IBM Technical Computing

Technology Overview and Outlook

CernVM Workshop

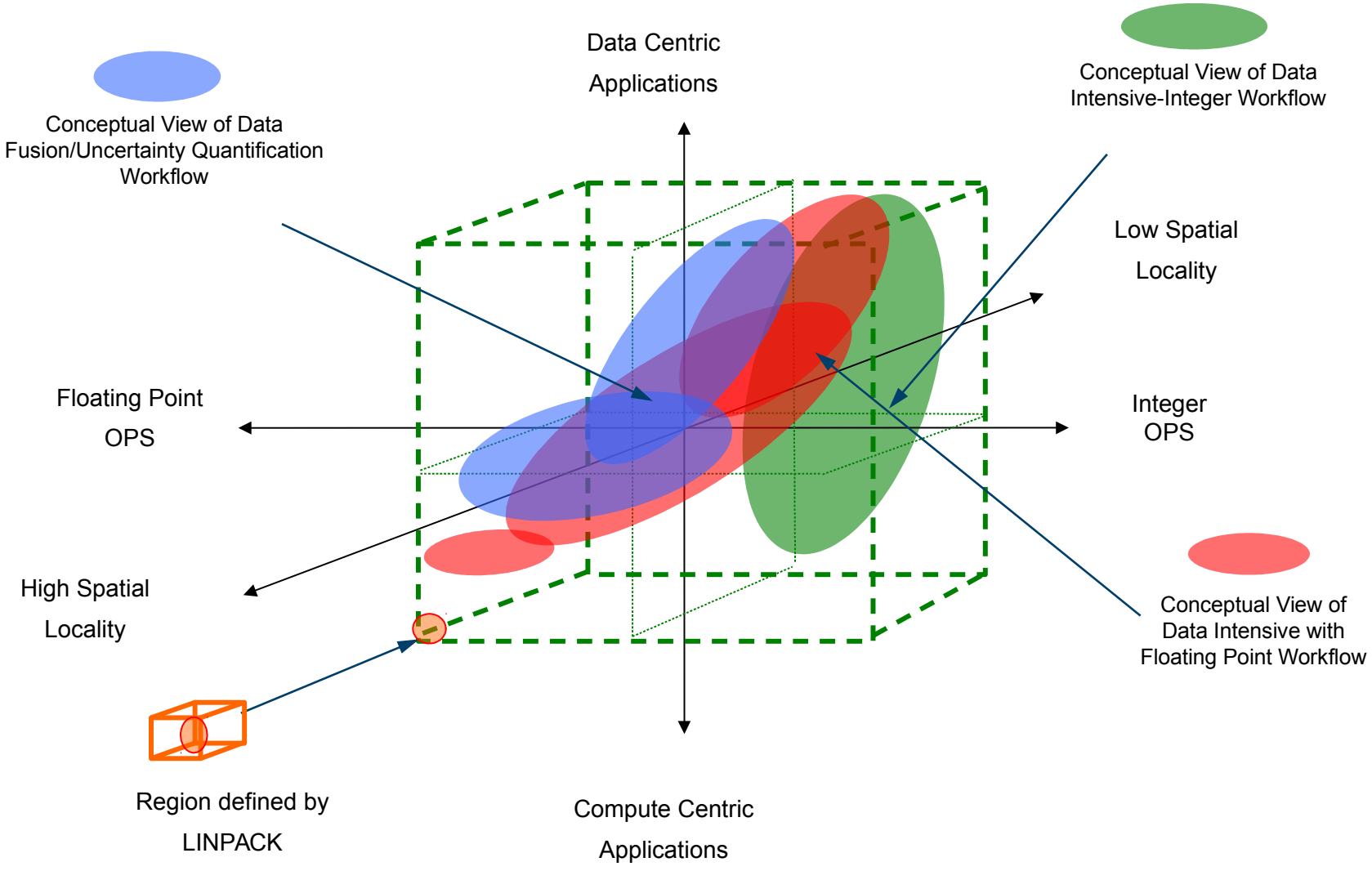
Dr. Oliver Oberst
07 June 2016

Virtualization in Technical Computing

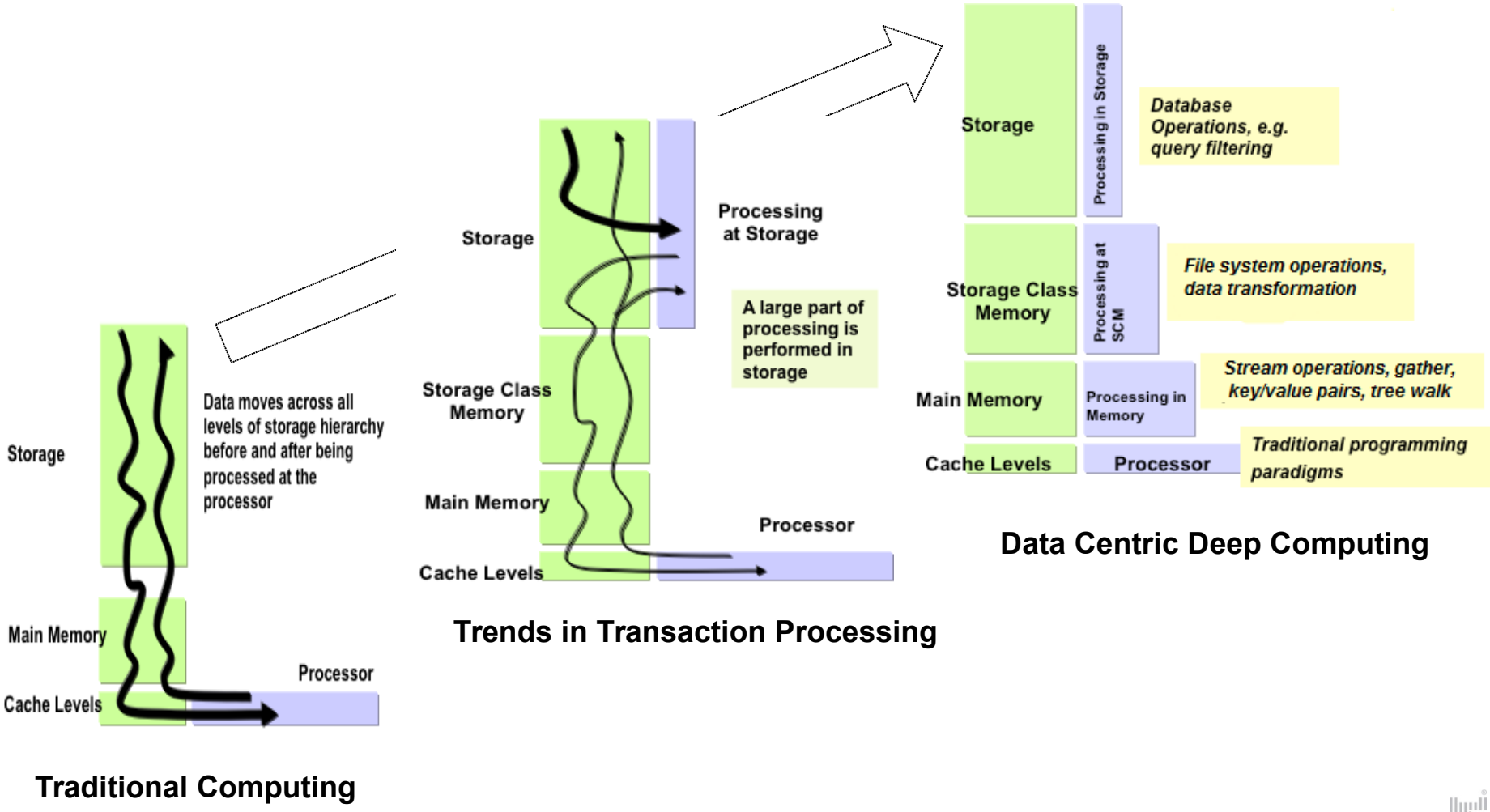
- ETP4HPC Strategic Research Agenda 2015 Update
  - ***"Virtualisation, data security at hardware and system level becomes a critical challenge for exascale infrastructure..."***
  - ***"...virtualisation is making its way into the HPC system design and is essential for a more flexible usage of HPC systems"***
  - ***"The improved flexibility will also facilitate access to HPC as a cloud resource, enabling new business and usage models through agile, on-demand infrastructures."***

Data Centric

Different Solutions for Different Parts of the Cube



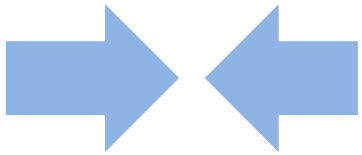
Optimized System Design for Data Centric Computing



Big Data Driving Common Requirements

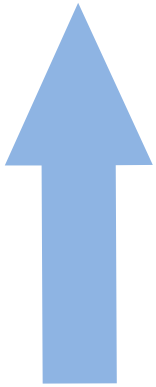
High Performance Analytics
- Unstructured data
- Primarily data mining

High Performance Computing
- Structured data
- Primarily scientific calculations/Simulation



