IBM Solutions for High Performance Computing looking at Large Scale Infrastructures

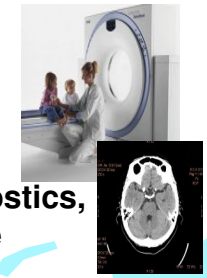
Euro-Par 2010, Ischia (Italy), Aug 31st – Sept 3rd 2010

Marco Briscolini
IBM Italy
Deep Computing Sales
http://www.ibm.com/systems/deepcomputing/
marco_briscolini@it.ibm.com

Digital Media

Digital content creation, management and distribution, online gaming, surveillance

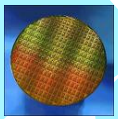
Life Sciences

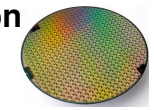


Research, drug discovery, diagnostics, information-based medicine

Electronic Design Automation

Chip design and fabrication





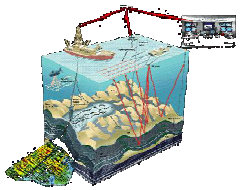
High Performance Supercomputing driving leading edge applications

Financial Services

Optimizing IT infrastructure, risk management and compliance, analytics

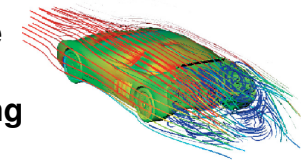
Petroleum

Oil and gas exploration and production



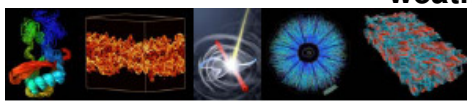
Automotive/Aerospace Engineering

Automotive, Aerospace and Defense Electronics & Engineering



Government & Higher Education

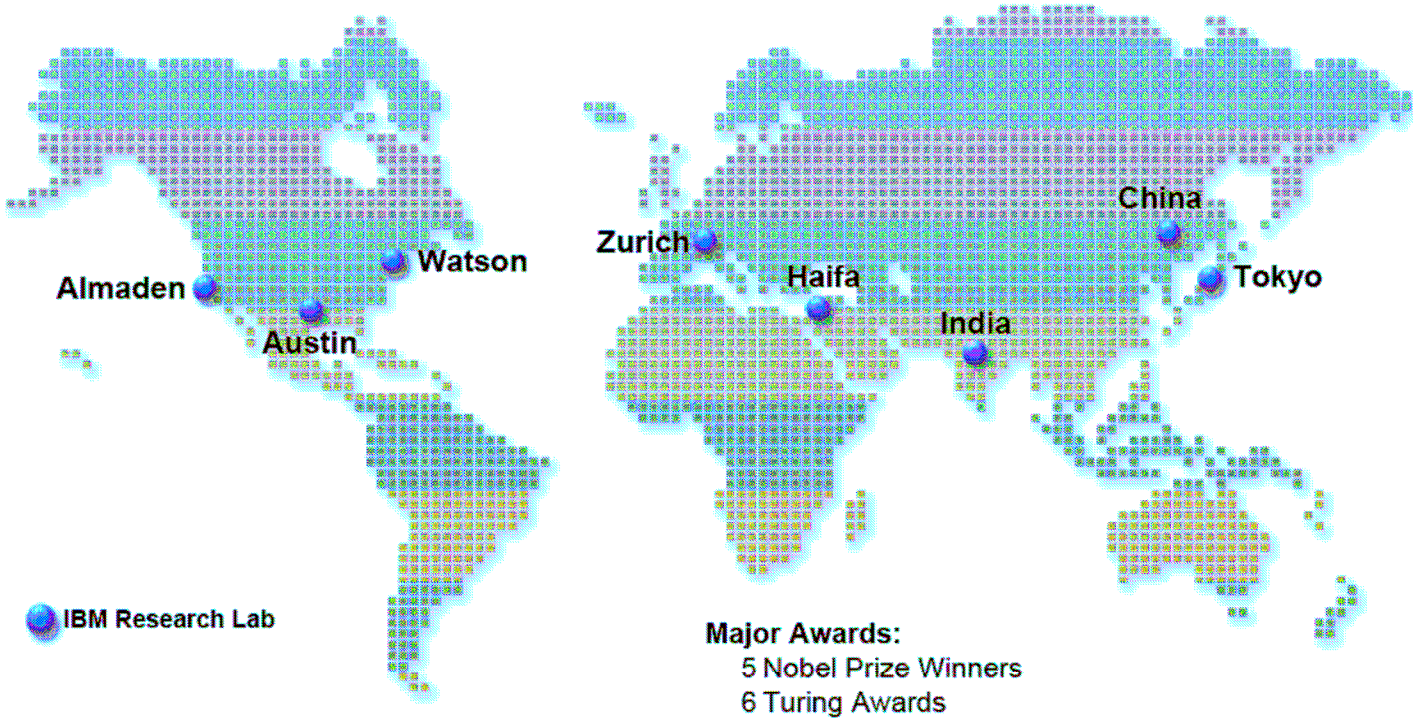
Scientific research, classified/defense, weather/environmental sciences



IBM Research Labs around the World

IBM Research

~3,000 employees



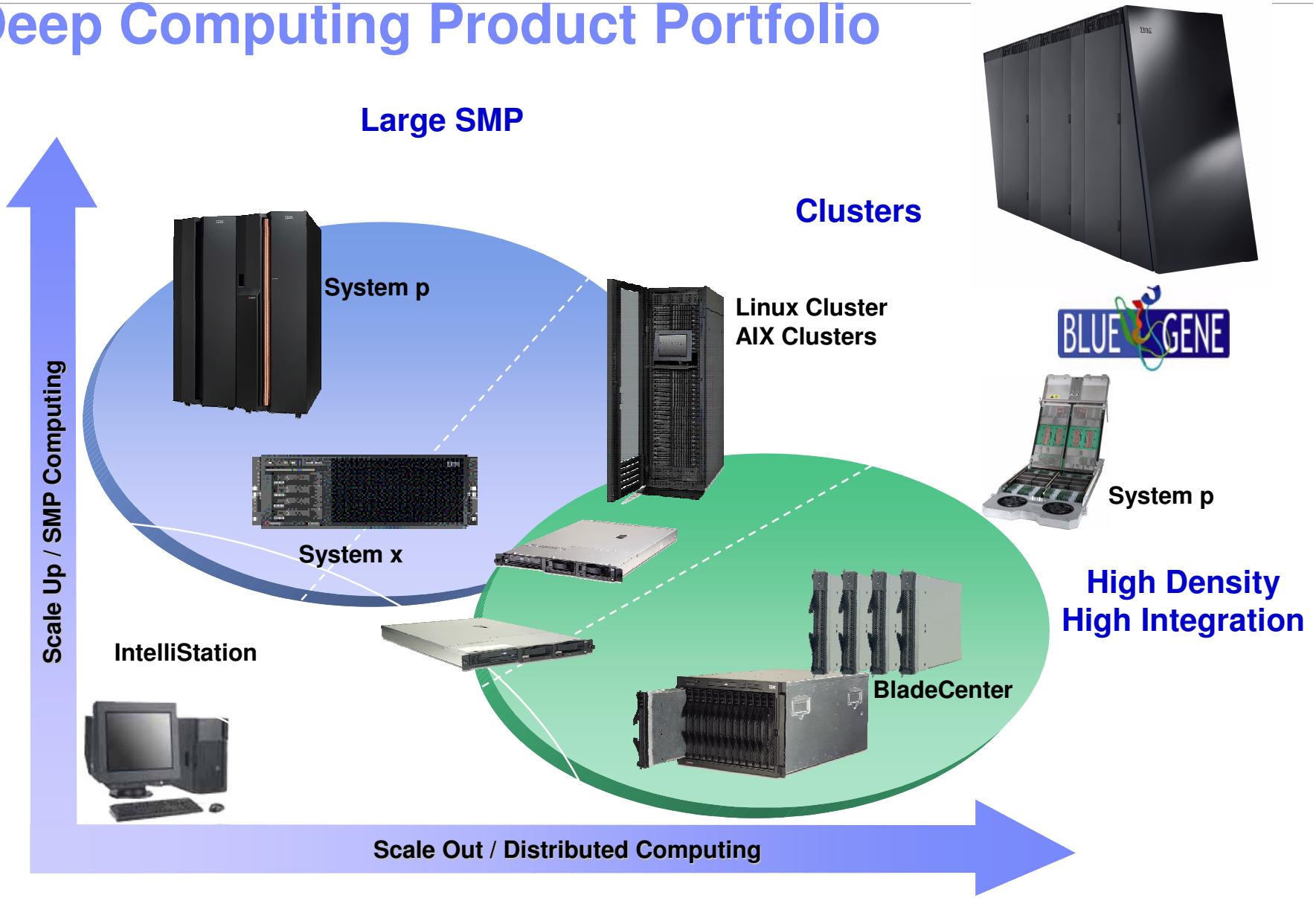
Major Awards:

5 Nobel Prize Winners
6 Turing Awards
14 National Medals
11 Inductees in Inventors' Hall of Fame

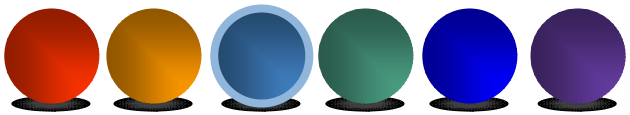
Major Initiatives: 2010 Big Bets

- Healthcare Transformation
- Smarter Cities
- Service Quality
- Mobile Web
- Massive Scale Analytics
- Cloud Computing
- Workload-Optimized Systems
- Nanotech

