CPU Pooling Workload

- **Live Guest Relocation (LGR) Workload**

- For a more details about performance results see:
  http://www.vm.ibm.com/perf/reports/zvm/html/1q5smt.html

SMT Ideal Application

Enable SMT →

Left box:

32 sessions	32 sessions		
Linux Guest 4 vIFL	Linux Guest 4 vIFL	Linux Guest 4 vIFL	Linux Guest 4 vIFL

non-SMT LPAR

4 Logical Processors

with 4 IFL Cores

Right box:

32 sessions	32 session		
Linux Guest 4 vIFL	Linux Guest 4 vIFL	Linux Guest 4 vIFL	Linux Guest 4 vIFL

SMT LPAR

8 Logical Processors

with 4 IFL Cores

SMT Ideal Application Configuration workload

- A Laboratory-designed Linux workload designed to show that under the right conditions, SMT provides significant benefit
- The workload consists of highly parallel activity:
  - 64 parallel sessions (work items)
  - 16 virtual processors
  - 4 cores in one LPAR
- Without SMT, the workload is processor constrained
- With SMT, total number of logical processors doubles, but they're slower

SMT Ideal Application Results

Multithreading	Disabled	Enabled
Logical Cores	4	4
Logical Processors	4	8
Virtual Processors	16	16
Application Sessions	64	64
Run ID	AMPDGLD0	AMPDGLD1
--------------------------	------------	------------
ETRR	1.000	1.362
ITRR	1.000	1.539
Application response time ratio	1.000	0.743
--------------------------	------------	------------
Total Util/Proc (p)	95.6	84.6
--------------------------	------------	------------
SMT Core Busy %	95.6	95.5
SMT Avg. Thread Density	na	1.83

- Throughput increased by 36% and Response Time decreased by 25%
- **Enabling SMT allowed us to push more work through the cores without increasing the number of cores.**

The Processor Log Screen for the nonSMT run:

```
FCX304  Run 2015/03/04 15:15:30          PRCLOG
                                         Processor Activity, by Time

From 2015/02/14 16:04:29
To   2015/02/14 16:14:59
For    630 Secs 00:10:30                  "This is a performance report f
_____

                                    <--- Percent Busy ----> <-- Rates
            C                  Pct
Interval    P                 Park                            Inst
End Time    U Type PPD Ent. DVID Time  Total  User  Syst  Emul  Siml
>>Mean>>    0 IFL  VhD  100 0000    0  95.7  95.5    .2  88.2 38153
>>Mean>>    1 IFL  VhD  100 0001    0  95.7  95.5    .2  88.2 37536
>>Mean>>    2 IFL  VhD  100 0002    0  95.6  95.4    .2  88.0 38178
>>Mean>>    3 IFL  VhD  100 0003    0  95.5  95.3    .2  87.8 38532
>>Total>   4 IFL  VhD  400   MIX    0 382.5 381.6    .9 352.1  152k
```

The Processor Log Screen for the SMT run:

```
FCX304  Run 2015/03/04 15:16:28        PRCLOG
                                       Processor Activity, by Time

From 2015/02/14 16:31:32
To   2015/02/14 16:42:02
For    630 Secs 00:10:30               "This is a performance repo
_____
```

| | | C | | | | Pct Park | <--- Percent Busy ----> | | | |
Interval End Time	P U	Type	PPD	Ent.	DVID	Time	Total	User	Syst	Emul
>>Mean>>	0	IFL	VhD	100	0000	0	84.7	84.5	.2	77.0
>>Mean>>	1	IFL	VhD	100	0000	0	84.3	84.1	.2	76.8
>>Mean>>	2	IFL	VhD	100	0001	0	84.5	84.4	.2	76.8
>>Mean>>	3	IFL	VhD	100	0001	0	84.6	84.4	.2	77.0
>>Mean>>	4	IFL	VhD	100	0002	0	84.5	84.3	.2	77.0
>>Mean>>	5	IFL	VhD	100	0002	0	84.9	84.7	.2	77.5
>>Mean>>	6	IFL	VhD	100	0003	0	84.8	84.6	.2	77.3
>>Mean>>	7	IFL	VhD	100	0003	0	84.7	84.5	.2	77.3
>>Total>	8	IFL	VhD	800	MIX	0	677.0	675.5	1.5	616.6

