

HIGH PERFORMANCE COMPUTING INTRO (SHORT)

Sorin Cristian CHERAN
HPC Competency Center Grenoble

©2010 Hewlett-Packard Development Company, L.P.
The information contained herein is subject to change without notice





High Performance Computing

Introduction

HPC @ HP

Changing Economics in HPC

Costs are moving: IT, Infrastructure, Energy

1990: IT is 80% of the costs. Infrastructure + Energy is about 20% of the costs.

2001: The cost of the Infrastructure + Energy (I&E) exceeds the cost of the server

2004: Infrastructure costs alone exceed the cost of the server

2008: Energy costs alone exceed the cost of the servers.

2014: Infrastructure and Energy Cost (I&E) will be 75% of the cost.
IT will be only 25%

➔ Optimized Data Centres, reduce costs, avoid over-provisioning

Example: PODs, SL product line

➔ HPC as a service, emerging Cloud HPC

Where can we find High Performance Computing ?

Mechanical Engineering (Virtual Prototyping)



- Structures
- Fluids
- Impact

Life and Materials Sciences



- R & D
- Clinical Trials
- Data Analysis
- Material Sciences

Government / Classified Defense



- Cryptography
- Custom
- Classified
- Weather

Scientific Research



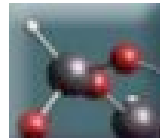
- Custom
- LS R&D

Electronic Design Automation



- Verification
- Layout

Product Lifecycle Management



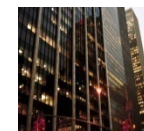
- PDM
- PLM

Oil & Gas Geo Sciences



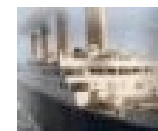
- Seismic
- Reservoir Modeling

Finance/ Securities



- Financial Modeling

High-End Film and Video



- Rendering

HP technology in High End rendering



Weta Digital, a special effects studio in Miramar, New Zealand, was responsible for bringing to life the beautiful and mesmerizing virtual world of the planet Pandora.

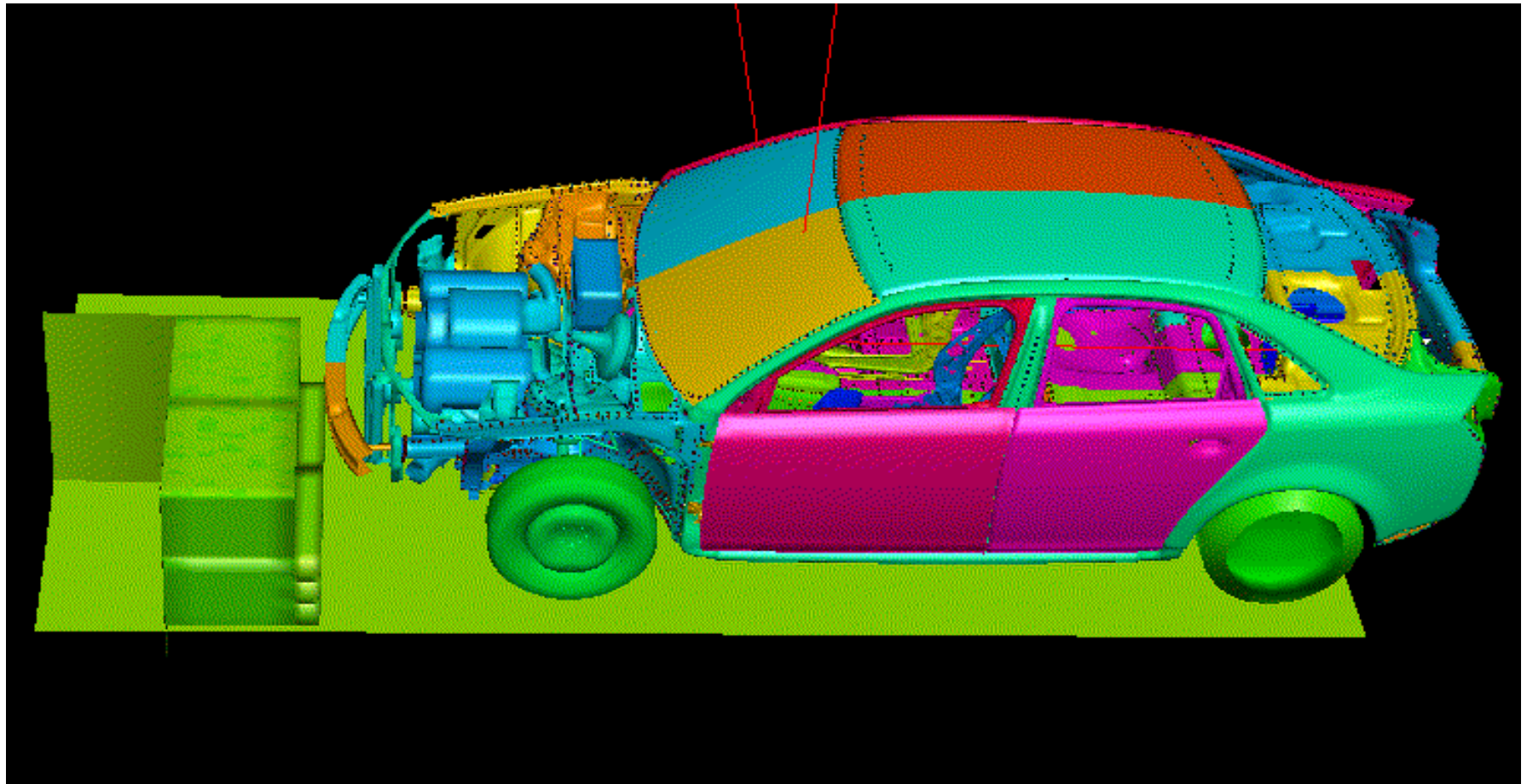
And Weta's computing core ran on 2,176 **HP ProLiant BL2x220c** Blade Servers.
(Weta used thirty four racks comprise the computing core, with 40,000 processors and 104 terabytes of memory.

