

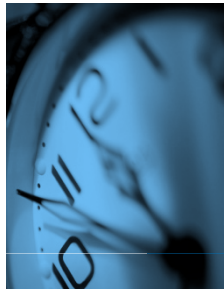
HPC Clusters and Solutions

Alberto Galli
HPC Consultant - Presales Italy
Mobile: +39 335 6322966
alberto.galli@hp.com
September 2010



CHALLENGES TO INNOVATION

CUSTOMER OBJECTIVES



Deeper
insight

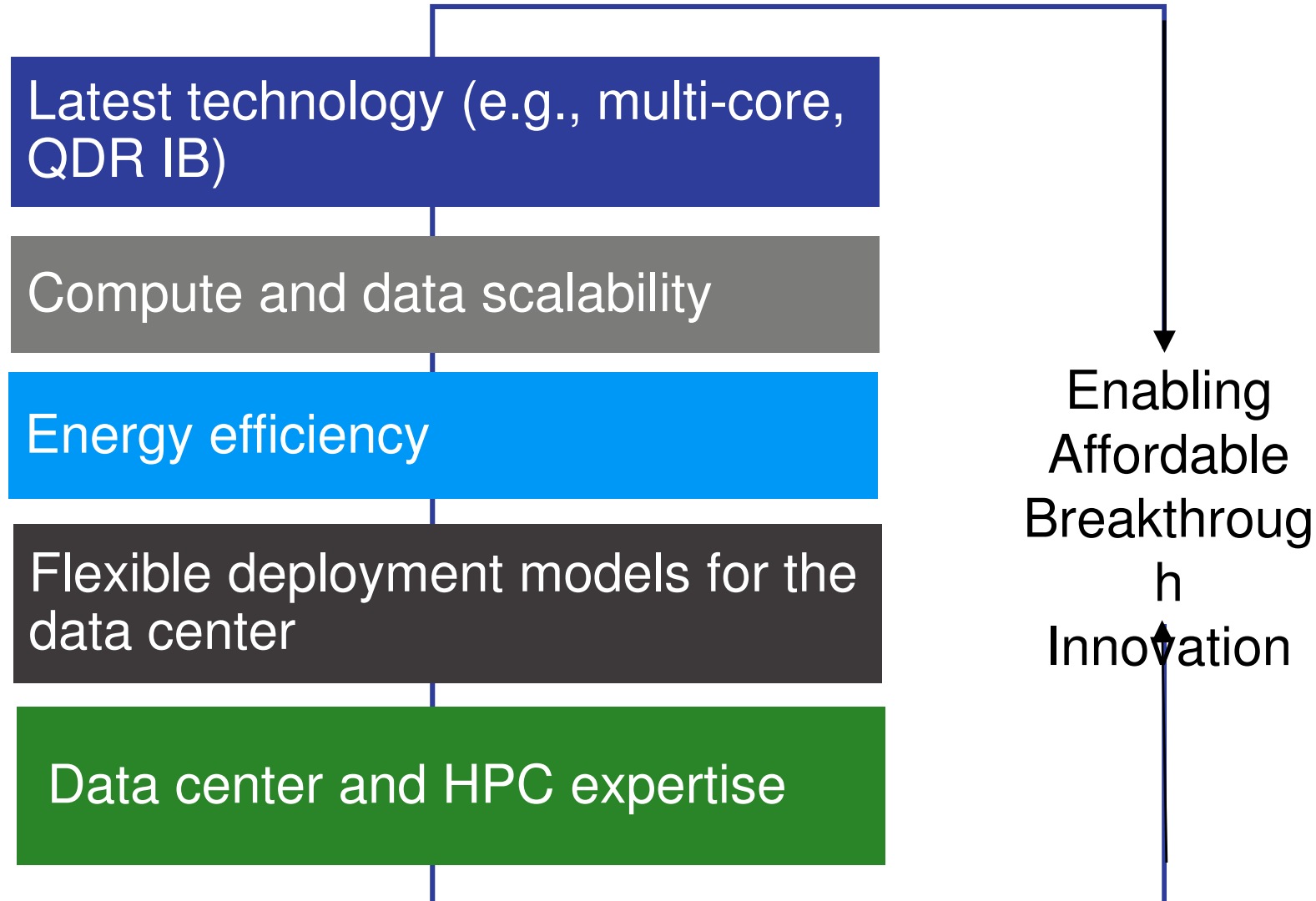
Smarter
decisions

Create
profit