$SMTMET Resultant File

D0R2 Per-Core Report for file: AMPDGLD1 MONDATA

Interval __Ended_	Core _ID_	Core Type	___Secs___	Pct Core Prodctvity	Pct MT Utilztion_	Average Thread Den	Pct Core ___Busy___
>>Mean>>	00	IFL	30.0	93.6	86.0	1.83	92.06
>>Mean>>	01	IFL	30.0	93.5	86.0	1.83	91.92
>>Mean>>	02	IFL	30.0	93.7	86.3	1.83	92.20
>>Mean>>	03	IFL	30.0	93.6	85.9	1.84	91.86
21:32:02	00	IFL	30.0	93.4	74.0	1.84	79.26
21:32:02	01	IFL	30.0	93.1	74.4	1.83	79.91
21:32:02	02	IFL	30.0	93.8	74.2	1.85	79.11
21:32:02	03	IFL	30.0	93.8	75.4	1.85	80.39
21:32:32	00	IFL	30.0	94.1	91.2	1.86	96.86
21:32:32	01	IFL	30.0	93.3	88.8	1.84	95.16
21:32:32	02	IFL	30.0	92.7	88.3	1.82	95.28
21:32:32	03	IFL	30.0	94.0	90.7	1.86	96.42

SMT Single Processor Serialization Application with Mitigation Results

Run ID	APNDGLD0	APNDGLD1	APNDGLDF	APNDGLDD
Multithreading	Disabled	Enabled	Enabled	Enabled
Logical Cores	3	3	3	3
Logical Processors	3	6	6	6
Virtual Processors	15	15	18	18
ETRR	1.000	0.649	1.065	1.303
ITRR	1.000	1.018	1.101	1.306
Application Response Time Ratio	1.000	1.450	0.896	1.701
Linux guests	3	3	3	6
Virt Processors per guest	1	1	2	1
Total Util/Proc (p)	95.8	61.1	92.6	95.6
Linux Users Serialized Util	70.460	95.413	143.794	73.698
SMT Core Busy %	95.8	95.5	95.1	95.7
SMT Avg. Thread Density	na	1.28	1.95	1.99
SMT Productivity %	na	90	98	na
SMT Capacity Factor	na	104	149	na
SMT Max. Capacity Factor	na	115	151	na

Performance Measurements: Live Guest Relocation

25 Linux guests relocated while running three workloads

- PING – to simulate network traffic

- BLAST– to simulate I/O

- PFAULT- to simulate referencing storage

Relocation was done synchronously using the SYNC option of VMRELOCATE command
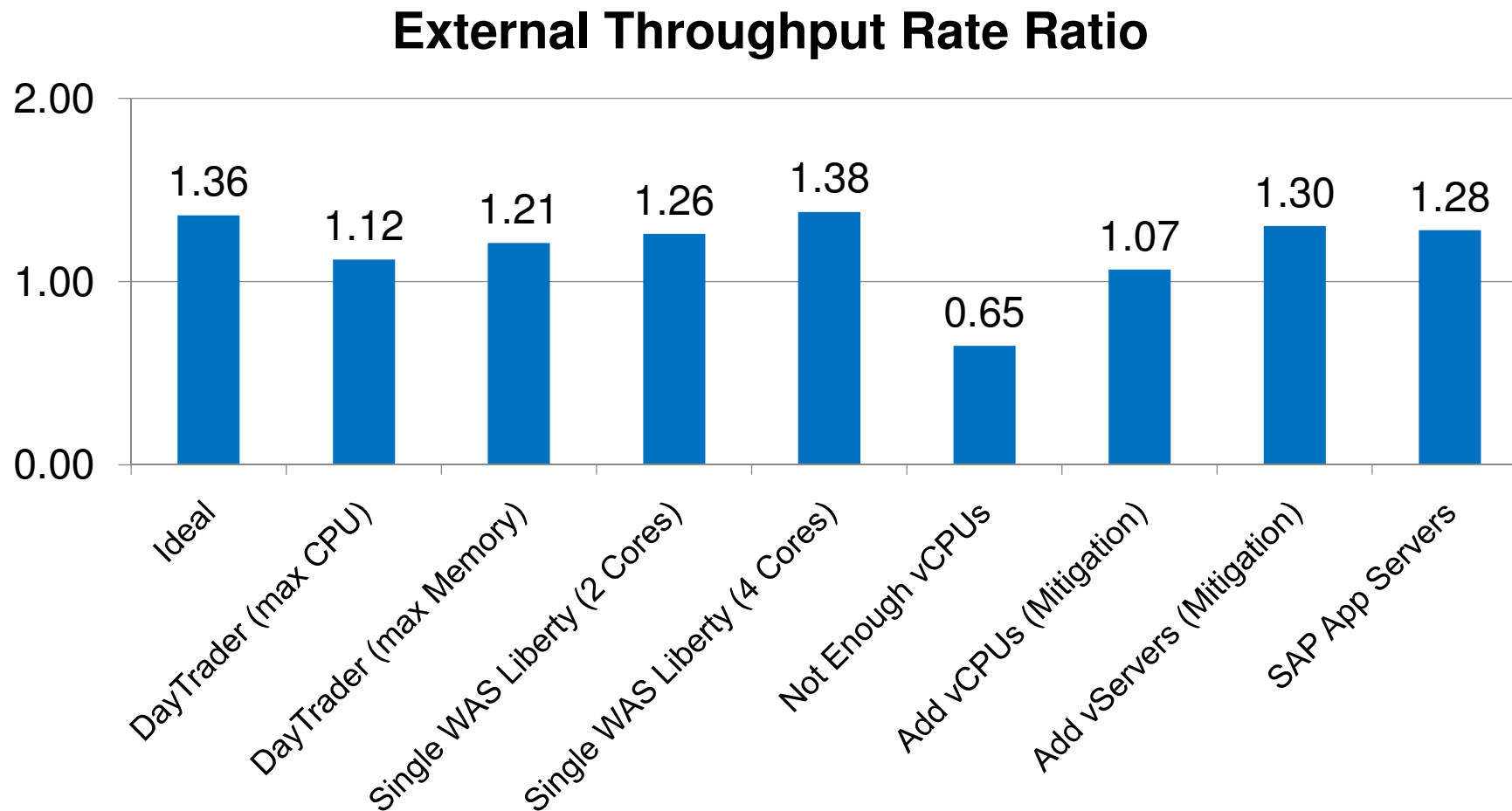
Results:

- Relocation time increased by 10%

- Quiesce time increased by 26%

- PFAULT (71%) and BLAST (34%) completions increased

- Total number of pages relocated during quiesce increased by 51%

Conclusion:

With SMT2, the BLAST and PFAULT workloads were changing pages more frequently, thus causing more pages to be moved during quiesce time.

Variation in Impact of SMT

External Throughput Rate Ratio

Base measurement is with SMT disabled.

Performance Measurements: Conclusion

- Results in measured workloads **varied widely**.

- Best results were observed for applications having highly parallel activity and no single point of serialization.

- No improvements were observed for applications having a single point of serialization.

- To overcome bottlenecks, workload adjustment should be done where possible.

- While very rare, workloads that have a heavy dependency on the z/VM master processor are not good candidates for SMT-2.

- The multithreading metrics (provided by the SMTMET tool) provide information about how well the cores perform when SMT is enabled. There is **no direct relationship with workload performance** (ETR, transaction response time)

- **Measuring workload throughput and response time is the best way to know whether SMT is providing value to the workload.**

Summary

Summary

- **SMT provides potential throughput improvements**
  - **Especially if workloads experiencing CPU Wait or cache misses**

- **Before moving to SMT**
  - **Gather throughput and response time data**
  - **Gather monitor and CPU MF data**
  - **Evaluate and adjust virtual machines (i.e., add virtual CPUs) that are approaching virtual CPU utilization limits (e.g., 85%) at peak times**

- **After moving to SMT**
  - **Compare throughput and response time data**
  - **Validate virtual CPU configurations**
  - **Evaluate the new SMT metrics**
  - **Assess both core resources and thread (logical processor) resources**

Backup Slides

Enhanced RACF Password Encryption Algorithm for z/VM
January 14, 2015

Enhanced RACF® password encryption algorithm for z/VM: In a future deliverable an enhanced RACF/VM password encryption algorithm is planned. This support will be designed to provide improved cryptographic strength using AES-based encryption in RACF/VM password algorithm processing. This planned design is intended to provide better protection for encrypted RACF password data in the event that a copy of RACF database becomes inadvertently accessible.

- z/OS support for this currently exists.

- Lack of this support in z/VM complicates sharing RACF databases with z/OS where the support is used.

Satisfied

GDPS/PPRC Multiplatform Resiliency Capability
January 14, 2015

In the first half of 2015, IBM intends to deliver a **GDPS/Peer to Peer Remote Copy (GDPS/PPRC) multiplatform resiliency capability** for customers who do not run the z/OS operating system in their environment. This solution is intended to provide IBM z Systems customers who run z/VM and their associated guests, for instance, Linux on z Systems, with similar high availability and disaster recovery benefits to those who run on z/OS. This solution will be applicable for any IBM z Systems announced after and including the zBC12 and zEC12.

- Lower the skill expense of running a GDPS environment, particularly for those customers with little, or no, z/OS background.

