

Форум

**"Современный подход
к построению ИТ-инфраструктуры.
Практика IBM."**

17 марта 2015 года, Санкт-Петербург

IBM POWER

Алексей Бойко

IBM Systems Product Manager

Alexey.Boyko@ru.ibm.com



TOP 10 US Patent Grants Totals

▪ IBM	7 534
▪ Samsung	4 952
▪ Canon	4 055
▪ Sony	3 224
▪ Microsoft	2 829
▪ Toshiba	2 608
▪ QUALCOMM	2 590
▪ Google	2 566
▪ LG Electronics	2 122
▪ Panasonic	2 095

Intel #16 1 578

HP #17 1 474

Oracle не входит в Top50

Source: IFI Patent Intelligence

http://www.ificlaims.com/index.php?page=misc_top_50_2014.



Количество патентов IBM за 2014 больше патентов Accenture, Amazon, Google, HP, Intel и Oracle вместе взятых

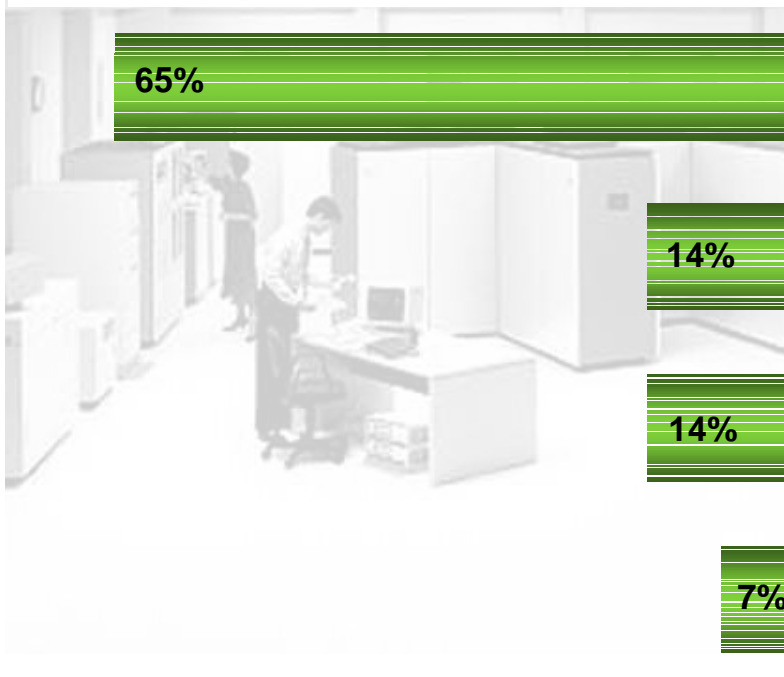
Система ценностей клиента



Сегодняшние реалии отличаются от вчерашних

1996

2011



Hardware

11%

Software

33%

People

46%

Other

10%

Структура затрат на работающую систему очень сильно изменилась. Использование прогрессивных технологий «железа» позволяет сократить расходы на ПО и обслуживающий персонал.

Какие критерии применяются при выборе платформы для важных задач?

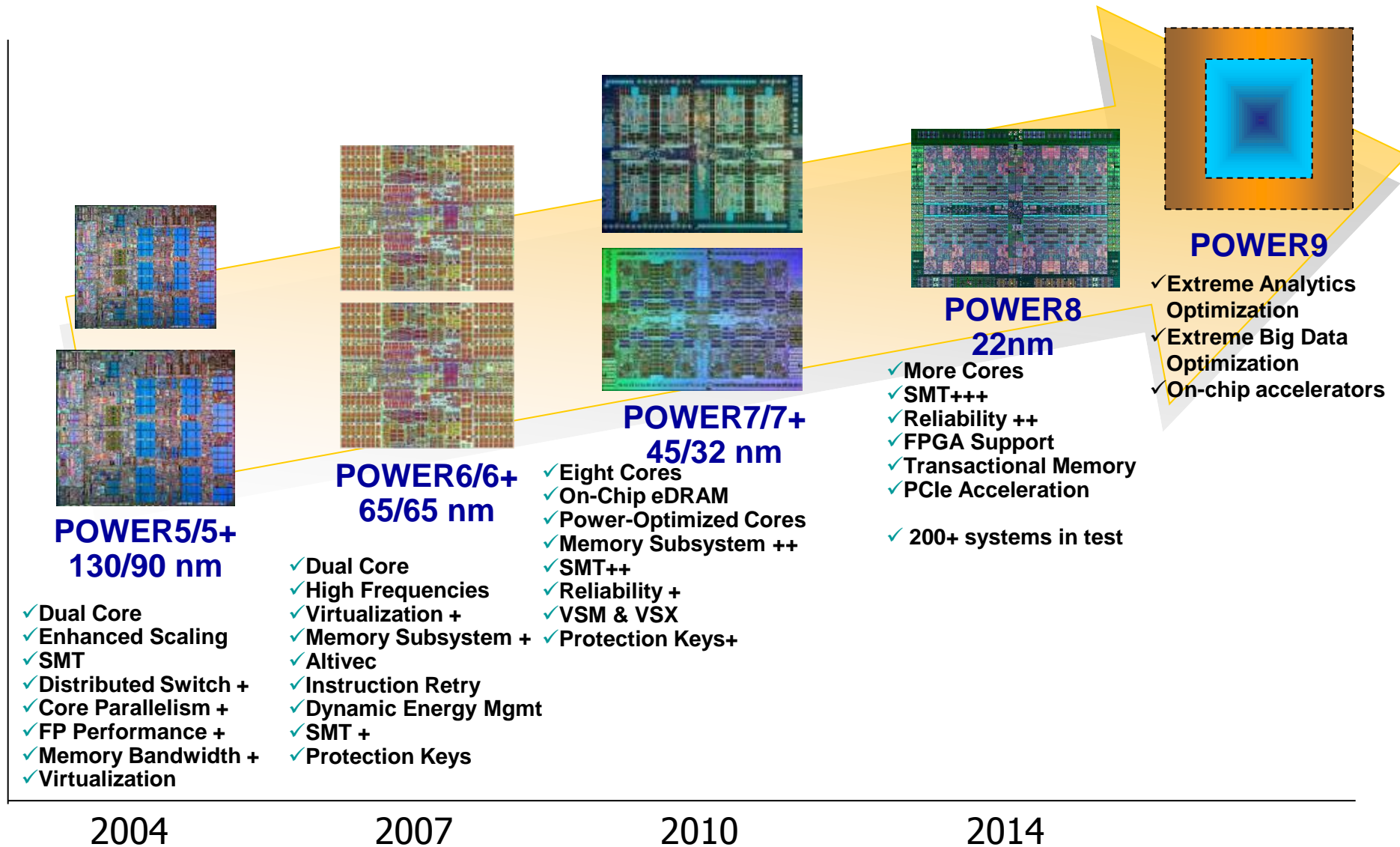
- **Защита инвестиций** – минимизация изменений в будущем (как «железа» так и программного обеспечения)
- **RAS-Reliability (надежность), Availability (доступность) и Serviceability (сервисность)**
- **Производительность и «плотность» нагрузок**
- **Масштабируемость** как для планируемого, так внепланового роста
- **Безопасность**
- **И снижение стоимости владения**



Немного о процессорах

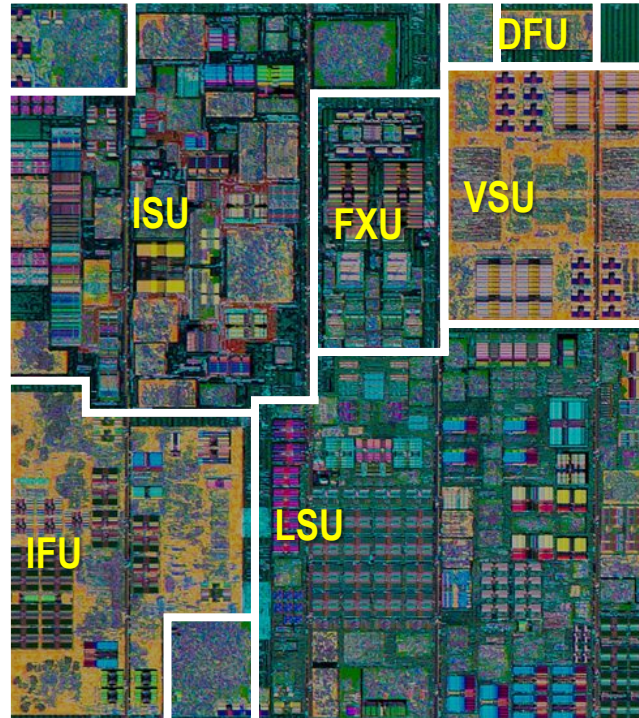
IBM Power





Execution Improvement vs. POWER7

- SMT4 → SMT8
- 8 dispatch
- 10 issue
- 16 execution pipes:
 - 2 FXU, 2 LSU, 2 LU, 4 FPU, 2 VMX, 1 Crypto, 1 DFU, 1 CR, 1 BR
- Larger Issue queues (4 x 16-entry)
- Larger global completion, Load/Store reorder
- Improved branch prediction
- Improved unaligned storage access