For the last month or more of production those 40,000 processors were handling 7 or 8 gigabytes of data per second, running 24 hours a day.
A final copy of Avatar equated to 17.28 gigabytes per minute of storage.

full story : <http://hpnow.corp.hp.com/news/10q1/100111fs2.htm>

HPC : Automotive Example

	10 years ago	6 years ago	4 years ago	Today	2 years from now
Time from Concept to Manufacturing	60	36	24	8-12	6-8



1979: CFD Contributions to the 767

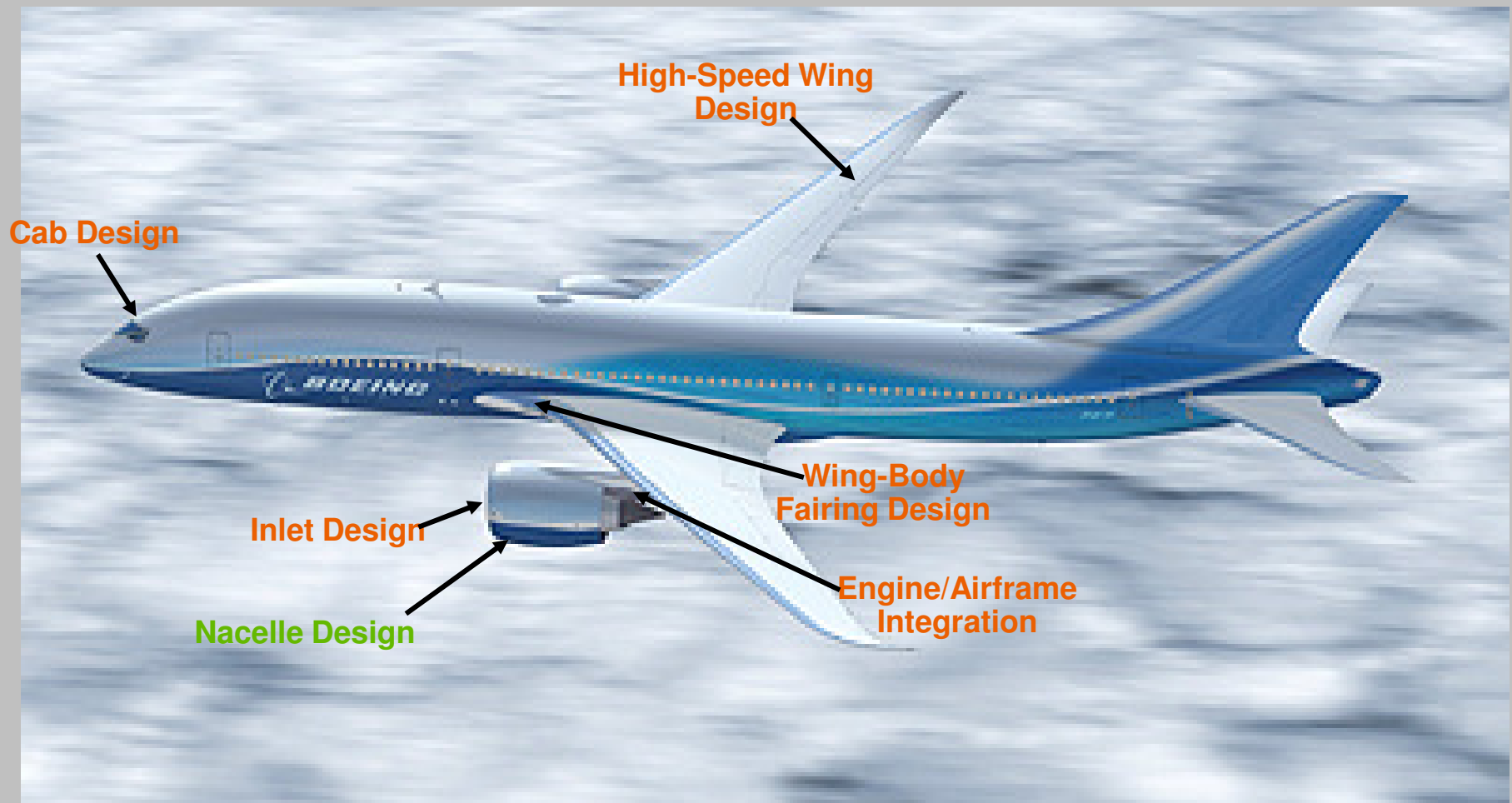
■ Much CFD penetration

Opportunities exist for higher accuracy and expanded complexity

■ Some CFD penetration

Opportunities exist for large increases in design process speed and application

■ CFD penetration opportunity



2005: CFD Contributions to the 787

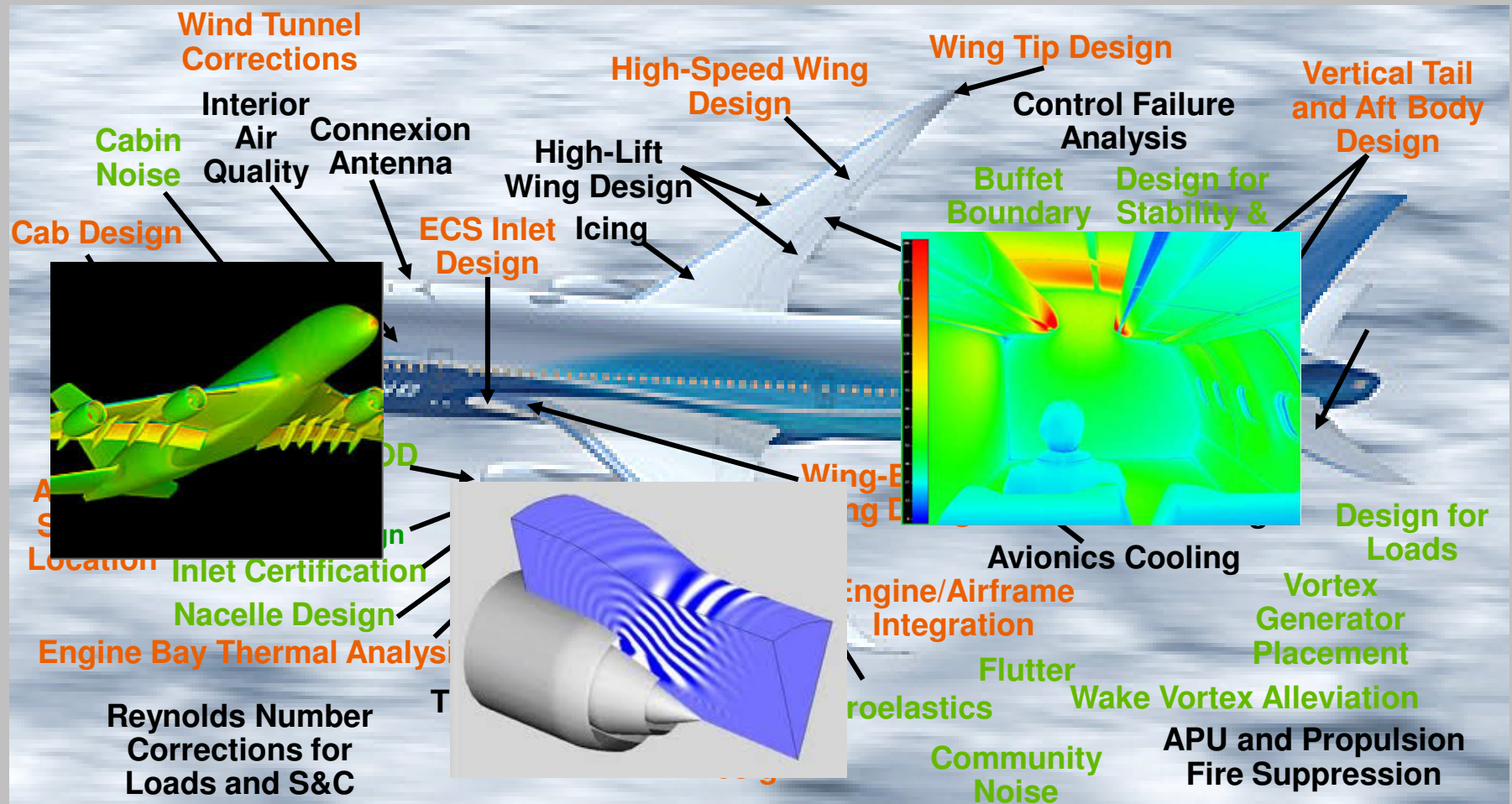
■ Much CFD penetration

Opportunities exist for higher accuracy and expanded complexity

■ Some CFD penetration

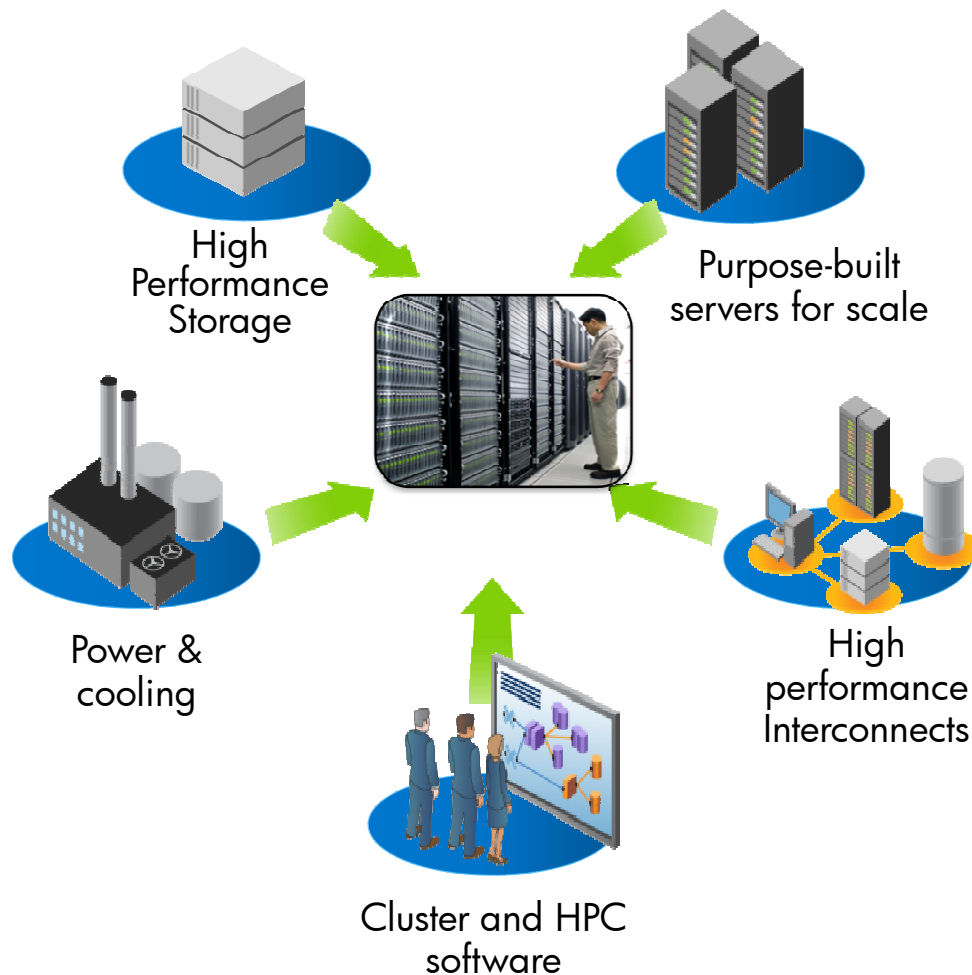
Opportunities exist for large increases in design process speed and application