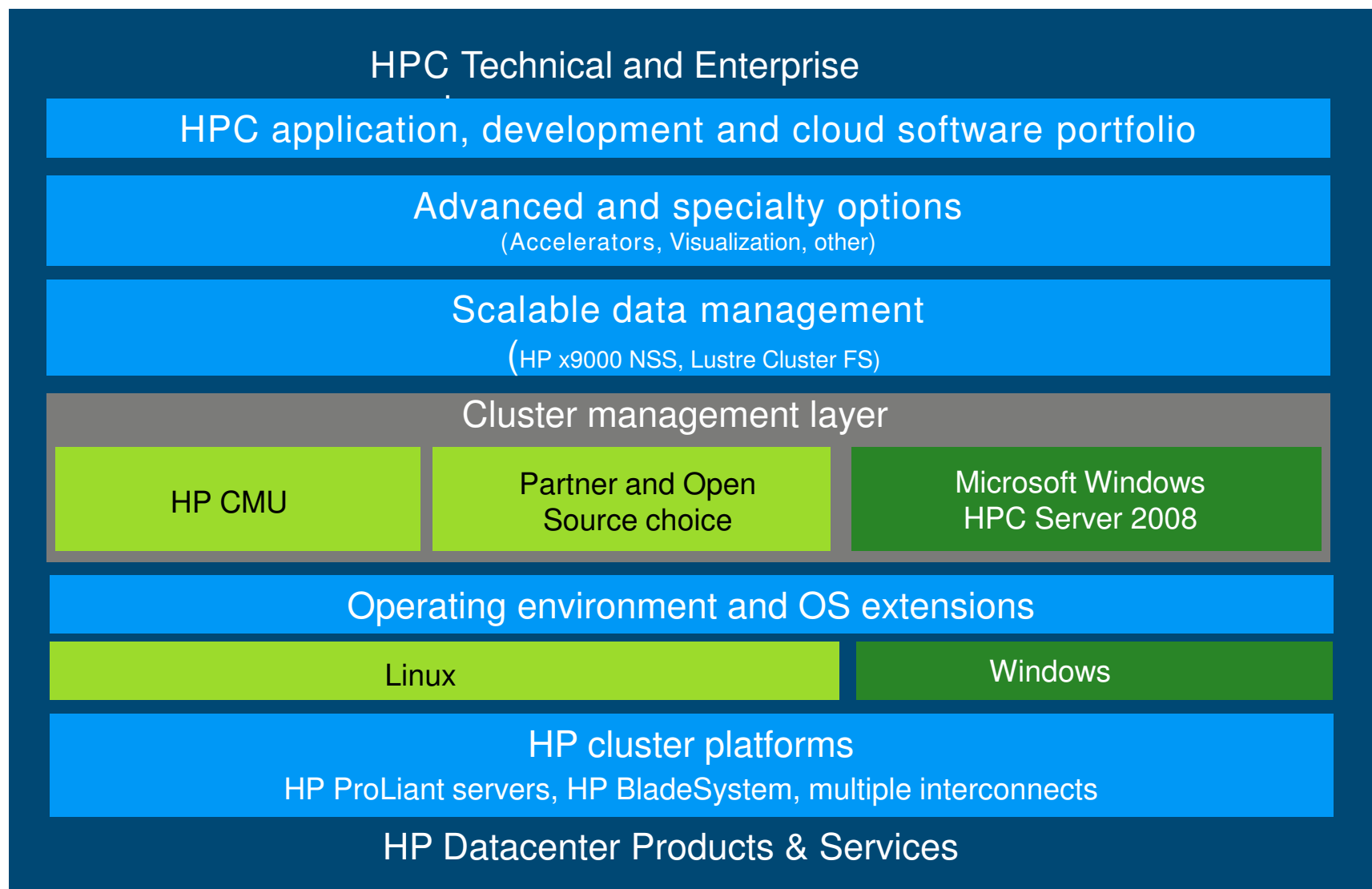
OBSTACLES

- Increasing scale and complexity of deployments and applications
- Exploding volumes of data
- Pressure to provide instantaneous results
- Affordability
- Power and space constraints
- Technology and market risks and uncertainty
- Adapting to new business models and technologies