Evolving requirements

Driver: Enhanced context Improves decision making
- Incorporate modeling and simulation for better predictions
- Incorporate sensor data

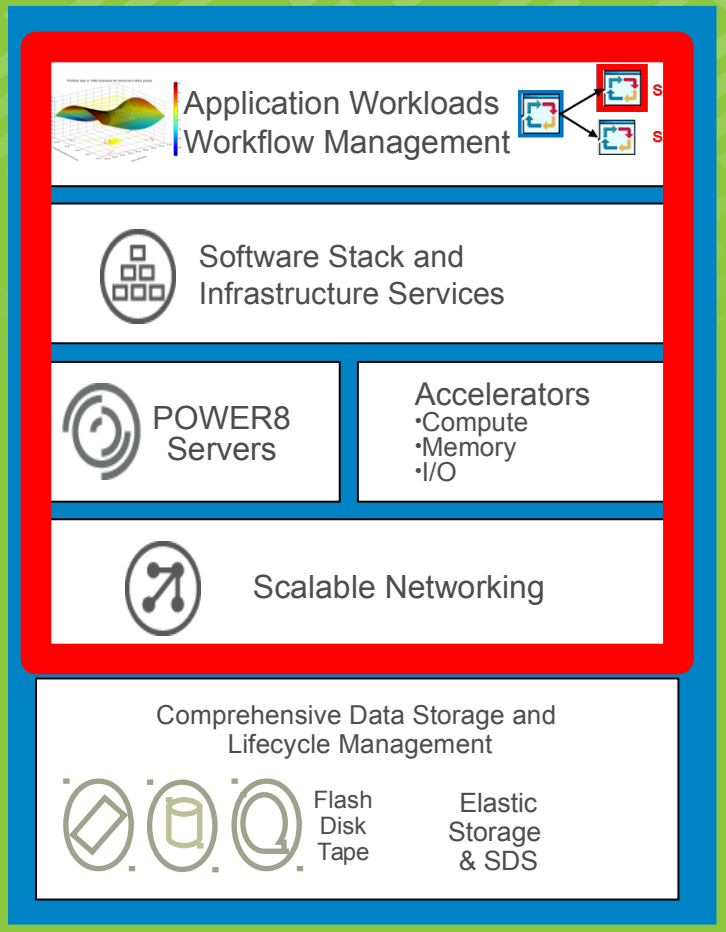


Driver: Doing more with models
- Real-time decision making
- Uncertainty quantification
- Sensitivity analysis
- Metadata extraction

Data Centric Systems

Compute

OpenPower



OpenPOWER: Open Architecture for HPC & Big Data

Processor IP Licensing	Licensing IP to enable semiconductor partners like Suzhou Powercore to build POWER chips
Open Interfaces	Tight integration using CAPI & NVLink with Accelerators (NVIDIA, Xilinx, Altera), Networking (Mellanox), Storage (CAPI Flash)
Systems & Software	Enabling Innovative POWER-based servers from Partners & OpenCompute and Sharing Open Source Software including Firmware & Hypervisor

OpenPOWER Key Strategies & Market Segments

Cloud Computing
Hyper-Scale Data Centers

Drive POWER into Domestic IT Agendas

Technical Computing
(HPC, Big Data, & Machine Learning)

US & UK Research Establishments Select OpenPOWER-Based Supercomputers

IBM, Mellanox, and NVIDIA awarded $325M U.S. Department of Energy's CORAL Supercomputers

CORAL: Leadership Class Supercomputers

5X – 10X HIGHER APP PERF THAN CURRENT SYSTEMS

OAK RIDGE National Laboratory

Lawrence Livermore National Laboratory

IBM & UK's STFC Partner for Big Data & Cognitive Computing Research

Science & Technology Facilities Council

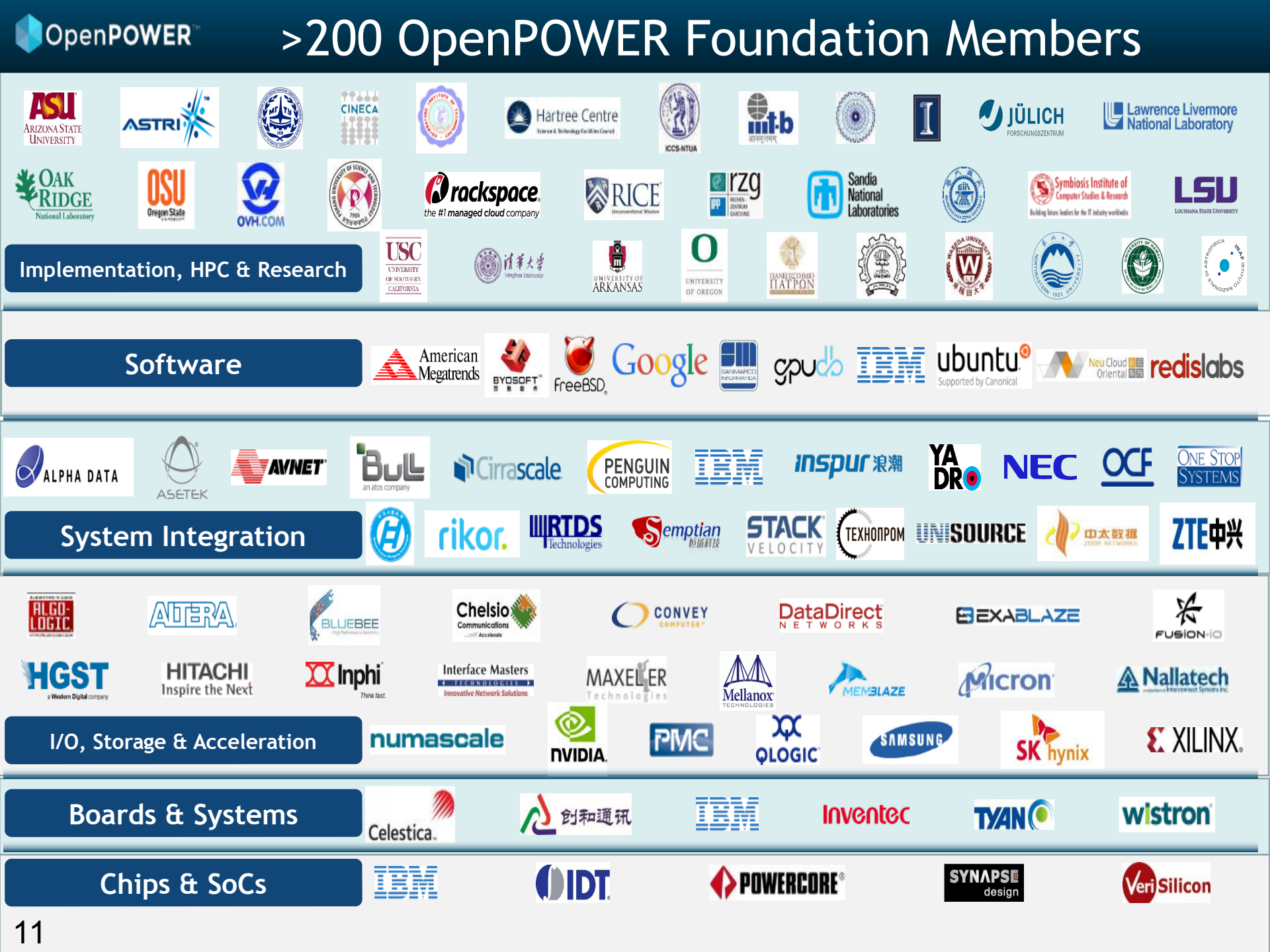
HM Government

OpenPOWER™

IBM Research

IBM Watson

>200 OpenPOWER Foundation Members



1600+ Applications on POWER

HPC

CHARMM
GROMACS
NAMD
AMBER
RTM
GAMESS
WRF
HYCOM
HOMME
LES
MiniGhost
AMG2013
OpenFOAM
miniDFT
CTH
BLAST
Bowtie
BWA
FASTA
HMMER
GATK
SOAP3
STAC-A2
SHOC
Graph500
Ilog

Cloud

openstack
TOMCAT
Apache
Chef
STORIX
REFLEXIS
DIASOFT
OneView Commerce
POWERSENSE
IFS
KANA.
Fis
YUCHENG Technologies Limited
sky solutions
betasystems
EXIGEN
ProjectWare
ALPHINAT

Big Data Analytics Machine Learning

hadoop
PostgreSQL
MariaDB
redislabs
SUGARCRM
gpudb
Spark
Datameer
neo4j
EDB ENTERPRISEDB
AsiaINFO
InterSystems
CROSSVIEW.
POLARIS
ZAO
Valogix

Mobile Enterprise

zend
亿阳信通 BOCO Inter-Telecom
php
CSC
RabbitMQ Open Source Enterprise Messaging
Kingdee 金蝶
SYBASE An SAP Company
Fis
NARI 南瑞集团公司 NARI GROUP CORPORATION
FINEOS
ALPHINAT
comptel
Corent
TEMENOS
KUTIR