■ CFD penetration opportunity



What do we need for HPC?

A converged infrastructure for HPC



Enabling breakthrough innovation

- Any application, anywhere
- Powerful and robust solutions
- Compelling economics
- Faster time to impact

Which platform for HPC Applications ?

- quad core ? Eight core ? many cores ?
- Address space, data set size
- Is application cache sensitive
- Memory/latency bandwidth
- NUMA aware (local/remote memory, cpu affinity)
- Integer / floating point units
- On which OS is the application available
- Serial or parallel version for the application



<http://techmktg.atl.hp.com>

HPC Applications

- Serial : single thread



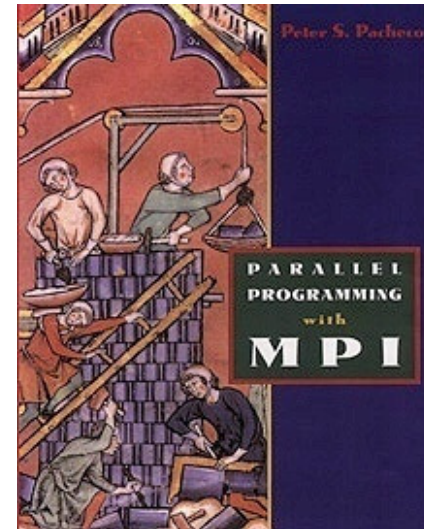
- Parallel

- **OpenMP (Open Multi Processing)**

- *multi-threaded, shared memory parallelism, X threads on 1 node*
- implicit programming
- <http://www.openmp.org/>

- **MPI (Message Passing Interface)**

- *targeted for distributed memory systems, X processes on Y nodes*
- Explicit programming
- <http://www.mpi-forum.org/>



Tomorrow...

