ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

CSCS
Swiss National Supercomputing Centre

Installation and Operational Needs of Multi-purpose GPU Clusters

Biozentrum, UNI-Basel, Room 106, 27 October 2011,

HPC Solutions
System Manager
Vincenzo Annaloro

Abstract

CSCS, in the last couple of years has started developing a basic experience on installing, configuring, managing, and operating multi-gpgpu cluster systems for several projects needs.

I will try to show you today some installation, configuration and operational choices adopted on one CSCS production visualization cluster and one upcoming GPGPU R & D cluster solution.

Index

1. DALCO Visualization, R&D Cluster EIGER

- Introduction
- System description overview
- Remote visualization service
- GPGPU computing service
- SLURM Resource Manager
- System Administration and Monitoring Tools
- Open Issues

Index

2. IBM iDataPlex R&D Cluster CASTOR

- Introduction
- Project goals
- Phase 1 & 2 project plans
- System description overview
- Cluster SW infrastructure

1. DALCO Visualization, R&D Cluster EIGER

Introduction

- First CSCS GPGPU production computing facility
- Q2 2010 integrated into the CSCS computing ecosystem
- opened to the swiss scientific user community:
  - Visualization/data analysis services
  - Remote Visualization and collaboration services
  - Hybrid multicore/multi-vendor GPU computing services
  - General purpose pre/post processing services

DALCO Visualization, R&D Cluster EIGER

System description overview (1)

- tightly coupled computing cluster system, IB QDR interconnect
- Novell SUSE Linux Enterprise Server 11 SP1 OS
- 19 nodes available to the users, 25 total in the cluster
- dual-socket six-cores AMD Opteron 2427 processor architecture running at 2.2 GHz (*Istanbul*)
- offering 24 GB DDR2 of main system memory per node
- total of 276* cpu cores and 648* GB aggregate memory
- 8 out of 19 cluster nodes offers a larger main system memory capacity up to 48 GB (*Fat Nodes*)
- 4 out of 19 cluster nodes offers dual-socket twelve-cores AMD Opteron 6174 processor architecutre (*Magny-cours*)

DALCO Visualization, R&D Cluster EIGER

System description overview (2)

Cluster nodes offer different GPU architectures **(NVIDIA CUDA vs AMD Radeon)**, with 1 or 2 GPUs per node of the same kind :

- **NVIDIA** GeForce GTX 285, 2 GB GDDR3
- **NVIDIA** Tesla S1070, 4 GB GDDR3
- **NVIDIA** GeForce Fermi GTX 480, 1.5 GB GDDR5
- **NVIDIA** Tesla Fermi M2050, 3 GB GDDR5
- **NVIDIA** Tesla Fermi C2070, 6 GB GDDR5
- **AMD** HD Radeon 6990, 4 GB GDD5

DALCO Visualization, R&D Cluster EIGER

System description overview (3)

Cluster network infrastructure :