Major Linux Distros

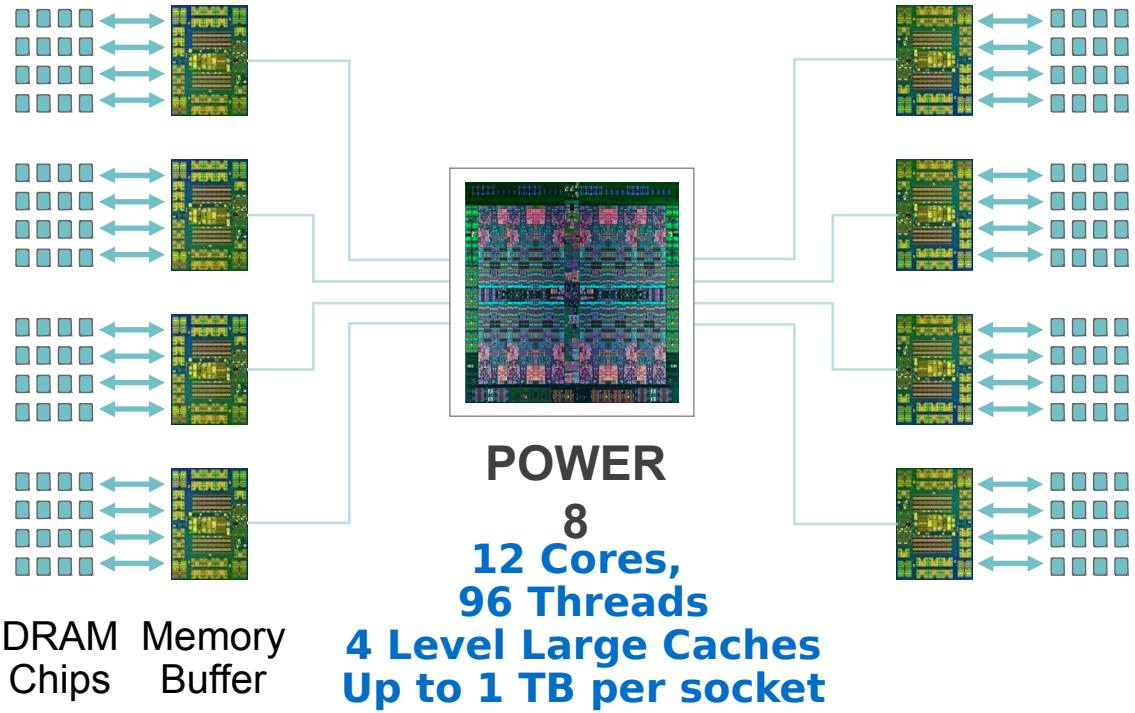
ubuntu
suse
redhat
Linux
Power
PowerVM

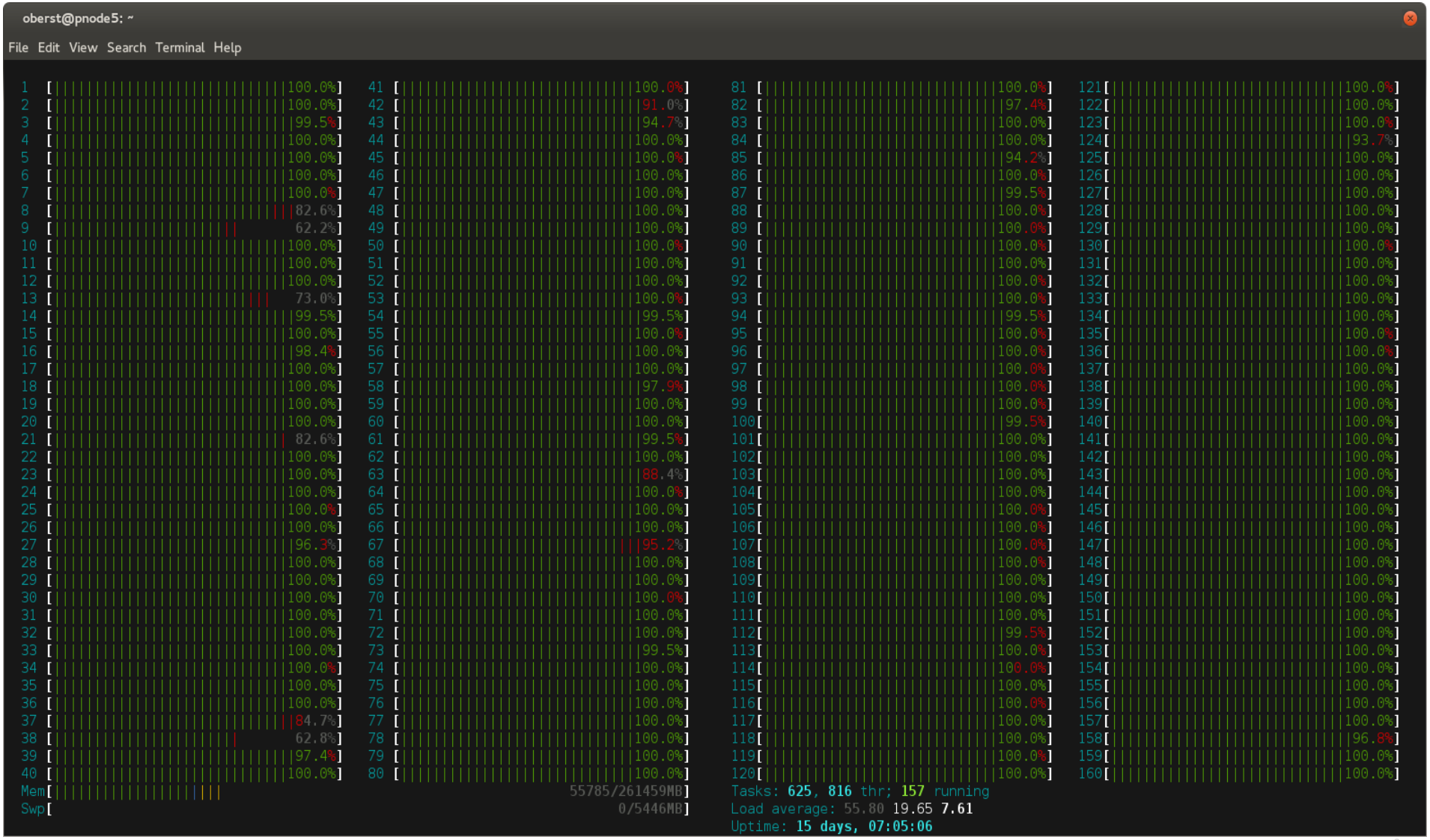
POWER8: Processor Performance Leadership

Faster Cores
8 Threads Per Core

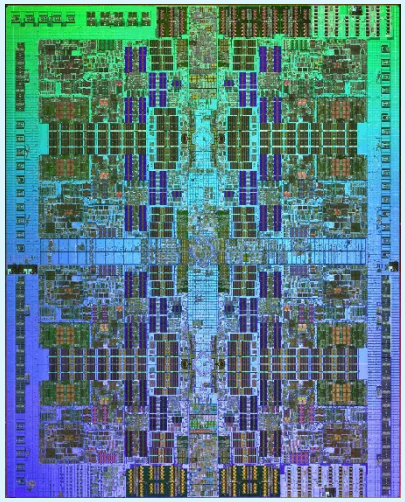
Larger Caches
Direct Accelerator Interconnect