Deep Computing Product Portfolio

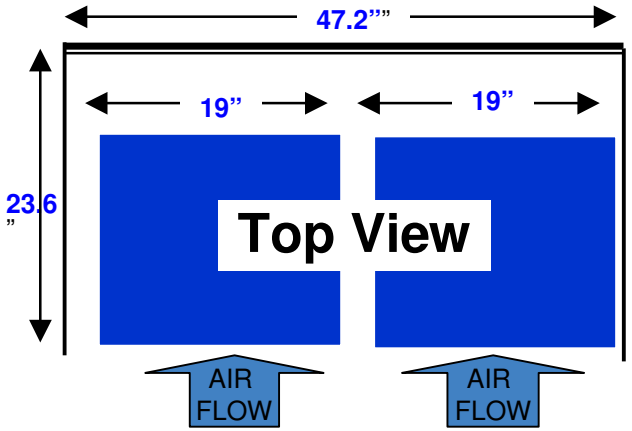


IBM iDataPlex
higher density for Intel x86 Cluster solution

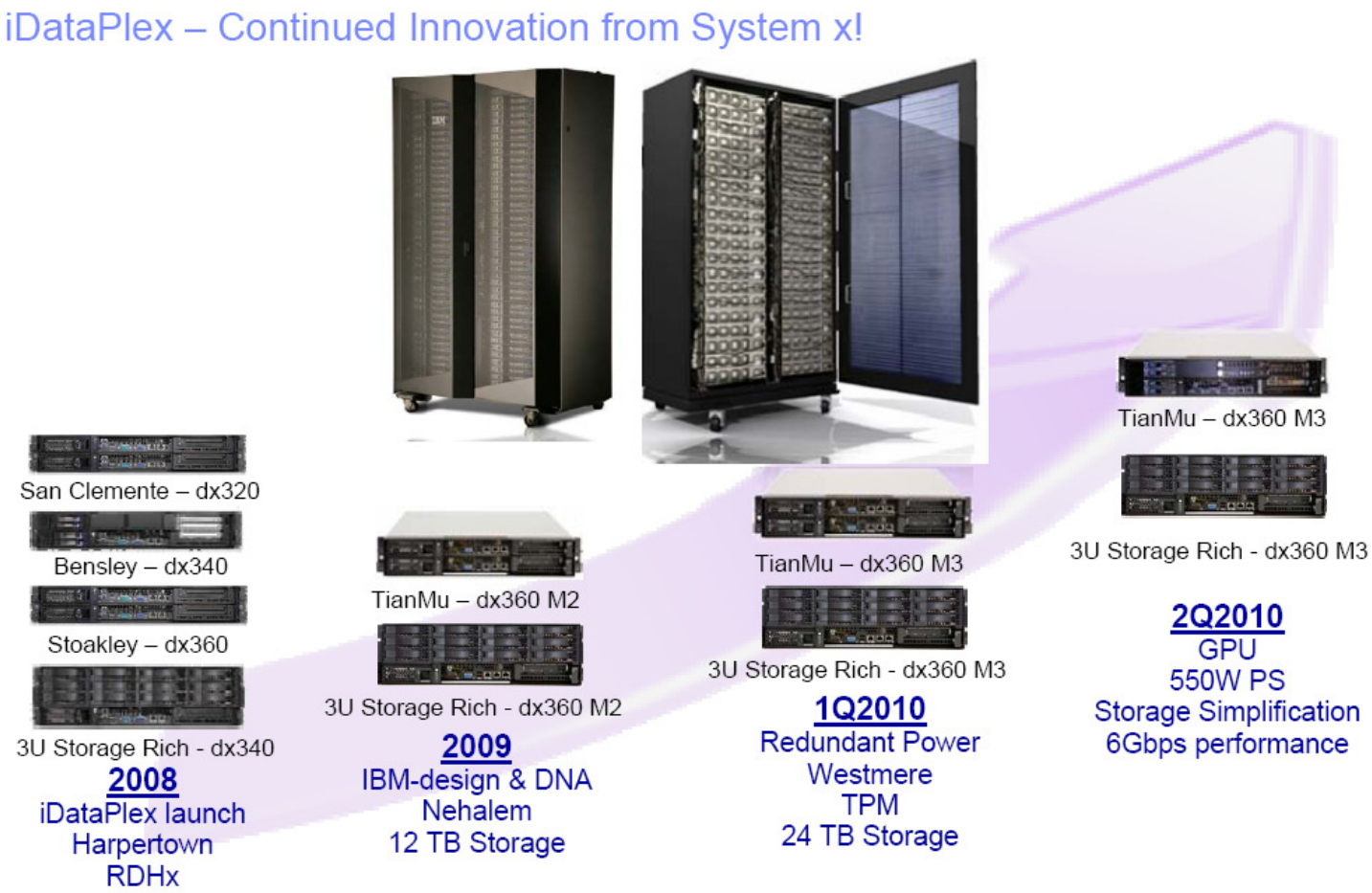


iDataPlex Rack Design

- **Energy efficiency The iDataPlex rack structure is optimized for datacenter cooling efficiency, density and deployment flexibility**
  - Half-depth rack optimizes airflow for cooling efficiency
  - Reduces pressure drop to improve chilled air efficiency
- **Leadership density - 100U Rack: 84 U of server and storage and 16 U of switch and PDU space in standard rack footprint**
  - Dual column / Half depth rack
  - Std 2 floor tile rack footprint
  - Up to 168 physical nodes in 8 sq ft
- **Flexibility: Fits in today's data center, optimized for tomorrow's**
  - Matches US & European data center floor tile standards
  - Compatible with standard forced air environments
- **Ease of use: All service and cabling from the front**



iDataPlex Server Design

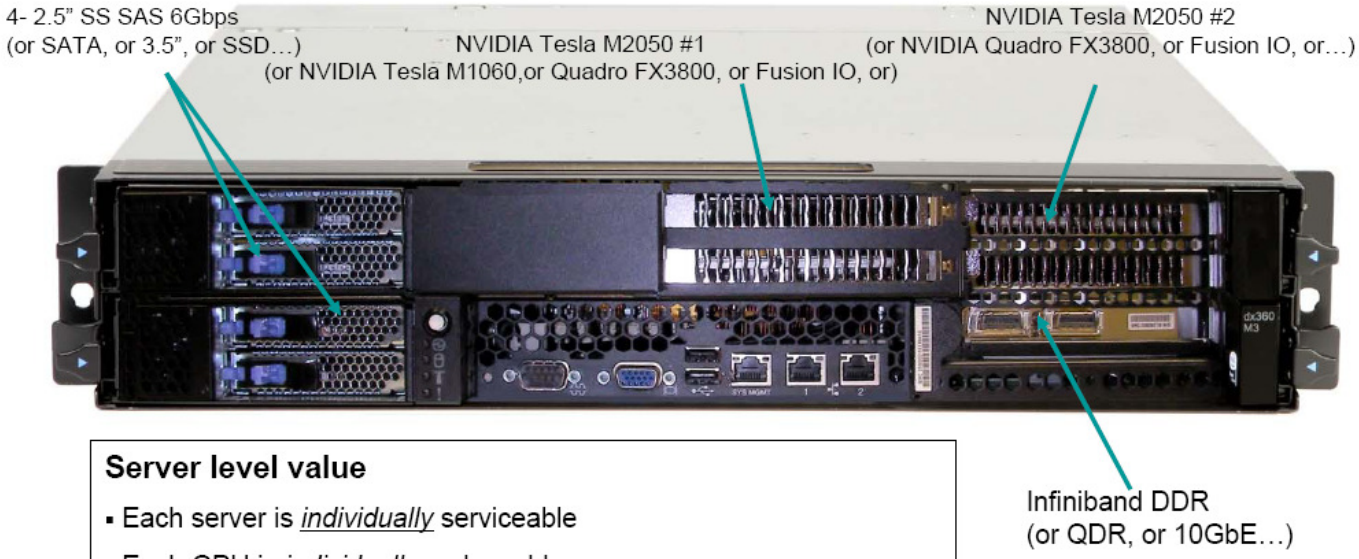


IBM dx360 M3 integrates 2 x GPGPU

May 18, 2010 Launch

dx360 M3 Refresh - Server GPU Configuration

Announce – May 18, 2010 : Ship Support – July 1, 2010



Server level value

- Each server is *individually* serviceable
- Each GPU is *individually* replaceable
- 6Gbps SAS drives and controller for *maximum performance*
- *Service and support* for server and GPU from IBM

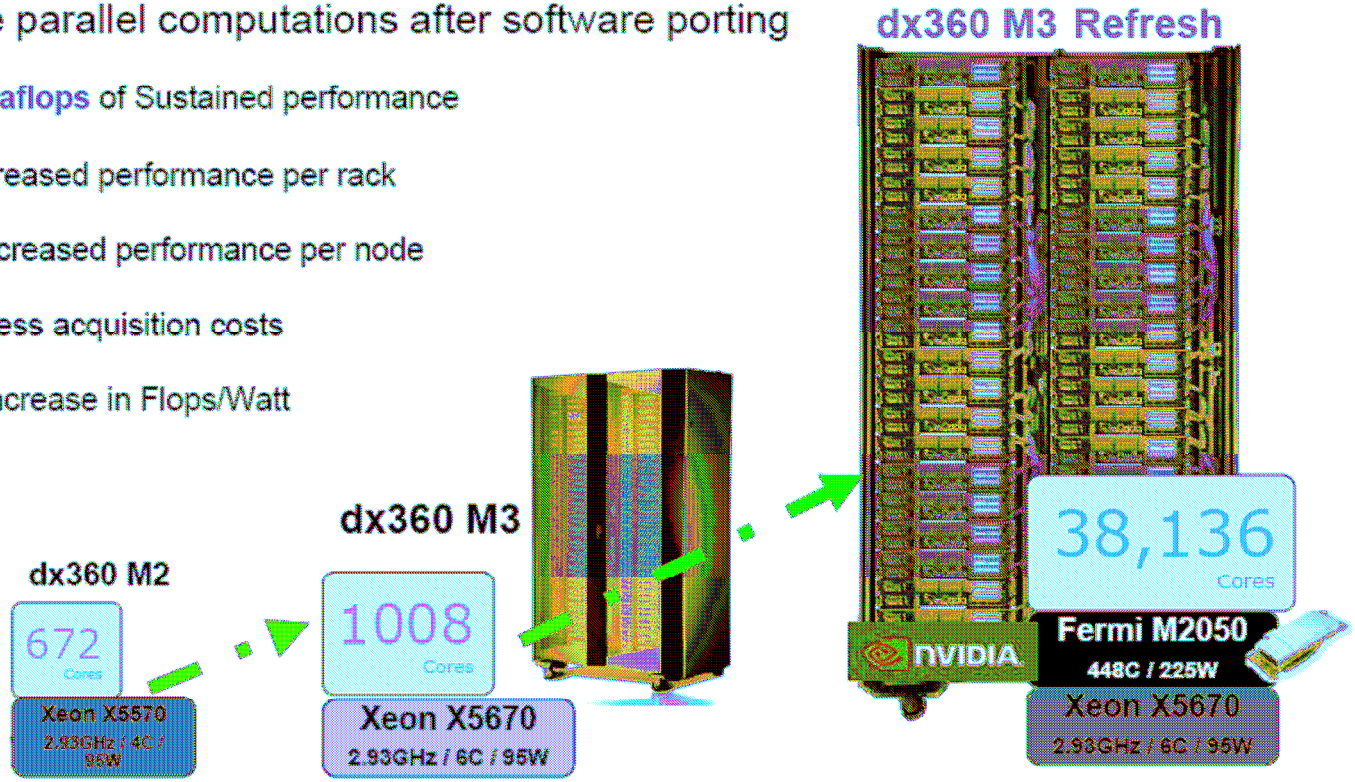
iDPX: Do More with Maximum Performance Density