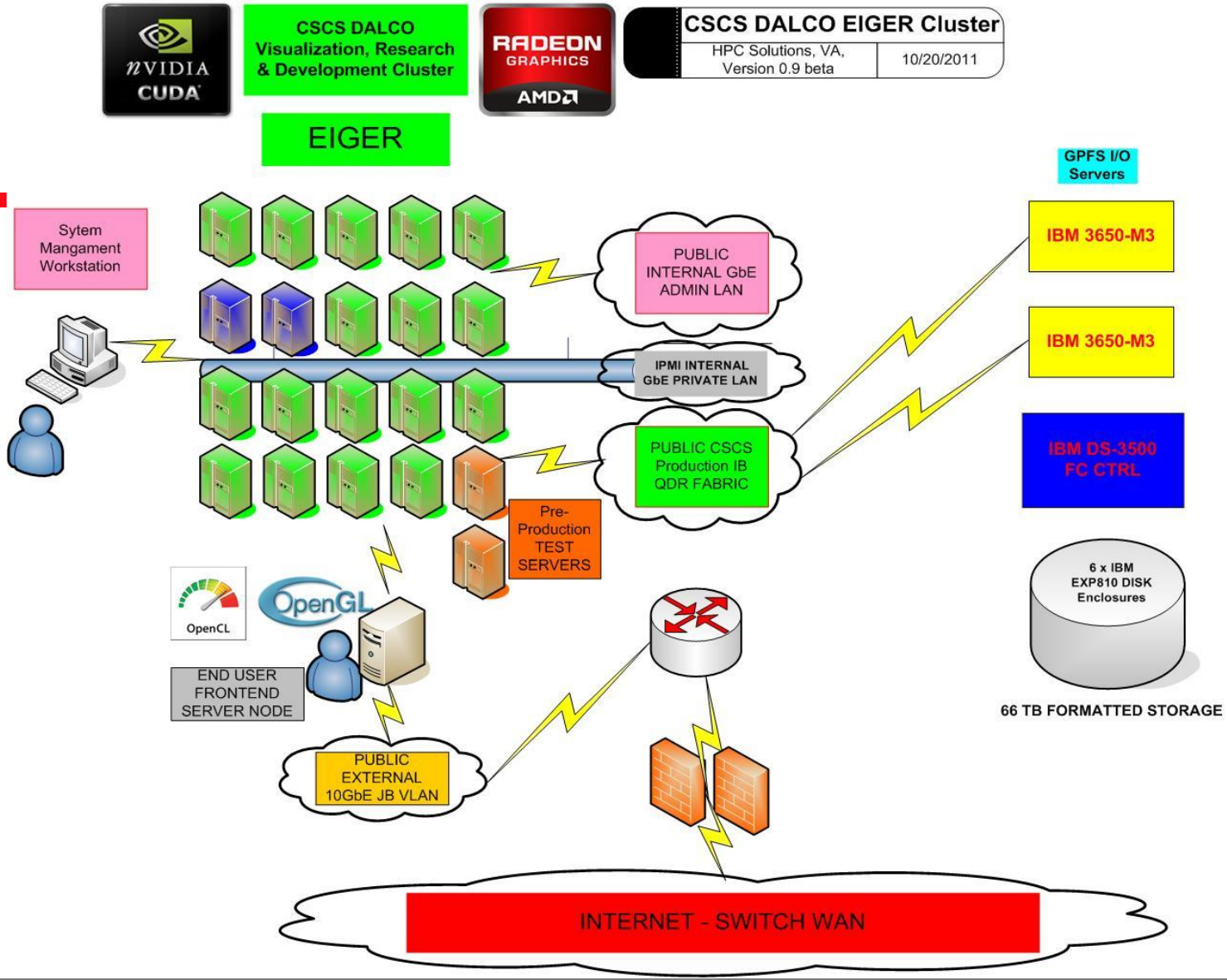
- dedicated **Infiniband QDR** fabric :
  - supporting both parallel-MPI traffic (*MVAPICH2 and OpenMPI*)
  - the internal parallel scratch file system I/O data traffic (*GPFS*)
- commodity **10 GbE JB-enabled login** LAN :
  - ensures interactive login access
- standard **1 GbE administration** LAN:
  - reserved for cluster management purposes (*IPMI VLAN*)

DALCO Visualization, R&D Cluster EIGER

System description overview (4)

Several **class of nodes** have been defined with dedicated system functionalities (roles) :

- Class 0: Administration Node (1x)
- Class 1: Login Node (1x)
- Class 2: Visualization Nodes (7x)
- Class 3: Fat Visualization Nodes (4x)
- Class 4: Advanced Development Nodes (4x)
- Class 5: Storage Nodes (2x)
- Class 6: Pre-production test nodes (2x)


nVIDIA CUDA
CSCS DALCO Visualization, Research & Development Cluster
RADEON GRAPHICS AMD
CSCS DALCO EIGER Cluster
HPC Solutions, VA, Version 0.9 beta
10/20/2011
EIGER
GPFS I/O Servers
Sytem Mangament Workstation
PUBLIC INTERNAL GbE ADMIN LAN
IPMI INTERNAL GbE PRIVATE LAN
PUBLIC CSCS Production IB QDR FABRIC
IBM 3650-M3
IBM 3650-M3
IBM DS-3500 FC CTRL
6 x IBM EXP810 DISK Enclosures
66 TB FORMATTED STORAGE
Pre-Production TEST SERVERS
OpenCL
OpenGL
END USER FRONTEND SERVER NODE
PUBLIC EXTERNAL 10GbE JB VLAN
INTERNET - SWITCH WAN