3x Higher Memory Bandwidth, 1 TB Memory per Socket



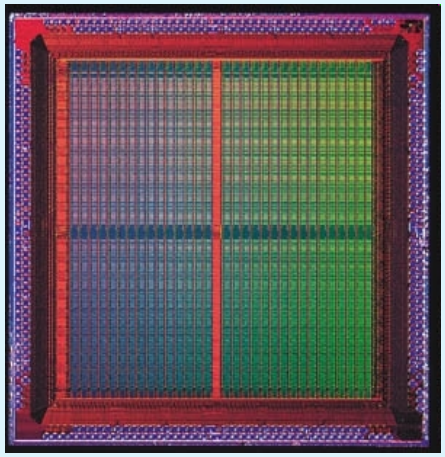

oberst@pnode5: ~
File Edit View Search Terminal Help
1 [||||||||||||||||||||||||100.0%] 41 [||||||||||||||||||||||||100.0%] 81 [||||||||||||||||||||||||100.0%] 121[||||||||||||||||||||||||100.0%]
2 [||||||||||||||||||||||||100.0%] 42 [||||||||||||||||||||||||91.0%] 82 [||||||||||||||||||||||||97.4%] 122[||||||||||||||||||||||||100.0%]
3 [||||||||||||||||||||||||99.5%] 43 [||||||||||||||||||||||||94.7%] 83 [||||||||||||||||||||||||100.0%] 123[||||||||||||||||||||||||100.0%]
4 [||||||||||||||||||||||||100.0%] 44 [||||||||||||||||||||||||100.0%] 84 [||||||||||||||||||||||||100.0%] 124[||||||||||||||||||||||||93.7%]
5 [||||||||||||||||||||||||100.0%] 45 [||||||||||||||||||||||||100.0%] 85 [||||||||||||||||||||||||94.2%] 125[||||||||||||||||||||||||100.0%]
6 [||||||||||||||||||||||||100.0%] 46 [||||||||||||||||||||||||100.0%] 86 [||||||||||||||||||||||||100.0%] 126[||||||||||||||||||||||||100.0%]
7 [||||||||||||||||||||||||100.0%] 47 [||||||||||||||||||||||||100.0%] 87 [||||||||||||||||||||||||99.5%] 127[||||||||||||||||||||||||100.0%]
8 [||||||||||||||||||||||||||82.6%] 48 [||||||||||||||||||||||||100.0%] 88 [||||||||||||||||||||||||100.0%] 128[||||||||||||||||||||||||100.0%]
9 [||||||||||||||||||||||||62.2%] 49 [||||||||||||||||||||||||100.0%] 89 [||||||||||||||||||||||||100.0%] 129[||||||||||||||||||||||||100.0%]
10 [||||||||||||||||||||||||100.0%] 50 [||||||||||||||||||||||||100.0%] 90 [||||||||||||||||||||||||100.0%] 130[||||||||||||||||||||||||100.0%]
11 [||||||||||||||||||||||||100.0%] 51 [||||||||||||||||||||||||100.0%] 91 [||||||||||||||||||||||||100.0%] 131[||||||||||||||||||||||||100.0%]
12 [||||||||||||||||||||||||100.0%] 52 [||||||||||||||||||||||||100.0%] 92 [||||||||||||||||||||||||100.0%] 132[||||||||||||||||||||||||100.0%]
13 [||||||||||||||||||||||||73.0%] 53 [||||||||||||||||||||||||100.0%] 93 [||||||||||||||||||||||||100.0%] 133[||||||||||||||||||||||||100.0%]
14 [||||||||||||||||||||||||99.5%] 54 [||||||||||||||||||||||||99.5%] 94 [||||||||||||||||||||||||99.5%] 134[||||||||||||||||||||||||100.0%]
15 [||||||||||||||||||||||||100.0%] 55 [||||||||||||||||||||||||100.0%] 95 [||||||||||||||||||||||||100.0%] 135[||||||||||||||||||||||||100.0%]
16 [||||||||||||||||||||||||98.4%] 56 [||||||||||||||||||||||||100.0%] 96 [||||||||||||||||||||||||100.0%] 136[||||||||||||||||||||||||100.0%]
17 [||||||||||||||||||||||||100.0%] 57 [||||||||||||||||||||||||100.0%] 97 [||||||||||||||||||||||||100.0%] 137[||||||||||||||||||||||||100.0%]
18 [||||||||||||||||||||||||100.0%] 58 [||||||||||||||||||||||||97.9%] 98 [||||||||||||||||||||||||100.0%] 138[||||||||||||||||||||||||100.0%]
19 [||||||||||||||||||||||||100.0%] 59 [||||||||||||||||||||||||100.0%] 99 [||||||||||||||||||||||||100.0%] 139[||||||||||||||||||||||||100.0%]
20 [||||||||||||||||||||||||100.0%] 60 [||||||||||||||||||||||||100.0%] 100[||||||||||||||||||||||||99.5%] 140[||||||||||||||||||||||||100.0%]
21 [||||||||||||||||||||||| 82.6%] 61 [||||||||||||||||||||||||99.5%] 101[||||||||||||||||||||||||100.0%] 141[||||||||||||||||||||||||100.0%]
22 [||||||||||||||||||||||||100.0%] 62 [||||||||||||||||||||||||100.0%] 102[||||||||||||||||||||||||100.0%] 142[||||||||||||||||||||||||100.0%]
23 [||||||||||||||||||||||||100.0%] 63 [||||||||||||||||||||||||88.4%] 103[||||||||||||||||||||||||100.0%] 143[||||||||||||||||||||||||100.0%]
24 [||||||||||||||||||||||||100.0%] 64 [||||||||||||||||||||||||100.0%] 104[||||||||||||||||||||||||100.0%] 144[||||||||||||||||||||||||100.0%]
25 [||||||||||||||||||||||||100.0%] 65 [||||||||||||||||||||||||100.0%] 105[||||||||||||||||||||||||100.0%] 145[||||||||||||||||||||||||100.0%]
26 [||||||||||||||||||||||||100.0%] 66 [||||||||||||||||||||||||100.0%] 106[||||||||||||||||||||||||100.0%] 146[||||||||||||||||||||||||100.0%]
27 [||||||||||||||||||||||||96.3%] 67 [||||||||||||||||||||||||95.2%] 107[||||||||||||||||||||||||100.0%] 147[||||||||||||||||||||||||100.0%]
28 [||||||||||||||||||||||||100.0%] 68 [||||||||||||||||||||||||100.0%] 108[||||||||||||||||||||||||100.0%] 148[||||||||||||||||||||||||100.0%]
29 [||||||||||||||||||||||||100.0%] 69 [||||||||||||||||||||||||100.0%] 109[||||||||||||||||||||||||100.0%] 149[||||||||||||||||||||||||100.0%]
30 [||||||||||||||||||||||||100.0%] 70 [||||||||||||||||||||||||100.0%] 110[||||||||||||||||||||||||100.0%] 150[||||||||||||||||||||||||100.0%]
31 [||||||||||||||||||||||||100.0%] 71 [||||||||||||||||||||||||100.0%] 111[||||||||||||||||||||||||100.0%] 151[||||||||||||||||||||||||100.0%]
32 [||||||||||||||||||||||||100.0%] 72 [||||||||||||||||||||||||100.0%] 112[||||||||||||||||||||||||99.5%] 152[||||||||||||||||||||||||100.0%]
33 [||||||||||||||||||||||||100.0%] 73 [||||||||||||||||||||||||99.5%] 113[||||||||||||||||||||||||100.0%] 153[||||||||||||||||||||||||100.0%]
34 [||||||||||||||||||||||||100.0%] 74 [||||||||||||||||||||||||100.0%] 114[||||||||||||||||||||||||100.0%] 154[||||||||||||||||||||||||100.0%]
35 [||||||||||||||||||||||||100.0%] 75 [||||||||||||||||||||||||100.0%] 115[||||||||||||||||||||||||100.0%] 155[||||||||||||||||||||||||100.0%]
36 [||||||||||||||||||||||||100.0%] 76 [||||||||||||||||||||||||100.0%] 116[||||||||||||||||||||||||100.0%] 156[||||||||||||||||||||||||100.0%]
37 [||||||||||||||||||||||||84.7%] 77 [||||||||||||||||||||||||100.0%] 117[||||||||||||||||||||||||100.0%] 157[||||||||||||||||||||||||100.0%]
38 [||||||||||||||||||||||62.8%] 78 [||||||||||||||||||||||||100.0%] 118[||||||||||||||||||||||||100.0%] 158[||||||||||||||||||||||||96.8%]
39 [||||||||||||||||||||||||97.4%] 79 [||||||||||||||||||||||||100.0%] 119[||||||||||||||||||||||||100.0%] 159[||||||||||||||||||||||||100.0%]
40 [||||||||||||||||||||||||100.0%] 80 [||||||||||||||||||||||||100.0%] 120[||||||||||||||||||||||||100.0%] 160[||||||||||||||||||||||||100.0%]
Mem[|||||||||||||||||| 55785/261459MB]
Swp[0/5446MB]
Tasks: 625, 816 thr; 157 running
Load average: 55.80 19.65 7.61
Uptime: 15 days, 07:05:06

Accelerated Technical Computing - FPGAs



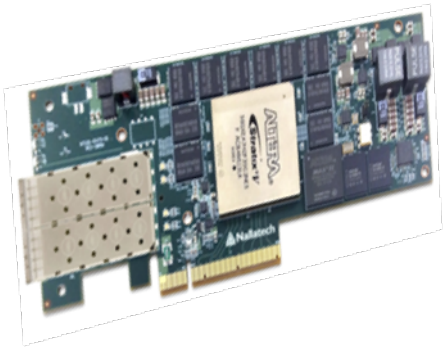
CPU's

+



Reconfigurable FPGA

Nallatech



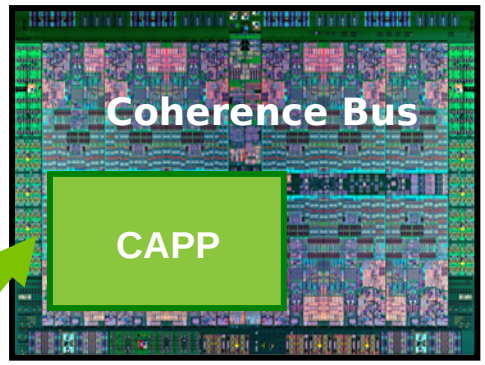
- Example workloads:
  - Monte Carlo
  - Data compression
  - Streaming compression/decompression

