ENEA, Roma 3 Lug 2012

Workshop " Il calcolo scientifico ad alte prestazioni – Aggiornamento tecnologico"

HPC clusters Overview

Panoramica sulle attuali tecnologie hardware e software per la realizzazione di cluster HPC

Simone Tinti
HPC team leader
simone.tinti@e4company.com

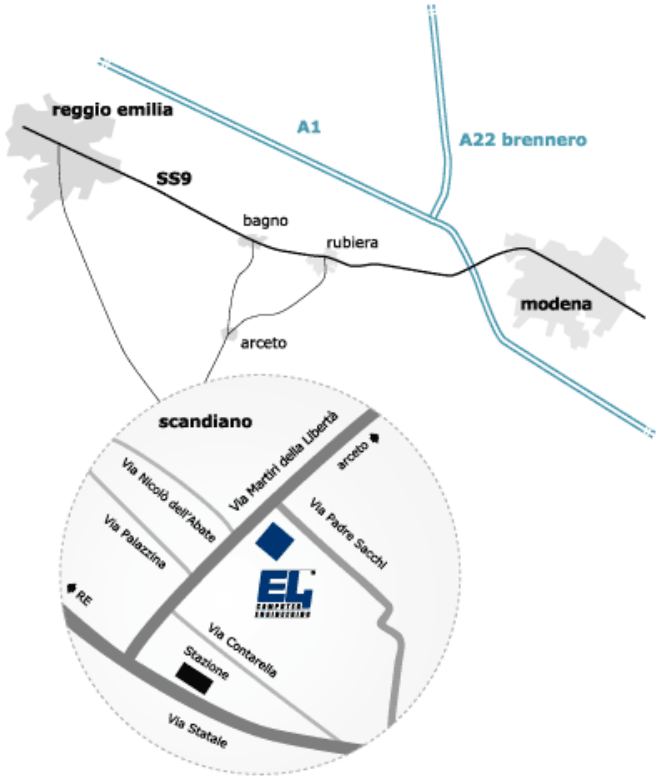
Agenda

- **E4 computer Engineering**
- **Calcolo**
  - CPUs
  - Compunting nodes
  - Hardware Accelerators (GPGPUs & C.)
- **Reti ad alte prestazioni**
  - 10 & 40 GbE
  - InfiniBand
- **Storage**
  - Storage solutions
  - Storage: analisi costo/prestazioni
- **HPC Middleware & Tools**

E4 Computer Engineering ?!?

Excellence 4 Computer Engineering

La Sede di E4®



- **2.600 m²**
- **Impianto solare fotovoltaico 100kWp (2012)**
- **Nuovi uffici destinati al reparto Tecnico (2010)**
- **Nuova area riservata alla produzione(2010)**
- **Due sale dedicate ai test (ampliate nel 2010)**

La Nostra Gamma Prodotti

PC Industriali (elaborazione immagini, biomedica…)

Workstation (fluido-dinamica, montaggio video…)

Server (firewall, nodi di calcolo, applicazioni scientifiche…)

Storage

HPC Cluster

Fino al …

… Datacenter completo

Pomezia data center

E4®: la garanzia di un partner altamente qualificato

Competenze Tecniche

Ottime relazioni coi Fornitori

Referenze

Soluzioni «su misura»

Assistenza

Certificazioni

E4®: la garanzia di un partner altamente qualificato

- Competenze Tecniche
- Relazioni coi Fornitori
- Referenze
- Soluzioni «su misura»
- Assistenza
- Certificazioni

Competenze Tecniche

Ufficio Interno di Ricerca &Sviluppo

Analisi di mercato: prodotti e campi applicativi

Analisi dei componenti

Analisi e test prestazionali

Tuning del Sistema

Validazione delle soluzioni HPC

Affidabilità

Le relazioni con i Fornitori

E4®: la garanzia di un partner altamente qualificato

- Competenze Tecniche
- Relazioni coi Fornitori
- Referenze
- Soluzioni «su misura»
- Assistenza
- Certificazioni

N° 1 Intel Southern Europe Channel Partner

AMD Fusion Partner Premier

N° 1 Italian Supermicro Partner

N° 1 Italian Qlogic/Mellanox Premium Partner

Nvidia Cuda/Tesla Preferred Partner

DotHill, Panasas, DDN Partner

Cray XD1 Partner

First Italian Asus Server Partner

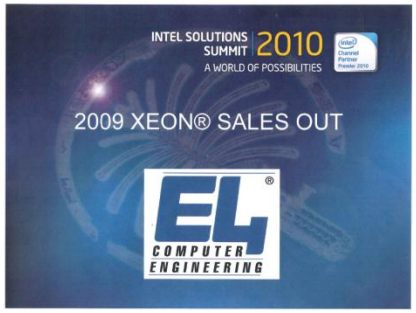
Numascale Partner

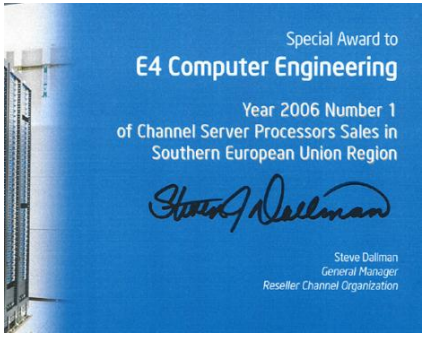
Nexenta Partner

Solarflare Partner

LSI, Adaptec, Areca Partner

Rittal Partner





Referenze Università & Ricerca

E4®: la garanzia di un partner altamente qualificato

- Competenze Tecniche
- Relazioni con i Vendors
- Referenze
- Soluzioni «su misura»
- Assistenza
- Certificazioni

CERN (20.000 cores and 18 PB di installato storage)

ESRF (European Synchrotron Radiation Facility)

SCL (Scientific Computing Laboratory - Serbia)

EPFL (École Polytechnique Fédérale de Lausanne)

ETH Zurich

I.N.A.F.

I.N.F.N.

S.I.S.S.A.

E.N.E.A.

CRS4

Caspur

CNR

I.N.G.V.

CMCC

"Ancora una volta, CERN sceglie l'Italiana E4 Computer Engineering… "

PRIMEUR Weekly — The monthly news service for the European HPCN community

(the largest Grid and Supercomputer information source in the world)

"E4 e Supermicro Blades scelti per un Progetto al CERN "

HPC wire

(The Leading Source for Global News and Information Covering the Ecosystem of High Productivity Computing)

ORACLE **Leading-Edge Business Intelligence Solution**

http://www.oracle.com/global/it/customers/pdf/snapshot_gruppodarco_e4.pdf

Referenze Industria, Governo, Difesa

E4®: la garanzia di un partner altamente qualificato

- Competenze Tecniche
- Relazioni coi Fornitori
- Referenze
- Soluzioni «su misura»
- Assistenza
- Certificazioni

Merck Group

SACMI

Sigma Tau

Ministero della Giustizia

Ministero dell'Interno

NATO-NURC

Thales Alania Space

Alenia Aeronautica

Ansaldo

Autodesk

Banca d'Italia

Biblioteca Apostolica Vaticana

Chiesi Farmaceutici

Enginsoft

Feltrinelli Editore

Fondazione Telethon

Novartis

Telecom Italia

Tiscali

Assistenza

E4®: la garanzia di un partner altamente qualificato

- Competenze Tecniche
- Relazioni coi Vendor
- Referenze
- Soluzioni su misura
- Assistenza
- Certificazioni

Partner Tecnici in Italia(oltre 200 tecnici)

Partner Tecnici in Europa

Servizio di assistenza On Site garantita entro Next Business Day

CERN: assistenza garantita entro 4 ore dalla chiamata

Call Center Tecnico interno

Sistemi E4 @ CERN

Il nostro team

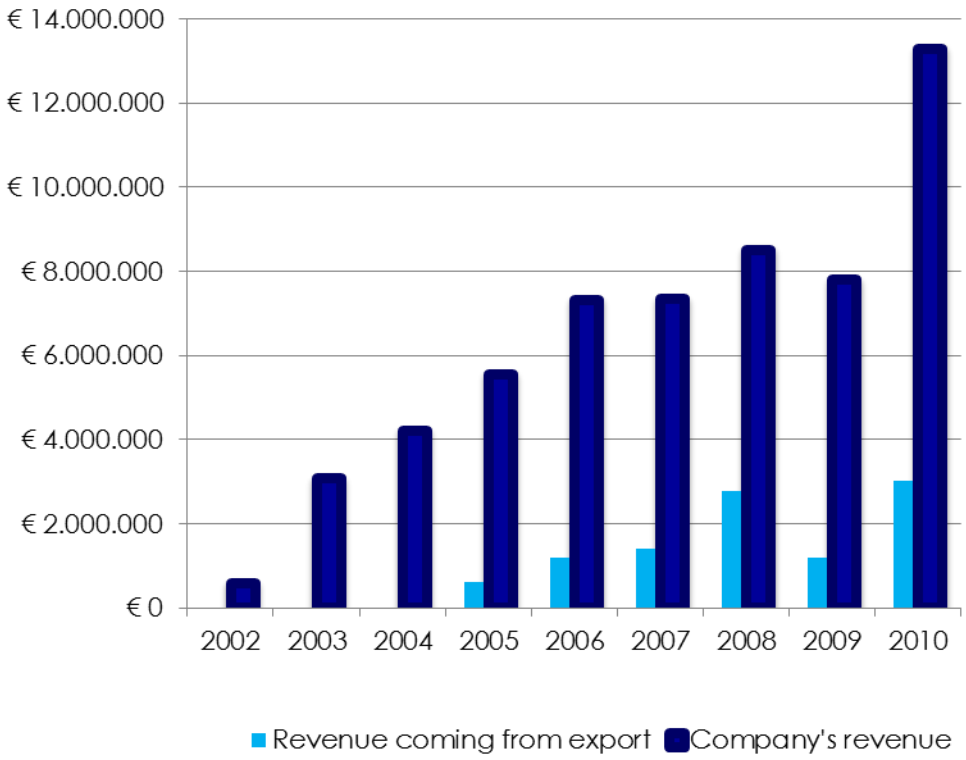
35 impiegati suddivisi tra

- ✓ Ingegneri
- ✓ Tecnici
- ✓ Ufficio Vendite (Nazionale e Internazionale)

E4® = partner solido come la roccia

- Ottimo rating Basilea2
- Ottime prestazioni finanziarie
- Crescita continua attraverso gli anni

La crescita di E4® dal 2002



Certificazioni

E4®: la garanzia di un partner altamente qualificato

Competenze Tecniche
Relazioni con i Vendors
Referenze
Soluzioni «su misura»
Assistenza
Certificazioni

ISO 9001: 2000


CSQ
www.imq.it
CISQ is a member of IQNet
www.iqnet-certification.com
CERTIFICATO N.
CERTIFICATE N. 9150.EQCE
SI CERTIFICA CHE IL SISTEMA QUALITA' DI
WE HEREBY CERTIFY THAT THE QUALITY SYSTEM OPERATED BY
E4 COMPUTER ENGINEERING SPA
VIA MARTIRI DELLA LIBERTA', 66 - 42019 SCANDIANO (RE)
UNITA' OPERATIVE
OPERATIVE UNITS
VIA MARTIRI DELLA LIBERTA', 66 - 42019 SCANDIANO (RE)
E' CONFORME ALLA NORMA
IS IN COMPLIANCE WITH THE STANDARD
ISO 9001:2000
PER LE SEGUENTI ATTIVITA'
FOR THE FOLLOWING ACTIVITIES
Progettazione, realizzazione, commercializzazione ed assistenza di workstation e server per l'utente professionale, le PMI e l'industria
Design, execution, commerce and technical support of workstation and server for professional use, small business and large industrial concerns
Riferirsi al manuale della qualità per l'applicabilità dei requisiti della norma ISO 9001:2000
Refer to quality manual for details of applications to ISO 9001:2000 requirements

In via di certificazione:

- ISO 14001: sistema di gestione ambientale

Prossima certificazione:

- Standard SA8000: responsabilità sociale d'impresa

Customer References

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

INFN

arpa

NURC

CRS4

CASPUR
Consorzio interuniversitario per le Applicazioni di Supercalcolo Per Università e Ricerca

UNIVERSITÀ DEGLI STUDI DI TRIESTE

PARTNERING FOR MARITIME INNOVATION

Ministero della Giustizia
DIPARTIMENTO DELL'AMMINISTRAZIONE PENITENZIARIA
PROVVEDITORATO REGIONALE DELLA BASILICATA

tiscali

UNIVERSITÀ DEGLI STUDI DI NAPOLI PARTHENOPE

UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA

BANCA D'ITALIA

MINISTERO DELL'INTERNO

SAPIENZA UNIVERSITÀ DI ROMA

ROMA TRE UNIVERSITÀ DEGLI STUDI

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

CERN

FONDMETAL

UNIVERSITÀ DEGLI STUDI DI PADOVA

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Ansaldo Sistemi Industriali
RESULTS TO THE POWER OF THREE

SCIENTIFIC COMPUTING LABORATORY

INAF

sigma-tau PHARMACEUTICALS, Inc.

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

MERCK

eni

ICTP the abdus salam international centre for theoretical physics

POLITECNICO DI MILANO

ma per seguir virtute e canoscenza
SISSA

ENEA

Chiesi
People and ideas for innovation in healthcare

ThalesAlenia Space

ESRF

POLITECNICO DI TORINO

UNIVERSITÀ DEGLI STUDI DI UDINE

issia

NOVARTIS

EL COMPUTER ENGINEERING

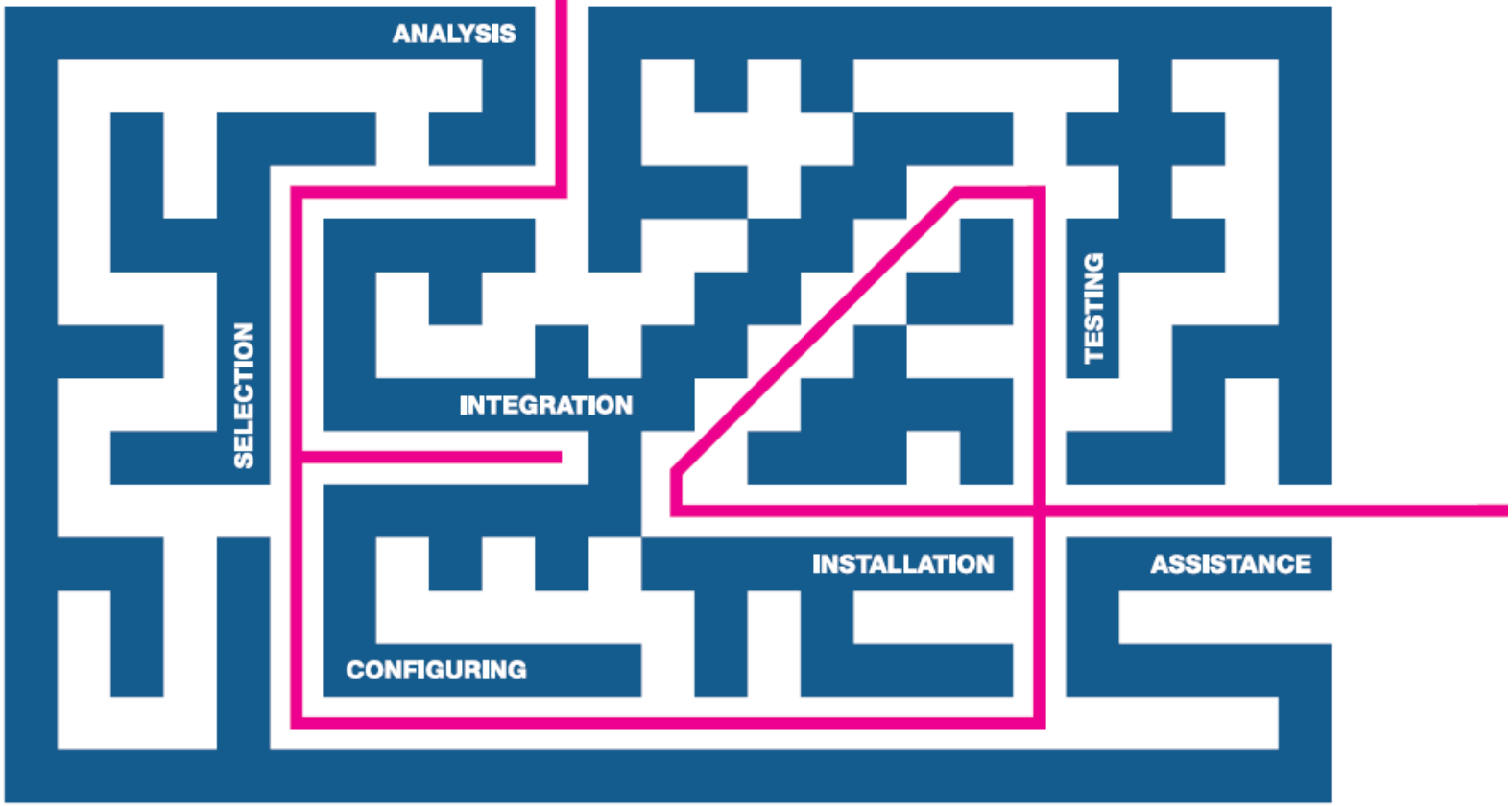
I princìpi di E4®: etica e legalità

- economia d’impresa nella piena osservanza delle leggi
- massimo supporto alle Istituzioni Accademiche
- Supporto attivo e costante a diverse organizzazioni Onlus
- Tra i principali sostenitori economici di

LIBERA
ASSOCIAZIONI, NOMI E NUMERI
CONTRO LE MAFIE

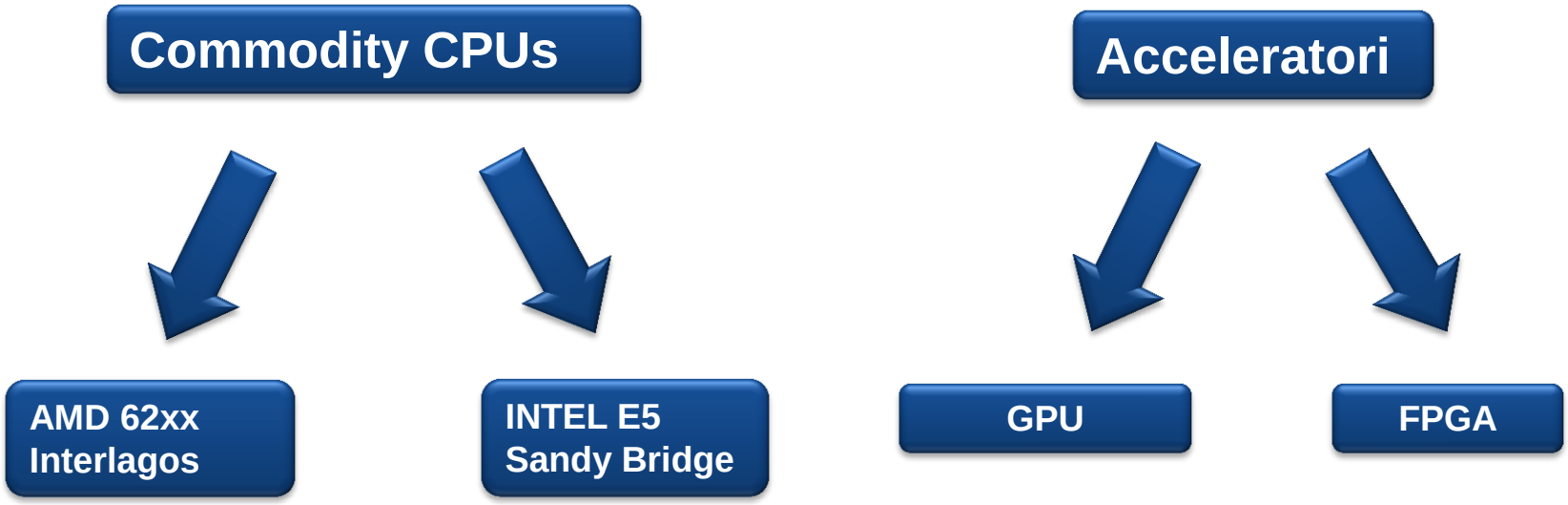
EL COMPUTER ENGINEERING

leads your way



from A to Z

Soluzioni per calcolo



AMD Interlagos

AMD OPTERON™ 6200 SERIES PROCESSOR

Product Specifications

Number of Processors per Server
- 1, 2, and 4 socket support

Frequency
- Up to 3.0 GHz base frequency
- Up to 3.6 GHz frequency using AMD Turbo CORE technology*

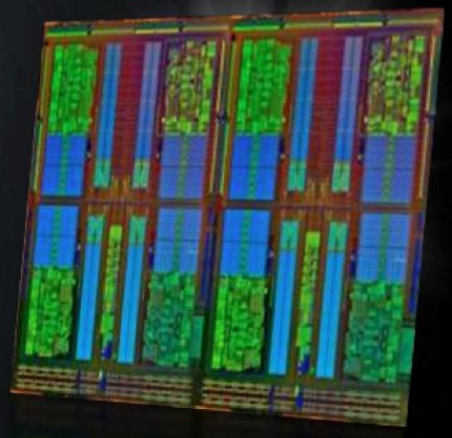
Cores
- 4, 8, 12, and 16 core options

Memory
- 4 memory channels
- Up to 1600 MHz memory
- Supports R/U DDR3, LV DDR3, ULV DDR3 and LR DDR3
- Up to 1.5 TB memory capacity
- Supports up to 12 DIMMs per CPU

Cache Size
- Total Cache: 32MB (16 core), 28MB (12 core), 24MB (8 core)
- L1 Cache: 16KB Data per core + 64KB Instruction (*per module*)
- L2 Cache: 1MB (*per core*)
- L3 Cache: 16MB *(per socket)*

HyperTransportTM 3.0 technology (HT3) Links
- Four x16 links @ up to 6.4GT/s per link

Power
- 85W to 140W TDP



Ideal for cloud, virtualization, HPC, Database and other highly scalable applications

ased on internal AMD engineering projections of AMD Opteron 6200 Series processors with up to 500 MHz in P1 boost state and up to 1.4 GHz in P0 boost state over base P2 ock frequency.