- Cloud computing becomes real : requires new delivery models
- There will be major new entrants at the low end
 - Microsoft and desk-side supercomputers will drive growth
- Processors industry is entering a new phase in which CPU improvements are driven by the addition of multiple cores on a single chip, rather than higher frequencies.

AND

Multicore → Manycore?

- GPU offers a massively parallel environment



- GPUs address :
 - HPC endless thirst for SPEED and PERFORMANCE
 - At lower power and cooling costs
- GPU offers a high (potential) peak performance
 - 665 GFlops DP for FERMI (1 TF for Kepler) compare to 140 GFlops of dual socket X5670 @ 2.93 GHz or 370GFlops for a SB at 2.9



Step 1 :

How to build a HPC cluster ?

What is a HPC cluster ?

HPC clusters are scalable systems based on

Commodity hardware

Private system network

(install, monitoring, job submission)

a high speed network

(application communication)

Software infrastructure

(open source or proprietary)

HPC Clusters are not :

General purpose clusters

Highly available clusters

Do not run Mission Critical Applications

2 types of Nodes in HPC clusters

Admin node(s)

(head node, service node, control node, utility node)

➔ Central administration point

➔ nodes may have roles (login, IO...)

Compute nodes

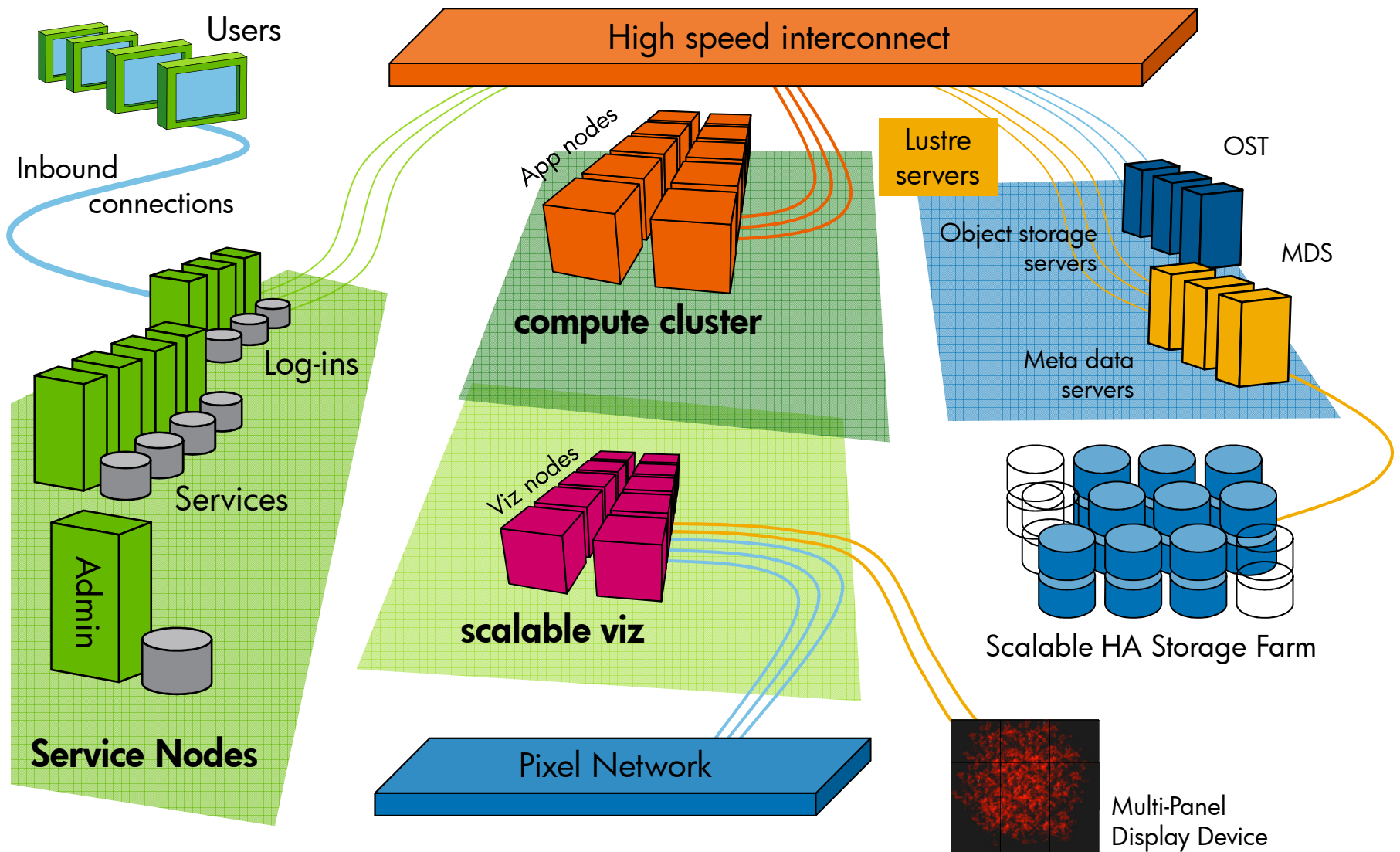
(application nodes)