DALCO Visualization, R&D Cluster EIGER

Remote visualization service (1)

- **Remote interactive 2D/3D OpenGL application portfolio:**
  - ParaView, AVS Express, TecPlot, R, Mathematica, MATLAB, Gaussian 09, Chimera and VMD
  - VisIt : free interactive parallel visualization and graphical analysis tool for viewing scientific data
  - Other ISV OpenGL applications on request

- **Remote batch parallel rendering farm**
  - ParaView: open-source, multi-platform data analysis and visualization application

DALCO Visualization, R&D Cluster EIGER

Remote visualization service (2)

- **Enabling technologies :**
  - VirtualGL / TurboVNC / TurboJPEG
  - Server-side 3D HW GPU Accelerators
  - OpenGL APIs and libraries
  - High Performance Interconnect: Infiniband QDR fabric infrastructure

DALCO Visualization, R&D Cluster EIGER

Remote visualization service (3)

- VirtualGL :
  - *VirtualGL* is an open source package which gives any Unix or Linux remote display software the ability to run OpenGL applications with full 3D hardware acceleration (**remote OpenGL application**)
  - the OpenGL commands and 3D data are instead redirected to a 3D graphics accelerator on the application server, and only the rendered 3D images are sent to the client machine (**3D hardware virtualization**)
  - *VirtualGL* also allows 3D graphics hardware to be shared among multiple users

DALCO Visualization, R&D Cluster EIGER

Remote visualization service (4)

- TurboVNC/TurboJPEG/libjpegturbo :
  - a free remote control software package, which allows you to **see the desktop of a remote machine and control it** with your local mouse and keyboard, just like you would do it sitting in the front of that computer
  - TurboVNC was developed to **address the performance limitations** of off-the-shelf VNC and performs server-side X11 rendering, so that only images are sent to the client,and thus it performs very well on high-latency networks
  - TurboVNC is based on TightVNC and borrows part of TightVNC's hybrid image encoding scheme, allowing it to use the most efficient image encoding method for each individual tile. TurboVNC **also provides rudimentary built-in collaboration capabilities**.

DALCO Visualization, R&D Cluster EIGER

Remote visualization service (5)

- Server-side 3D HW GPU Accelerators :
  - NVIDIA Tesla FERMI C2070, 515 Gigaflops of double-precision peak performance in the GPU, SP peak performance is over a Teraflop per GPU, 6 GB GDD5 memory clocked at 1.5 Ghz, 144 GB/s memory BW, 448 CUDA cores.

  - Dual-gpu AMD HD RADEON 6990 card is configured with a total of 4 GB GDDR5 memory clocked at 1250MHz, ensuring up to 320 MB/s gpu memory bandwidth, and providing a total of 3072 Stream Processors

DALCO Visualization, R&D Cluster EIGER

GPGPU computing service (1)

Provided by multi-vendor GPU architectures and custom vendor software solutions :

- **NVIDIA CUDA-enabled GPUs :**
  - NVIDIA CUDA Toolkit V 4.0
  - NVIDIA CUDA SDK V 4.0
  - NVIDIA GPU Driver V 270.41.19 (testing with 280.13 and **285.05.09** **Certified)**

- **AMD Radeon GPUs :**
  - AMD Catalyst GPU Driver V 11.9
  - AMD Accelerated Parallel Processing (APP) SDK (APP-SDK) V 2.5
  - AMD Core Math Library for Graphic Processors (ACML-GPU)
  - AMD Accelerated Parallel Processing Math Libraries (APPML)
  - AMD APP Profiler

DALCO Visualization, R&D Cluster EIGER

GPGPU computing service (2)

NVIDIA CUDA Toolkit provides a compiler for NVIDIA GPUs, math libraries, and tools for debugging and optimizing the performance of end user applications :

- **CUDA Libraries**
  - cuFFT, cuBLAS, cuSPARSE, cuRAND, NPP, Thrust
- **Development Tools**
  - NVIDIA CUDA C/C++ Compiler (NVCC)
  - Visual Profiler, CUDA-GDB Debugger, CUDA-MEMCHECK
- **Developer Documentation**
  - Getting Started Guides, Release Notes, and more..
  - CUDA C Programming Guide, CUDA Best Practices Guide
  - OpenCL Programming Guide, OpenCL Best Practices Guide
- **Support** for Windows, **Linux** and MacOS

