Power Systems

Redbooks

IBM

International Technical Support Organization

POWER8 Technology and System Technical Deep Dive

WRZ13G

Alexandre Bicas Caldeira
abicas@br.ibm.com
ATS Power Systems
Brazil

Agenda

- POWER8 processor technology and architecture
- POWER8 processor-based systems design points and advanced features
- Cache architecture
- Memory infrastructure
- PCIe Gen3
- CAPI
- Power S822, S814, S824, S812L, S822L systems
- Speeds and feeds, form factors, feature options
- Performance information
- POWER8 Solutions overview
- Traditional Enterprise workloads
- Big Data/Analytics
- Cloud
- HMC version 8
- Requirements and system support
- New features in systems management and monitoring
- PowerVM new capabilities
- AIX exploitation of POWER8
- New and improved features
- IBM i exploitation of POWER8
- New and improved features
- POWER8 and Linux
- OpenPOWER Foundation
- Linux support - SuSE/redHat/Ubuntu
- PowerKVM

The Power Systems ITSO team

- Scott Vetter, ITSO Systems Team Leader
- Dino Quintero – ITSO Project Leader - BAO and Infrastructure solutions
- Chrys Rayns – ITSO Manager - Systems and Storage redbooks

The Power Systems/AIX Extended ITSO team

- Alexandre Caldeira - Consulting IT Specialist, ATS Power Systems, Latin America
- Andrei Socoliuc - IT Specialist, AIX/Power Systems, IBM Romania
- César Diniz Maciel - Executive IT Specialist, Power Systems, Global Techline, US
- Daniel Martin-Corben - Technical Solution Designer, IBM UK
- Marcos Quezada - Consulting IT Specialist, ATS Power Systems, Latin America
- Octavian Lascu - Complex IT Project Team Leader, IBM Romania
- Tony Steel - Senior IT Specialist, ATS Power Systems, IBM Australia

- ... and some others that contribute teaching workshops around the world.

Your instructor

- Alexandre Bicas Caldeira – abicas@br.ibm.com
- I work for ATS Power LA – We provide pre-sales consulting for IBM and Business Partners
- Background info
  - Worked 10 years as a business partner with IBM solutions
  - Started with AIX 4.3.2, RS/6000 Model F50
  - Joined IBM in 2010 – Power System Pre-sales
- BIG ITSO fan!! Workshops and residencies
- 3 residencies so far
  - 6 redbooks
  - 1 workshop
- Started with Power Virtualization back in 2004 (POWER4 – Regatta)

Power Systems

POWER8 is ready for the next generation of enterprise applications

Designed for Big Data

First server processor generation optimized for big data & analytics with POWER8 innovative design
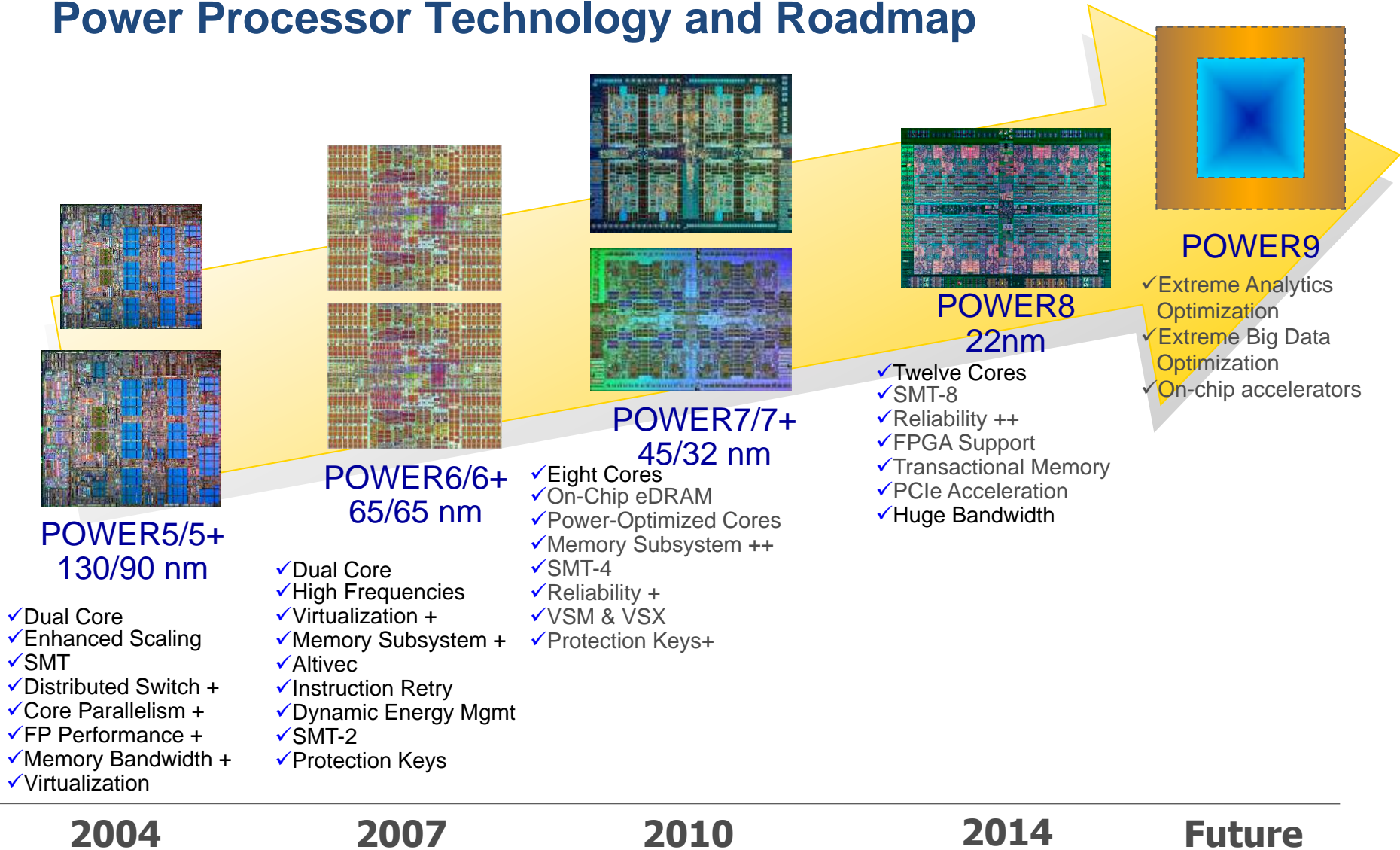
Superior cloud economics

Superior cloud price/ performance advantages & security to move data-centric applications to the cloud

Open Innovation Platform

Delivering the world's first open server ecosystem revolutionizing the way IT is developed & delivered

'IBM Introduces the first processor designed for big data'

Power Processor Technology and Roadmap

POWER9

✓Extreme Analytics
Optimization
✓Extreme Big Data
Optimization
✓On-chip accelerators

POWER8
22nm

✓Twelve Cores
✓SMT-8
✓Reliability ++
✓FPGA Support
✓Transactional Memory
✓PCIe Acceleration
✓Huge Bandwidth

POWER7/7+
45/32 nm

✓Eight Cores
✓On-Chip eDRAM
✓Power-Optimized Cores
✓Memory Subsystem ++
✓SMT-4
✓Reliability +
✓VSM & VSX
✓Protection Keys+

POWER6/6+
65/65 nm

✓Dual Core
✓High Frequencies
✓Virtualization +
✓Memory Subsystem +
✓Altivec
✓Instruction Retry
✓Dynamic Energy Mgmt
✓SMT-2
✓Protection Keys

POWER5/5+
130/90 nm

✓Dual Core
✓Enhanced Scaling
✓SMT
✓Distributed Switch +
✓Core Parallelism +
✓FP Performance +
✓Memory Bandwidth +
✓Virtualization

2004	2007	2010	2014	Future

POWER8 Highlights

Announced 2013 at Hot Chips Conference

12 Cores per Socket/Chip

Significantly Strengthened Cores

> **8 threads per core (SMT8)**
> **Wider fetch / dispatch/issue of instructions**
>   **(8 fetch / dispatch, 10 issue)**
> **Doubled highly utilized execution units**

Larger Caches per core/region

> **L1: 64K data Cache, 32K instruction**
>   **Cache   per core**
> **L2: 512K private   per core**
> **L3: 8M per core (up to 96MB per chip)**
> **L4: external to the chip**

2 Integrated Memory Controllers w/ Improved
 Latency & BW

> **~ 25% memory latency improvement via**
> **on-chip fastpath interconnect**
>   **16MB mem cache / buffer chip**

Integrated SMP Interconnect w/

 improved "Flatness"
>   **2-Hop fabric topology**

Integrated IO Subsystem
>   **On Chip PCIe Controller**

 Fine Grained Power Management
>   **On Chip Power Management Controller & Power gating**
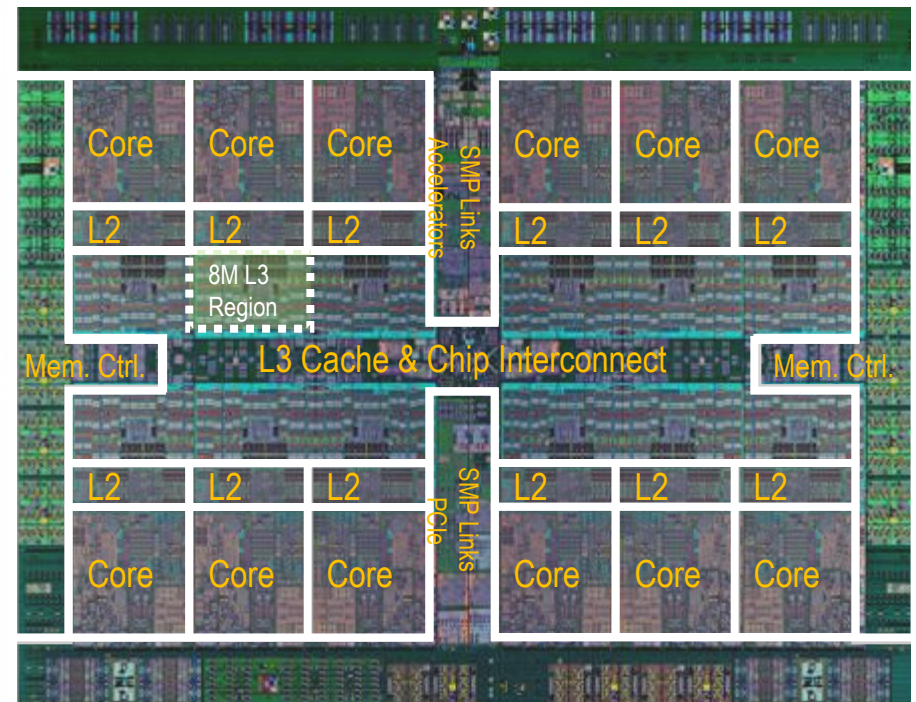
Scale-out Systems - DCMs and POWER8 Chips

1S & 2S servers use DCM (Dual Chip Module)

- 1 DCM fills 1 socket
- 1 DCM has two Scale-out POWER8 chips
- 1 DCM can provide 6-core, 8-core, 10-core or 12-core sockets

6-core Processor Chip
   two 3-cores for 6-core DCM
   two 4-cores for 8-core DCM
   two 5-cores for 10-core DCM
   **two 6-cores for 12-core DCM**
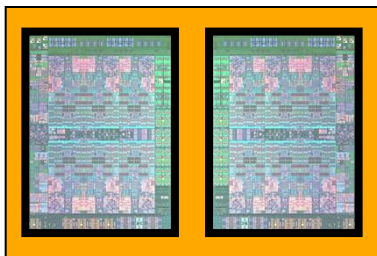
**Max 8 Threads per Core**
Excellent I/O bandwidth per socket
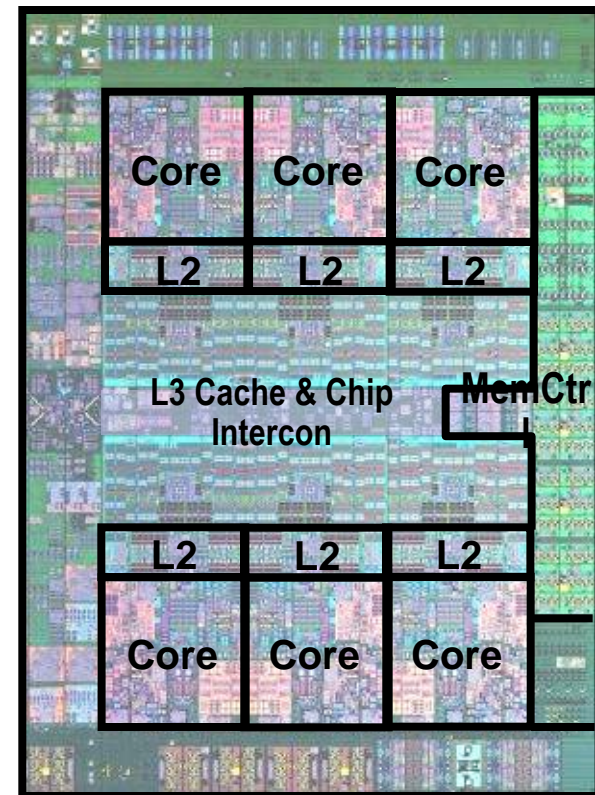
POWER8 core

- Execution Improvement (vs POWER7)
    - SMT4 -> SMT8
    - 16 execution units:
        - 2 FXU, 2 LSU, 2 LU, 4 FPU, 2 VMX, 1 Crypto, 1 DFU, 1 CR, 1 BR

- Larger Caching Structures (vs POWER7)
    - 2x L1 Data Cache (64 KB)
    - 2x Outstanding Data Cache misses
    - 4x Translation Cache
    - 4x Table walk machines
- Per core power/thermal management
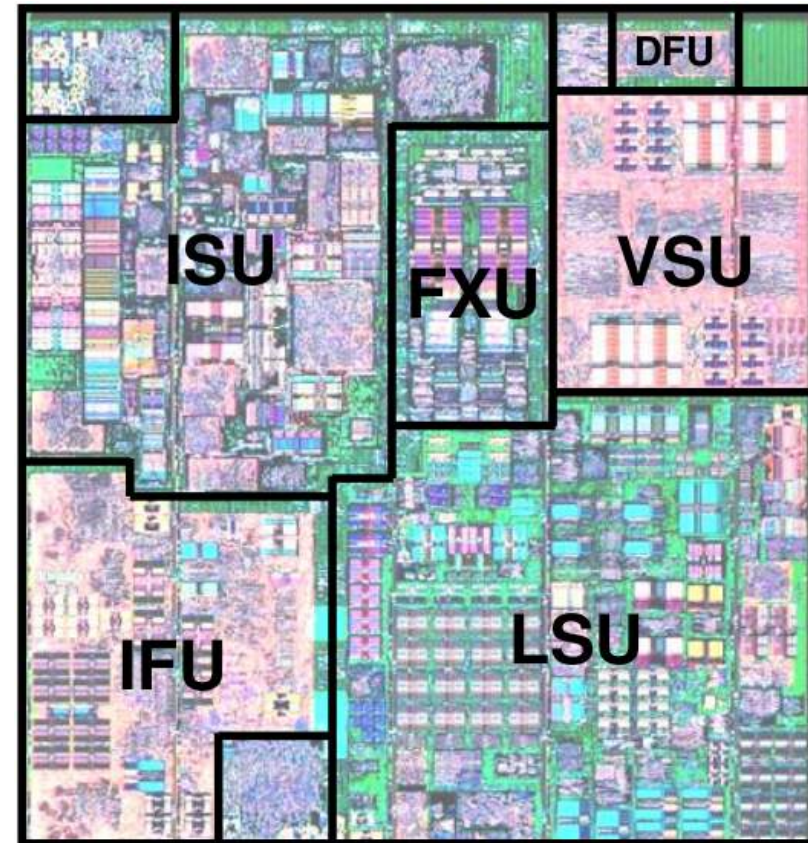
POWER8 core

- Execution Improvement (vs POWER7)
  - SMT4 -> SMT8
  - 16 execution units:
    - 2 FXU, 2 LSU, 2 LU, 4 FPU, 2 VMX,1 Crypto, 1 DFU, 1 CR, 1 BR
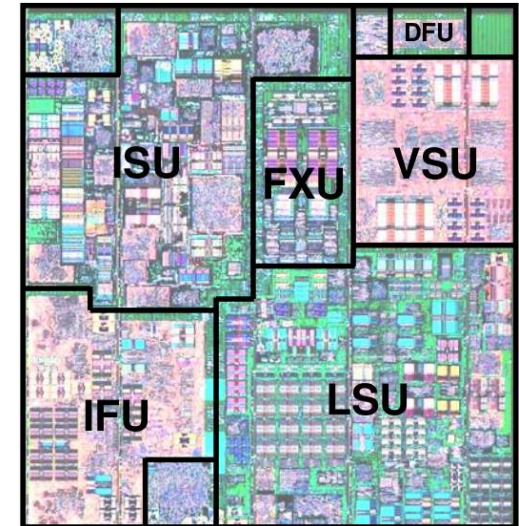
	POWER7	POWER8
Max Cores	8	12
SMT	4	8
L1D	32 K	64 K
L1I	32 K	32 K
L2	256 K	512 K
L3	8 MB	8 MB
L4 (off-chip)	-	128 MB
Memory Controllers	2	2

eDRAM technology

IBM's eDRAM technology benefits: Greater density, Less power requirements, Fewer soft errors, and Better performance

eDRAM enables on POWER8 96MB on-chip L3 cache

eDRAM enables 128MB off-chip L4 cache (on the DIMM)
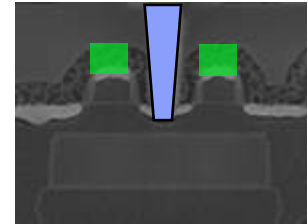
eDRAM Cell

L3 Cache critical to balanced design / performance:
- 6:1 Latency improvement for L3 accesses vs external L3
- 2X Bandwidth improvement with on chip interconnect. 32B busses to and from each core
- No off chip driver or receivers in L3 access path.

eDRAM is nearly as fast as conventional SRAM but requires far less space
- 1/3 the space of conventional 6T SRAM memory cell implementation
- 1/5 the standby power
- Soft Error Rate 250x lower than SRAM (Better availability)

IBM is effectively doubling microprocessor performance beyond what classical scaling alone can achieve," said Dr. Subramanian Iyer, IBM Fellow
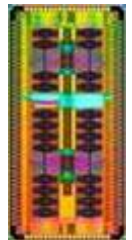
POWER8 Chip Caches

- **L2: 512 KB 8 way per core**

- **L3: 96 MB (12 x 8 MB 8 way Bank)**

- **"NUCA" Cache policy (**Non-Uniform Cache Architecture)
    Scalable bandwidth and latency
    Migrate "hot" lines to local L2, then local L3 (replicate L2 contained footprint)

- **Chip Interconnect: 150 GB/sec x 12 segments per direction = 3.6 TB/sec**
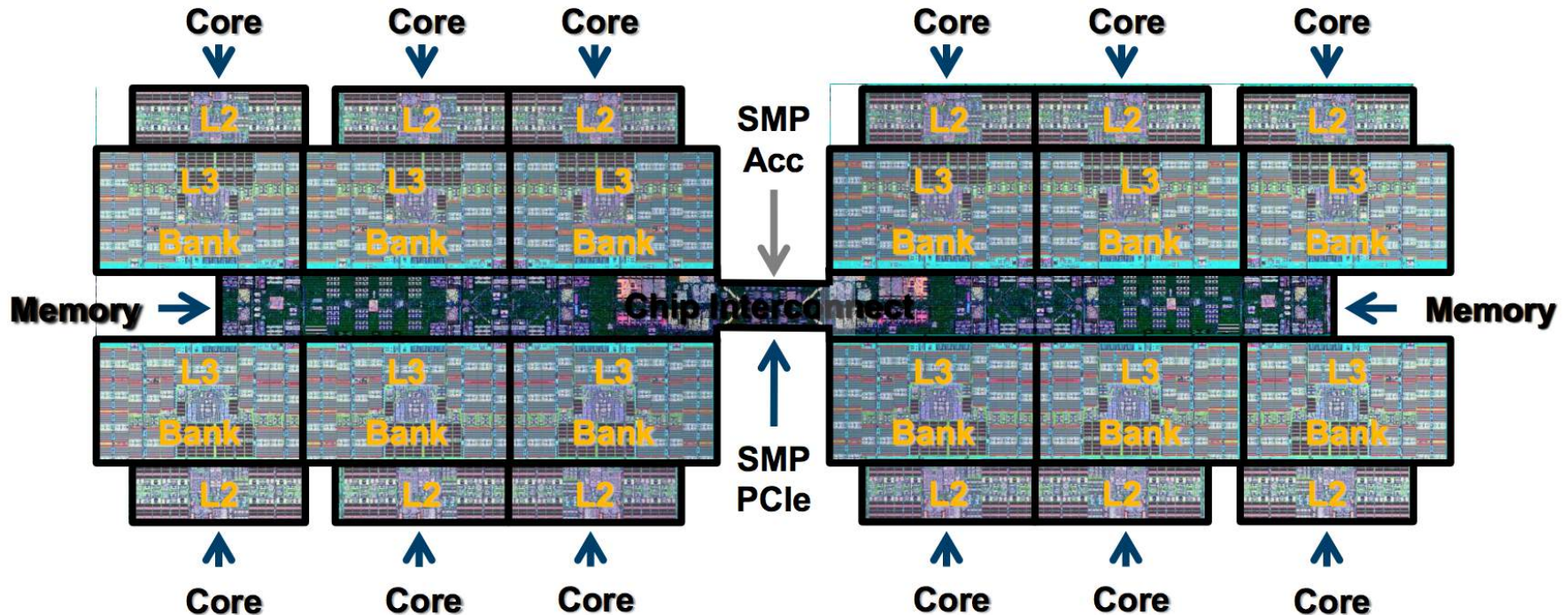
POWER8 Multi-threading Options

- **ST:** Largest unit of execution work
- **SMT2:** Smaller unit of work, but provides greater amount of execution work per cycle
- **SMT4:** Smaller unit of work, but provides greater amount of execution work per cycle
- **SMT8:** Smallest unit of work, but provides the maximum amount of execution work per cycle
- Can dynamical shift between modes as required: ST / SMT2 / SMT4 / SMT8
- Mixed SMT modes supported within same partition
  - Requires use of "Resource Groups"
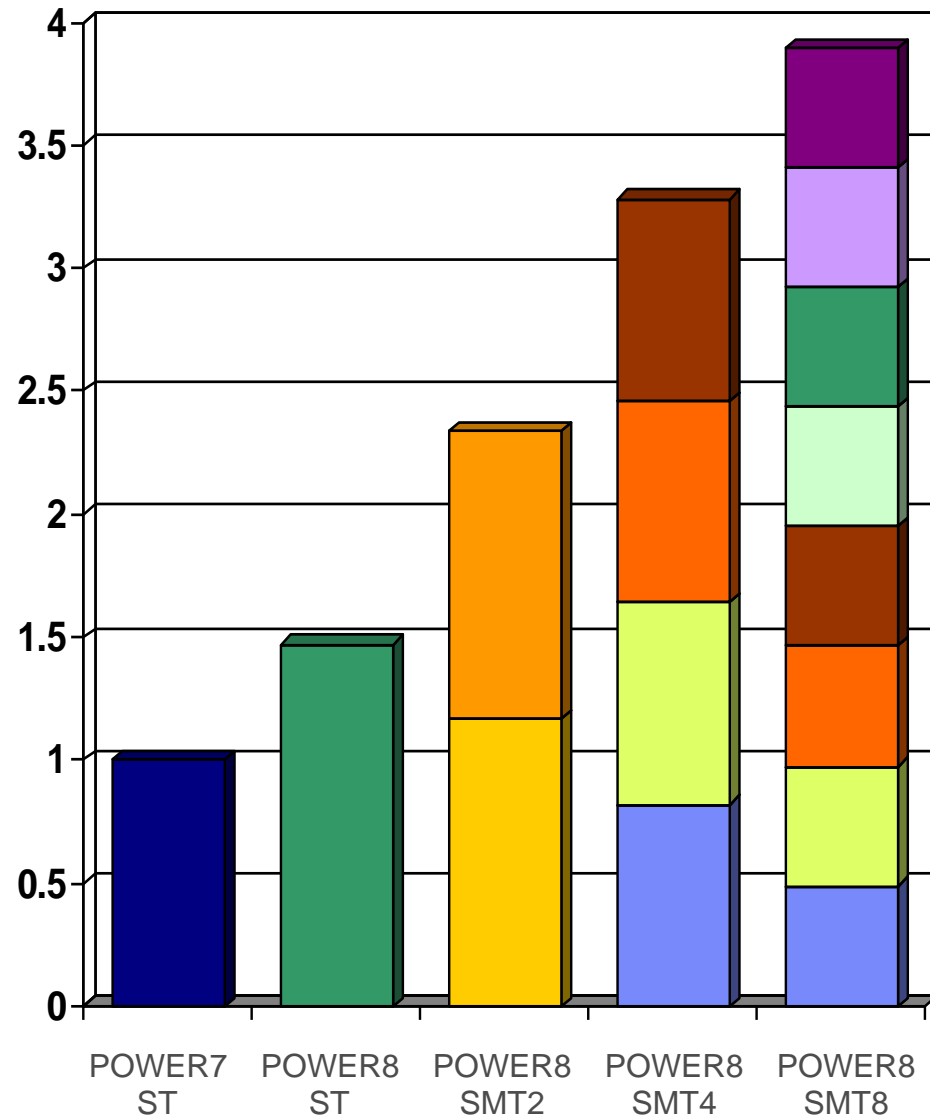
POWER8 "Turbo"

12 Core Sockets
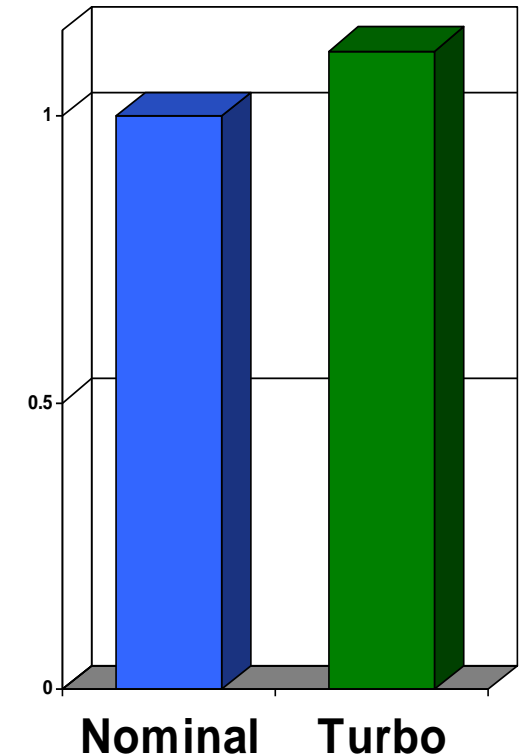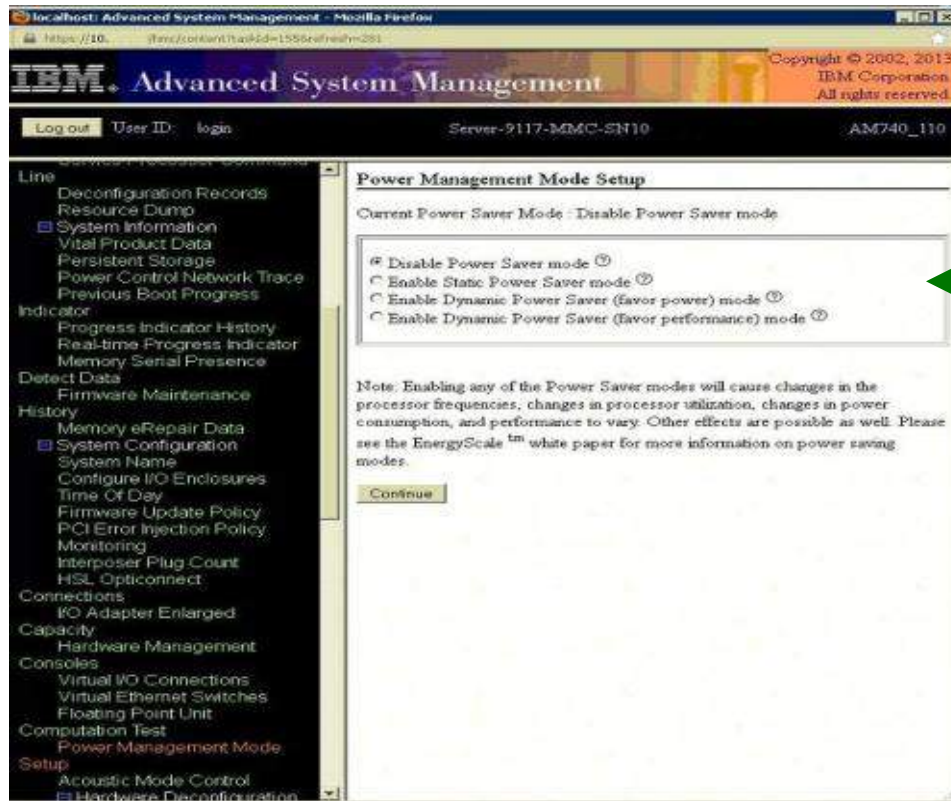
- Configure via ASMI menu
- Potential increase in processor frequency: ~ 11+%
- Requires firmware 810 (POWER8 support)
- CPW & rPerf measured using nominal
- Note: fans run faster increasing noise level

This is not POWER7 Turbo-core !!! But the same option exists for POWER7 systems
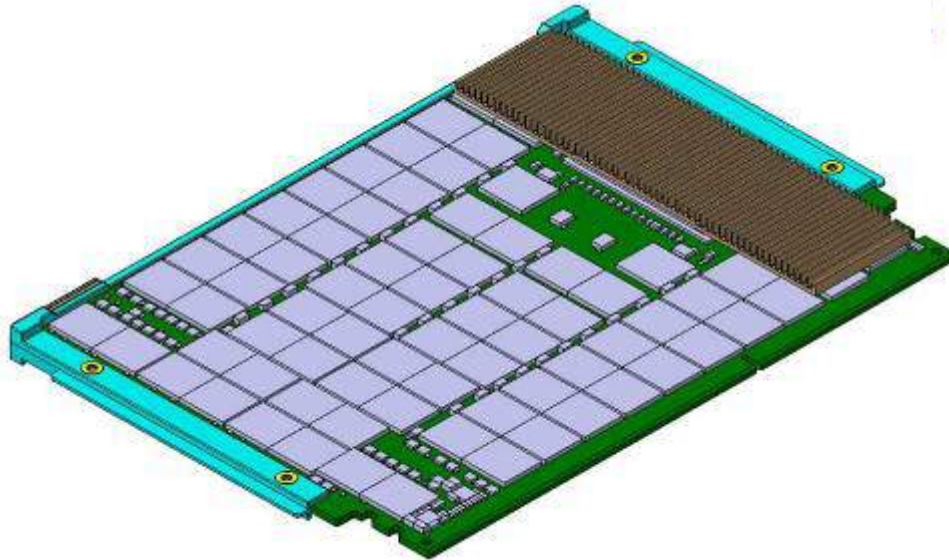
POWER8 Memory Chip

- IBM Memory DIMM
  - DDR3 RAM 1.600 MHz
- Intelligent Memory
  - Tasks performed by cores now done by DIMM
    - RAS decision point
    - Energy Management
    - Scheduling Logic
    - Writes Optimization
- L4 Memory Buffer
  - 16MB eDRAM / DIMM
  - Write Cache
  - Partial Updates
  - Reduces Latency
- 2 DIMM types
  - Low Profile
    - 2U & 4U Servers
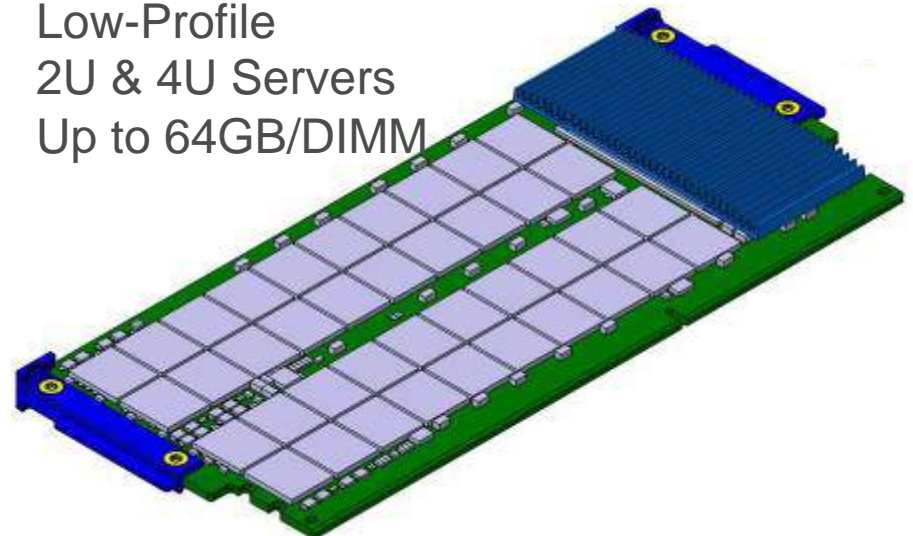 - High Profile – 4U Servers
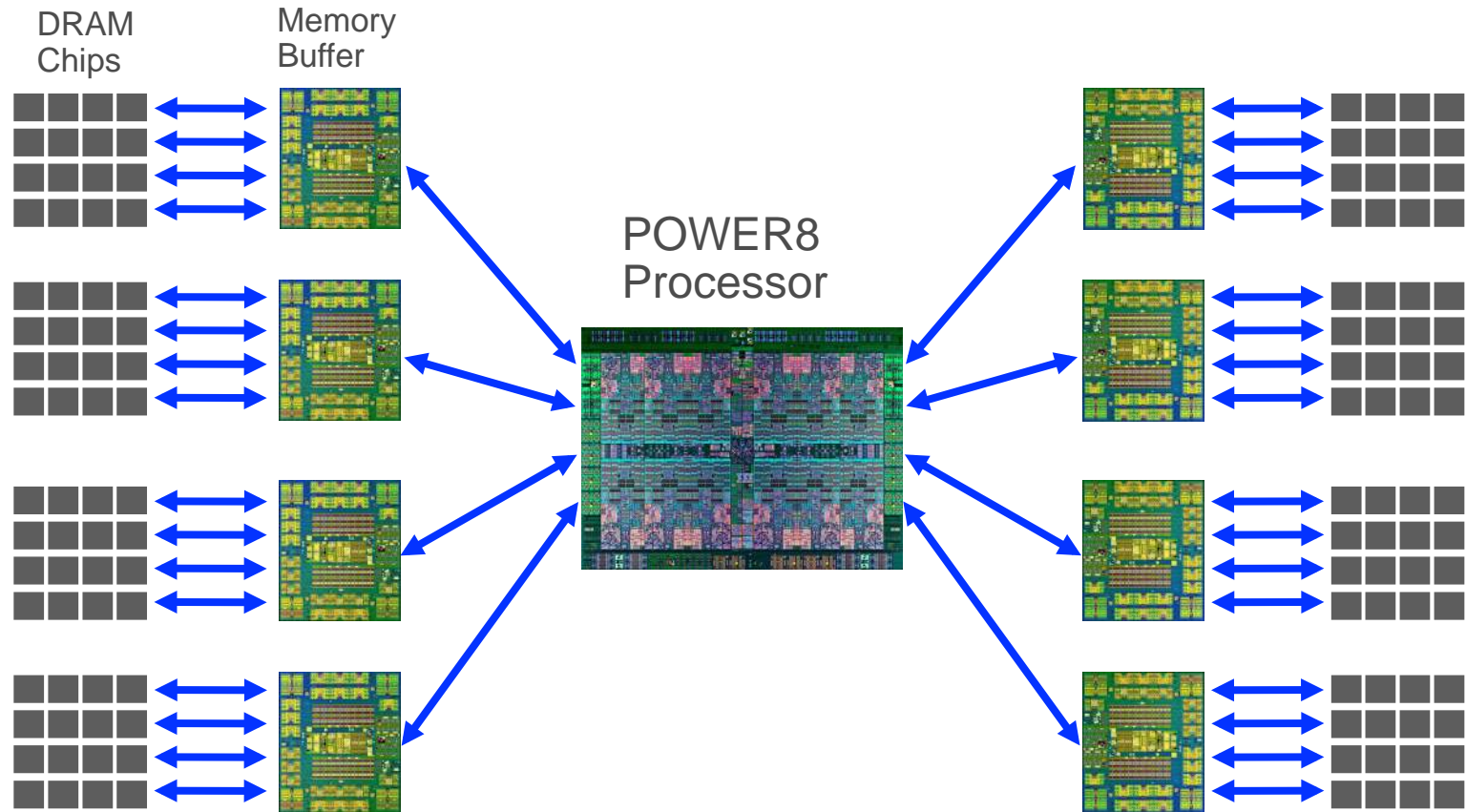   and future Enterprise servers
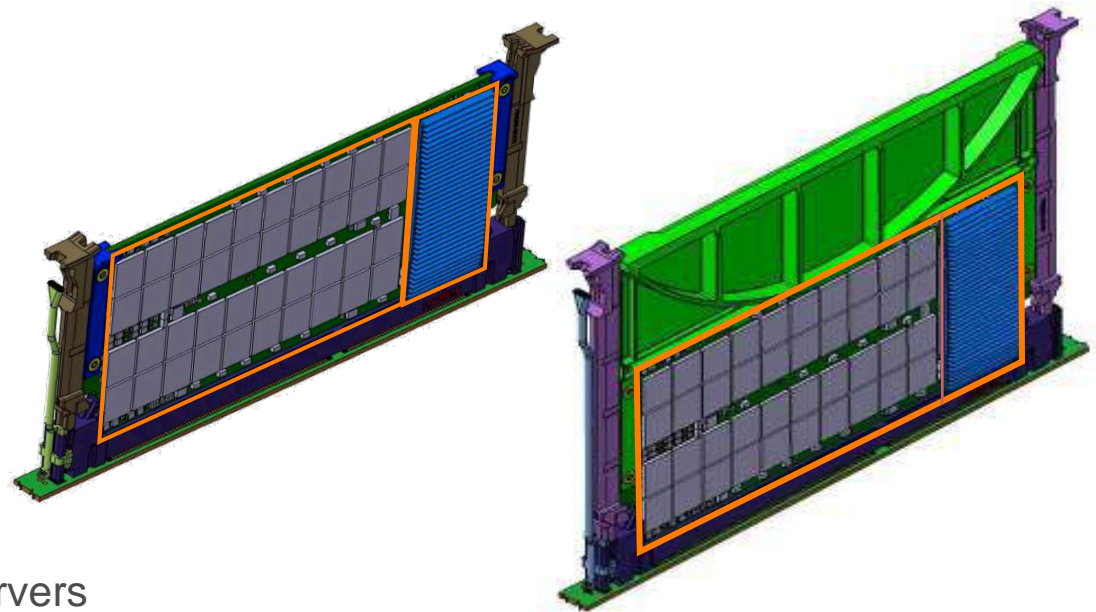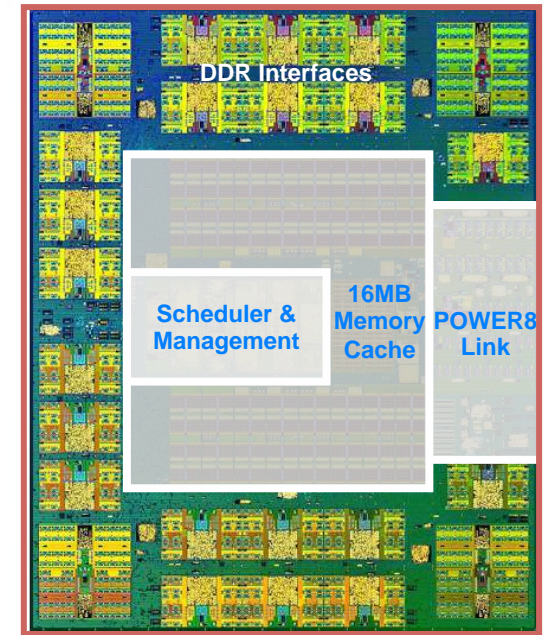
POWER8 DIMM