May 18, 2010 Launch

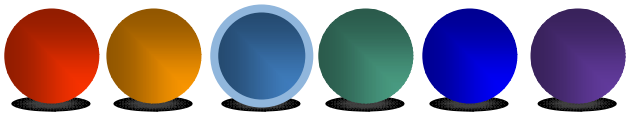
Do More with Maximum Performance Density

Compared to first-generation Intel® Xeon® processor-based iDataPlex servers, the dx360 M3 server with 2 GPUs improves performance density in the data center for massive parallel computations after software porting

- **49 Teraflops** of Sustained performance
- **4X** increased performance per rack
- **10X** increased performance per node
- **65%** Less acquisition costs
- **3.7X** increase in Flops/Watt



IBM Power Technology in Large Scale System



Power Systems HPC Roadmap

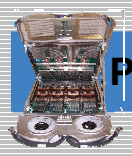
Power 755, Blue Gene, Power 575

Extreme Scaling

Blue Gene/P

Blue Gene/Q

Mid to High-end Capability



Power 575 (POWER6)

POWER7 IH **P7+ IH**

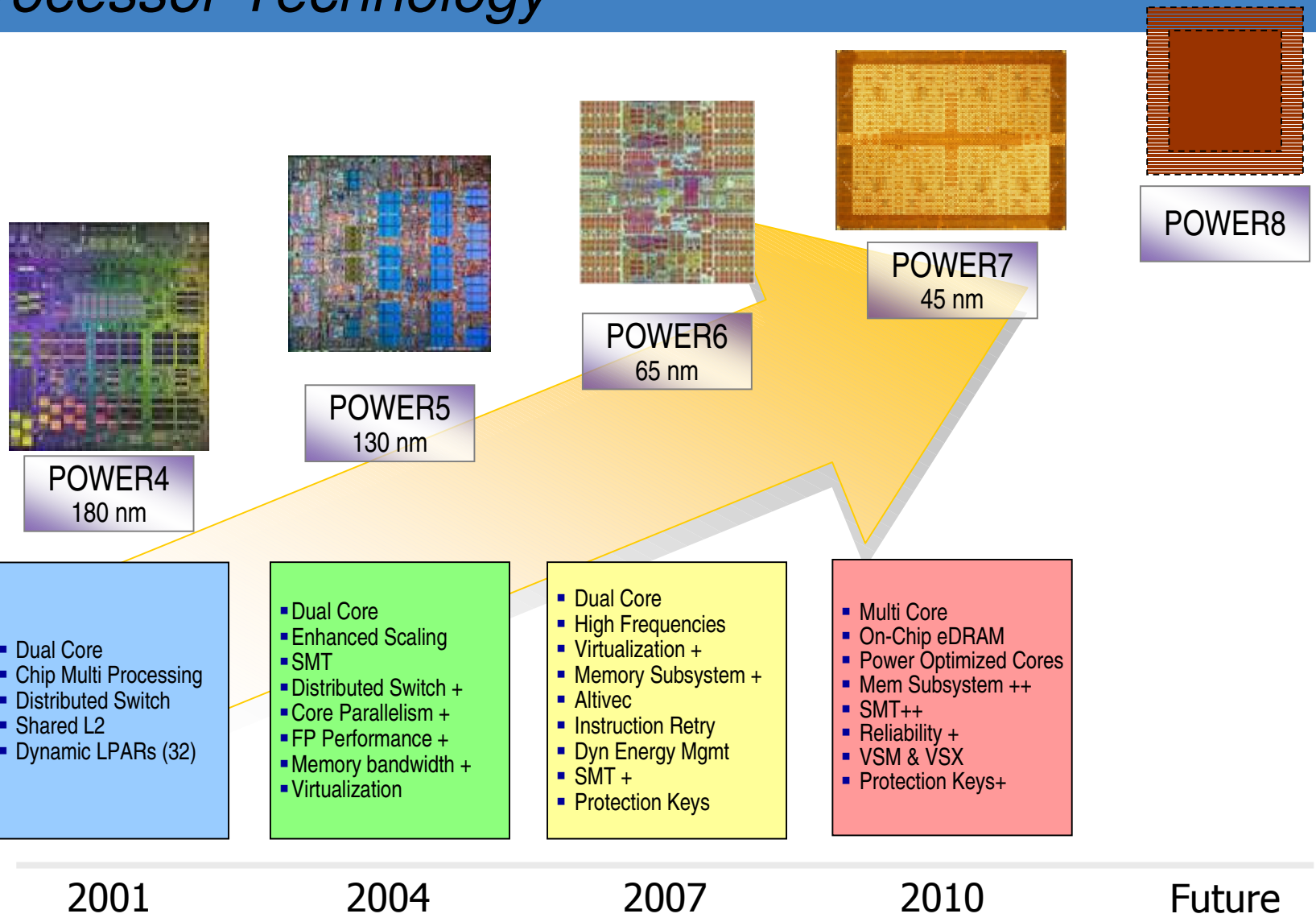
Power 755

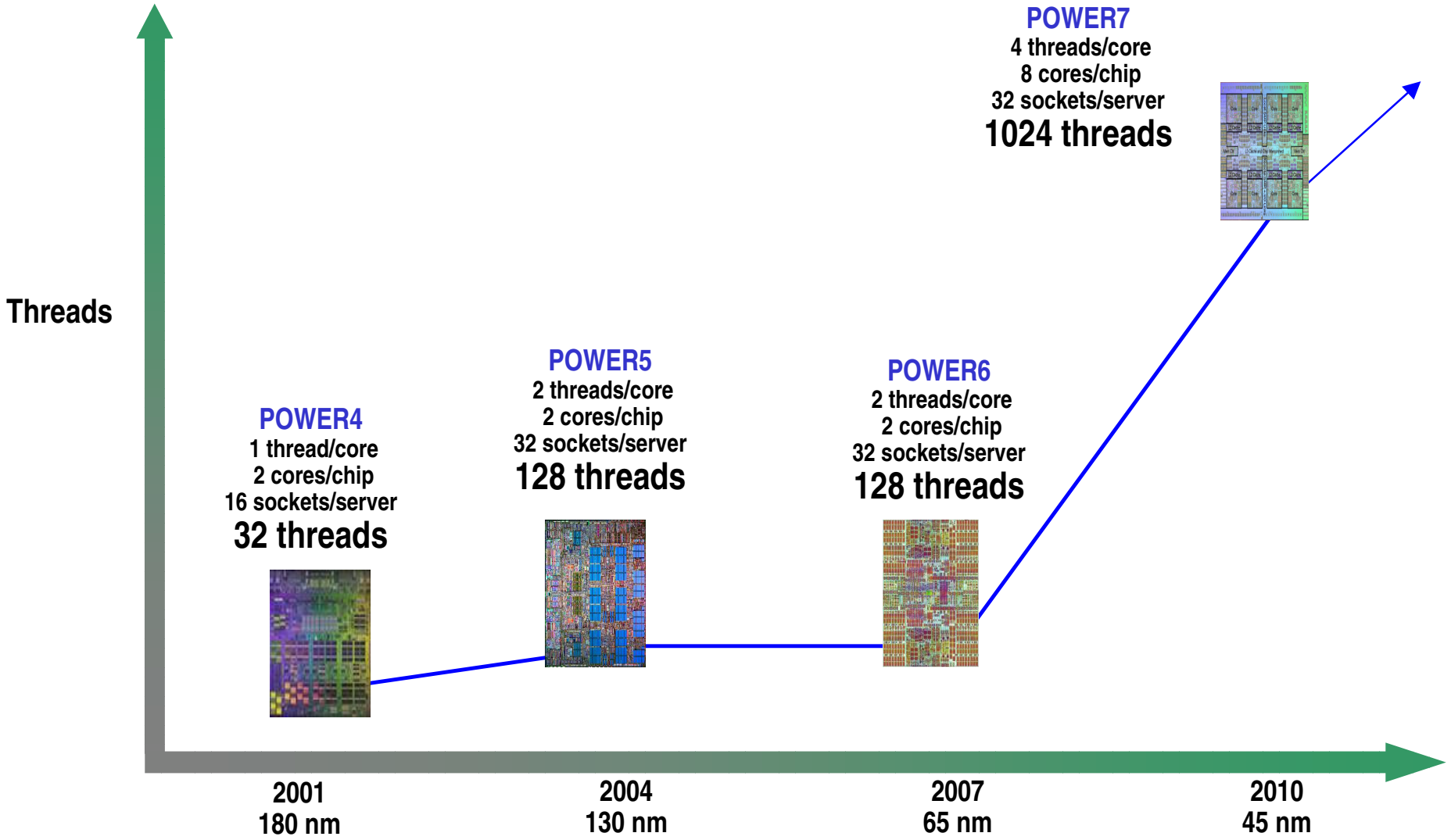
Mid Range Departmental & Divisional Systems