➔ dedicated to user applications

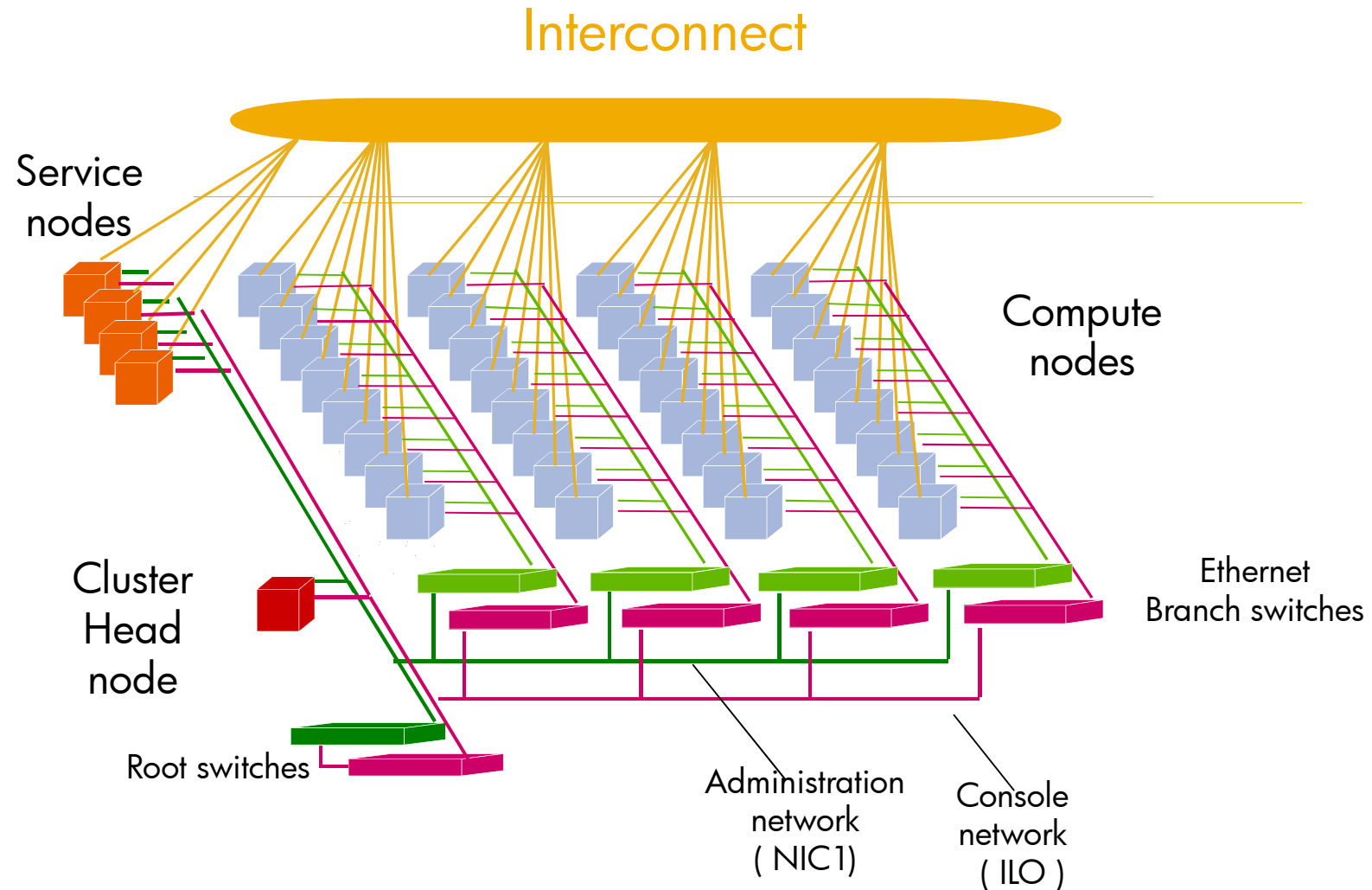
4 types of Networks in HPC clusters

- 1) Console Network
- 2) Admin Network
- 3) Interconnect (Application Network)
- 4) External Network