High-Profile – planned for
4U Servers and Enterprise
servers
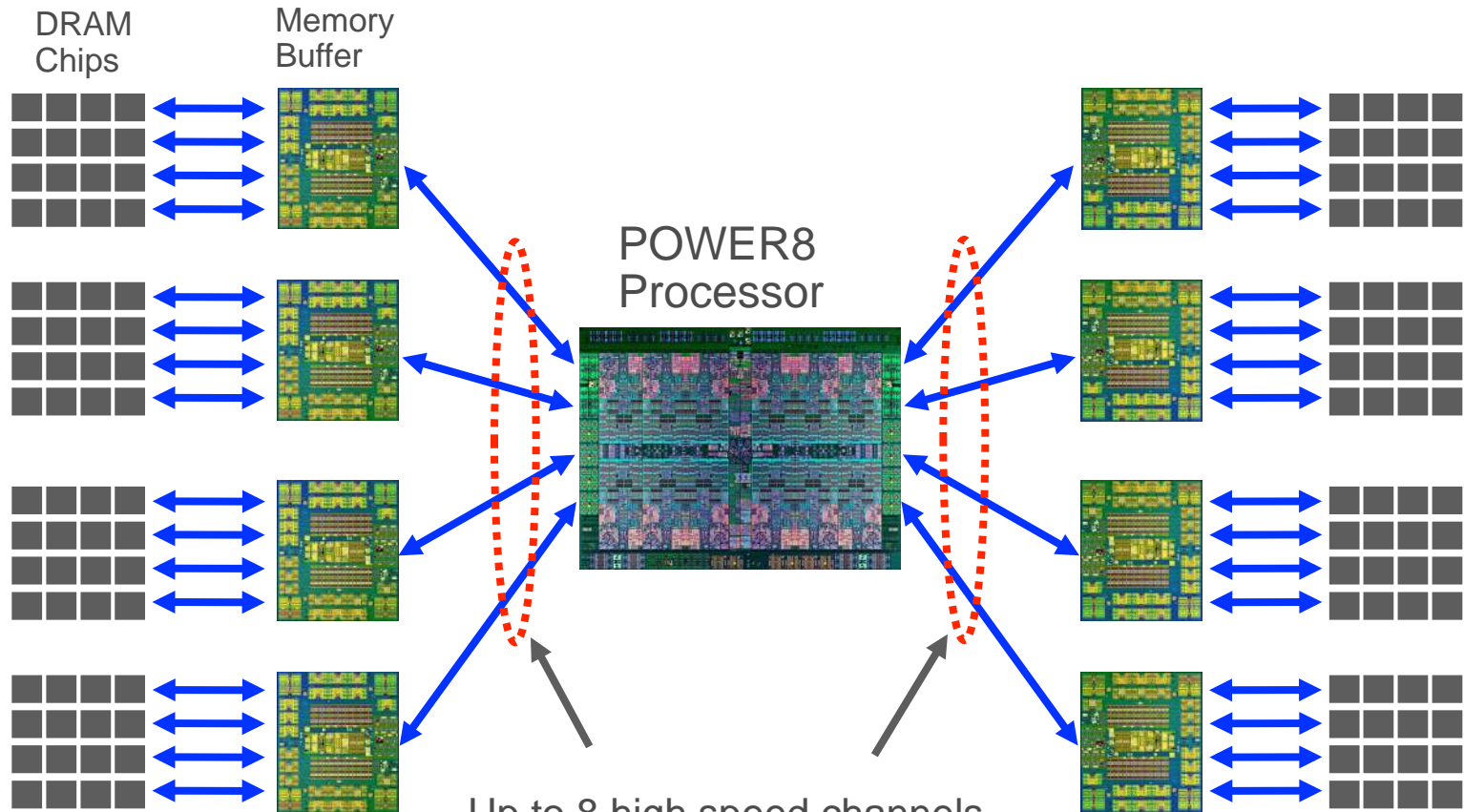128GB/DIMM and above

Low-Profile
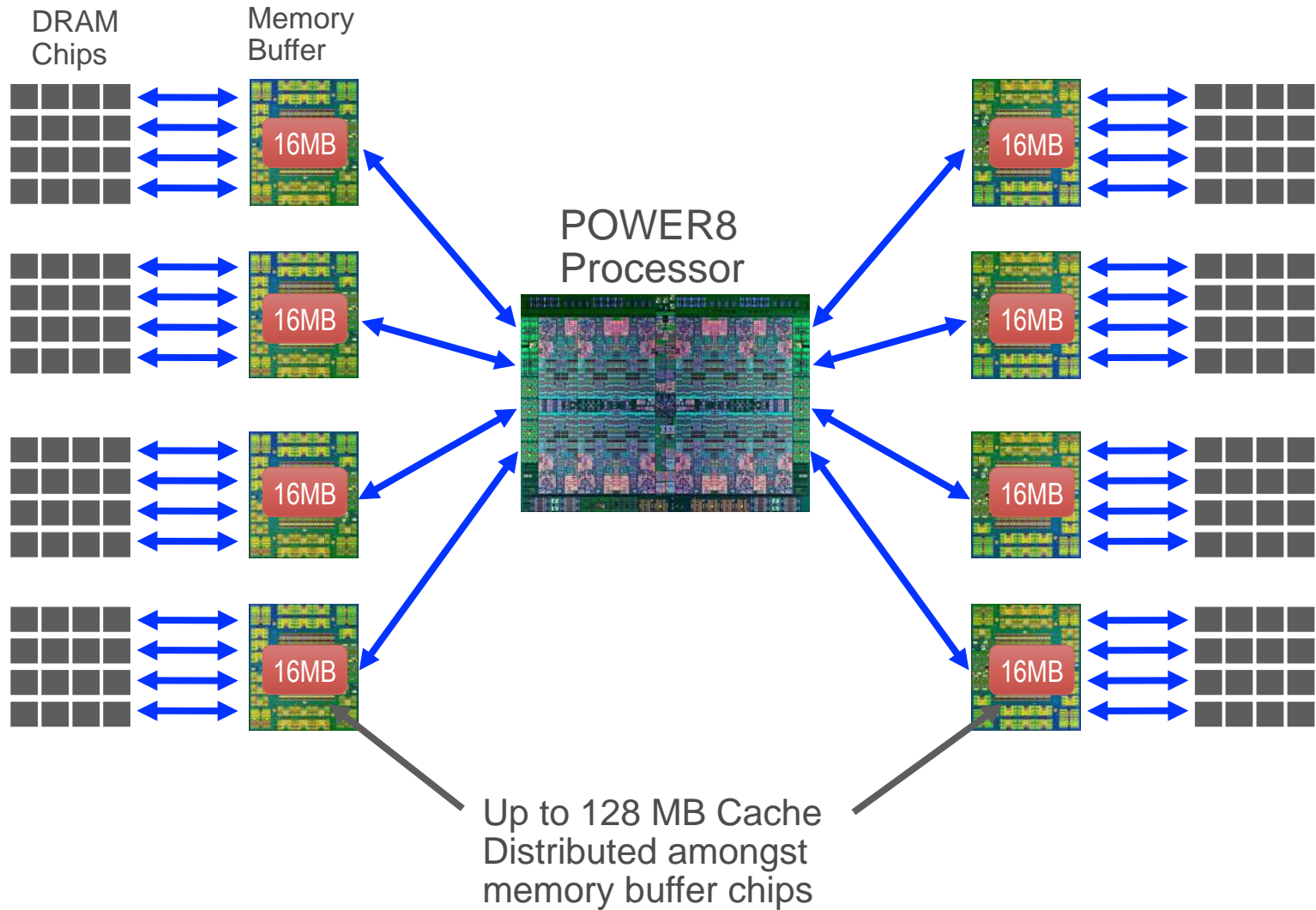2U & 4U Servers
Up to 64GB/DIMM

POWER8 Memory

DRAM
Chips

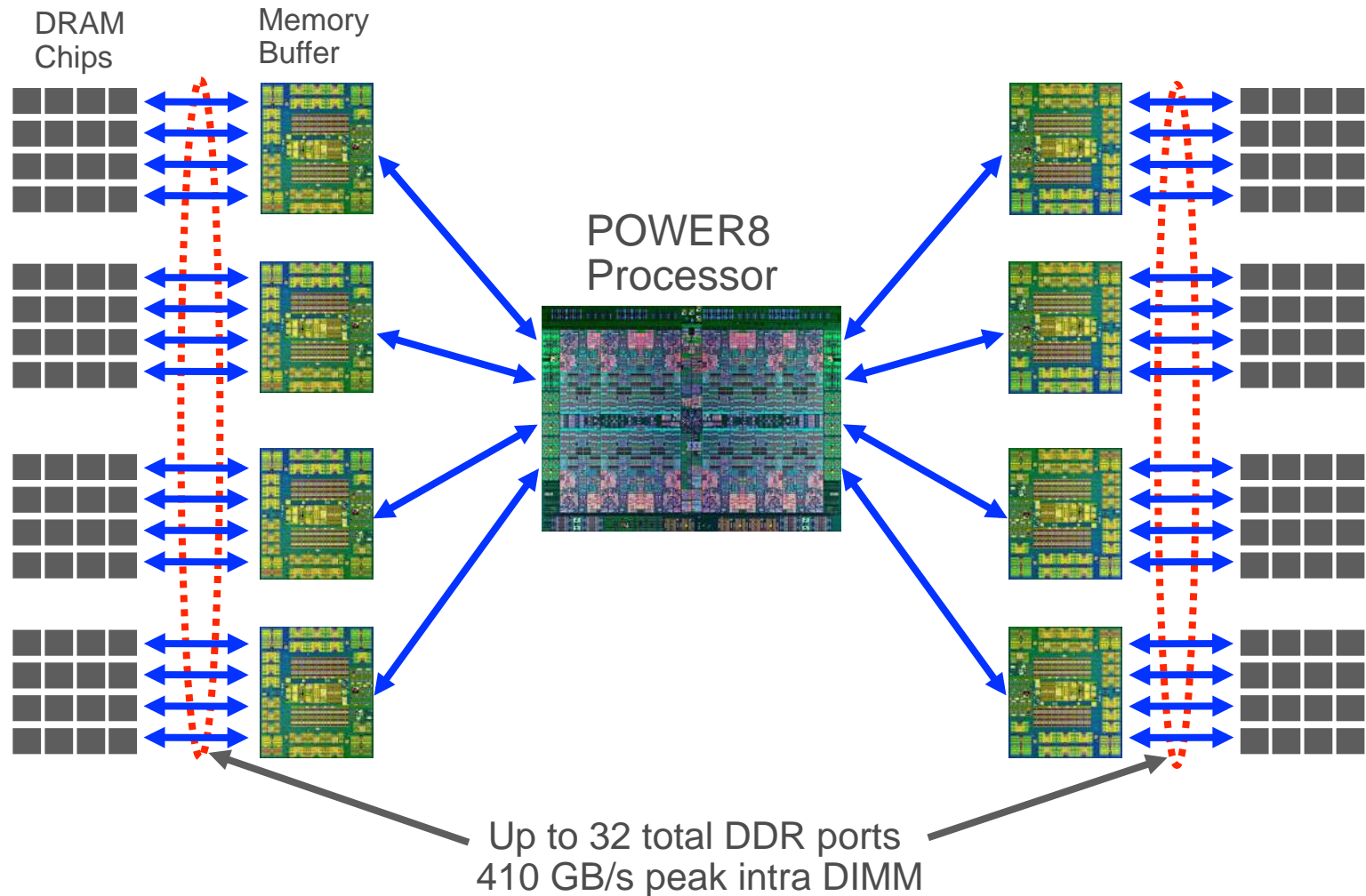Memory
Buffer

POWER8
Processor

POWER8 Memory

DRAM Chips

Memory Buffer

POWER8 Processor

Up to 8 high speed channels,
Up to 9.6 GB/s per channel
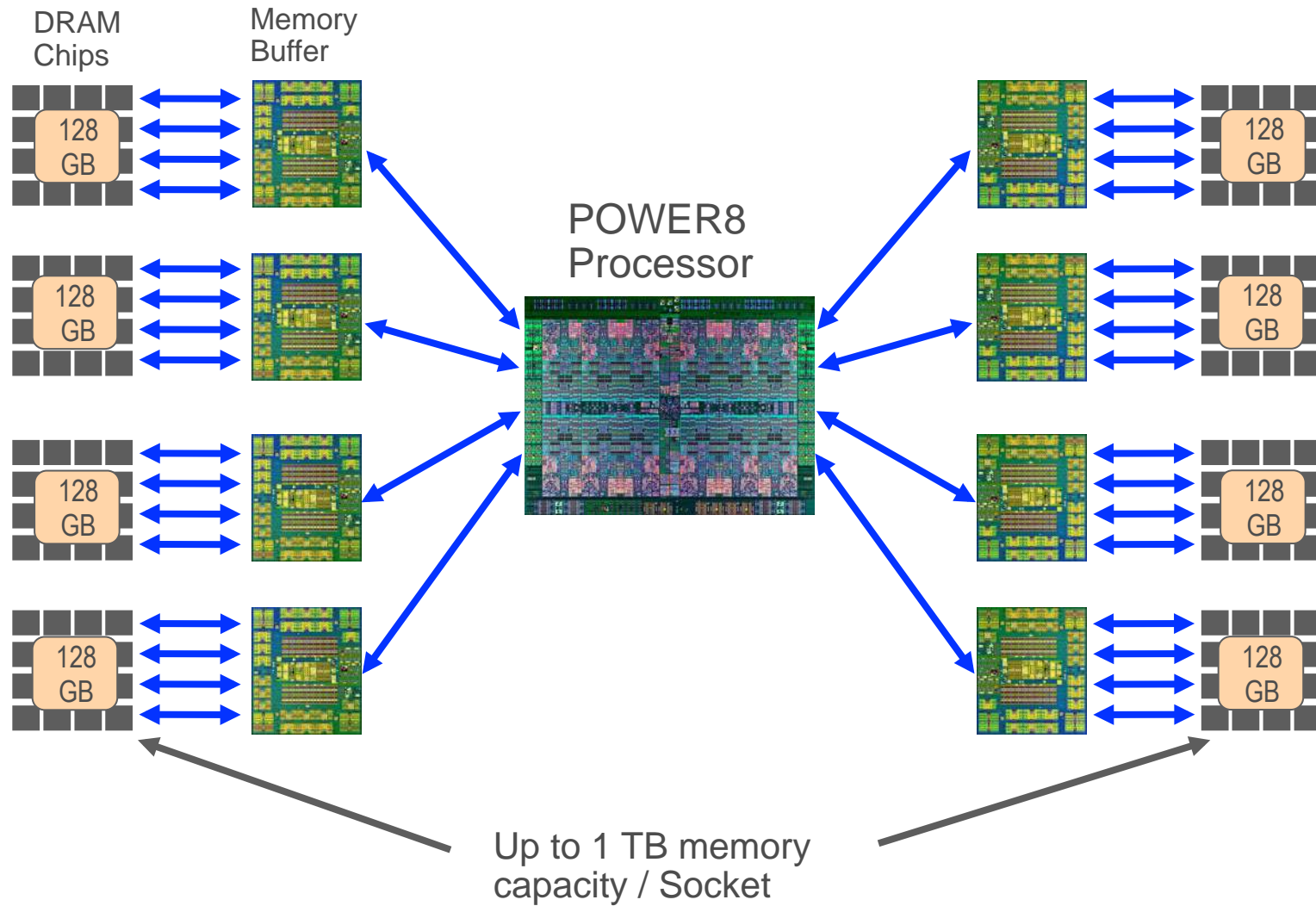(2B read / 1B write + cmd)
for up to 230 GB/s sustained

POWER8 Memory

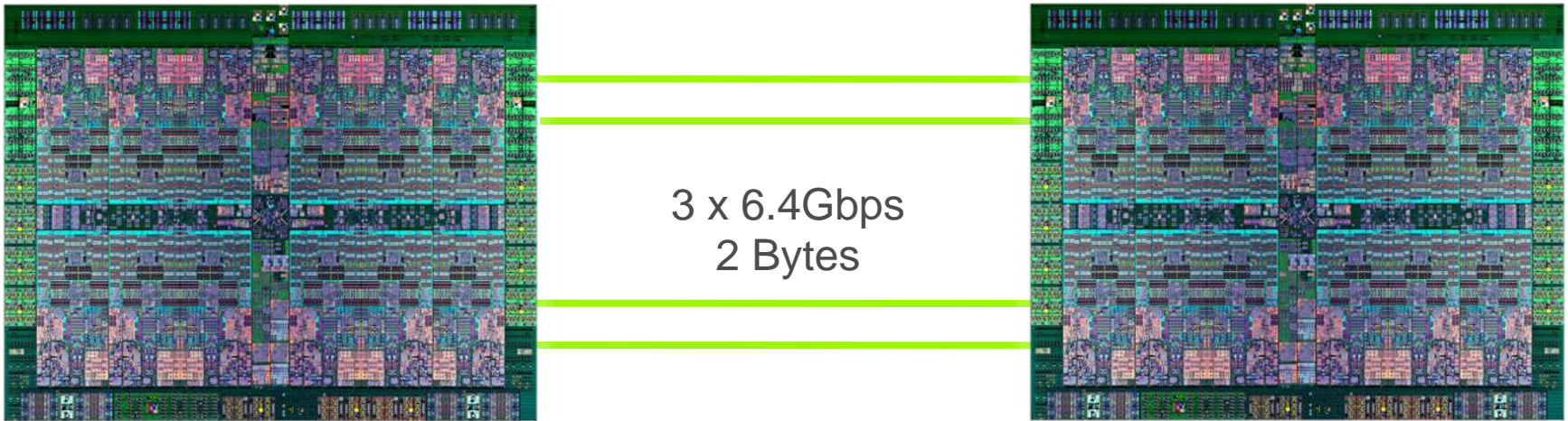DRAM Chips

Memory Buffer

16MB

16MB

16MB

16MB

16MB

16MB

16MB

16MB

POWER8 Processor

Up to 128 MB Cache Distributed amongst memory buffer chips

POWER8 Memory

DRAM
Chips

Memory
Buffer

POWER8
Processor

Up to 32 total DDR ports
410 GB/s peak intra DIMM

POWER8 Memory

DRAM
Chips

Memory
Buffer

POWER8
Processor

128 GB

Up to 1 TB memory
capacity / Socket

POWER8 2 Sockets SMP Interconnect

3 x 6.4Gbps
2 Bytes

3 12.8 GB/s channels = 38.4GB/s
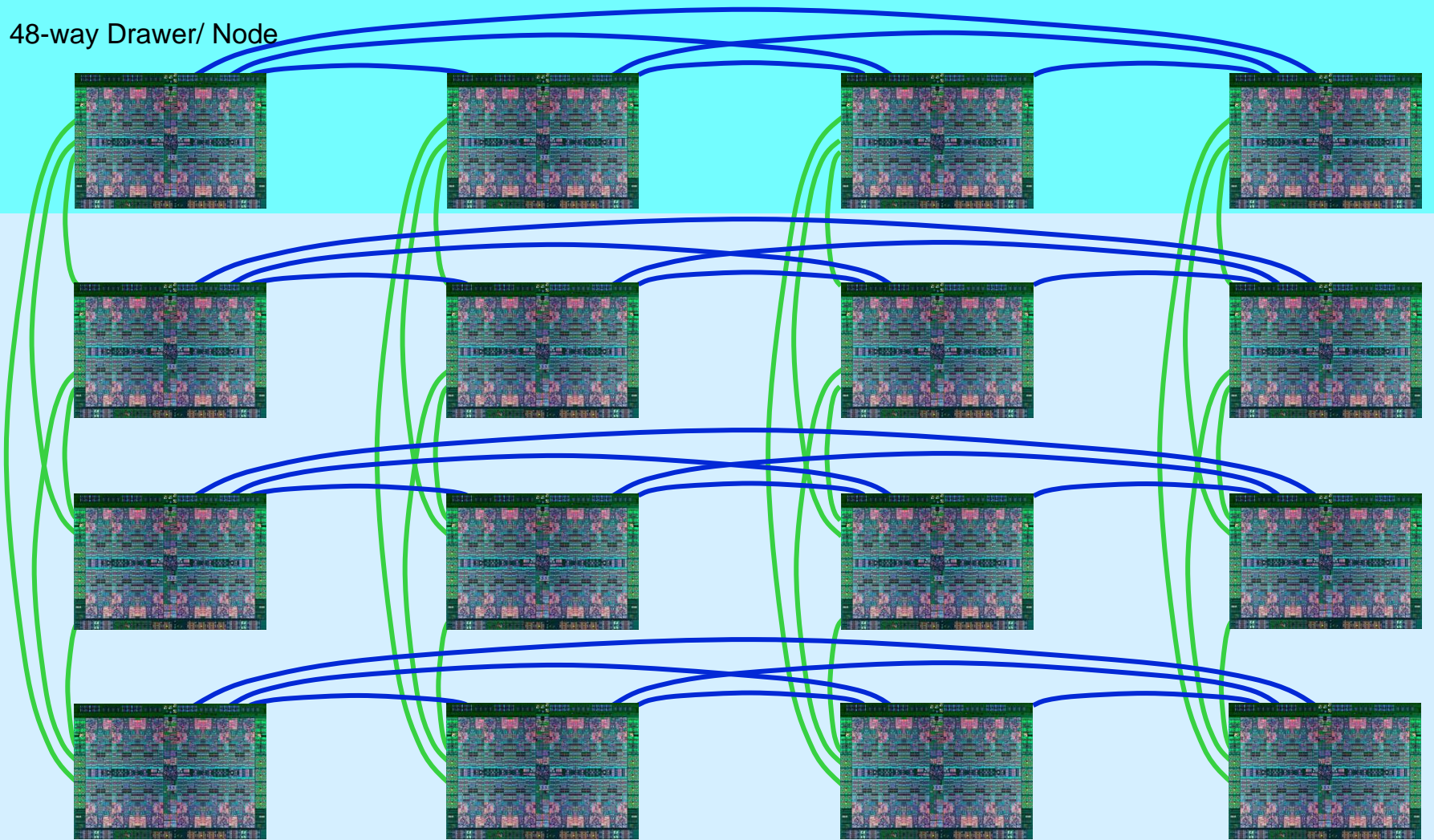+ Spare Lane

POWER8 4 Sockets SMP Interconnect - Future
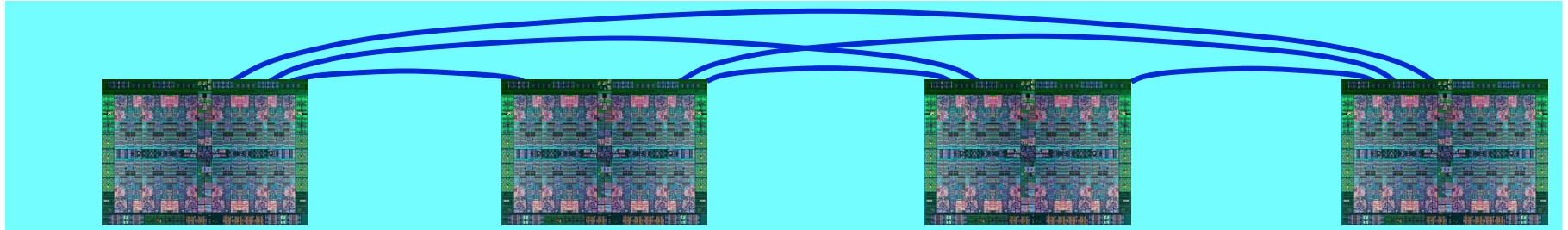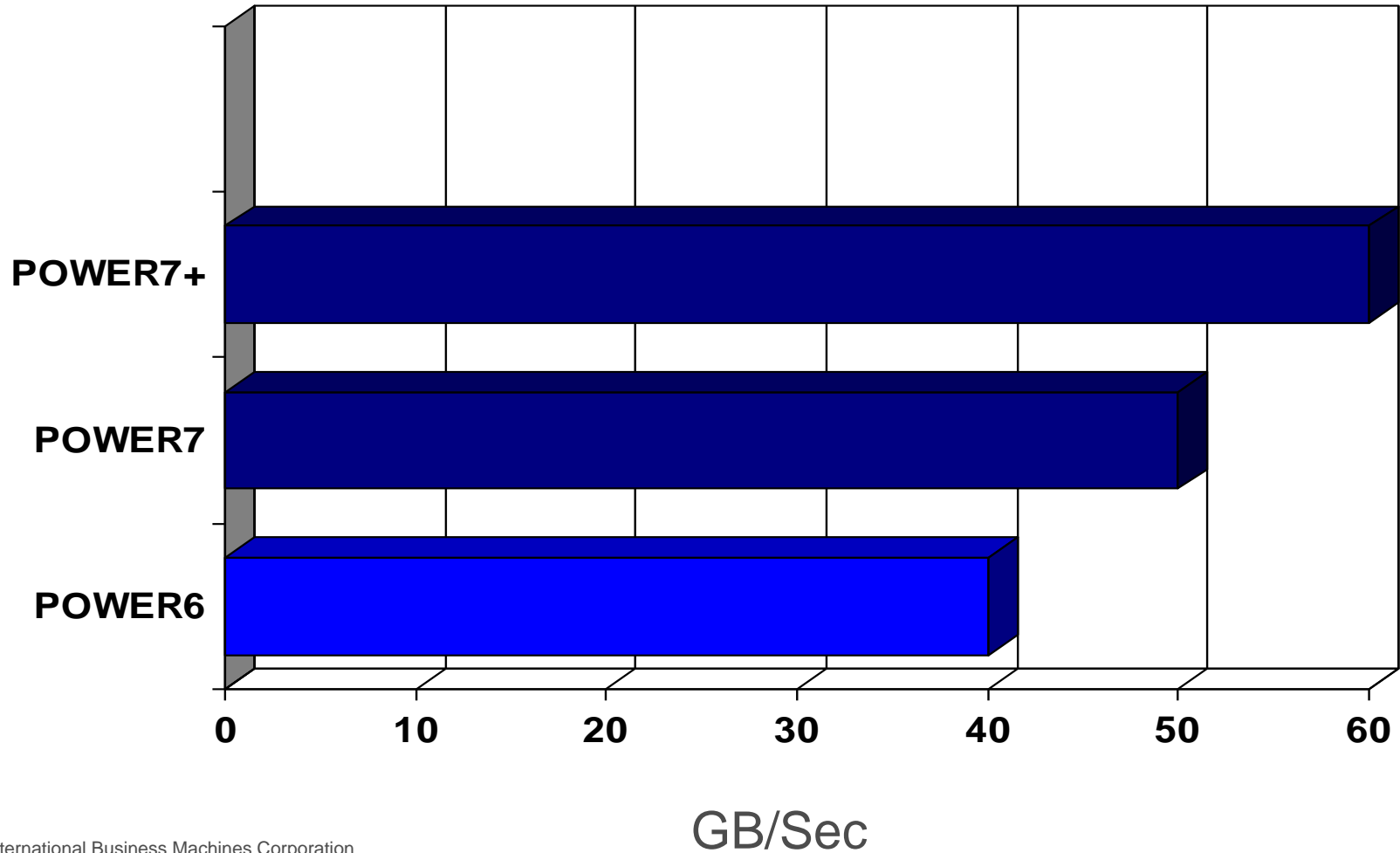

38.4 GB/s

POWER8 Max SMP Interconnect – Future
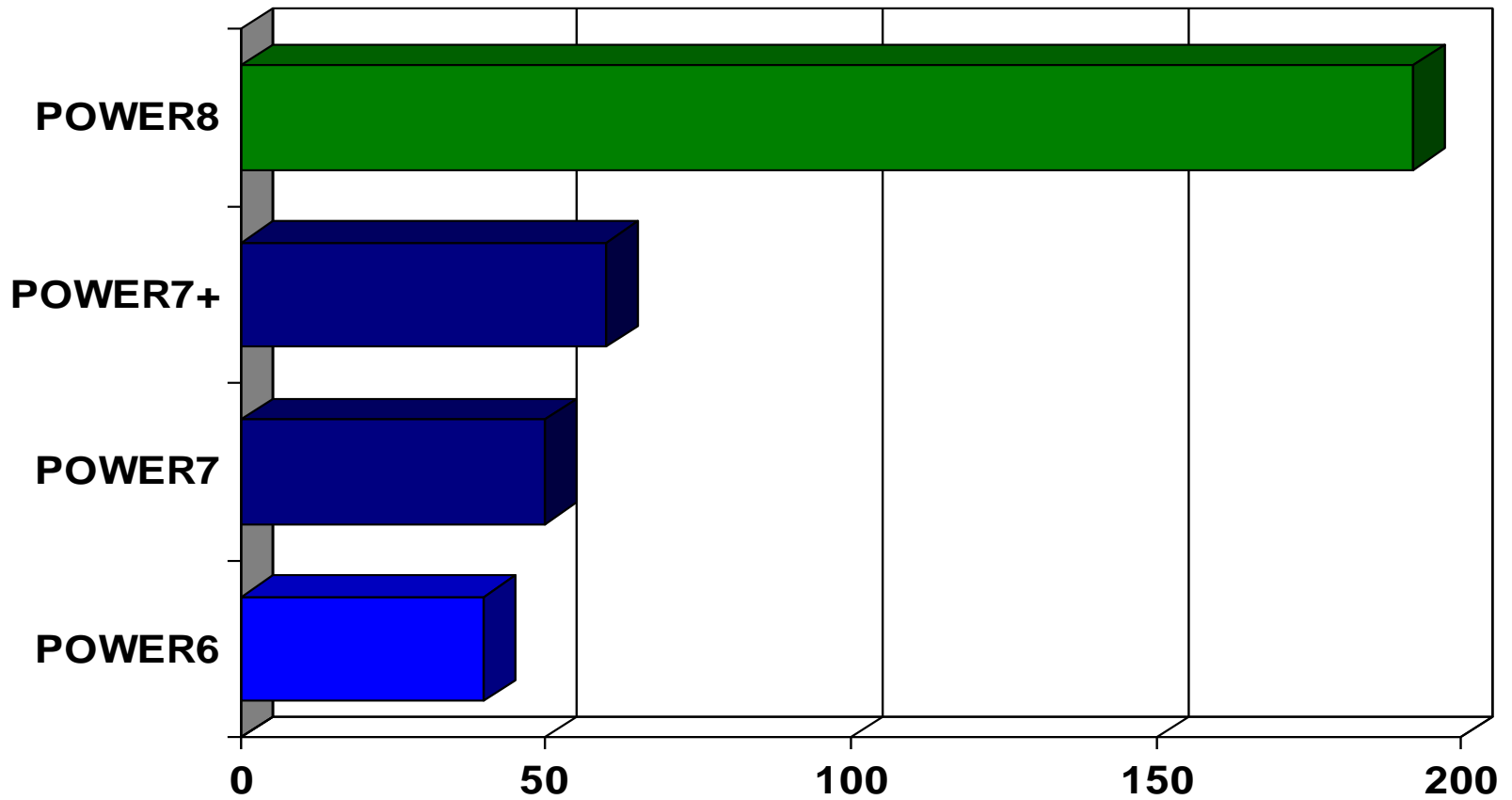
48-way Drawer/ Node