TODAY'S REQUIREMENTS FOR THE HPC DATA CENTER



HP UNIFIED CLUSTER PORTFOLIO

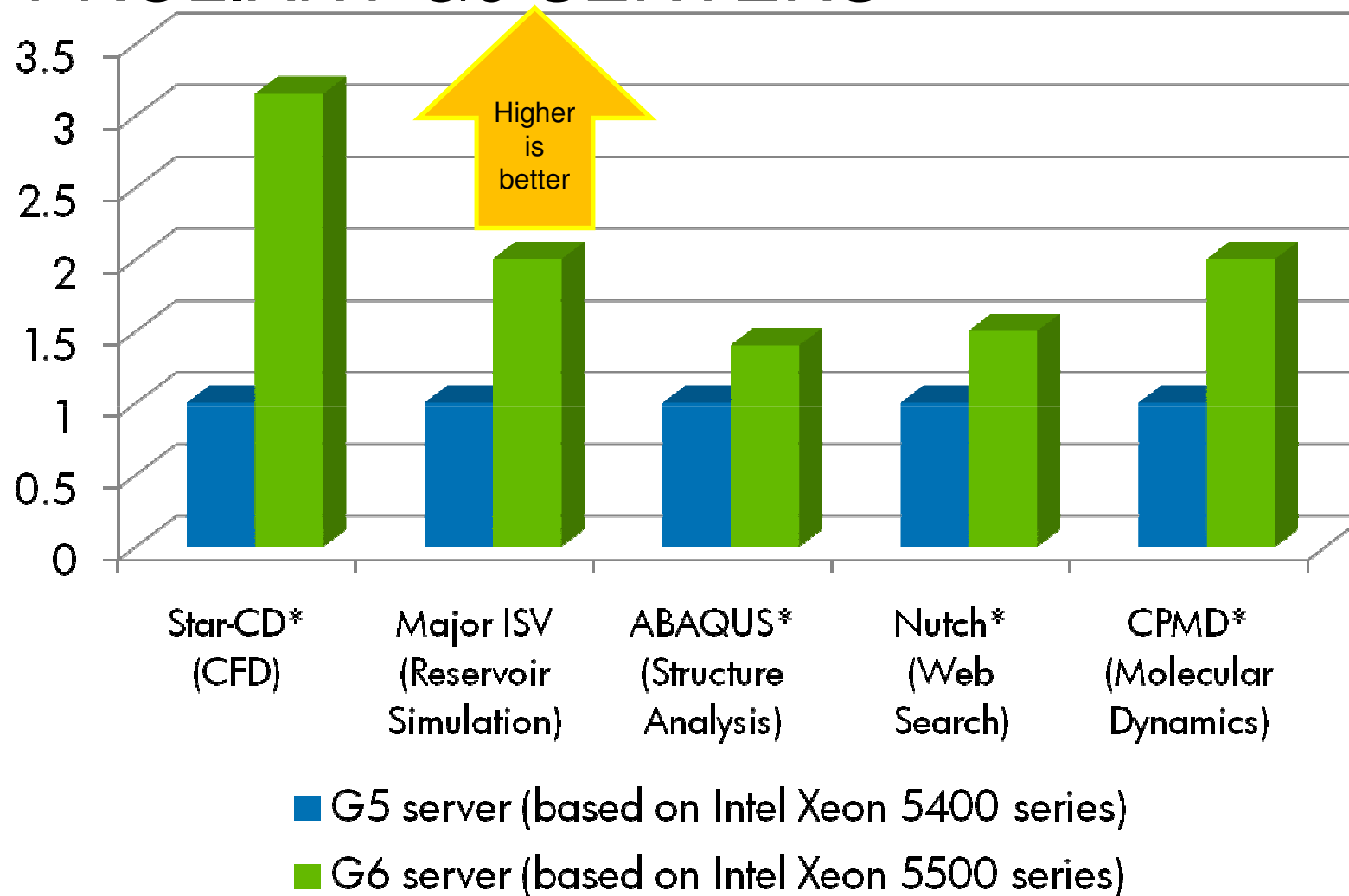


HPC CLUSTER MANAGEMENT OPTIONS FOR CLUSTER PLATFORMS

HP Cluster Management Utility (CMU)* 	Lightweight and low-cost tool for provisioning, managing and monitoring HPC Clusters . Highly scalable.
Platform HPC Workgroup Manager	‘All-in-One’ suite from Platform Computing: Platform Cluster Manager, MPI, and LSF resource management for up to 32 nodes Support directly from Platform.
Open source 	Choices vary by country: In U.S., Platform Cluster Manager (from Platform) and ROCKS+ (from ClusterCorp) are available for HP systems from HP SLMS
Windows HPC Server 2008* 	Cluster solution based on Windows Server 2008, includes management, Windows MPI, job scheduling. Designed to bring the affordable power of clusters to Windows users.