Power 755

2010 **2011** **2012**

Processor Technology

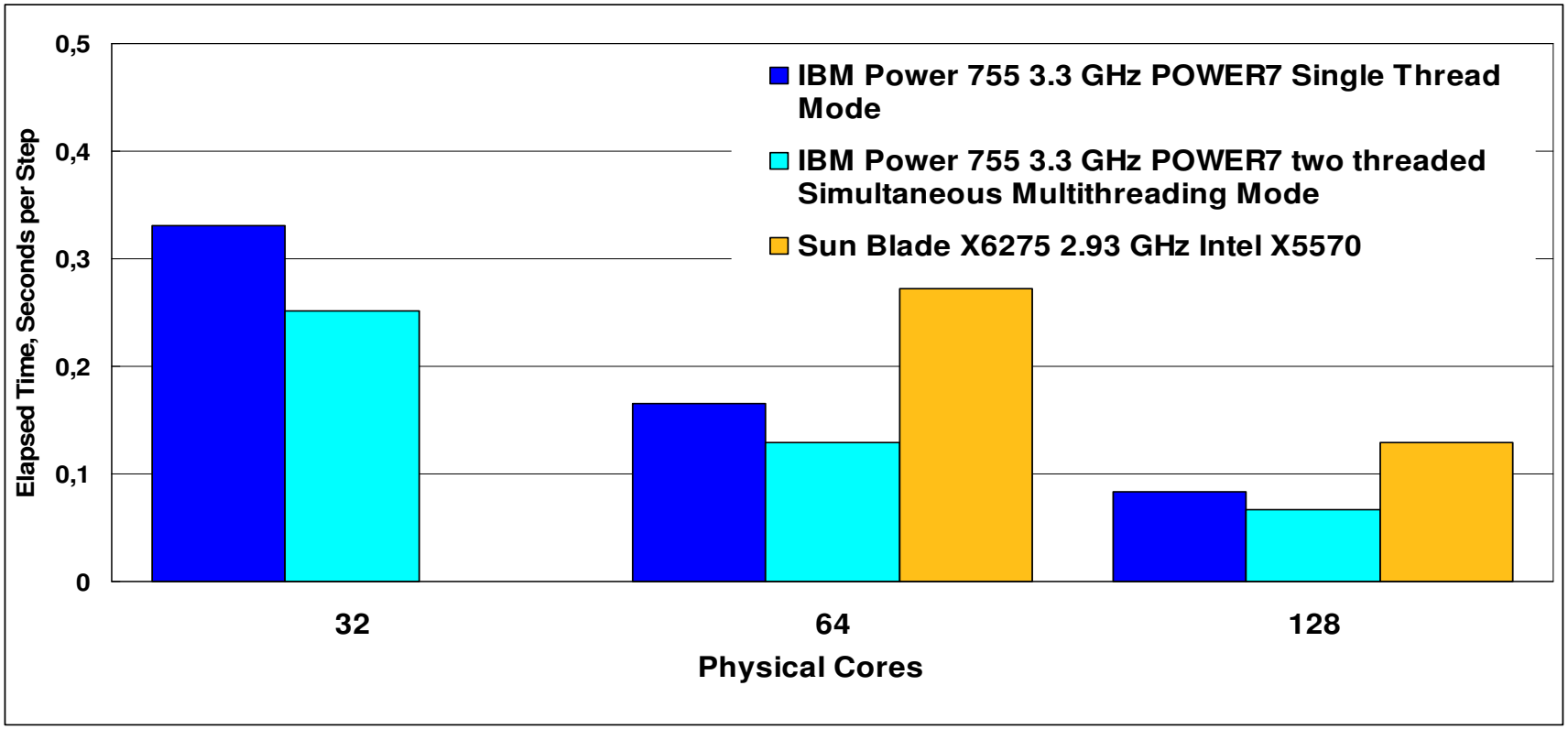




NAMD 2.7b1

STMV Benchmark: 1,066,628 atoms, 12A cutoff + PME every 4 Steps, periodic, total 500 Steps

Elapsed Time in seconds per step – A lower number indicates better performance

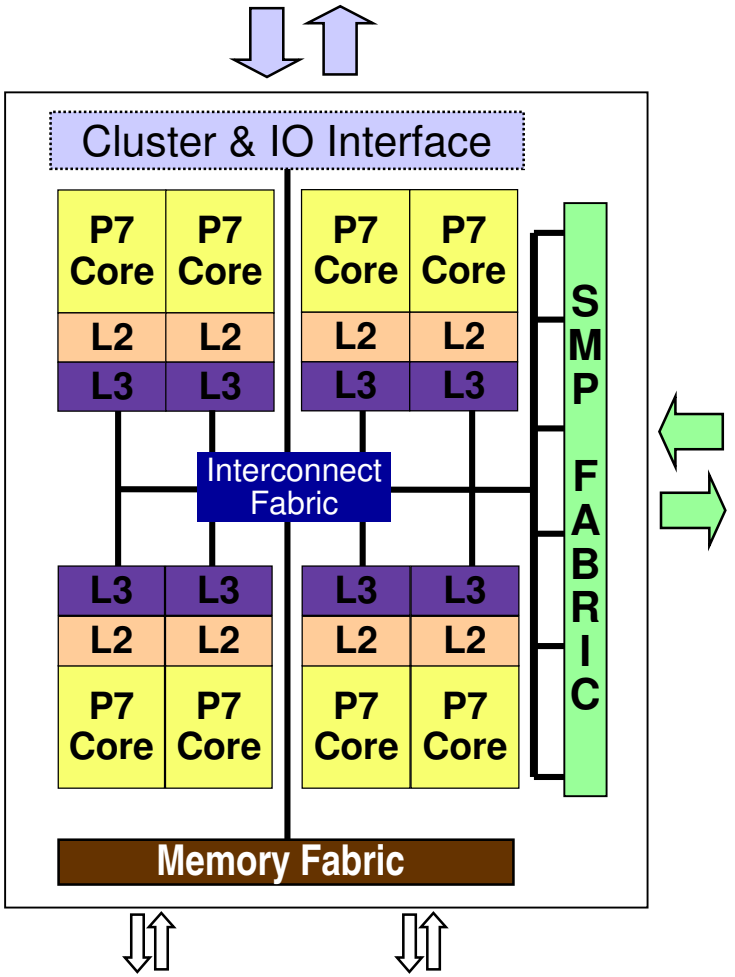


- Power 755 benchmarks were performed in Single Thread Mode and two-threaded Simultaneous Multithreading Mode
- Sun X6275 data current as of 1/24/2010, http://blogs.sun.com/BestPerf/entry/sun_blade_6048_and_sun1
- IBM data current as of 1/24/2010.

POWER7 Architecture – Key features for HPC

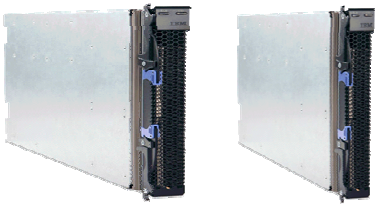
Key Features:

- 8 Cores
- Core frequency : 3 ~ 4 GHz
- On Chip 4MB L3 Cache/core
- Extended SIMD Support
  - Altivec – same as in Power6 and PPC 970
  - VSX – 145 instruction set
    - 4 DP FMAs /cycle
- Multiple Memory Controllers
- 3rd Generation Multi-Threading
  - Enhanced performance
- DDR3 memory support (1066, 1333 MHz)
- 4th Generation SMP Fabric Bus
- Other:
  - Stride N prefetching

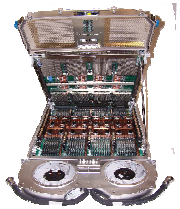


Entry (HV)
Up to 32 Cores

Blade
Up to 16 Cores
Up to 8 Cores

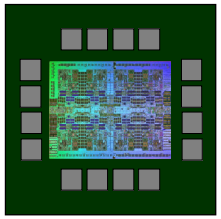


Compute / Cluster (575)
PERCS
2U Building Blocks



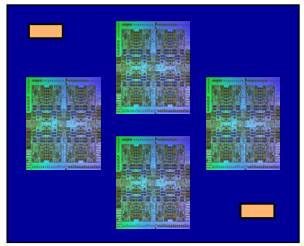
High End
Up to 200+ Cores

2/4s Blades and Racks
Single Chip Organic



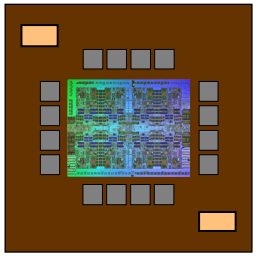
1 Memory Controller
3 4B local links

Compute Intensive
Quad-chip MCM



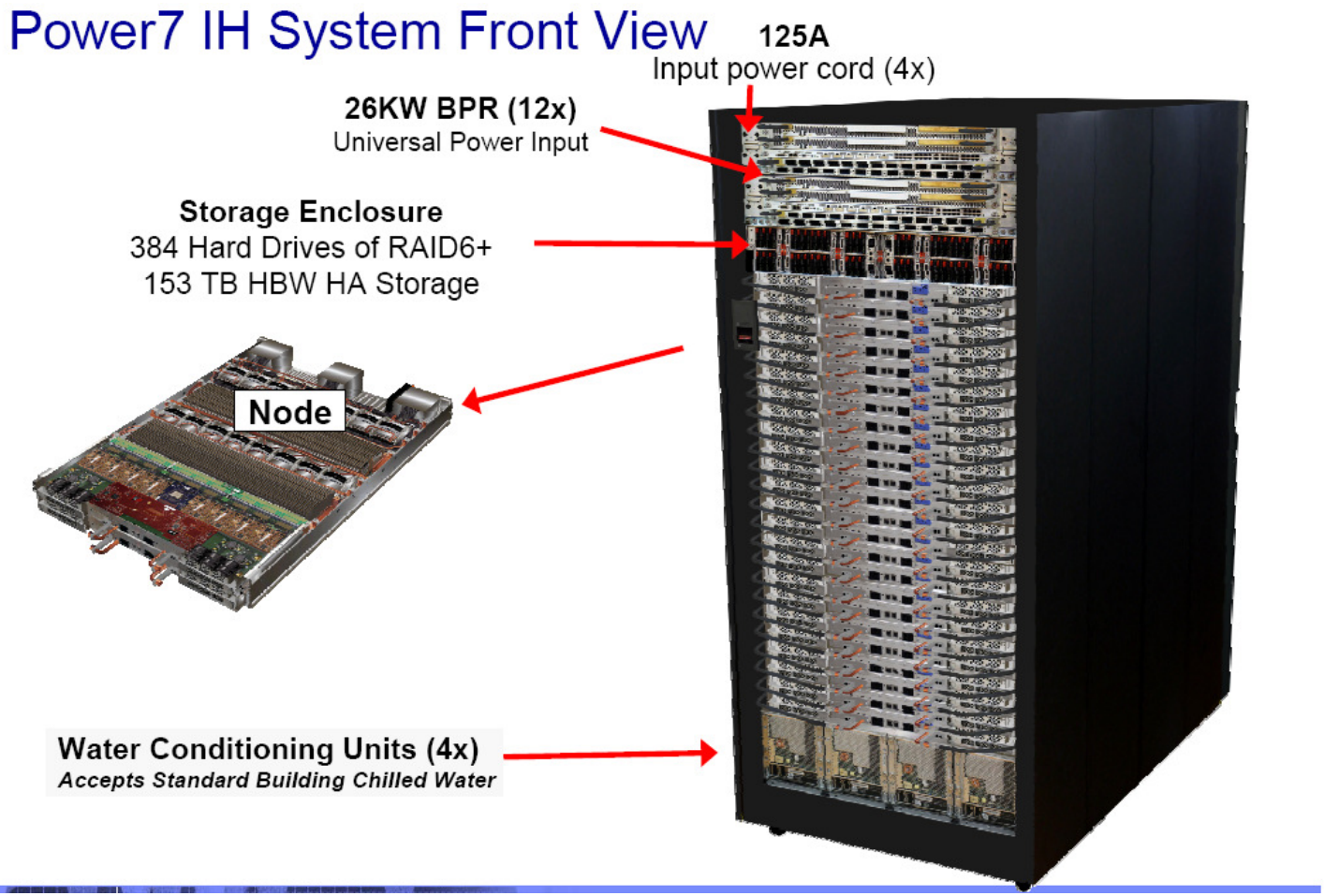
8 Memory Controller
3 16B local links (on MCM)