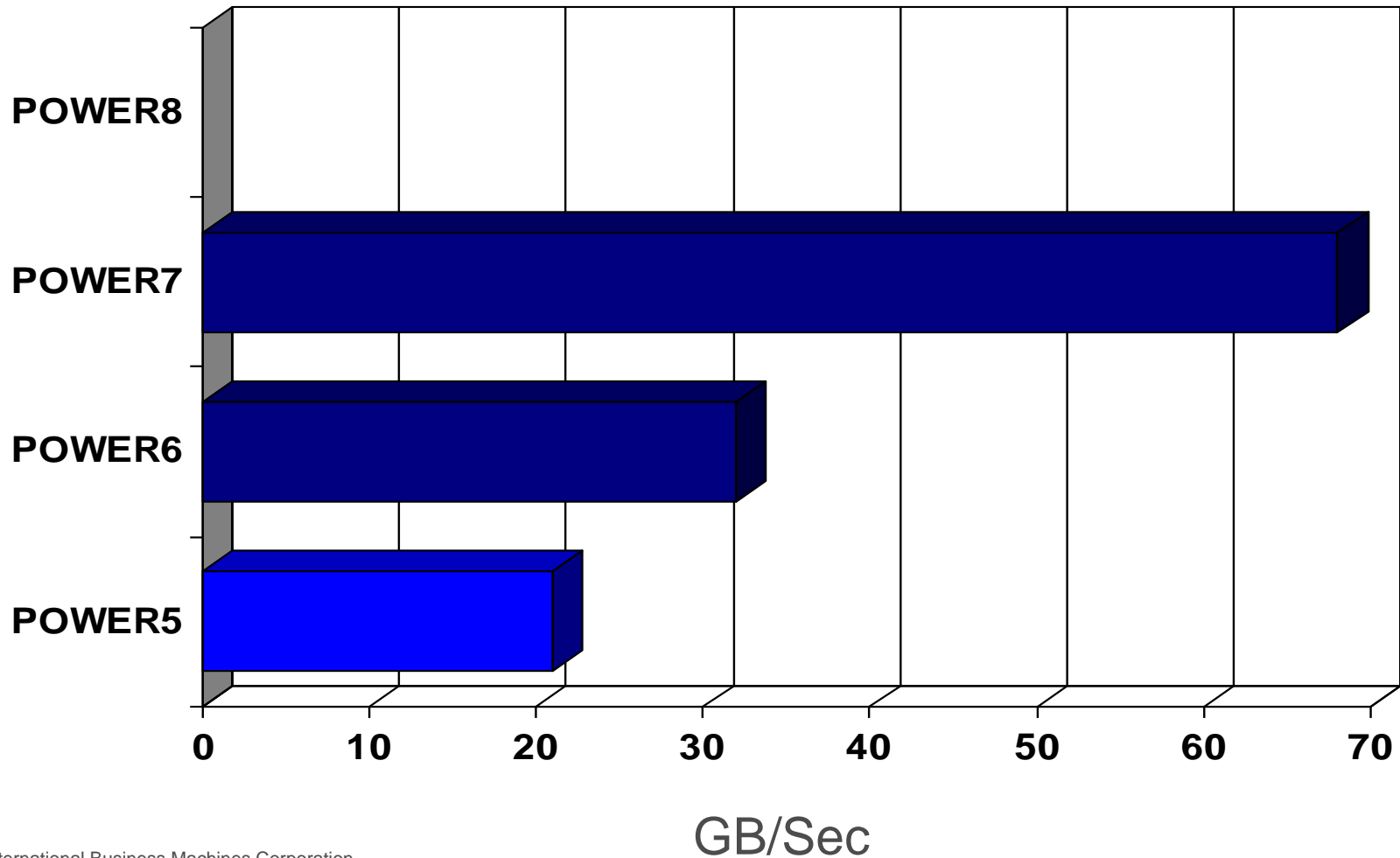192-way SMP system

—— 38.4 GB/s   —— 12.8 GB/s

IO Bandwidth

GB/Sec

IO Bandwidth

GB/Sec

Memory Bandwidth per Socket

Memory Bandwidth per Socket

GB/Sec
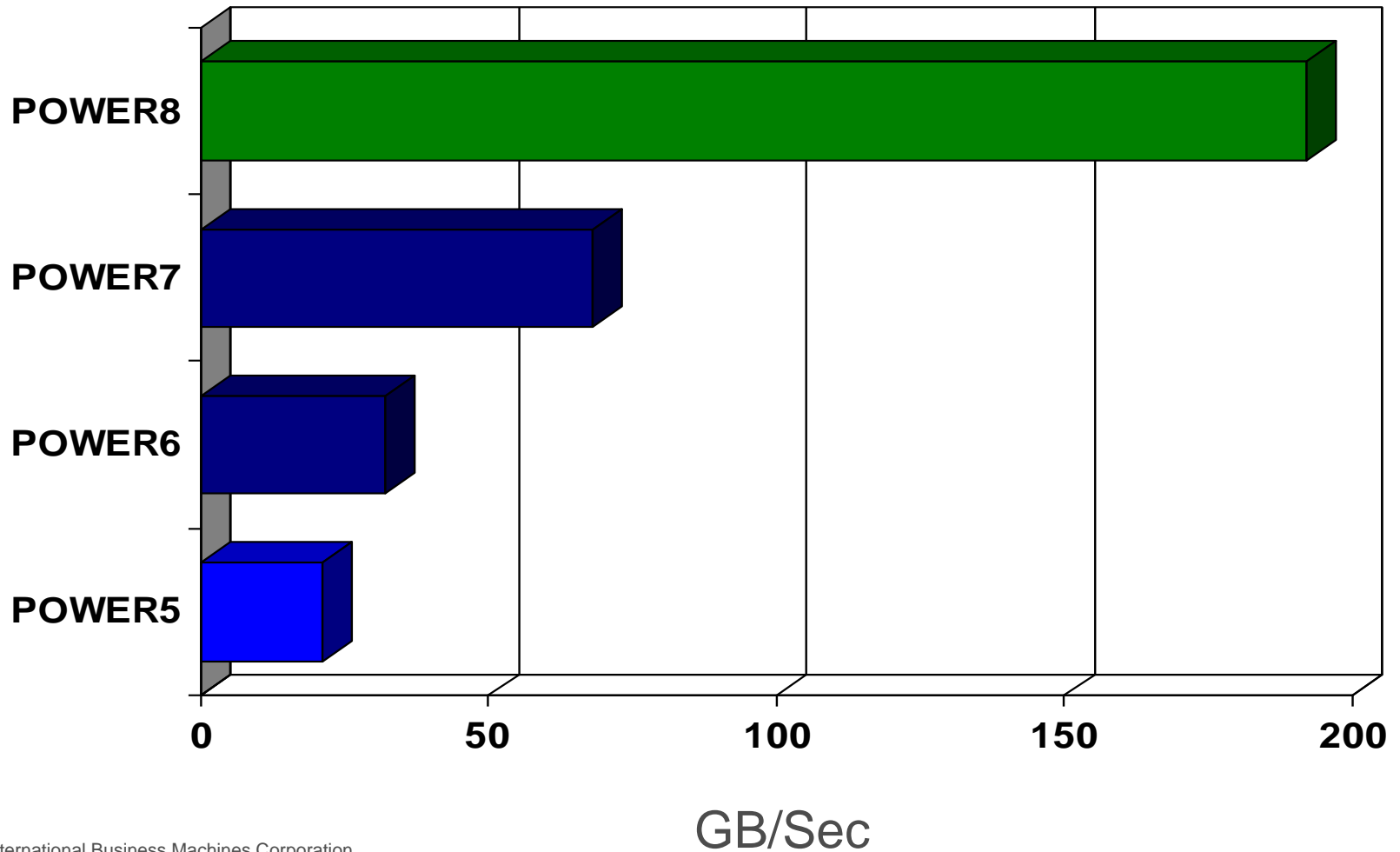
POWER8 Integrated PCI Gen 3

POWER8

POWER7

Native PCIe Gen 3 Support
- Direct processor integration
- Replaces proprietary GX/Bridge
- Low latency
- Gen3 x16 bandwidth (32 GB/s)

Transport Layer for CAPI Protocol
- Coherently Attach Devices connect to processor via PCIe
- Protocol encapsulated in PCIe
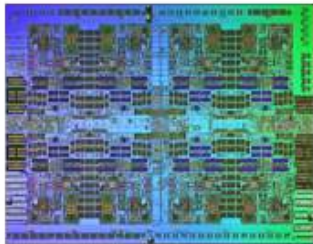
GX
Bus

I/O
Bridge

PCIe G2

PCIe G3

PCI
Devices

PCI
Device
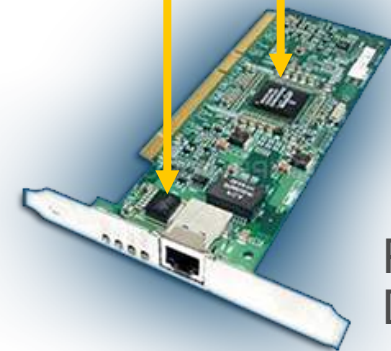
POWER8 CAPI (Coherent Accelerator Processor Interface)

Virtual Addressing
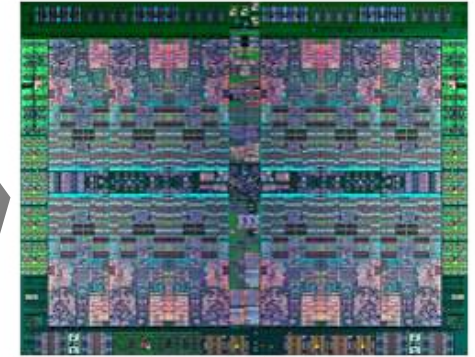- Accelerator can work with same memory addresses that the processors use
- Pointers de-referenced same as the host application
- Removes OS & device driver overhead

Hardware Managed Cache Coherence
- Enables the accelerator to participate in "Locks" as a normal thread Lowers Latency over IO communication model

Customizable Hardware
Application Accelerator
• Specific system SW, middleware, or user application
• Written to durable interface provided by PSL

CAPI Technology

Power Processor

Ecosystem
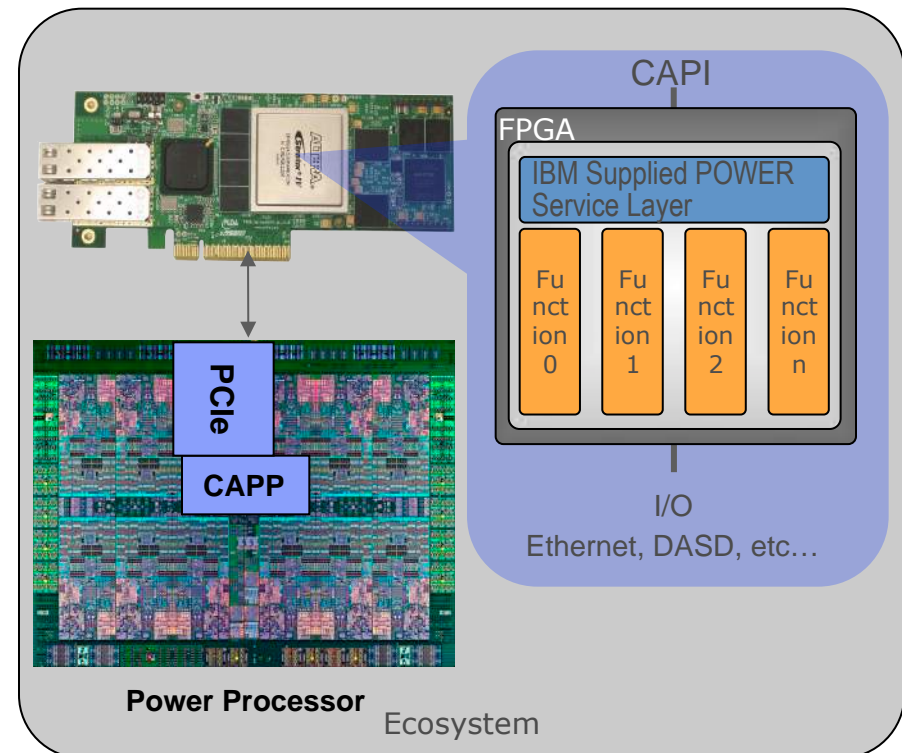
Traditional Accelerator Behavior

Application

Operating System

Device Driver

RAM
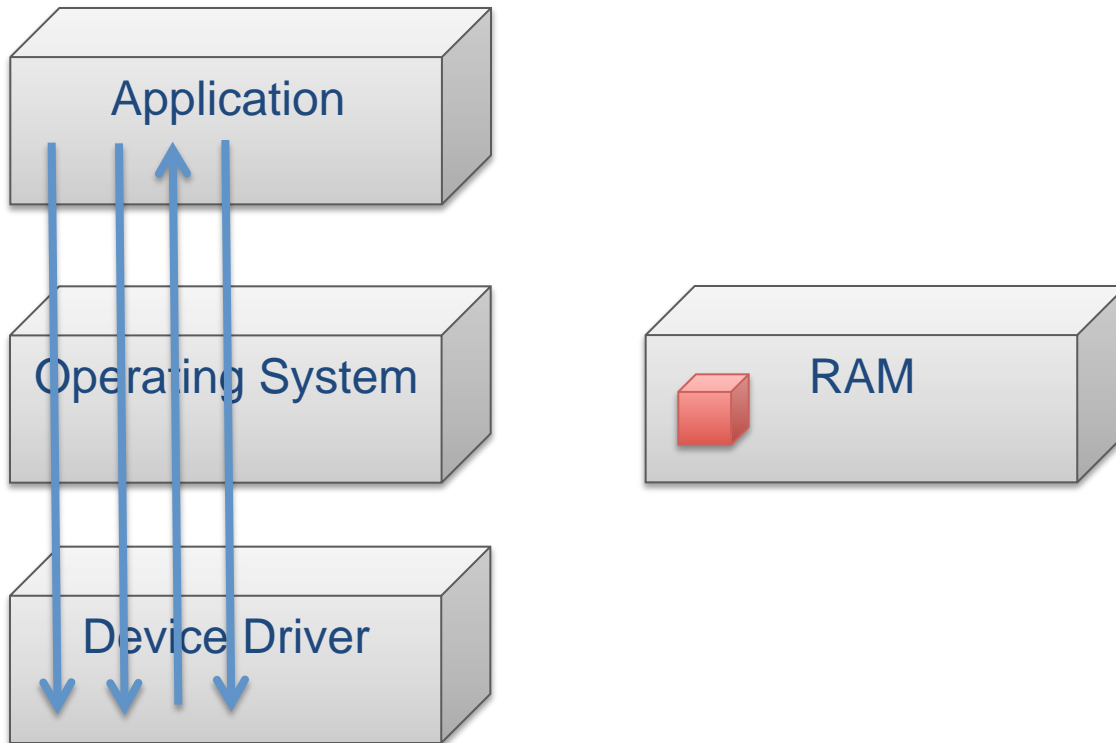
1. Device driver Call

2. Move data to Accelerator

3. Operation Start Interrupt

4. Acceleration

5. Operation Finish Interrupt

6. Get data from Accelerator

7. Device Driver Close

Coherent Accelerator Behavior

Application

Operating System

RAM

Device Driver

1. Notify Accelerator with Shared Memory Address

2. Acceleration on RAM

3. Operation Finished

Coherent Accelerator Processor Interface (CAPI) Overview

Power Processor

CAPI

FPGA

IBM Supplied POWER Service Layer

Function 0 | Function 1 | Function 2 | Function n

CAPP | PCIe

Typical I/O Model Flow

DD Call → Copy or Pin Source Data → MMIO Notify Accelerator → Acceleration → Poll / Int Completion → Copy or Unpin Result Data → Ret. From DD Completion
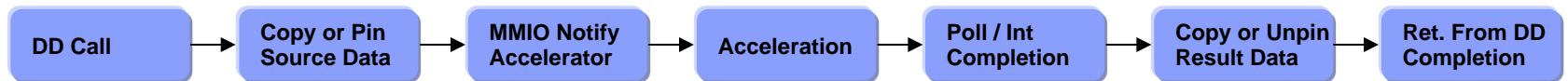
Flow with a Coherent Model

Shared Mem. Notify Accelerator → Acceleration → Shared Memory Completion
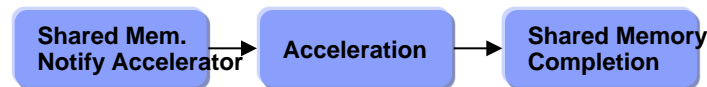
Advantages of Coherent Attachment Over I/O Attachment

- **Virtual Addressing & Data Caching**
  - Shared Memory
  - Lower latency for highly referenced data
- **Easier, More Natural Programming Model**
  - Traditional thread level programming
  - Long latency of I/O typically requires restructuring of application
- **Enables Applications Not Possible on I/O**
  - Pointer chasing, etc…

Advantages of Coherent Connected Accelerator

- **Virtual Addressing**
  - Removes the requirement for pinning system memory for PCIe transfers
    - Eliminates the copying of data into and out of the pinned DMA buffers
    - Eliminates the operating system call overhead to pin memory for DMA
  - Accelerator can work with same addresses that the processors use
    - Pointers can be de-referenced same as the host application
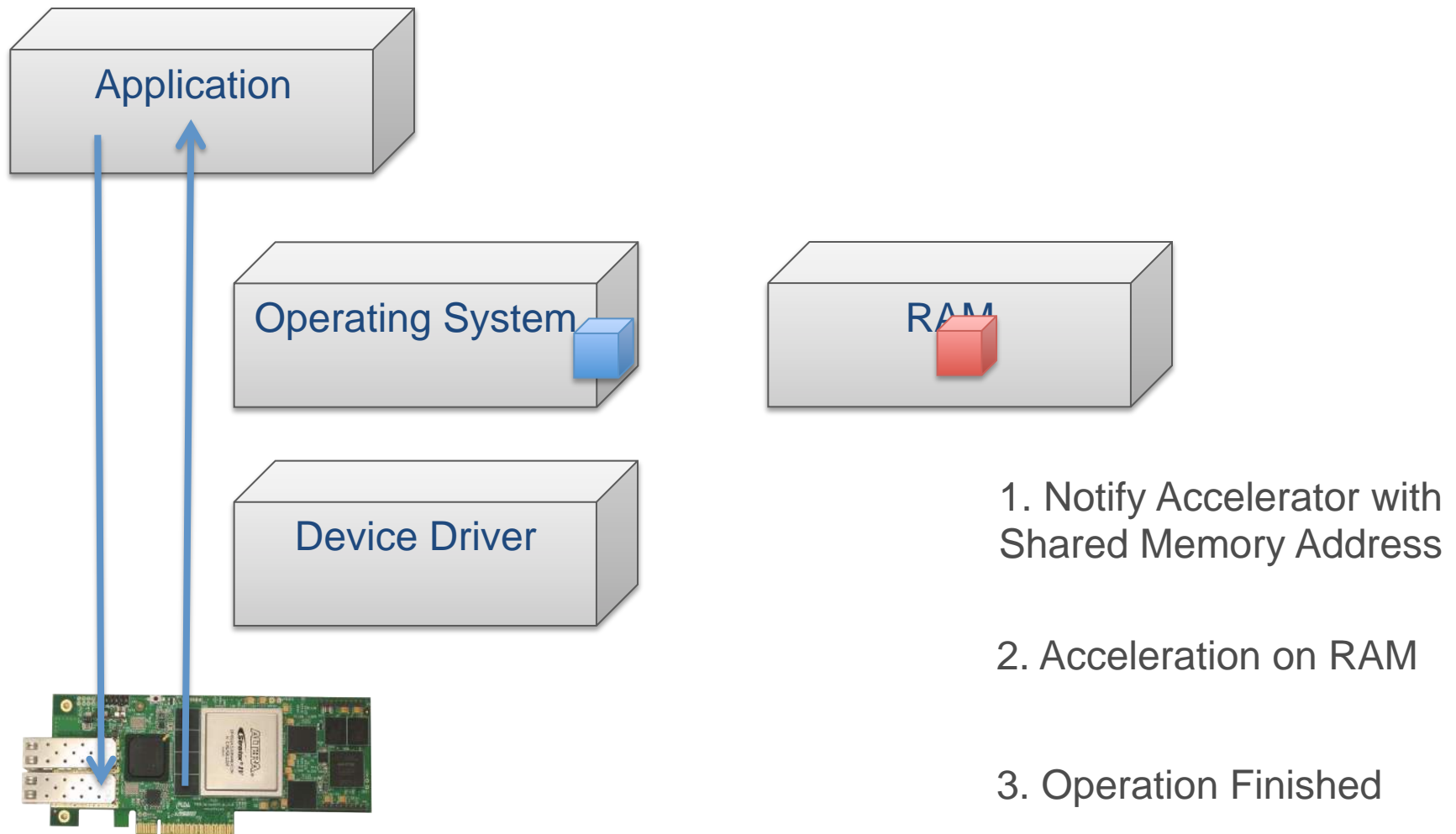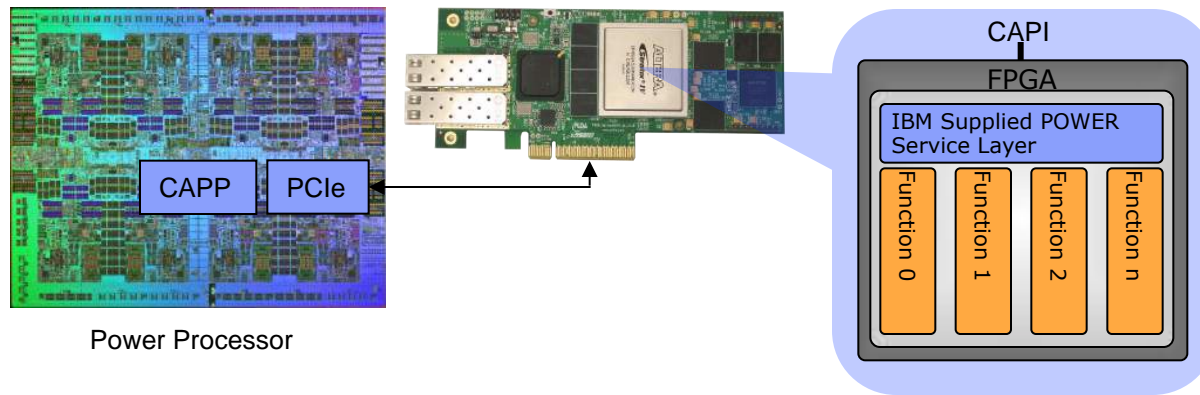   - Example: Enables the ability to traverse data structures

- **Coherent Caching of Data**
  - Enables an accelerator to cache data structures
  - Enables Cache to Cache transfers between accelerator and processor
  - Enables the accelerator to participate in "Locks" as a normal thread

- **Elimination of Device Driver**
  - Direct communication with Application
  - No requirement to call an OS device driver or Hypervisor function for mainline processing

- **Enables Accelerator Features not possible with PCIe**
  - Enables efficient Hybrid Applications
    - Applications partially implemented in the accelerator and partially on the host CPU
  - Visibility to full system memory
  - Simpler programming model for Application Modules

PCIe Java Gzip Compression FPGA

Transparent zlib HW acceleration integrated into JAVA 7.1

Corsa 5 A7
- Altera Stratix V PCIE Gen 2/3 card
- Currently supported as PCIe adapter - CAPI support possible in the future
- Initial offering with gzip use case for Java
- Dual SFP+ cages for networking - not used
- 8 GB DDR3 1600 SDRAM
- Full Card Level RAS

Attributes
- FPGA base accelerator card
- PCIe gen2 x8, ranges between 25W & 40W depending on application
    - ~21W for gzip application
- Follows Altera Stratix roadmap

PCIe base I/O adaptor
- JAVA 7.1 gzip acceleration use case
- Supports multi-threaded environment

OS Support
- AIX 7.1
- RHEL 7.0 with patches. Full distro support on RHEL 7.1
- SLES 12 later in 2014 for LE support

Programming model
- Altera Quartos 13.1
    - OpenCL Power run time support

RAS
- Follows same model as Z-EDC use case
- Full error detection scheme
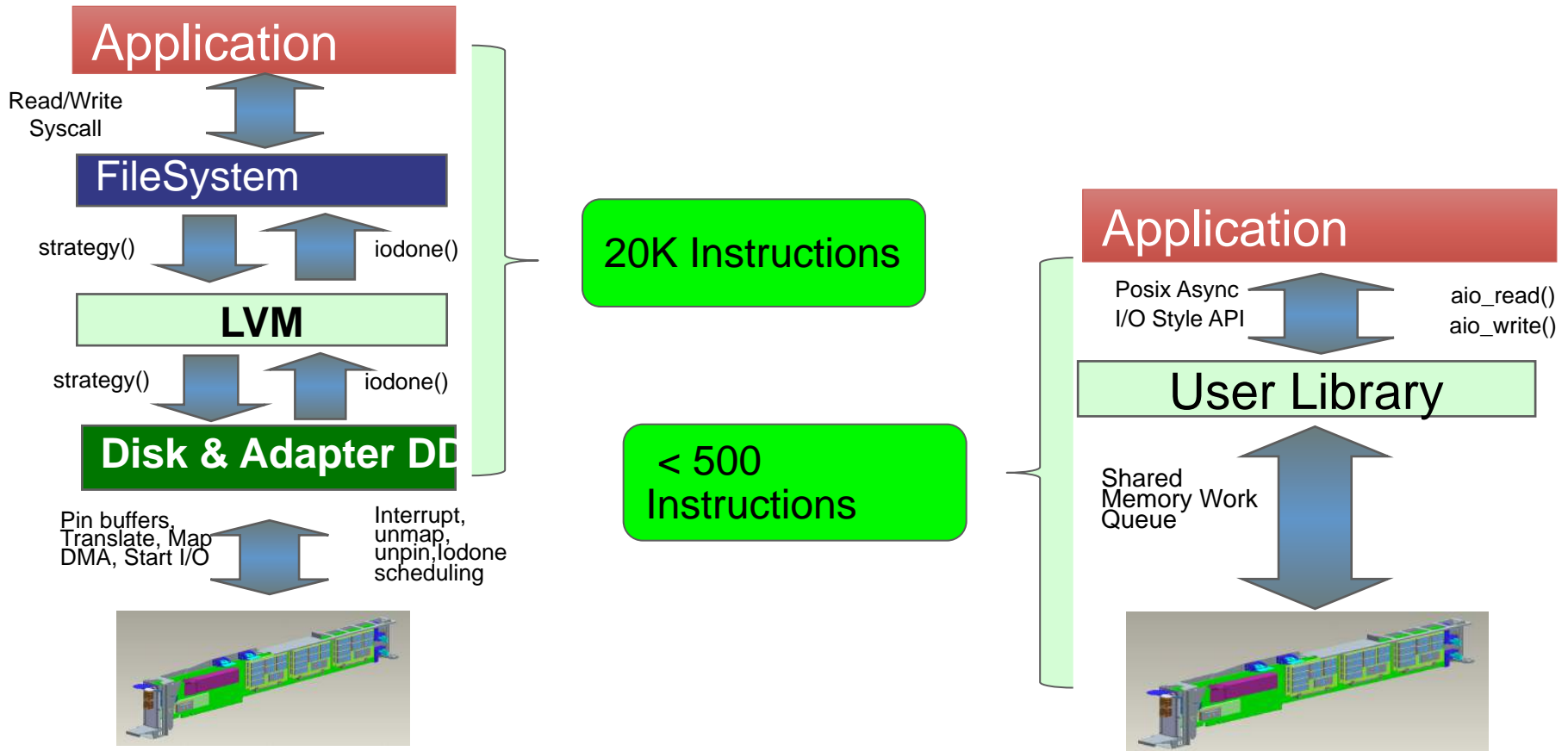- This application not reprogammabe by customers
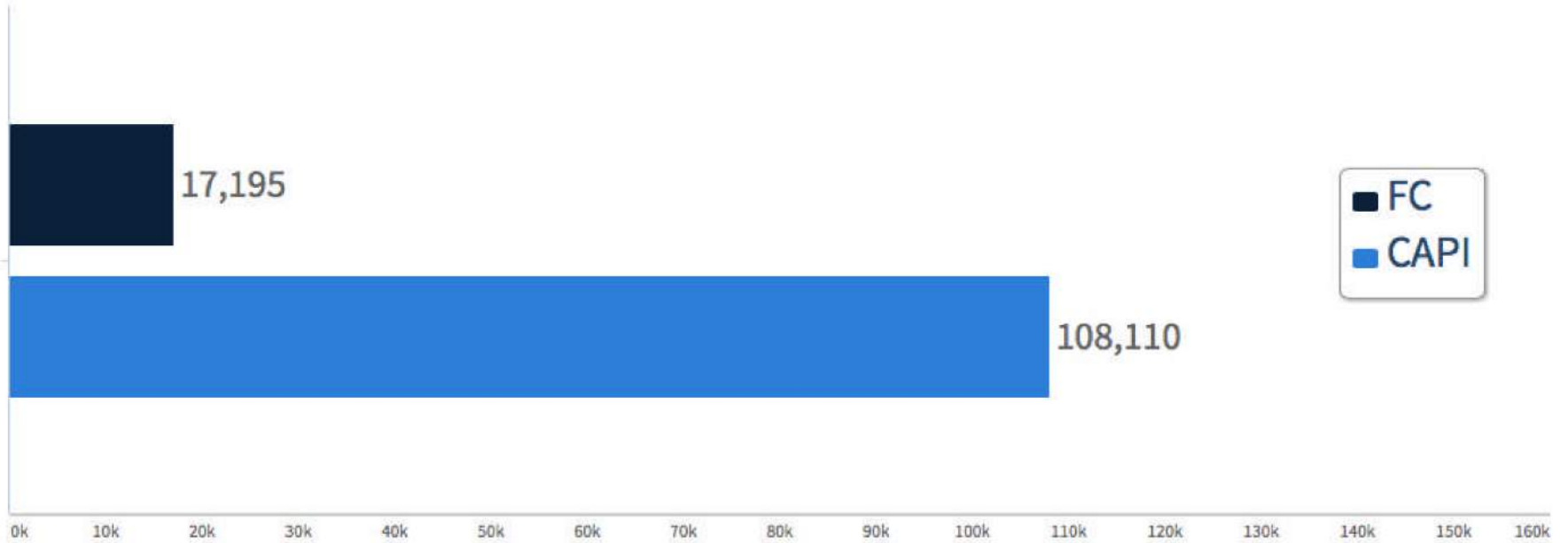
36

PCIe Java GZIP FPGA

- Compression tests

File	Original File Size	Compressed File Size	Compress Ratio	zlib Inflate Time (ms)	FPGA Inflate Time (ms)	FPGA Throughput	FPGA Speedup
asyoulik.txt	125179	59320	2.11	0.59	0.040	3.16 GB/s	14.99X
alice29.txt	152089	65188	2.33	0.67	0.047	3.21 GB/s	14.06X
lcet10.txt	426754	172770	2.47	1.77	0.129	3.31 GB/s	13.75X
plrabn12.txt	481861	241161	2.0	2.33	0.147	3.28 GB/s	15.87X
ptt5	513216	67529	7.6	1.69	0.150	3.42 GB/s	11.27X
pi.txt	1000000	662248	1.51	4.67	0.334	2.99 GB/s	13.98X
world192.txt	2473400	840481	2.94	9.01	0.739	3.35 GB/s	12.20X
bible.txt	4047392	1421406	2.85	15.17	1.205	3.36 GB/s	12.58X
e.coli	4638690	1891365	2.45	16.76	1.391	3.33 GB/s	12.04X

CAPI Attached Flash Optimization - planned

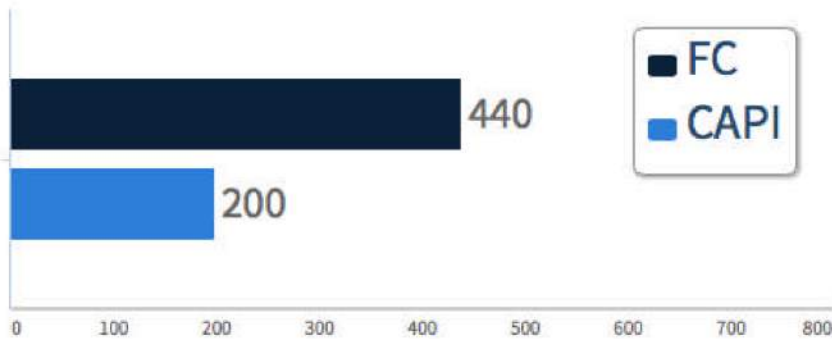- **Attach FlashSystem to POWER8 via CAPI coherent Attach**
- **Issues Read/Write Commands from applications to eliminate 97% of instruction path length**
  - **CAPI Flash controller Operates in User Space**
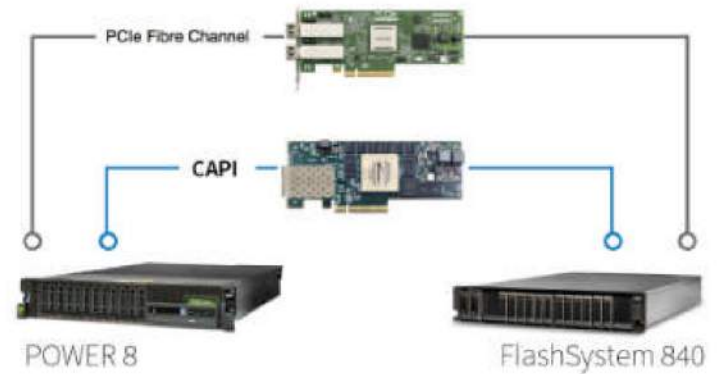- **Saves 10 Cores per 1M IOPs**

Application

Read/Write
Syscall

FileSystem

strategy()   iodone()

LVM

strategy()   iodone()

Disk & Adapter DD

Pin buffers,
Translate, Map
DMA, Start I/O

Interrupt,
unmap,
unpin,Iodone
scheduling

20K Instructions

< 500
Instructions

Application

Posix Async
I/O Style API

aio_read()
aio_write()

User Library

Shared
Memory Work
Queue

CAPI Flash vs Fibre Channel Flash

17,195 — FC

108,110 — CAPI

IOPS per HW thread →

(axis: 0k 10k 20k 30k 40k 50k 60k 70k 80k 90k 100k 110k 120k 130k 140k 150k 160k)

440 — FC

200 — CAPI

Latency [us] →

(axis: 0 100 200 300 400 500 600 700 800)

PCIe Fibre Channel

CAPI

POWER 8

FlashSystem 840

Power Systems: Innovation to put data to work

IBM

- **Processor and Memory**
  - Up to 12-core POWER8 processor card
  - Up to 8 Simultaneous Multi Thread per core
  - Transactional Memory
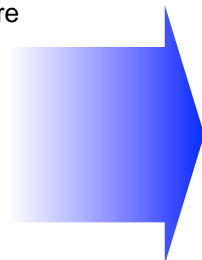 - Java Code Optimization w/HW Assist
- **I/O innovation**
  - PCIe Gen3
  - Coherent Accelerator Processor Interface
- **RAS**
  - RAID 0, 5, 6, 10 in the base
  - Concurrent maintenance PCIe Gen3 slots
  - Standard redundant power/cooling

POWER Scale-out Systems - Created for New Workloads & Enhanced Core Values

✓ **Provide efficient consolidation of business workloads**

✓ **Designed for demanding environments where uptime is critical and downtime is costly**

✓ **Deliver a higher value, more secure, and more reliable platform than competitors**

New 4-core

Power Systems S824

- 2-socket, 4U
- Up to 24 cores POWER8
- 1 TB memory
- 11 PCIe Gen 3
- AIX, IBM i, Linux
- Up to 1 CAPI support
- PowerVM

Power Systems S814

- 1-socket, 4U
- Up to 8 cores POWER8
- 512 GB memory
- 7 PCIe Gen 3
- AIX, IBM i, Linux
- Up to 1 CAPI support
- PowerVM

Power Systems S822

- 2-socket, 2U
- Up to 20 cores POWER8
- 1 TB memory
- 9 PCIe Gen 3
- AIX & Linux
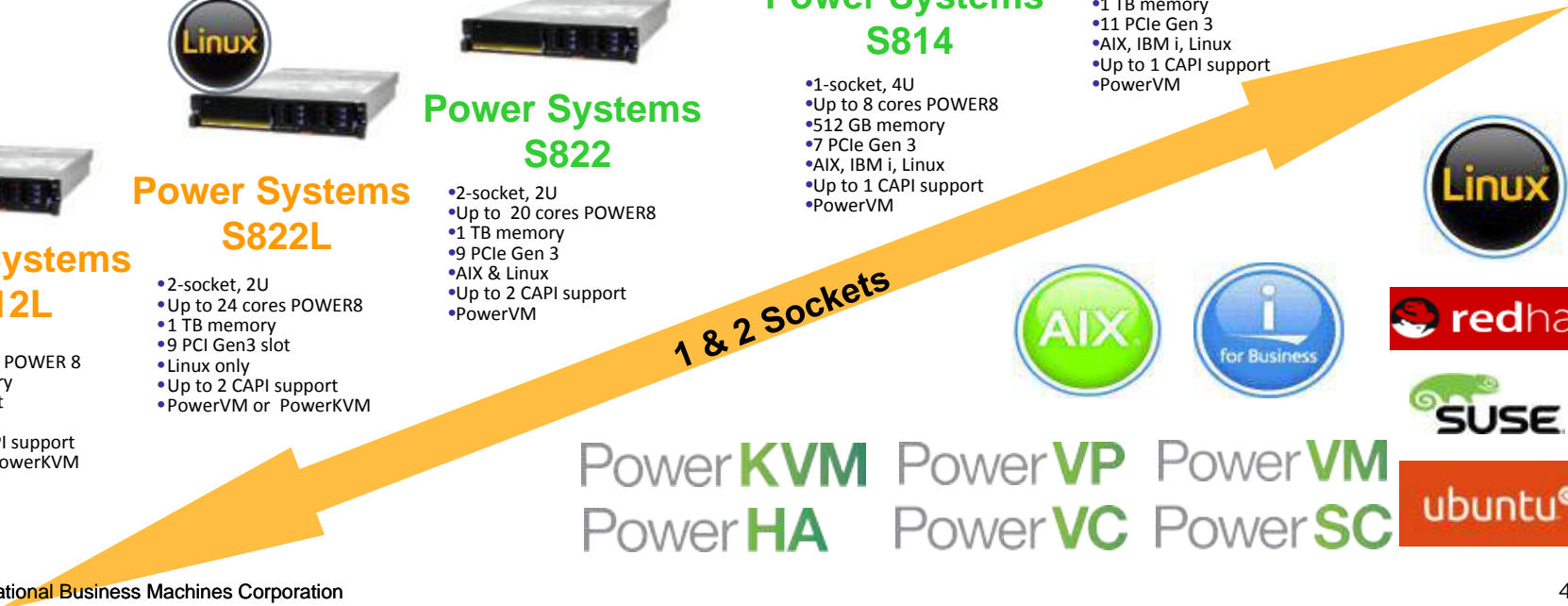- Up to 2 CAPI support
- PowerVM