Larger Caching Structures vs. POWER7

- 2x L1 data cache (64 KB)
- 2x outstanding data cache misses
- 4x translation Cache

Wider Load/Store

- 32B → 64B L2 to L1 data bus
- 2x data cache to execution dataflow

Enhanced Prefetch

- Instruction speculation awareness
- Data prefetch depth awareness
- Adaptive bandwidth awareness
- Topology awareness

Core Performance vs . POWER7

~1.6x Thread
~2x Max SMT

Technology

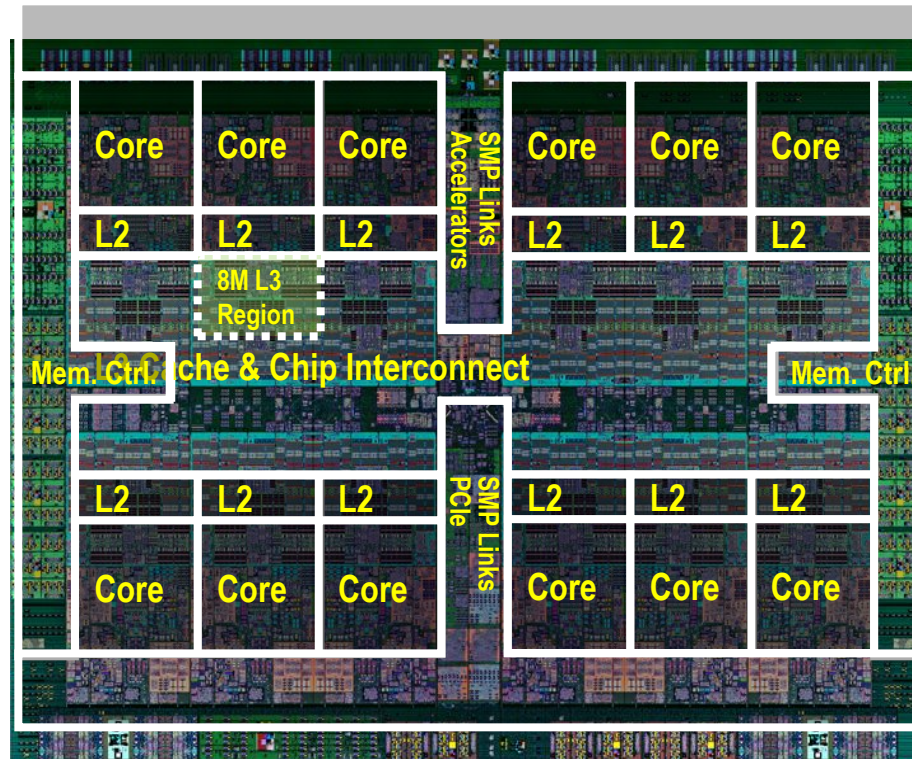
22nm SOI, eDRAM, 15 ML 650mm²

Cores

- 12 cores (SMT8)
- 8 dispatch, 10 issue, 16 exec pipe
- 2X internal data flows/queues
- Enhanced prefetching
- 64K data cache, 32K instruction cache

Accelerators

- Crypto & memory expansion
- Transactional Memory
- VMM assist
- Data Move / VM Mobility



Energy Management

- On-chip Power Management Micro-controller
- Integrated Per-core VRM
- Critical Path Monitors

Larger Caches

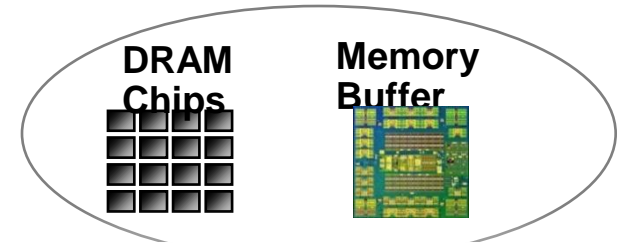
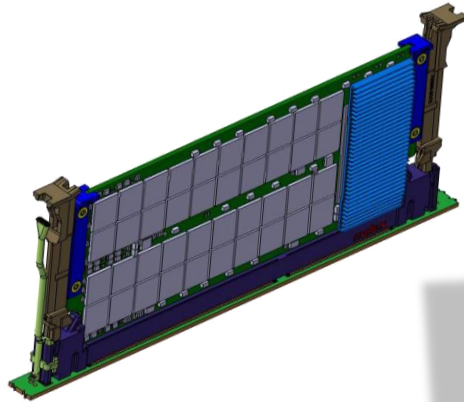
- 512 KB SRAM L2 / core
- 96 MB eDRAM shared L3
- Up to 128 MB eDRAM L4 (off-chip)

Memory

- Up to 230 GB/s sustained bandwidth

Bus Interfaces

- Durable open memory attach interface
- Integrated PCIe Gen3
- SMP Interconnect
- CAPI (Coherent Accelerator Processor Interface)



Intelligence Moved into Memory

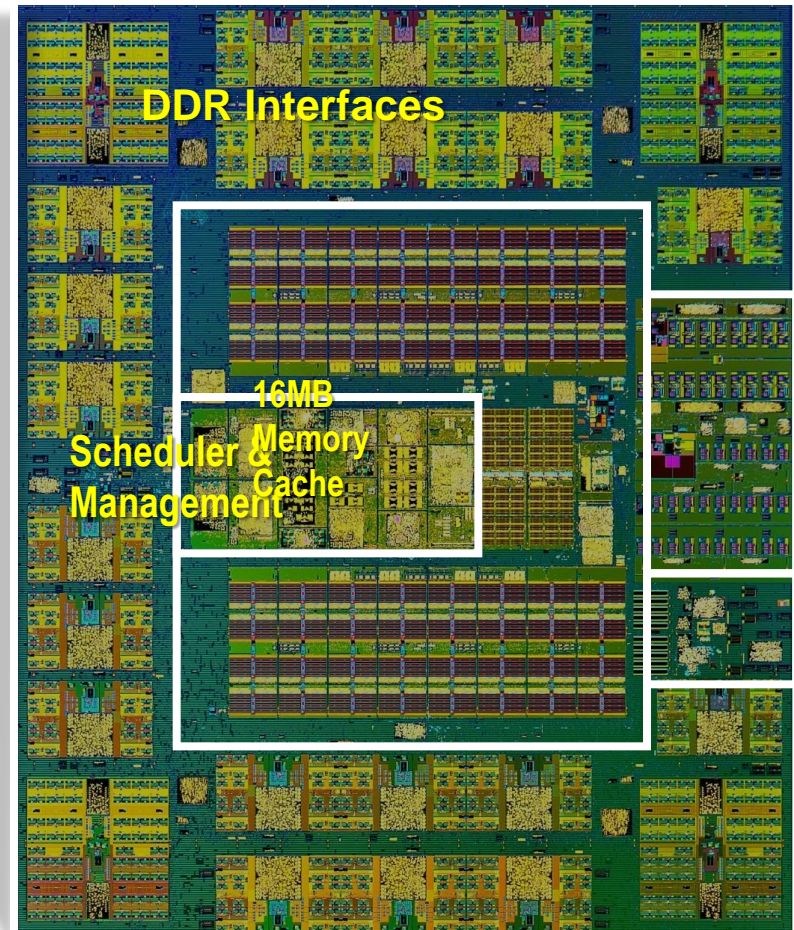
- Scheduling logic, caching structures
- Energy Mgmt, RAS decision point
 - Formerly on Processor
 - Moved to Memory Buffer

Processor Interface

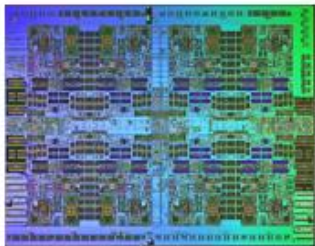
- 9.6 GB/s high speed interface
- More robust RAS
- “On-the-fly” lane isolation/repair
- Extensible for innovation build-out

Performance Value

- End-to-end fastpath and data retry (latency)
- Cache → latency/bandwidth, partial updates
- Cache → write scheduling, prefetch, energy
- 22nm SOI for optimal performance / energy
- 15 metal levels (latency, bandwidth)