- Available factory-installed with Operating System and HP-MPI

DRAMATIC PERFORMANCE BOOST WITH PROLIANT G6 SERVERS



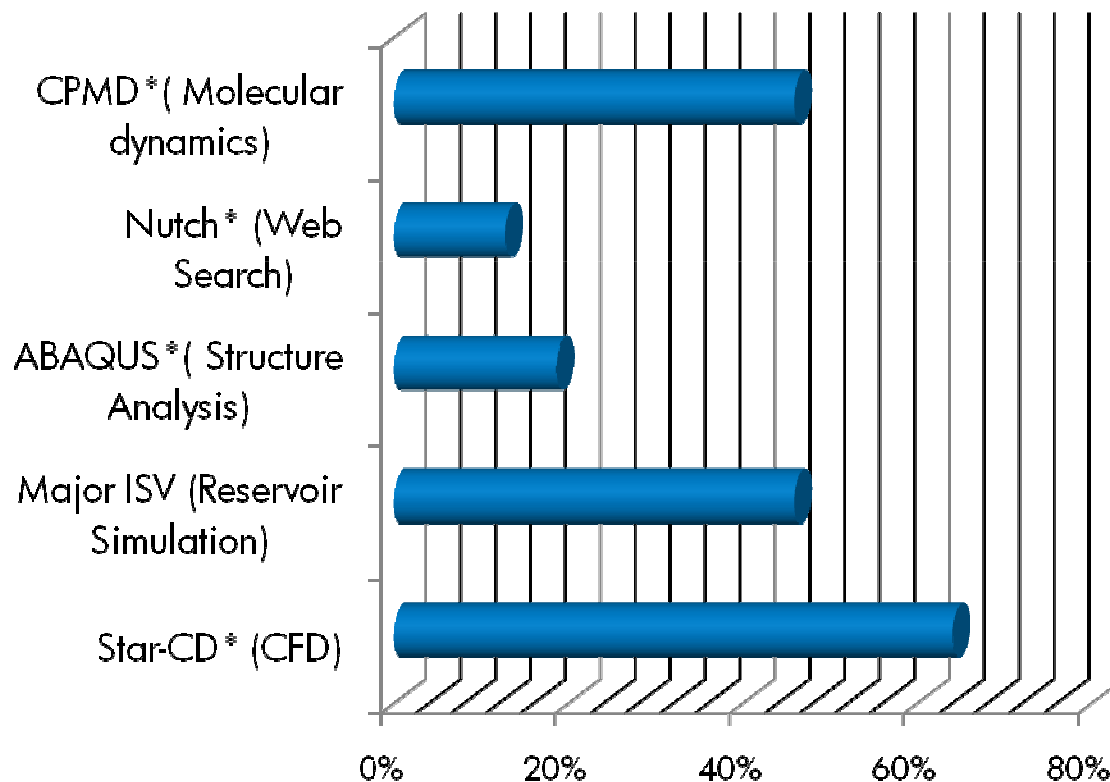
Performances are measured on specific HP server configurations and specific data sets. Any difference in system configurations or data sets may affect actual performance.

* Other names and brands may be claimed as the property of others.



OPERATION SAVINGS WITH PROLIANT G6 SERVERS

Power Saving over G5 for the same amount of work**



• For the same amount of work, operating savings from

- less power consumed
- less floor space required
- fewer software licenses

• fewer servers to manager

** Power savings are estimated based on BladeSystem Power Calculator.

* Other names and brands may be claimed as the property of others.



INTERCONNECT SPECTRUM FOR SCALE-OUT SOLUTION

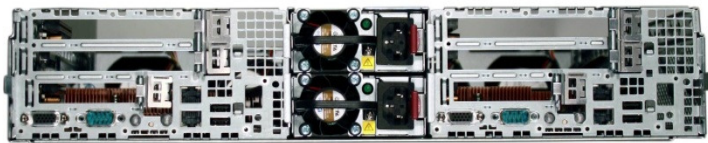
- Gigabit Ethernet dominates in total volume
 - Pervasive mature technology
 - Most cost effective for workloads not interconnect latency and bandwidth bound
- 10GE being increasingly deployed for scale-out solutions
 - Decreasing cost: 10G LOM, switch cost, etc
 - Performance improvements and acceleration technologies (TOE, RNIC, RoEE?)
 - Most for higher bandwidth requirement of network traffics
- InfiniBand remains the choice for HPC clusters
 - Extremely low latency, high bandwidth, scale to 1000s nodes
 - Most for MPI-based applications, low latency market data systems, scalable data warehouse applications



HP PROLIANT SERVERS OPTIMIZED FOR EXTREME SCALE

— DL1000

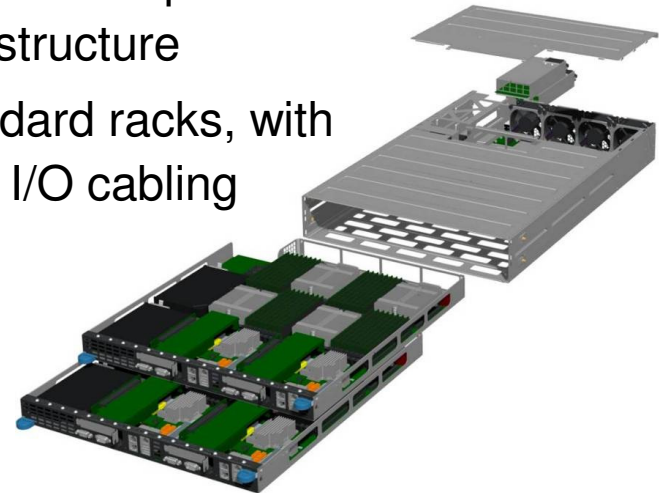
- Designed for scale-out from SMB to extreme scale-out
- Up to 4 nodes in a high efficiency 2U chassis
- Non-plugable servers, with front hot-plug hard drives
- Standard racks, with traditional rear cabling