Power Systems S822L

- 2-socket, 2U
- Up to 24 cores POWER8
- 1 TB memory
- 9 PCI Gen3 slot
- Linux only
- Up to 2 CAPI support
- PowerVM or PowerKVM

Power Systems S812L

- 1-socket, 2U
- Up to 12 cores POWER 8
- 512 GB memory
- 6 PCI Gen3 siot
- Linux only
- Up to One CAPI support
- PowerVM or PowerKVM

1 & 2 Sockets

Linux

AIX

i for Business

redhat

SUSE

ubuntu

Power**KVM** Power**VP** Power**VM**
Power**HA** Power**VC** Power**SC**

Servers

2U Servers

1&2 Socket Servers

New Scale-out Servers with POWER8 technology

- **1 socket :**   **4U**   **S814**
- **2 socket:**   **2U and 4U**   **S822 and S824**

Linux-only Power Systems (Not called "PowerLinux")

- **1 socket:**   **2U**   **S812L**
- **2 socket:**   **2U**   **S822L**

1 Socket	*1 Socket*	*2 Socket Option 1S*	*2 Socket Option 1S*	*2 Socket*
S814	*S812L*	*S824*	*S822*	*S822L*
8286-41A	*8247-21L*	*8286-42A*	*8284-22A*	*8247-22L*
4U	*2U*	*4U*	*2U*	*2U*

Order number = "number EIA/U" then "socket"

For example 8286 - 4 2 A

Power S812L

- Form Factor: 2U

- Sockets: 1S

- Single Socket
  - Cores: 10 / 12
  - Memory: Up to 512 GB / 8 DIMMs
  - Slots: 6 PCI Gen3 LP   (Hotplug)

- Ethernet: Quad 1 Gbit / (x8 slot)

- Media: DVD

- Internal Storage
  - Up to 12 SFF Bays
  - Up to 8 SFF Bays & Up to 6 1.8" SSD Bays with Easy Tier
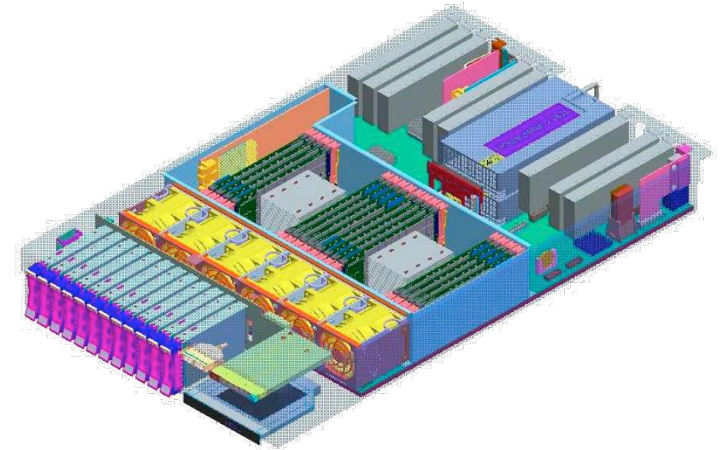
- Hypervisor: PowerVM or PowerKVM

Power S812L

Power Systems

IBM

PCIe Gen3 x8

PCIe Gen3 x16

Memory

POWER8

Memory

2xHMC
Serial
USB

8 SFF Bays

6 SSD Slots

Slim DVD

44
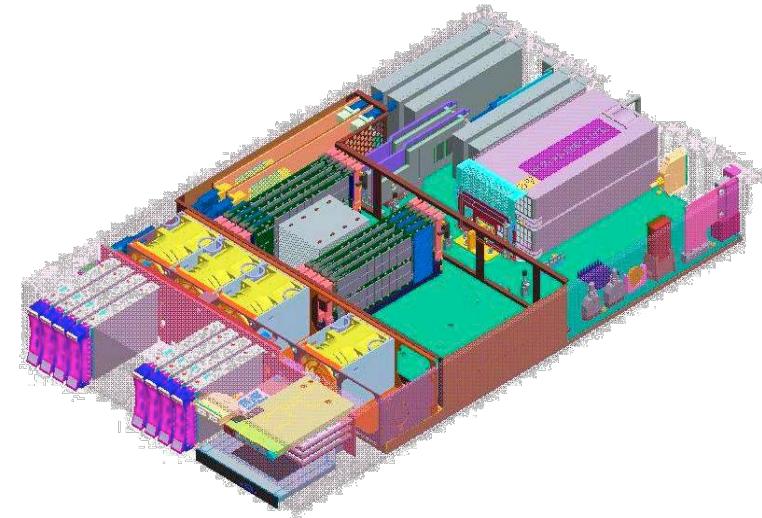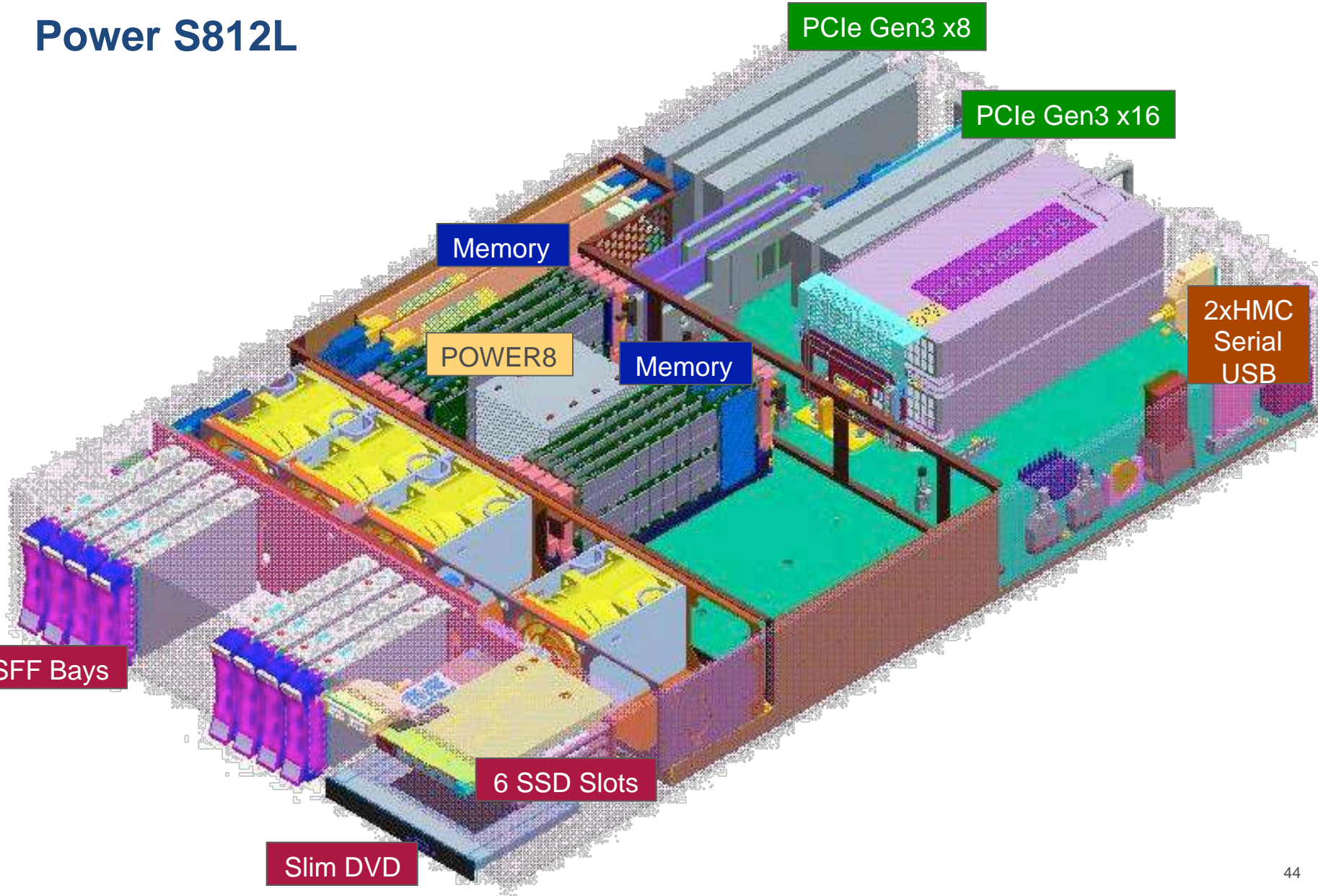
POWER S822 & Power S822L

- Sockets: 2S (1S upgradeable)
- Single Socket
  - Cores: 6 or 10
  - Memory: Up to 512 GB / 8 DIMMs
  - Slots: 6 PCI Gen3 LP   (Hotplug)
- Dual Sockets
  - Cores: 12 or 20 (2 x 6 or 10)
  - Memory: Up to 1 TB / 16 DIMMs
  - Slots: 9 PCIe Gen3 LP   (Hotplug)
- SRIOV support (SoD)
- Ethernet: Quad 1 Gbit / (x8 slot)
- Native IO: USB (2), Serial (2), HMC (2)
- Internal Storage
  - DVD
  - Up to 12 SFF Bays
  - Optional: Split Backplane: 6 + 6
  - Optional: Up to 8 SFF Bays & Up to 6 1.8" SSD Bays with Easy Tier
- Power Supplies: 2 (200-240 AVC)

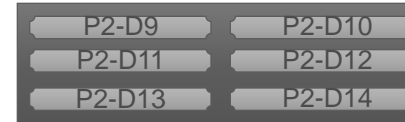# Power S822 - Storage Backplanes

P2-D1 P2-D2 P2-D3 P2-D4 P2-D5 P2-D6 P2-D7 P2-D8 P2-D9 P2-D10 P2-D11 P2-D12
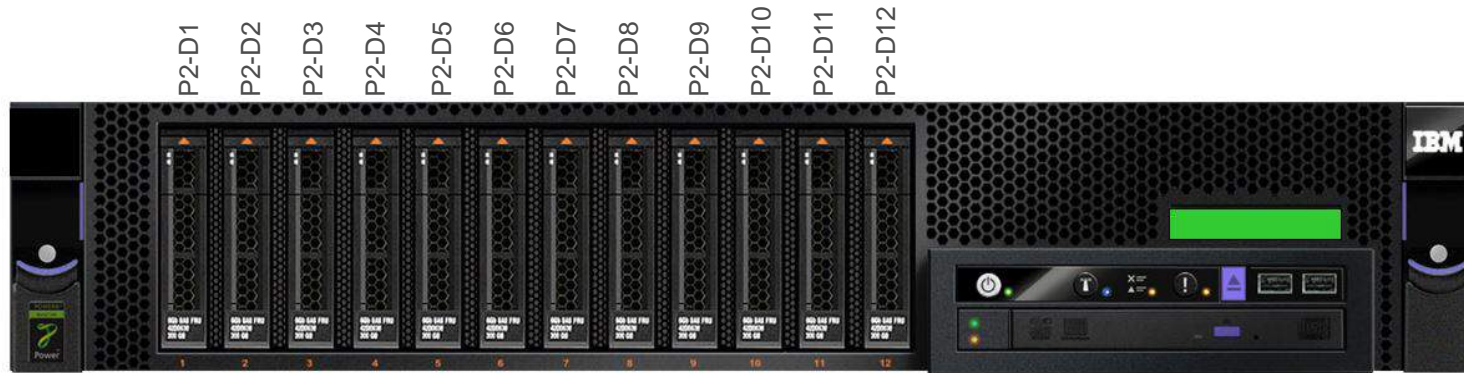
Backplane FC EJ0T - 12 SFF HDD/SSD Bays SFF-3

Op-panel + 2 USB Ports
DVD

- Optional Split Disk Backplane (FC EJ0V) – 6 + 6 Disks

SSD CAGE (FC EJTL)
(located behind the bezel)

P2-D9	P2-D10
P2-D11	P2-D12
P2-D13	P2-D14

P2-D1 P2-D2 P2-D3 P2-D4 P2-D5 P2-D6 P2-D7 P2-D8
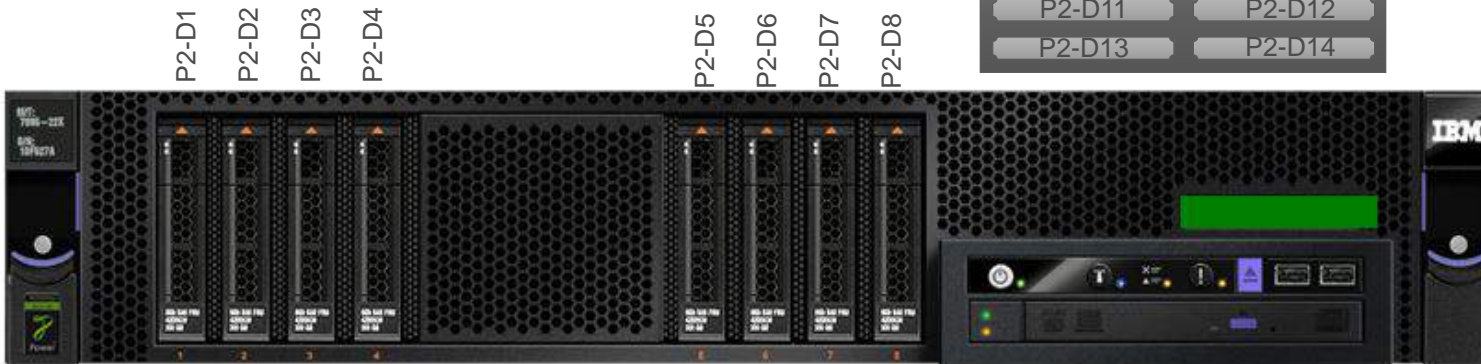
Backplane FC EJ0U - 8 SFF HDD/SSD Bays SFF-3 +
RAID 0,5,6,10,5T2,6T2,10T2

Op-panel + 2 USB Ports
DVD

- Optional 6-disk 1.8-inch SSD CAGE (FC EJTL)

46

Power S822 – SSD Cage

A

P8HAL513-0
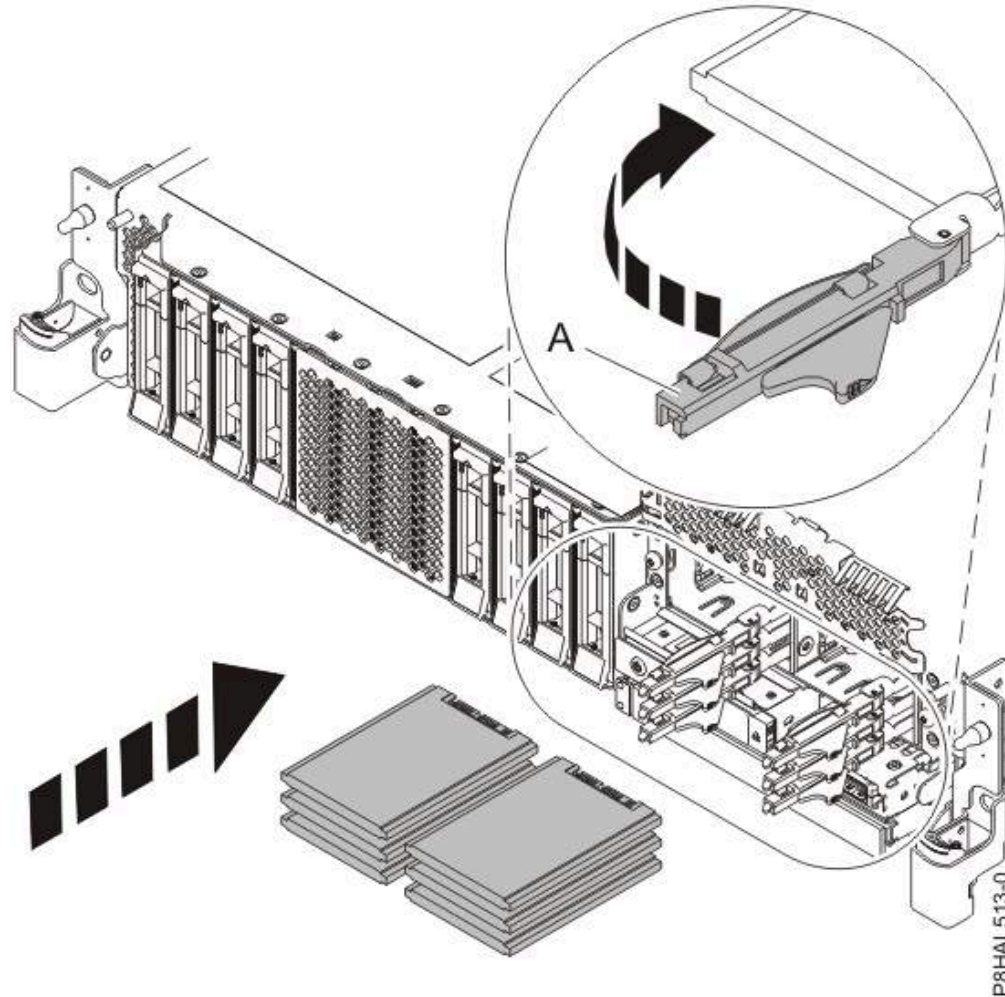
POWER S822 – Slots vs Standard I/O Backplane

Rear I/O Card
1x Serial / 2x USB / 2x HMC

P1-C2 – PCIe Gen3 x8 LP

P1-C3 – PCIe Gen3 x16 LP

P1-C5 – PCIe Gen3 x16 LP

2X Power Supplies

P1-C6 – PCIe Gen3 x16 LP

P1-C7 – PCIe Gen3 x16 LP

P1-C9– PCIe Gen3 x8 LP

4-Port
1Gbps LAN

P1-C10 – PCIe Gen3 x8 LP

P1-C11 – PCIe Gen3 x8 LP

P1-C12 – PCIe Gen3 x8 LP

Memory
4x DIMMs

POWER8

Memory
8x DIMMs

POWER8

Memory
4x DIMMs

P1-C14 – SAS Adapter Slot

Fan
Fan
Fan
Fan
Fan
Fan

Optional SSD
Slots

SFF Gen3
Slots

POWER S822 – Slots vs High-Performance I/O Backplane

Rear I/O Card
Serial / 2x USB / 2x HMC

P1-C2 – PCIe Gen3 x8 LP

P1-C3 – PCIe Gen3 x16 LP

P1-C5 – PCIe Gen3 x16 LP

2X Power Supplies

P1-C6 – PCIe Gen3 x16 LP

P1-C7 – PCIe Gen3 x16 LP

RAID & Dual Write Cache

2X SAS Ports

4-Port 1Gbps LAN

P1-C10 – PCIe Gen3 x8 LP

P1-C11 – PCIe Gen3 x8 LP

P1-C12 – PCIe Gen3 x8 LP

Memory 4x DIMMs

POWER8

Memory 8x DIMMs

POWER8

Memory 4x DIMMs

P1-C15 – SAS Adapter Slot

P1-C14 – SAS Adapter Slot

Fan

Fan

Fan

Fan

Fan

Fan

Optional SSD Slots

SFF Gen3 Slots

Power S822 – Back view

Not used

C2   C3   C5

Not used

C6   C7   C9   C10   C11   C12

2X USB
Serial Port
HMC Ports

Dual Power Supplies

SAS Port
or PCIe x8

Power S822 compared to Power 730

	Power 730	Power System S822
Processor	POWER7+	POWER8
Sockets	2	1 (Upgradeable) / 2
Cores	8 / 12 / 16	6 / 10 or 12 / 20
Maximum Memory	512 MB @ 1066 MHz	256 / 512 GB @ 1600 MHz
Memory Cache	No	Yes
Memory Bandwidth	136 GB/sec	192 / 384 GB/sec
Memory DRAM Spare	No	Yes
IO Expansion Slots	Dual GX++	SoD
PCIe slots	5 PCIe x8 LP	4 / 5 PCIe x8 LP 2 / 4 PCIe x16 LP
CAPI (Capable slots)	N / A	One / Two
PCIe Hot Plug Support	No	Yes
IO bandwidth	40 GB/sec	192 GB/sec
Ethernet ports	Quad 1 Gbt	Quad 1 Gbt
SFF	6	12
Easy Tier Support	No	Yes
Integrated split backplane	No	Yes (6 + 6)
SRIOV Support	No	SoD
Service Processor	Generation 1	Generation 2

Servers

4U Servers

Power S814

- Form Factor: 4U or Tower

- Sockets: 1S

- Single Socket
  - Cores: 4, 6 or 8
  - Memory: Up to 512 GB / 8 DIMMs
  - Slots: 6 PCI Gen3 LP   (Hotplug)

- SRIOV support (SoD)

- Ethernet: Quad 1 Gbit / (x8 slot)

- Native IO: USB (3), Serial (2), HMC (2)

- Internal Storage
  - DVD
  - Up to 12 SFF Bays
  - Optional: Split Backplane: 6 + 6
  - Optional: Up to 18 SFF Bays & Up to 8 1.8" SSD Bays with Easy Tier

- Power Supplies: 2   (110 or 200-240 AVC)

- 4-core specially attractive to IBM i customers because of P05 pricing tier
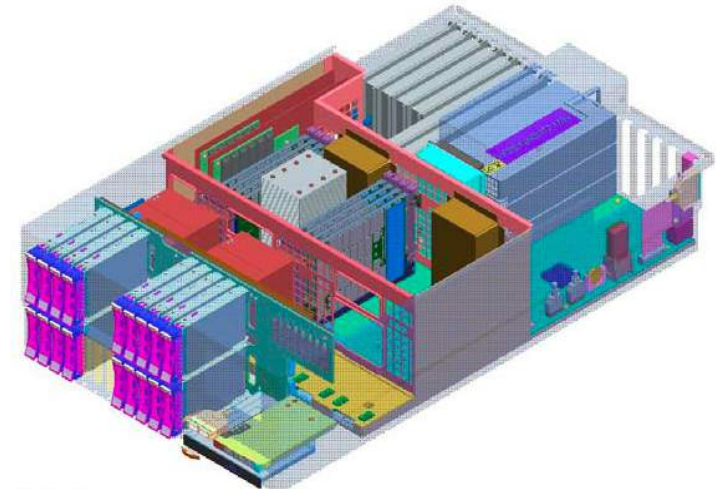
Power S814
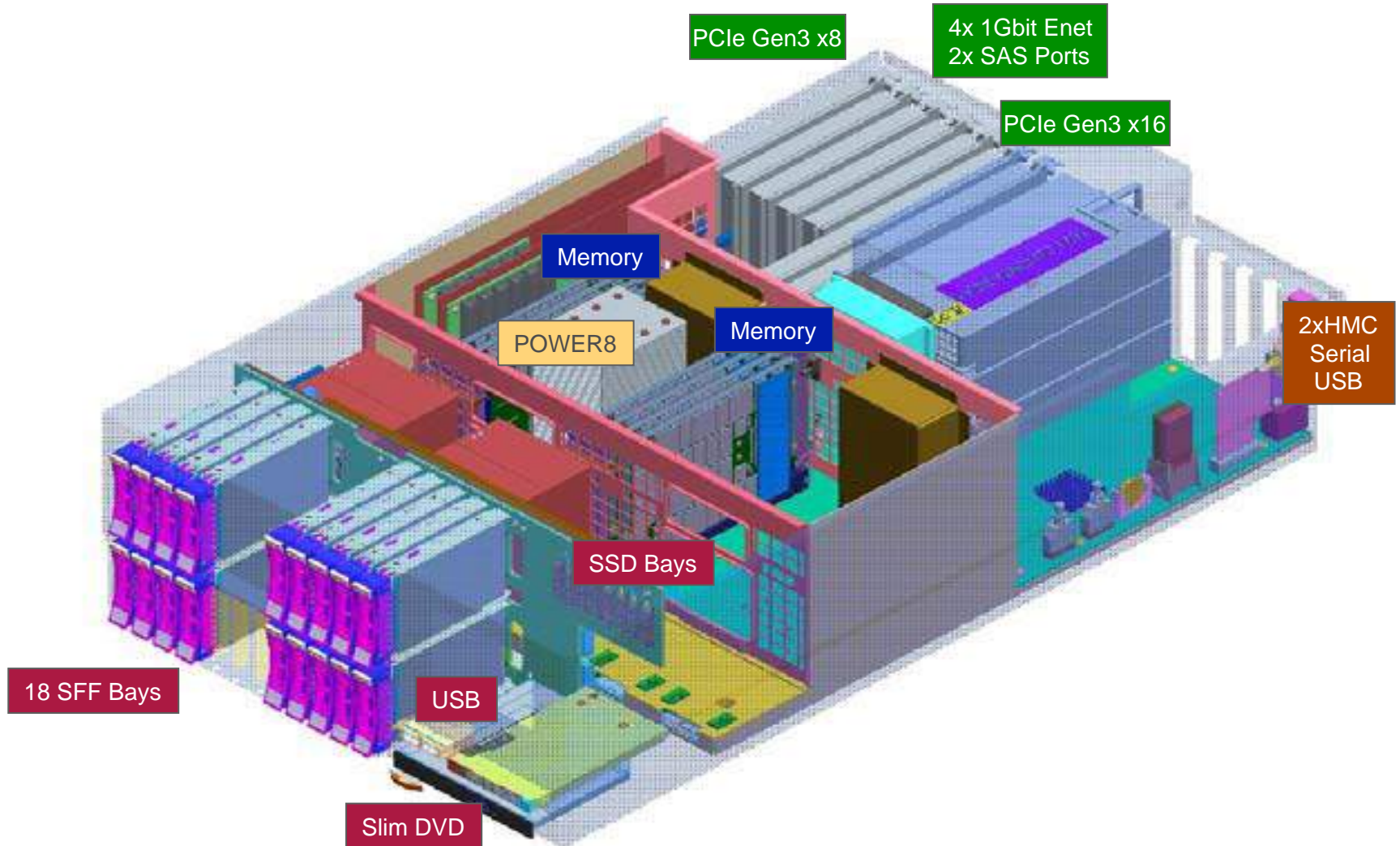
4-Core S814

- Four-core server for AIX, IBM i and Linux.
- Especially interesting to IBM i clients because of its P05 software tier
- 3 GHz cores
- Like the POWER7 4-core offering – no feature code I/O drawers like #5887 and no future PCIe expansion drawer. (Fibre Channel attached drawers/ SANs supported.)
- Like the POWER7 4-core – **64GB memory maximum.** 4-core POWER8 uses only 16GB memory features
- Max of ten SFF-3 drives in system unit (SSD or disk or mix) -- POWER7 4-core maximum is eight.   POWER8 4-core is 25% more drives as of July.
- Supports 15k 139/146GB and 15k 283/300GB disk drives and supporting 387GB SSD.   Not supporting other drive sizes.
- 4-core S814 offers better price/performance than 4-core 720 and higher function, but the absolute price of S814 4-core option higher than 720 4-core

POWER S824

- Form Factor: 4U
- Sockets: 2S (1S upgradeable)
- Single Socket
    - Cores: 6 or 8
    - Memory: Up to 512 GB / 8 DIMMs
    - Slots: 7 PCI Gen3 (Hotplug)
- Dual Sockets
    - Cores: 12, 16, or 24 (2 x 6, 8, or 12)
    - Memory: Up to 1 TB / 16 DIMMs
    - Slots: 11 PCIe Gen3 (Hotplug)
- SRIOV support (SoD)
- Ethernet: Quad 1 Gbit / (x8 slot)
- Native IO: USB (3), Serial (2), HMC (2)
- Internal Storage
    - DVD
    - Up to 12 SFF Bays
    - Optional: Split Backplane: 6 + 6
    - Optional: Up to 18 SFF Bays & Up to 8 1.8" SSD Bays with Easy Tier
- Power Supplies: 4 (200-240 AVC)
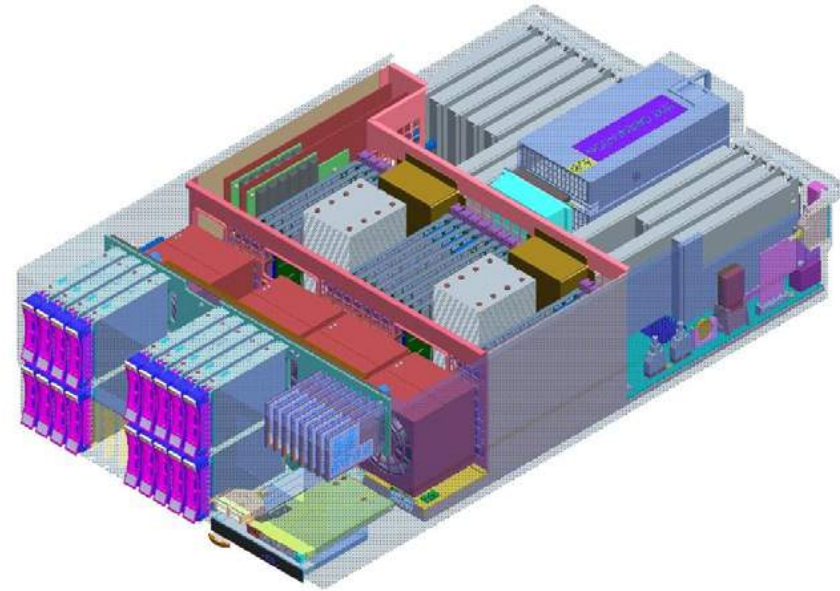
3 Yr Warranty

Power S824 - Storage Backplanes

Backplane FC EJ0N - 12 SFF HDD/SSD Bays SFF-3

- Optional Split Disk Backplane (FC EJ0S) – 6 + 6 Disks

P2-D1 P2-D2 P2-D3 P2-D4 P2-D5 P2-D6 P2-D7 P2-D8 P2-D9 P2-D10 P2-D11 P2-D12

Op-panel + 2 USB Ports
DVD

Power S824 – Storage Backplanes

Backplane FC EJ0P - 18 SFF HDD/SSD Bays SFF-3 + RAID 0,5,6,10,5T2,6T2,10T2

- Optional 8-disk 1.8-inch SSD CAGE (FC EJTM)

SSD CAGE (FC EJTM)
(located behind the bezel)
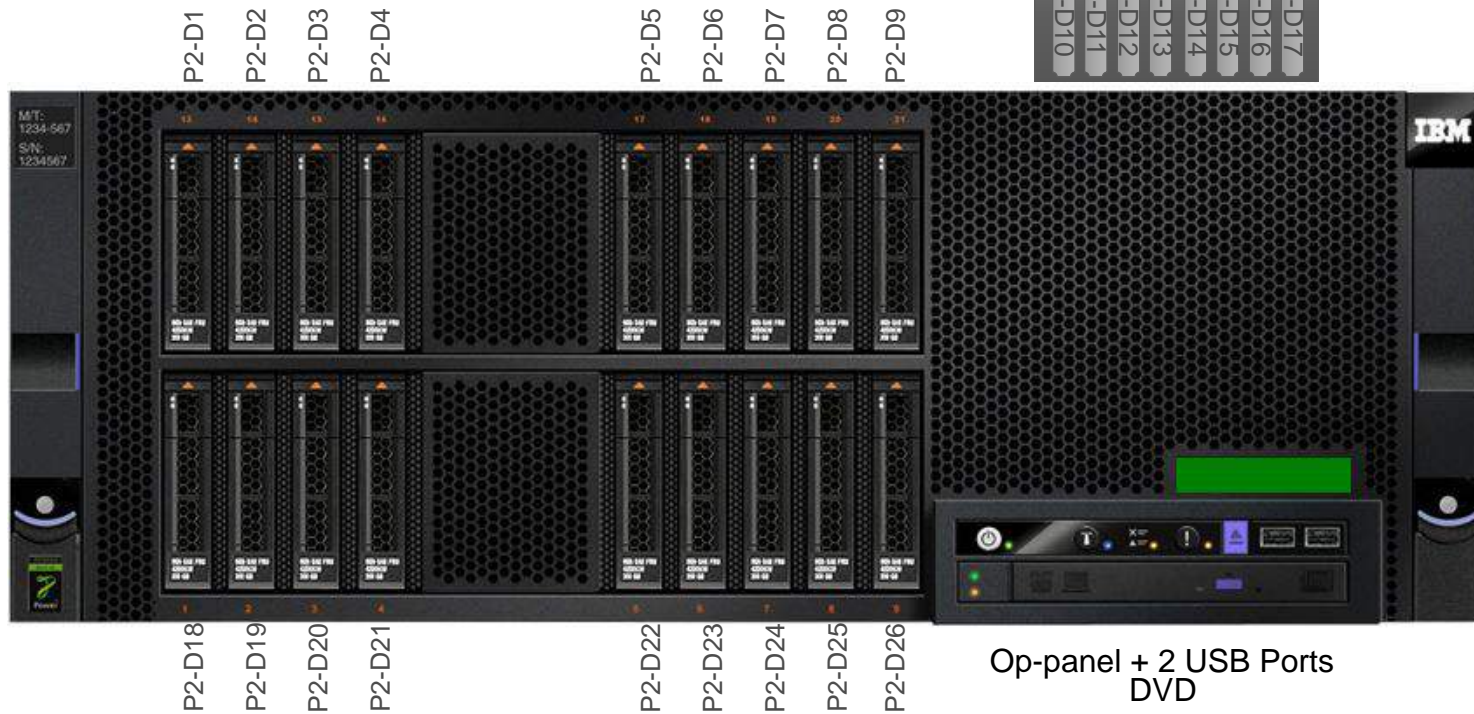
Op-panel + 2 USB Ports
DVD

Power S824 – Storage Backplane

Power S824 – Slots vs Standard I/O Backplane

Optional SSD Bays

Memory
4x DIMMs

P1-C2 – PCIe Gen3 x8

P1-C3 – PCIe Gen3 x16

P1-C4 – PCIe Gen3 x8

P1-C5 – PCIe Gen3 x16

FAN

POWER8

4X Power Supplies

FAN

P1-C6 – PCIe Gen3 x16

P1-C7 – PCIe Gen3 x16

FAN

P1-C8 – PCIe Gen3 x8

P1-C9 – PCIe Gen3 x8

POWER8

SFF Bays

4-Port 1Gbps
LAN

P1-C10 – PCIe Gen3 x8

Memory
4x DIMMs

P1-C11 – PCIe Gen3 x8

FAN

P1-C15 – SAS Adapter Slot

P1-C12 – PCIe Gen3 x8

P1-C14 – SAS Adapter Slot

Standard SAS Adapters

POWER S824 – Slots vs High-Performance I/O Backplane

Optional SSD Bays

P1-C2 – PCIe Gen3 x8

P1-C3 – PCIe Gen3 x16

P1-C4 – PCIe Gen3 x8

P1-C5 – PCIe Gen3 x16

Memory 4x DIMMs

POWER8

FAN

FAN

FAN

4X Power Supplies

P1-C6 – PCIe Gen3 x16

P1-C7 – PCIe Gen3 x16

P1-C8 – PCIe Gen3 x8

P1-C9 – PCIe Gen3 x8

P1-C10 – PCIe Gen3 x8

POWER8

SFF Bays

Memory 4x DIMMs

4-Port 1Gbps LAN

2X SAS Ports

P1-C15 – SAS Adapter Slot

P1-C14 – SAS Adapter Slot

FAN

P1-C12 – PCIe Gen3 x8

High Performance SAS cards
RAID & •Dual Write Cache

Power S814 & Power S824 – Back view

Power S814 and S824 compared to Power 740

	Power System 740	Power System S814	Power System S824
Processor Offerings	*4, 6, or 8 core @ 3.6 GHz* *1x or 2x 6-core 4.2 GHz* *1x or 2x 8-core 3.6 / 4.2 GHz*	*6-core 3.02 GHz* *8-core 3.69 GHz*	*6 core @ 3.02 GHz* *8 core @ 3.69 GHz* *1x or 2x 6 -core@ 3.89 GHz* *1x or 2x 8 -core 4.15 GHz* *2x 12 -core 3.52 GHz*
DDR3 Memory features	*8 / 16 / 32 / 64GB* *8GB to 256GB / socket*	*16 / 32 / 64GB* *16GB to 512GB*	*16 / 32 / 64GB* *32GB to 1024GB*
OS Support	*AIX: 5.3 / 6.1 / 7.1* *IBM i: 6.1 & 7.1* *Linux: RHEL & SLES*	*AIX: 6.1 & 7.1* *IBM i: 7.1 & 7.2* *Linux: RHEL & SLES*	*AIX: 6.1 & 7.1* *IBM i: 7.1 & 7.2* *Linux: RHEL & SLES*
DASD / Bays	*Up to 6 HDD or SSD*	*Up to 12 SFF or SSD*	*Up to 12 SFF or SSD*
PCIe Expansion Slots	*Five x8 G2 LP*	*Two x16 G3 LP Hot Plug* *Five x8 G3 LP Hot Plug* *1x CAPI capable*	*Four x16 G3 LP Hot Plug* *Seven x8 G3 LP Hot Plug* *2x CAPI capable*
Integrated SAS/SATA Cntrl	*Standard: RAID 0, 1, & 10* *Optional: RAID 5 & 6*	*Standard RAID 0, 5, 6, 10*	*Standard RAID 0, 5, 6, 10*
Ethernet	*Quad 10/100/1000*	*Quad 10/100/1000*	*Quad 10/100/1000*
Media Bays	*1 Slim-line* *& 1 Half Height (Optional)*	*1 Slim-line*	*1 Slim-line*
IO Drawers	*0 on 1S , 2 on 2S*	*N / A*	*N / A*
Power requirement	*100V to 240V AC 1S* *200 to 240V AC 2S*	*100V to 240V AC* *200V to 240V AC*	*200V to 240V AC*
Red Power & Cooling	*Optional / Standard*	*Optional*	*Standard*
Warranty	*3 Years*		

Physical Planning – Tower Acoustic Considerations

S814 Tower considerations: The dB level of the tower meets general business standards and can be installed in an open business environment in addition to a machine room. It may be perceived as noisy to some clients who wish to seat people nearby.