DALCO Visualization, R&D Cluster EIGER

GPGPU computing service (3)

AMD APP Software Development Kit (SDK) :

- Complete development platform created by AMD to allow you to quickly and **easily develop applications** accelerated by AMD APP technology, (AMD Accelerated Parallel Processing (APP) Technology)
- The AMD Accelerated Parallel Processing system includes a **software stack** and the AMD GPUs.
- The AMD SDK allows you to develop your applications in a high-level language, **OpenCL™** V 1.1 (Open Computing Language)

DALCO Visualization, R&D Cluster EIGER

GPGPU computing service (4)

Development environment provided:

- Compilers :
  - PGI Accelerator C/C++/Fortran V 11.9
  - Intel Composer C/C++/Fortran XE 2011
  - EKOPath PathScale V 4.x
  - GNU C/C++/Fortran V 4.6.1
  - Open64 V 4.2.5.2

- Integrated Development Environment (IDE):
  - Eclipse IDE for C/C++ Developers with incubating components INDIGO V 3.7.1
  - Eclipse IDE support for CUDA via plugin : the **Eclipse CUDA Plugin** is the first IDE available on Linux for CUDA Development, it allows you to develop, debug, and run CUDA applications using the Eclipse IDE

DALCO Visualization, R&D Cluster EIGER

GPGPU computing service (5)

Development environment provided:

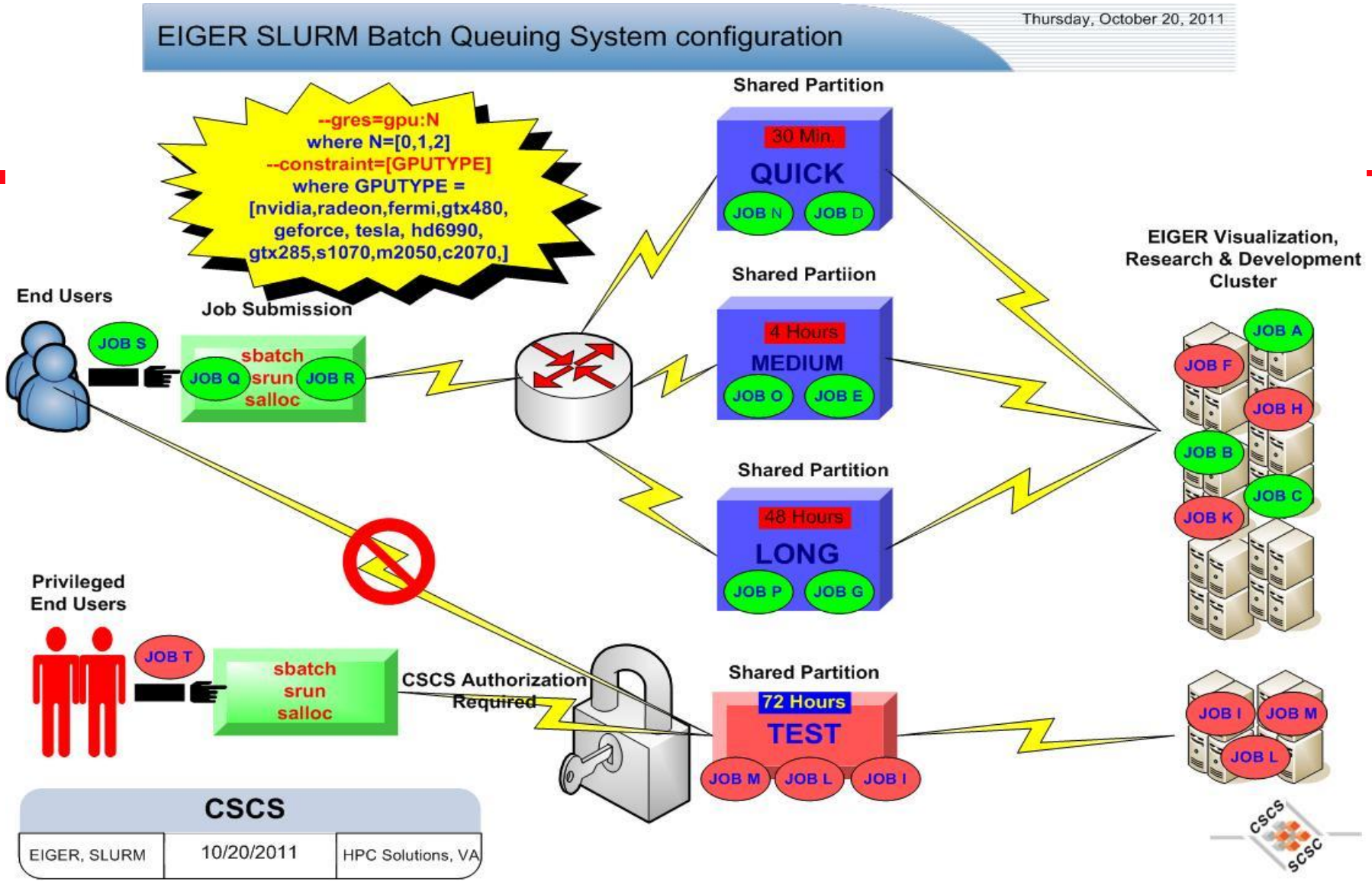
- Commercial Debugging tools (under evaluation) :
  - TotalView V 8.9.2 support for CUDA V 4.0
  - Allinea DDT with CUDA : Allinea DDT is able to provide application developers with a single tool that can debug hybrid MPI, OpenMP and CUDA applications on a single workstation or GPU cluster
  - NVIDIA *cuda-memcheck* and *cuda-gdb*
- Profiling tools :
  - Nvidia Visual Profiler (NVIDIA *cudaprof* and *openglprof)*
  - Nvidia Parallel Nsight (Windows only)
  - ATI Stream Profiler