HP ProLiant SL1000 Scalable System

— SL6000

- Designed for extreme scale-out deployments by the rack
- High efficiency, modular power and cooling chassis
- Front serviceable server/storage trays de-coupled from the infrastructure
- Standard racks, with front I/O cabling



HP ProLiant SL6000 Scalable System

Highly Flexible s6500 Chassis

Multi-node, Shared Power & Cooling Architecture

Benefits: Low cost, high efficiency chassis



- Shared Power & Fans
- Optional Hot-Plug Redundant PSU
- Energy efficient Hot Plug fans
- Single or 3 Phase Load Balancing
- 94% Platinum Common Slot Power Supplies
- N +1 Capable Power Supplies (up to 4)

- 4U Chassis for deployment flexibility
- Standard 19" racks, with front I/O cabling
- Unrestricted airflow (no mid-plane or I/O connectors)
- Reduced weight
- Individually Serviceable Nodes
- Variety of optimized Node Modules
- 1200mm rack to close doors

- SL Advanced Power Manager Support
 - Power Monitoring
 - Node Level Power Off/On



HP ProLiant SL390s Fast Fabric Compute

1/2 width, 1U tall, 2 hot plug nodes in 1U
Tray

2 LFF (3.5") or 4
SFF (2.5") Quick
Release HDD

Cx2 LOM
Dual GbE
Dedicated iLO3 port

PCIe Gen2 x8
LP Slot

VGA

Serial (RJ45)

Power
Button

SFP+

QSFP
(QDR IB)

Dual 1GbE

Health LED

Dedicated management iLO3
LAN & 2 USB ports

UID LED & Button

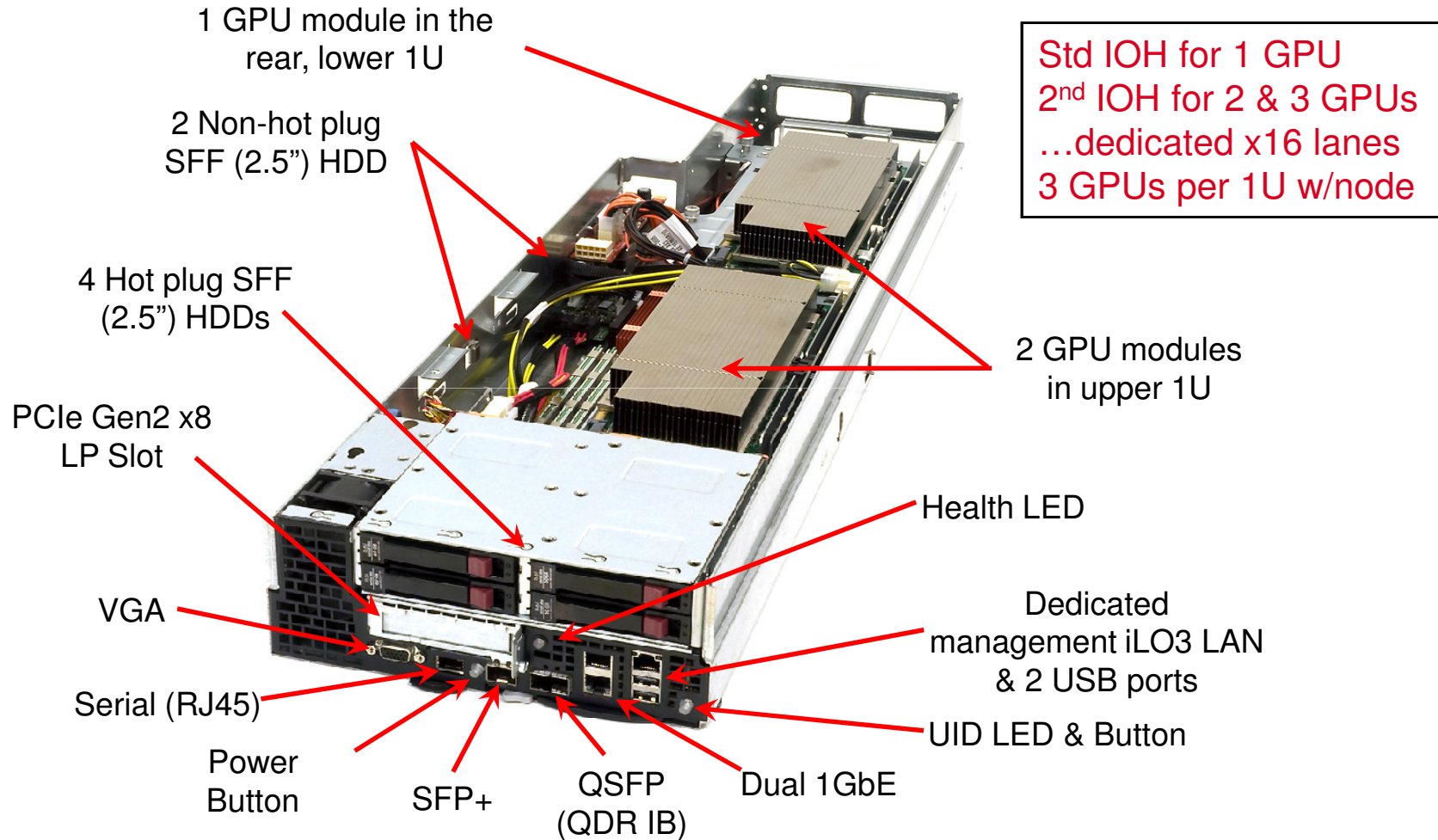
QDR IB or 10GbE embedded, 2 nodes per 1U