Also note that some PCIe adapters draw more power and require more cooling. The POWER8 server recognizes these adapters and speeds up fans to provide this cooling (airflow). **Faster fans SIGNIFICANTLY increases the noise level**. To avoid potential acoustic situations, these adapters are only supported in the S814 when it is in a rack configuration.

Adapters as of April 2014 requiring more cooling:

- PCIe3 SAS adapters (#EJ0J, #EJ0M, #EL3B, #EJ0L)
- PCIe2 SAS adapters (#ESA1, #ESA2, #ESA3, and #5913)

Physical Planning – Turbo Mode and Acoustics

If the server is enabled to run Turbo mode, the fans run a bit faster and there is a noticeably increased noise level (5-10 dB estimate). If the server actually enters Turbo mode, the fans speed increase again and the noise increases noticeably again.

Enabling Turbo mode on tower configurations is not recommended for acoustic reasons unless the tower is in a machine room or is some distance away from people in an office environment.

If higher energy PCIe adapters are also present, the fans REALLY significantly increase in speed and in noise level.

PCIe2 2-port 10/1GbE BaseT RJ45 Adapter
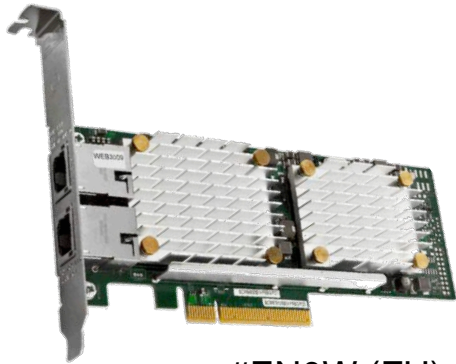
#EN0W (FH) / #EN0X (LP)
CCIN 2CC4

Only $950
Plus inexpensive cabling

Supported on POWER8 Systems

Operating System support:
- AIX 6.1 & AIX 7.1 or later
- IBM i via VIOS
- Linux
  - ❖ RHEL 6 or later
  - ❖ SLES 11 or later
- PowerKVM Host
- VIOS

Dual port 10GBase-T   RJ45 ports -   Ethernet NIC traffic

Ports default to auto negotiate the highest speed:
- 10Gb   (10GBaseT), 1Gb (1000BaseT) or 100Mb (100BaseT) full duplex.

Port's configuration are independent of the one other, but both ports owned by one partitions or one VIOS

CAT-6A cabling supported

NIM/Linux Install support announced July 2014 (planned GA 29 Aug)

Prices are USA list prices and are subject to change. Reseller prices may vary.

PCIe3 12GB Cache RAID SAS Adapter Quad-port 6Gb

- Dedicated Slots
- 3GB Cache
  - Cache compression
  - Up to 12GB Logical Cache
- Easy Tier
- Redundant Controllers
- External SAS Ports for ONE EXP24S
- RAIDs 0,1,5,6,10 & 5T2, 6T2 & 10T2
- No support for:
  - Split backplane
  - Tape
- Physically capable of more than 1M IOPS on RAID0
  - 900k IOPS on RAID5
- Limited by the number of disks we attach to a server

Easy Tier for SAS RAID Adapters

- Provides the benefits of Easy Tier and performance advantages of DAS
- Accentuates the Integrated Tier Storage (SSD/HDD) design of POWER8
- The RAID storage subsystem tunes itself based on the workload characteristics of the system to optimize performance
- Adapter dynamically swaps data between tiers to achieve lowest response times and highest throughput
- Swaps happen automatically, without manual user intervention
- Hottest (best) data moved first
- Reacts to workload changes in Real Time, and in seconds/minutes (not hours)
- Improves cost/performance tradeoff between HDDs/SSDs
- GA 6/2014
  - Supported in POWER7+/POWER8 with FC=EJOL / CCIN=57CE (PCIe3 12 GB Cache RAID SAS adapter Quad-port 6 Gb x8)
  - Supported in POWER8 with FC=EJ0P / CCIN=57D8 (PCIe3 x8 Cache SAS RAID internal adapter 6 Gb)
- Supported by AIX, Linux, and VIOS
  - IBM i supported only via VIOS, requires IBM I 6.1 or later and VIOS 2.2.3.2 or later

Easy Tier for SAS RAID Adapters

Hot Data Tier

SSDs

Tiered Array

Cold Data Tier

HDDs

Data Bands (1MB or 2MB Blocks)

Hot Data

Cold Data
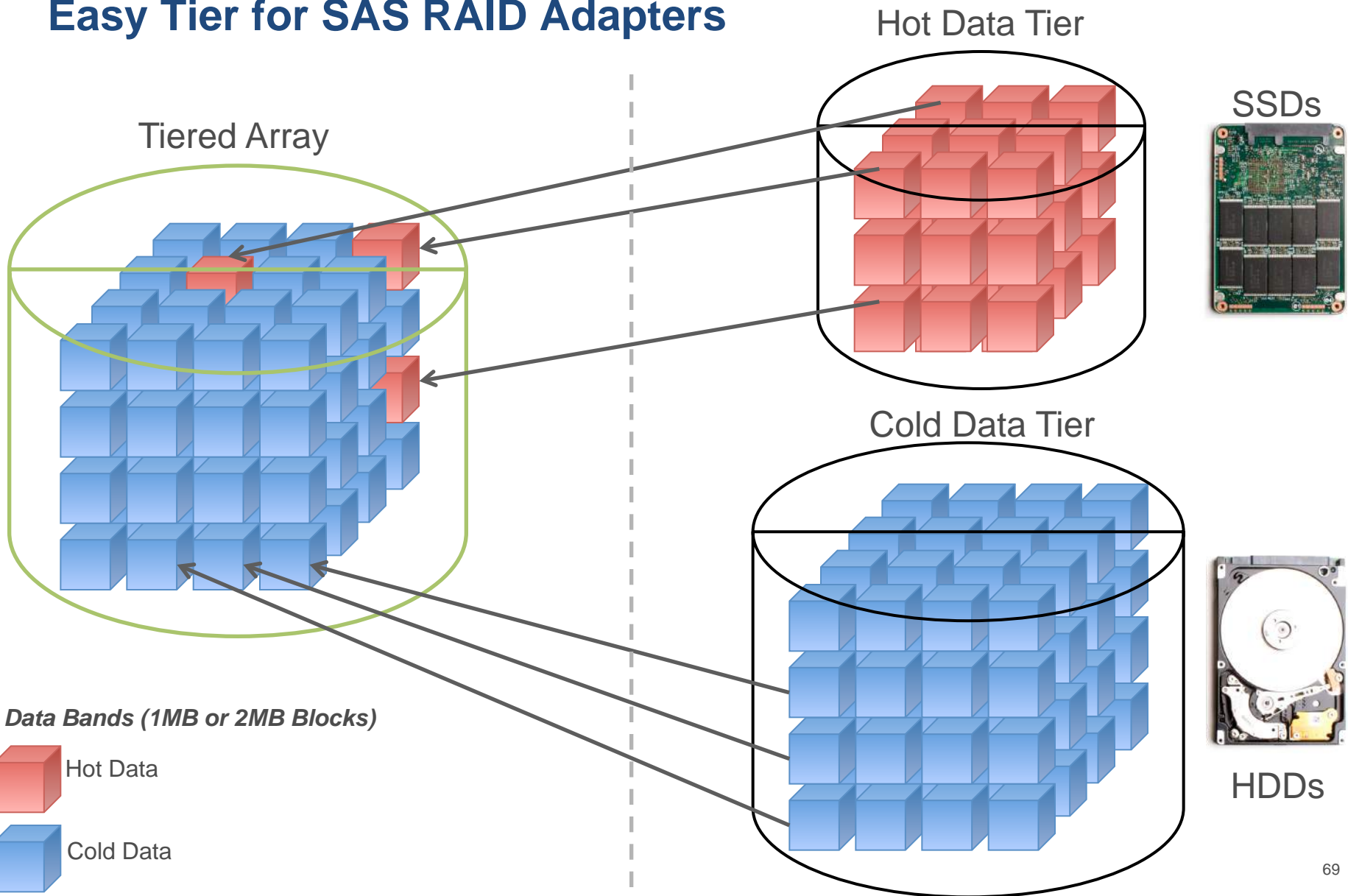
Easy Tier Performance

AIX Stock Trading Application Performance (RAID 5T2, 3 SSDs, 16 HDDs)

ET array Tp started ~2X higher than all HDD array @ application's start
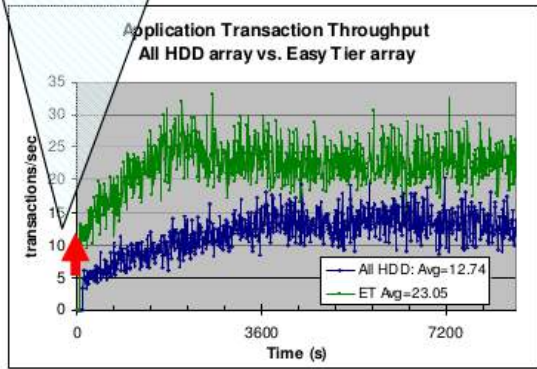
ET array reached peak throughput ~24 minutes sooner
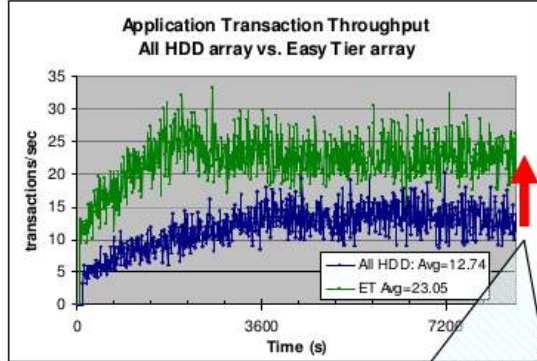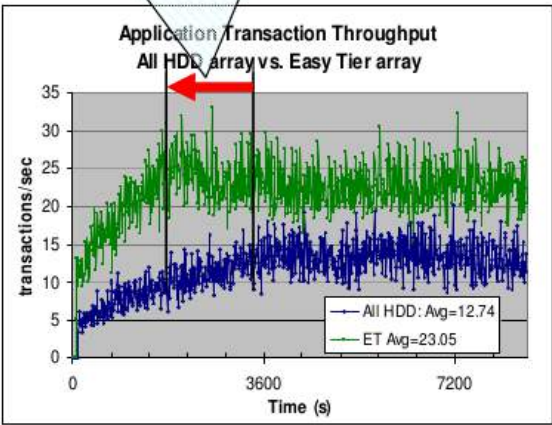
At steady state ET array Tp was ~2X all HDD Array
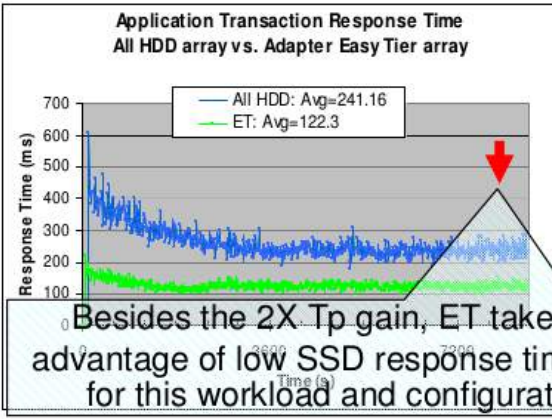
Besides the 2X Tp gain, ET takes full advantage of low SSD response times, 2X for this workload and configuration.

Creating a Tiered Array

```
                    IBM SAS Disk Array Manager

Move cursor to desired item and press Enter.

   List SAS Disk Array Configuration
   Create an Array Candidate pdisk and Format to RAID block size
   Create a SAS Disk Array
   Delete a SAS Disk Array
   Add Disks to an Existing SAS Disk Array
   Migrate an Existing SAS Disk Array to a New RAID Level
   Configure a Defined SAS Disk Array
   Change/Show Characteristics of a SAS Disk Array
   Manage HA Access Characteristics of a SAS Disk Array
   Reconstruct a SAS Disk Array
   Change/Show SAS pdisk Status
   Diagnostics and Recovery Options

   +-----------------------------------------------------------------+
   |                    Select a RAID Level                          |
   |    Move cursor to desired item and press Enter.                 |
   |                                                                 |
   |      0                                                          |
   |      5                                                          |
   |      10                                                         |
   |      6                                                          |
   |      5T2                                                        |
   |      6T2                                                        |
   |      10T2                                                       |
   |                                                                 |
   |      F1=Help              F2=Refresh            F3=Cancel       |
   |      F8=Image             F10=Exit              Enter=Do        |
 F1=Help  /=Find              n=Find Next                           |
 F9=Shel -------------------------------------------------------------
```

Tiered arrays on AIX

```
----------------------------------------------------------------------------
Name       Resource   State       Description                       Size
----------------------------------------------------------------------------
sissas1    FEFFFFFF   Primary     PCIe3 12GB Cache RAID SAS Adapter Quad-port 6Gb x8
sissas0    FEFFFFFF   HA Linked   Remote adapter SN  00325001

hdisk1     FC0000FF   Optimal     RAID 5T2 Array (N/N)             773.5GB
 pdisk0    000400FF   Active      Array Member                     139.6GB
 pdisk1    000401FF   Active      Array Member                     139.6GB
 pdisk2    000402FF   Active      Array Member                     139.6GB
 pdisk3    000403FF   Active      Array Member                     139.6GB
 pdisk7    000407FF   Active      SSD Array Member                 177.8GB
 pdisk6    000406FF   Active      SSD Array Member                 177.8GB
 pdisk8    000408FF   Active      SSD Array Member                 177.8GB

hdisk2     FC0100FF   Optimal     RAID 6T2 Array (N/N)              1090GB
 pdisk10   00040AFF   Active      SSD Array Member                 387.9GB
 pdisk11   00040BFF   Active      SSD Array Member                 387.9GB
 pdisk4    000404FF   Active      Array Member                     139.6GB
 pdisk20   000414FF   Active      SSD Array Member                 387.9GB
 pdisk21   000415FF   Active      SSD Array Member                 387.9GB
 pdisk9    000409FF   Active      SSD Array Member                 177.8GB
 pdisk5    000405FF   Active      Array Member                     139.6GB
 pdisk12   00040CFF   Active      Array Member                     139.6GB
 pdisk13   00040DFF   Active      Array Member                     139.6GB
 pdisk14   00040EFF   Active      Array Member                     139.6GB
 pdisk15   00040FFF   Active      Array Member                     139.6GB

hdisk3     FC0200FF   Optimal     RAID 10T2 Array (0/0)            666.6GB
 pdisk22   000416FF   Active      SSD Array Member                 387.9GB
 pdisk23   000417FF   Active      SSD Array Member                 387.9GB
 pdisk16   000410FF   Active      Array Member                     139.6GB
 pdisk17   000411FF   Active      Array Member                     139.6GB
 pdisk18   000412FF   Active      Array Member                     139.6GB
 pdisk19   000413FF   Active      Array Member                     139.6GB
```

PCIe Slots - Memorize This

	4U		2U (no IBM i)	
	1S 4U	2S 4U	1S 2U	2S 2U
Total PCIe slots **All hot swap** **All Gen3**	7	11	6	9

- Use 1 of the above slots for a LAN adapter
- Use 1 of the above for expanded function backplane
- PCIe expansion drawers are not yet available.

PCIe Slots - High Level

	4U		2U (no IBM i)	
	1S 4U	2S 4U	1S 2U	2S 2U
Total PCIe slots (all hot swap)	7	11	6	9
Required* LAN adapter (available for client use)	1	1	1	1
PCIe slots after required* LAN adapter	6	10	5	8
However if use high performance, expanded function backplane	-1	-1	-1	-1
PCIe slots after required* LAN and if using high performance backplane	5	9	4	7

* required for IBM Manufacturing

- PCIe slots are all Gen3 slots
- 2U are all low profile   and   4U are all full high
- There is no PCI expansion drawer announced. There is an SOD.

PCIe Gen3

Gen1 x8

Gen2 x8

Gen3 x8

- Though these cards physically look the same … and fit in the same slots
- Gen3 cards/slots have up to 2X more **bandwidth** than Gen2 cards/slots
- Gen3 cards/slots have up to 4X more **bandwidth** than Gen1 cards/slots
    - More virtualization
    - More consolidation ⎬ saving PCI slots and I/O drawers
    - More ports per adapter
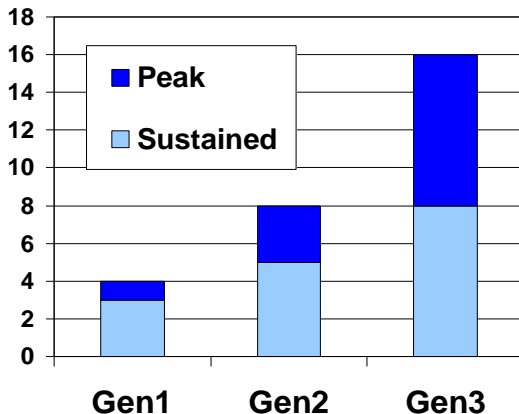
A Gen1 x8 PCIe adapter has a theoretical max (peak) bandwidth of 4 GB/sec.
A Gen2 x8 adapter has a peak bandwidth of 8 GB/sec.
A Gen3 x8 adapter has a peak bandwidth of 16 GB/sec.

PCIe x8 and x16

POWER8 servers have x8 AND x16 PCIe slots

Compared to POWER7+ PCIe Gen2 x8 slot, a POWER8 PCIe Gen3 x16 slot has a peak bandwidth of 4X   (2X going Gen2 to Gen3 plus 2X going x8 to x16)

x8

x16

x1   x4   x8   x16

POWER8

POWER7+
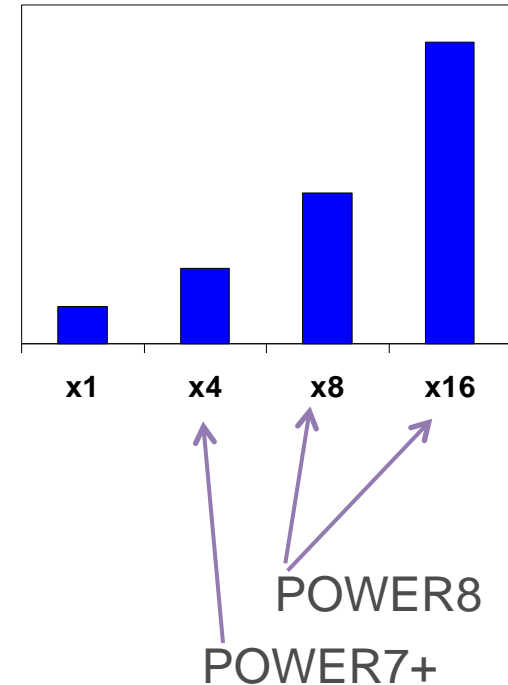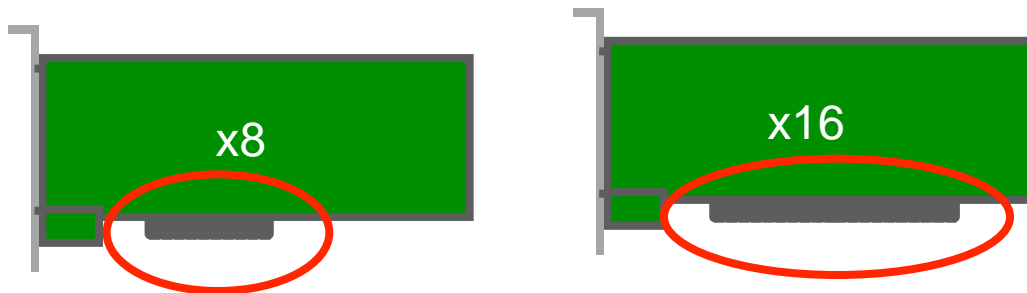
- x16 slot/card has more connections than a x8 slot/card
- "x16" or "x8" refers to the number of lanes. 16Gbps/lane (Full Duplex)
- A x8 card can be placed in a x16 slot, but only uses half the connections

Standard Backplane

SAS

SAS

SAS PCIe Slot

SFF
SFF
SFF
SFF
SFF
SFF

SFF
SFF
SFF
DVD
SFF
SFF
SFF

Standard Backplane + SPLIT

SAS

SAS

SAS PCIe Slot

SFF
SFF
SFF
SFF
SFF
SFF

SAS

SAS

SAS PCIe Slot

SFF
SFF
SFF
DVD
SFF
SFF
SFF

2U – High Performance Backplane

RAID with Write Cache

SAS **SAS**

SAS PCIe Slot

RAID with Write Cache

SAS **SAS**

SAS PCIe Slot

SAS Expander

SAS Expander

Slot P1-C9

SAS External SAS Ports

SAS

Internal Connections

Support to ONE EXP24

SFF
SFF
SFF
SFF
SFF
SFF
SFF
SFF

DVD

SSD
SSD
SSD
SSD
SSD
SSD

4U – High Performance Backplane

Power Systems

RAID with Write Cache

SAS

SAS

SAS PCIe Slot

RAID with Write Cache

SAS

SAS

SAS PCIe Slot

SAS Expander

SAS Expander

DVD

SFF
SFF
SFF
SFF
SFF
SFF
SFF
SFF
SFF
SFF
SFF
SFF
SFF
SFF
SFF
SFF
SFF
SFF
SSD
SSD
SSD
SSD
SSD
SSD
SSD
SSD

Slot P1-C11

SAS

External
SAS Ports

SAS

Internal Connections

Support to ONE EXP24

PERFORMANCE

Benchmarks POWER8

- For the 1 and 2 socket POWER8 Systems, IBM has run several benchmarks

  - SAP SD 2-Tier
    - Database + App
  - SPECjEnterprise
    - Database + Java + Storage
  - Oracle Payroll
  - Siebel
  - SPEC
    - Integer
    - Float-point
  - rPerf
    - SMT-1
    - SMT-2
    - SMT-4
    - SMT-8

- All benchmarks show leadership performance for the POWER8 systems.

SAP SD 2-Tier

CERTIFICATION
SAP® Standard Application Benchmarks

The SAP Sales and Distribution (SD) Standard Application Benchmark performed on March 3, 2014, by IBM in Beaverton, OR, USA, was certified by SAP on behalf of the SAP Benchmark Council on April 28, 2014, with the following data:

Number of SAP SD benchmark users:	21,212
Average dialog response time:	0.98 seconds
Throughput:	
Fully processed order line items per hour:	2,317,330
Dialog steps per hour:	6,952,000
SAPS:	115,870
Average database request time (dialog/update):	0.011 sec / 0.019 sec
CPU utilization of central server:	99%
Operating system, central server:	AIX 7.1
RDBMS:	DB2 10.5
SAP Business Suite software:	SAP enhancement package 5 for SAP ERP 6.0

> 21.212 Users
> 4.828 SAPs/CORE
> 99% server usage
> 0,011 sec dialog

Configuration:
Central server: IBM Power System S824, 4 processors / 24 cores / 192 threads, POWER8, 3.52 GHz, 32 KB (I) and 64 KB (D) L1 cache and 512 KB L2 cache per core, 8 MB L3 cache per core, 512 GB main memory

SAP SD 2-Tier per Core

SAP SD 2-Tier - Performance / Core

System	Value
M6-32 - 384 x SPARC M6 3,6GHz	2.068
M5-32 - 192 x SPARC M5 3,6GHz	2.461
T5-8 - 128 x SPARC T5 3,6GHz	1.726
T5-4 - 64 x SPARC T5 3,6GHz	1.726
INTEL - 120 x E7-8890 v2 2,8GHz	2.164
INTEL - 60 x E7-4890 v2 2,8GHz	2.204
INTEL - 24 x E5-2697 v2 2,7GHz	2.255
Power S824 - 24 x POWER8 3,52GHz	4.828
Power 780 - 96 x POWER7+ 3,72GHz	3.247
Power 740 - 16 x POWER7 3,8GHz	2.975
Flex p270 - 24 x POWER7+ 3,41GHz	2.849
Flex p260 - 16 x POWER7+ 4,1GHz	3.419

2x-2,8x

2,15x

1,4x-1,7x

SAP SD 2-Tier – POWER8 vs x86

Total Server Performance - Best x86 vs POWER8
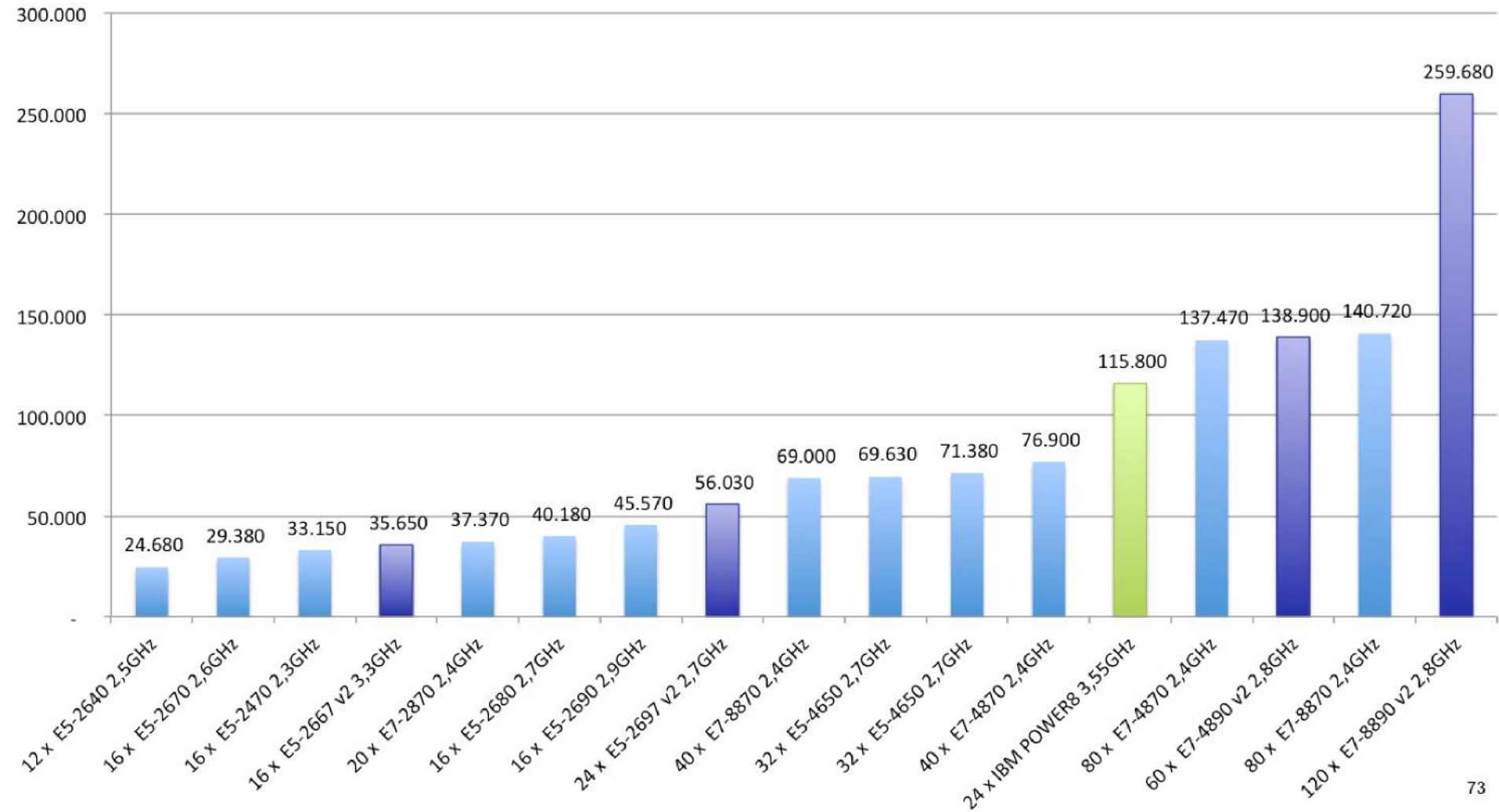
rPerf – Multiple SMT Levels

	SMT1	SMT2	SMT4	SMT8
Power S814				
6-core 3.0 GHz	48.3	70.1	91.1	97.5
8-core 3.7 GHz	71.4	103.5	134.5	143.9
Power S824				
6-core 3.8 GHz	59.9	86.9	112.9	120.8
12-core 3.8 GHz	116.8	169.4	220.2	235.6
8-core 4.1 GHz	82.3	119.3	155.1	166.0
16-core 4.1 GHz	160.4	232.7	302.4	323.6
24-core 3.5 GHz	209.1	303.2	394.2	421.8
Power S822				
6-core 3.8 GHz	59.9	86.9	112.9	120.8
12-core 3.8 GHz	116.8	169.4	220.2	235.6
10-core 3.4 GHz	88.2	127.8	166.2	177.8
20-core 3.4 GHz	171.9	249.3	324.0	346.7

rPerf / Core – POWER7 vs POWER8

rPerf / Core - P7 vs P7+ vs P8 @ 3.5GHz

Category	Value
POWER8 3.5GHz (SMT-8)	18
POWER8 3.5GHz (SMT-4)	16
POWER7+ 3.6GHz	12
POWER7 3.55GHz	11

33%   50%

45%   65%

rPerf / Core - P7+ vs P8 @ 4.1GHz

Category	Value
POWER8 4.1GHz (SMT-8)	20
POWER8 4.1GHz (SMT-4)	19
POWER7+ 4.2GHz	14

36%   44%
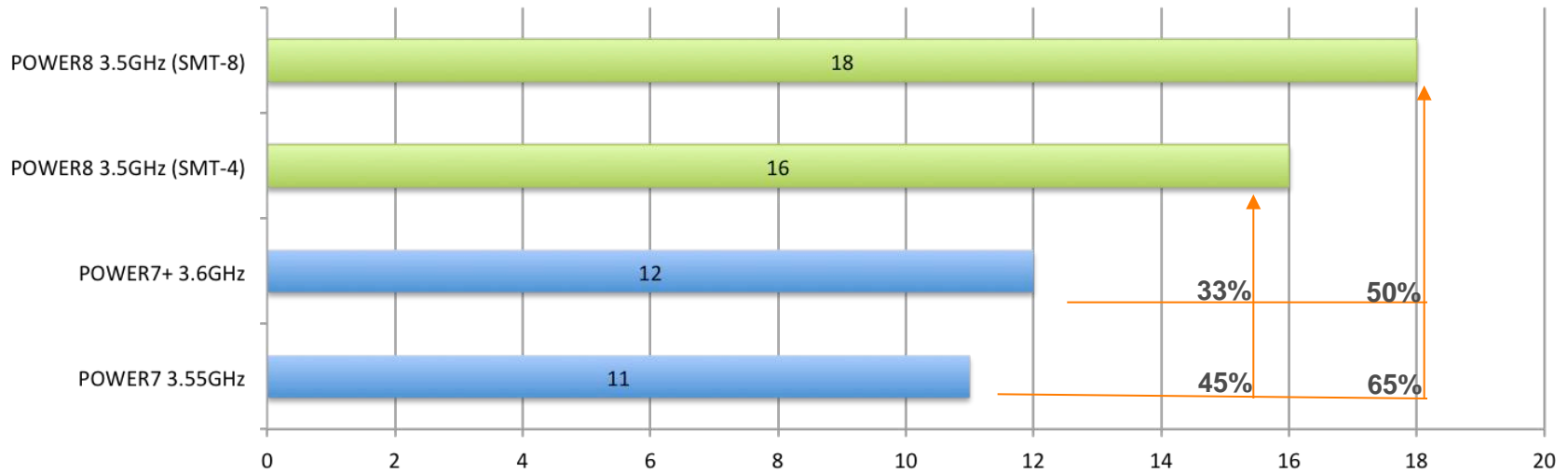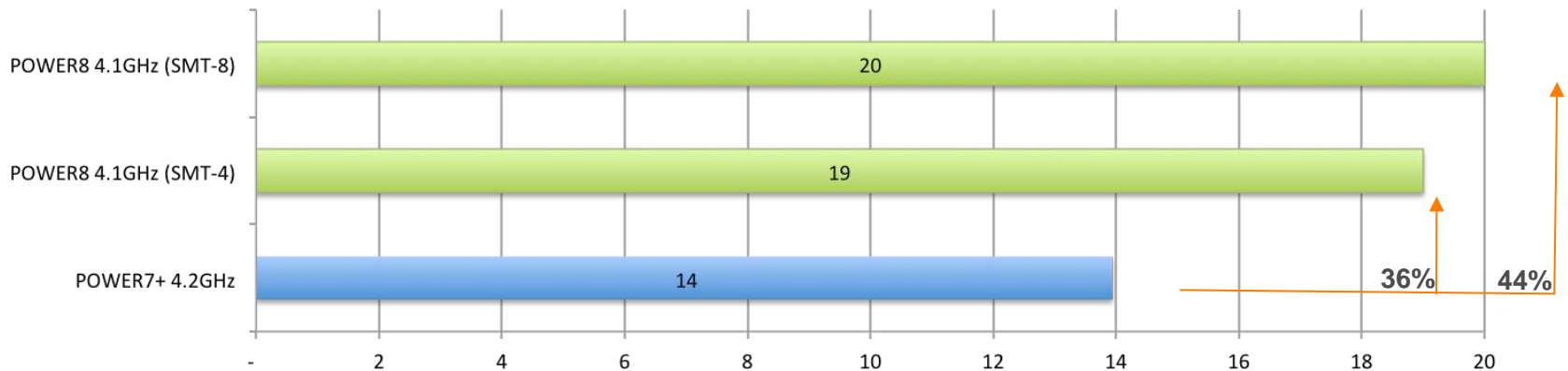
Power 750+ vs POWER8 4U2S