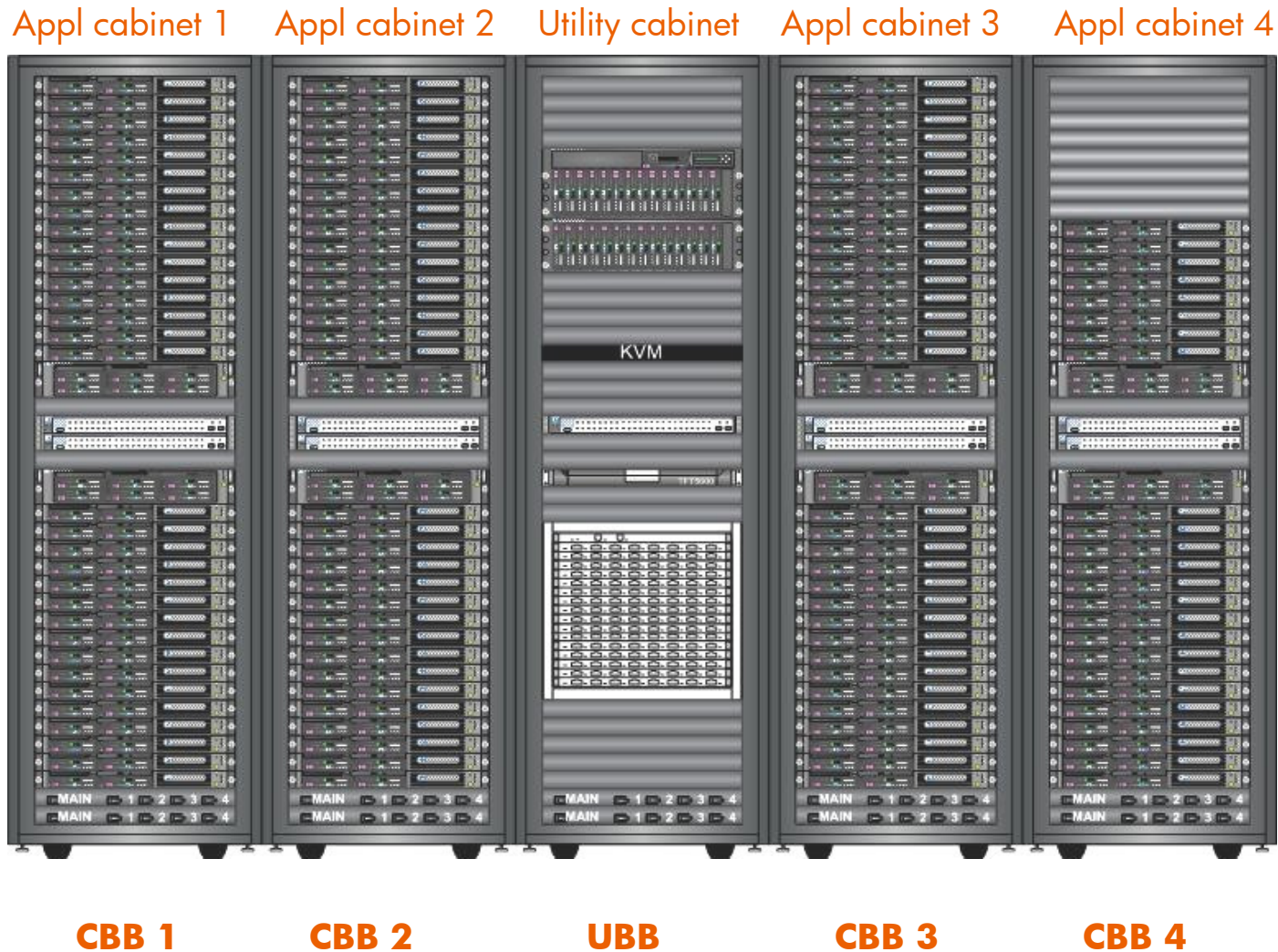
Delivering the vision



HPC Cluster : high level view



Hardware Layout: Cabinets



CBB = Compute Building Block - UBB = Utility Building Block

Utility Building Block

Procurve Switches

Voltaire Switch

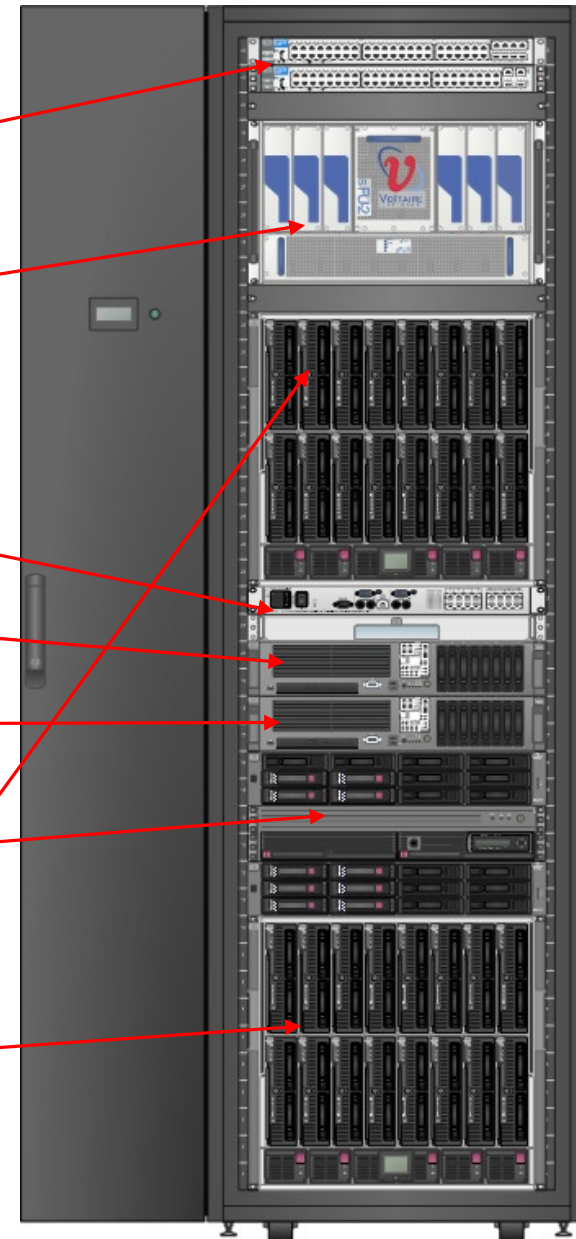
KVM - Display

Storage Node

Control Node

6 TB of Storage

2x C7000 with 16 BL460c





Step 2 :

HP Components to build the cluster

What do you need to build a HPC Cluster ?

- Platforms / processors

- Management node(s) / Compute nodes
- BLs , DLs , SLs
- Management card (ILO)
- PDU

- Networks

- Admin, Mgt
- Application network
 - GigE
 - InfiniBand

- HPC software

1. Protocols
2. Operating systems
3. File Systems
4. Middleware
5. Management software
6. Applications

- Storage

- Local disks
- NFS
- Parallel File Systems
- Enclosures :
 - P series, EVA, MSA, etc

Industry's most complete portfolio for HPC

- Workload optimized, engineered for any demand



**ProLiant DL
Family**

*Versatile, rack-optimized
servers with a balance of
efficiency, performance
and management*



**ProLiant BL
Family**

*Cloud-ready converged
infrastructure engineered to
maximize every hour, watt
and dollar*



**ProLiant SL
Family**

*Purpose built for the
world's most extreme
data centers*

Application/compute network

Low Latency / High Bandwidth network

- **Ethernet 1 gigE, 10 gigE**



Procurve switches

- **Infiniband DDR, QDR, FDR/EDR**

Double Data Rate = 20 Gb/s , Quadruple Data Rate = 40 Gb/s



HP Cluster Interconnect Strategy

- Ethernet
 - Best standards-based choice for cost-driven applications
- InfiniBand
 - Best standards-based choice for performance-driven applications
- Proprietary
 - Alternatives for specific solutions
 - Myrinet from Myricom
 - QsNet from Quadrics
 - Cray Interconnect

Which interconnect to select ?

There is no BEST cluster interconnect

you must investigate :

Type of problem to be solved with the cluster

How the cluster interconnect will perform in that situation

Some criteria :