AMD

5 I AMD Opteron™ 6200 Series Customer Presentation I Aug 2011 I NDA only until launch

AMD Interlagos

GENERATIONAL COMPARISONS

	AMD Opteron™ 6100 Series Processors	AMD Opteron™ 6200 Series Processors
Cores	8 or 12 core	4, 8, 12 or 16 core
Cache (L2 per core / L3 per die)	512KB / 6MB	2MB (shared between 2 cores) / 8MB
Memory Channels and speed	four; up to 1333MHz	four; up to 1600MHz
Floating point capability	128-bit FPU per core (FADD/FMUL)	128-bit dedicated FMAC per core or 256-bit AVX shared between 2 cores
Integer Issues Per Cycle	3	4
Turbo CORE Technology	No	Yes (+500MHz with all cores active)
Power (ACP)	65W, 80W, 105W	TBD (planned 65W, 80W, 105W)
New Instruction Sets		SSSE3, SSE 4.1/4.2, AVX, AES, FMA4, XOP, PCLMULQDQ
Power Gating	AMD CoolCore™, C1E	AMD CoolCore™, C1E, C6
Process / Die Size	45nm SOI	32nm SOI (smaller overall die size)
Performance		35% higher processing throughput*

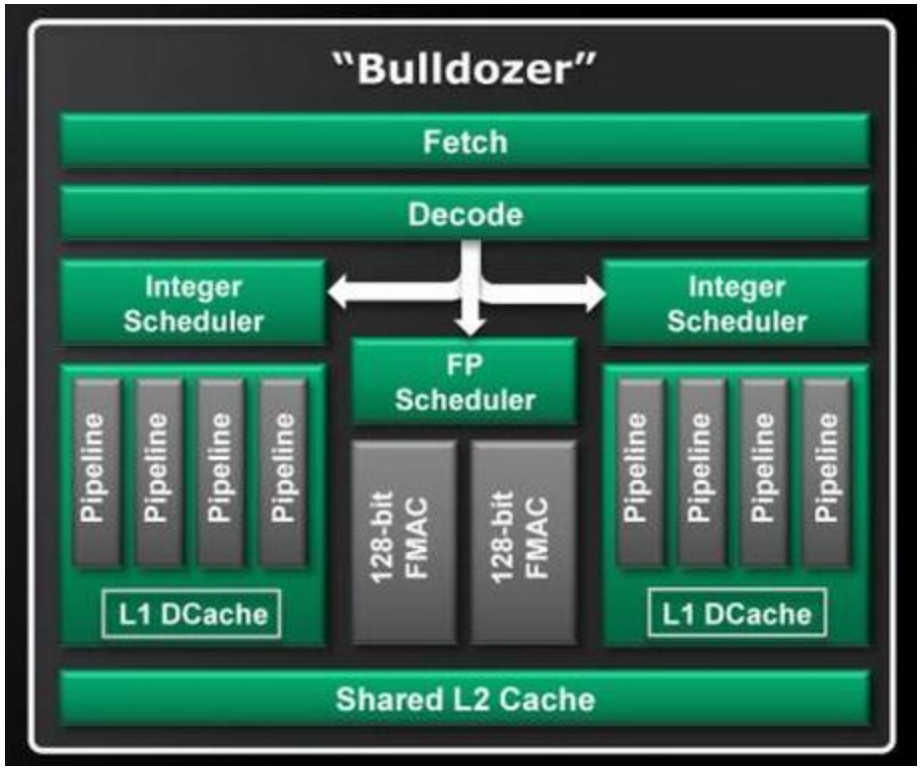
The above reflect current expectations regarding features and performance and is subject to change.

*Based on pre-production measurements by AMD labs as of 8/9/11 and comparison of top bin AMD Opteron™ 6200 Series processors to top bin AMD Opteron 6100 Series processors

21 I AMD Opteron™ 6200 Series Customer Presentation I Aug 2011 I NDA only until launch

AMD

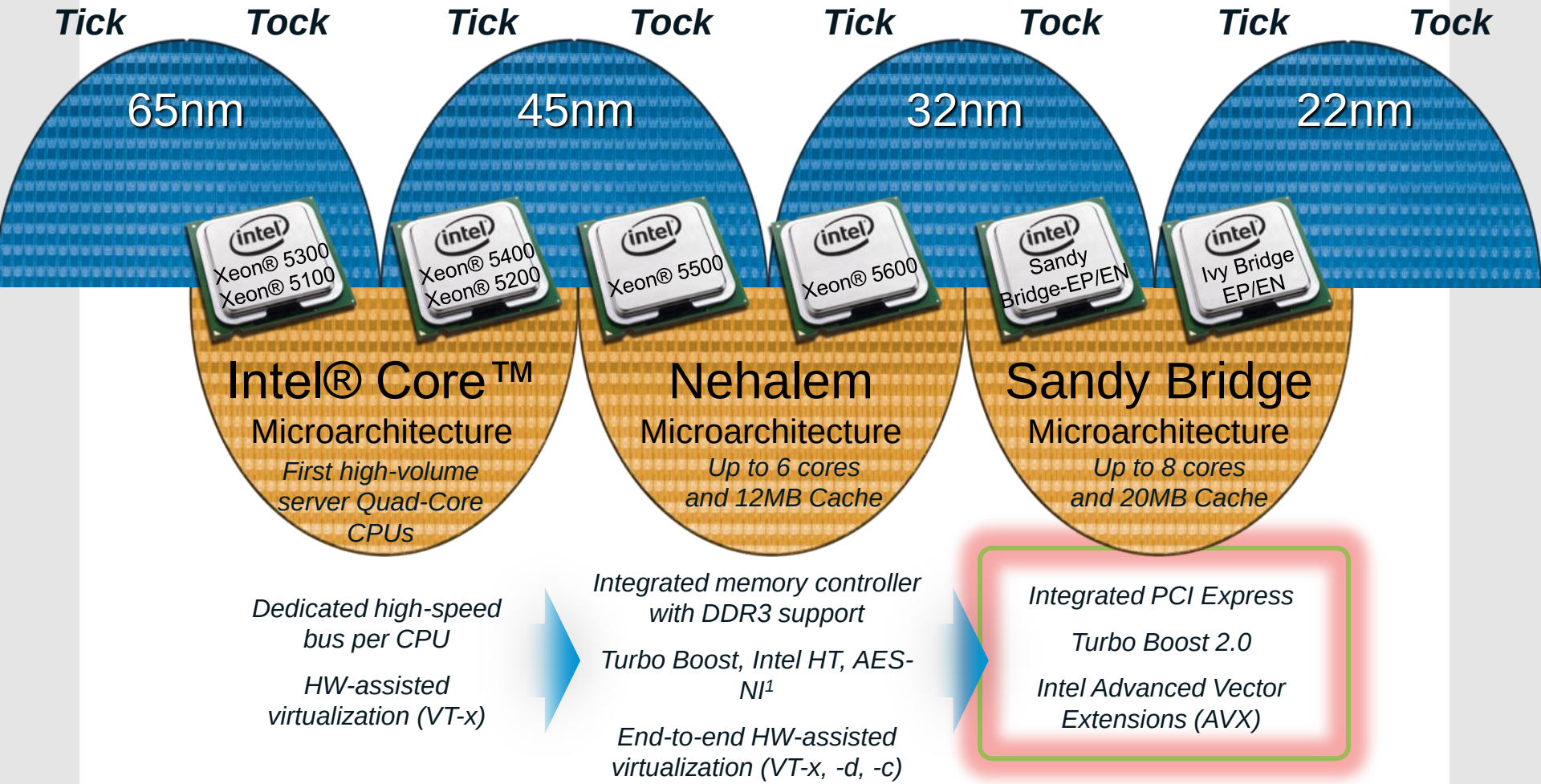
AMD Bulldozer Architecture



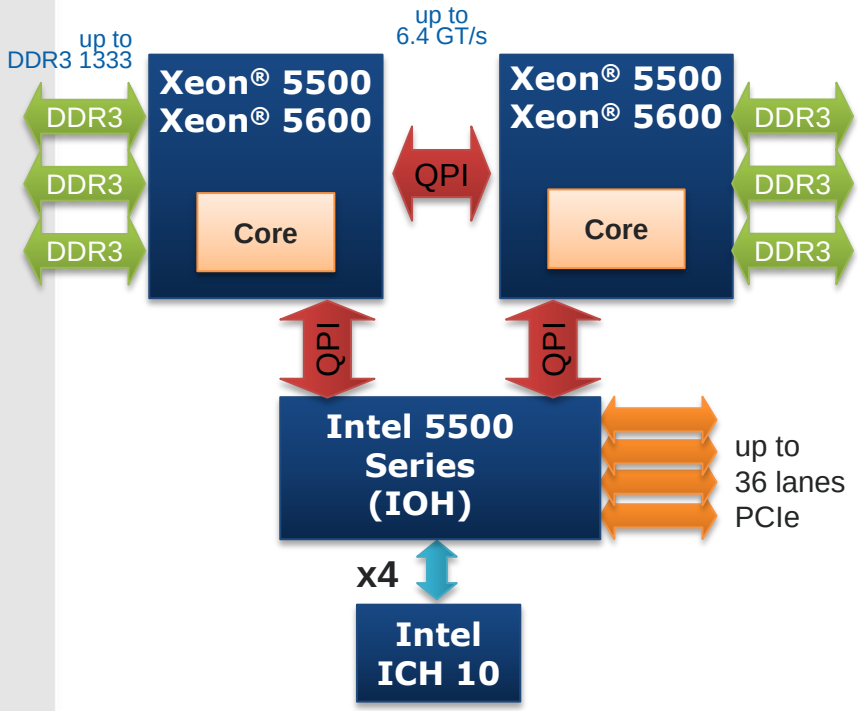
AMD Interlagos

AMD Opteron™ 6200 Series Processor Product Specifications								
Model Number	Core Count	Core Frequency	All Core Boost Frequency	AMD Turbo CORE Max Frequency	North Bridge	L2 Cache	L3 Cache	TDP
6284 SE*	16	2.7 GHz	3.1 GHz	3.4 GHz	2.0 GHz	8 x 2 MB	16 MB	140W
6282 SE	16	2.6 GHz	3.0 GHz	3.3 GHz	2.0 GHz	8 x 2 MB	16 MB	140W
6278*	16	2.4 GHz	2.7 GHz	3.3 GHz	2.0 GHz	8 x 2 MB	16 MB	115W
6276	16	2.3 GHz	2.6 GHz	3.2 GHz	2.0 GHz	8 x 2 MB	16 MB	115W
6274	16	2.2 GHz	2.5 GHz	3.1 GHz	2.0 GHz	8 x 2 MB	16 MB	115W
6272	12	2.1 GHz	2.4 GHz	3.0 GHz	2.0 GHz	8 x 2 MB	16 MB	115W
6262 HE	16	1.6 GHz	2.1 GHz	2.9 GHz	1.8 GHz	8 x 2 MB	16 MB	85W
6238	12	2.6 GHz	2.9 GHz	3.2 GHz	2.0 GHz	6 x 2 MB	16MB	115W
6234	12	2.4 GHz	2.7 GHz	3.0 GHz	2.0 GHz	6 x 2 MB	16 MB	115W
6220	8	3.0 GHz	3.3 GHz	3.6 GHz	2.0 GHz	4 x 2 MB	16 MB	115W
6212	8	2.6 GHz	2.9 GHz	3.2 GHz	2.0 GHz	4 x 2 MB	16 MB	115W
6204	4	2.3 GHz	N/A	N/A	2.0 GHz	2 x 2 MB	16 MB	115W

INTEL Sandy Bridge

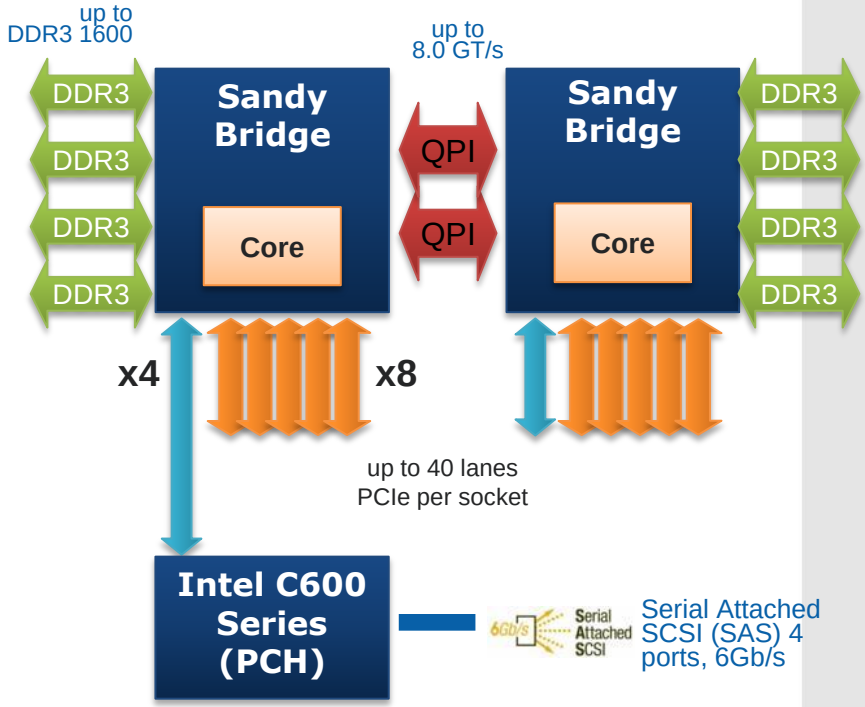


Xeon® 5500 / 5600 based Platform



- Up to 18 DIMMs per 2S platform
- Up to 36 PCIe lanes
- Two-chip IOH / ICH

E5-2600 Product Family based Platform



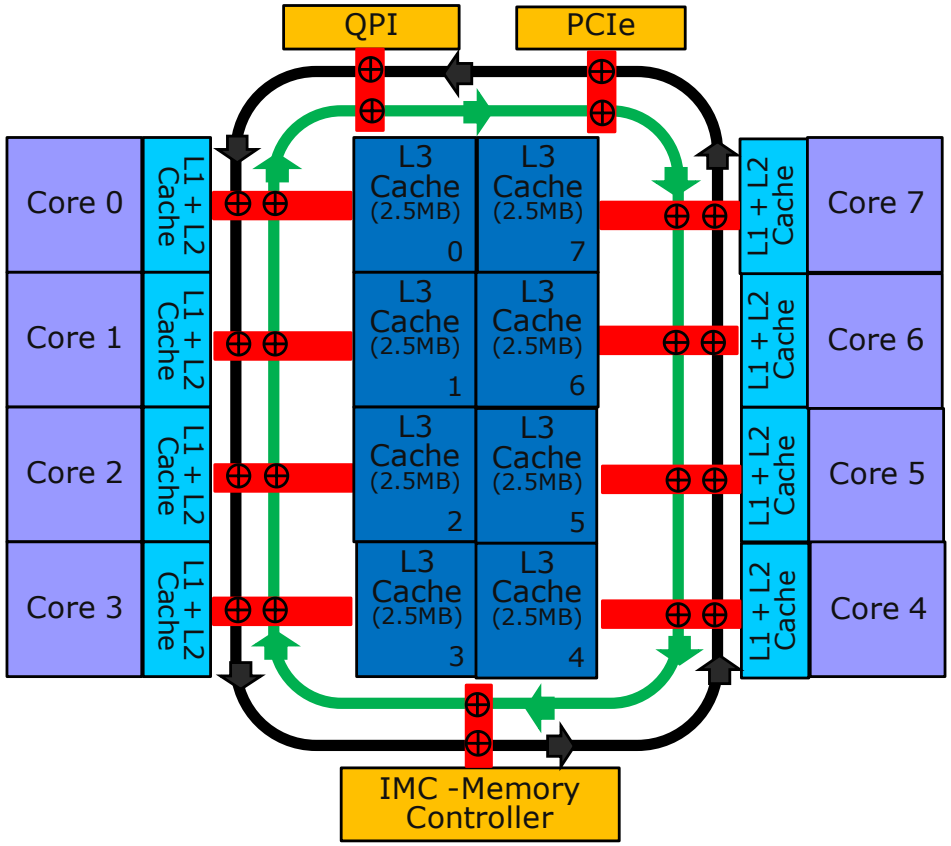
- Up to 24 DIMMs
- Up to 80 PCIe lanes
- Two QPI links between CPUs
- One-chip Platform Controller Hub (PCH)

Sandy Bridge "Ring Bus" Optimizes Platform Data Movement

Ring bus architecture delivers an efficient bi-directional highway for data movement