Performance
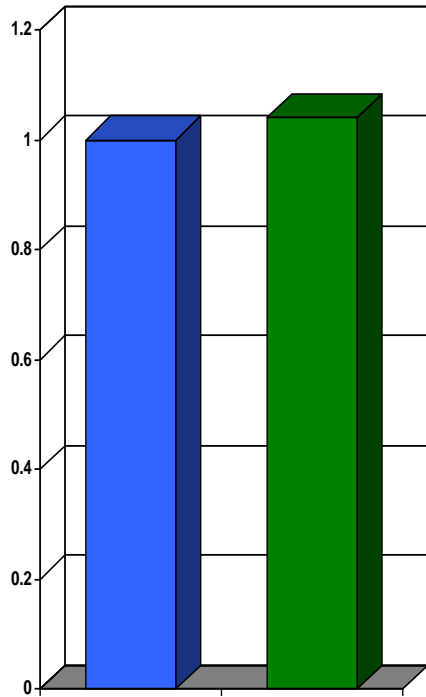
P 750+	P8 4U2S

POWER8 ≥ Performance

Total # Cores
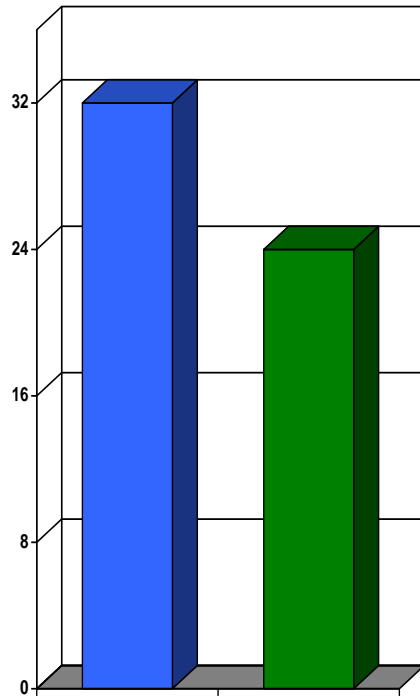
P 750+	P8 4U2S

Fewer cores means Lower Software Costs

Does not reflect lower PVU savings

TCA
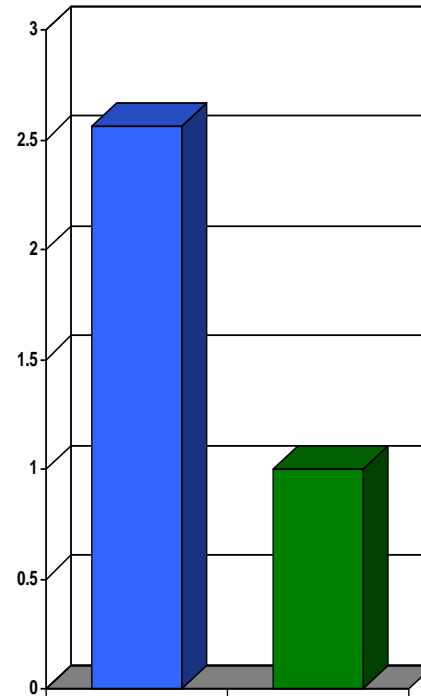
P 750+	P8 4U2S

Lower investment costs

Performance per $
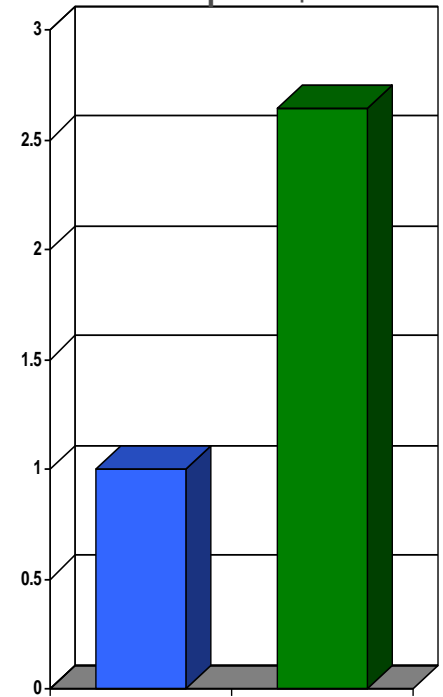
P 750+	P8 4U2S

Better performance per dollar spent

Oracle EBS Payroll Batch XL

Oracle eBS 12.1.3 Payroll Bacth XL Overall Performance

- IBM Power S824 12 x POWER8: 1.090.909
- IBM Flex p260 8 x POWER7+: 626.741
- ORACLE X3-2 16 x E5-2690: 789.515
- CISCO UCS 24 x E5-2697v2: 1.017.639
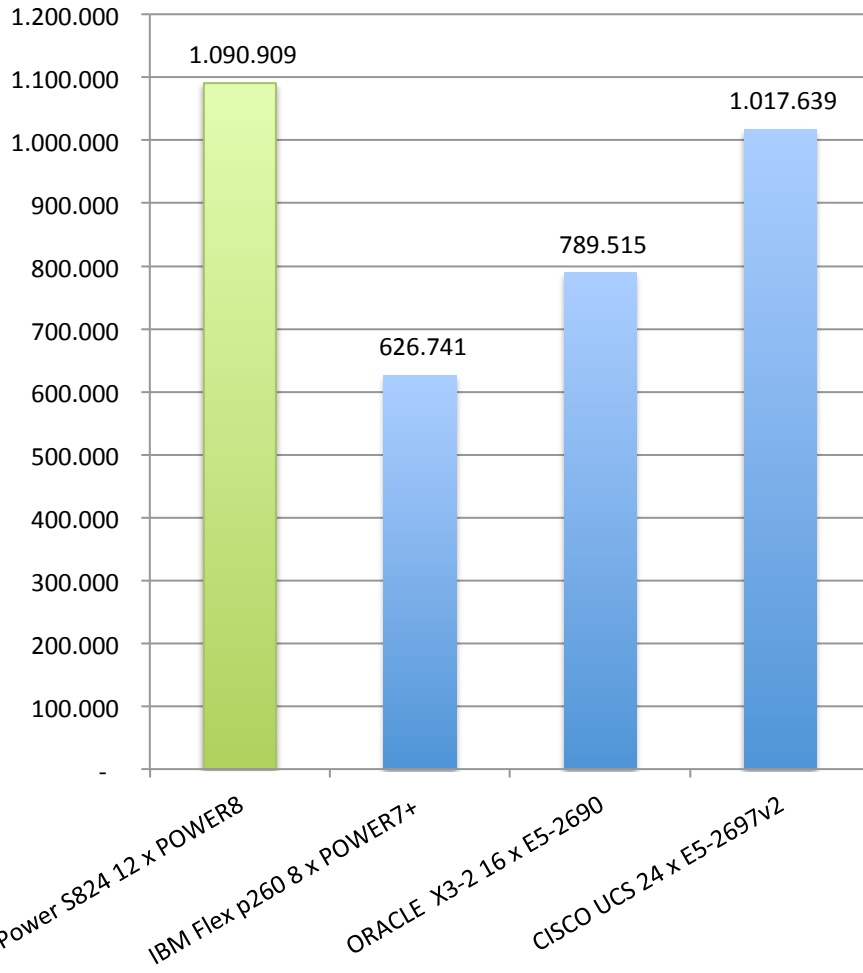
Oracle eBS 12.1.3 Payroll Bacth XL Performance/core

- IBM Power S824 12 x POWER8: 90.909
- IBM Flex p260 8 x POWER7+: 78.343
- ORACLE X3-2 16 x E5-2690: 49.345
- CISCO UCS 24 x E5-2697v2: 42.402

2,15x
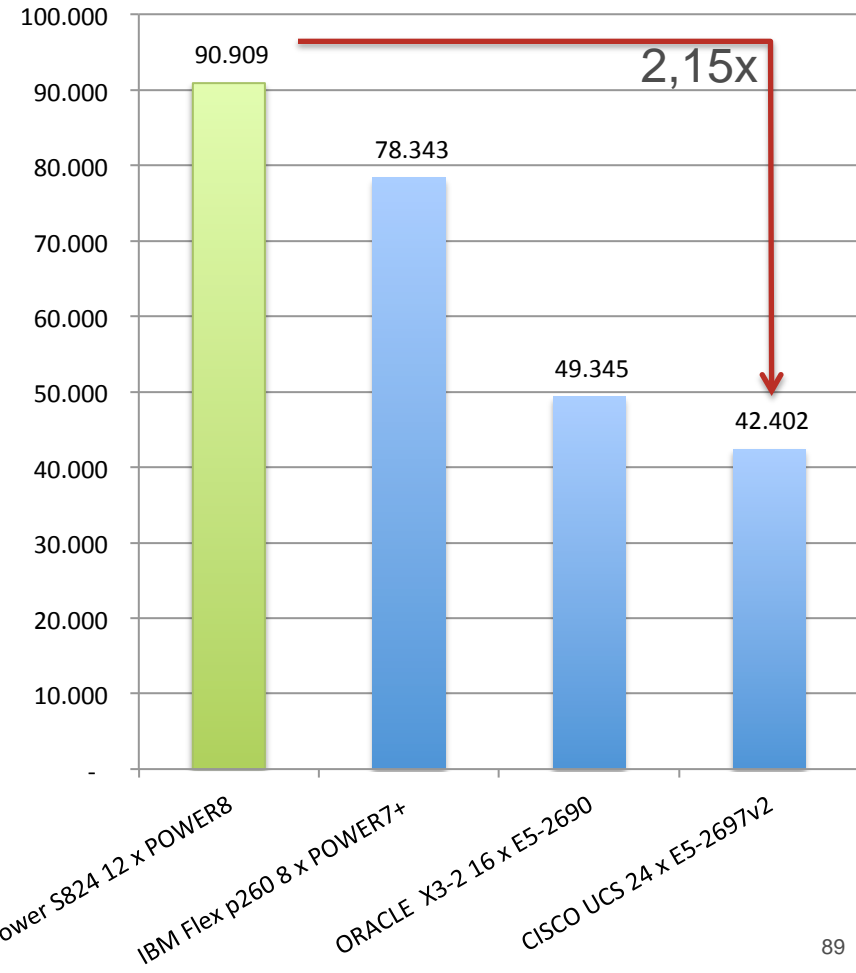
RAS

RAS Feature Overview

- ● Standard
- ◉ Optional
- — Not Available

RAS Item	POWER7+ 710 / 730	POWER7+ 720 / 740	POWER8 models
Redundant / Hot Swap Fans & Blowers	●	●	●
Hot Swap DASD & Media	●	●	●
Hot Swap PCI Adapters	—	—	●
Concurrent Firmware Update	●	●	●
Redundant / Hot Swap Power Supplies	●	●	●
Dual disk controllers (split backplane)	—	◉	◉
Processor Instruction Retry	●	●	●
Alternate Processor Recovery	●	●	●
Storage Keys	●	●	●
PowerVM™/Live Part. Mobility/Live App Mobility	◉	◉	◉
Dynamic Processor Sparing	—	—	—
Redundant Service Processors	—	—	—
Redundant System Clocks	—	—	—
Hot GX Adapter Add and Cold Repair	—	—	*N/A*
Dynamic Service Processor &System Clock Failover	—	—	—
Enterprise Memory (Memory Sparing)	—	—	●
Integrated TPMD Function	—	—	●
Hot GX Adapter Repair	—	—	*N / A*
Active Memory Mirroring for Hypervisor	—	—	—
Power Pools	—	—	—
Dynamic Processor Sparing	—	—	—

RAS Features – Always On

- L1/L2/L3 Cache ECC
  - Correct Single Bit Error
  - Detect Double Bit Error
  - Can delete a cache line with a persistent issue
  - Can repair a column of an L3 cache dynamically by using spare capability
- Cores
  - Processor Instruction Retry
  - Alternate Processor Recovery
  - Partition Availability Priority
- Memory
  - Retry Buffer on Controllers
  - Spare data lanes on memory buses
  - L4 on DIMM – Correct single error, Detect double error
  - Memory Scrubbing
  - ChipKill
  - Spare DRAM modules - can handle up to 4 DRAM module ChipKills

Intel Memory – Performance or RAS not Both

- 2 memory modes
- Lockstep
  - RAS
  - Redundant memory channels
  - 1600MHz
- Performance
  - Performance
  - Non redundant memory channels
  - 1333MHz
- Setting via BIOS
- Reboot

Intel® C10x[1] Scalable Memory Buffer Performance vs. Lockstep Mode BIOS Options

Two operation modes supported, default setting is Performance Mode

Lockstep Mode (1:1)
- Up to 1600 : 1600

Performance Mode (2:1)
- Up to 2667:1333
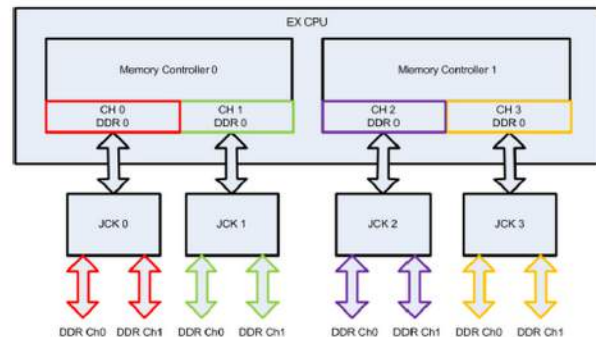
1:1 bus ratio
DDR channels @ DDR3-1600
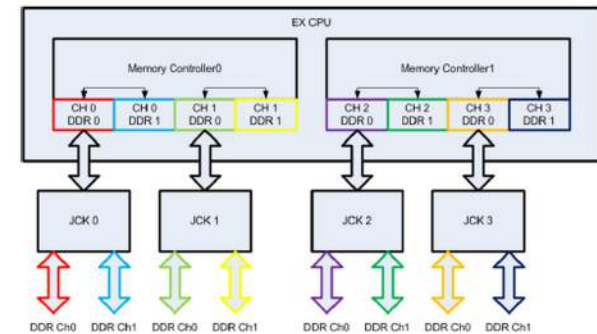
2:1 bus ratio
DDR channels @ DDR3-1333

[1] Intel® C10x is the external product name for the JordanCreek Memory Extension Buffer Chip. Two versions exist, one that supports up to 8 DIMMs and a second one that supports up to 12 DIMMs, hence the suffix "x" is used.

Intel Confidential

Memory Bandwidth & Latency

Massive increase in throughput[1]

Results have been measured based on internal Intel tools and are provided for informational purposes only. Any difference in system hardware or software design or configuration may affect actual performance.

3Read 1Write traffic mix

Up to 54% BW increase by enabling independent channel mode

Near 100% BW scaling efficiency 50% memory speed increase giving back up to 49% memory BW increase

Lockstep Mode
Perf/socket
27GB/s @ 100ns
44GB/s @ 300ns

Performance Mode
Perf/socket
40GB/s @ 100ns
70GB/s @ 300ns

Lower idle latency by 8-10ns, latency stays flat longer

Up to 130% increase in overall BW from WSM-EX to IVB-EX

4 Socket Memory Bandwidth (in MB/sec)

IVB-EX Lockstep - 1600 DDR [1]　WSM-EX 1067 DDR [2]　IVB-EX Performance Mode - 1333-DDR [3]

Performance mode use recommended for highest system memory bandwidth (default setting with released BIOS)

Solutions

New Power Systems Solutions for Big Data & Analytics

Bringing insight to the point of impact with systems and solutions optimized for Big Data & Analytics performance

IBM Solution for Hadoop – Power Systems Edition

Integrated platform solution for Hadoop analytics software
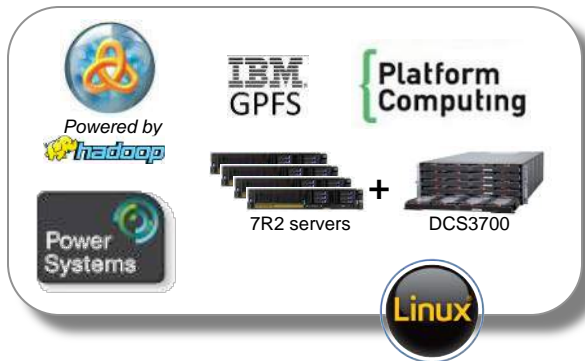
BLU Acceleration – Power Systems Edition

Integrated platform solution for analytics software

IBM Solution for Analytics – Power Systems Edition

Integrated solution pre-configured with Cognos BI and/or SPSS and/or BLU Acceleration
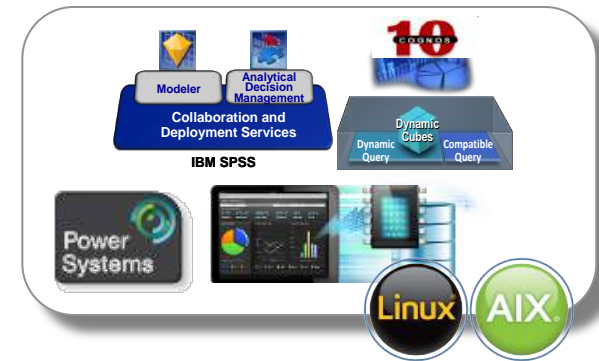
IBM

Power Systems can deliver insight to the point of impact with big data & analytics **accelerators**

75% less storage[1]

IBM FlashSystem

82X faster insights[2]

Next Generation In-Memory

IBM DB2
BLU Acceleration

24:1 consolidation[3]

POWER8 with CAPI Flash Accelerators

CAPI

IBM DB2 with BLU Acceleration
Unmatched Innovation from IBM Research & Development

Next Generation In-Memory

In-memory columnar processing with
dynamic movement of data from storage

C1 C2 C3 C4 C5 C6 C7 C8

Actionable Compression

Patented compression that preserves order
so data can be used without decompressing

0 DATA 1
11001101

Encoded

Instructions   Data

SIMD

Results

Parallel Vector Processing

Multi-core and SIMD parallelism
(Single Instruction Multiple Data)
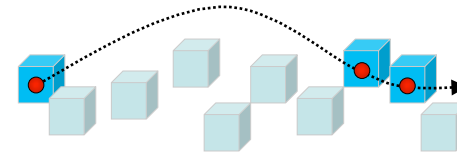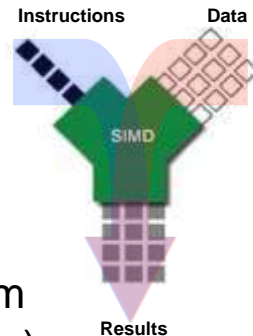
Data Skipping

Skips unnecessary processing of
irrelevant data

BLU Acceleration Illustration: 10TB query in seconds or less

- The System: **32 cores, 1TB memory, 10TB table with 100 columns and 10 years of data**
- The Query: **How many "sales" did we have in 2010?**
  - SELECT COUNT(*) from MYTABLE where YEAR = '2010'
- The Result: **In seconds or less as each CPU core examines the equivalent of just 8MB of data**

10TB data

Actionable Compression
reduces to **1TB**
In-memory

Column Processing
reduces to **10GB**

Data Skipping
reduces to **1GB**

Parallel Processing
32MB linear scan
on each core via

Vector Processing
Scans as fast as
8MB through SIMD

**Result in
seconds or less**
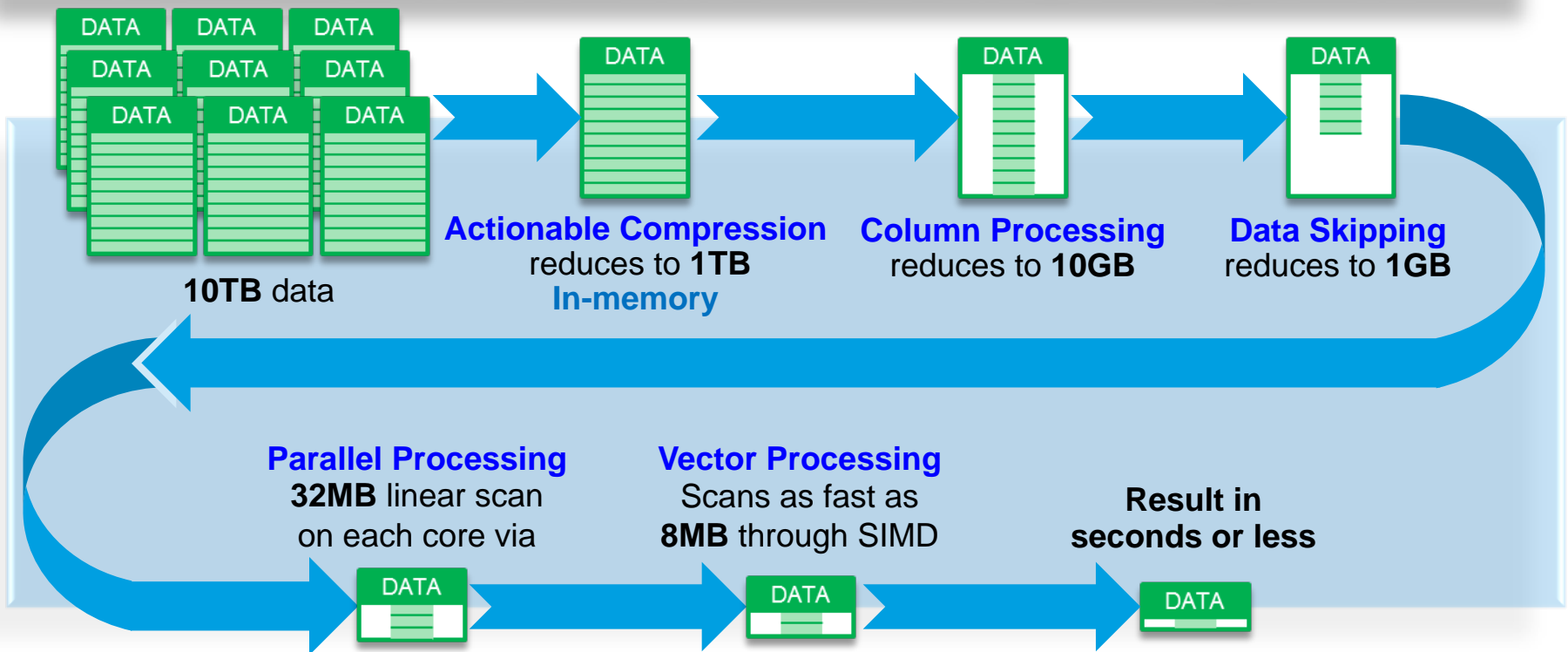
Power Systems clients are seeing Remarkable Results

*"We tested some representative queries taken from our existing SAP application and tested them on DB2 with BLU Acceleration and observed performance improvements in many of our query response times. For example**, one of our most time consuming queries experienced a 50x performance improvement**."*

- Richard Simms, Director of Infrastructure, Fossil

*"I was asked to help one of our analysts who gave up on something he was executing across a billion records, because **after 3 hours, the answer didn't come back**. We took that query (which involved a big join) & ran it on DB2 with BLU Acceleration and it **finished in 10 seconds – 1080x faster.**"*

- Randy Wilson, Lead DB2 for LUW DBA

"The BLU Acceleration technology makes our analytical queries run **4-15x faster and decreases the size of our tables by 10x**. But it's all the things I don't have to do with BLU that make me appreciate the technology even more: **no tuning, no partitioning, no indexes, no aggregates.**"

- Andrew Juarez, Lead SAP Basis and DBA

DB2 BLU Sizing Validation – System Recommendations

The following table shows various uncompressed database sizes and systems to consider using. A larger system can always contain an optimally configured LPAR.

Total Uncompressed DB Size	Active Uncompressed Data Size	# of Cores	RAM
<1TB	< 330GB	8 -12	64 - 128
1 – 2.5	330 – 800GB	8 - 16	128 - 256
2.5 – 5	800 – 1700GB	16 - 24	256 - 512
5 – 7.5	1.7 – 2.5TB	16 - 32	512 - 1024
7.5 – 10	2.5 – 3.3TB	32 - 48	1024 - 2048
10 – 15	3.3 – 5TB	48 - 64	1024 - 2048
15 – 20	5 – 6.7TB	64 - 128	2048 - 4096
20 +	6.7TB >	64 +	2048 +

These values may be adjusted once the Best Practices and sizing guidelines are published.

IBM Solution for Analytics - Power Systems Edition

- **POWER8 server options**: Power S814, Power S822, Power S824

- **System Software**: AIX Standard or Enterprise Edition, PowerVM Standard or Enterprise Edition

- **BLU Acceleration options**: IBM DB2 Advanced Workgroup or Advanced Enterprise Edition

- **Analytics options**: Cognos BI, SPSS Modeler, SPSS C&DS, SPSS ADM, Infosphere DataStage

Simple to Acquire
Order server, storage, software and support from a single vendor

Simple to Deploy
Pre-installed and pre-optimized server, storage & software

Simple to Implement
Highly scalable to grow as your analytics need change

http://www.ibm.com/systems/power/solutions/analytics/editions/preload-analytics.html

IBM Solution for Hadoop - Power Systems Edition

NEW. A storage-dense integrated platform optimized to simplify and accelerate unstructured big data & analytics

Powered by hadoop

Platform Computing

IBM GPFS

POWER7+ BUILT ON Power

7R2 servers   +   DCS3700
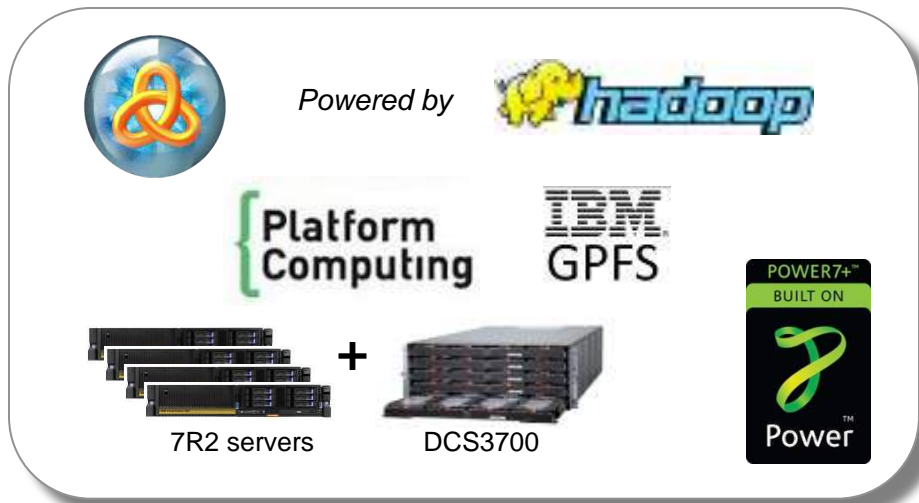
Speed Matters
Higher ingest rates delivers 37% faster insights than competitive Hadoop solutions with 31% fewer data nodes.[1]

Availability Matters
Better reliability and resiliency with 73% fewer outages and 92% fewer performance problems over x86.[2]

Integrated platform solution for Hadoop ready for analytics software

1) Based on STG Performance testing comparing to Cloudera/HP published benchmark
2) CLAIMS: Solitaire Interglobal Paper - Power Boost Your Big Data Analytics Strategy – http://www.ibm.com/systems/power/solutions/assets/bigdata-analytics.html?LNK=wf

IBM Solution for Hadoop – Power Systems Edition

Key requirements & design parameters – focused on customer value

- Best-in-class hardware
- Dense storage subsystem
- Advanced software capabilities
- Better reliability & management
- Best in class file system
- Automated cluster provisioning
- IBM Platform Symphony

IBM InfoSphere BigInsights
or Open-source Hadoop
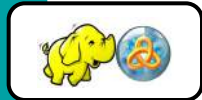
IBM Platform Symphony
IBM Platform Cluster Manager

Distributed File System
IBM Elastic Storage, HDFS

Linux Operating Environment
RHEL SUSE

IBM Power Systems
IBM Power 7+, Power8

Standard Configurations – POD – based design

Power Big Data clusters are built using a simple building block approach to tailor the mix of CPU and storage to application requirements.

Compute Dense
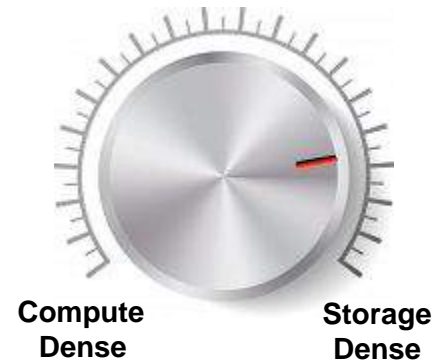
Storage Dense

Compute Dense Data POD

Four servers per storage subsystem

Balanced Data POD

Two servers per storage subsystem

Storage Dense Data POD

One server per storage subsystem

Software benefits and performance

SQL on Hadoop – BigInsights (BigSQL) on Power8

- BigSQL allows you to run SQL on your Hadoop data
  - All data stays in a Hadoop filesystem (HDFS or GPFS-FPO)
- Uses IBM's mature relational MPP database technology
- Provides for enterprise-level security
  - Authentication, authorization (fine-grain access), auditing, etc.
- Other enterprise features:
  - Federation
  - Use of standard SQL, and all the SQL / Business Analytics tools
  - Mixed workload cluster management

- Performance results
  - Ran TPC-DS queries only on a 10TB data set
  - All 99 queries run without any changes
  - Ran a multi-user throughput test
  - Comparison against Hive running on Intel based cluster

**ETA 3Q14
BigInsights 3.0.0.1**

POWER8 delivers **2.5x performance** on Big Data/Hadoop
POWER8 requires 60% fewer servers based on the best x86 published result

- POWER8 S822L delivers over 2.5x the performance of the best published x86 system

  … and continues to offer far superior RAS

- POWER8 exploits additional cores, more threads, larger caches, memory bandwidth

- Terasort is a popular benchmark to measure the performance of a Hadoop solution

  - Sorts a large dataset (10 TB) in parallel

  - Exercises the Map-reduced framework and Hadoop Distributed File System (HDFS)

Relative System Performance
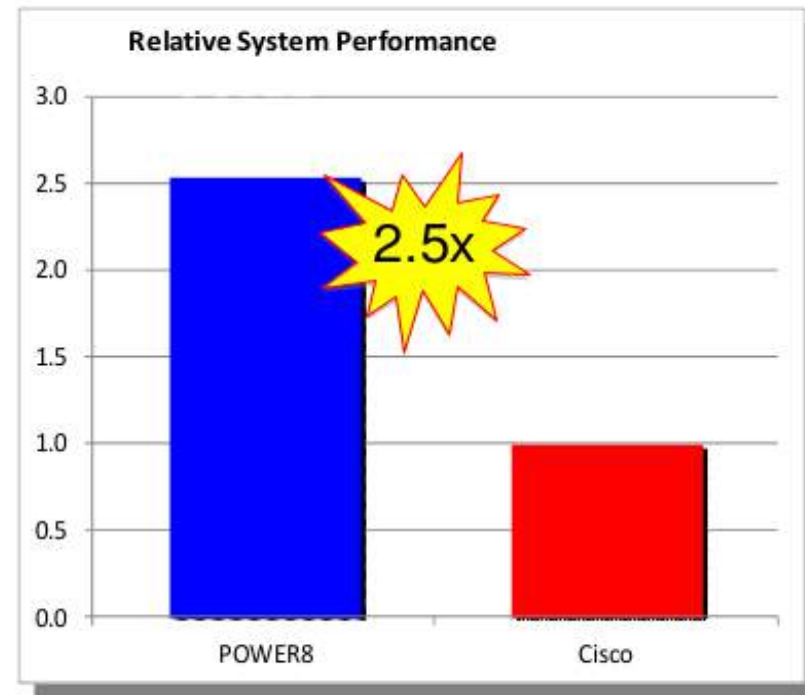
IBM Analytics Stack: IBM Power System S822L; 24 cores / 192 threads, POWER8; 3.0GHz, 512 GB memory, RHEL 6.5, InfoSphere BigInsights 3.0.0.0

http://www.cisco.com/en/US/solutions/collateral/ns340/ns517/ns224/ns944/le_tera.pdf

"In-Memory" NoSQL/KVS Infrastructure via CAPI-attached Flash

12TB, 6K users, 2GB/User Example – Power Systems delivers 24:1 consolidation over x86[1]

Today (x86)

WWW

10Gb Uplink

Load Balancer

500GB Cache Node

500GB Cache Node

500GB Cache Node

500GB Cache Node

500GB Cache Node

500GB Cache Node

24 nodes

Backup Node

Backup storage req'd for data persistence
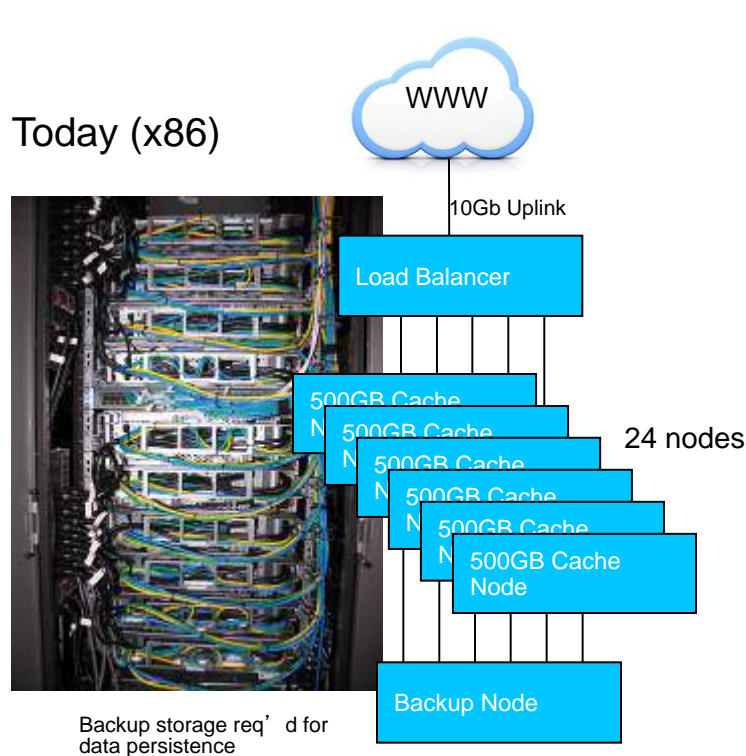
26 servers, switches, etc. = >1 rack
In-memory database w/ 12TB DRAM
~18kW

WWW

10Gb Uplink

POWER8 + CAPI Flash

POWER8 4u Server

2u Flash w/ 12TB

4u, 1.5kW
In-memory database with 12TB Flash
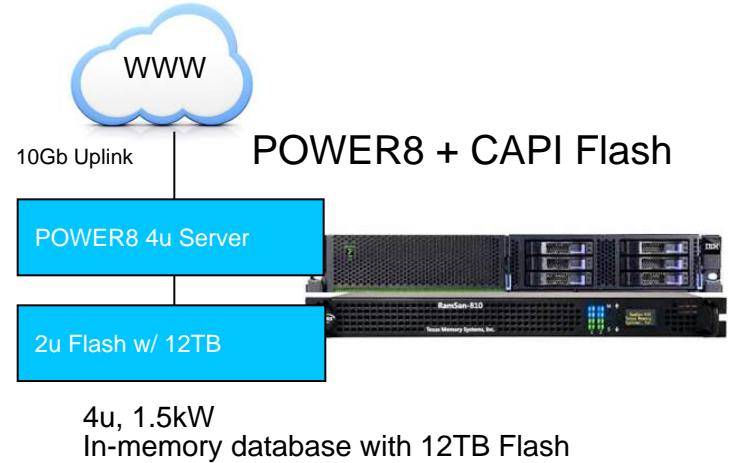
Sizings

Power + CAPI Flash Advantage

- 24:1 physical server consolidation
- >1 rack to 4U Density (>12x)
- 18kW ◊ 1.5kW Energy (12x)
- $24/user ◊ $7.5/user Cost (3.2x)
- 4.5k ◊ 60k Users per rack (13x)[2]

redis

1) 24:1 system consolidation ratio (12:1 rack density improvement) based on a single IBM S824, (24 cores, POWER8 3.5 GHz), 256GB RAM, AIX 7.1 with 40 TB memory based Flash replacing 24 HP DL380p, 24 cores, E5-2697 v2 2.7 GHz), 256GB RAM, SuSE Linux 11SP3 . Inbound network limits performance to 1M IOPs in both scenarios, equal capacity (#user, data) in both cases. x86 cost includes 10k$ for 2x 1U switches

2) Normalized to a single rack

NC State analyzes unstructured big data with Linux on Power

Largest technical college in the Southeast deploys Cloud based big data & analytics solution on IBM PowerLinux

Need

- NC State University tries to teach students how to look at a business problem, analyze them, and come to a decision-based understanding of where they're going to go with a new business.

Solution

- IBM Power Linux servers
- IBM InfoSphere BigInsights
- IBM Content Analytic Studio
- Deployed in an IBM Cloud

    NCSU (Case Study, YouTube)

Results

- Establish a Cloud computing environment that allows students to use products such as IBM Content Analytic Studio for doing natural language processing.