High-End and Mid-Range
Single Chip Glass Ceramic

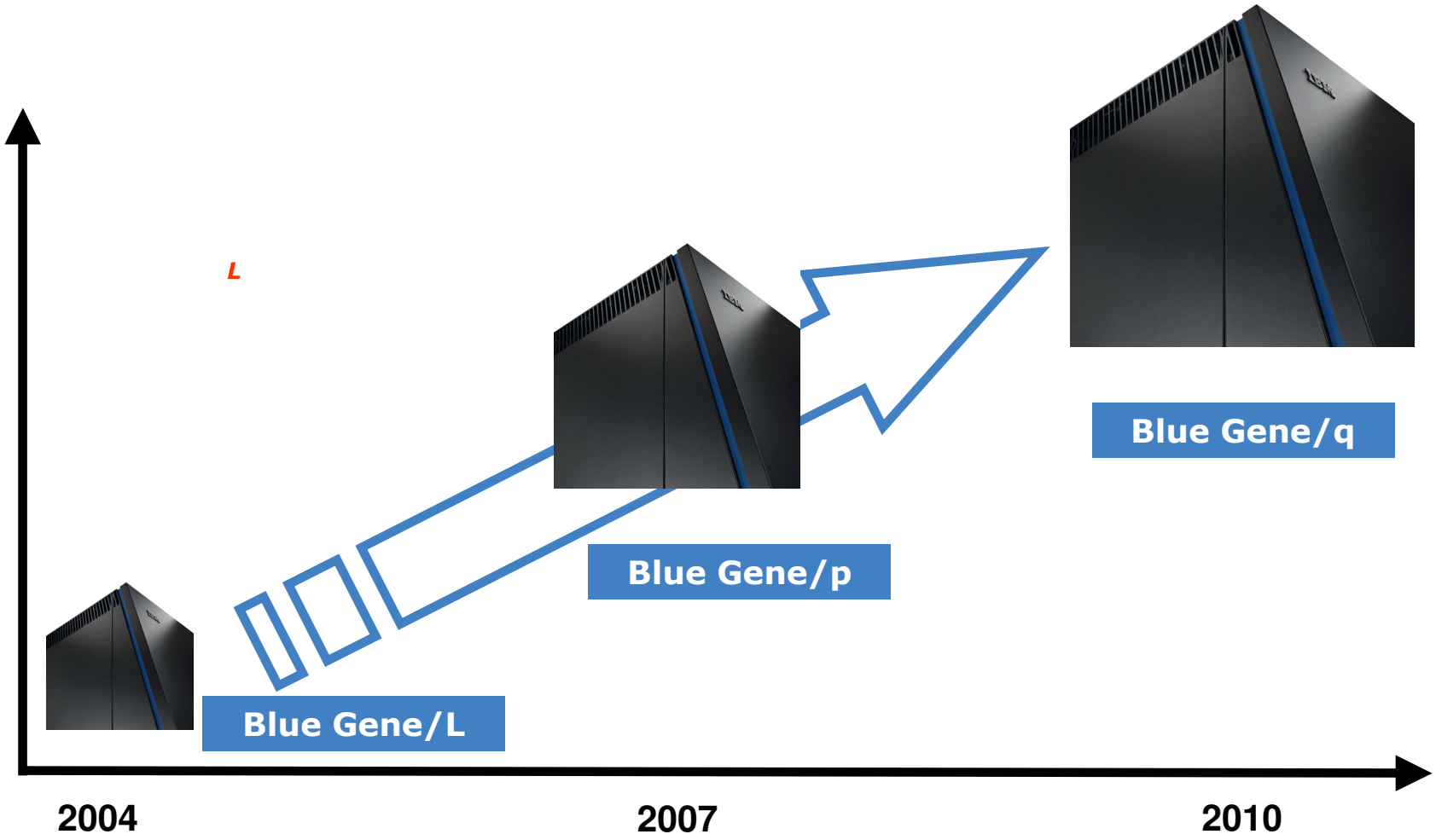


2 Memory Controllers
3 8B local links
2 8B Remote links

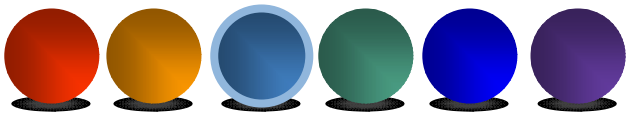
Compute cluster (575) Building Blocks: around 100TFLOPs in one rack