Latency, Bandwidth

Scalability

Topology Issues

Reliability

Price versus Performance



You need to know your applications

Analyze the ratio

Computation / Communication

Process1 on Node 1

Compute

Process2 on Node 2

Compute

Processes exchange MPI messages



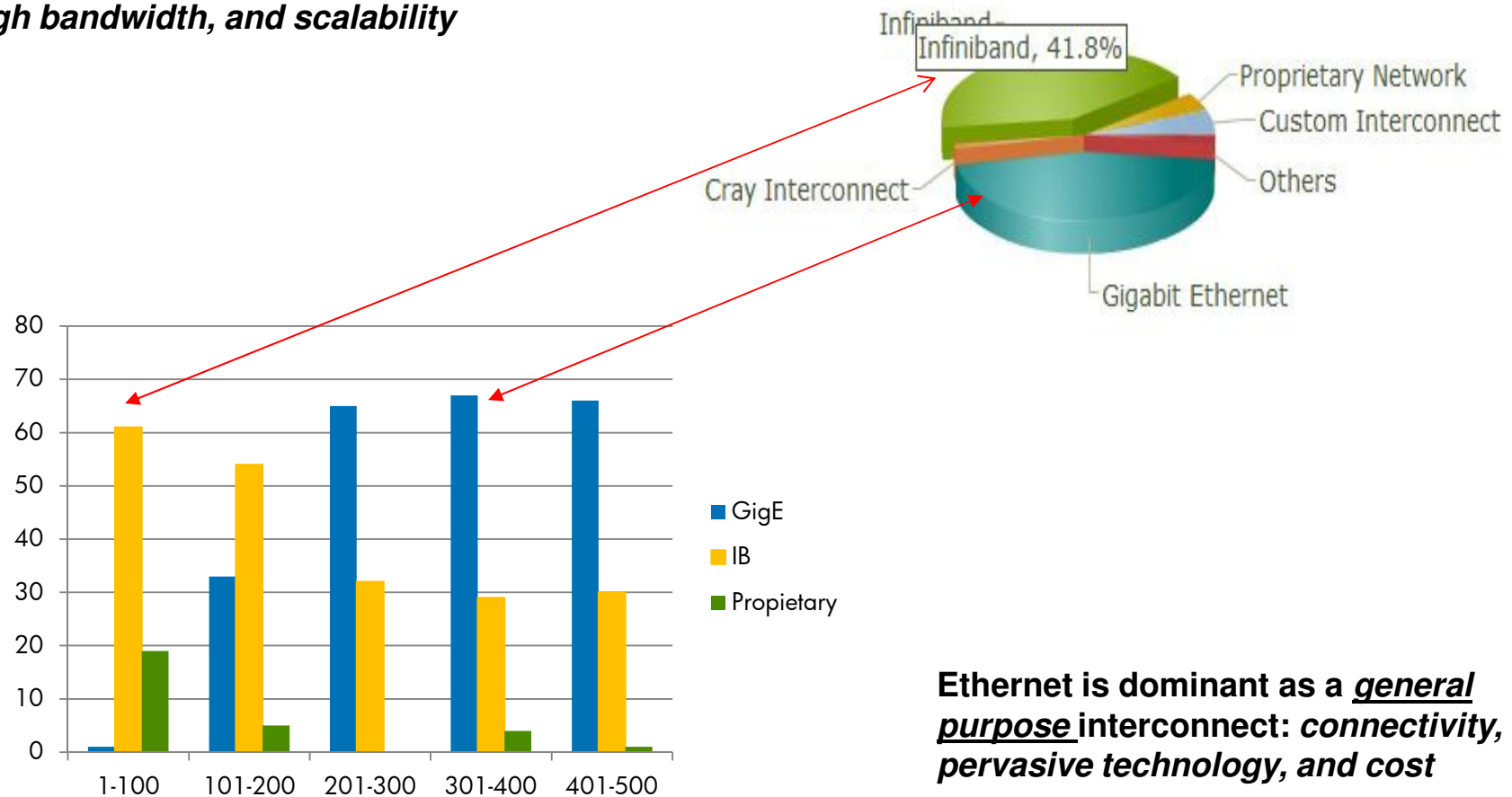
if ratio high, no big communications, 'slow' interconnect may suit

if ratio low, lot of communications, 'fast' interconnect needed

if ratio very low, a cluster does not suit, need a SMP machine

Top500 list breakdown by interconnects

InfiniBand is dominant as a ***high speed*** interconnect: ***low latency, high bandwidth, and scalability***

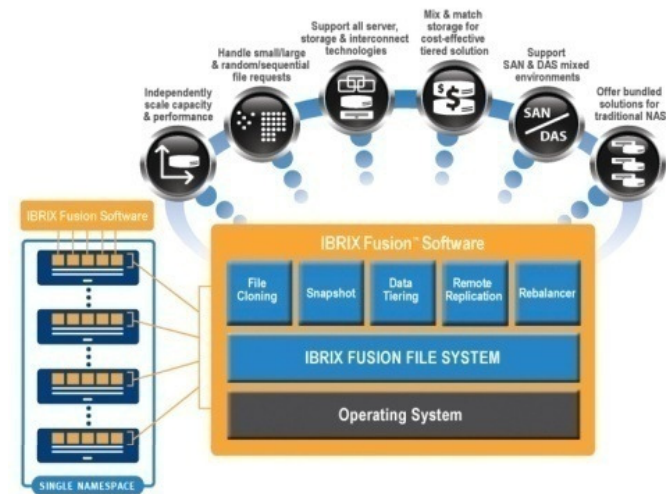


Ethernet is dominant as a ***general purpose*** interconnect: ***connectivity, pervasive technology, and cost***

Source: TOP500 list, November 2011

Mix HP Storage in HPC cluster

- HP X9000 Network Storage System
 - Small files
 - High metadata operation rates
 - Wide access
 - **/home typically...**



- Lustre file system with optimized HPC focused hardware
 - Extreme sequential bandwidth
 - "True" parallel I/O
 - several writers to same file
 - Or high single stream throughput
 - **/scratch, /work typically...**



Software Architecture for HPC Cluster

Operating System

- **Linux** (dominates largely)
- **Windows HPC 2008**

Cluster Management tools

- **HP** : CMU (Linux) or Windows HPC
- **Platform** : Open Cluster Stack from Platform Inc.
- Open Source (ROCKS, OSCAR, ...)

Libraries used by applications

- Communication Message Passing Interface – MPI
- Math libraries (ScalaPack, BLAS, MKL ...), parallel debuggers