POWER7



GX Bus

I/O Bridge

PCIe Gen2

PCI Devices

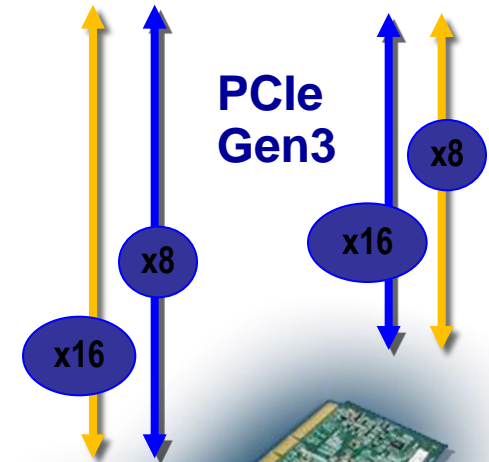
Native PCIe Gen 3 Support

- 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

POWER8



PCI Devices



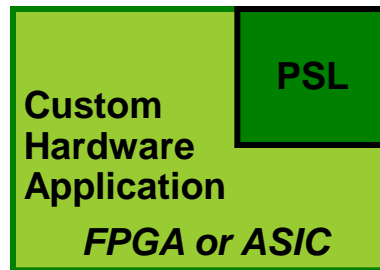
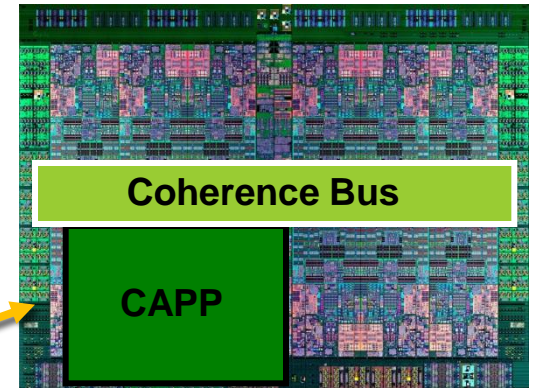
Virtual Addressing

- 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

POWER8



PCIe Gen 3

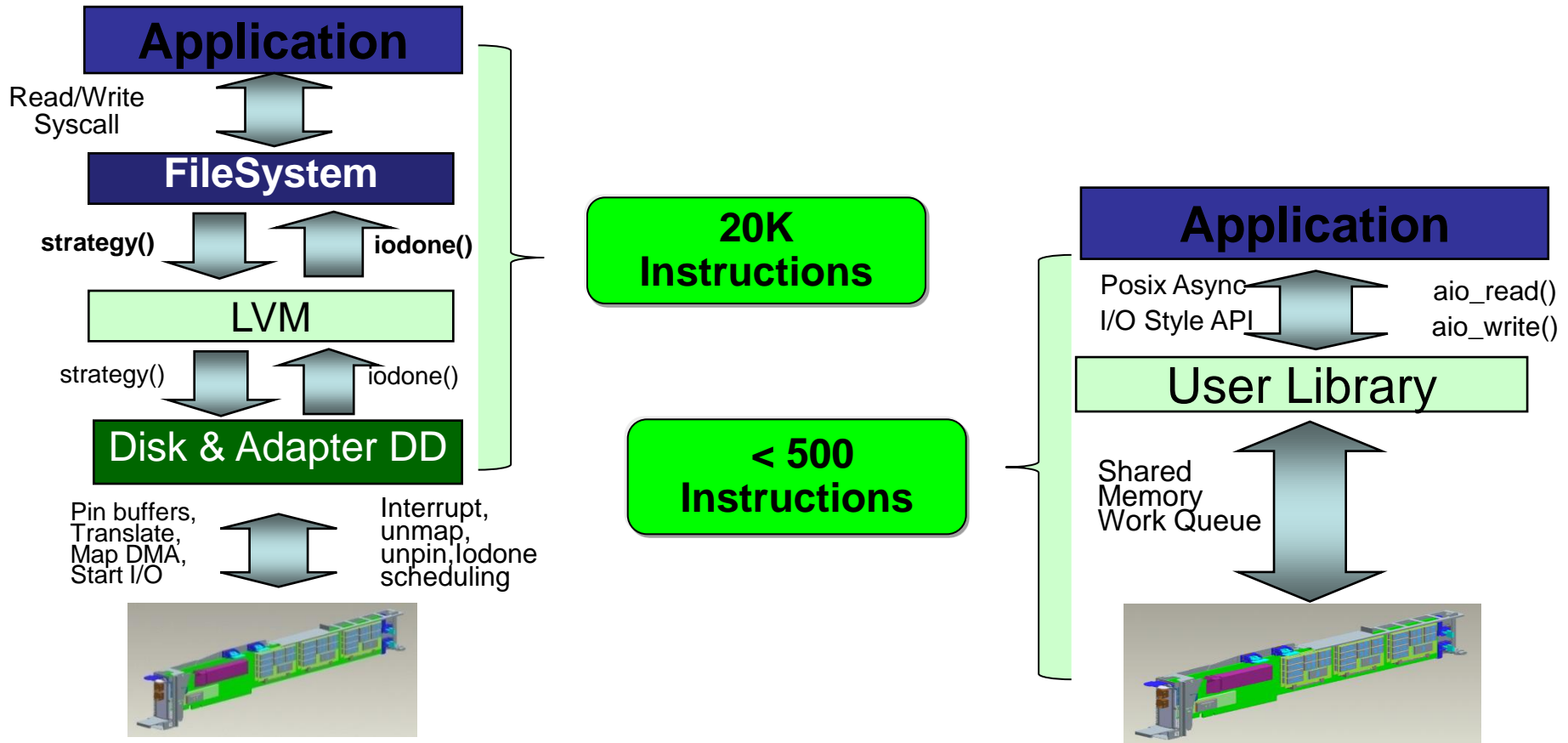
Transport for encapsulated messages

Processor Service Layer (PSL)

- Present robust, durable interfaces to applications
- Offload complexity / content from CAPP

Customizable Hardware Application Accelerator

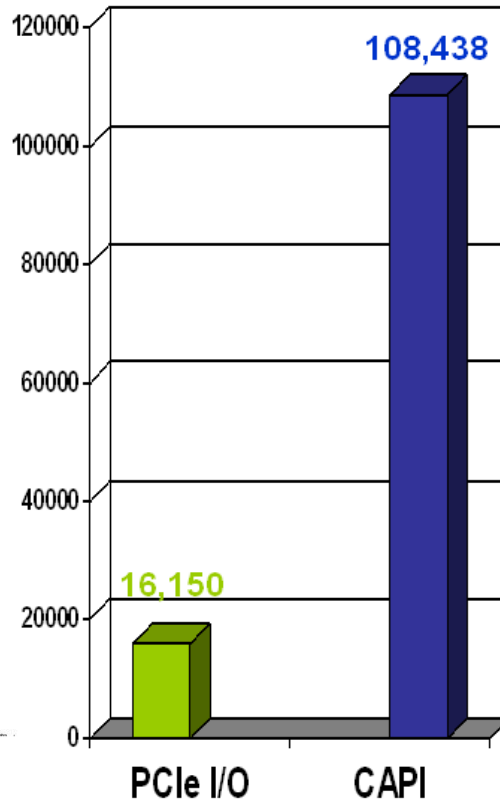
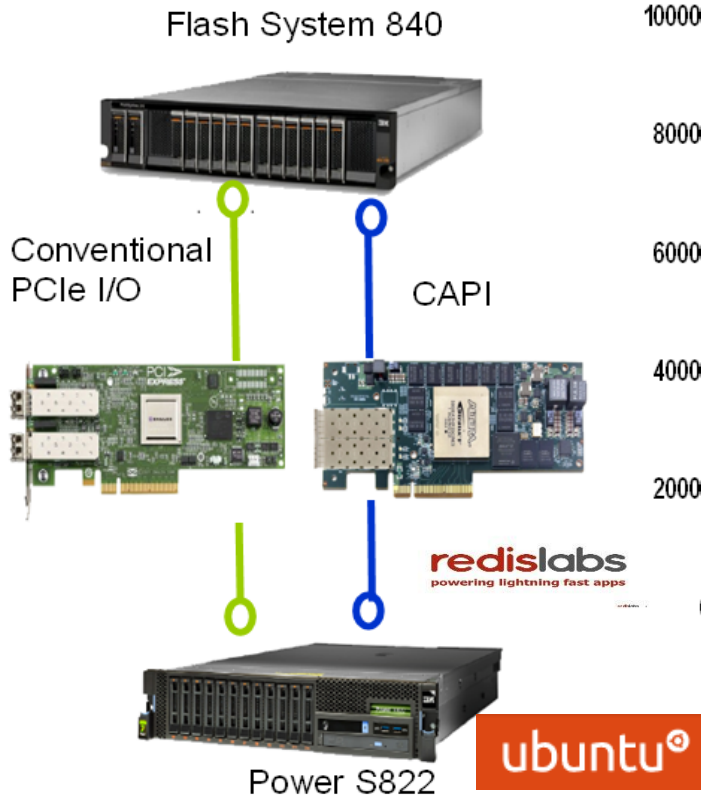
- Specific system SW, middleware, or user application
- Written to durable interface provided by PSL