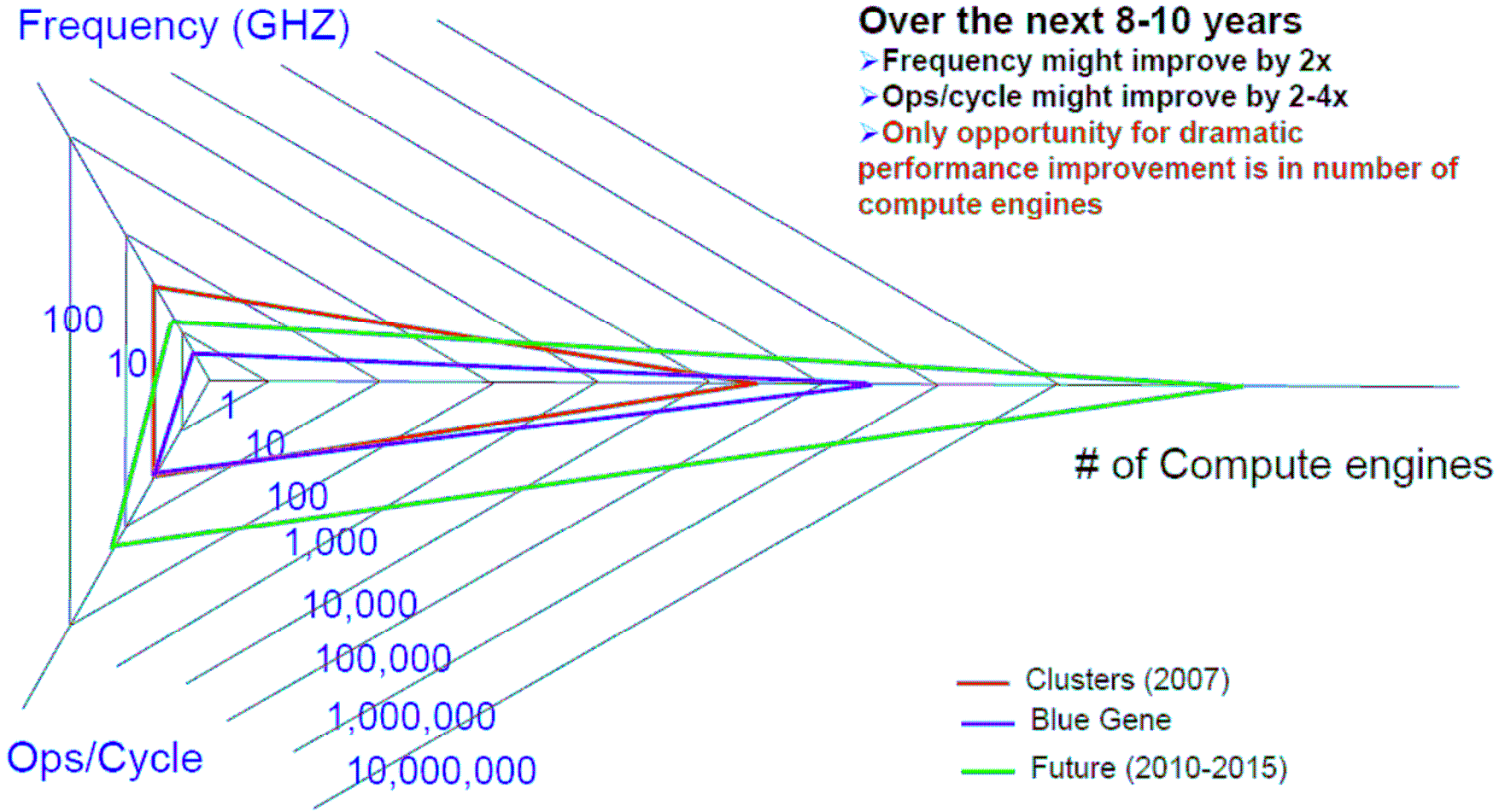
Blue Gene technology roadmap



Manycore Technology Trends



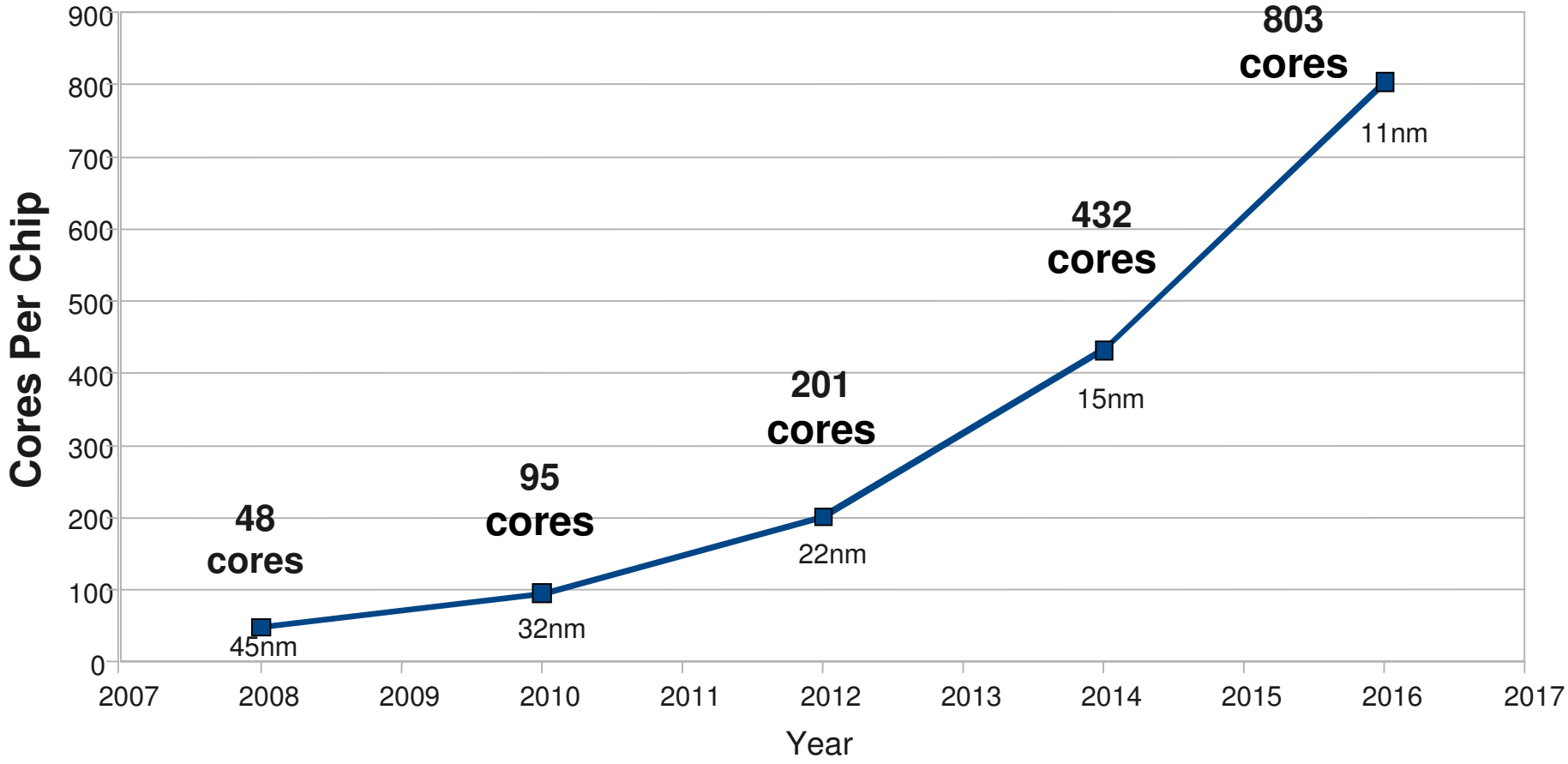
Increased Transistor Count is Driving Performance



Over the next 8-10 years

- Frequency might improve by 2x
- Ops/cycle might improve by 2-4x
- Only opportunity for dramatic performance improvement is in number of compute engines

Servers will have thousands of execution threads available

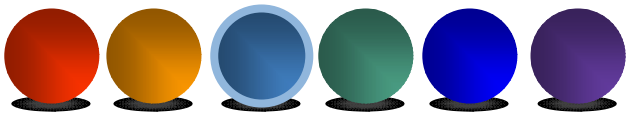


Linear projection based on Intel's 48 core datacenter chip (27 Million transistors per core) and lithographic improvements

A new programming language to drive programmer productivity and scaling in the Multicore era

- **Introduce X10, a parallel programming language that has been funded by DARPA to achieve high productivity and high performance for the science/engineering community**
  - IBM chose a broader programming model for productivity to enable use by middleware and commercial HPC programmers
  - 6x productivity improvement using X10 and its development environment over C/MPI (2009 productivity study at Rice University)
- **X10 provides**
  - Java-like language
  - Ability to specify fine-grained concurrency
  - Ability to represent heterogeneity at language level
  - Single programming model for computation offload
  - Migration path
    - X10 concurrency idioms can be realized in other languages, Java, C, Fortran, via library annotations that communicate with the APGAS runtime
- **Leverages 5+ years of research and development via PERCS/HPCS**
- **Community building activities already underway (Columbia, CMU, Rice, etc.)**
  - Tutorials and graduate classes using X10 in Fall'09
  - Open Collaborative Research and grants

IBM GPFS™
Parallel File System in Large Scale Infrastructure



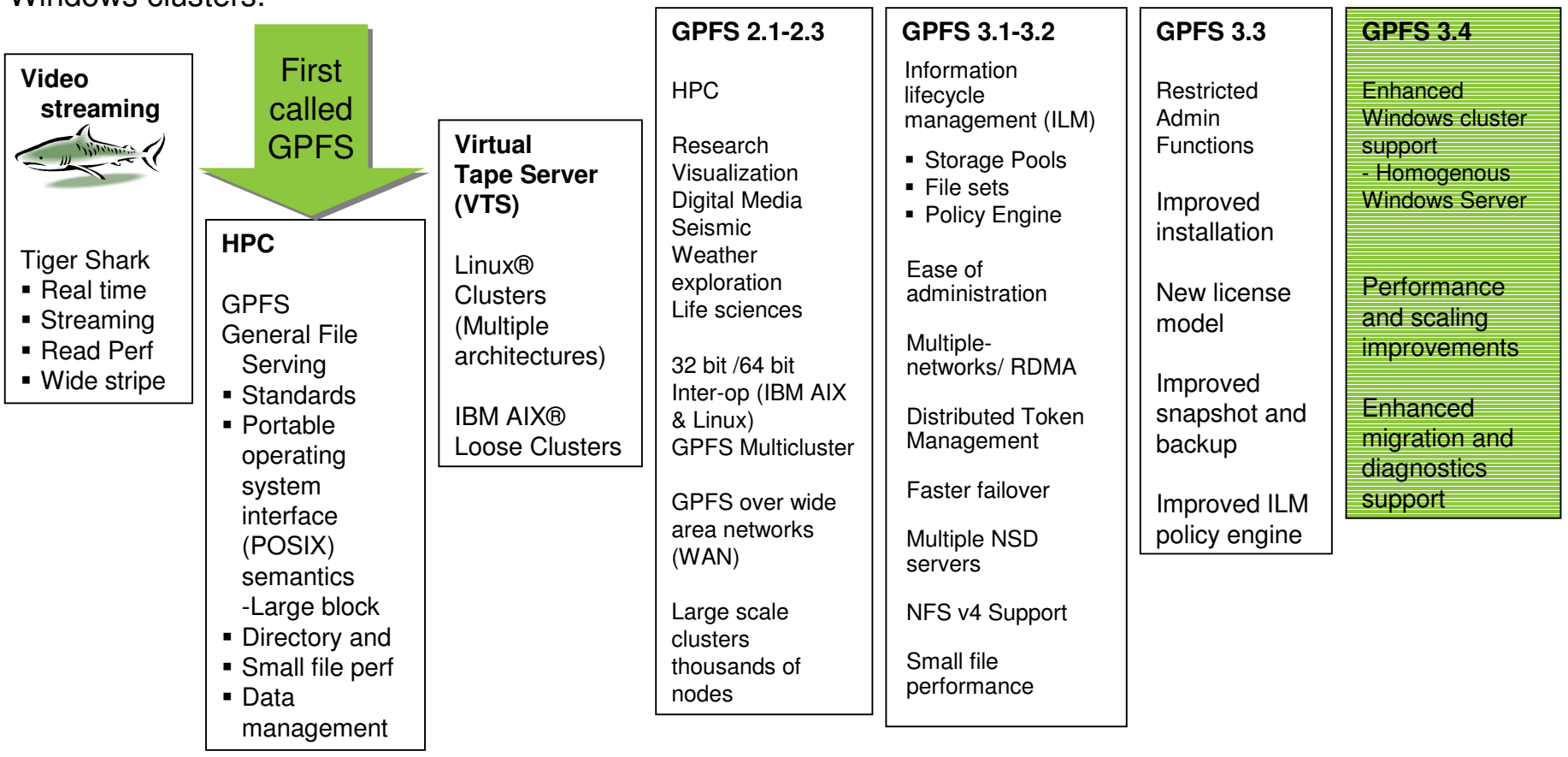
IBM General Parallel File System (GPFS™)

GPFS is a scalable, high-performance file management infrastructure for IBM AIX®, Linux® and Windows™ systems.