DALCO Visualization, R&D Cluster EIGER

SLURM Resource Manager (1)

Batch queuing system / resource manager:

- SLURM V 2.3.0-pre5 versus other BQS :
  - No licensing fees, developer support available
  - Highly customizable, extensible via plugin/APIs
  - Widely adopted within the whole CSCS Supercomputing Center
  - Fine-grain granularity in task placement
- Customized by CSCS for project budget-based CPU accounting and priority scheduling
- Job back-filling enabled
- Ad-hoc GPU job accounting, reporting, GPU resource enforcement currently under development


EIGER SLURM Batch Queuing System configuration
Thursday, October 20, 2011
--gres=gpu:N
where N=[0,1,2]
--constraint=[GPUTYPE]
where GPUTYPE =
[nvidia,radeon,fermi,gtx480,
geforce, tesla, hd6990,
gtx285,s1070,m2050,c2070,]
Shared Partition
30 Min.
QUICK
JOB N
JOB D
Shared Partiion
4 Hours
MEDIUM
JOB O
JOB E
Shared Partition
48 Hours
LONG
JOB P
JOB G
Shared Partition
72 Hours
TEST
JOB M
JOB L
JOB I
EIGER Visualization,
Research & Development
Cluster
JOB A
JOB F
JOB H
JOB B
JOB C
JOB K
JOB I
JOB M
JOB L
End Users
JOB S
Job Submission
sbatch
JOB Q
srun
JOB R
salloc
Privileged
End Users
JOB T
sbatch
srun
salloc
CSCS Authorization
Required
CSCS
EIGER, SLURM
10/20/2011
HPC Solutions, VA
CSCS
SCSC

DALCO Visualization, R&D Cluster EIGER

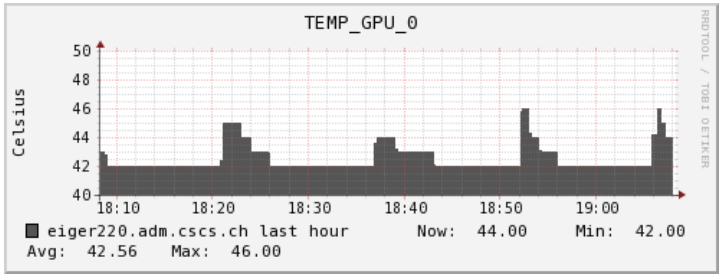
System Administration and Monitoring Tool (1)