©2009 HP



HP ProLiant SL390s, 0 to 3 GPUs

1/2 width, 2U tall (effective density=1U); 4 trays per 4U chassis



Dedicated x16 lanes to each GPU

WHY LOOK AT ACCELERATORS?

- Facing computing challenges
- Need to innovate to meet computing needs
- Need to increase performance
 - While improving power, cooling, and space usage
 - While improving price/performance
- Accelerators offer an alternative to multi-core for meeting performance needs
- Accelerators add heterogeneous co-processors to standard systems to improve performance
 - More silicon doing what you need



PROS AND CONS OF HPC ACCELERATORS

Pros

- Hundreds of functional units executing in parallel
- Speedup applications by 2x, 10x, 30x, 100x or more!
- Really fast
- Excellent for a growing number of highly parallel applications
 - When it works, it can really fly!

Cons

- Can be hard to program
 - But getting easier to program every quarter
- Not useful for many general purpose applications

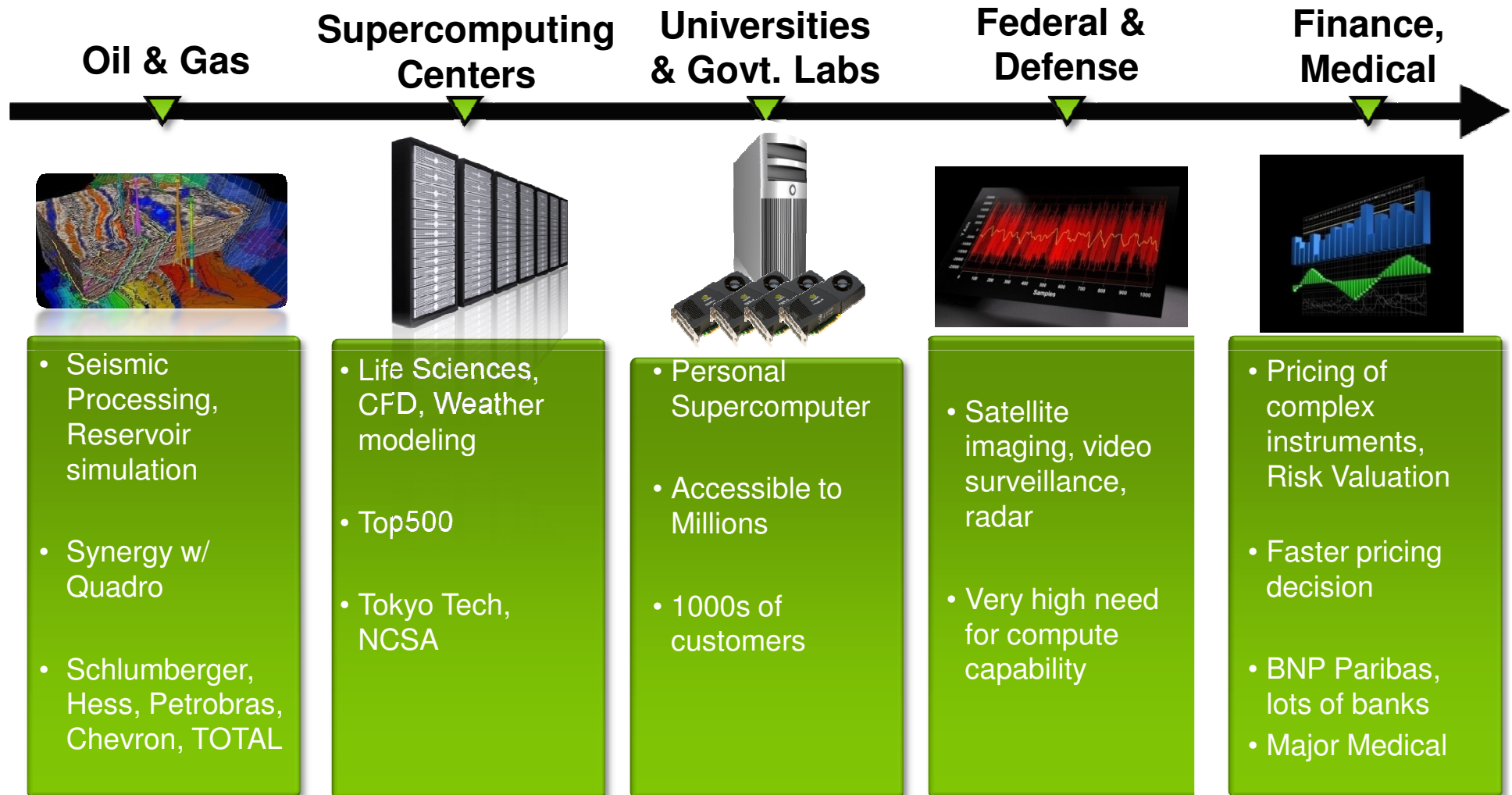


HP SCI ACCELERATOR PROGRAM

- HW Accelerators provide substantial speed-up for certain applications
 - GPUs, FPGAs, custom processors
- HP Accelerator Program
 - partner with accelerator vendors
 - enable accelerators in ProLiant platforms
 - PCIe x16 slots, power, BIOS
 - qualify selected accelerators
 - Nvidia and AMD/ATI GPUs, Nallatech & XDI FPGAs, ClearSpeed and Mercury Cell custom processors
 - deliver integrated solutions
 - ProLiant DL160seG6 (1U 2P Xeon) + Nvidia Tesla S1070 (1U 4 GPUs)
 - Configure DL160seG6 with 2, 4, or 6 GPUs



WHERE WE SEE ACCELERATORS



MULTI-CORE AND APPLICATIONS: THE OPPORTUNITY AND THE RISK

- Opportunity - more raw compute power
 - Multi-core processors can deliver significant performance benefits and relieve compute cluster sprawl.
 - Each core contains its own set of execution resources
 - Results in very low latency parallel execution of application threads within a single physical CPU package.