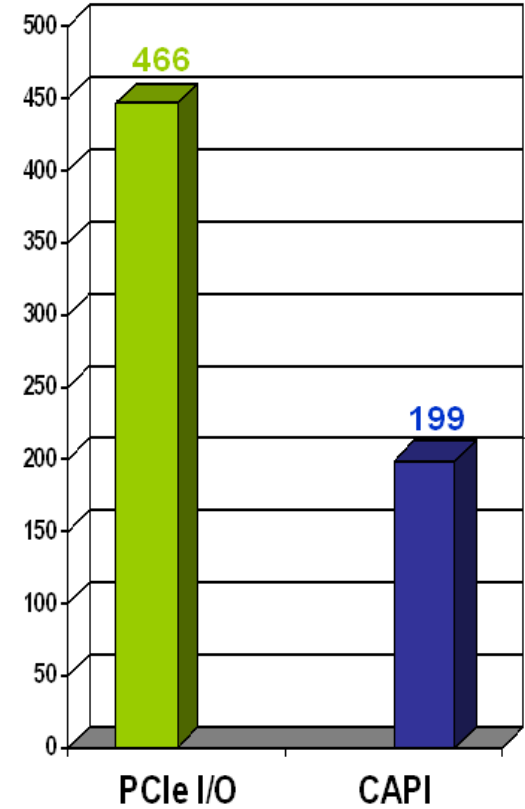
- Attach flash memory to POWER8 via CAPI coherent Attach
- Issues Read/Write Commands from applications to eliminate 97% of instruction path length
 - CAPI Flash controller Operates in User Space
- Saves 10 Cores per 1M IOPs

IBM Data Engine for NoSQL

Identical hardware with 2 different paths to data



IOPs per HW Thread



Latency (us)

RAS

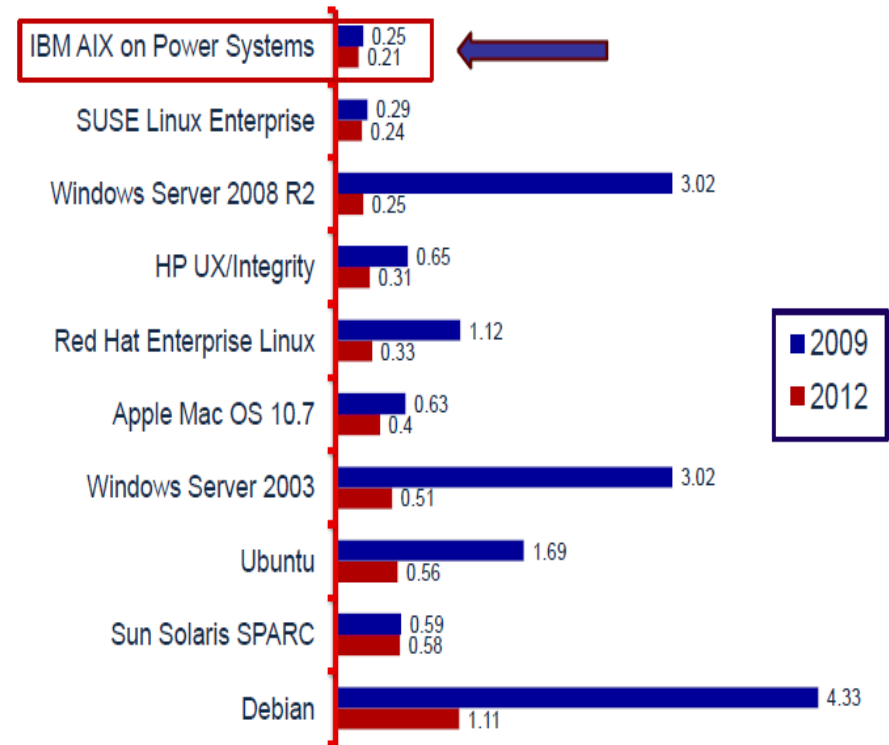


Высокая доступность - ключевое требование

Решения IBM Power обеспечивают доступность **99.997+%**

- Almost five 9's availability, the gold standard
- 67% of corporations now require a minimum of 99.99% uptime or better for mission critical hardware, operating systems and main line of business (LOB) applications
- AIX on Power consistently has the least amount of downtime in ITIC studies for several years
- Industry leading availability for all workloads, including SAP

Corporate Enterprise Downtime 2009 vs. 2012 (Hours per Year)



Source: ITIC 2013 Global Server Hardware, Server OS Reliability Survey, ITIC, (All rights reserved); January 2013.

Тестовые компоненты и регистры изоляции ошибок составляют **First Failure Data Capture**

FFDC ключевой компонент надежности IBM Power

	Checkers	FIRs
POWER8 S812L	30,068	13,200
POWER8 S824	60,136	26,400
Power 740	37,720	5,800
Power 770 (one drawer)	39,800	6,500
Power 795 (Maximum configuration)	598,000	96,000
Intel Ivy Bridge EX	5,500	none

Power Service 3
Processor Fixes!



FIR: 2
(Fault Isolation Register)
Error data from the checkers are collected and stored here. The service processor monitors the contents



Checkers: 1
Built-in HW error diagnostic probes to identify error conditions



Power HW / FW



Power Systems обеспечивают лучшую надежность



RAS Feature	Power Systems	x86
Application/Partition RAS		
Live Partition Mobility (vMotion)	Yes	Yes
Live Application Mobility	Yes	No
Partition Availability priority	Yes	No
System RAS		
OS independent First Failure Data Capture	Yes	No
Memory Keys (including OS exploitation)	Yes	No
Processor RAS		
Processor Instruction Retry	Yes	No
Alternate Processor Recovery	Yes	No
Dynamic Processor Deallocation	Yes	No
Dynamic Processor Sparing	Yes	No
Memory RAS		
Chipkill™	Yes	Yes
Survive Double Memory Failures	Yes	No
Selective Memory Mirroring	Yes	No
Redundant Memory	Yes	Yes
I/O RAS		
Extended Error Handling	Yes	No
I/O Adapter Isolation (PCI-Bus and TCEs)	Yes	No

See the following URLs for addition details:

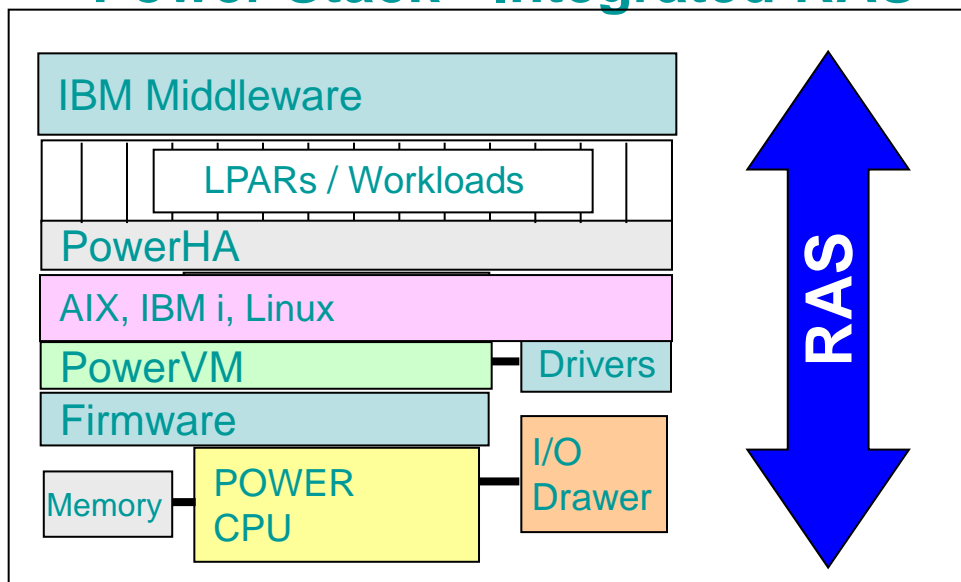
<http://www-03.ibm.com/systems/migratetoibm/systems/power/availability.html>

<http://www-03.ibm.com/systems/migratetoibm/systems/power/virtualization.html>

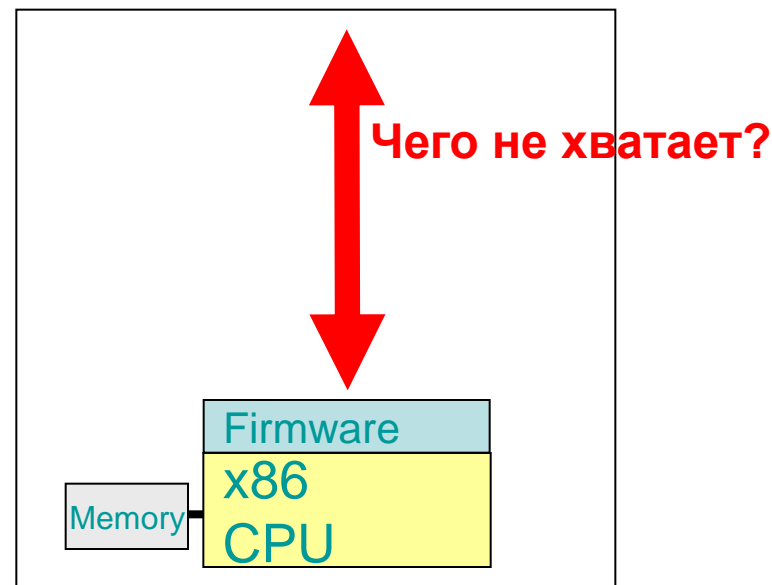
RAS не только на уровне процессора

- Разработка, тестирование и интеграция всего комплекса для достижения наивысших характеристик RAS
- Для того, чтобы RAS x86 приблизить к RAS Power нужно, как минимум использовать дорогостоящее кластерное ПО (например, Oracle RAC)