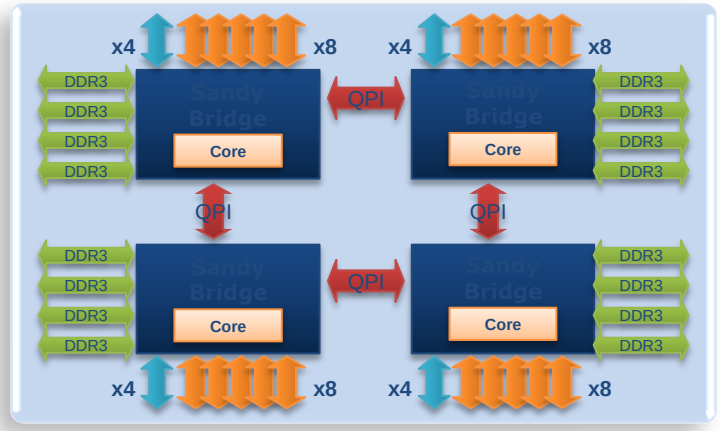
Compared to Xeon 5500/5600:

- Lower L3 cache latency (~20%)
- Higher bandwidth between cores, L3, Memory & I/O
- Up to 8x more L3 to core bandwidth



E5-4600

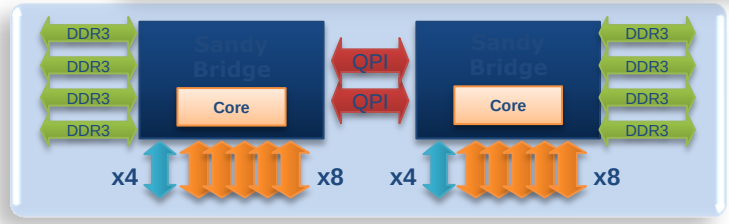
Density and Cost Optimized 4S



- Up to 4 CPUs
- Up to 48 DIMMs
- Up to 160 PCIe lanes
- Two QPI links per CPU (ring topology)

E5-2600

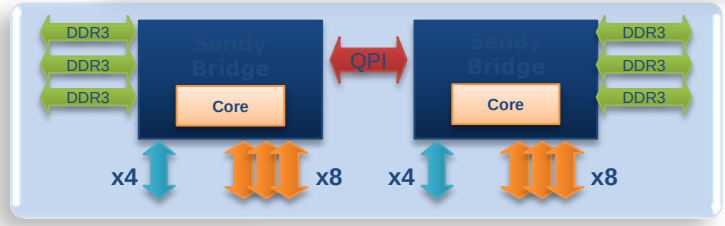
Highest Performance Max Memory



- Up to 24 DIMMs
- Up to 80 PCIe lanes
- Two QPI links

E5-2400

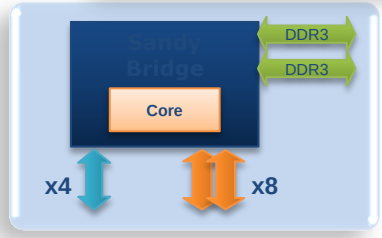
Density and Cost Optimized 2S



- Up to 12 DIMMs
- Up to 48 PCIe lanes
- One QPI link

E3-1200

Lowest Cost (1S only)



- Up to 4 DIMMs
- Up to 20 PCIe lanes

	Intel® Xeon® processor E5-2400 product family Socket-B2	Intel® Xeon® processor E5-2600 product family Socket-R
Advanced ▪ 8C-20M cache ▪ 8.0 GT/s QPI ▪ Hyper-Threading ▪ DDR3-1600 ▪ Turbo +	8C 95W 2.3 GHz — E5-2470 8C 95W 2.1 GHz — E5-2450	8C 135W 2.9 GHz — E5-2690 8C 130W 2.7 GHz — E5-2680 8C 115W 2.6 GHz — E5-2670 8C 115W 2.4 GHz — E5-2665 8C 95W 2.2 GHz — E5-2660 8C 95W 2.0 GHz — E5-2650
Standard ▪ 6C-15M cache ▪ 7.2 GT/s QPI ▪ Hyper-Threading ▪ DDR3 1333 ▪ Turbo	6C 95W 2.4 GHz — E5-2440 6C 95W 2.2 GHz — E5-2430 6C 95W 1.9 GHz — E5-2420	6C 95W 2.5 GHz — E5-2640 6C 95W 2.3 GHz — E5-2630 6C 95W 2.0 GHz — E5-2620
Basic ▪ 4C-10M, 2C-5M cache ▪ 6.4 GT/s QPI ▪ DDR3 1066	4C 80W 2.2 GHz — E5-2407 4C 80W 1.8 GHz — E5-2403	4C 80W 2.4 GHz — E5-2609 4C 80W 1.8 GHz — E5-2603

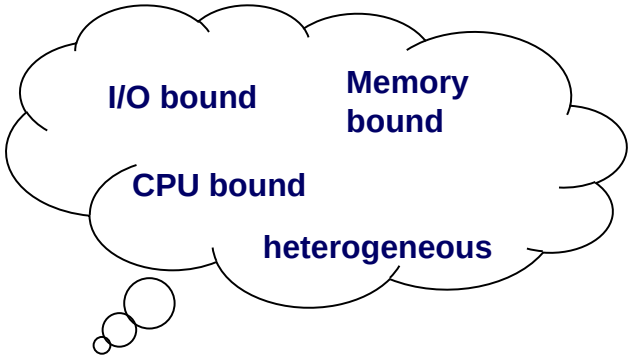
Addt'l R skt E5-2600 SKUs

E5-2687W	8C 150W 3.1 GHz
E5-2667	6C 130W 2.9 GHz
E5-2643	4C 130W 3.3 GHz
E5-2637	2C 80W 3.0 GHz

Low Power

Skt-B2	Skt-R
8C 70W 1.8 GHz — E5-2450L	8C 70W 1.8 GHz — E5-2650L
6C 60W 2.0 GHz — E5-2430L	6C 60W 2.0 GHz — E5-2630L

Computing nodes



Requirements analysis

SMP

Cluster

Interconnection

Storage

Software/Middleware

Computing nodes

AMD	INTEL
•**Form factor:** [1U to 4U, blade, TWIN, TWIN2]	•**Fattore di forma:** [1U to 4U, blade, TWIN, TWIN2]
•**Nsocket:** [1,4(8)]	•**Nsocket:** [1,4,(8)]
•**Ncore:** 6/8/12/16	•**Ncore:** 2/4/6/8 (+ Multithreading)

Too many combinations to check them all, rely on:
•Analysis
•Experience

… but the best way is to test them!!!

AMD & INTEL systems

1U TWIN: 2 computing nodes

7U: 10 - 14 - 20 blades

2U TWIN²: 4 computing nodes
New 8nodes **4U Fat TWIN**

sgi

BULL

EL COMPUTER ENGINEERING

Workstations

E7095

CPU: 2x INTEL
RAM: 12 slots (48GB/4GB modules, up to 192GB)
HD: 8x3,5”, SATA
Connectivity: 2Gb,1 NIC for management
Expansion slots: 4xPCI-Ex x16 Gen 2.0 + others

GPU: up to 4 or 8 using external GPU box + 1 nVIDIA low performance VGA

E8095

CPU: 2x ADM
RAM: 16 slots (64GB/4GB modules DDR2)
HD: 8x3,5”, SATA
Connectivity: 2Gb,1 NIC for management
Expansion slots: 4xPCI-Ex x16 Gen 2.0

GPU: up to 4 or 8 using external box

EL COMPUTER ENGINEERING

Interconnessioni

Ethernet

10GbE

40GbE

Kernel (TCP/IP) bypass

iWARP (RDMA)

iWARP (RDMA)

SOLARFLARE®

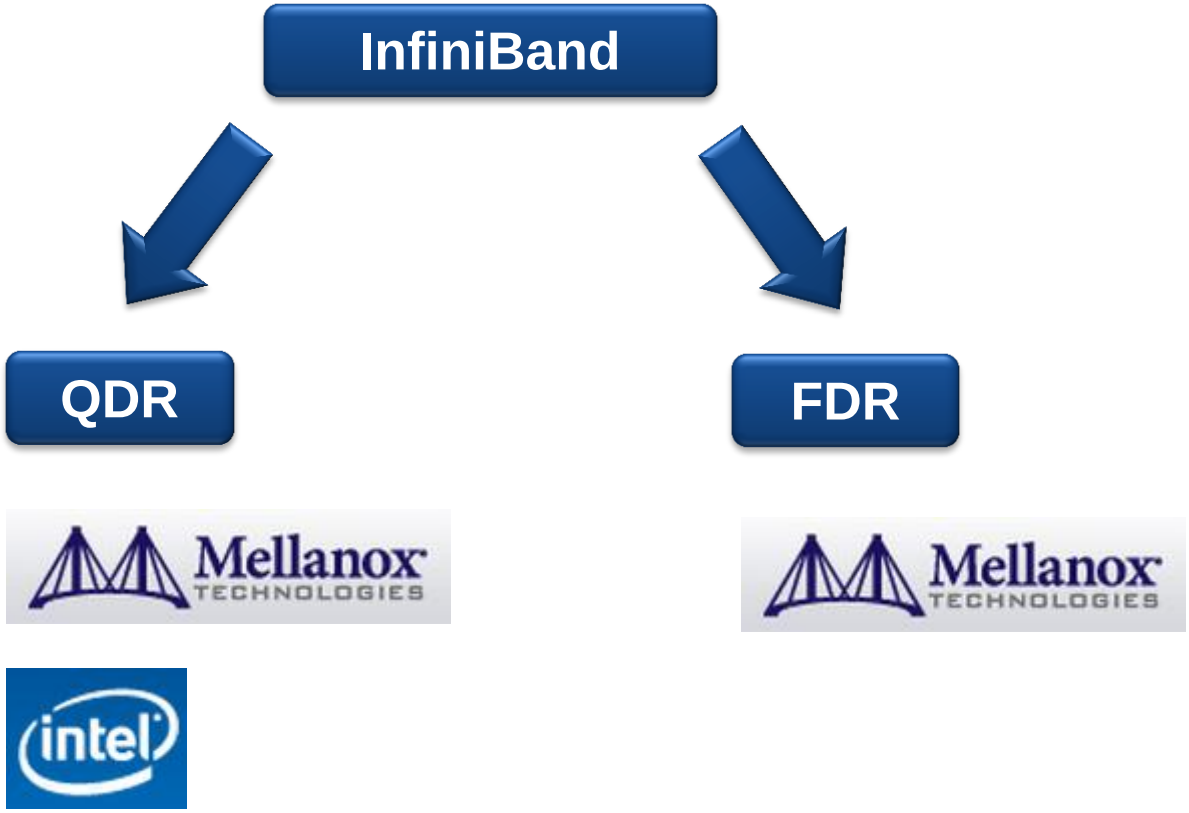
Chelsio Communications
Accelerate

Mellanox TECHNOLOGIES

Mellanox TECHNOLOGIES

EL COMPUTER ENGINEERING

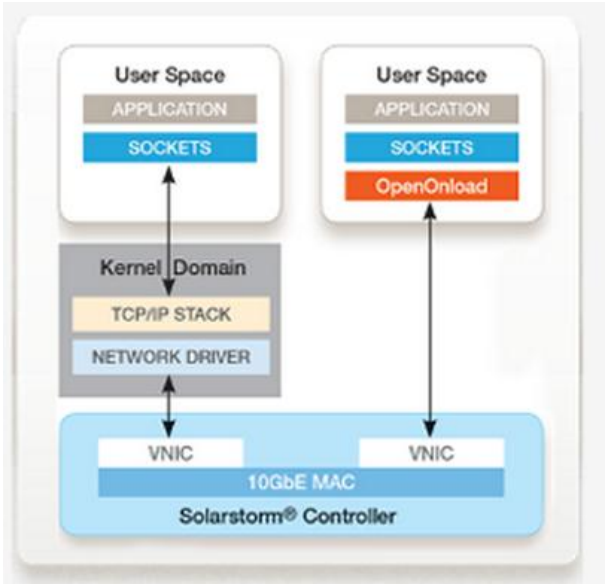
Interconnessioni



10GbE: Solarflare

Kernel (TCP/IP stack) bypass

~6usec MPI latency using OpenOnload



SOLARFLARE®

10GbE: Chelsio

iWARP (**Internet Wide Area RDMA Protocol**)

~4usec MPI latency (T4 adapter)

Unified Wire

TCP
UDP
iSCSI
iWARP
FCoE
The T4 adapters also fully support:
SR-IOV
EVB/VNTag
DCB
Traffic Management
Filtering

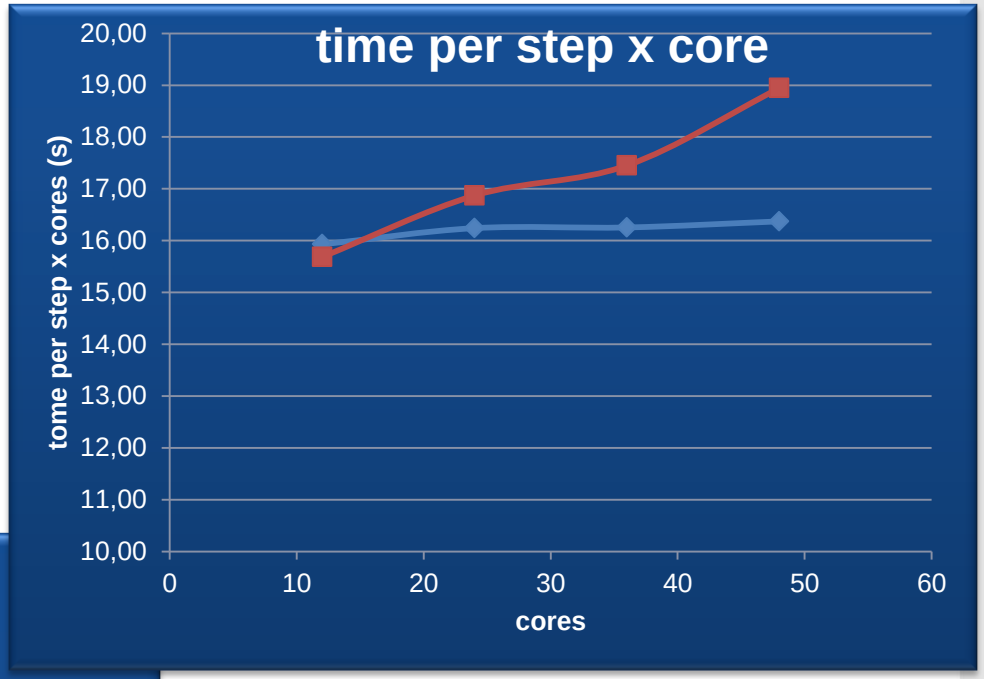
10/40 GbE, InfiniBand: Mellanox

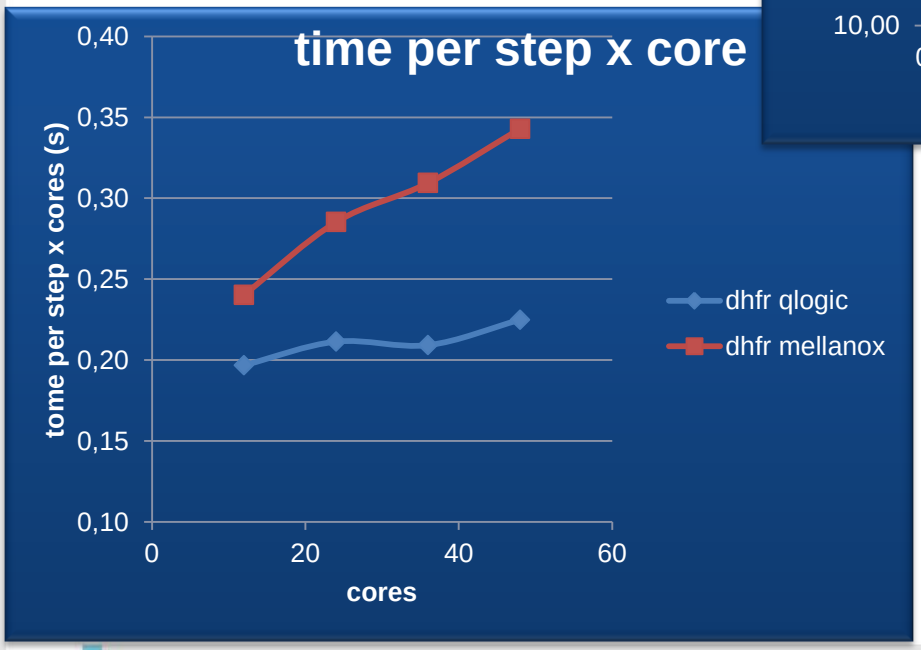
ConnectX-2 VPI: 40 Gb/s (DQR) IB & 10GbE