OpenPOWER CAPI Developer Kit for POWER 8

CAPI – Coherent Accelerator Processor Interface

- Virtual Addressing
  - Accelerator can work with same memory addresses that the processors use
- Hardware Managed Cache Coherence
  - Enables the accelerator to participate in “Locks” as a normal thread Lowers Latency over IO communication model

POWER8

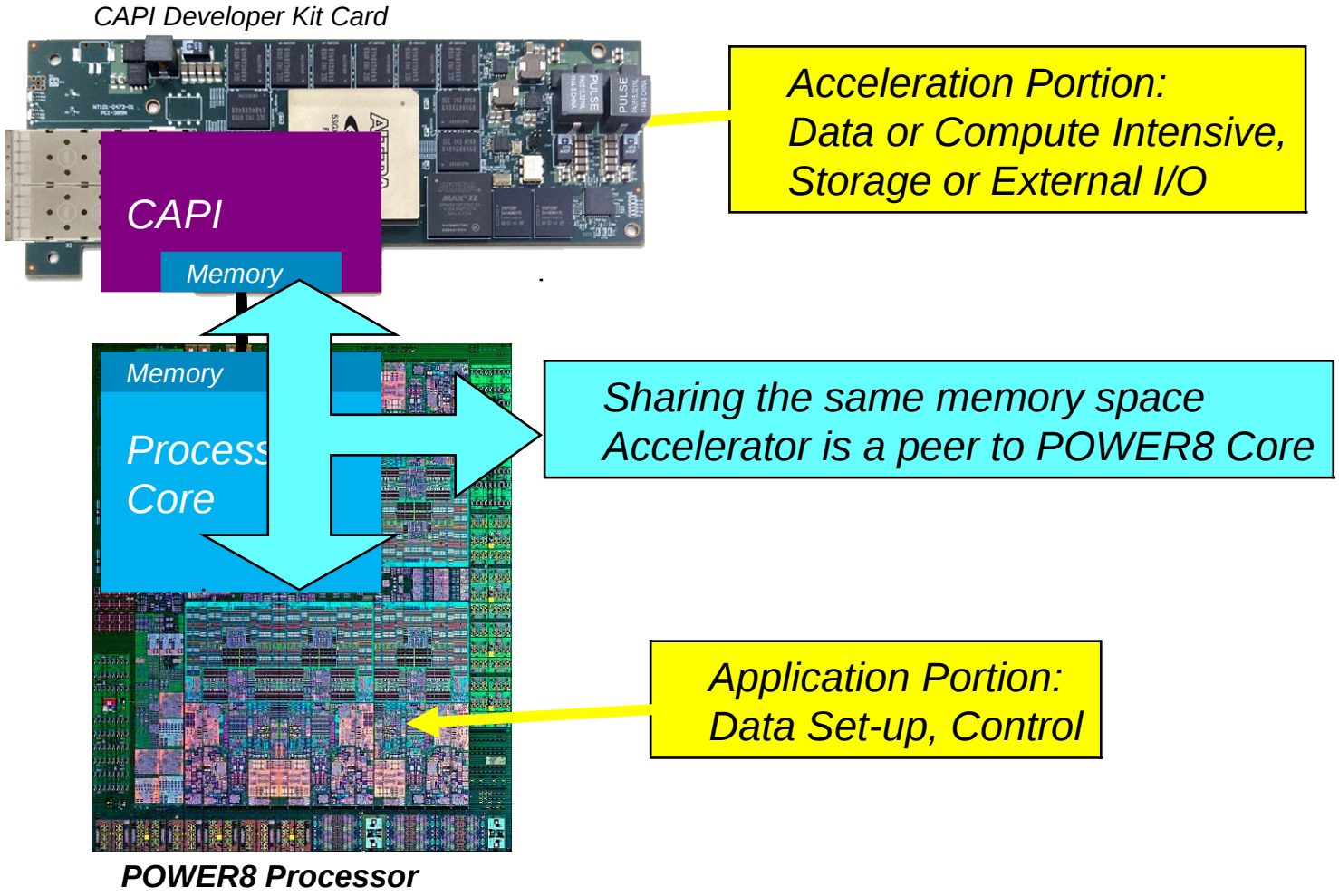


PCIe Gen 3
Transport for encapsulated messages

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

How CAPI Works



IBM Accelerated GZIP Compression

What it is:

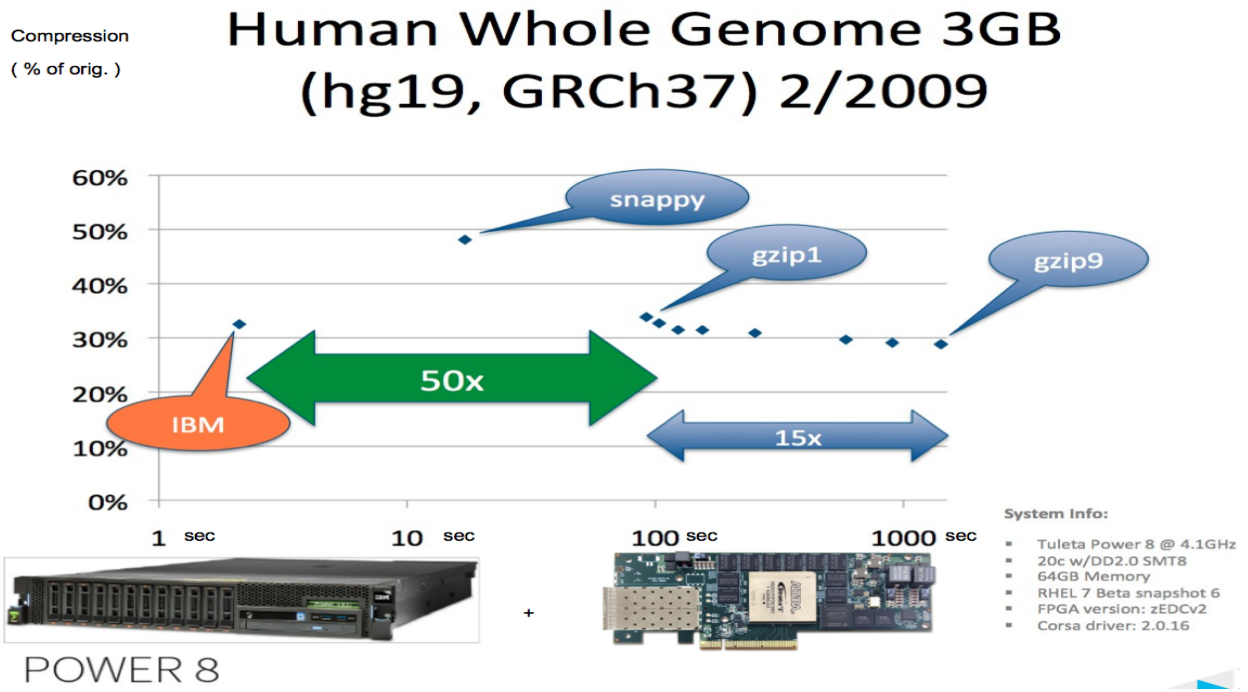
- *An FPGA-based low-latency GZIP Compressor & Decompressor with.*

Results:

- ***Single-thread** througput of ~2GB/s and a compression rate significantly better than low-CPU overhead compressors like snappy*

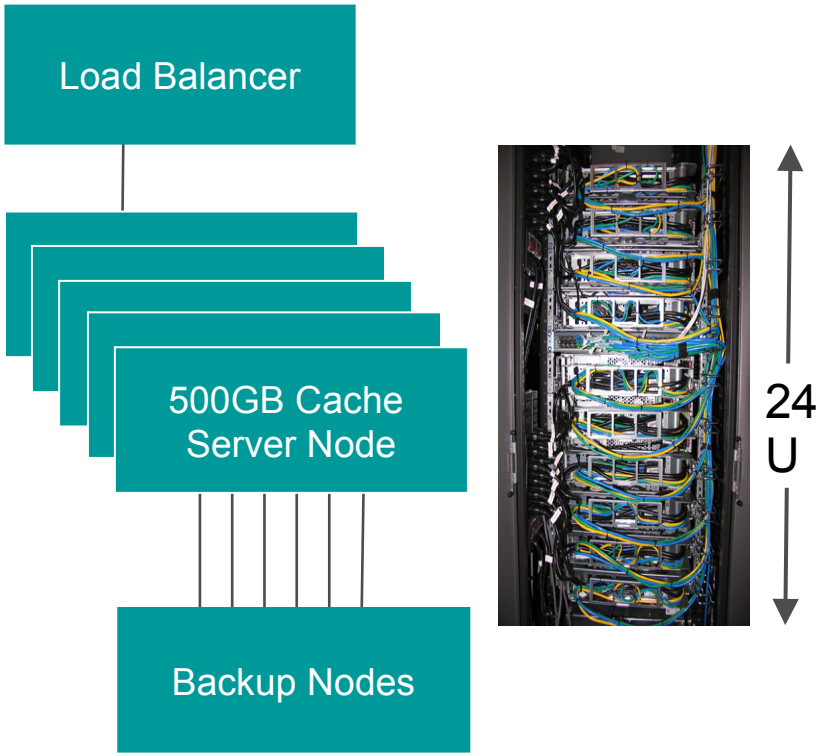
Source:

- *Non-published results*

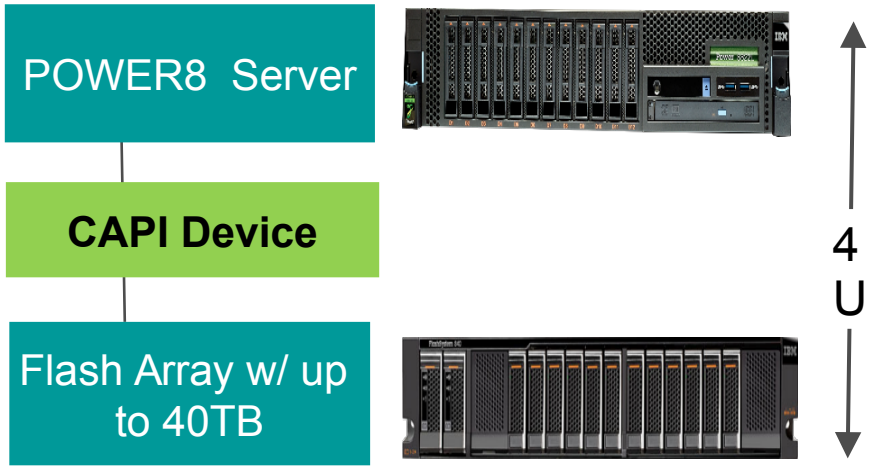


3x Lower Cost for NoSQL Databases using CAPI-Attached Flash

Before: NoSQL In-Memory (x86)



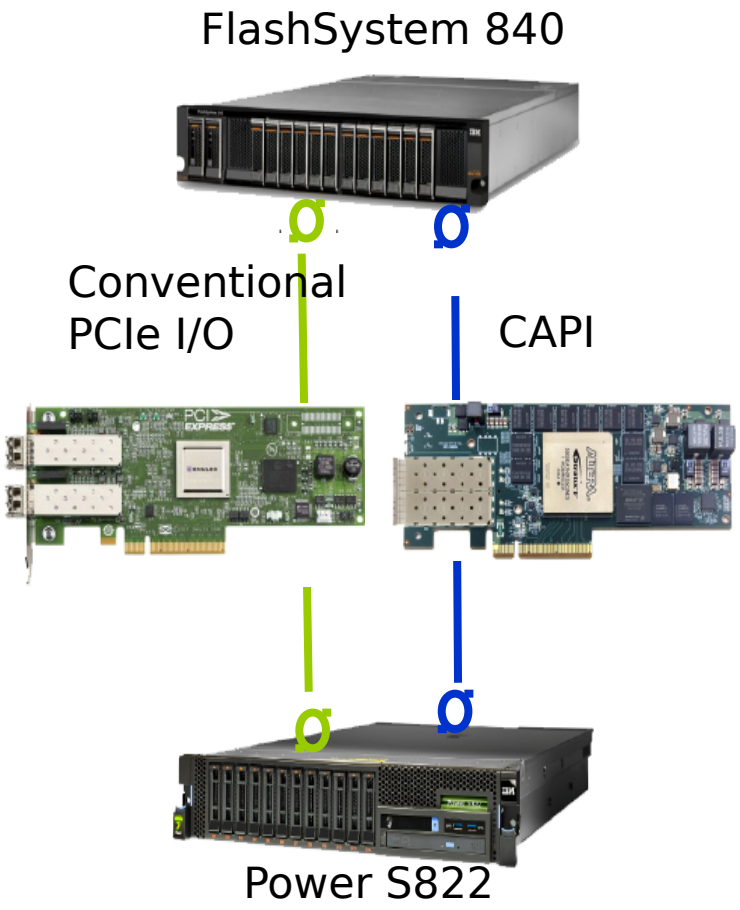
After: NoSQL POWER8 + CAPI Flash

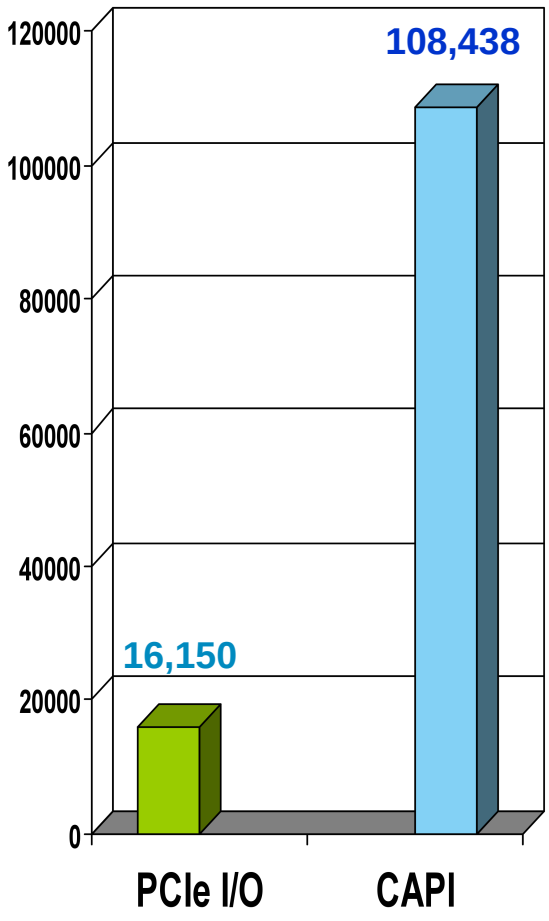


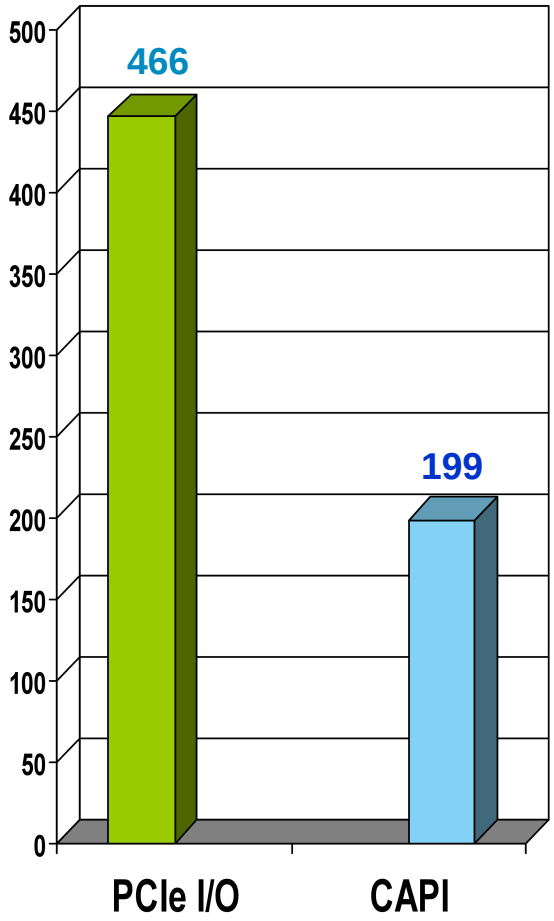
Flash Acts As Extension of System Memory

Demonstrating the Value of CAPI Attachment

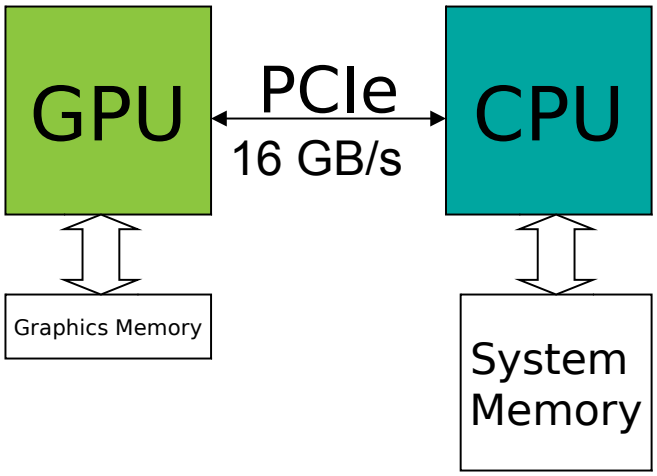
Identical hardware with
2 different paths to data