Power Stack - Integrated RAS



Intel Processor RAS



Безопасность
никто не хочет быть
целью атаки



TARGET®

Безопасность ограничена «слабым звеном»



Безопасность системы виртуализации критична для баз данных

0 зафиксированных взломов PowerVM hypervisor

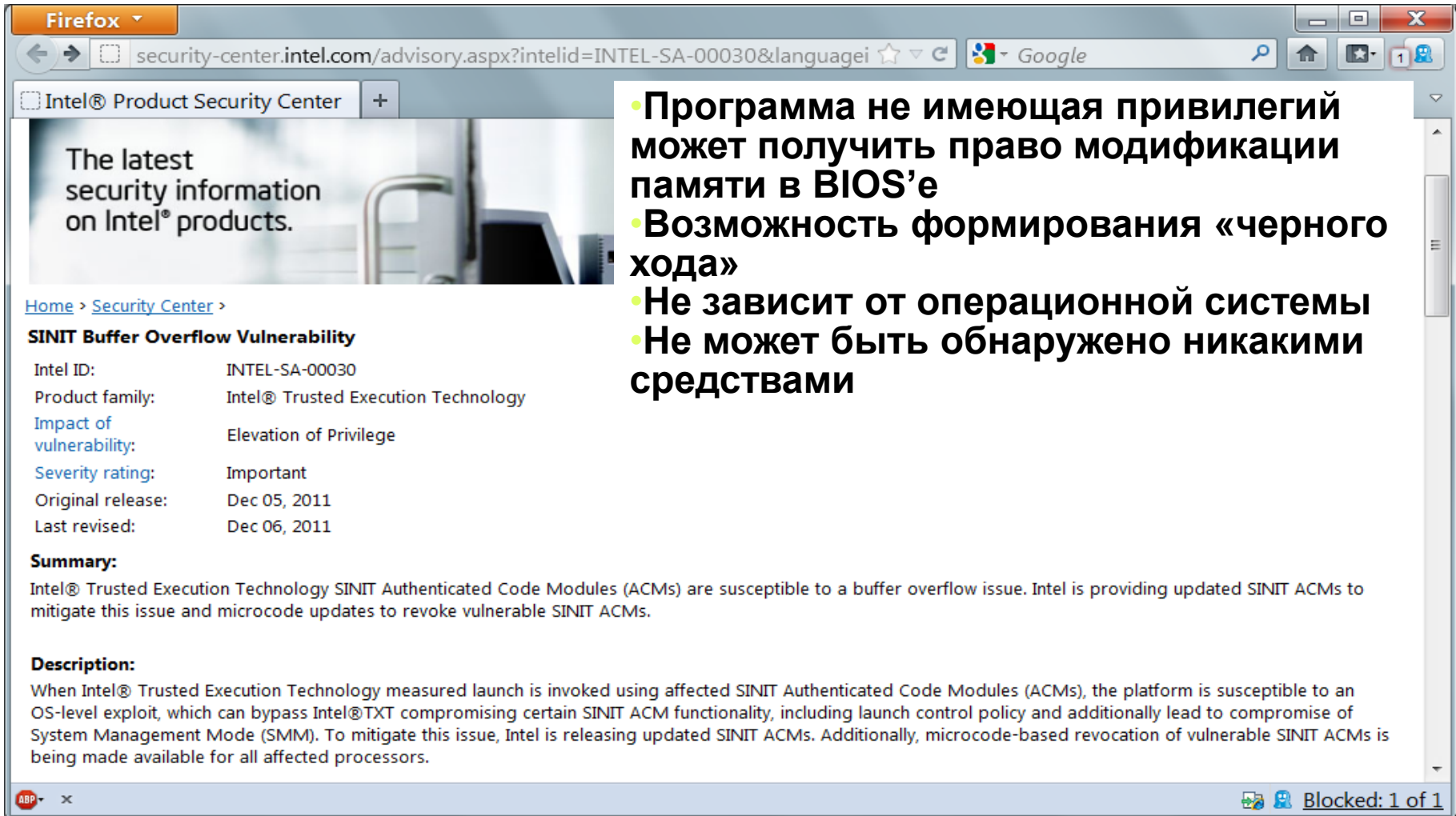
- The PowerVM hypervisor has never had a reported security vulnerability and provides the bullet-proof security that customers demand for mission-critical workloads
- The VIOS, which is part of the overall virtualization has had 0 reported security vulnerabilities
- **Dare to compare – search any security tracking DB and compare Power against x86**

Search term or Hypervisor (unfiltered)	NIST NVD Results	Processor Architecture
VMware	640	x86
Xen	153	x86
VMware ESX	95	x86
KVM	58	x86
VMware vSphere	48	x86
Windows Server 2012	43	x86
Oracle VM	24	x86
Hyper-V	3	x86
PowerVM	0	POWER

Source: National Vulnerability Database, <http://nvd.nist.gov/home.cfm>, July 2013.

NVD is the U.S. government repository of standards based vulnerability management data. This data enables automation of vulnerability management, security measurement, and compliance. NVD includes databases of security checklists, security related software flaws, misconfigurations, product names, and impact metrics. NVD is a product of the NIST [Computer Security Division](#), Information Technology Laboratory and is sponsored by the Department of Homeland Security's [National Cyber Security Division](#).

CVE is co-sponsored by the National Cyber Security Division of the U.S. Department of Homeland Security. Copyright © 1999–2012, The MITRE Corporation. CVE and the CVE logo are registered trademarks and CVE-Compatible is a trademark of The MITRE Corporation. This Web site is sponsored and managed by The MITRE Corporation to enable stakeholder collaboration.



The latest security information on Intel® products.

[Home](#) > [Security Center](#) >

SINIT Buffer Overflow Vulnerability

Intel ID: INTEL-SA-00030
Product family: Intel® Trusted Execution Technology
Impact of vulnerability: Elevation of Privilege
Severity rating: Important
Original release: Dec 05, 2011
Last revised: Dec 06, 2011

Summary:
Intel® Trusted Execution Technology SINIT Authenticated Code Modules (ACMs) are susceptible to a buffer overflow issue. Intel is providing updated SINIT ACMs to mitigate this issue and microcode updates to revoke vulnerable SINIT ACMs.

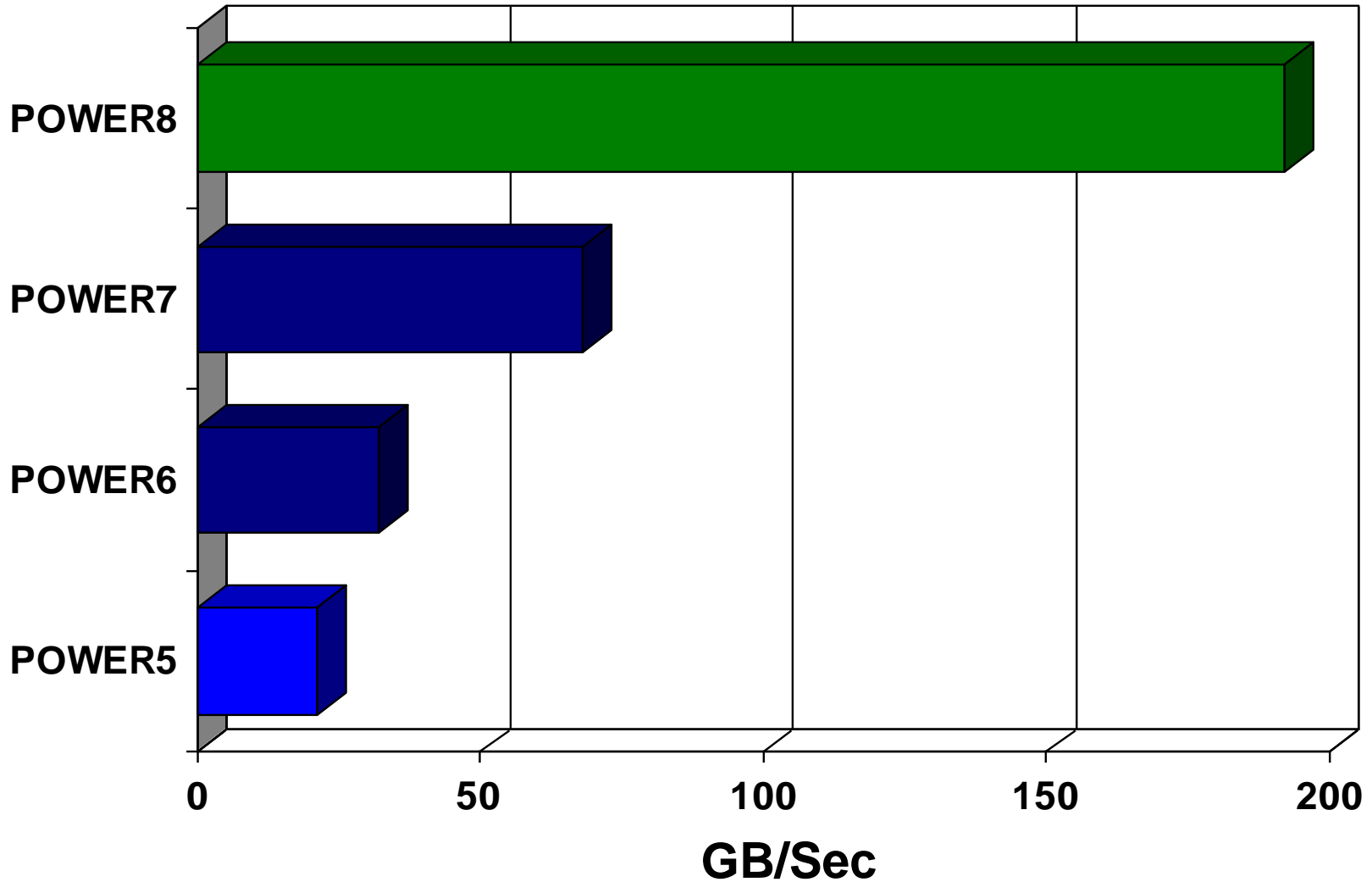
Description:
When Intel® Trusted Execution Technology measured launch is invoked using affected SINIT Authenticated Code Modules (ACMs), the platform is susceptible to an OS-level exploit, which can bypass Intel®TXT compromising certain SINIT ACM functionality, including launch control policy and additionally lead to compromise of System Management Mode (SMM). To mitigate this issue, Intel is releasing updated SINIT ACMs. Additionally, microcode-based revocation of vulnerable SINIT ACMs is being made available for all affected processors.