Satisfied

KVM offering for IBM z Systems

January 14, 2015

In addition to the continued investment in z/VM, IBM intends to support a Kernel-based Virtual Machine (KVM) offering for z Systems that will host Linux on z Systems guest virtual machines.

The KVM offering will be software that can be installed on z Systems processors like an operating system and can co-exist with z/VM virtualization environments, z/OS, Linux on z Systems, z/VSE, and z/TPF.

The KVM offering will be optimized for z Systems architecture and will provide standard Linux and KVM interfaces for operational control of the environment, as well as providing the required technical enablement for OpenStack for virtualization management, allowing enterprises to easily integrate Linux servers into their existing infrastructure and cloud offerings.

- An additional option for virtualization on z Systems.
- The IBM commitment to z/VM remains steadfast.

Satisfied

Security Evaluation of z/VM 6.3

July 23, 2013

IBM intends to evaluate z/VM V6.3 with the RACF Security Server feature, including labeled security, for conformance to the Operating System Protection Profile (OSPP) of the Common Criteria standard for IT security, ISO/IEC 15408, at Evaluation Assurance Level 4 (EAL4+).

- Evaluation is with inclusion of **RACF Security Server** and **Single System Image** priced features enabled.
  - Evaluated configuration supports clusters of *1-n* z/VM systems.
  - No claims made about standalone systems, or systems without RACF for VM.

- See http://www.vm.ibm.com/security/ for current z/VM Security information.

Satisfied

FIPS Certification of z/VM 6.3

July 23, 2013

IBM intends to pursue an evaluation of the Federal Information Processing Standard (FIPS) 140-2 using National Institute of Standards and Technology's (NIST) Cryptographic Module Validation Program (CMVP) for the System SSL implementation utilized by z/VM V6.3.

- Federal Information Protection Standard (FIPS) 140-2
  - Target z/VM 6.3 System SSL is FIPS 140-2 Validated*
  - Enablement requirements for certificate database and servers
    - http://csrc.nist.gov/groups/STM/cmvp/documents/140-1/1401val2012.htm#1735

- See http://www.vm.ibm.com/security/ for current z/VM Security information.

Satisfied

*A Certification Mark of NIST, which does not imply product endorsement by NIST, the U.S. or Canadian Governments.

Support of the 10GbE RoCE Express Feature
July 23, 2013

In a future z/VM deliverable IBM plans to offer support for guest exploitation of the 10GbE RoCE Express feature (#0411) on the IBM zEnterprise EC12 and IBM zEnterprise BC12 systems. This is to allow guests to utilize Remote Direct Memory Access over Converged Ethernet (RoCE) for optimized networking.

- RoCE is high bandwidth, low latency link layer protocol

- Guest support for devices dedicated to z/VM guests that support RoCE

- Requires 10GbE RoCE Express feature on either the IBM zEC12 or IBM zBC12

Satisfied

Support of the zEDC Express Feature

July 23, 2013

In a future z/VM deliverable IBM plans to offer z/VM support for guest exploitation of the IBM zEnterprise Data Compression (zEDC) Express feature (#0420) on the IBM zEnterprise EC12 and IBM zEnterprise BC12 systems.

- New data compression hardware feature to improve ability to do compression by offloading to zEDC

- Support is planned for guest usage

- Requires zEDC Express feature on either the IBM zEC12 or IBM zBC12

Satisfied

Stabilization of z/VM 5.4 Support
July 23, 2013

The IBM zEnterprise EC12 and IBM zEnterprise BC12 are planned to be the last System z servers supported by z/VM V5.4 and the last System z servers that will support z/VM V5.4 running as a guest (second level). z/VM V5.4 will continue to be supported until December 31, 2016, or until the IBM System z9® Enterprise Class (z9 EC) and IBM System z9 Business Class (z9BC) are withdrawn from support, whichever is later. Refer to Withdrawal Announcement 912-144, (RFA56762) dated August 7, 2012.

- While support will continue for z/VM 5.4, support for new function and processors is being stabilized.

- z/VM 5.4 will not be supported on processors after the zEC12 and zBC12.
  - This includes running as a guest of a supported z/VM Version 6 release.

- Plan now to avoid a migration which would involve both hard~~wa~~ ~~t~~ the same time.

Satisfied

z/VM Support for Single Instruction Multiple Data (SIMD)
January 14, 2015

In a future deliverable IBM intends to deliver support to enable z/VM guests to exploit the Vector Facility for z/Architecture (SIMD).

- The Single Instruction Multiple Data (SIMD) was introduced as part of the z13, allowing use of the new Vector Facility.

- The initial z/VM support for z13 does not contain the virtualization of SIMD, which would allow guests to exploit it and gain potential performance benefits.

Satisfied