- The risk - it's a lot more complicated
 - An example:
 - It's possible for one memory-intensive job to saturate the shared memory bus resulting in degraded performance for the system.
 - As the number of cores per processor and the number of threaded applications increase, the performance of more and more applications will be limited by the processor's memory bandwidth, I/O etc



HP MULTI-CORE OPTIMIZATION PROGRAM



- It is a vehicle to identify, scope and then address the issues that Industry Standard Multi-core Systems present to our customers, partners and the industry.
- Best of breed program drawing resources from:
 - Across HP including HP labs
 - Our Technology partners The intent of the Multi-core program is to
- Increased visibility and effective usage of Multi-core high performance systems
- Maximizing of total application/job throughput in multi-core systems
- Maximizing application performance in Multicore systems, using
 - Traditional parallelization and multi thread techniques for compilers and debuggers
 - New and emerging technologies
 - Enhancing legacy HPC math and libraries for mutli-core



HP X9000 ADDRESSES:

Scalable tiers of storage

Example: Health and Life Sciences

Problem

- Massive scalability required at a moments notice
- Data availability during system expansions
- Performance degradation over time
- Infrequently accessed data on high performance disks

Parallel workloads

Example: Media & Entertainment

Problem

- Lots of small files require concurrent access
- Limited bandwidth into storage system
- Controller hardware bottlenecks
- Performance - manageability tradeoff

Disaster recovery

Example: Financial Services

Problem

- Copying very large files concurrently to a remote site
- Cost and complexity
- System performance impacts
- Unused storage at remote location



COMPLEMENTARY SCALABLE STORAGE SOLUTIONS FOR HPC

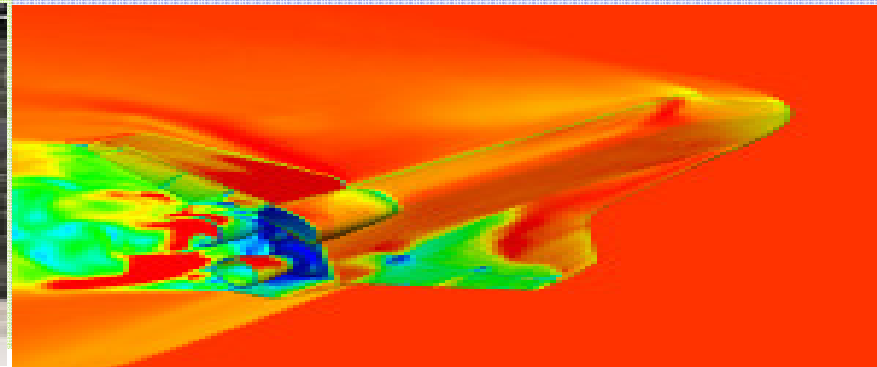
X9000 Network Storage System and Fusion File System