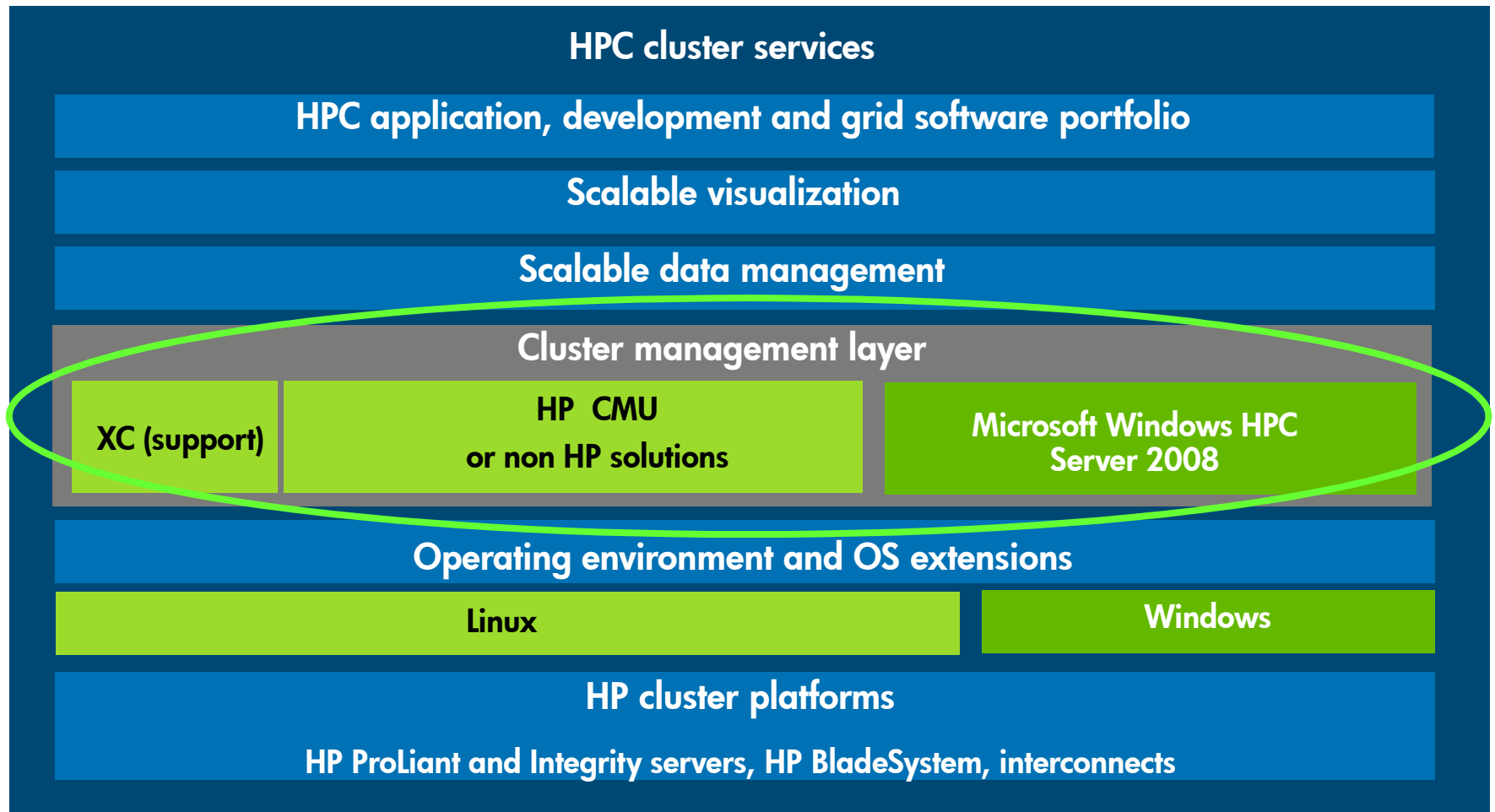
Workload manager

- Torque, Lava
- Slurm
- HPC-LSF
- LSF (Platform), PBSpro (Altair)
- MOAB (Adaptive Computing)

Applications

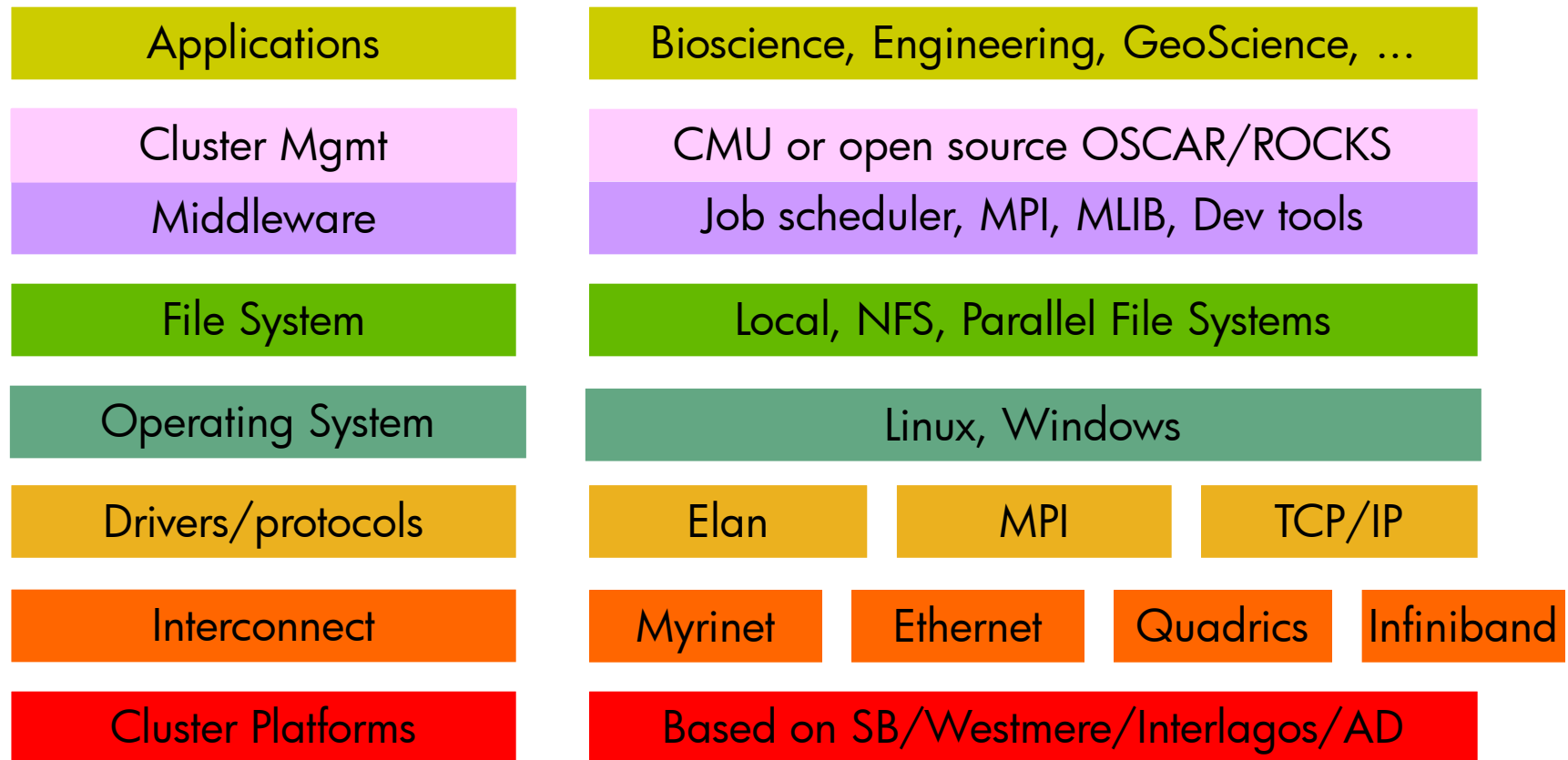
LS-Dyna, PAM Crash, RADIOSS,
Nastran, Abaqus, elsA,
LionBioscience, Gaussian,
and hundreds more

Updated Unified Cluster Portfolio framework



Cluster Solution Design Choices

Top level view of SW and HW components of an HPC cluster



THANK YOU