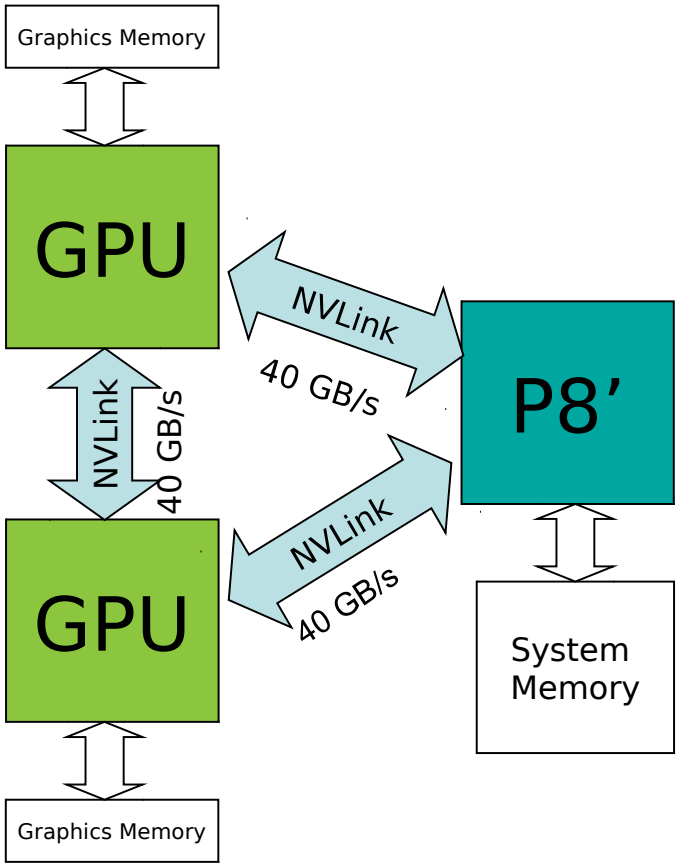




2.5x Faster CPU-GPU Connection via NVLink

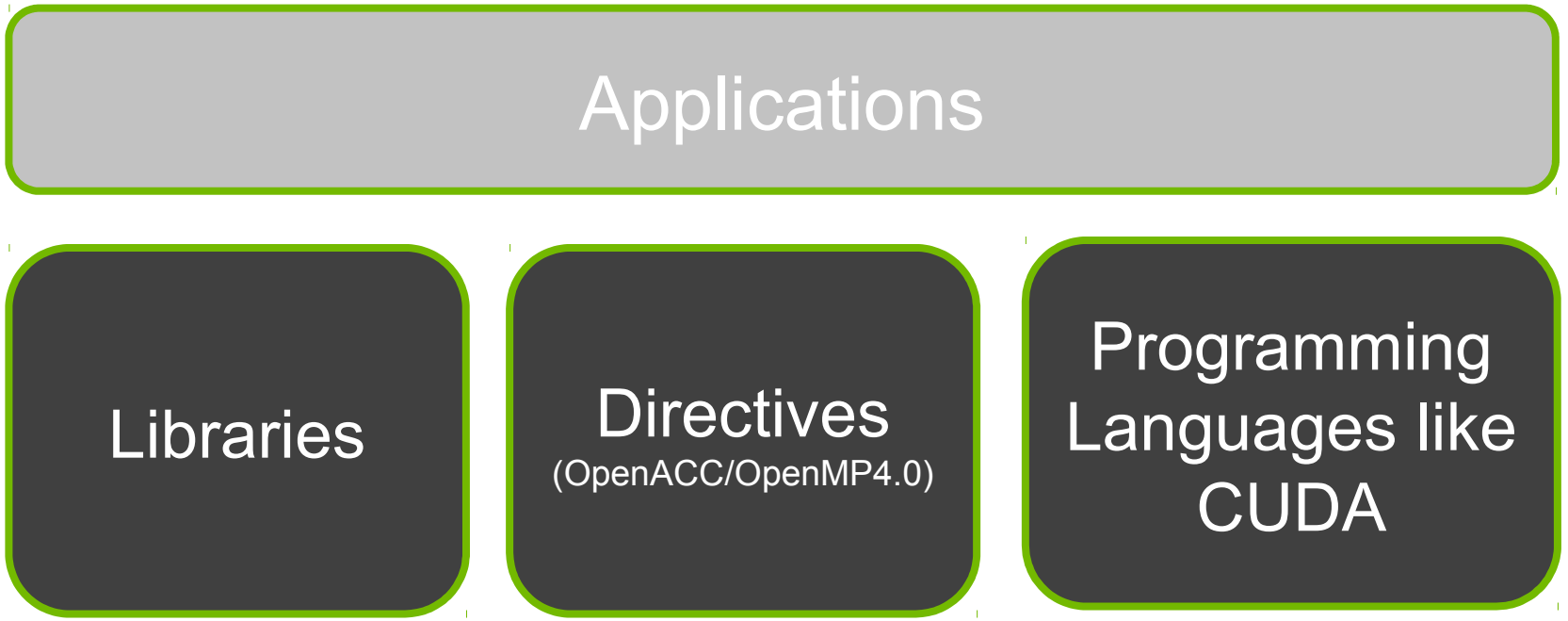


GPUs Bottlenecked by PCIe Bandwidth From CPU-System Memory

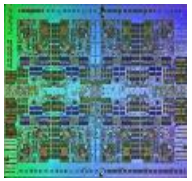
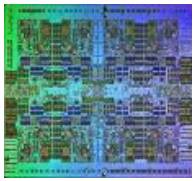
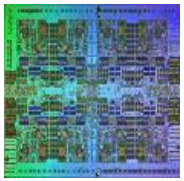
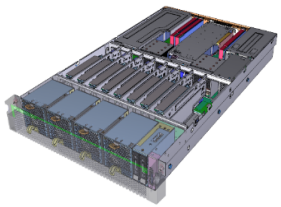
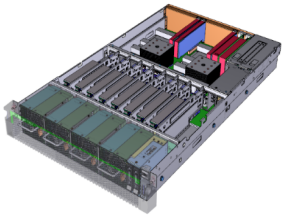
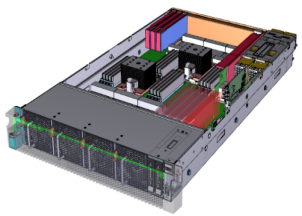


NVLink Enables Fast Unified Memory Access between CPU & GPU Memories

3 Ways to Accelerate Applications



IBM OpenPOWER-based HPC Roadmap

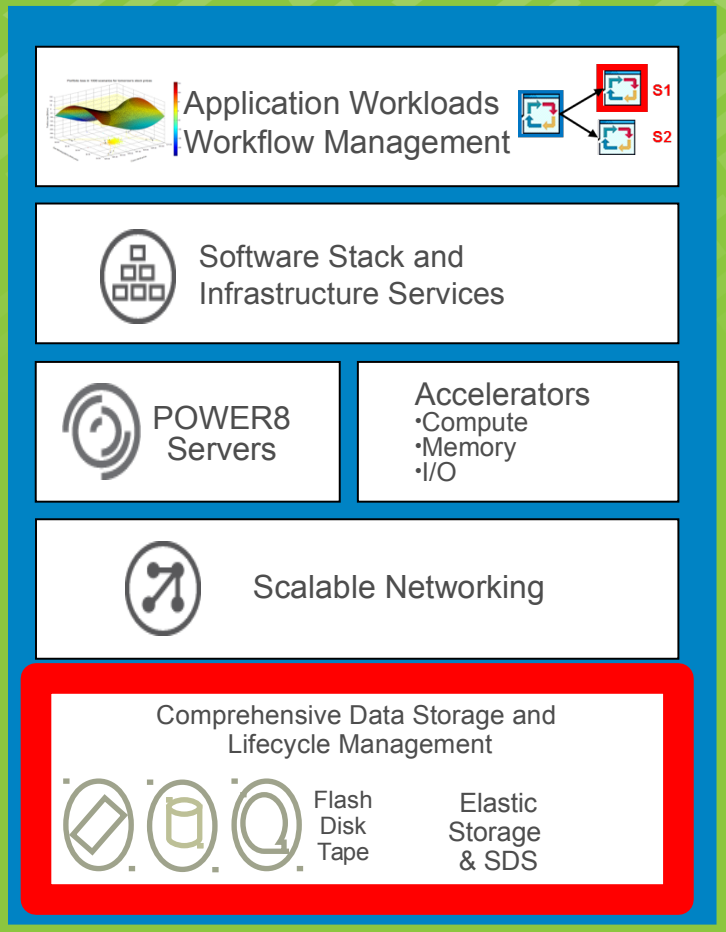
	2015	2016	2017
Mellanox Interconnect Technology	Connect-IB FDR Infiniband PCIe Gen3	ConnectX-4 EDR Infiniband CAPI over PCIe Gen3	ConnectX-5 Next-Gen Infiniband Enhanced CAPI over PCIe Gen4
NVIDIA GPUs	Kepler PCIe Gen3	Pascal NVLink	Volta Enhanced NVLink
IBM CPUs	POWER8  OpenPower CAPI Interface	POWER8+  NVLink	POWER9  Enhanced CAPI & NVLink
IBM Nodes			

Virtualization Support
IBM Power 8 (822LC)