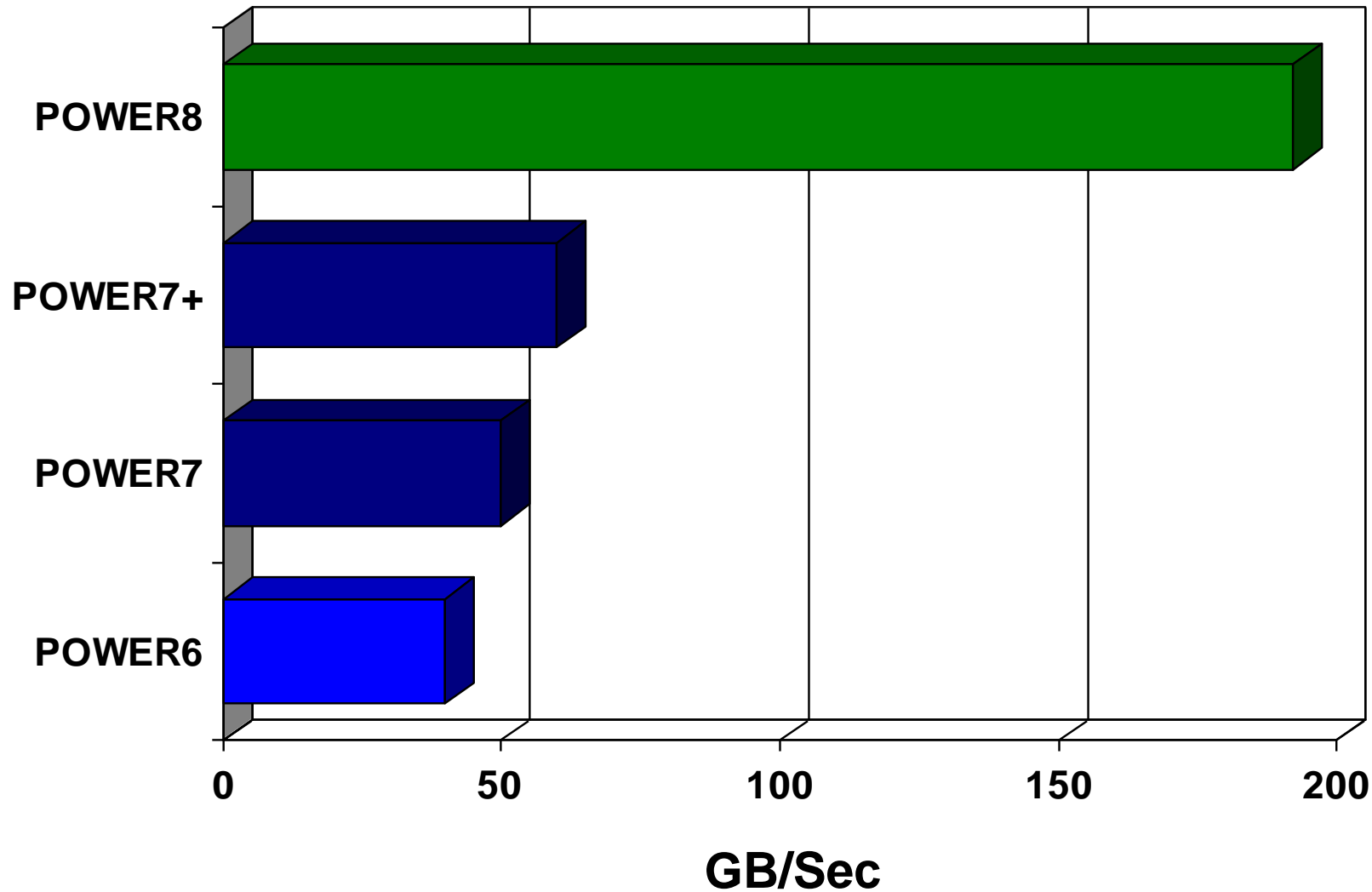
- Программа не имеющая привилегий может получить право модификации памяти в BIOS'е
- Возможность формирования «черного хода»
- Не зависит от операционной системы
- Не может быть обнаружено никакими средствами

Blocked: 1 of 1

Производительность

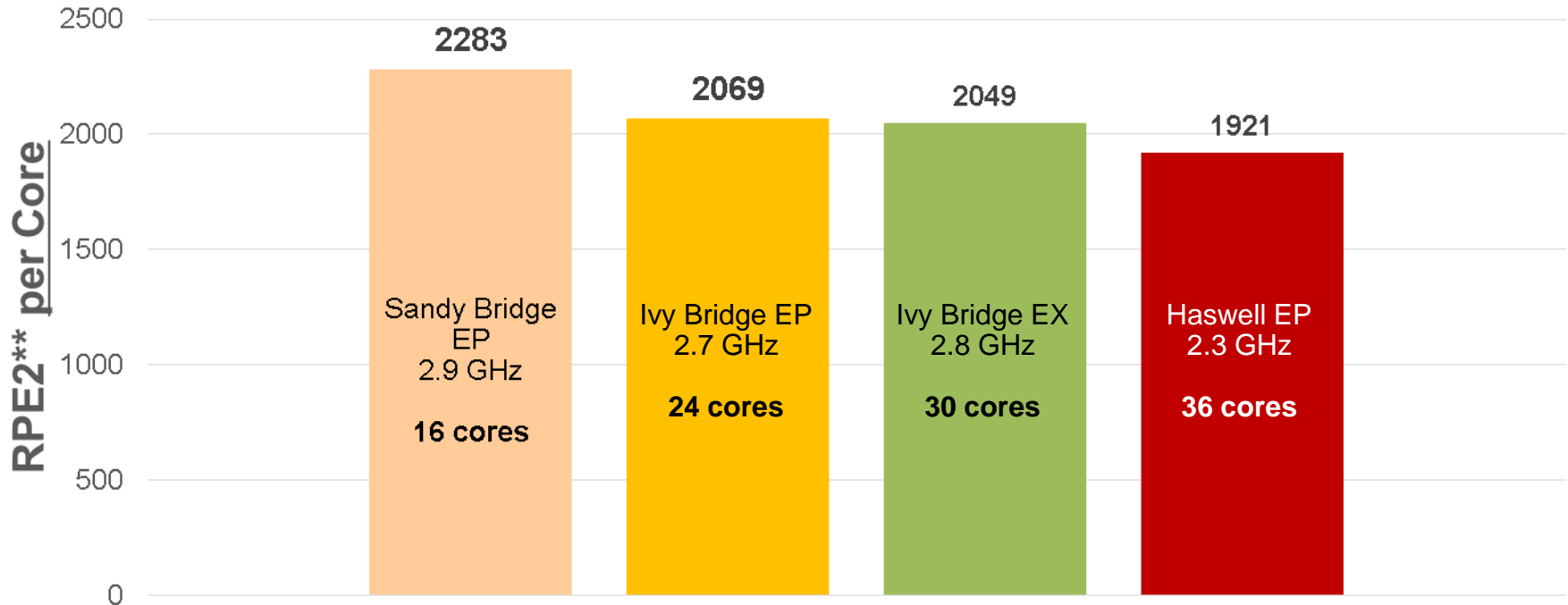






	Sandy Bridge EP E5-x6xx	Ivy Bridge EP E5-26xx v2	Ivy Bridge EX E7-88xx v2	Haswell EP E5-26xx v3	POWER8
Clock rates (GHz)	1.8–3.6	1.7-3.7	1.9-3.4	1.6-3.5	3.0-4.35 GHz
SMT options	1,2*	1, 2*	1, 2*	1, 2*	1, 2, 4, 8
Max Threads / sock	16	24	30	36	96
Max L1 Cache	32KB	32KB*	32KB*	64 KB	64KB
Max L2 Cache	256 KB	256 KB	256 KB	256KB	512 KB
Max L3 Cache	20 MB	30 MB	37.5 MB	45 MB	96 MB
Max L4 Cache	0	0	0	0	128 MB
Memory Bandwidth	31.4-51.2 GB/s	42.6-59.7 GB/s	68-85** GB/s	51-68 GB/s	192 GB/sec