North Carolina State University
is one of the largest technical colleges in the Southeast
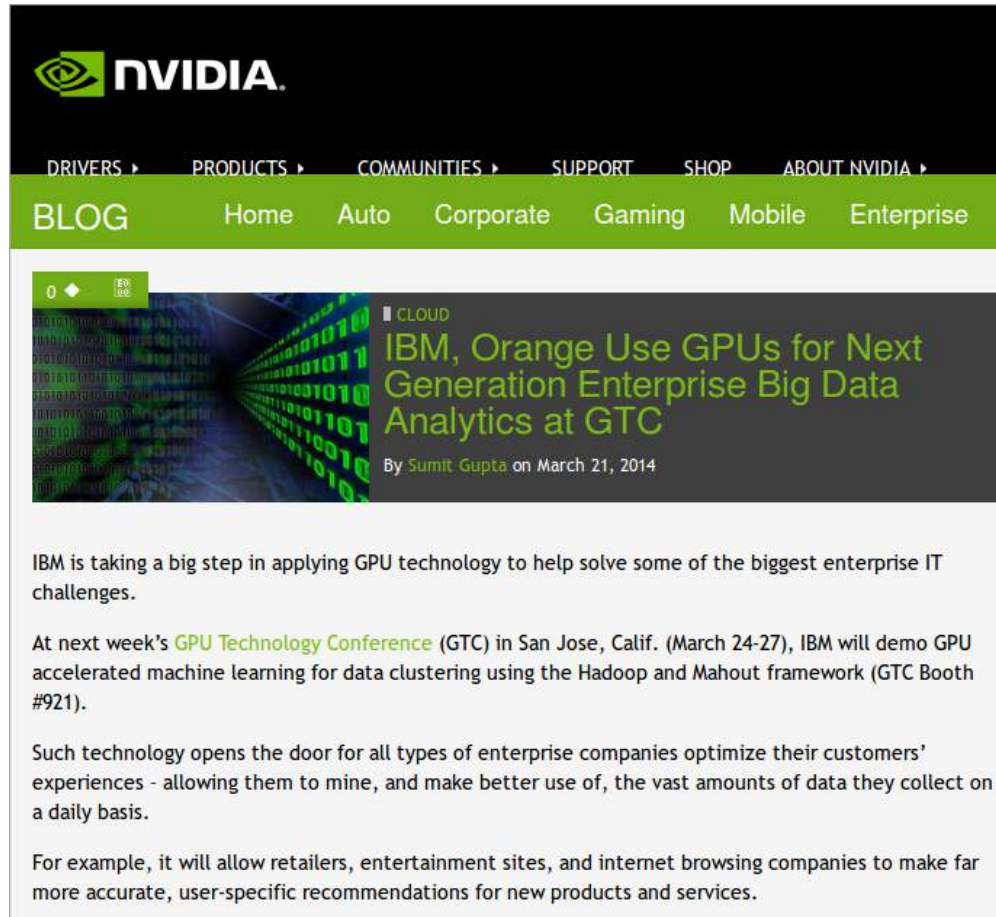
"By working smarter and doing smarter computing, you'll be able to optimize your systems and, again, leverage your talent for looking for important knowledge."

- Dr. Michael Kowolenko, Senior Research Scholar College of Management, NCSU

GPU processing and Big Data

IBM is demonstrating the use of GPU accelerators on a distributed computing system (required for such an enormous data set) for clustering using Hadoop. With GPU accelerators working alongside IBM Power CPUs, the demo runs eight-times faster than with a Power system without GPUs.

HMC Version 8
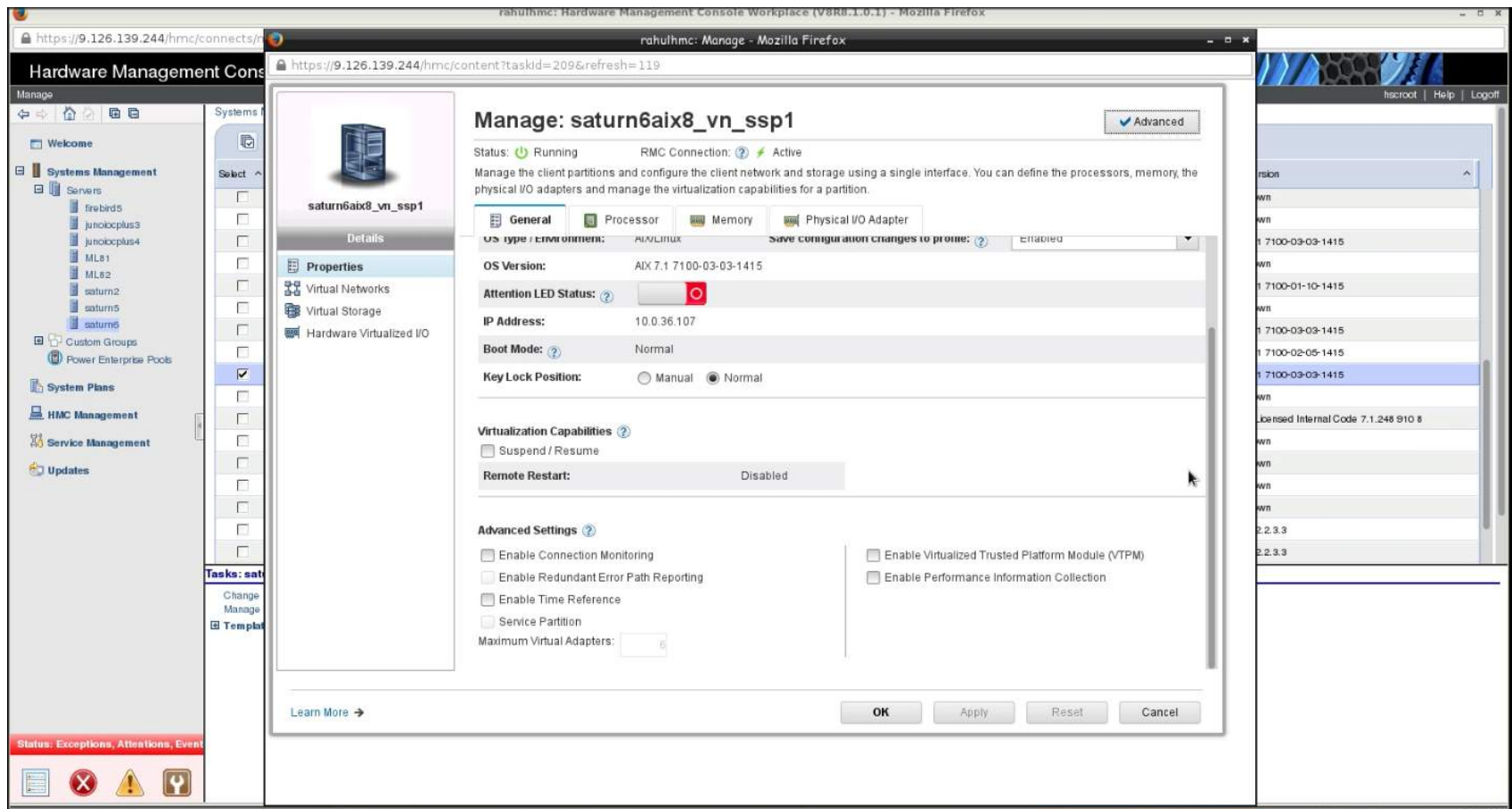
- Hardware Support / Serviceability
  - Support for new POWER8 processor-based systems
  - Support for new I/O adapters
  - New rack-mount HMC – 7042-CR8
  - Call Home Serviceable Event Manager (SEM)
  - Server Firmware Entitlement
  - Additional tracing during Serviceable Event creation

- PowerVM Support
  - Integrated Facility for Linux
  - HMC CLI for Remote Restart
  - Dynamic reset of Partition Remote Restart Attribute
  - IBM i 5250 virtual Terminal changes
  - Schedule DPO operation from GUI

- HMC Console Management
  - Security: Support of NIST 800-131a
  - HMC Installation Changes

- HMC Usability Enhancements
  - PowerVM Management
  - Performance Management

- Redbook: *IBM Power Systems Hardware Management Console Version 8 Release 8.1.0 Enhancements* - http://www.redbooks.ibm.com/redpieces/pdfs/sg248232.pdf

HMC GUI Highlights

- **Seamless integration with existing HMC dashboard**
  - New GUI tasks are available right inside current HMC dashboard
  - New GUI tasks can launch existing HMC tasks from itself.
  - Look and feel in accordance with IBM OneUI guidelines.
  - GUI design after regular feedback from real customers and client council meetings.

- **Change in Concepts**
  - The VIOS is no more a partition, its considered part of Virtualization layer.
  - Storage assignments made from a Logical Partition perspective
  - Introduction of new concepts such Virtual Network and Network Bridge.
  - Abstracting low level details such as adapters in Virtual storage assignment.

- **New Features**
  - Introduction of Shared Storage Pool management.

HMC New Management GUI - Managing partitions

HMC New Management GUI - Managing VIOS

HMC New Management GUI - Managing Virtual Networks

HMC New Management GUI - Managing Virtual Storage

HMC Performance Monitoring

HMC Today

HMC w/ new enhancements

- Disjoint set of tools
- Multiple agents need to be installed in OS
- Minimal or Lack of Visualization

- Integrated Visual Monitor in HMC
- Standard set of Interfaces (REST APIs) for external APIs to consume data

Integrated Performance & Capacity Metrics

Provide full PowerVM performance and capacity metrics via a single touch-point (HMC).

- Data collection and aggregation of performance metrics from Hypervisor and VIOS

- Identify performance bottlenecks and facilitate early problem detection and notification

- REST based API to access all platform (Hypervisor & VIOS) metrics for IBM Internal and Third Party tools

- Comprehensive Visualization to view **aggregated historical performance data**

- It is not a real time performance monitoring tool. It provides snapshots of resource utilization every minute

- Main objective is performance management and capacity planning - not monitoring of specific process/application in detail

-

Performance & Capacity Metrics Details
Provides both Compute & IO Resources Utilization

- Following new metrics available and exposed through new HMC APIs
  - Physical System Level Processor & Memory Resource Usage Statistics
    - System Processor Usage Statistics (w/ LPAR, VIOS & Power Hypervisor usage breakdown)
    - System Dedicated Memory Allocation and Shared Memory Usage Statistics (w/ LPAR, VIOS & Power Hypervisor usage breakdown)
  - Advanced Virtualization Statistics
    - Per LPAR Dispatch Wait Time Statistics
    - Per LPAR Placement Indicator (for understanding whether the LPAR placement is good / bad based on score)
  - Virtual IO Statistics
    - Virtual IO Server's CPU / Memory Usage (Aggregated, Breakdown)
    - SEA Traffic & Bandwidth usage Statistics (Aggregated & Per Client, Intra/Inter LPAR breakdown)
    - NPIV Traffic & Bandwidth usage Statistics (HBA & Per Client breakdown)
    - vSCSI Statistics (Aggregated & Per Client Usage)
    - VLAN Traffic & Bandwidth usage Statistics (Adapter & LPAR breakdown)
    - SRIOV Traffic & Bandwidth usage Statistics (Physical & Virtual Function Statistics w/ LPAR breakdown)

120

Integrated Performance Monitoring in HMC
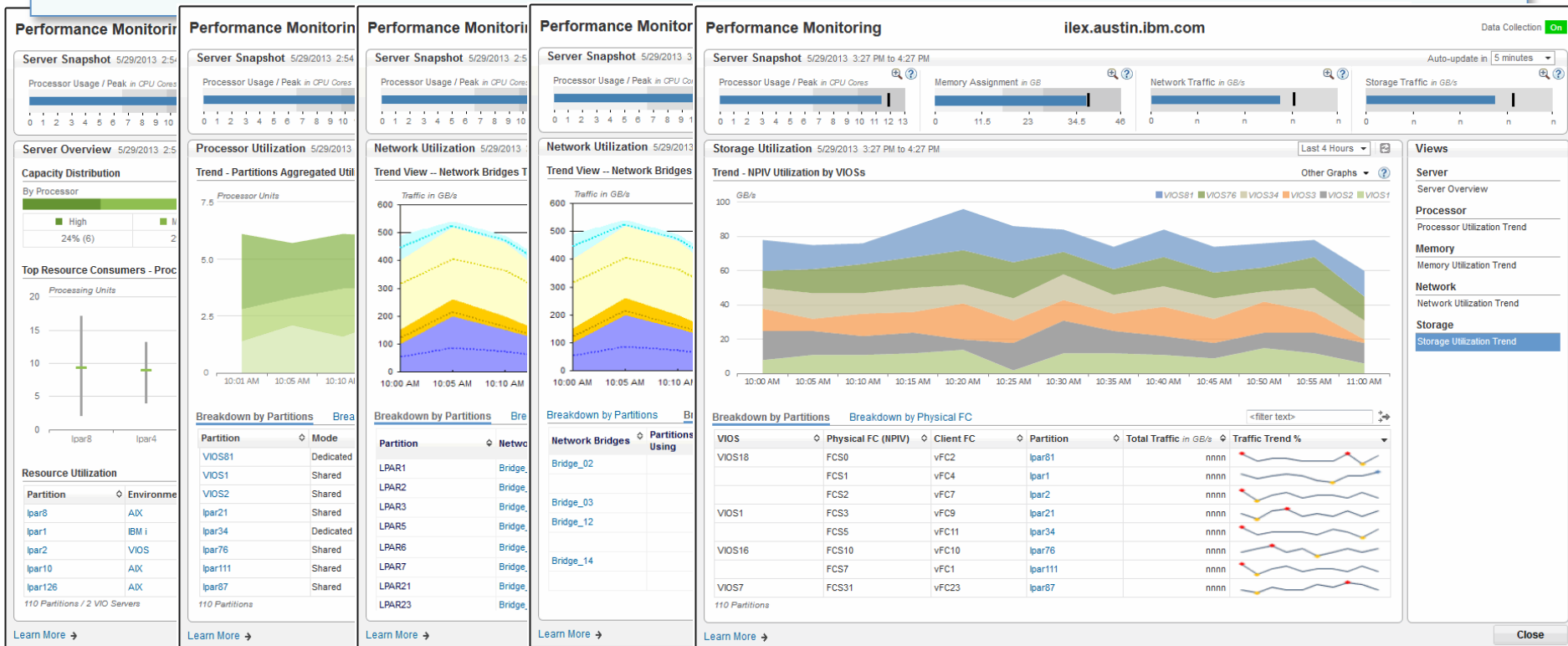
- Resource utilization dashboard
  - Last few minutes Processor, memory & I/O utilization snapshot (1, 5, 10, 15 minutes refresh)
  - Server & LPAR level Capacity Distribution, Top Resource consumption information
- Basic Utilization Trends for all resources w/ LPAR & Resource level breakdowns

HMC Management Demo Videos

- Managing partitions
  - ManageLpar-1-4.avi
- Managing VIOS
  - ManagePowerVM-1.avi

PowerVM Live Partition Mobility Performance Enhancements

- LPM performance has been continuously improving

  - 5 times faster in 2014 than in 2011
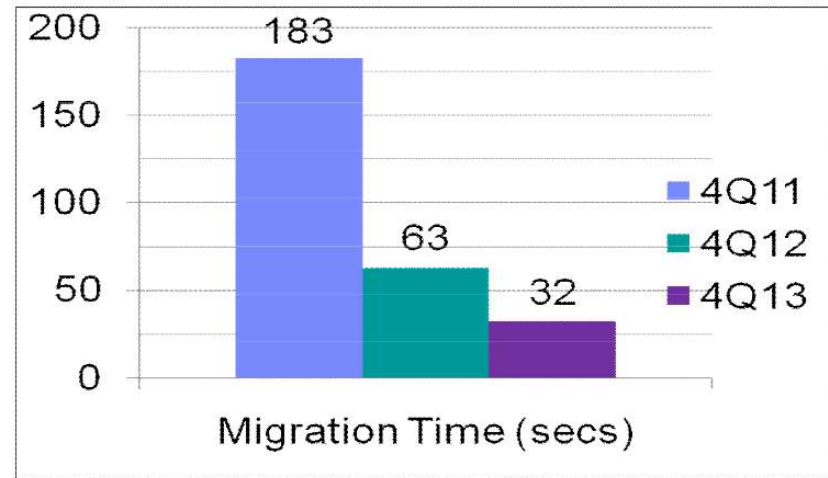
- VIOS 2.2.3 brought significant significant performance enhancements for LPM (2013)

  - Ability to achieve full bandwidth on a 10G adapter when transferring memory
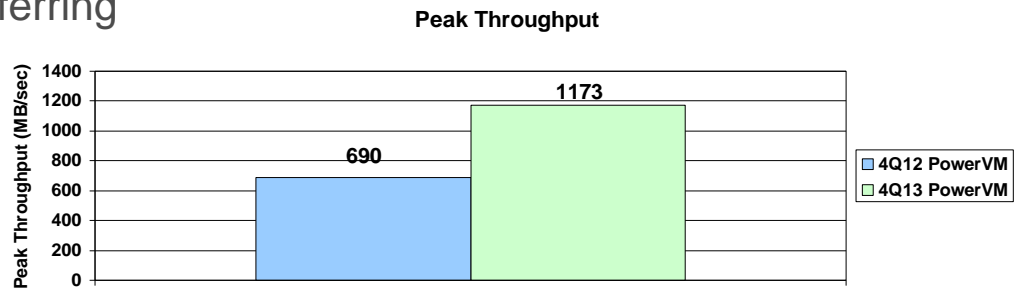
PowerVM Live Partition Mobility Performance Enhancements

- HMC V8R8.1.0 brings further enhancements by optimizing the device selection for LPM
    - During a Live Partition Mobility activity the HMC makes *find_device* calls to the Virtual I/O Server involved in the migration when virtual SCSI (vSCSI) device mapping is specified.
    - In previous versions of HMC the number of *find_device* calls was equal to the number of vSCSI adapters in each Virtual I/O Server.
    - In HMC V8R8.1.0 the logic has been improved so that only one *find_device*
    - call is made to each Virtual I/O Server. Depending on the version of Virtual I/O Server and the configuration of the backing devices each *find_device* call could range between one and 15 seconds tocomplete, so a reduction in the number of
    - *find_device* calls improves Live Partition Mobility performance.

Fix Level Recommendation Tool - LPM Validation

- The FLRT Live Partition Mobility (LPM) report provides recommendations for LPM operations based on source and target input values. These recommendations may include recommended fixes, including interim fixes, for known LPM issues. These recommendations are based on testing and field experience. The levels provided in the input screen are based on the supported levels for the selected machine type models.

- https://www14.software.ibm.com/webapp/set2/flrt/home

Fix Level Recommendation Tool - LPM Validation

Date: 2014.06.24
Report name: LPM
Source MTM: 9117-MMA
Target MTM: 8284-22A

Database last updated 2014.06.22

+ Expand all — Collapse all   PDF URL   PDF page

Source system — ☐

		Input level	Recommended level and updates
🛈	System firmware **Release date** **EoSPS**	EM350_159 2014.06.25 Last	None
🛈	HMC Software **Release date** **EoSPS**	V7 R770 2013.02.20 2015.02.28	V7 R770 SP4 2014.07.21 2015.02.28
✔	AIX **Release date** **EoSPS**	6100-09-01 2013.11.19 2016.11.19	rmcpkt5.140306.epkg.Z 2014.04.22

▼ Notes

rmcpkt5.140306.epkg.Z : rmcpkt5.140306.epkg.Z sets the default RMC packet size to 1500 bytes to avoid MTU imbalance problems. This fix does not require a reboot.

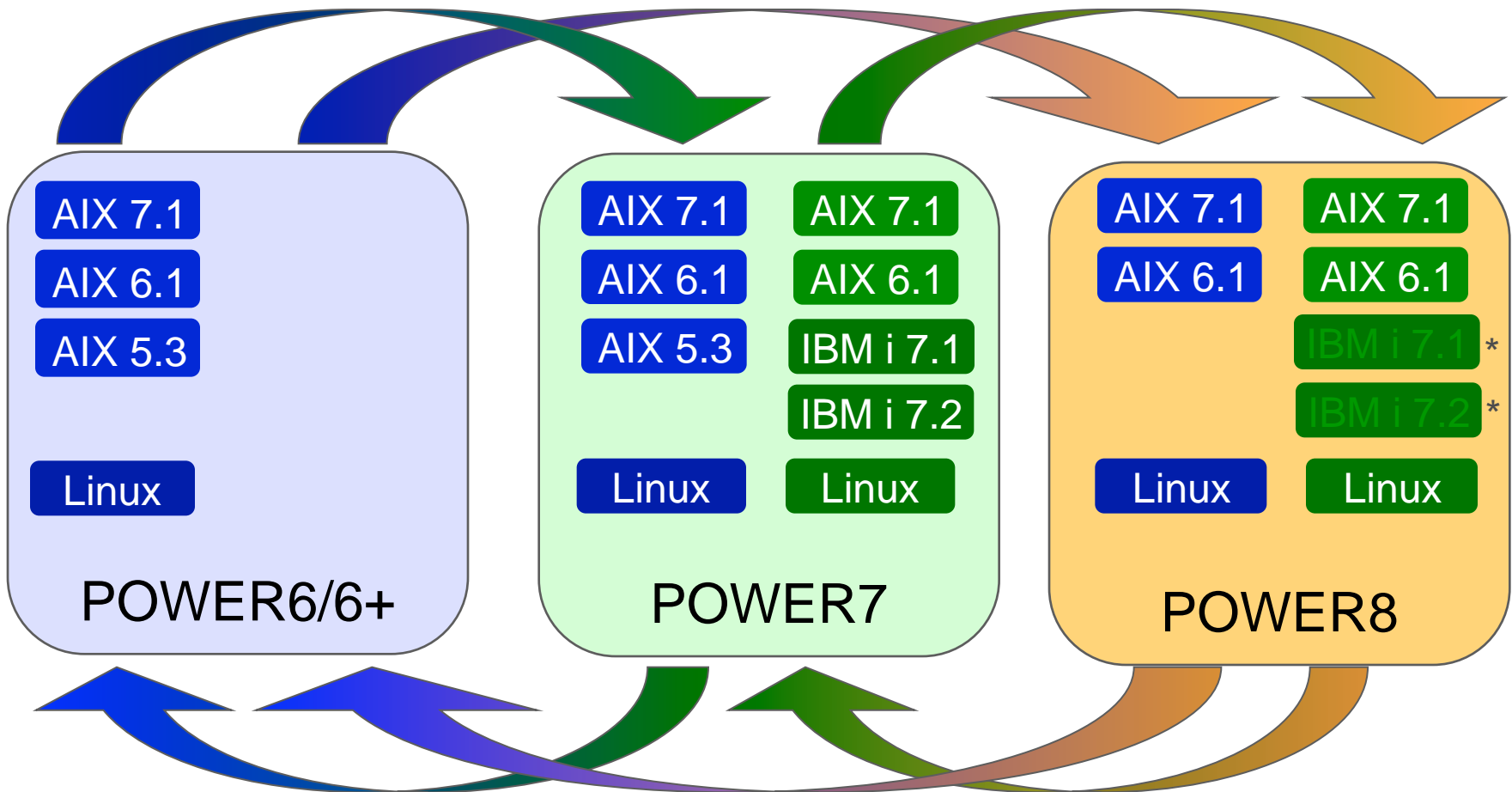|		Input level	Recommended level and updates
🛈	VIOS **Release date** **EoSPS**	2.2.2.2 2013.02.20 2015.10.01	2.2.2.4 2014.02.21 2015.10.01
			IV55956m4a.140226.epkg.Z 2014.04.22 rmcpkt4.131211.epkg.Z 2014.04.22

▼ Notes

IV55956m4a.140226.epkg.Z : IV55956m4a.140226.epkg.Z is only required if using NPIV.

rmcpkt4.131211.epkg.Z : rmcpkt4.131211.epkg.Z sets the default RMC packet size to 1500 bytes to avoid MTU imbalance problems. This fix does not require a reboot.

Target system — ☐

		Input level	Recommended level and updates
✔	System firmware **Release date** **EoSPS**	SV810_061 2014.07.29 Not Announced	None
✔	HMC Software **Release date** **EoSPS**	V8 R810 2014.06.10 Not Announced	MH01421 2014.06.10
✔	VIOS **Release date** **EoSPS**	2.2.3.3 2014.05.12 2016.10.01	None

POWER6 / POWER7 / POWER8 Partition Mobility

POWER6/6+
- AIX 7.1
- AIX 6.1
- AIX 5.3
- Linux

POWER7
- AIX 7.1
- AIX 6.1
- AIX 5.3
- Linux
- AIX 7.1
- AIX 6.1
- IBM i 7.1
- IBM i 7.2
- Linux

POWER8
- AIX 7.1
- AIX 6.1
- Linux
- AIX 7.1
- AIX 6.1
- IBM i 7.1 *
- IBM i 7.2 *
- Linux

Leverage POWER6 / POWER7 Compatibility Modes
LPAR Migrate between POWER6 / POWER7 / POWER8 Servers
Can not move POWER8 Mode partitions to POWER6 or POWER7 systems.

OS Support

POWER6 and POWER6+ MODES	POWER7 MODE *	POWER8 MODE
2-Thread SMT	4-Thread SMT, IntelliThreads	8-Thread SMT
8 Protection Keys *(16 in P6+ Mode)	32 Protection Keys User Writeable AMR	32 Protection Keys User Writeable AMR
VMX (Vector Multimedia Extension / AltiVec)	VSX (Vector Scalar Extension)	VSX2, In-Core Encryption Acceleration
Affinity OFF by Default	CPU/Memory Affinity Enhancements ON by Default, HomeNode, 3-tier Memory, MicroPartition Affinity	HW Memory Affinity Tracking Assists, Micropartition Prefetch, Concurrent LPARs per Core
64-core/128-thread Scaling	64-core / 256-thread Scaling 256-core / 1024-thread Scaling	> 1024-thread Scaling Hybrid Threads Transactional Memory Active System Optimization HW Assists
N/A	Active Memory Expansion	HW Accelerated/Assisted Active Memory Expansion
N/A	P7+ : AME compression acceleration and Encryption acceleration	Coherent Accelerator / FPGA Attach

* POWER7+ mode does not exist - POWER7 and POWER7+ processors use the same processor mode

AIX Support of Scale-out Servers

	11 / 2012	2 / 2012	3 / 2013	5 / 2013	8 / 2013	9 / 2013	10 / 2013	12 / 2013	2Q / 2014	3Q / 2014
AIX 6 TL7	SP6		SP7			SP8		SP9		**SP10**
AIX 6 TL8	SP1	SP2				SP3		SP4		**SP5**
AIX 6 TL9							SP1		**SP3**	
AIX 7 TL1	SP6			SP7	SP8			SP9		**SP10**
AIX 7 TL2	SP1	SP2								**SP5**
AIX 7 TL3							SP1		**SP3**	

P7 or P6 Modes with Virtual I/O

P7 or P6 Modes with Full I/O Support

P8, P7 or P6 Modes with Full I/O Support

AIX 7.1 TL 3 SP 3 Exploitation of POWER8 Features (June 2014)

- Security Features
  - Random number generator
  - On-chip crypto
  - In-core crypto
- Performance Optimization
  - Transactional Memory enablement for middleware and applications
  - New POWER8 instructions
  - Dynamic Code Optimization (for Java and IBM JIT technology)
  - SMT 8 (SMT 4 as default mode)
  - 24x7 Monitor
- Virtualization
  - Compression accelerator (used by Active Memory Expansion)

AIX exploitation of POWER8 accelerators

AIX OpenSSL Acceleration with POWER8 in-core Crypto Support
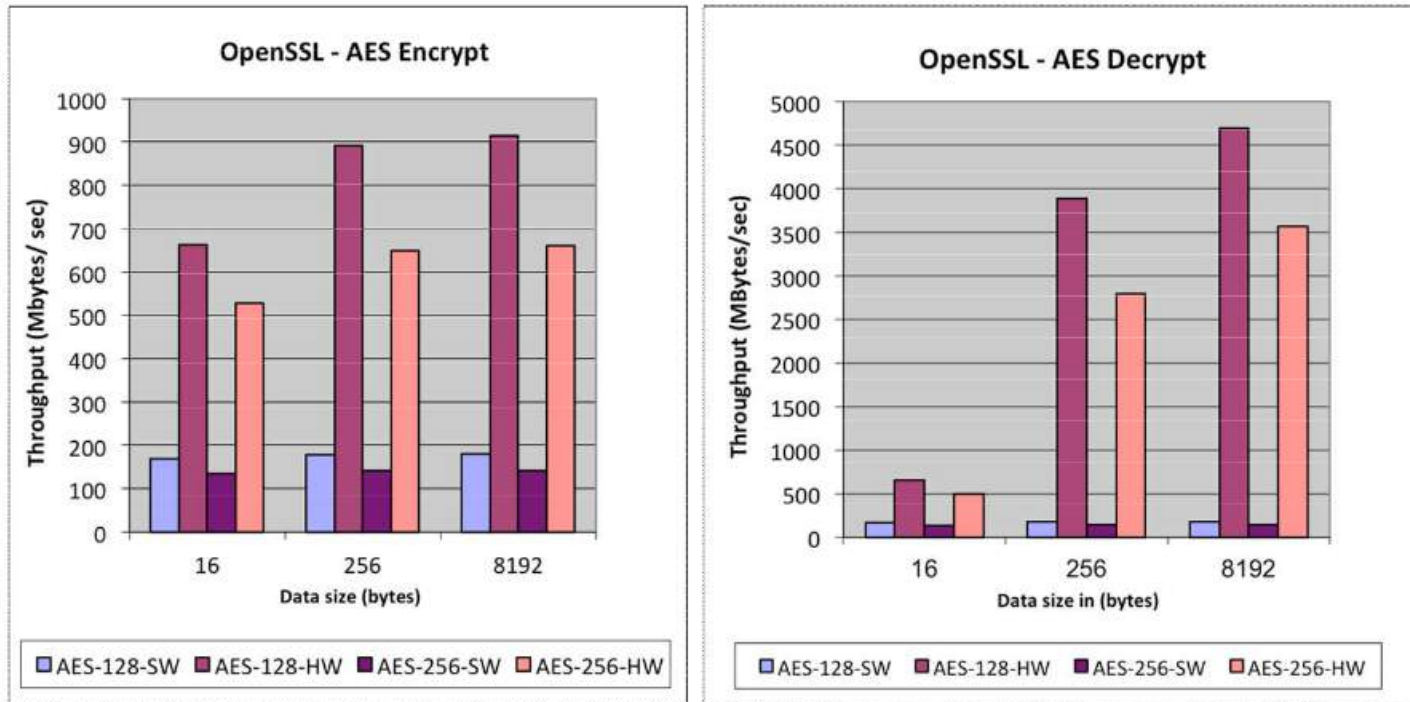
Results from AIX Development internal testing

- In-core Crypto is enabled on AIX 7.1.3.3 with OpenSSL version 1.0.1.510
- IBM GSKit 8.0.50.X with HW crypto Support Planned for 3Q 2014
- IBM Software products which embed GSKit will pick up HW acceleration as they qualify 8.0.50.X

IBM i 7.1 TR8   Cores and Threads for P7/P8

	IBM i 7.1
POWER7	Max Scale = 32 cores (SMT4) Max Partition = **64 cores (SMT4)** Threads = ST, SMT2, SMT4 up to 256 threads in single partition
POWER8	Max Scale = 32 cores (SMT8) Max Partition = **64 cores (SMT4)** Threads = ST, SMT2, SMT4, SMT8 up to 256 threads in single partition

Thread Model:
IBM i 7.1 will support ST, SMT2, SMT4, and SMT8 and will remain limited to 256 threads in a partition. The ability to switch SMT mode without IPL will be PTFed back to IBM i 7.1.

IBM i v7.2   Cores and Threads for P7/P8

	IBM i 7.2
POWER 7	Max Scale = 32 cores (SMT4) Max Partition = **96 cores (SMT4)** Threads = ST, SMT2, SMT4 up to 384 threads in single partition
POWER 8	Max Scale = 48 cores (SMT8) Max Partition = **96 cores (SMT8)** Threads = ST, SMT2, SMT4, SMT8 up to 768 threads in single partition

Thread Model:
IBM i 7.2 will support ST, SMT2, SMT4 and SMT8, up to a limit of 768 threads in a single partition. It will support switchable SMT (not requiring a partition IPL to change modes)

IBM i 7.2

What's new:

- DB2 for i significant enhancements in the areas of Security, Performance and Usability
- IBM Navigator for i provides system and message monitoring as well as batch process capture and modeling
- Rational Developer for i supports Free Format RPG
- IBM Integrated Application Server built on WebSphere Liberty profile
- POWER8 and POWER7+ refreshed firmware - 770 & 780 supported by IBM i 7.2
- Support added for USB 3.0, SRIOV, WAN over LAN
- PowerHA Express Edition with HyperSwap
- BRMS Hub Function
- SSO application suite extended to include FTP and Telnet authenticating with Kerberos

Features / Business Value:

- Client enjoy new features and functions of DB2, WebSphere, Security, Systems Management, Business resiliency and development tools/compilers
    - ❖ Fully integrated and tested
- Secures critical business data

Client Benefits:

- Provides support for new features of POWER8 and POWER7 systems
- Protects critical business data
    - ❖ Ensures security of critical business data which is critical in today's environment of mobile, web and social access
- Improves system management – ease-of-use and and easier problem resolution
- Improved business resiliency – BRMS and PowerHA
- Many additional benefits

IBM i 7.2 Highlights

Powerful new features of DB2® for i ensure security of the data in a modern environment of mobile, social and network access
- Role based security
- Temporal support now included in DB2 for i

IBM Navigator for i extends system management capabilities to manage and monitor performance services

Announcing a new resiliency option; PowerHA Express for i with HyperSwap

Additional flexibility in configuration for I/O with new virtualization: Ethernet through SRIOV and WAN over LAN"

Backup and Recovery Management Systems (BRMS introduces
a Hub function for easier management

Support for new POWER8 systems, and new features for
POWER8 and POWER7
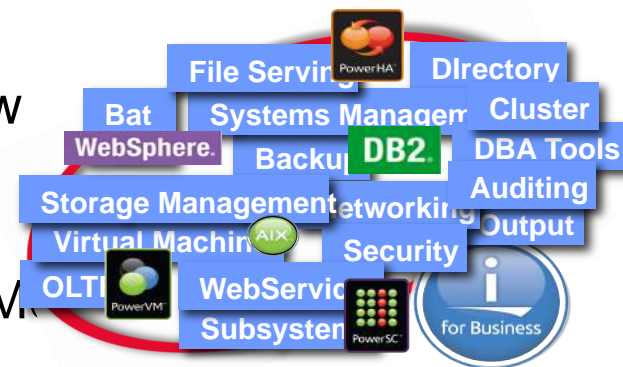
Linux on Power Systems: A powerful combination….

IBM Power Systems
Optimized for big data and analytics

Workload-based resource allocation

Automated, policy-based security

Linux Operating System
Rich opportunities for innovation

Significant cost savings

Industry-leading flexibility & performance

Delivering leadership solutions across IBM
- Optimized for mobile, big data, analytics and social
- Industry specific capabilities and reference configurations

Building a robust ISV ecosystem
- Power Systems Linux Centers worldwide
- Strong development support for community

Delivering industry leading virtualization
- Commitment to OpenStack & wide range of server deployment options
- Platform, virtualization & cloud management solutions

IBM Software

openstack
CLOUD SOFTWARE

KVM

PowerVM

<u>Linux</u> on Power Systems Solutions

IFL: Power Cores & Memory dedicated for Linux workloads

New

Power 795 IFL

Power 780 IFL

Power 770 IFL

PowerLinux 7R4

- 4-socket, 5U
- Up to 32 POWER7+ cores
- 1 TB memory
- Hot-swap PCI adapters
- Linux only

PowerLinux 7R2

- 2-socket, 2U
- POWER7+ processor
- Up to 16 cores
- 512 GB memory
- Linux only

PowerLinux 7R1

- 1-socket, 2U
- POWER7+ processor
- Up to 8 cores
- 256 GB memory
- Linux only

1, 2, and 4 socket systems

Power Enterprise IFLs

Available for POWER7 & POWER7+ Models Power 770 / 780 / 795

Power Enterprise Linux Engines

Power IFL: Power Integrated Facility for Linux

- Power Cores dedicated for Linux workloads
- Discounted pricing for Cores, Memory, and PowerVM

Structure of the offering

- Single Feature Code: 4 Cores & 32 GB memory
- Available for Power 770, Power 780 and Power 795
- PowerVM EE licensed for Power IFLs
- HWMA and SWMA for PowerVM for Linux priced separately
- Can have multiple FC's on a single server

T's & C's

- Clients must sign a contract addendum agreeing to segregation of cores and memory for Linux activities
- No converting to AIX & IBM i environments

Linux license acquired separately

Power 770

Power 780

Power 795

Power IFL Structure

Power IFL Cost Savings

Lower Processor core activations

Lower Memory activations

Lower PowerVM Costs

$

Lower Software Cost   PVU = 70

$

Lower Software Maint

For PowerVM for PowerLinux

$

Lower cost core Hardware Maint after warranty

$

Power Systems' Linux momentum is growing

Ecosystem Leadership

OpenPOWER Foundation

Open development alliance based on POWER hardware and software

Global Linux ISV Development

400+ Global Linux ISVs

200+ new ISVs across GCG with 400 applications

Power Systems Linux Centers

Beijing

Austin

New York

KVM

Montpellier

IBM Software Optimization

IBM WATSON

150+ SWG Optimized Offerings

WebSphere	Rational	Industry
Tivoli	Cognos	Solutions
	SPSS	
Information Management		Security

Industry-Leading Benchmarks

Over twice the throughput of x86/ VMware with TPoX on DB2

36% faster than x86 with SPEC Java on WebSphere

A new generation of **IBM Power Systems** optimized for scale-out data and cloud infrastructures

Open Innovation to put data to work

POWER8

Linux-only systems
offer better economics for Linux deployments

Ubuntu Server for Power
brings the fastest growing cloud OS to Power Systems

PowerKVM

simplifies management of heterogeneous data centers

Target workloads for Linux on Power Systems

Virtualized Application Infrastructure
Java, PHP, web, networking

Analytics & Research
Business intelligence, structured & unstructured data, predictive analytics, HPC

Data Services
Database, enterprise content management

Business Applications
Mobile data access, ERP, cloud delivery, application development

Workload Requirements

Highly threaded
Throughput oriented
Scale out capable
Economic efficiency

Compute intensive
High memory bandwidth
Floating point
High I/O rates

Handle peak workloads
Scalability
High quality of service
Resiliency and security

High quality of service
Scalability
Flexible infrastructure
Large memory footprint

What Linux Distributions in various Power Environments?