ConnectX-3 VPI: 56 Gb/s (FDR) IB & 40GbE

InfiniBand: INTEL

Intel® TrueScale InfiniBand 7300 Series
-QDR
-PCI x8 Gen 2.0



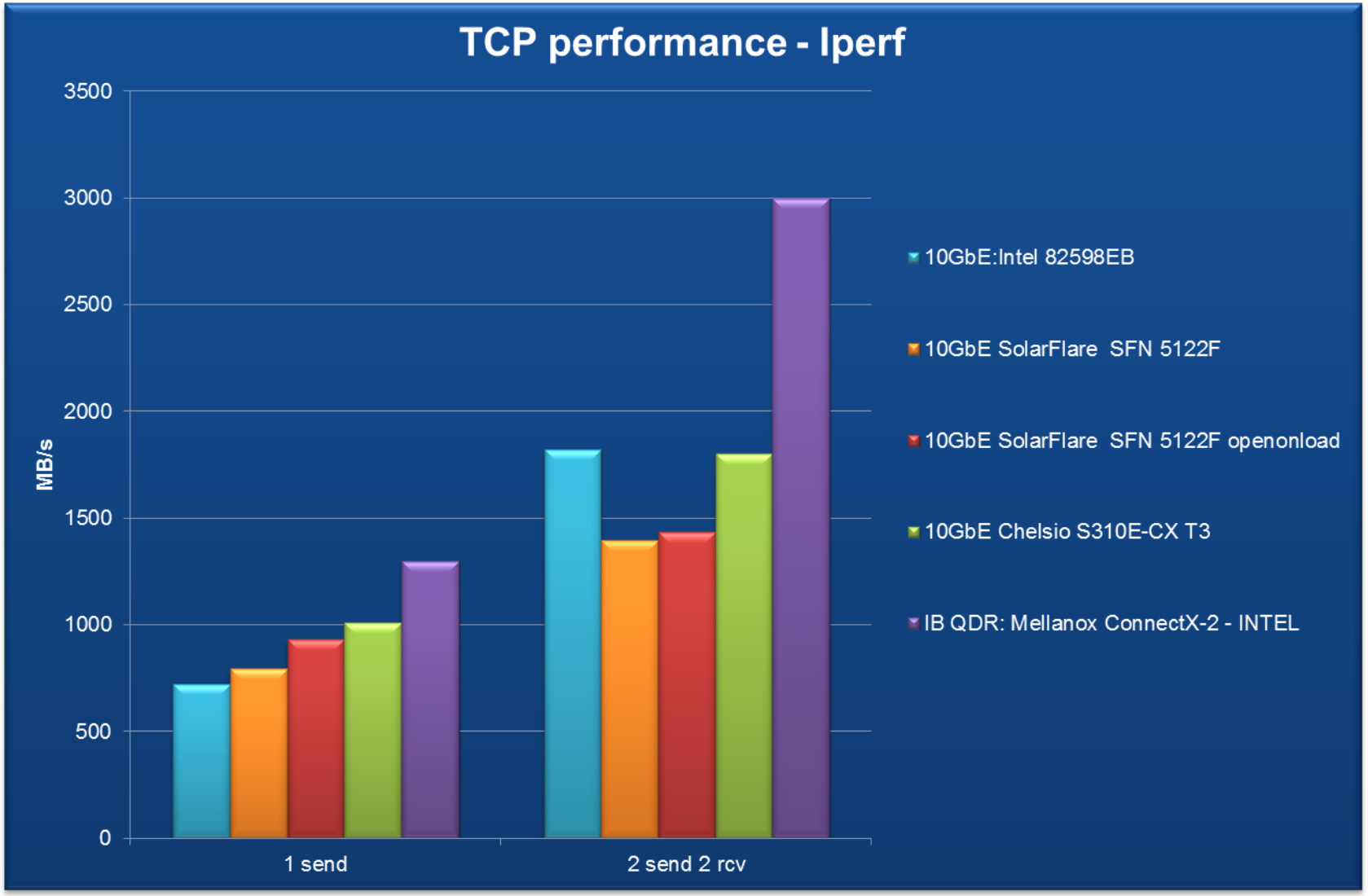


↓

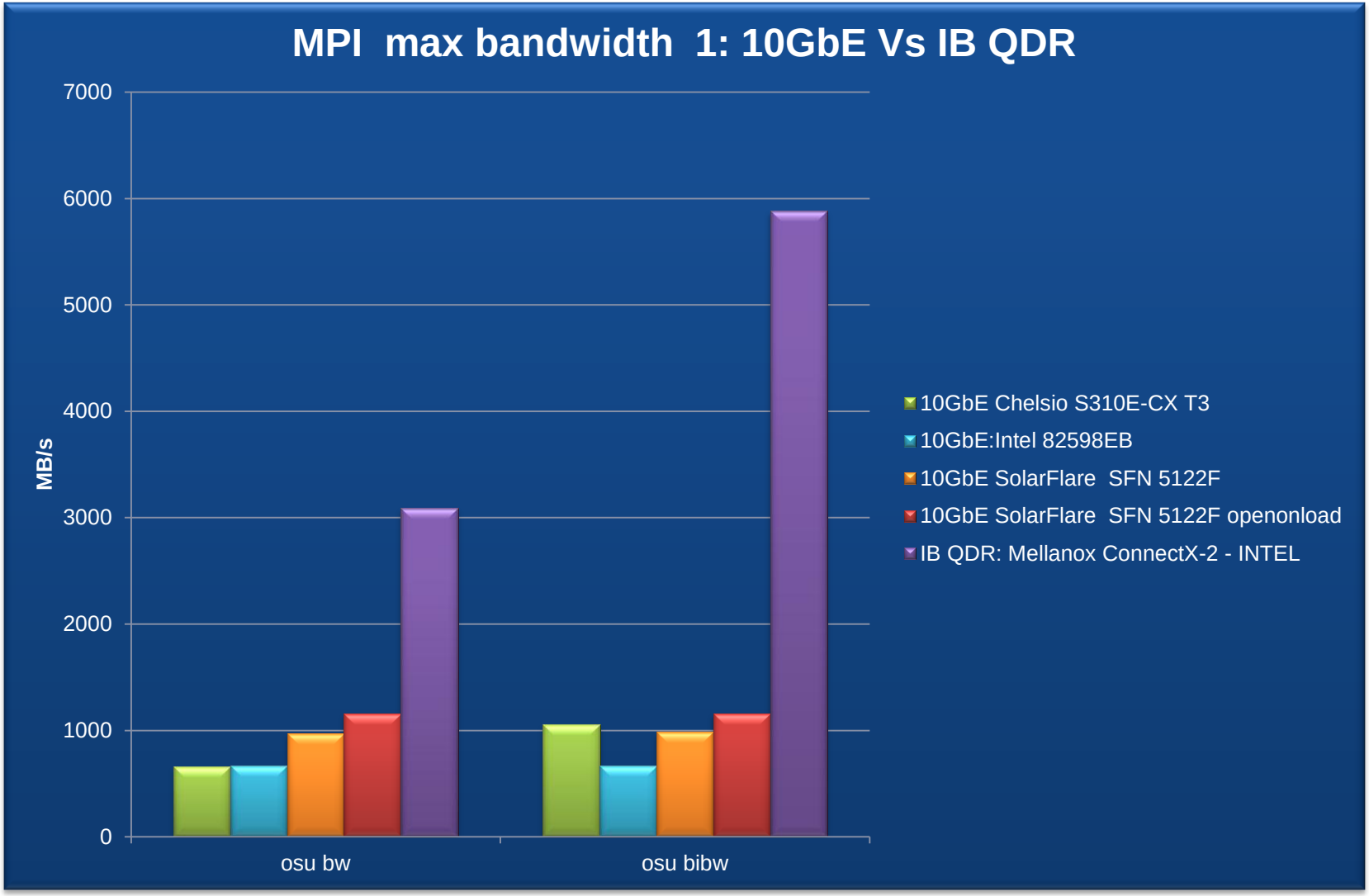
SMTV 1 milione di atomi

NAMD: INTEL Vs. Mellanox

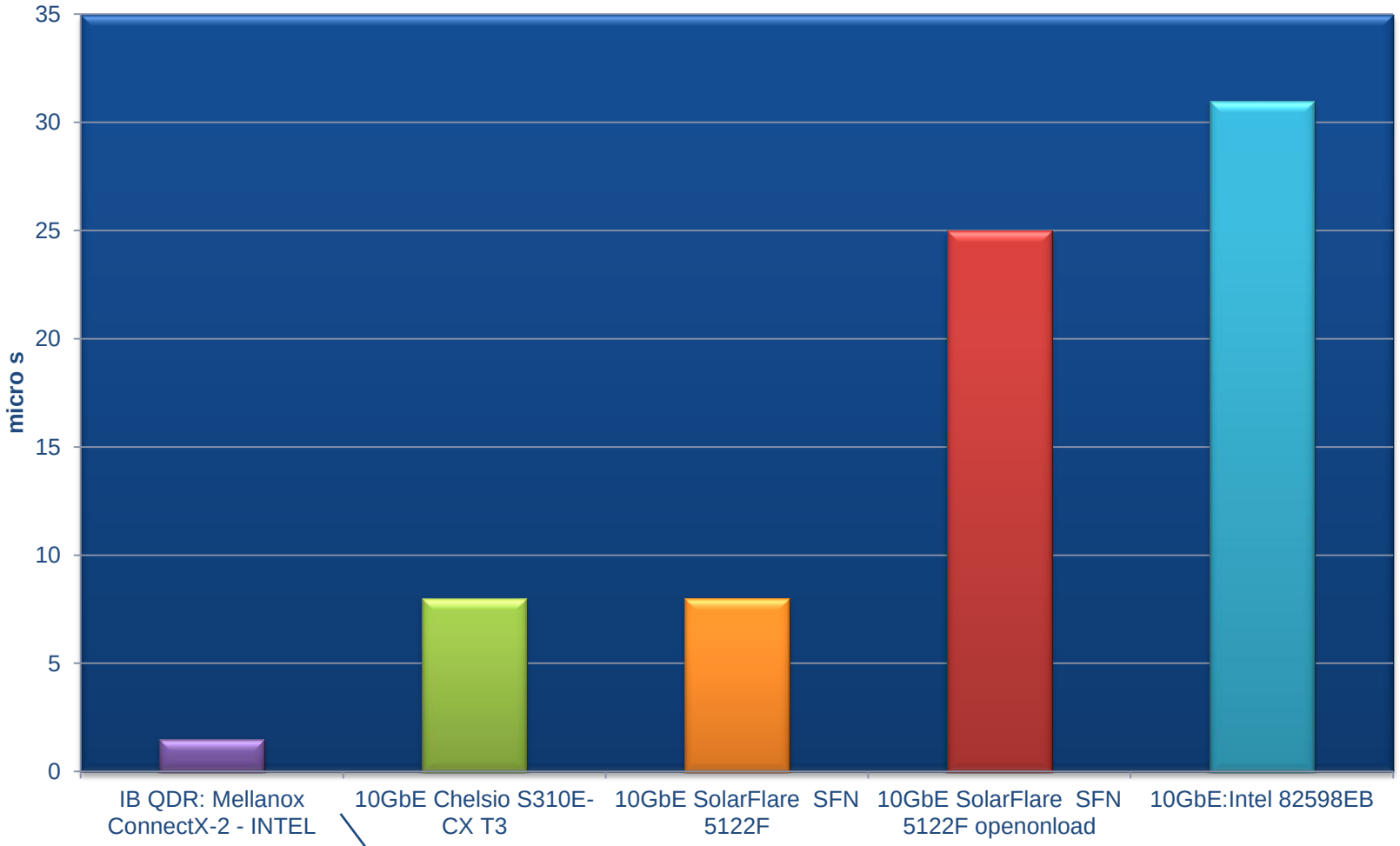
→ DHFR 2300 atomi



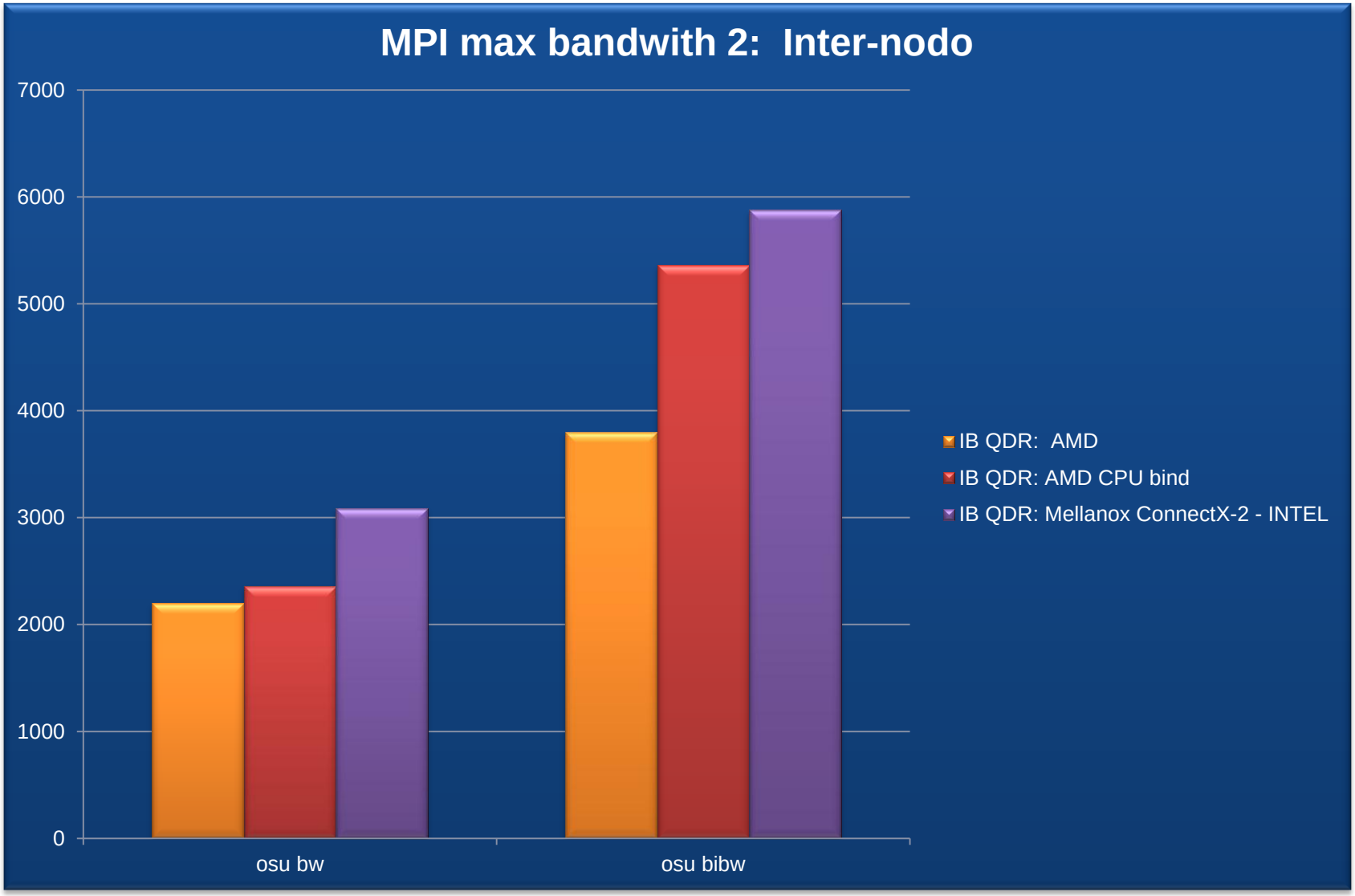
Chelsio T4: **2300 MB/s** 2 send 2 rcv


MPI max bandwidth 1: 10GbE Vs IB QDR
MB/s
7000
6000
5000
4000
3000
2000
1000
0
osu bw
osu bibw
10GbE Chelsio S310E-CX T3
10GbE:Intel 82598EB
10GbE SolarFlare SFN 5122F
10GbE SolarFlare SFN 5122F openonload
IB QDR: Mellanox ConnectX-2 - INTEL

MPI min latency (OSU_lat)

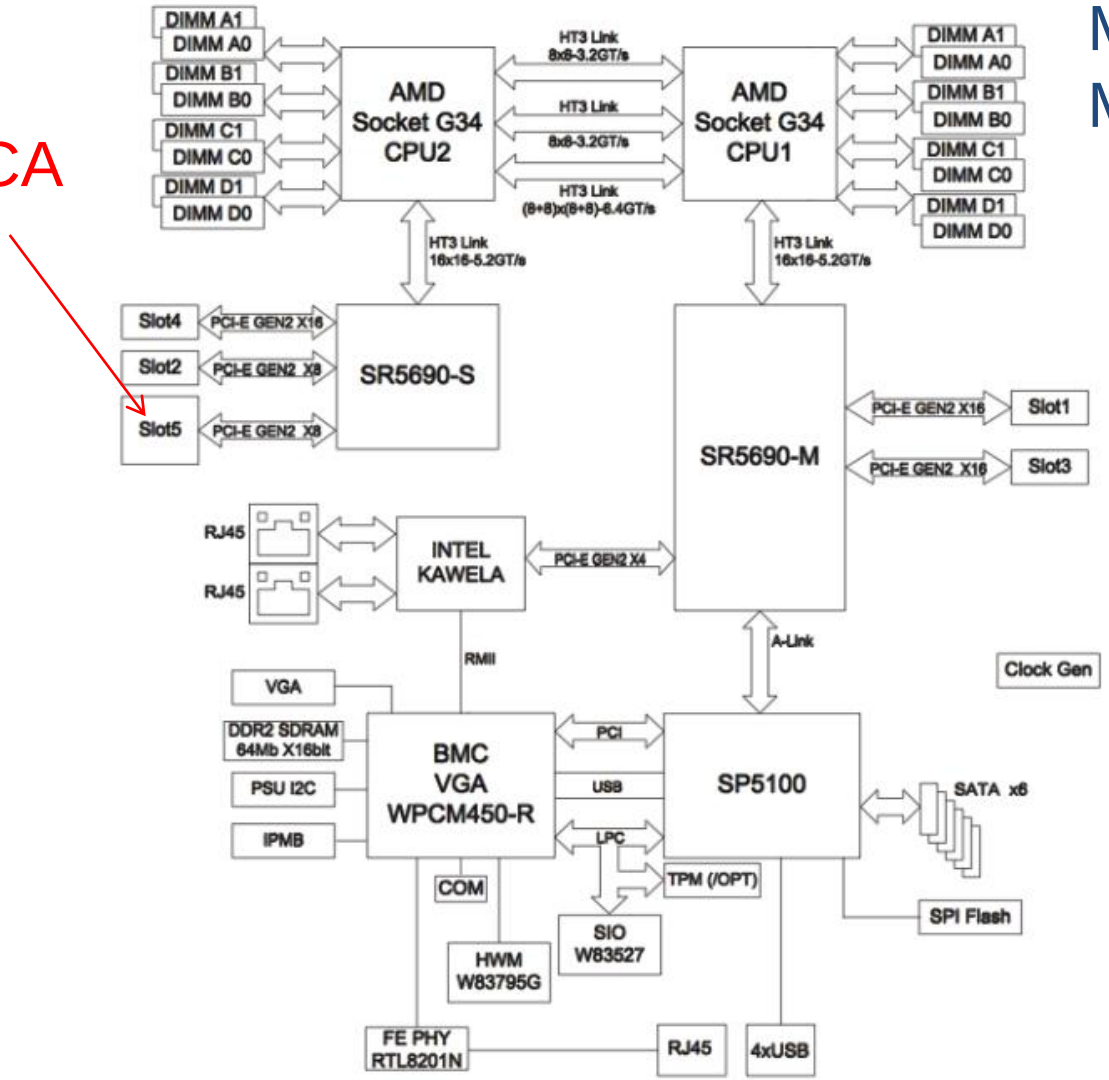


Chelsio T4: 4,8 usec

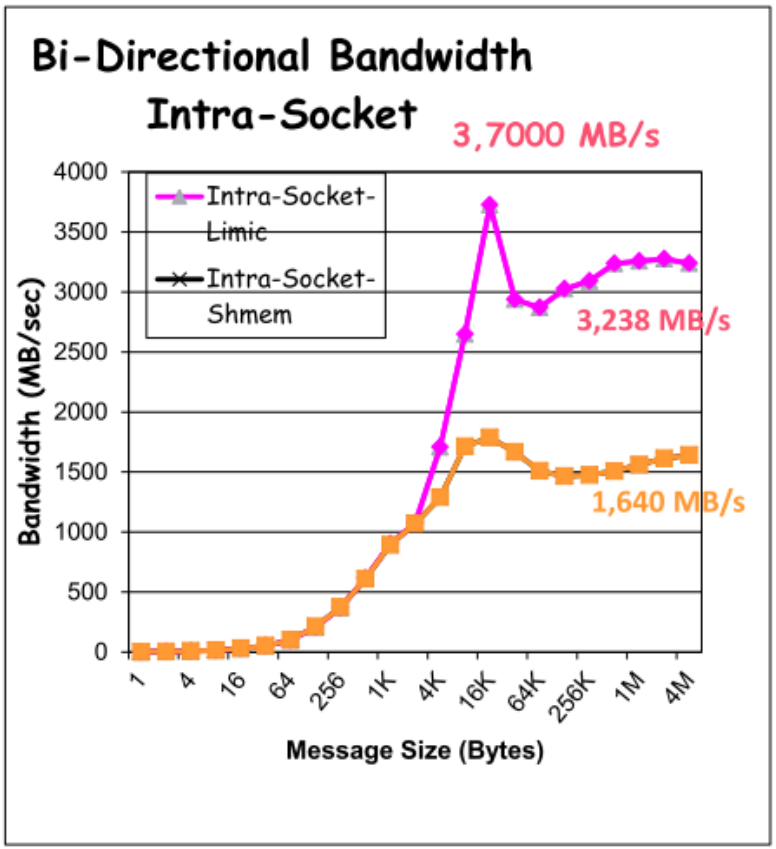

MPI max bandwith 2: Inter-nodo
7000
6000
5000
4000
3000
2000
1000
0
osu bw
osu bibw
IB QDR: AMD
IB QDR: AMD CPU bind
IB QDR: Mellanox ConnectX-2 - INTEL