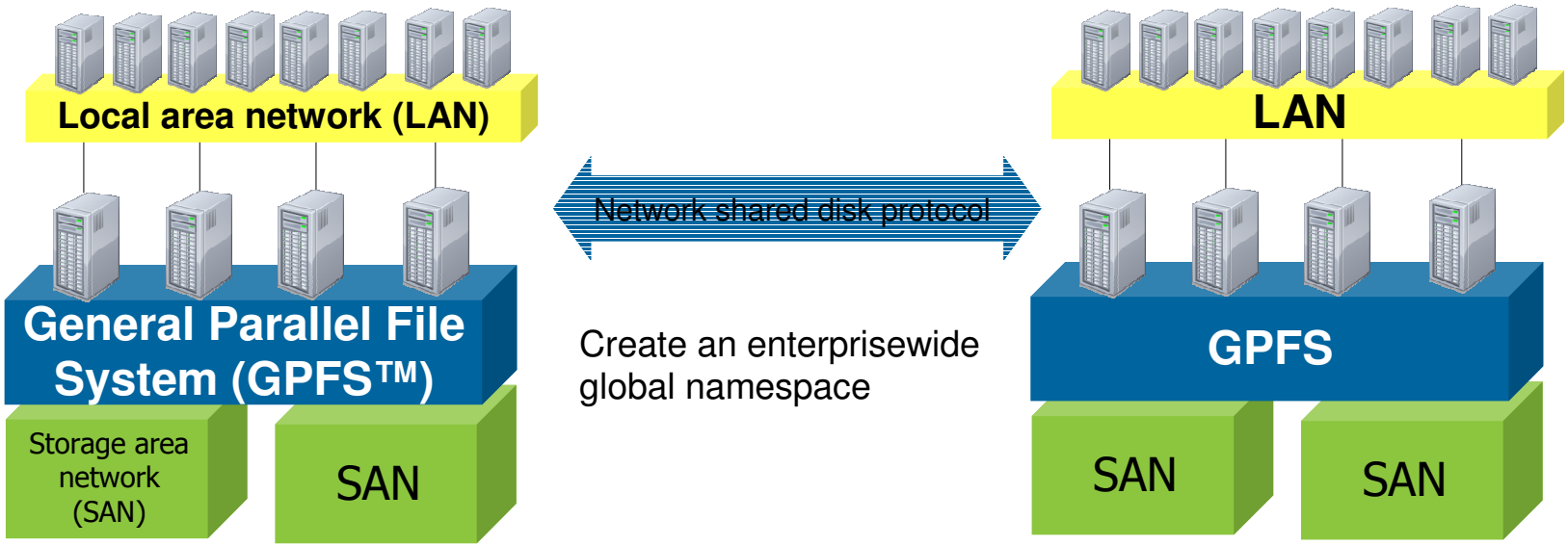
- A highly available cluster architecture
- Concurrent shared disk access to a single global namespace
- Capabilities for high-performance parallel workloads

IBM General Parallel File System (GPFS™) – History and evolution

GPFS 3.4 introduces improvements in performance, scalability, migration and diagnostics and enhanced Windows™ high performance computing (HPC) server support, including support for homogenous Windows clusters.



Enhance organizationwide collaboration through multiclustering



Why?

- Tie together multiple sets of data into a single namespace
- Allow multiple application groups to share portions or all data
- Help enable security-rich, highly available data sharing that's also high performance

Extreme capacity and scale

General Parallel File System (GPFS™) already is running at data sizes most companies will start supporting five years from now.

File system

- 2^{63} files per file system
- 256 file systems
- Maximum file system size: 2^{99} bytes
- Maximum file size equals file system size
- Production 3 PB file system

Disk input and output:

- IBM AIX® 134 GB/sec
- Linux® 66 GB/sec

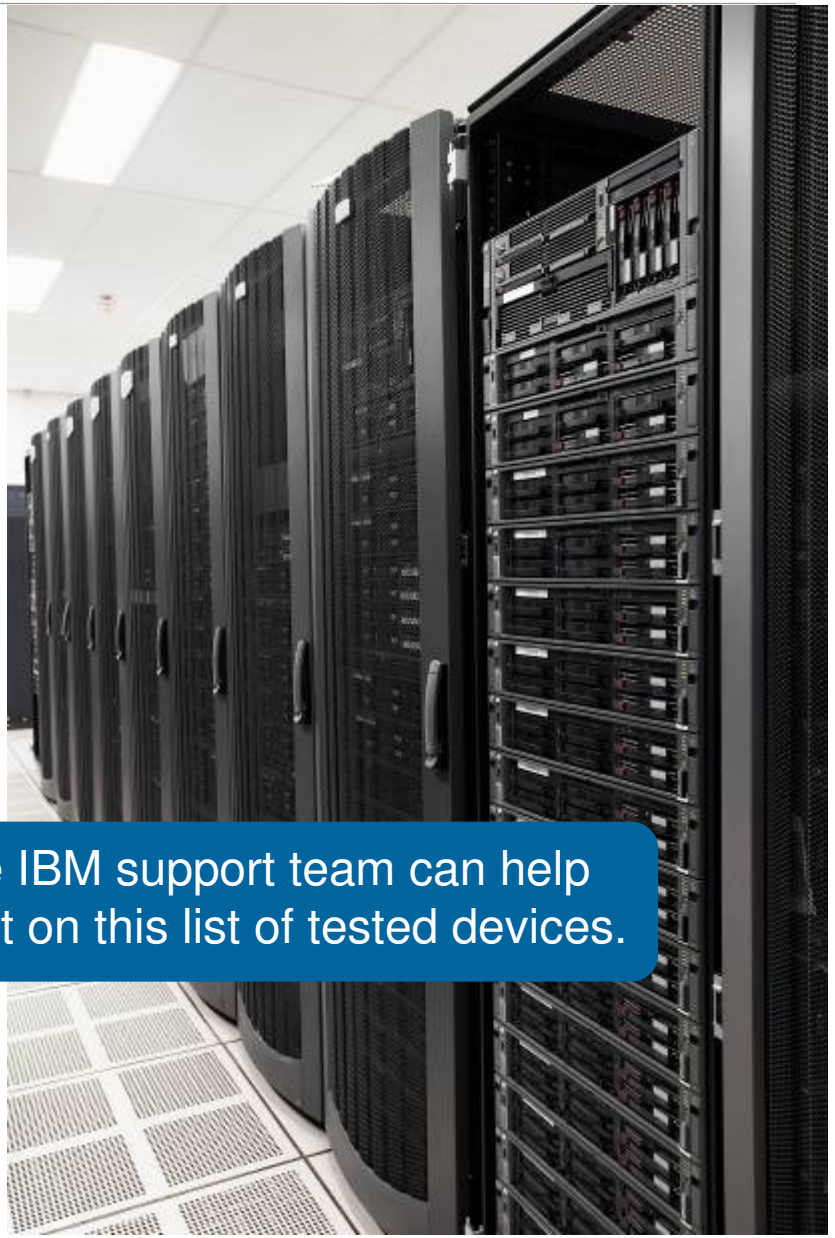
Number of nodes:

- 1 to 8192

Supported storage hardware

In addition to IBM Storage, IBM General Parallel File System (GPFS™) supports storage hardware from these vendors:

- EMC
- Hitachi
- Hewlett Packard
- DDN



GPFS supports many storage systems, and the IBM support team can help customers using storage hardware solutions not on this list of tested devices.

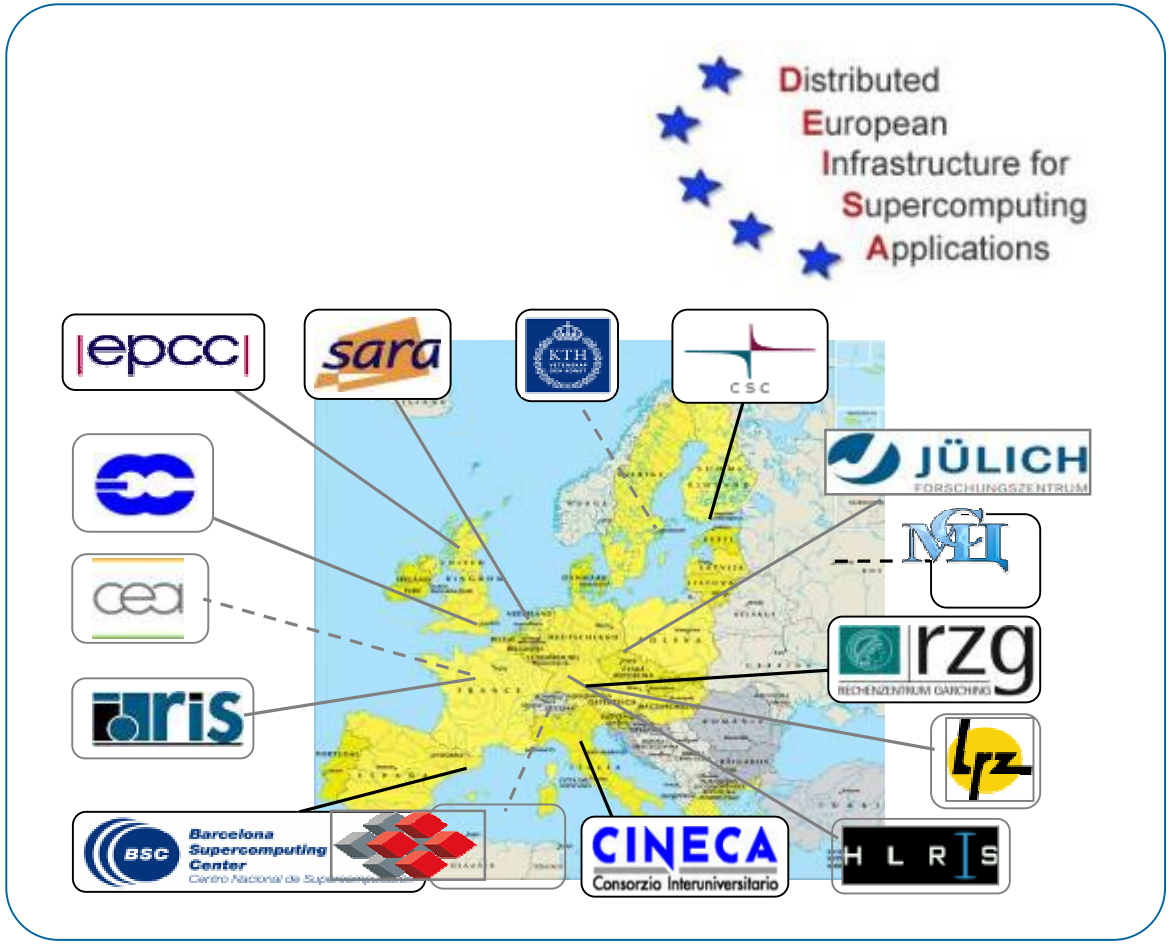
DEISA Partners and Associate Partners

Business need:

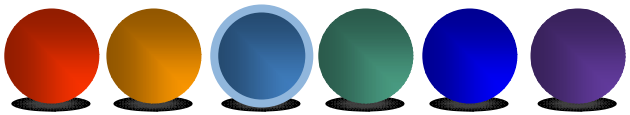
As a research infrastructure comprised of leading nation super computers in Europe, high band-width network connectivity is required to guarantee the high performance of the distributed services, and to avoid performance bottlenecks.