- Shared datacenter multipurpose storage
 - Desktops, clusters, farms, clouds
 - Windows, Linux; CIFS/NFS
- High performance and scale with distributed metadata
- Data tiering
- Continuous replication
- Popular for web 2.0; bioInformatics, FSI, GeoScience apps with many small files



Cluster File System with DirectData Networks

- Tightly coupled HPC storage
 - For large HPC Linux clusters with large single files/single stream requirements (traditional HPC, such as CAE)
- High parallel bandwidth
- High capacity
- High scalability and reliability
- Lustre-open source technology



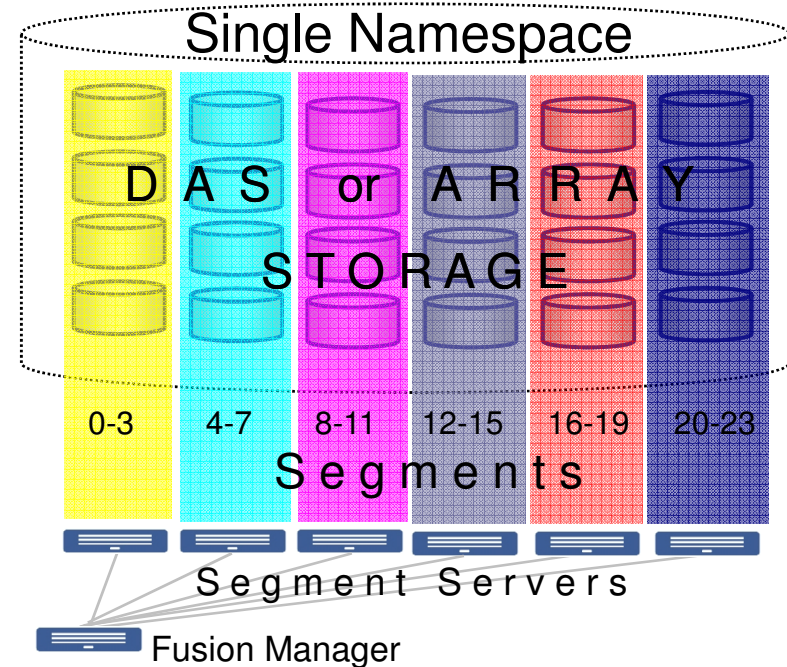
IBRIX Fusion Glossary

Segment: Storage Element

- An atomic storage element in the IBRIX File System
- A logical storage volume
- A file and directory storage bucket
- Segments can be added dynamically

Segment Servers: File Serving to clients, compute nodes and app servers

- Manage elements for specific segments
 - Data allocation and layout
 - Metadata
 - Locks
- Provide file system NAS functions
 - Protocols: NFS, CIFS, IBRIX Client
 - File System and/or Segment Cache
- Provides file system wide coherent cache
- Communicates with other segment servers
 - For metadata information/operations
 - For file information and access (NFS & CIFS)
 - For cluster status and other information



Fusion Manager: Management

- Cluster management system
- Web based and CLI interface
- Allocates segments to servers
- Manages faults & segment servers
- Provides logging and notification

SCALABLE VISUALIZATION FOR HPC

Remote visualization

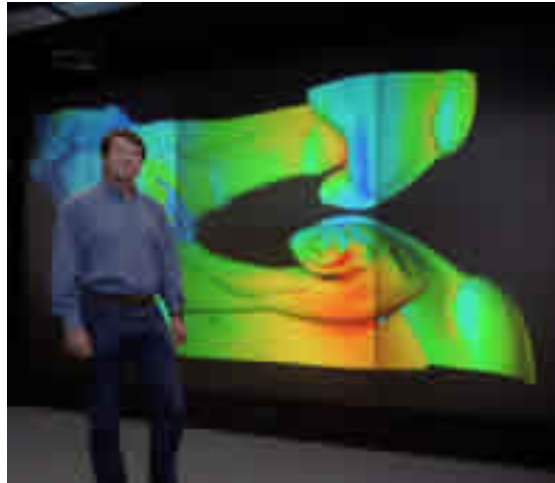
Remote access to graphic enabled servers over a standard network
Simplify management and share resources

Scale-up beyond the workstation

Process large models
Drive display to multi-tile panels and caves

Simplify HPC visualization

VizStack manages surFace displays
Supports launch of parallel visualization applications



Thank you