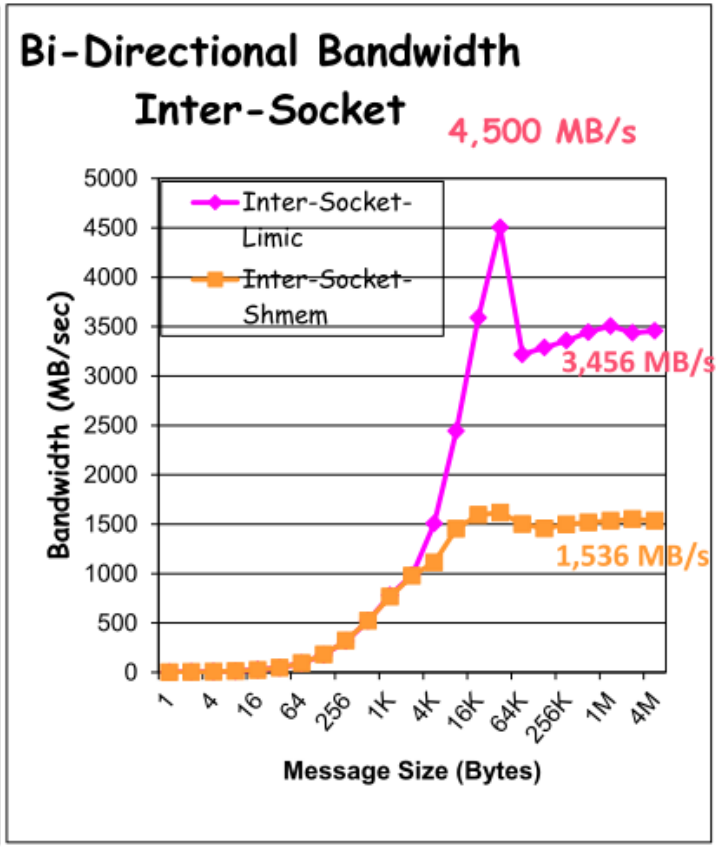
Multi chipset Motherboard

IB HCA

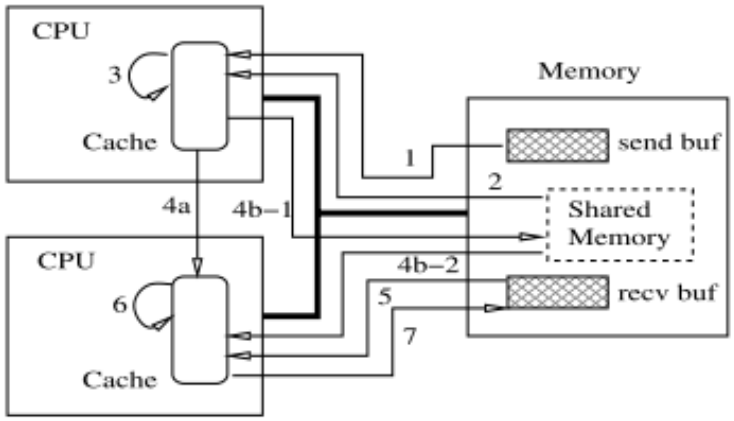


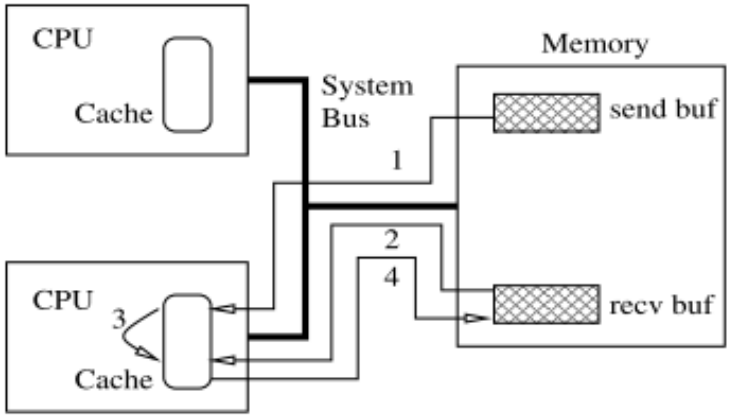
MPI bandwidth 3: Intra-nodo





Fonte: Dhabaleswar K. (DK) Panda, Sayantan Sur (The Ohio State University) - MPI Performance and Optimizations, Tutorial at HPC Advisory Council Workshop, Lugano 2011





Shared-Memory based two-copy intra-node communication

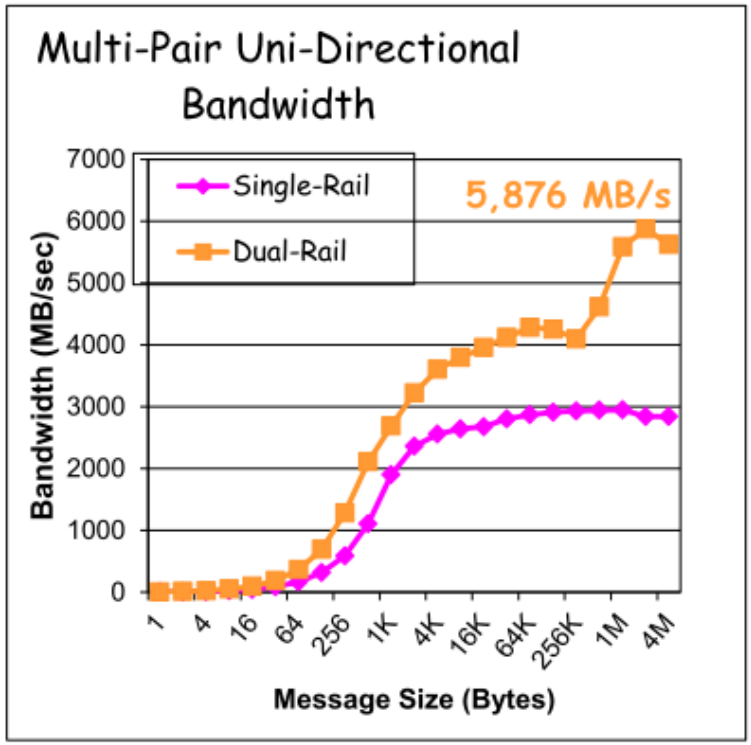
– Copy from the sender's user buffer to the shared buffer
– Copy from the shared buffer to the receiver's user buffer

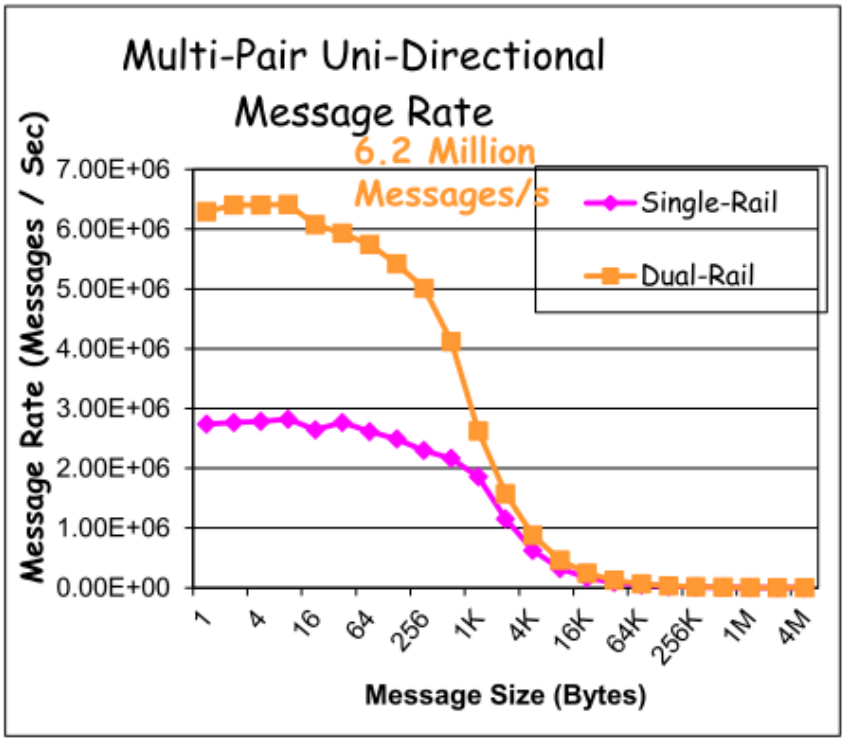
LiMIC2 on modern multi-core platforms

Linux kernel Module for MPI Intra-node communication

– Motivation: Eliminate one memory copy
– Kernel-level module for achieving efficient intra-node communication

MPI Bidirectional bandwidth 5: Multirail





Fonte: Dhabaleswar K. (DK) Panda, Sayantan Sur (The Ohio State University) - MPI Performance and Optimizations, Tutorial at HPC Advisory Council Workshop, Lugano 2011

Storage

Ampia gamma prodotti

Disk Server
FC storage
HPC storage (DDN, PANAS, EMC2, Hitachi)

per una vasta gamma di soluzioni:

-NAS, DAS, SAN

-Appliance ZFS nexenta
Enterprise class storage for everyone

Disks server

- Based on commodity hardware
- Tailor made upon customer requirements
- 1-2 CPUs
- 108TB in 4U
- Interconnection: Gb, 10GbE, IB, FC

E6550

- **Chassis 2U**
  - 12 HD SAS/SATA/SSD

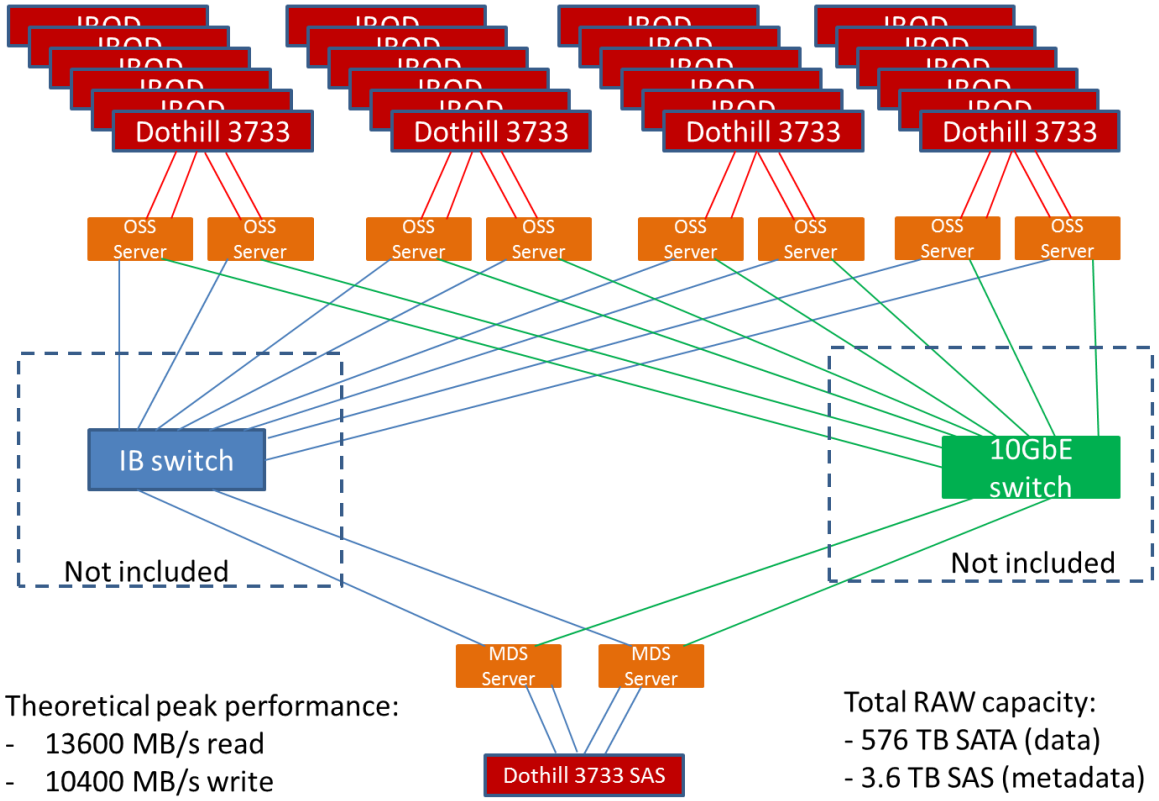
Full system, 18U – 9 chassis

- **up to 108 HD**
- Mix SATAII/SAS
- Modular desing
- Host interface:
  - 8 x 8Gb/s Fibre channel
- **Performance: 3,2GB / read 2,9 GB/write**

- Super capacitor cache back up
- Mirrored cache (Simulcache)
- Assured snapshot
- Volume copy

www.e4company.com simone.tinti@e4company.com

E6550 @ ETHZ



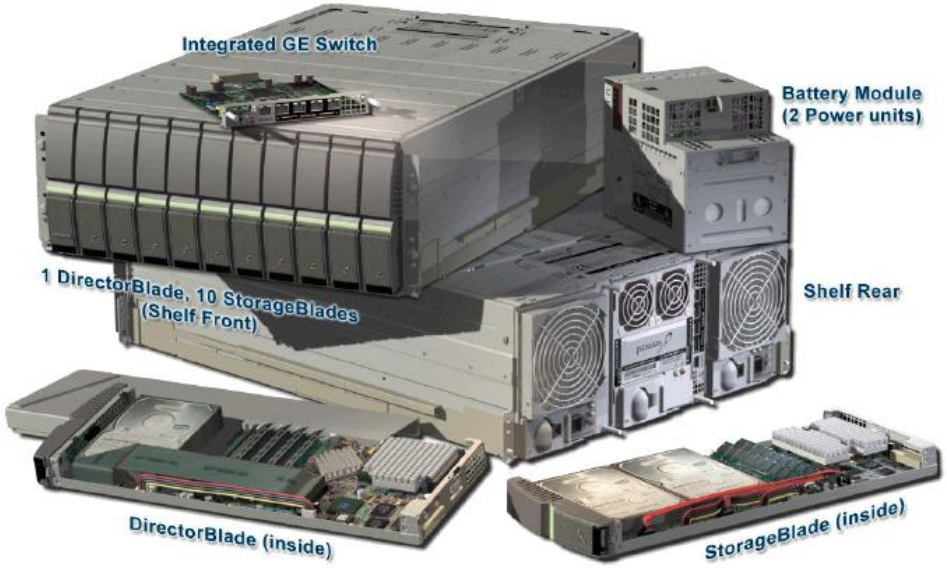
Theoretical peak performance:
- 13600 MB/s read
- 10400 MB/s write

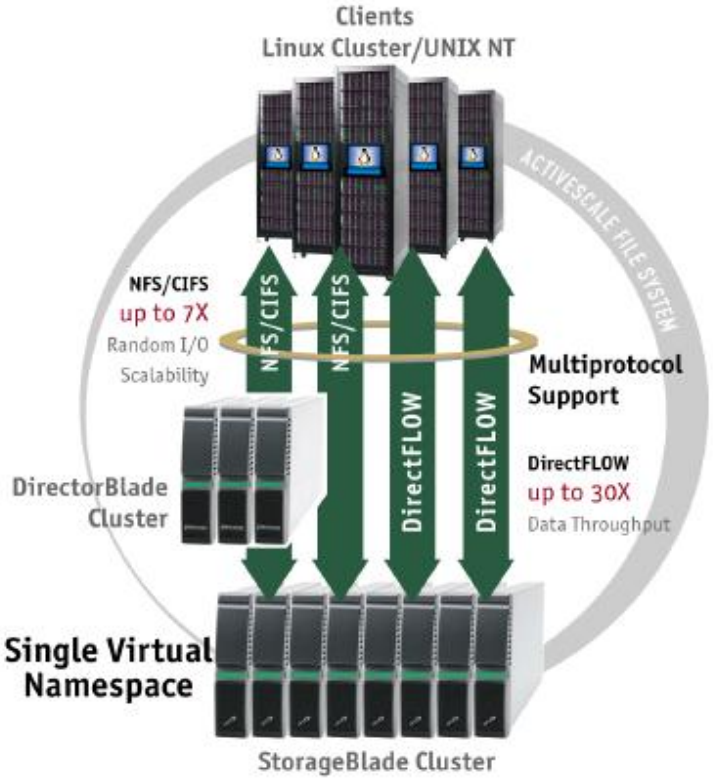
Total RAW capacity:
- 576 TB SATA (data)
- 3.6 TB SAS (metadata)

576 TB – 10,2 GB/s utilizzando midrange storage & Lustre parallel file system

PANASAS Active Stor

- High Performance NAS
- Parallel file system (pnfs)
- Easy to manage
- Based on appliances





PANASAS Active Stor

ActiveStor Appliance Model	ActiveStor 11	ActiveStor 12
Supported Blade Configurations (Director Blade + Storage Blade)	1+10, 2+9, or 3+8. Also 0+11 for expansion.	1+10 or 2+9. Also 0+11 for expansion.
Capacity (TB)*	40 or 60	40 or 60
Hard Drives (3.5" SATA)*	20	20
ECC Memory (GB of Cache)*	48	92
Max. Throughput, Write*	950MB/s	1600MB/s
Max. Throughput, Read*	1150MB/s	1500MB/s
Networking Switch Modules	One (second optional)	Two
Networking Uplinks per Switch Module	1 x 10GbE SFP+/CX4 or 4 x GbE Copper	1 x 10GbE SFP+/CX4 or 4 x GbE Copper
Additional Networking per Director Blade	2 x 10GbE SFP+	2 x 10GbE SFP+
Network Failover	Optional	Standard
High Availability Link Aggregation	No	Yes
QDR InfiniBand Router Compatibility	Yes	Yes
Power Supply	950W 1+1 redundant auto-switching 100–240VAC (47–63Hz)	950W 1+1 redundant auto-switching 100–240VAC (47–63Hz)
Power Consumption at 208VAC, Typical/Max. In-rush/Max.	4.8A/30A/7.0A	4.8A/30A/7.0A
Thermal Rating, Typical (BTU/hr)	3500	3500


Energy


Automotive


Aerospace


Bio Science
Pharma


Finance


University
Research


Government

EL COMPUTER ENGINEERING

Data Direct Network

SFA10K-X

- Up to 1200 SATA/SAS/SSDdrives
- 12GB/s
- Max Capacity: 3,6PB
- 8 x 40 Gb/s IB QDR
- 16 x 8Gb/s FC

SFA12K

- Up to 1680 SATA/SAS/SSD drives
- 40GB/s
- 1.4 M IOPS
- 16 x 56Gb/s IB, FC, 40Gb/Eth

Storage

DDN S2A900 (up to 6GB/s)
SAS + SATA 2TB
host Interface: FC 4/8 <-> 10GbE file server <-> clients

new DDN 10K: up to 12GB/s

E7412 disk server (~ 500 -1000MB/s)
HD: SATA 2TB
Host Interface: 10GbE <-> clients