Производительность ядер процессоров Intel не увеличивается, а уменьшается 2 Socket HP Servers



The number shown is best in each category (sockets and number of cores)

**Gartner RPE2 Details:
<http://www.gartner.com/technology/research/RPE2-methodology-details.jsp>

RPE2** numbers are derived from the following six benchmark inputs:

SAP SD Two-Tier, TPC-C, TPC-H, SPECjbb2006 and two SPEC CPU2006 components

Системы IBM POWER продолжают расти опережающими темпами и использовать системы более эффективно

- Infrastructure Software Price-performance has been REDUCED on Intel servers by up to 14%
 - Assumes flat system pricing
- Software Licensing has increased by 1.5x
 - 12 cores versus 8 cores
- Performance has increased only 1.29-1.40x (IBM increased by 1.71-2.31x)
 - x86 publishes on 2-socket Sandy Bridge and 2-socket Ivy Bridge

	x86 "Sandy Bridge"	x86 "Ivy Bridge"	System Performance Ratio	POWER7+	POWER8	System Performance Ratio
	2-socket E5-2690	2-socket E5-2697		2-socket POWER7+	2-socket POWER8	
# Cores	16	24	1.50	16	24	1.50
ERP SAP 2-Tier	7960	10240	1.29	10000	21212	2.12
SPECint_rate	693	967	1.40	884	1701	1.92
SPECfp_rate	510	701	1.37	602	1301	2.16
SPECjEnterprise2010	8310	11260	1.35	13161	22543	1.71

¹ Based on generational comparisons of SW that utilizes per core pricing and 50% more cores in per system (Power: 8c POWER7 to 12c POWER8; x86: 8c E5-2690 to 12c E5-2697 v2)

² Performance is based on published x86 data and published/projected POWER7+ & POWER8. Workloads are ERP, Integer, Floating Point, Java

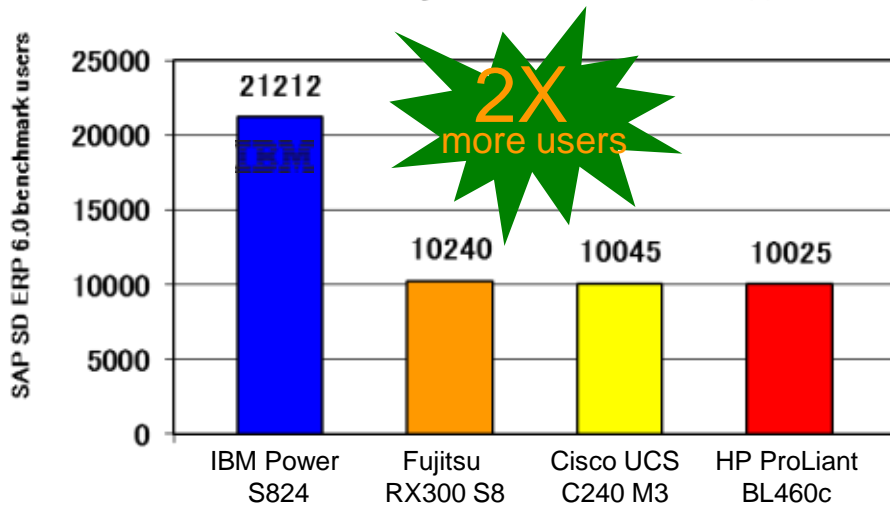
SAP Sales & Distribution 2-Tier ERP 6 Benchmarks

IBM Power System S824 using DB2 10.5 vs. Competition

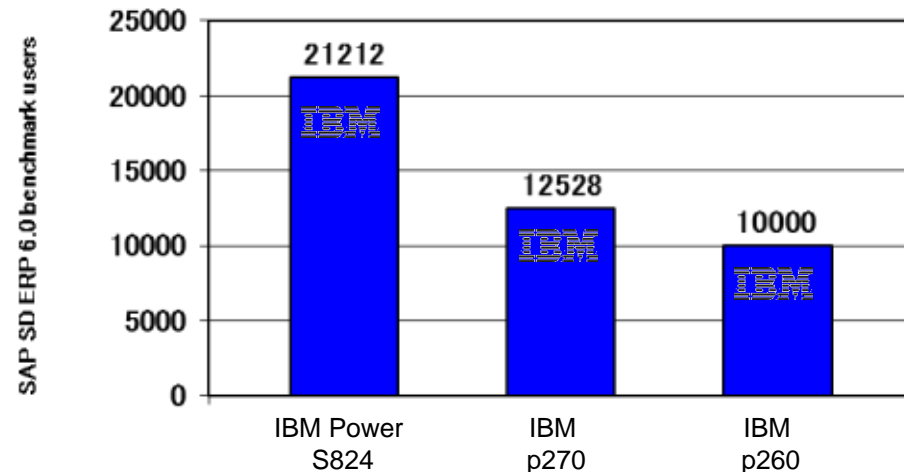
Более чем в 2 раза лучшая производительность на 24 ядрах по сравнению с лучшими показателями Intel

Прмерно в 2раза лучше чем предидущее поколение IBM Power

IBM Power System S824 Performance(1)



IBM Power System S824 Performance(2)
(per core)



(1.0) IBM Power System S824 on the two-tier SAP SD standard application benchmark running SAP enhancement package 5 for the SAP ERP 6.0 application; 4 processors / 24 cores / 192 threads, POWER8; 3.52GHz, 512 GB memory, 21,212 SD benchmark users, running AIX® 7.1 and DB2® 10.5, dialog response: 0.98 seconds, line items/hour: 2,317,330, dialog steps/hour: 6,952,000 SAPS: 115,870 database response time (dialog/update): 0.011 sec / 0.019sec, CPU utilization: 99%, Certification #: 2014016 Results valid as of 3/24/14. Source: <http://www.sap.com/benchmark>.

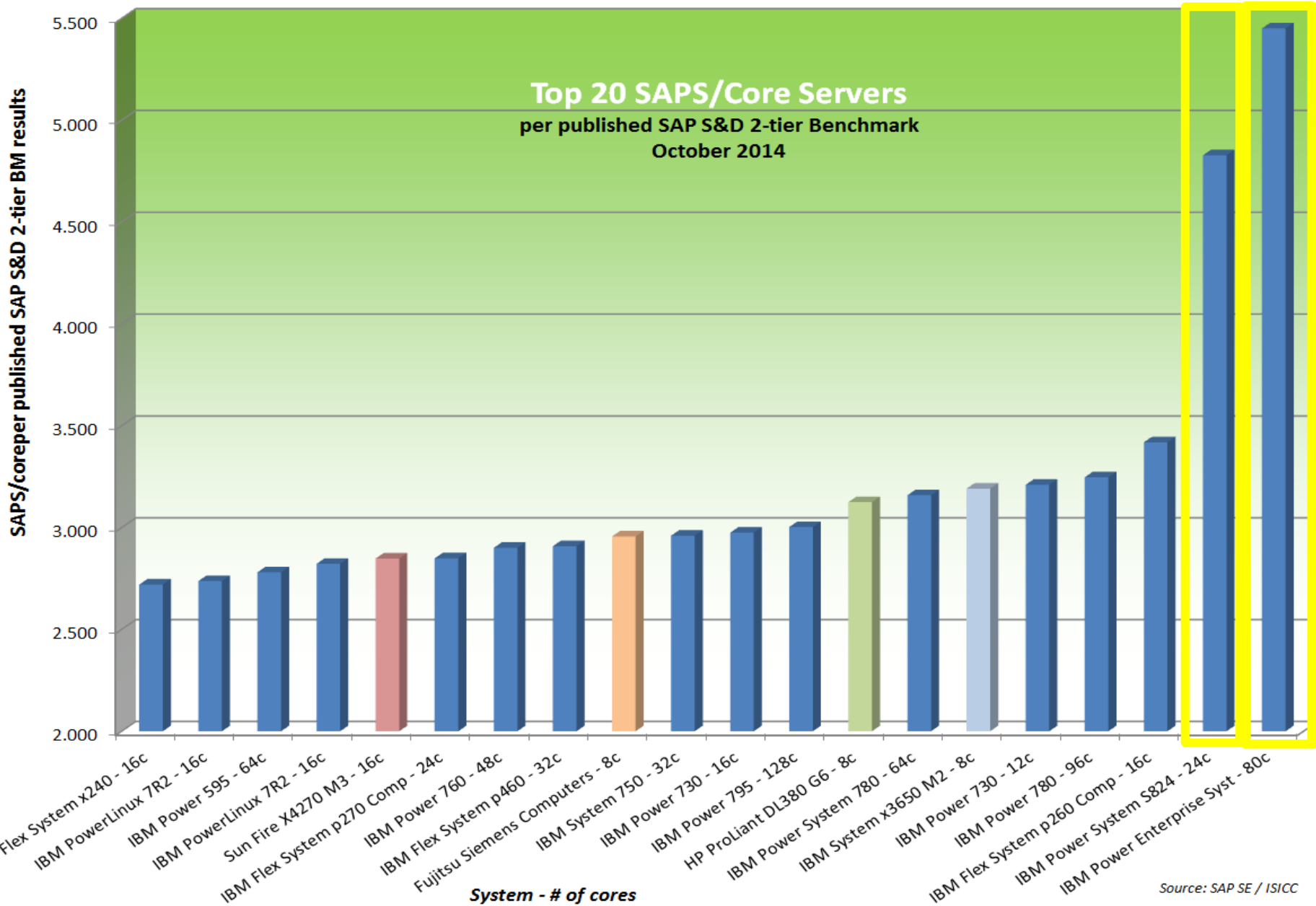
(1.1) Fujitsu RX300 S8 on the two-tier SAP SD standard application benchmark running SAP enhancement package 5 for the SAP ERP 6.0 application; 2 processors / 24 cores / 48 threads. Intel Xeon E5-2697 processor 2.70 GHz, 256 GB memory, 10,240 SD benchmark users, running Windows Server 2012 SE and SQL Server 2012, Certification #: 2013024