Linux	Release	Endian	Dedicated LPAR	PowerVM Guest	PowerKVM Guest
Redhat	5.10	Big	Y	Y	N
Redhat	6.4	Big	Y	Y	N
Redhat	6.5	Big	Y	Y	Y
SUSE	11 SP3	Big	Y	Y	Y
Ubuntu*	14.04	Little	N	N	Y

*Exploits P8

1. Select the applications you want to run on Linux on Power
2. Then look at the Linux distributions that are available for those apps
3. Pick your Linux distribution of choice

Linux for Scale-out Servers

	1H / 2014	2H / 2014
RHEL6	*RHEL 6.5* *POWER7 Mode on POWER8*	*RHEL 6.6* *POWER7 mode on POWER8*
RHEL 7	*RHEL 7.0*	
SLES 11	*SLES 11 + SP3* *POWER7 Mode on POWER8*	
SLES 12		*SLES 12 (LE)* *POWER8 Support*
Ubuntu (LE)	*14.04.00/01* *POWER8 Support*	*14.04.00/02*
Debian	*LE Introduction* *POWER8 Support*	*LE Update*
VIOS	*2.2.3.3*	*2.2.3.X*

Learn more about Linux on Power Systems

Power Systems Linux Portal
(Product Information)

www.ibm.com/systems/power/software/linux/

The PowerLinux Community
(developerWorks)

www.ibm.com/developerworks/group/tpl/

@ibmpowerlinux

plus.google.com/communities/100156952249293416679

OpenPower
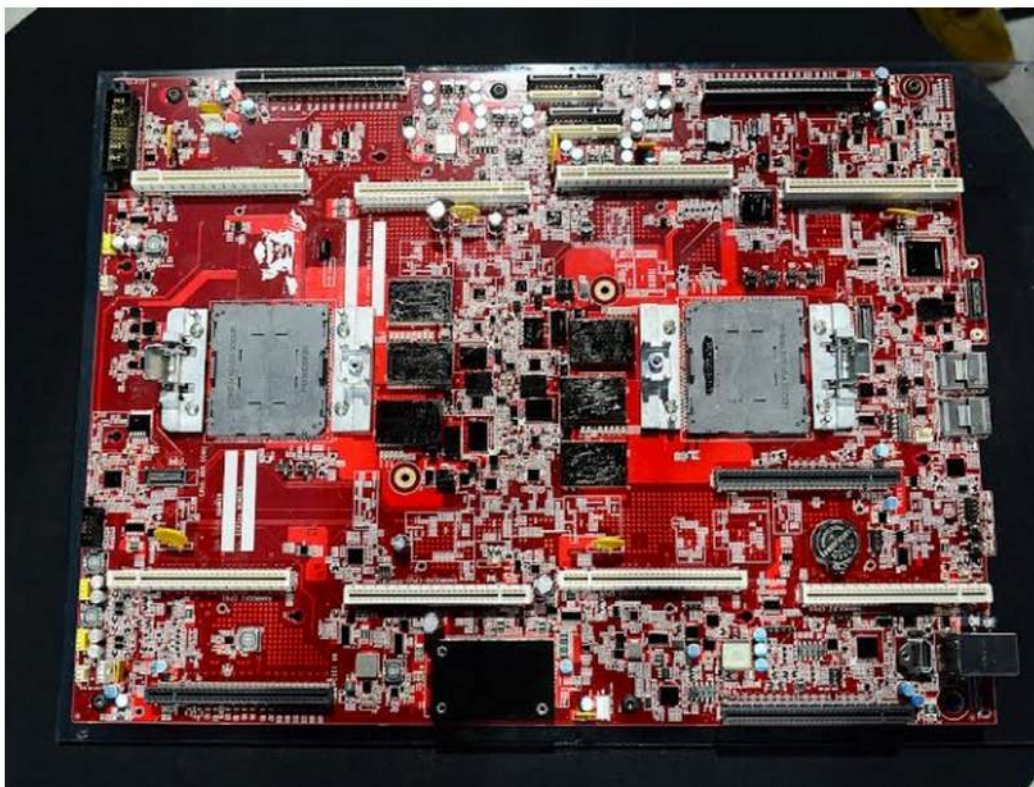
Image: Google

OpenPOWER™ Foundation Members

How will the OpenPOWER Consortium benefit clients?

- New Innovators on the Power platform will create more value
- OpenPOWER technology creates greater choice for our customers
- Innovation will increase adoption of the Power Systems platform

What does this mean to the industry?

- Game changer on the competitive landscape of the server industry
- Will enable and drive innovation in the industry
- Provide more choice in the industry

PowerKVM

Power Virtualization Options

PowerKVM

Initial Offering: Q2 2014

- **PowerKVM: Open Source option for virtualization on Power Systems for Linux workloads.**
- **For clients that have Linux centric admins. (RHEL 6.5 & SLES 11.3)**

PowerVM

Initial Offering: 2004

PowerVM: Provides virtualization of Processors, Memory, Storage, & Networking for AIX, IBM i, and Linux environments on Power Systems.

Architecture
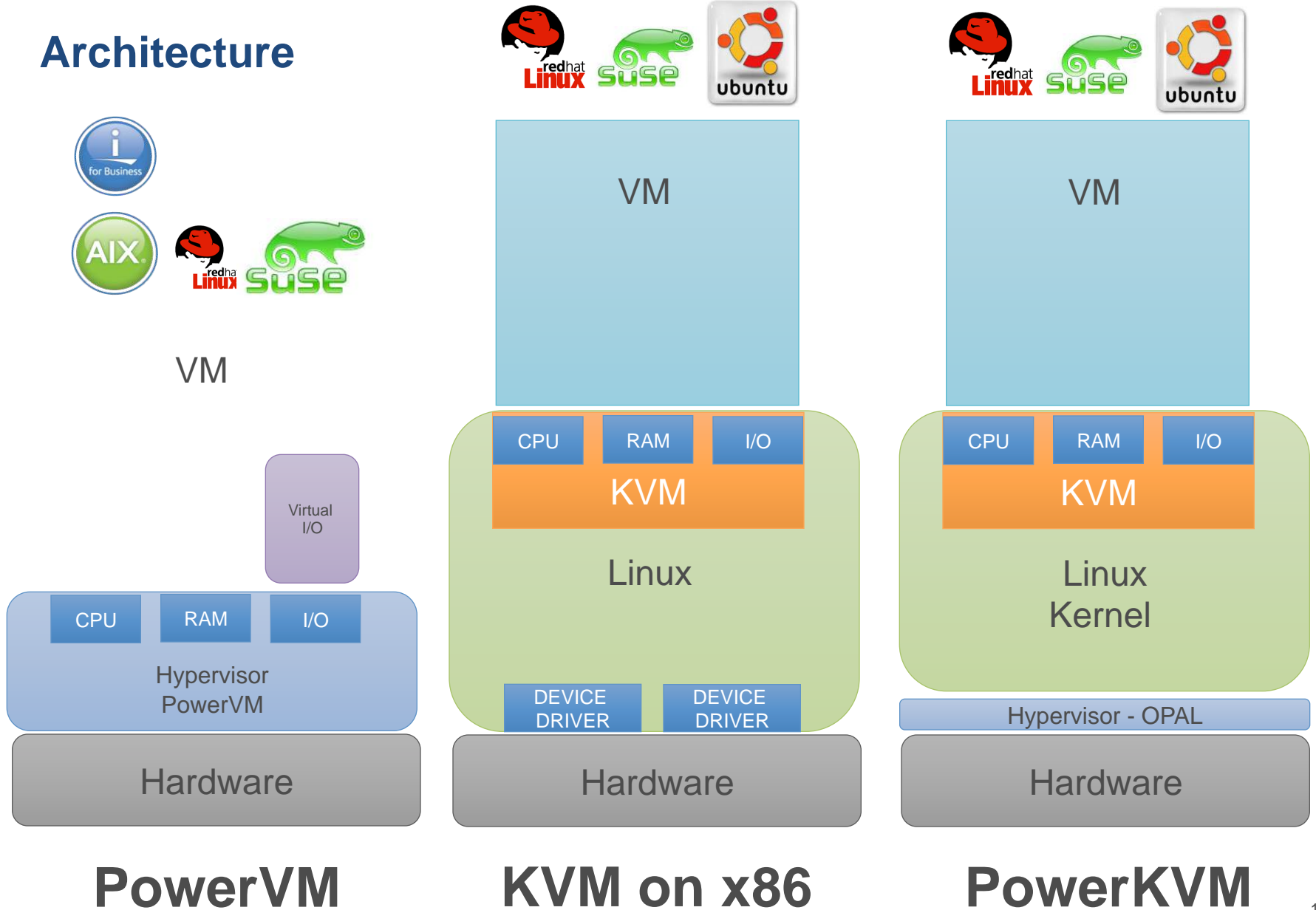
PowerVM

KVM on x86

PowerKVM

PowerKVM v2.1

Open Virtualization Choice for Linux-only Scale-out Servers

NEW

KVM

- ✓ Optimize **Linux Workload Consolidation and scale out** of workloads at a lower cost of ownership
- ✓ Maintain flexibility and agility by exploiting **Open Source** Community
- ✓ **Leverage** traditional **Linux admin skills** on Power Systems to administer virtualization
- ✓ Use **open source tools like OpenStack** to manage virtualization

Announce – 4/28 GA – 6/10

- ✓ *Reduces IT Infrastructure costs*
- ✓ *Optimize Linux workload consolidation at a lower cost*
- ✓ *Simplify your virtualization management using open source tools*

- • Kernel-Based Virtual Machine(KVM) Linux based virtualization For Scale Out POWER8 Linux Servers
- • Processor and memory sharing and over commitment enables higher VM and workload consolidation
- • Dynamic addition & removal of virtual devices
- • Live VM Migration enables higher availability and allows workload balancing
- • Exploits P8 Features like Micro-Threading providing greater scheduling granularity vs x86 virtualization
- • Exploits performance, scalability and security built into Linux
- • Managed by PowerVC and open source tools which provides flexible familiar Linux admin tools
- • Supports Redhat, SUSE, Ubuntu Linux Guests

PowerKVM

Core Handling

Full Core

2 Threads	4 Threads	1 Thread	1 Thread
VM A 2 vCPUs	VM B 4 vCPUs	VM C 1 vCPU	VM D 1 vCPU

KVM

8 Threads — 8 Threads

Linux Kernel

Hypervisor - OPAL

SMT1-8   1 core

SPLIT Core

2 Threads	2 Threads	2 Threads	2 Threads
VM A 2 vCPUs	VM B 2 vCPUs	VM C 2 vCPUs	VM D 2 vCPUs

CPU   CPU   CPU   CPU

KVM

2 Threads | 2 Threads | 2 Threads | 2 Threads

Linux Kernel

Hypervisor - OPAL

SMT1-2

1 core

158
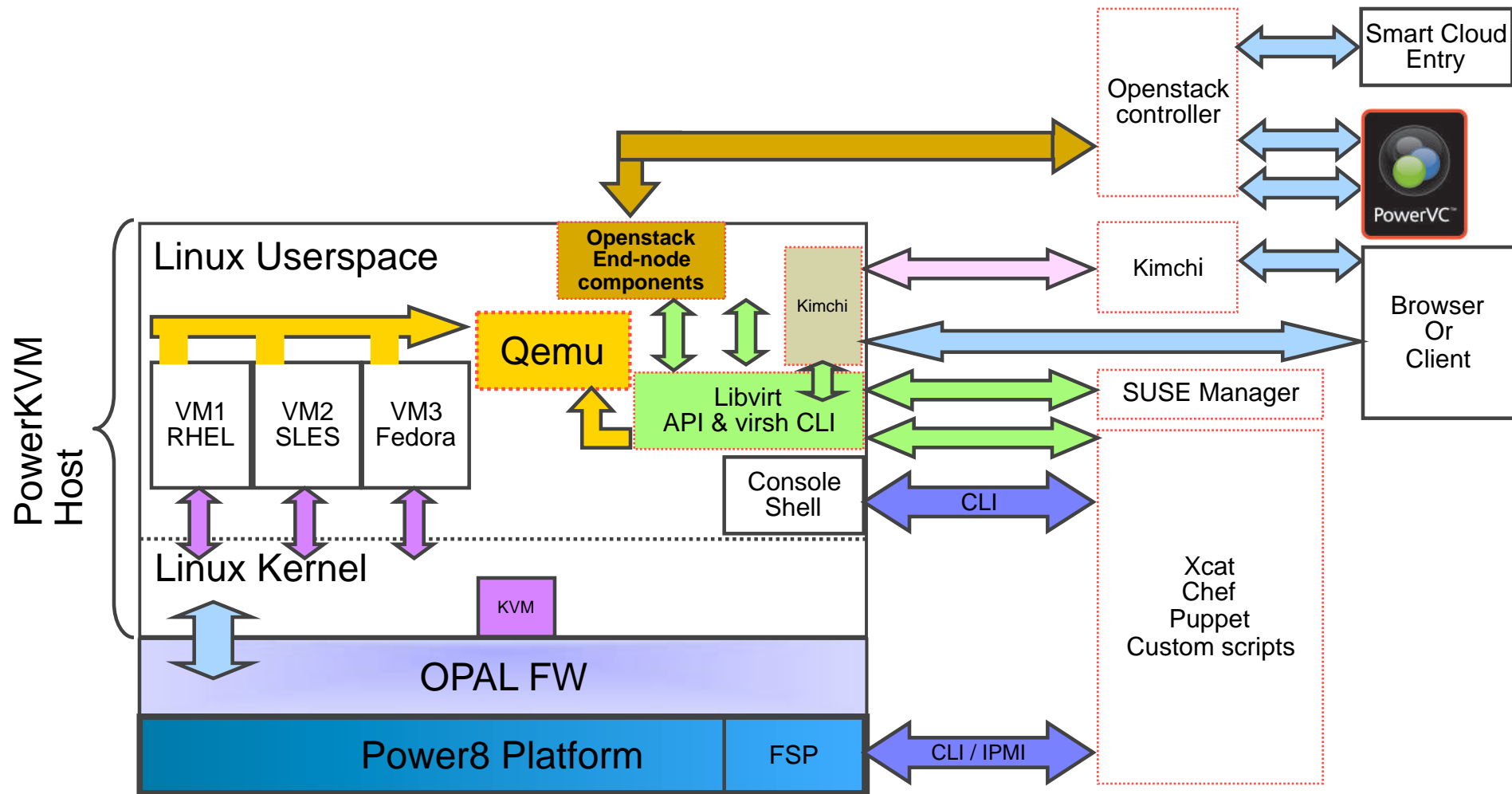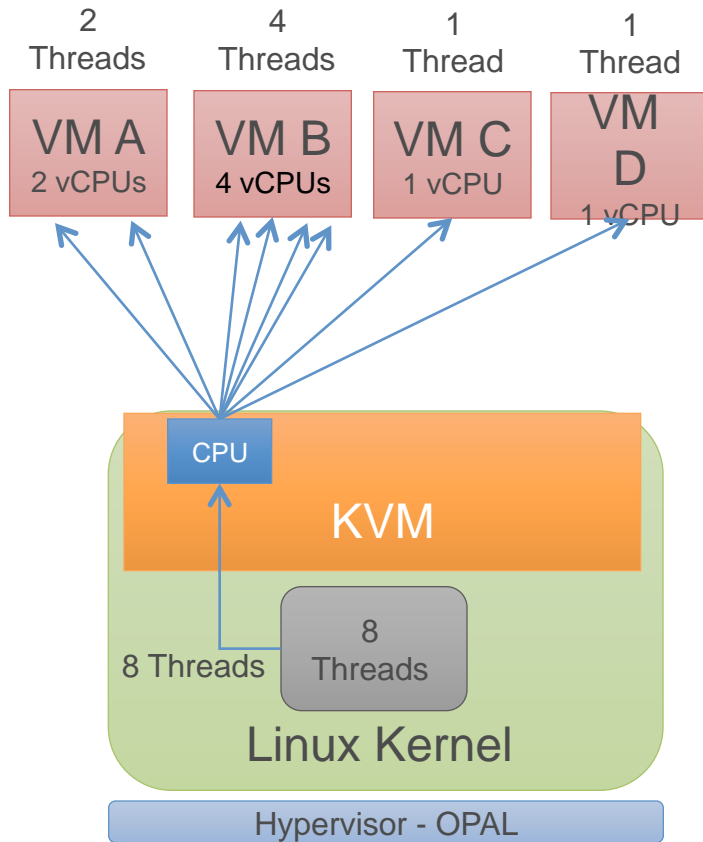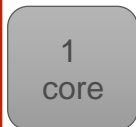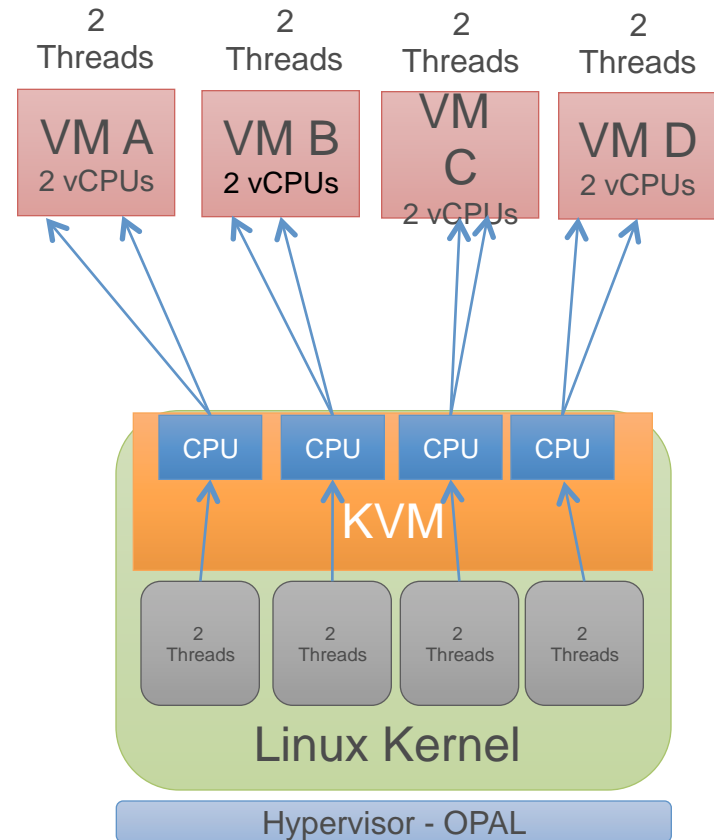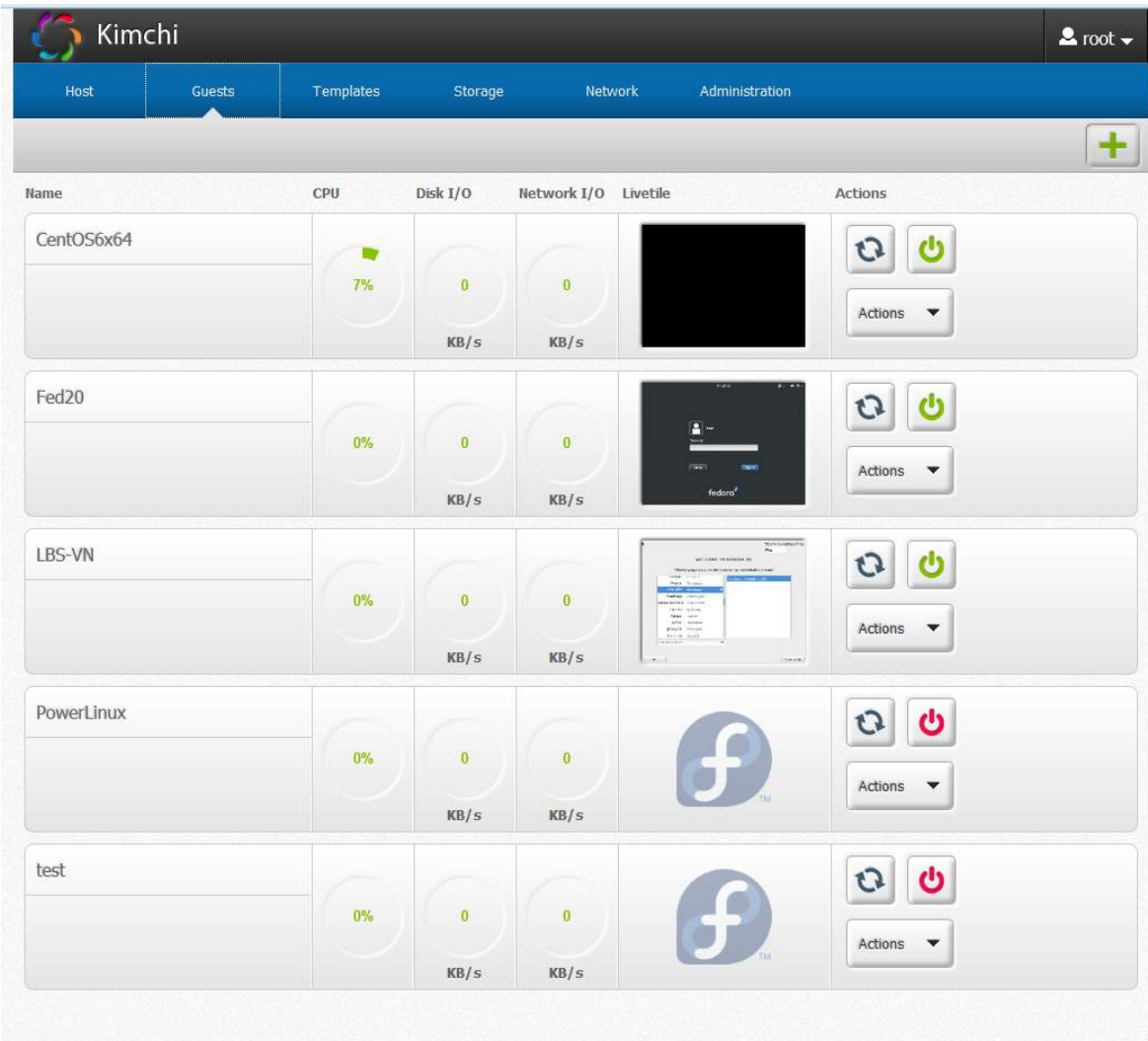
PowerVM vs PowerKVM Comparison

	PowerVM	PowerKVM
GA Availability	2004	Q2 2014
Supported Hardware	All P6, P7, P7+, P8 Systems	PowerLinux P8 Systems S812L & S822L
Supported OS	AIX, IBM i & Linux	Linux
Workload Mobility	Supports AIX, IBM i & Linux	Linux
Basic Virtualization Management	IVM / HMC / FSM	Virtman/libvirt
Advanced Virtualization Management	PowerVC/VMControl	PowerVC, Vanilla OpenStack
Admin Type	Power Centric	Linux/x86 Centric
Established Security Track Record on Power	Yes	No
Open Source Hypervisor	No	Yes

© 2

Kimchi Local Administration

Provides simple graphical web interface to initially configure the PowerKVM Host and to manage basic virtualization for a small configuration.
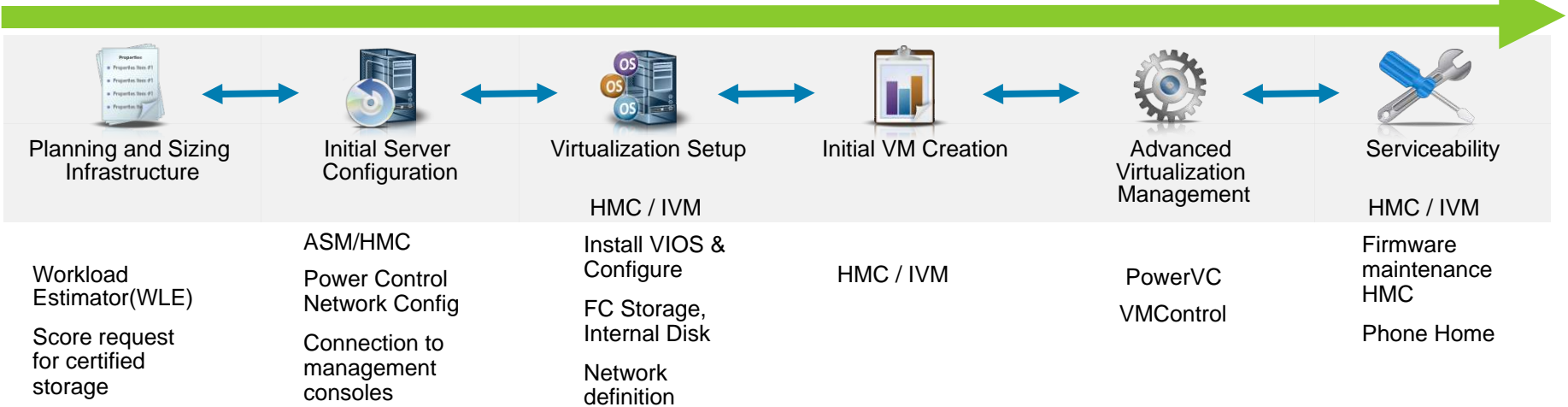
Included in PowerKVM distribution.
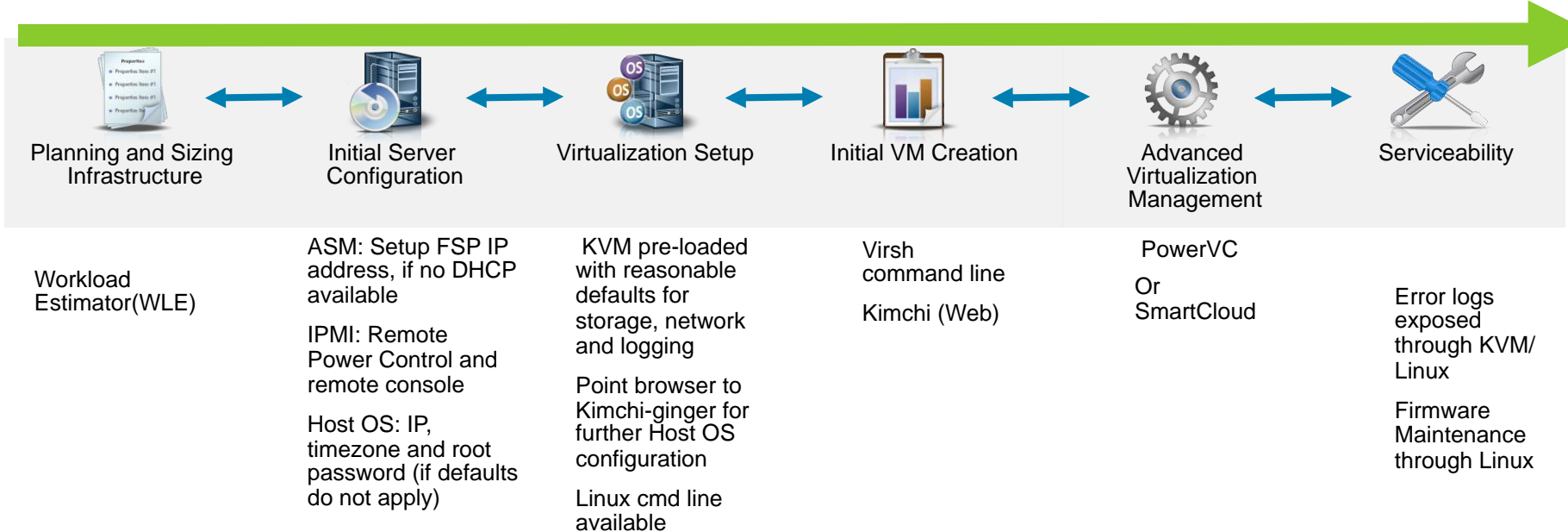
Function includes
✓Initial host setup
✓Firmware update
✓Backup of configuration
✓Simple VM setup
✓Start and stop of VMs
✓Host monitoring
✓Use of Templates
✓View VM guest console

© 2014 IBM Corporation

PowerVM VS KVM Out of Box Experience

Power Systems

IBM

PowerVM

Planning and Sizing Infrastructure	Initial Server Configuration	Virtualization Setup	Initial VM Creation	Advanced Virtualization Management	Serviceability
		HMC / IVM			HMC / IVM
Workload Estimator(WLE)	ASM/HMC	Install VIOS & Configure	HMC / IVM	PowerVC	Firmware maintenance HMC
Score request for certified storage	Power Control Network Config	FC Storage, Internal Disk		VMControl	Phone Home
	Connection to management consoles	Network definition			

PowerKVM

| Planning and Sizing Infrastructure	Initial Server Configuration	Virtualization Setup	Initial VM Creation	Advanced Virtualization Management	Serviceability
Workload Estimator(WLE)	ASM: Setup FSP IP address, if no DHCP available	KVM pre-loaded with reasonable defaults for storage, network and logging	Virsh command line	PowerVC	Error logs exposed through KVM/ Linux
	IPMI: Remote Power Control and remote console		Kimchi (Web)	Or SmartCloud	
	Host OS: IP, timezone and root password (if defaults do not apply)	Point browser to Kimchi-ginger for further Host OS configuration			Firmware Maintenance through Linux
		Linux cmd line available			

PowerKVM Offering Structure

- PowerKVM has a single Standard Edition = PID 5765-KVM

- Socket Based Pricing with Subscription model for the PowerKVM host
    - Pricing is socket based and requires buying a socket pair subscription 8x5 or 24x7 offering

- POWER8 Scale-out Linux Only Servers have a choice of either PowerKVM or PowerVM
    - Only one virtualization option can be chosen
    - Either PowerVM or PowerKVM can run on the system but not both

- When PowerKVM is ordered the feature code drives proper firmware defaults
    - PowerKVM is preloaded on internal disk with option to disable preload
    - Guest Linux ISO can be preloaded for convenience to build guest VMs

Endianess

- Endianess is how a system handles data
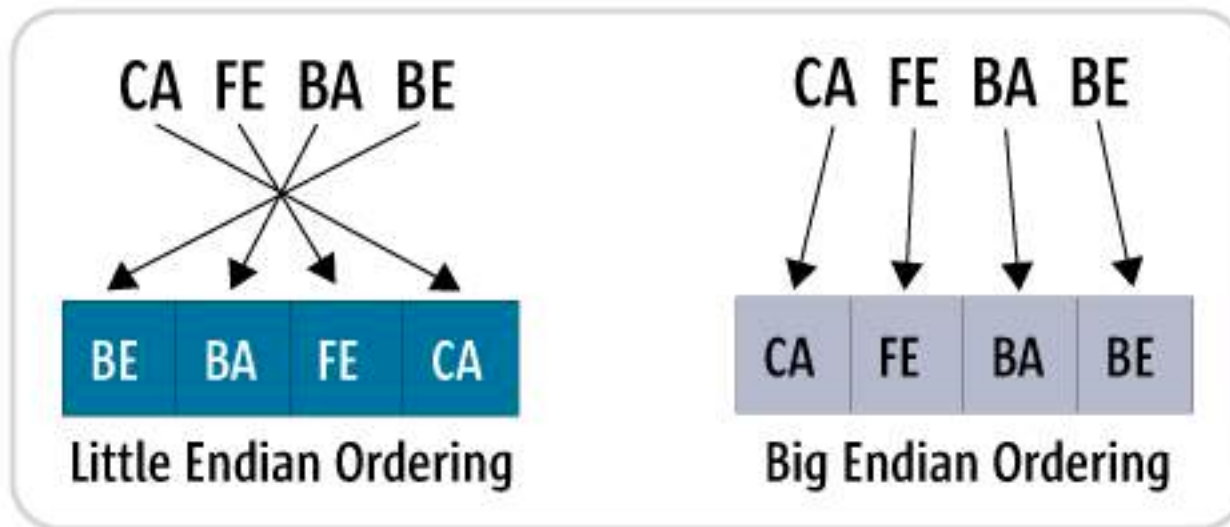
- x86 is Little Endian
- PowerVM is Big Endian
  - AIX, IBM i, RHEL, SUSE on top of PowerVM must be Big-Endian

- POWER is Bi-endian !!!

Endianess

- Endianess is how a system handles data

```
#include <stdio.h>
int main(void) {
int val;
unsigned char *ptr;

ptr = (char*) &val; /* pointer â€˜ptrâ€™ points to â€˜valâ€™ */
val = 0x89ABCDEF; /* four bytes constant */
printf("%X.%X.%X.%X\n", ptr[0], ptr[1], ptr[2], ptr[3]);
exit(0);
}
```

On an x86 system, the result is:

```
EF.CD.AB.89
```

On a POWER system, the result is:

```
89.AB.CD.EF
```

- Linux Apps with this kind of programming logic had to have its code changed to run on Power properly

PowerKVM Positioning

- **First release available in 2014**
- **Focus:** New Linux workloads for Power Systems
- **Seamless transition for existing Linux admins** to adopt Power Linux Virtualization without any training
- **No HMC or other traditional IBM consoles**
  - Normal Linux management and OpenStack options
- **PowerKVM only supports Linux guest VMs**
- **Cloud potential:** Have many more small VMs than traditional Power Virtualization
- **POWER8 PowerLinux hardware only**
- **Live Workload mobility** support between PowerKVM servers
- **Open Source Hypervisor:** Hardware is abstracted by firmware
- **Managed by OpenStack(PowerVC) or by off the shelf OpenStack or local Linux Tools**

More Information on PowerKVM

ibm.com/systems/power/software/linux/powerlinux/powerkvm

IBM Systems > Power Systems > Software > Linux > PowerLinux solutions >

IBM PowerKVM – open virtualization

PowerKVM – the open virtualization choice for Power scale-out Linux Systems

| Overview | Features & benefits |

Highlights

- Open source server virtualization for IBM scale-out Linux servers

- Optimize Linux workload consolidation at a lower cost of ownership

- Exploit the advantage of performance, scalability and security built into Linux and the Kernel-Based Virtual Machine (KVM) hypervisor

- Avoids high cost proprietary x86 virtualization

- Managed just like any other KVM host – OpenStack, libvirt and open Linux tools help you avoid vendor lock-in

- KVM enables single cross platform virtualization which simplifies management

- Maintain flexibility, extensibility and agility by exploiting open source solutions

- Leverage traditional Linux admin skills on Power Systems

- Exploit new advanced POWER8 processor features

- Eliminate scheduled downtime by using live VM migration

IBM® PowerKVM™ provides an open virtualization choice for IBM scale-out Linux systems based on the POWER8™ technology. This solution includes the Linux open source technology of KVM virtualization, and is designed to complement the performance, scalability and security qualities of Linux. This provides an open extendable solution for running VMs on Linux scale-out servers that enables cloud deployments, scale-out processing, and big data solutions reducing complexity and cost.

Learn more

- View IBM PowerKVM Open Virtualization data sheet (763KB)
- → IBM PowerVM for IBM PowerLinux
- Get Adobe Reader

Contact IBM

- Chat now
- Email IBM
- → Find a Business Partner
- Call IBM: 1-866-883-8901
  Priority code: 101AR13W

Browse Power Systems

- Hardware
- Operating systems
- System software

- Solutions
- Migrate to Power
- Advantages

- → Community
- → Success stories
- → News

- → Support & services
- → Resources
- → Education