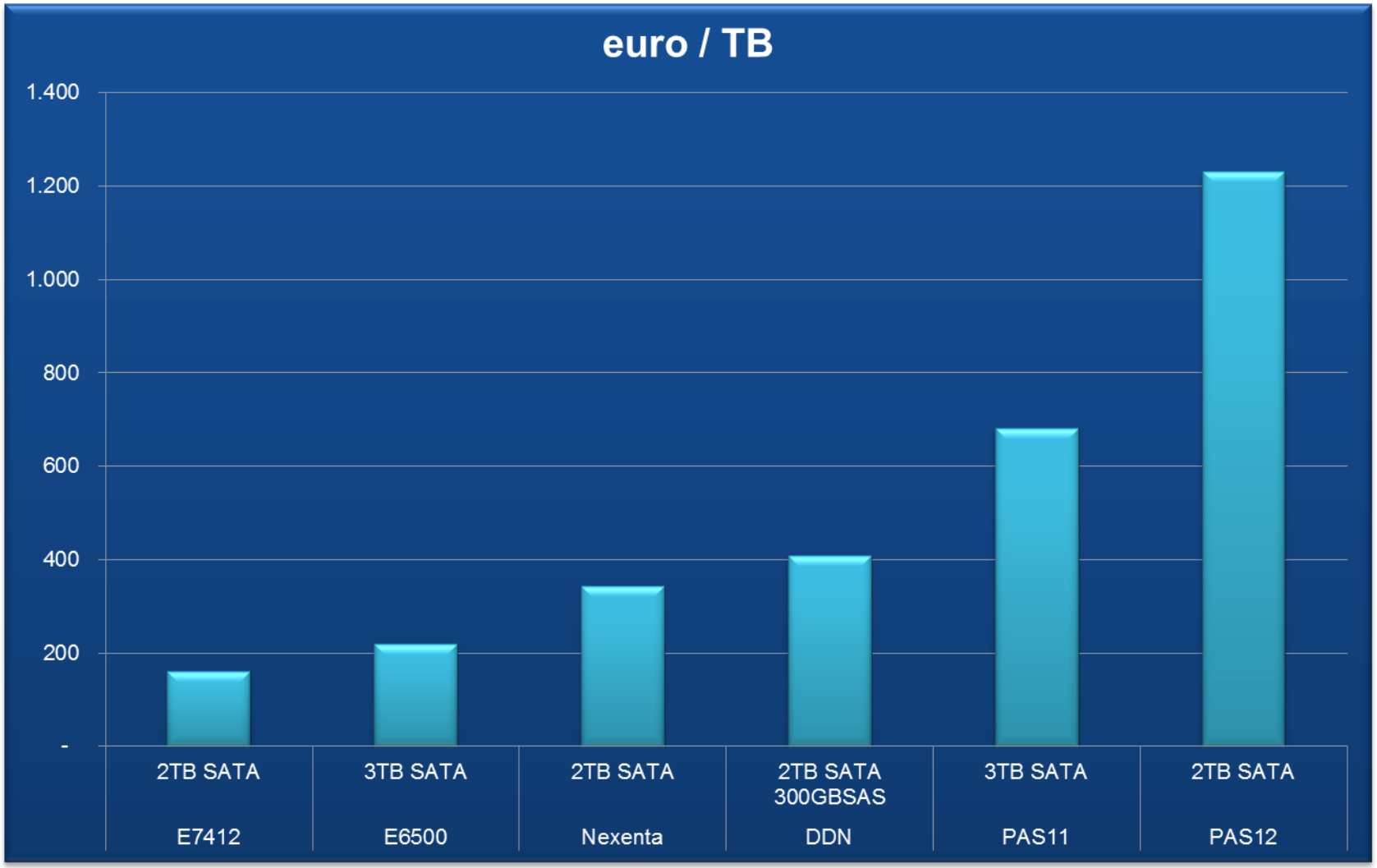
SAS JBOD + Nexenta (~600MB/s)
HD: SATA 2TB
Host Interface: 10GbE <-> clients

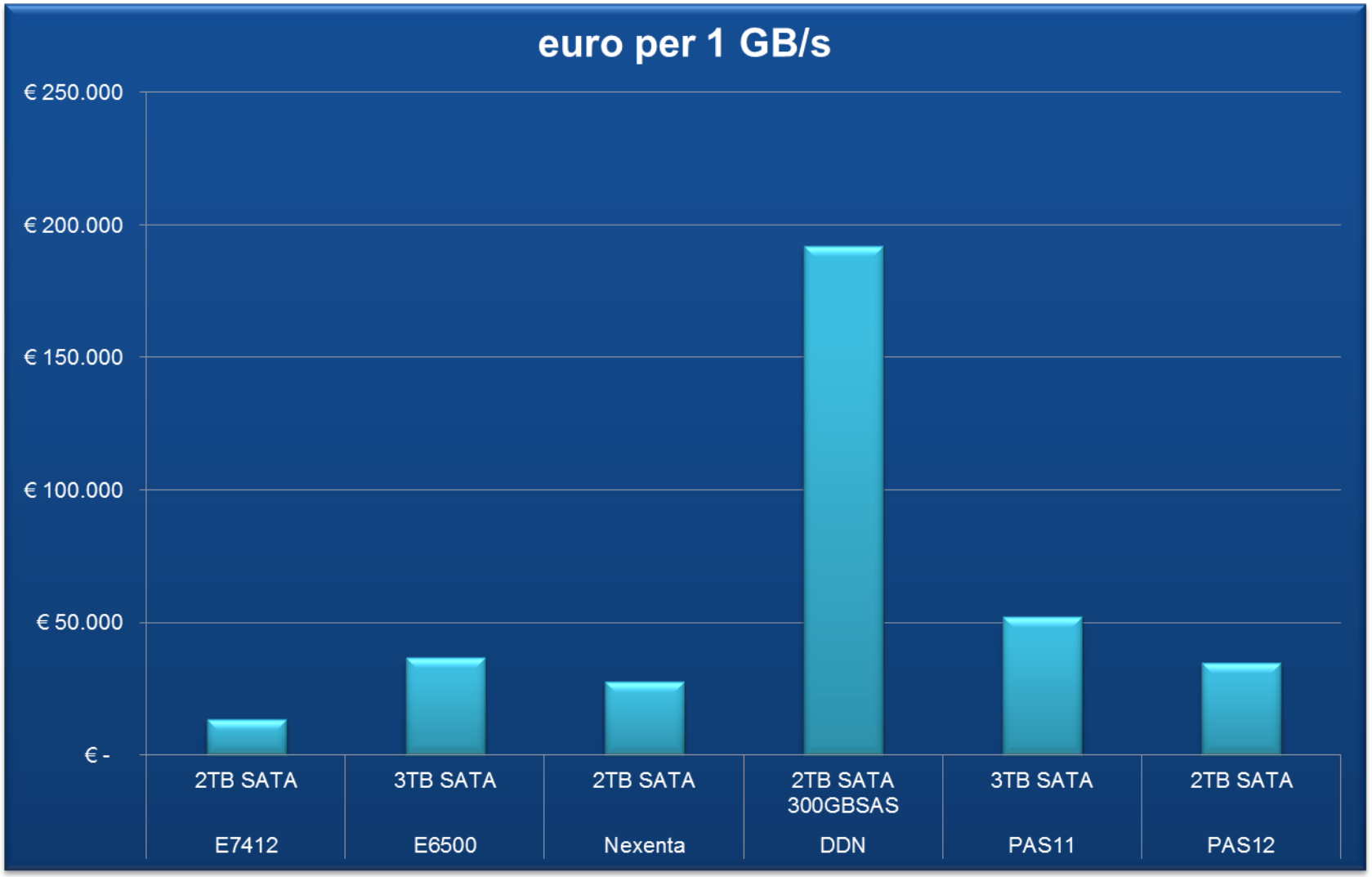
PAS11 (1GB/s per shelf)
SATA 3TB
host Interface: 10GbE <-> IB QDR file server <-> clients

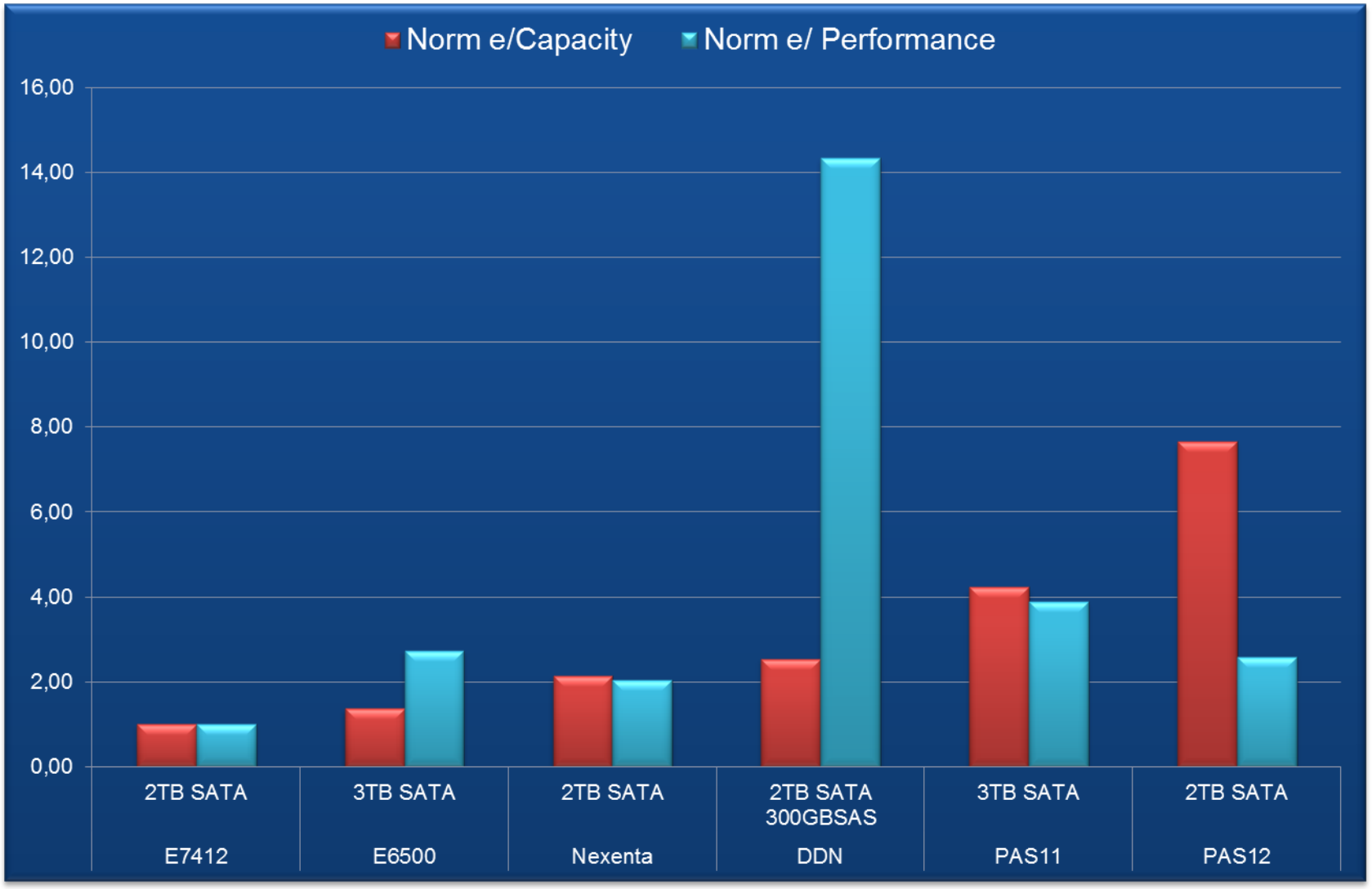
PAS12 (1,5 GB/s per shelf)
SATA 3TB
host Interface: 10+10GbE <-> IB QDR file server <-> clients

E6500 storage (up to 1.5GB/s per system)
HD: SATA 3TB
Host Interface: FC 8 <--> 10GbE file servesr <-> clients

new E6500 “Nitro” (up to 3 GB/s per system)


euro / TB
1.400
1.200
1.000
800
600
400
200
-
2TB SATA
E7412
3TB SATA
E6500
2TB SATA
Nexenta
2TB SATA
300GBSAS
DDN
3TB SATA
PAS11
2TB SATA
PAS12


euro per 1 GB/s
€ 250.000
€ 200.000
€ 150.000
€ 100.000
€ 50.000
€ -
2TB SATA
E7412
3TB SATA
E6500
2TB SATA
Nexenta
2TB SATA
300GBSAS
DDN
3TB SATA
PAS11
2TB SATA
PAS12


Norm e/Capacity
Norm e/ Performance
16,00
14,00
12,00
10,00
8,00
6,00
4,00
2,00
0,00
2TB SATA
E7412
3TB SATA
E6500
2TB SATA
Nexenta
2TB SATA
300GBSAS
DDN
3TB SATA
PAS11
2TB SATA
PAS12

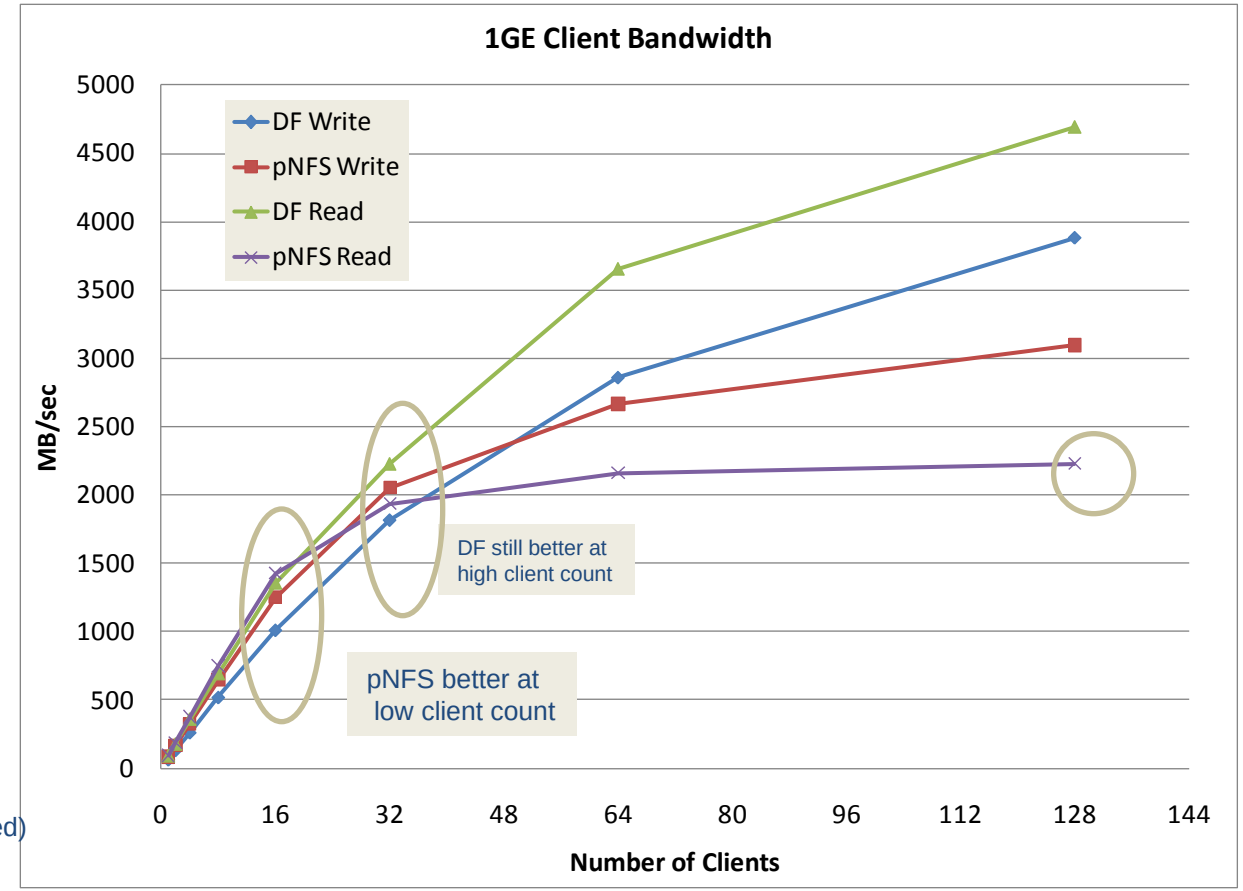
pNFS

2.6.XX

250 patches of remaining pNFS functionality divided into 4 batches by the developers and maintainers

Remaining files-based pNFS patches
Object-based pNFS
Block-based pNFS

Push from RedHat for 6.1 and beyond



Test Setup

Client
Fedora 12
pnfs-all-2.6.36.rc6-2010-10-7 (patched)
pnfs-utils-1-2-3

pNFS Metadata Server
Fedora 12
pnfs-all-2.6.36.rc6-2010-10-7 (patched)
pioneer-bh-546439 (Panasas gateway)

Panasas storage cluster
PanFS 4.0.0-540019

N-to-N streaming I/O

Acceleratori

GPU

x86_64
many core

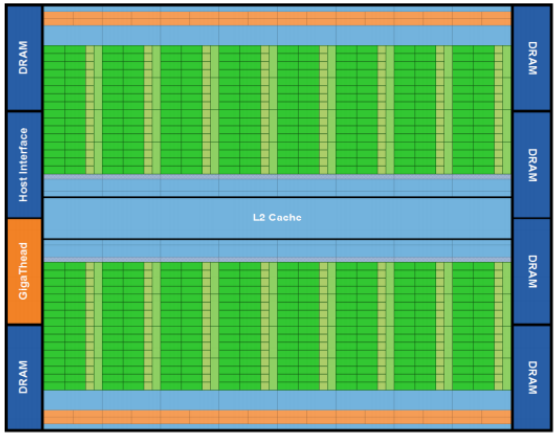
FPGA

NVIDIA

intel inside
Xeon Phi

CONVEY computer

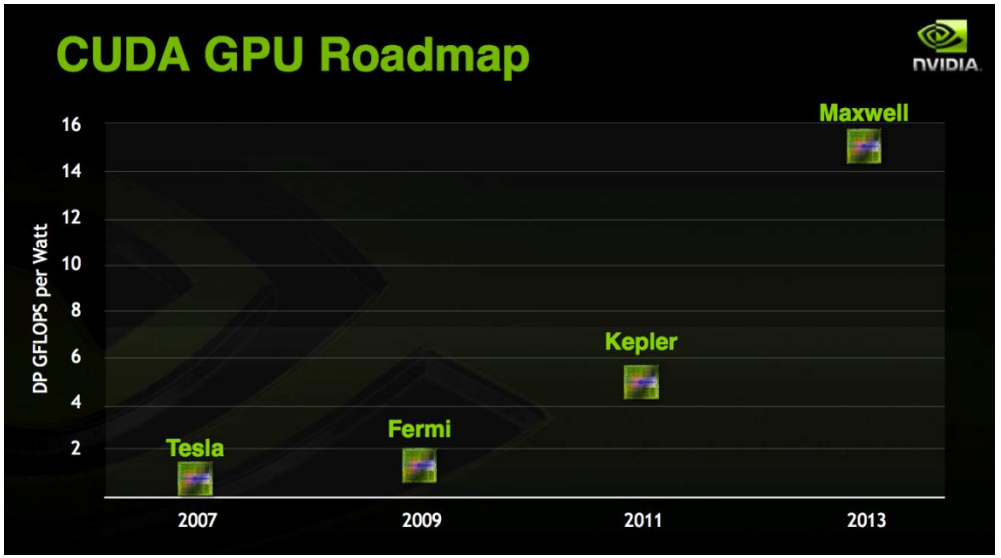
NVIDIA: Fermi Architecture



up to 512 CUDA cores (16 SM x 32 cores)
3/6 GB ECC
NVIDIA GigaThread™
NVIDIA Parallel DataCache™ Technology
4x DP (665 GFLOPS peak) Vs TESLA
2x SP (1,3 TFLOPS peak) Vs TESLA