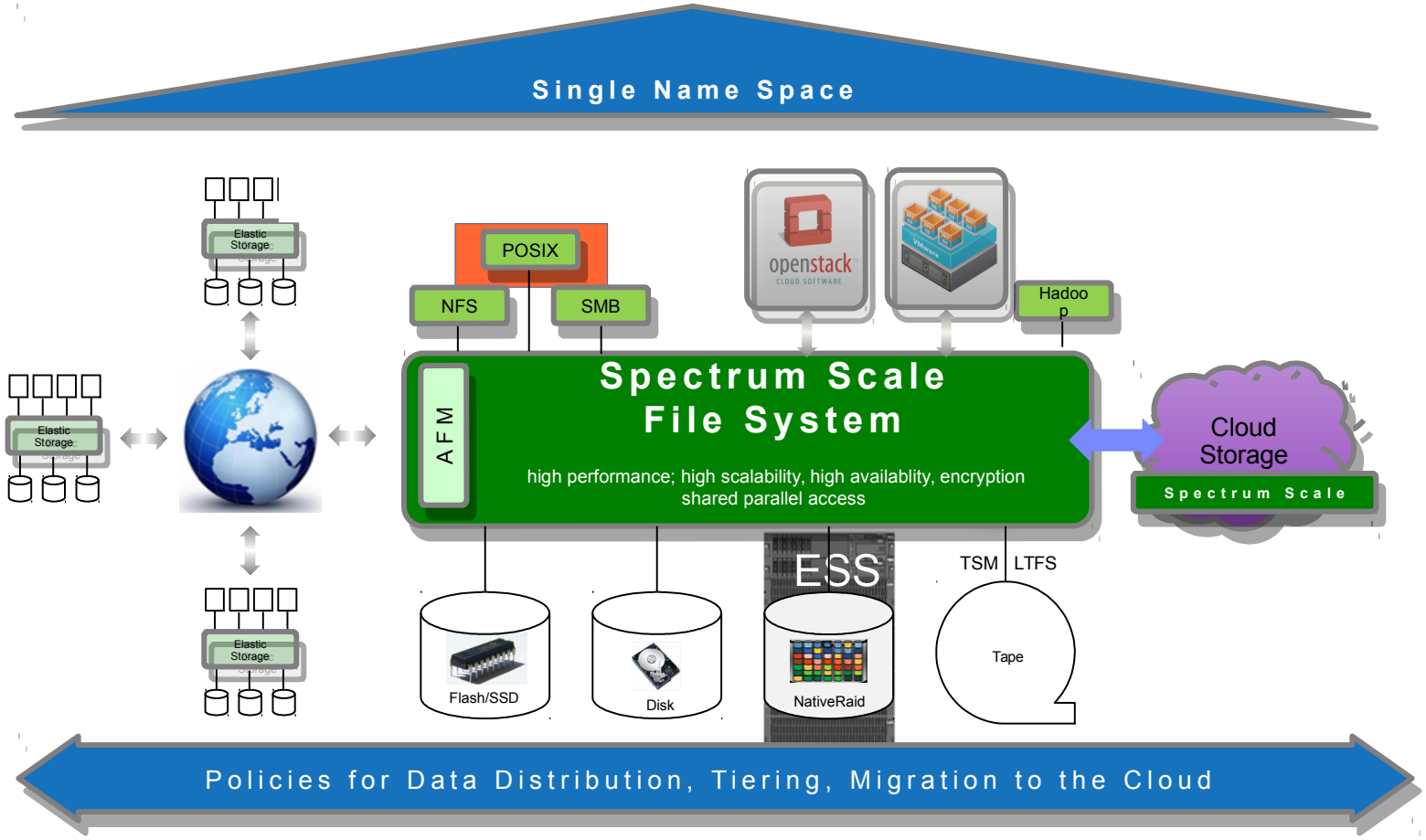
- PowerKVM 3.1
  - Big Endian:
    - Red Hat Enterprise Linux 7, any subsequent updates
    - Red Hat Enterprise Linux 6.6, any subsequent updates
    - SUSE Linux Enterprise Server 11 SP3, any subsequent updates
  - Little Endian:
    - Red Hat Enterprise Linux 7.1, any subsequent updates
    - SUSE Linux Enterprise Server 12, any subsequent updates
    - Ubuntu 14.04, any subsequent updates
    - Ubuntu 15.04 subsequent updates
- PowerVM (on 822L)
  - See IBM Knowledgebase

Data Management

Spectrum Scale

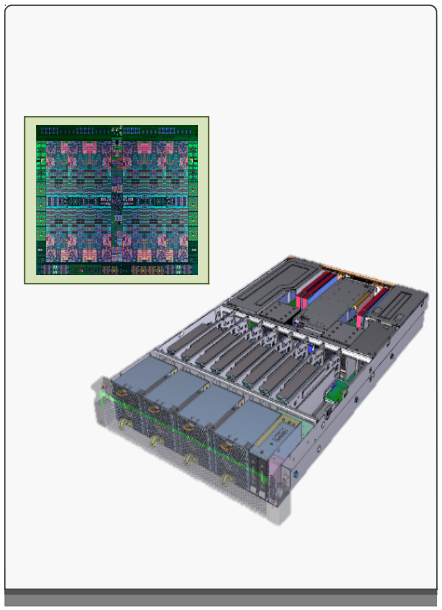


Software Defined Data Management

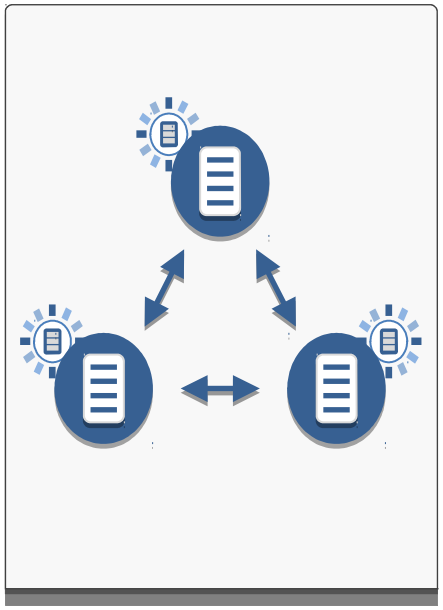


IBM Research Paving the Path to Next-Generation HPC

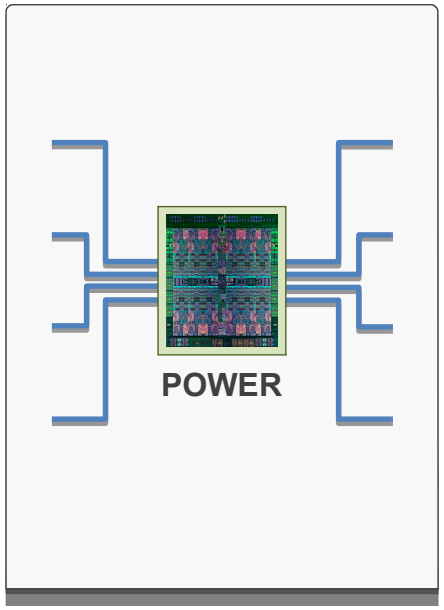
Data Centric System Node, & Processor Innovations



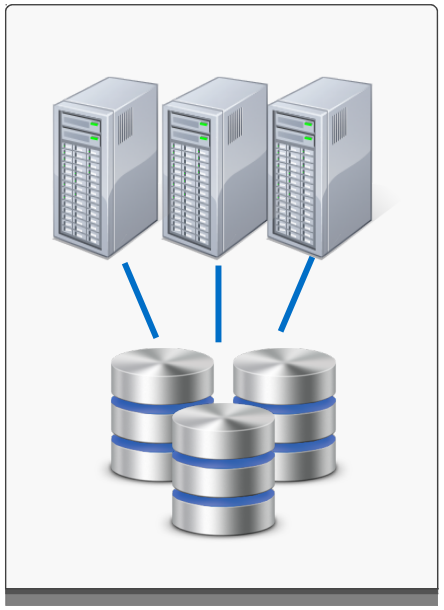
Programming Models for Exascale



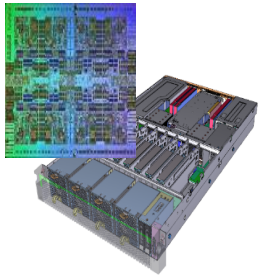
Enhancing Open Interconnects



Scalable High Performance Storage & File Systems

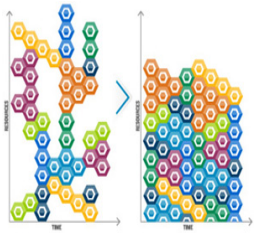


Portfolio of Technical Computing Solutions



Processors & Systems

- High Performance Processors & Systems
- Accelerator, networking, storage integration via CAPI & NVLink
- Innovative solutions like CAPI Flash



Software

- Platform LSF & Symphony workflow and resource management
- Compilers: gcc, IBM XLC, PGI Fortran/C/C++, Java, OpenACC, OpenMP
- Debuggers, Profilers, Math libraries, MPI & HPC apps
- Virtualisation through PowerKVM

High Performance File System & Storage

- High Performance Spectrum Scale (GPFS) Parallel File System
- Highest Performance HPC Storage: Elastic Storage Server
- Scalable Storage Solutions

Thank you