Solution:

A global shared file system based on IBM multicluster
General Parallel File System (GPFS™)
and a dedicated network provided by GEANT2

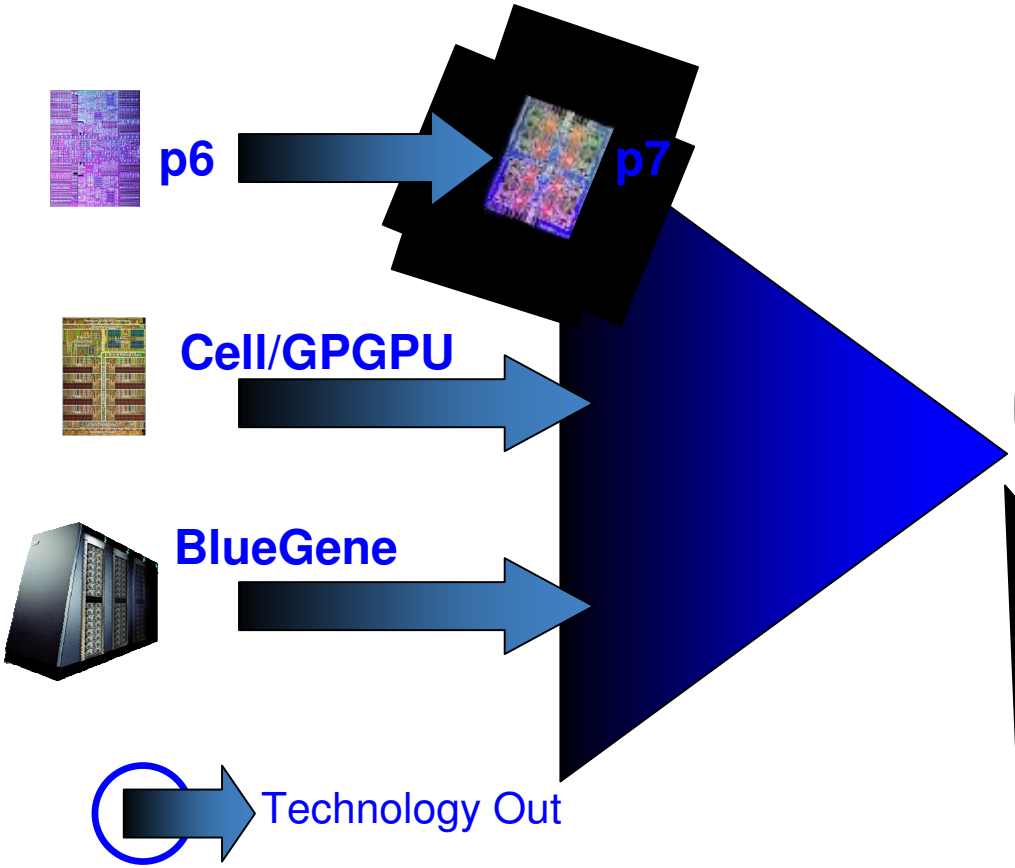


Next Era of Innovation – Hybrid Computing and Cloud



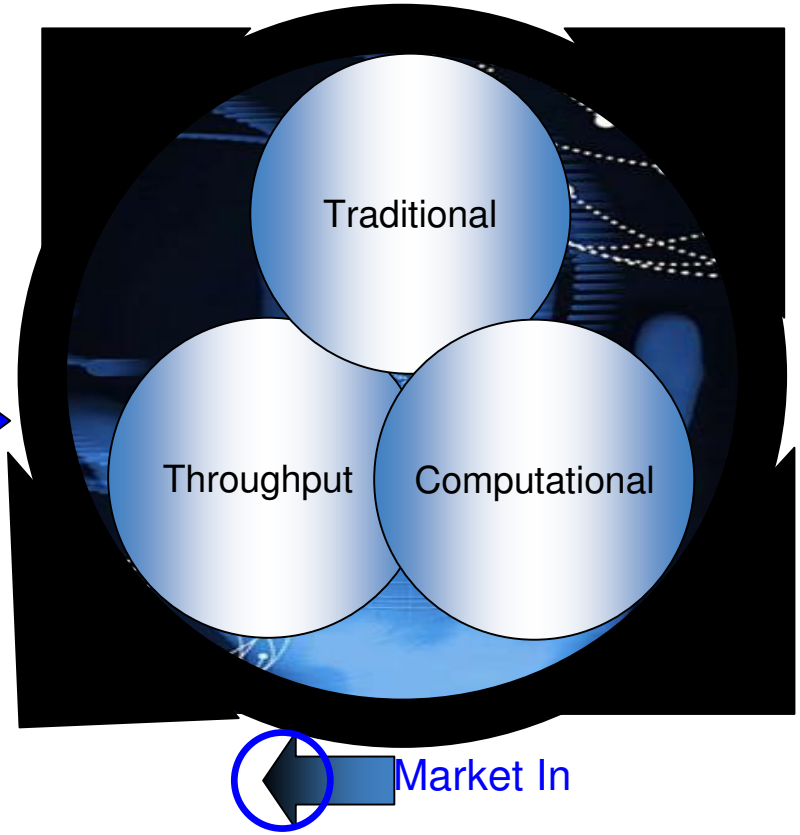
Next Era of Innovation – Hybrid Computing
The Next Bold Step in Innovation & Integration

Symmetric Multiprocessing Era



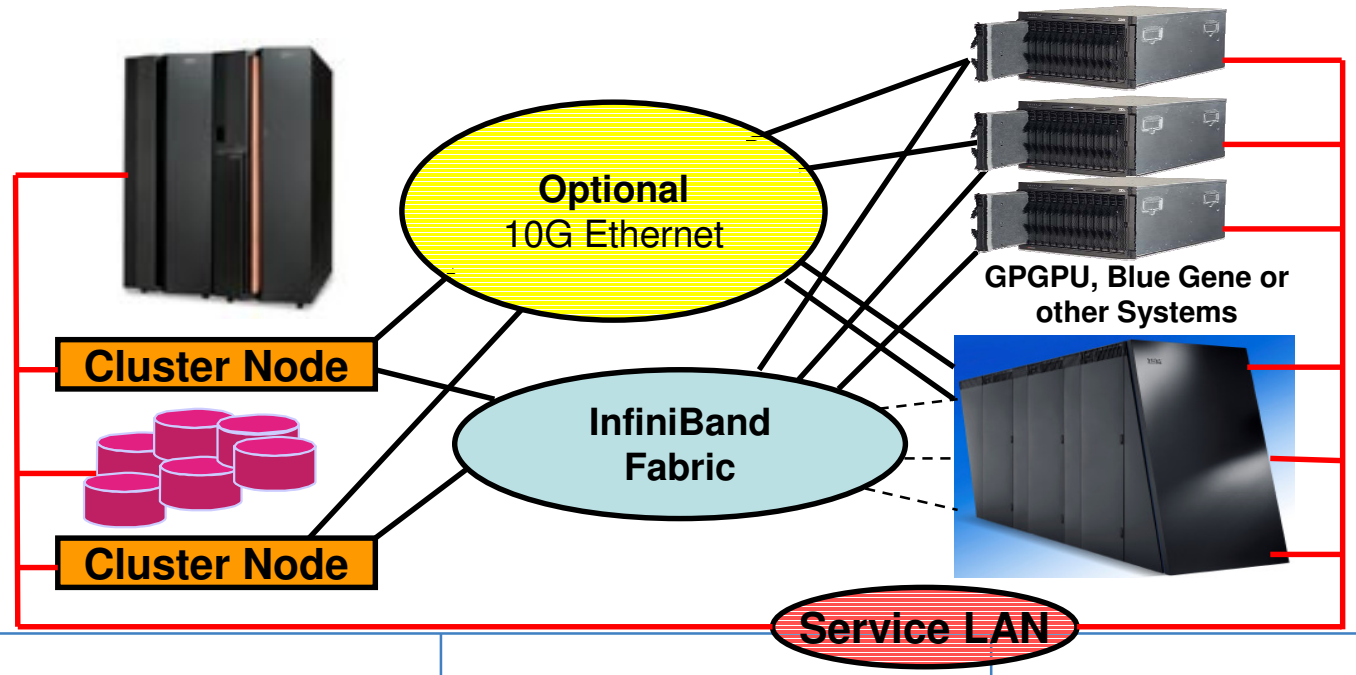
Driven by cores/threads

Hybrid Computing Era



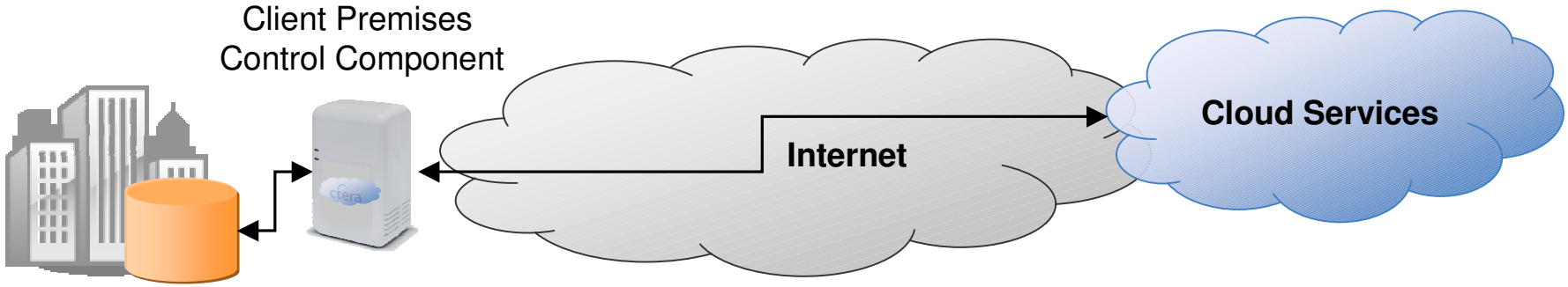
Driven by workload consolidation

Example architectures of system level accelerators



Base System	Interconnect	Accelerator
▪ Standard workloads, etc, Visualization clusters, workstations ▪ Power SMP Intel, AMD Clusters	▪ Fast networking and clever algorithms that maximize network bandwidth ▪ Ethernet / InfiniBand have emerged as standards	▪ Off load / accelerate compute intensive components of workload ▪ Blue Gene x86/GPGPU

Emerging solution: Client Controlled Cloud – separation of control components



Existing Applications & Data

- Component on the premises of the enterprise
- On premises control of sharing and composition of services and sharing of information

Control components

- Clients declare policies for sharing data and services
- Selection and secure composition of cloud services from a variety of providers
- Client specify how and when to get more IaaS or PaaS resources

C3 ensures secure composition of services, thus reducing data security and privacy issues