Kepler K20
x3 GFLOPS/watt

Kepler K10
4+ TFLOS SP



NVIDIA Tesla™ 20-series

M

PRODUCT SPECIFICATIONS			
	Tesla M2090	Tesla M2070/M2075	Tesla M2050 and S2050
Peak double precision floating point performance	665 Gigaflops	515 Gigaflops	515 Gigaflops
Peak single precision floating point performance	1331 Gigaflops	1030 Gigaflops	1030 Gigaflops
CUDA cores	512	448	448
Memory size (GDDR5)	6 GigaBytes	6 GigaBytes	3 GigaBytes
Memory bandwidth (ECC off)	177 GBytes/sec	150 GBytes/sec	148 GBytes/sec

PRODUCT SPECIFICATIONS			
	Tesla C2075	Tesla C2070	Tesla C2050
Peak double precision floating point performance	515 Gigaflops	515 Gigaflops	515 Gigaflops
Peak single precision floating point performance	1030 Gigaflops	1030 Gigaflops	1030 Gigaflops
CUDA cores	448	448	448
Memory size (GDDR5)	6 GigaBytes	6 GigaBytes	3 GigaBytes
Memory bandwidth *(ECC off)	144 GBytes/sec	144 GBytes/sec	144 GBytes/sec
Display support	Dual-Link DVI Max display resolution 1600x1200	Dual-link DVI-I Max display resolution @ 60 Hz: 2560x1600	

C

NVIDIA Tesla™ 20-series

Express

2 PCIe x16 Gen2 connections

Extreme

8 PCIe x16 Gen2 connections

vCORE Family	vCORE Model	# GPU	RAM	CUDA cores	TFLOPS DP
Express	Lite **2075**	2 x M2075	12 GB	896	1
Express	**2075**	4 x M2075	24 GB	1792	2
Express	**2090**	4 x M2090	24GB	2048	2,66
Extreme	**2075**	4-16 x M2075	96GB (up to)	7168	8,2
Extreme	**2090**	4-16 x M2090	96GB (up to)	8192	10,64

Code developement / porting

Approaches

- Developement (new software)
- GPU ready libraries
- Directive based tools

Tool / Development environment

	Platform	Programming Model	Parallel Libraries	Development Tools
New in CUDA 4.0	GPUDirect™ (v 2.0) Peer-Peer Communication	Unified Virtual Addressing C++ new/delete C++ Virtual Functions	Thrust C++ Library Templated Performance Primitives Library	Parallel Nsight Pro 1.5
	Hardware Features ECC Memory Double Precision Native 64-bit Architecture Concurrent Kernel Execution Dual Copy Engines 6GB per GPU supported	**C support** NVIDIA C Compiler CUDA C Parallel Extensions Function Pointers Recursion Atomics malloc/free	**NVIDIA Library Support** Complete math.h Complete BLAS Library (1, 2 and 3) Sparse Matrix Math Library RNG Library FFT Library (1D, 2D and 3D) Video Decoding Library (NVCUVID) Video Encoding Library (NVCUVENC) Image Processing Library (NPP) Video Processing Library (NPP)	**NVIDIA Developer Tools** Parallel Nsight for MS Visual Studio cuda-gdb Debugger with multi-GPU support CUDA/OpenCL Visual Profiler CUDA Memory Checker CUDA Disassembler GPU Computing SDK NVML CUPTI
	Operating System Support MS Windows 32/64 Linux 32/64 Mac OS X 32/64	**C++ support** Classes/Objects Class Inheritance Polymorphism Operator Overloading Class Templates Function Templates Virtual Base Classes Namespaces	**3rd Party Math Libraries** CULA Tools (EM Photonics) MAGMA Heterogeneous LAPACK IMSL (Rogue Wave) VSIPL (GPU VSIPL)	**3rd Party Developer Tools** Allinea DDT RogueWave /Totalview Vampir Tau CAPS HMPP
	Designed for HPC Cluster Management GPUDirect Tesla Compute Cluster (TCC) Multi-GPU support	**Fortran support** CUDA Fortran (PGI)		

CUDA toolkit

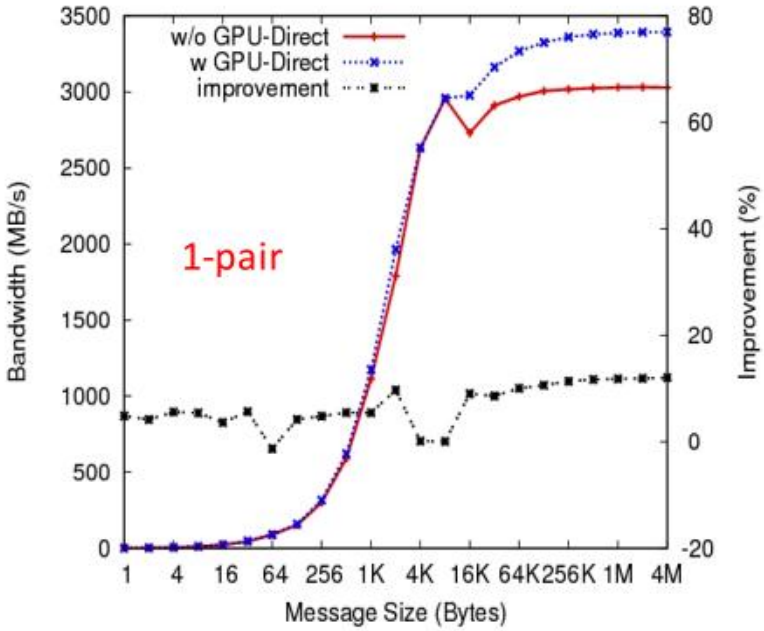
C/C++ compiler
CUDA-GDB debugger
Visual Profiler
GPU-accelerated **BLAS** library
GPU-accelerated **FFT** library
GPU-accelerated **Sparse Matrix** library
GPU-accelerated **RNG** library
(NVIDIA® Parallel Nsight™)

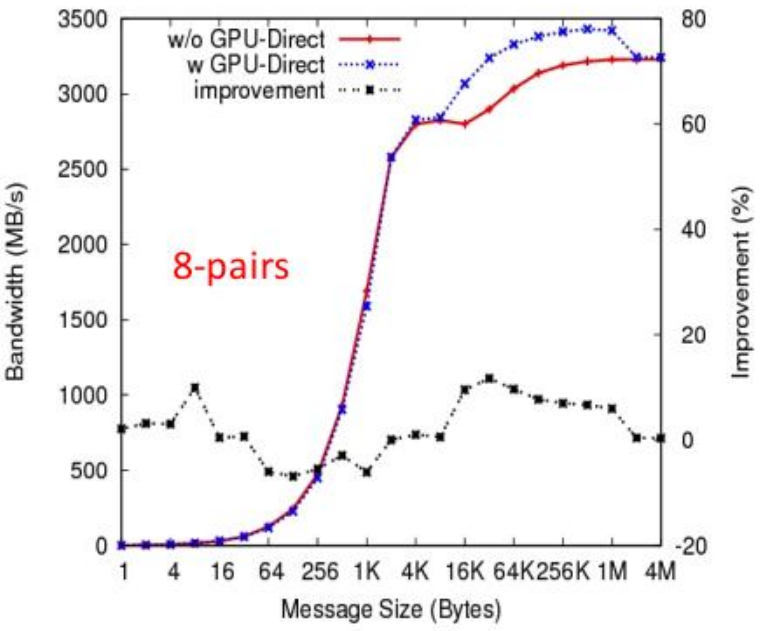
New in CUDA 4.0:

- **NVIDIA GPUDirect™ 2.0**
- Unified Virtual Addressing (UVA)
- Condivisione multi-thread delle GPU
- Condivisione multi-GPU di un singolo thread

NVIDIA GPUDirect™ 1.0

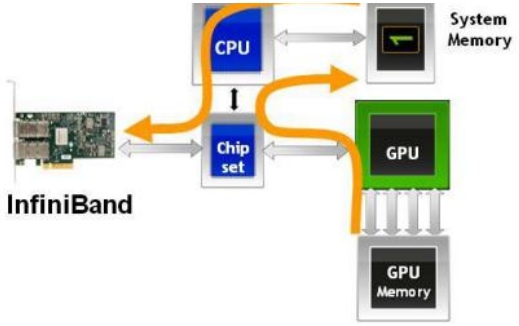
Comunicazione GPU to GPU attraverso IB





GPU direct: host buffer comune a GPU e HCA IB

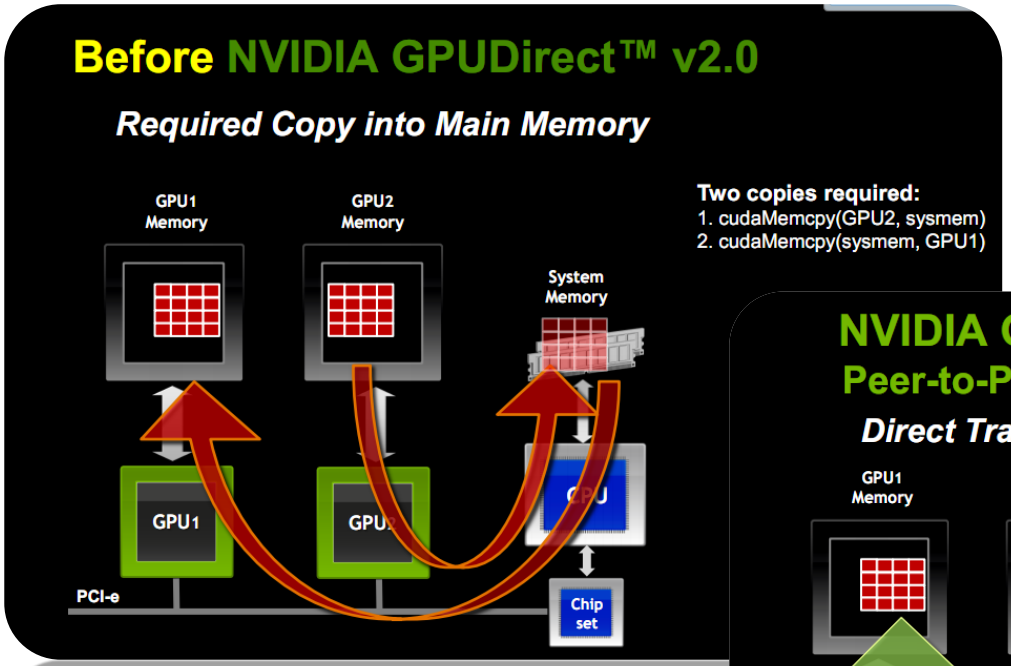
Incrementi fino al 12%

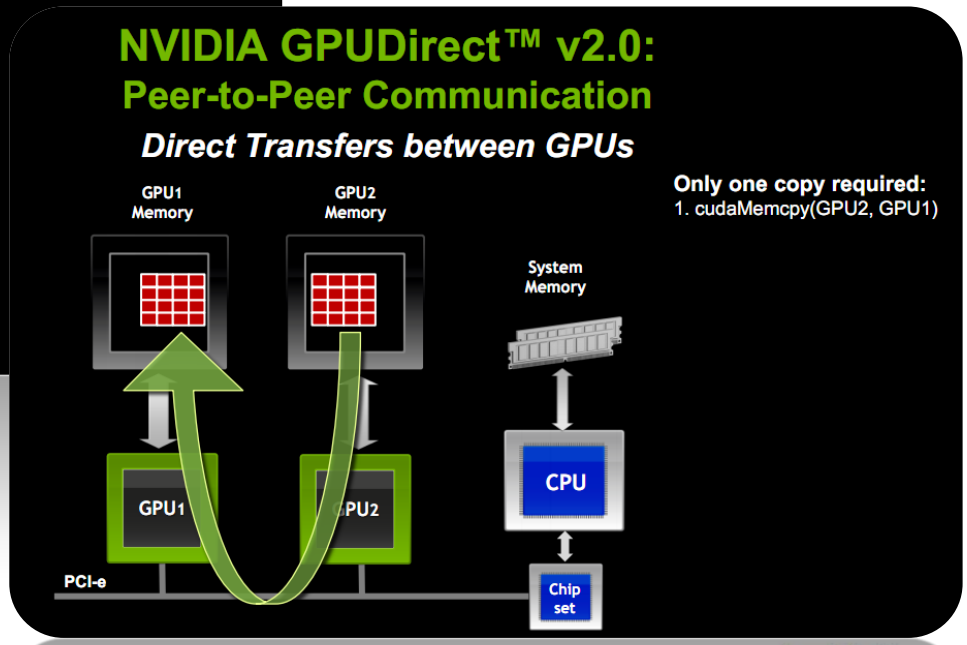


Fonte: Dhabaleswar K. (DK) Panda, Sayantan Sur (The Ohio State University) – Future of MPI, Tutorial at HPC Advisory Council Workshop, Lugano 2011

NVIDIA GPUDirect™ 2.0

Peer to peer commonication (GPU to GPU) over PCIe BUS





OpenCL

CAPS

HMPP Workbench

Portland Group

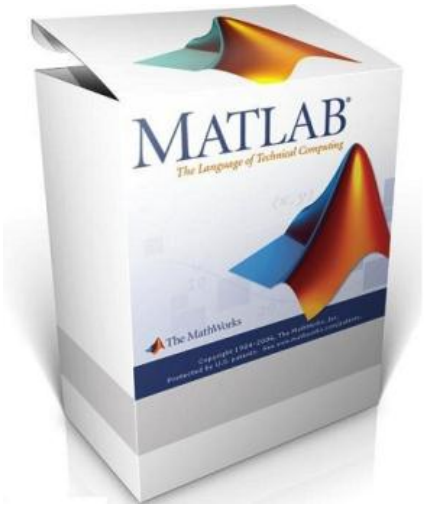
CUDA Fortran

PGI Accelerator Compilers

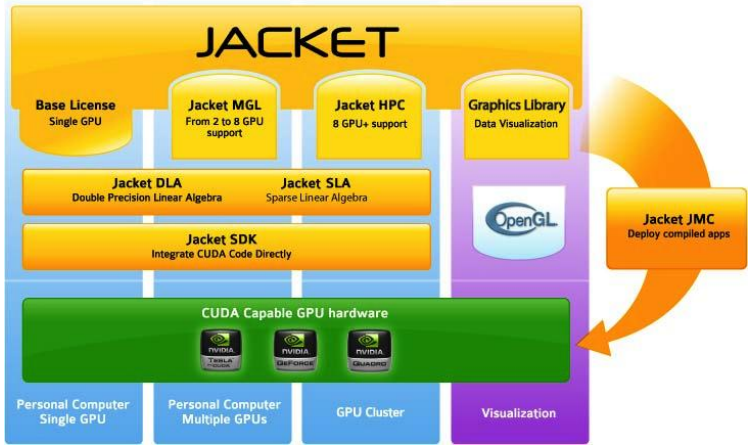
CUDA-x86

pyCUDA
JCUDA
CUDA.NET
Microsoft Direct Compute

MATLAB & GPU



Accelerys Jacket



GPU support is now included in Parallel Compute Toolbox

Systems for GPU computing

Workstations

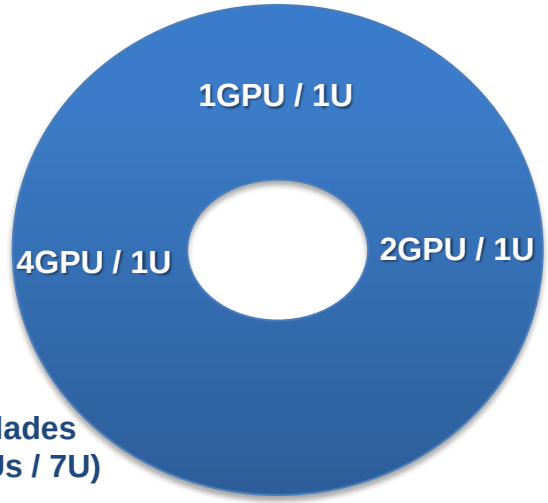
GPU optimized servers

vCORE external boxes

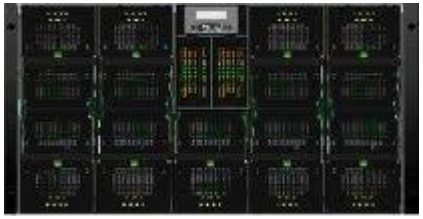
- #GPU / system
- # cores /GPU
- Shared / dedicated PCI bus
- # PCIe lanes

Hybrid systems

E7130



GPU blades
(18 GPUs / 7U)