(1.2) Cisco UCS c240 M3 on the two-tier SAP SD standard application benchmark running SAP enhancement package 5 for the SAP ERP 6.0 application; 2 processors / 24c ores / 48 threads. Intel Xeon E5-2697 processor 2.70 GHz, 256 GB memory, 10,045 SD benchmark users, running Windows Server 2012 DE and SQL Server 2012, Certification #: 2013038

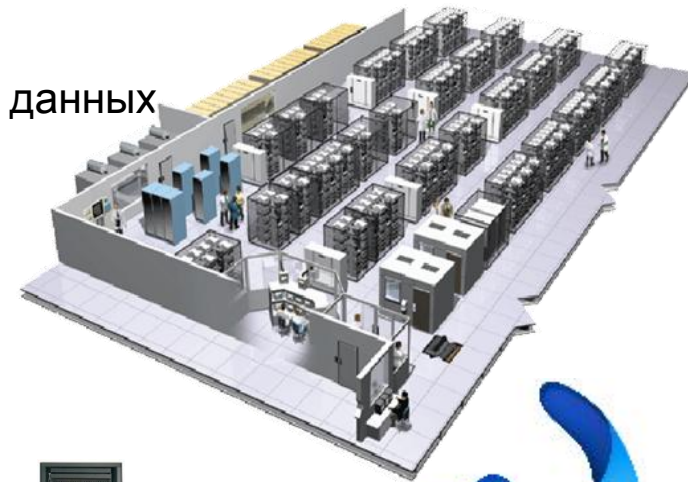
(1.3) HP ProLiant BL460c Gen8 on the two-tier SAP SD standard application benchmark running SAP enhancement package 5 for the SAP ERP 6.0 application; 2 processors / 24 cores / 48 threads. Intel Xeon E5-2697 processor 2.70 GHz, 256 GB memory, 10,025 SD benchmark users, running Windows Server 2012 DE and SQL Server 2012, Certification #: 2013025

(2.1) IBM Flex System p270 Compute Node on the two-tier SAP SD standard application benchmark running SAP enhancement package 5 for the SAP ERP 6.0 application; 4 processors / 24 cores / 96 threads, POWER7+; 3.4GHz, 256 GB memory, 12,528 SD benchmark users, running AIX® 7.1 and DB2® 10.5 Certification #: 3012019 Source: <http://www.sap.com/benchmark>.

(1.1) IBM Flex System p260 on the two-tier SAP SD standard application benchmark running SAP enhancement package 5 for the SAP ERP 6.0 application; 2 processors / 16 cores / 64 threads, POWER7+; 4.1GHz, 256 GB memory, 10,000 SD benchmark users, running AIX® 7.1 and DB2® 10, Certification #: 2012035



- Рост расходов и сложности (например, большее количество физических серверов, расширение сети)
- Чрезмерное энергопотребление и проблемы, связанные с перегревом
- Рост расходов на хранилища данных и синхронизацию данных
- Линейное увеличение расходов на труд
- Линейное увеличению расходов на программное обеспечение на процессор
- Частые простои



Выход:

использование меньшего количества более производительных POWER серверов в центрах обработки данных

*Industry
Best Practice*

Еще больше ядер

12 ядер в кристалле

- 1.5x больше, чем в Power7/7+
- лучшая производительность на ядро

Что в результате?

Почувствуйте производительность задач вертикального и горизонтального масштабирования

Industry Leading

Еще больше кэш-памяти

96МБ кэша

- 3X больше на кристалл, чем Power7
- плюс 128МБ off-chip cache – NEW!!

Что в результате?

Лучшая производительность для задач, требовательных к скорости памяти!

Industry Leading

Еще больше потоков

SMT8 – 8 динамических потоков на ядро. Динамически изменяемые режимы SMT1, 2, 4, 8 для каждой VM. Смешивание режимов в LPAR

Что в результате?

Лучшие режимы для каждой конкретной задачи.

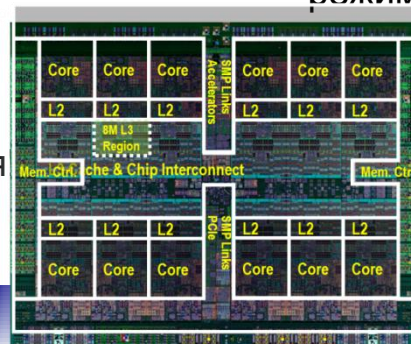
Industry Leading

Лучшая пропускная способность

В **3.5 раза** быстрее работа с памятью, в **2.8 раза** быстрее I/O.

Что в результате?

Работа больших данных и аналитики работают значительно эффективнее.



*Industry
Innovation*

CAPI

Технология Open interface позволяет устройствам работать со скоростью памяти.

Что в результате?

Новые возможности интеграции интеллектуальных устройств.

*Innovation
On Power*

Встроенный PCIe

Контроллер PCIe Gen3, встроенный в кристалл.

Что в результате?

Приложения, требовательные к вводу выводу, будут работать значительно быстрее и с меньшими задержками

*Innovation
Extended*

Transactional Memory

Технология класса Mainframe. Значительное ускорение записи и исключение конкуренции работы с памятью.

Что в результате?

Возможности «мэйнфреймов» для ускорения OLTP задач теперь доступны на Power.

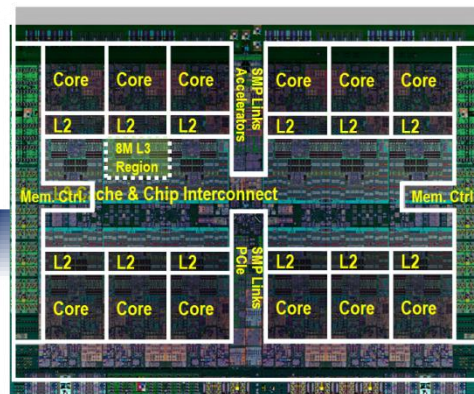
*Innovation
On Power*

PowerKVM

Встроенная поддержка KVM - open-source решение для виртуализации для Linux систем.

Что в результате?

Единый стандартизованный подход для построения облачных систем с единым центром управления.





IBM POWER
МОДЕЛЬНЫЙ ряд

Power
Systems

1Q 2015 Portfolio: POWER7/POWER7+/POWER8

**POWER8
S822**

**POWER8
S814/S824**

**POWER8
E870**

**POWER8
E880**

**Power
780+**

**Power
795**

POWER8 Scale-out Servers

POWER8 Scale-up Servers

**Power
750+ / 760**

**Power
710+/730+**

**Power
720+/740+**

PowerLinux

**IBM PureFlex
System**

P460+

p260+

POWER8 Scale-out Servers

**POWER8
S824L**

**POWER8
S812L/S822L**

OpenPOWER – открытое сообщество разработчиков



101 член консорциума и 8 рабочих групп

Implementation / HPC / Research

22 страны
25 университетов

System / Software / Integration

I/O / Storage / Acceleration

Boards / Systems

Chip / SOC



Power Systems



- ✓ Разработаны для Big Data
- ✓ Экономически выгодны для облаков
- ✓ Открытая инновационная платформа



Виртуализация без ограничений



PowerVC



PowerSC



PowerHA



PowerVM

PowerVP



Thanks! Спасибо!

Вопросы?

Alexey.Boyko@ru.ibm.com

+7(495)-775-8800 ext 2904