GPU price / power considerations

4 CPU
0 GPU

High Density GPU

2 CPU
4 GPU

GPU optmized

4 CPU
4 GPU

Upgrade

4 CPU
4 GPU

GPU price / power considerations

x 3,2

x 3,2

High Density GPU

2 CPU
4 GPU

GPU optmized

x 2,9

x 2,8

4 CPU
4 GPU

4 CPU
0 GPU

Upgrade

x 3,6

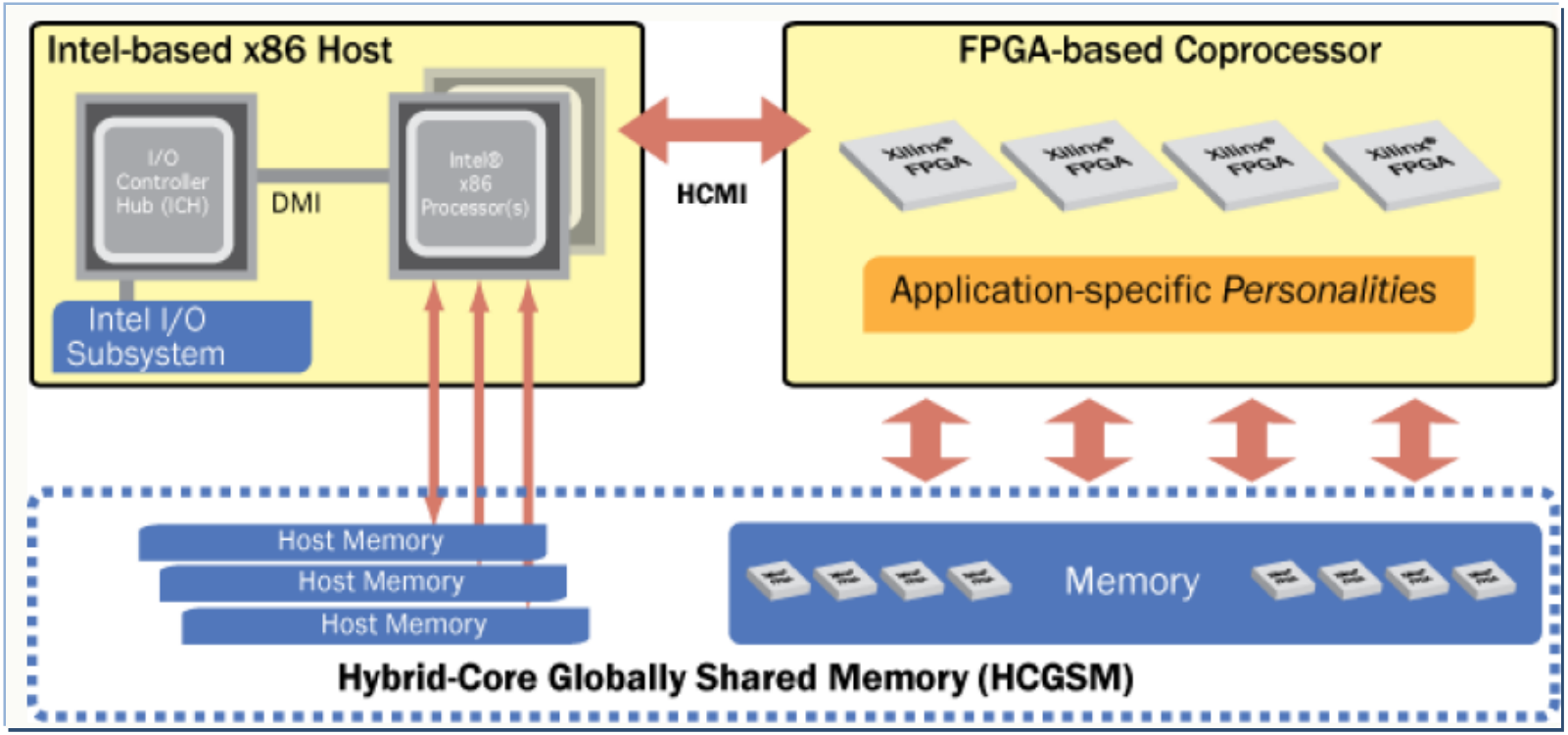
x 2,8

4 CPU
4 GPU

Prezzo

Prezzo + PWR
+ cooling

FPGA



CONVEY computer™

Cluster management suite

Hardware

↓

pre-checked/ configured Cluster

↓

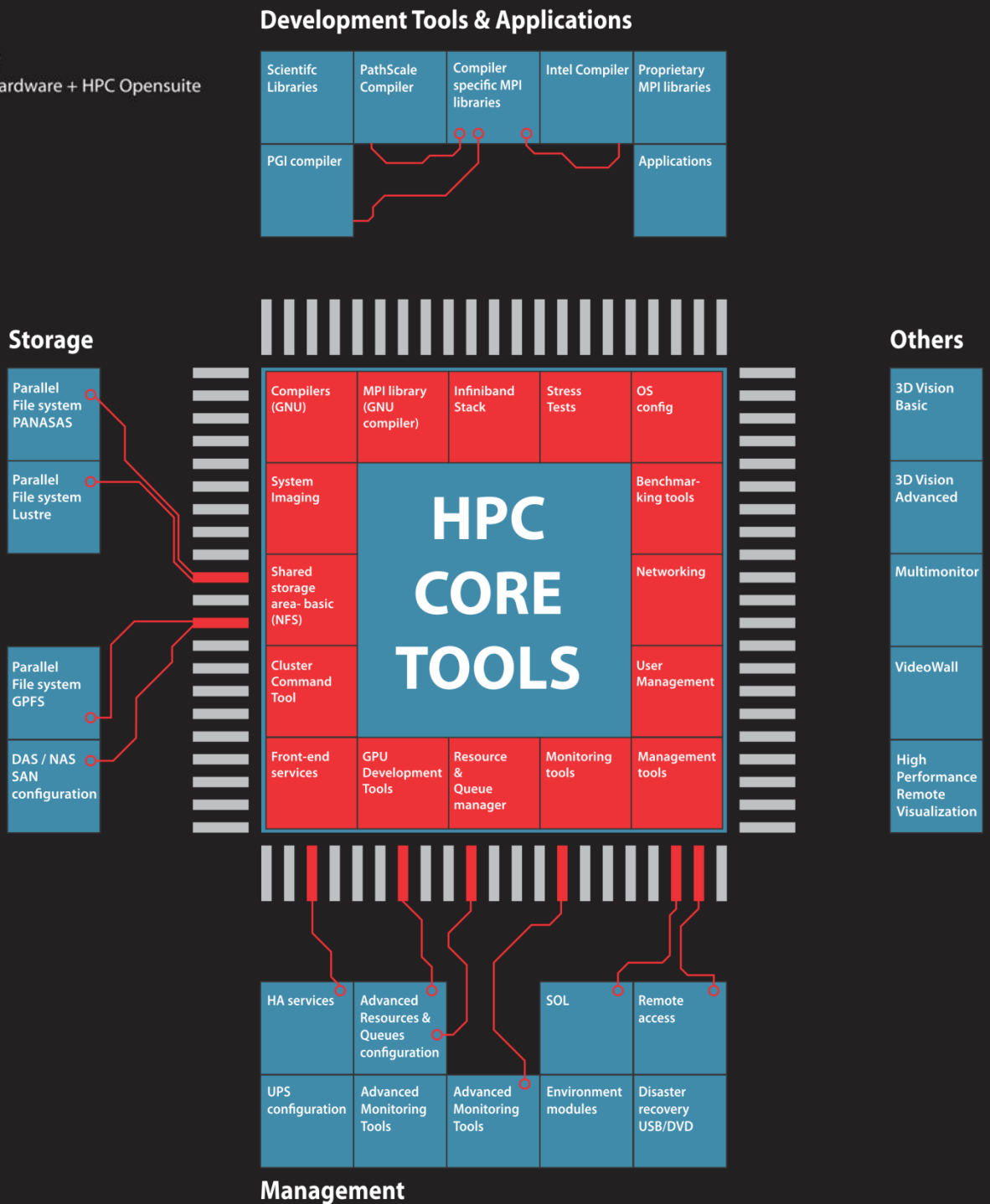
HPC Opensuite → **INTEL Cluster Ready**

↓

Enterprise class CMT: Concurrent COMMAND

HPC OpenSuite

The perfect Cluster recipe: E4® Hardware + HPC Opensuite



Key Features and Benefits:

- **Open source tools**
- **Linux based**
- **Easy cluster management and monitoring**
- **Support for GPU computing with CUDA ad OpenCL libraries and GPU monitoring and management**
- **Development environment for parallel applications with compilers, libraries and debuggers**
- **Support for high performance networks**
- **Support and deploy of 3D video solutions**
- **Fully Customizable**

E4 HPC OpenSuite simplifies and automates the installation, management and monitoring of HPC clusters.
E4 HPC OpenSuite is the perfect tool for every high performance computing scenario, from computational chemistry up to digital content creation.
The suite is made of a set of core tools, fundamental when building a cluster, as well as of various optional packages which may be installed on request and are fully customizable according to the user's requirements.

Feature	Standard	Premium
OS Installation & configuration	✔	✔
Node provisioning/deployment	✔	✔
User Management	✔	✔
Queue and Resource Manager	✔	✔
Compilers	✔	✔
Benchmark Tools	✔	✔
Front-End Services	✔	✔
Networking	✔	✔
Infiniband Stack & MPI Libraries	✔	✔
GPU development tools	✔	✔
Optional		
Scientific Libraries	✖	✔
Applications for HPC	✖	✔
Advanced Storage Area Configuration	✖	✔
HA Services	✖	✔
Remote and SOL access	✖	✔
Advanced Resource and Queue Configuration	✖	✔
Infrastructure management (UPS, rack sensors)	✖	✔
3D Configuration	✖	✔
Video Wall	✖	✔
Immersive 3D enviroment (cave & much more)	✖	✔
High performance Remote Visualization	✖	✔

Intel® Cluster Ready

Certified OS:
Red Hat, SLES, OpenSUSE, Scientifc Linux, CentOS, Fedora
Tools compliant with Intel Cluster Ready

EL COMPUTER ENGINEERING

Intel Cluster Ready

Hardware certification
+
Software certification

Benefit of ICR certification:

- Omogeneus software environment
- Performance check on all nodes
- Software compatibility

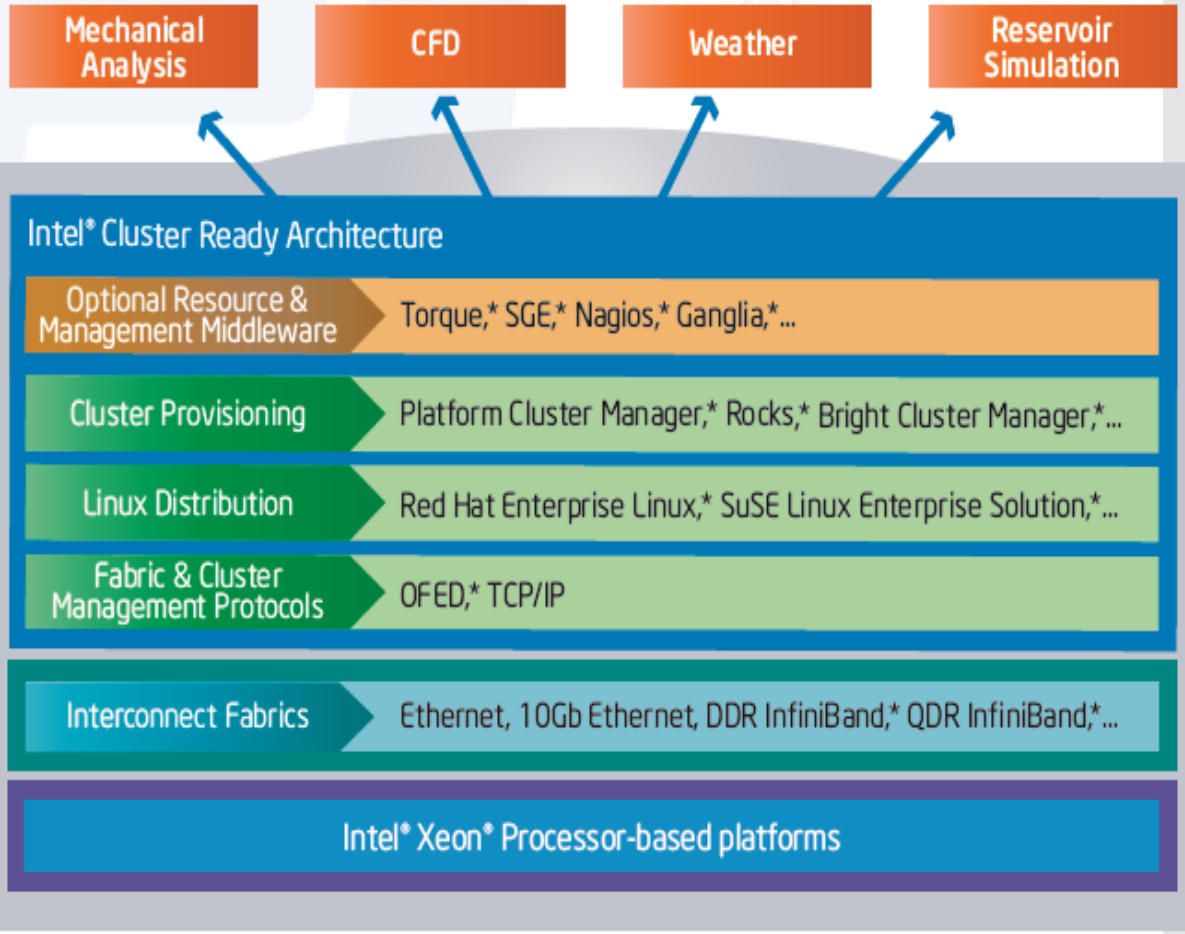
Intel Cluster Ready - Architecture

A wide range of industry-leading HPC applications

A choice of middleware and system software components that work together seamlessly

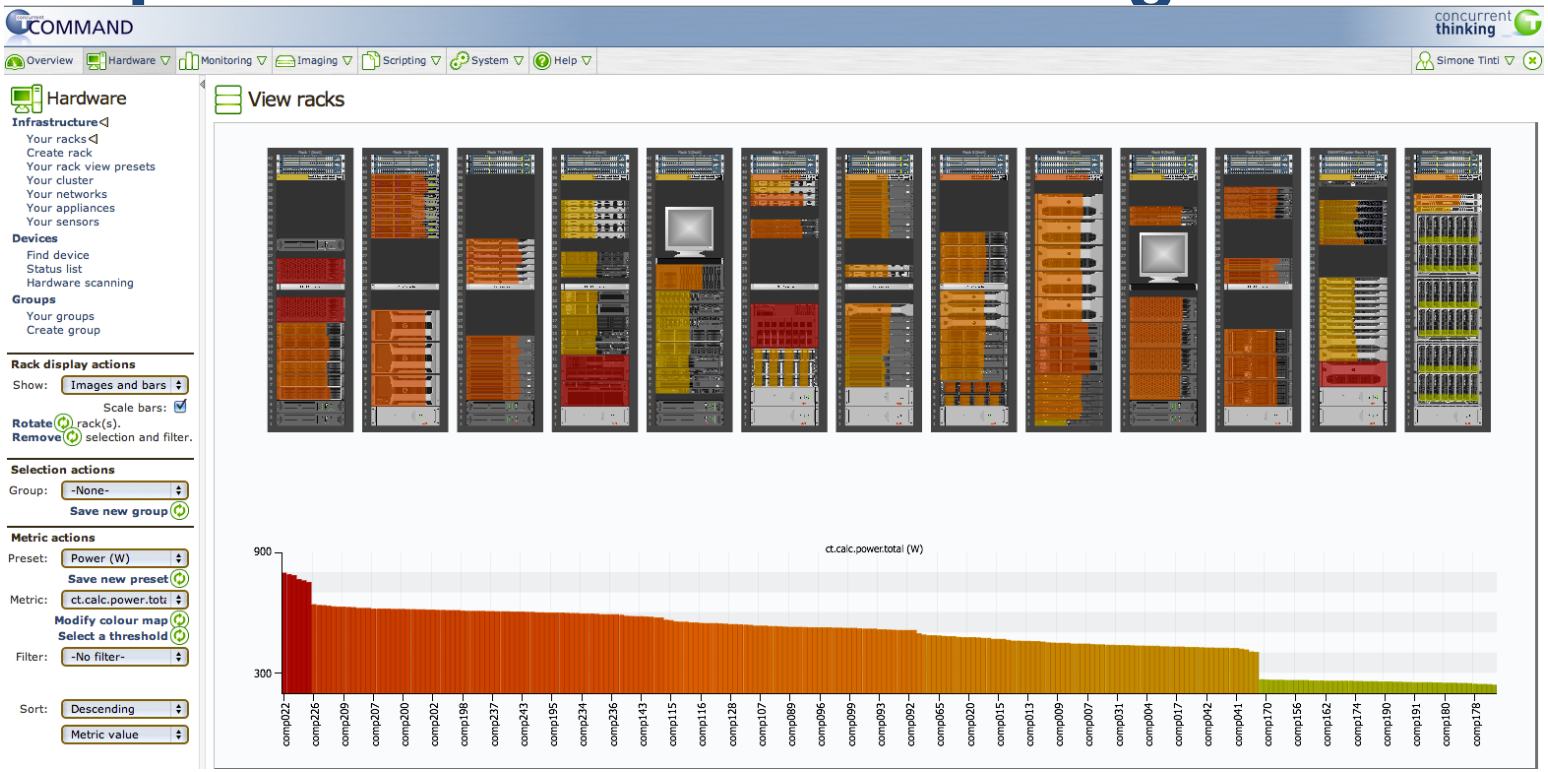
A broad selection of compatible inter-connect, storage and hardware options

The HPC industry's leading server architecture



Accelrys	Discovery Studio 2.0 Materials Studio 4.3 (see additional requirements)
Altair Engineering	OptiStruct* 11.0 RADIOSS 90 SA1 (see additional requirements) RADIOSS 10.0 (see additional requirements) AcuSolve 1.8a
ANSYS	ANSYS version 11 (including the MPI version Dansys)(see additional requirements) ANSYS FLUENT 6.3 (see additional requirements) ANSYS FLUENT 12.0 ANSYS FLUENT 13.0 ANSYS Mechanical 12.0.1 ANSYS Mechanical 13.0
CD-adapco	STAR-CD 4.02.001 (see additional requirements) STAR-CD 4.12 STAR-CCM+ 3.06 STAR-CCM+ 4.06 STAR-CCM+ 5.02
EM Software & Systems - S.A. (Pty) Ltd	FEKO Suite 5.3 FEKO Suite 5.5
ESI Group	PAM-CRASH 2G v2007 (see additional requirements) PAM-CRASH 2009
Flow Science, Inc.	FLOW-3D/MP Version 3.0 FLOW-3D/MP Version 4.0 FLOW-3D/MP Version 4.1
GNS mbH	INDEED 8.0.4
ICON Technology & Process Consulting Ltd.	ICON FOAMpro 1.5 (see additional requirements)
Landmark Software and Services	SeisSpace/ProMAX (see additional requirements) Nexus R5000.0.2 VIP R5000.0.2
LSTC - Livermore Software Technology Corporation	LS-DYNA LS971.7600.398 MPP for Intel64 MPI LS-DYNA LS971.7600.1224 MPP for Intel64 MPI LS-DYNA/MPP R3 or contact LSTC directly (see additional requirements) LS-DYNA LS971R3.2/MPP for Intel MPI (single precision) or contact LSTC directly (see additional requirements) LS-DYNA LS971R4.2/MPP
MAGMA Giessereitechnologie GmbH	MAGMA V4.4-p19 (see additional requirements) MAGMA5
The Mathworks	Distributed Computing Toolbox 3.1 and Distributed Computing Engine 3.1
Metacomp Technologies, Inc.	CFD++ 7.1.1
MSC.Software	MD Nastran R2 MD Nastran R2.1 MD Nastran R3 MD Nastran 2010 MD Nastran 2010.1 MSC Nastran 2010.1 MSC Marc 2007 MSC Marc 2008 R1 MSC Marc 2010.1
Schlumberger	Petrel Seismic Server 2008.1 (see additional requirements)
Siemens PLM Software	NX Nastran, Version 5.1 (see additional requirements) NX Nastran, Version 7.1
simufact	simufact.forming 9.0 (see additional requirements) simufact.forming 10.0
SIMULIA	Abaqus/Explicit 6.7-1, 6.7-2, 6.7-3, 6.7-4, 6.8-1, 6.8-2, 6.8-3, 6.8-4, 6.8-EF1, 6.8-EF2, 6.9-1, 6.9-2, 6.9-3, 6.9-EF1, 6.9-EF2, 6.10-1, 6.10-2, 6.10-3, 6.10-EF1 Abaqus/Standard 6.7-1, 6.7-2, 6.7-3, 6.7-4, 6.8-1, 6.8-2, 6.8-3, 6.8-4, 6.8-EF1, 6.8-EF2, 6.9-1, 6.9-2, 6.9-3, 6.9-EF1, 6.9-EF2, 6.10-1, 6.10-2, 6.10-3, 6.10-EF1 Abaqus/CAE 6.7-4, 6.8-1, 6.8-2, 6.8-3, 6.8-4, 6.8-EF1, 6.8-EF2, 6.9-1, 6.9-2, 6.9-3, 6.9-EF1, 6.9-EF2, 6.10-1, 6.10-2, 6.10-3, 6.10-EF1 Abaqus/Viewer 6.7-4, 6.8-1, 6.8-2, 6.8-3, 6.8-4, 6.8-EF1, 6.8-EF2, 6.9-1, 6.9-2, 6.9-3, 6.9-EF1, 6.9-EF2, 6.10-1, 6.10-2, 6.10-3, 6.10-EF1
Software Cradle Co., Ltd.	SC/Tetra V7 STREAM V3.15
Wolfram Research, Inc.	gridMathematica 2.2

Enterprise class Cluster Management Suite



- Easy, powerfull, customizable
- Supplied with hardware appliances for load sharing
- Ideal for cluster and datacenter
- CO_2, power, energy monitoring