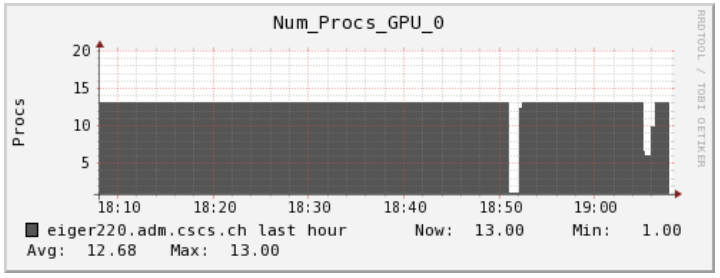
- Cluster Installation/Configuration/Profiling :
  - Autoyast Novell SUSE SLES 11 SP1
  - DHCP,BOOTP,TFTP, PXE
- Daily cluster administration: C3 tools and Parallel Distributed Shell:
  - cexec, cget, ckill, cpush, cname, etc., pdsh, pdcp
- GPU status check/report :
  - NVIDIA *nvidia-smi* command tool
  - AMD *aticonfig* command tool
- Cluster File Management:
  - Puppet (under integration) in the production environment

DALCO Visualization, R&D Cluster EIGER

System Administration and Monitoring Tool (2)

- **GANGLIA** Monitoring System (1 min. polling):
  - GPU temperature thresholds
  - GPU fan usage %
  - GPU memory usage %
  - GPU core usage %
  - End user processes using the GPU(s)
  - Other standard Ganglia cluster resource monitoring (cpu,network,swap,etc.)





DALCO Visualization, R&D Cluster EIGER

System Administration and Monitoring Tool (3)

- **NAGIOS** IT Infrastructure Monitoring Tool
  - Periodic critical service check and report in case of failure:
    - Cluster Nodes availability
    - IB fabric status
    - SLURM resource manager status
    - GPFS file system status
    - PGI License Server status
    - GPU driver status and appropriate ownership and file mode
    - X11 Server running on display :0.0 on all Viz Nodes
    - Simple CUDA kernel code execution on NVIDIA GPUs nodes
    - Simple OpenCL code execution on AMD HD Radeon GPUs nodes
    - Etc…

DALCO Visualization, R&D Cluster EIGER

Open Issues

- Automatic and transparent **self-recovery** after NVIDIA/AMD Radeon driver failure (nvidia/fglrx)
- Effective **GPU** detailed **accounting** measurement and reporting
- Heterogeneous multi-vendor **GPUs coexistence** within a node for OpenCL V 1.1 hybrid programming (NVIDIA+AMD Radeon)
- **Run time limit on GPU kernels** should be disabled for all NVIDIA GPUs
- **Concurrent GPU Kernels execution** should be allowed for all NVIDIA GPUs
- End user/job-based **ECC GPU memory control**
- **GPU compute operating mode** setting also for non-root user

Index

2. IBM iDataPlex R&D Cluster CASTOR

- Introduction
- Project goals
- Phase 1 & 2 project plans
- System description overview
- Cluster SW infrastructure

IBM iDataPlex R&D Cluster CASTOR

Introduction

CSCS, in partnership with several academic & industrial partners, wants to evaluate :

- Interconnect virtualization for scalable heterogeneous hybrid/multi-gpus platforms
- Provide selected HP2C projects with early access to a representative GPGPU prototype system

IBM iDataPlex R&D Cluster CASTOR

Project goals

- Improve the efficiency of hybrid applications by reducing the communication overhead between GPU devices (***GPUDirect technology***)
- Assess the feasibility of completely decoupling GPU and CPU resources in a hybrid computing environment (***r-CUDA***)
- Evaluate scalable interconnect technologies using industry-standard components (***IB fabric topologies***)

IBM iDataPlex R&D Cluster CASTOR

Phase 1 & 2 project plans

Phase 1 prototype:

- Based on technology currently available on the market, providing an evaluation platform for host memory bypass and 2D-Mesh/2D-Torus Infiniband fabric
- Time planning : Oct. 2011 – April 2012

Phase 2 prototype :

- Based on next generation technology, providing an evaluation platform for CPU-GPU decoupling and higher Infiniband dimensional fabric
- Time planning : April 2012 – Dec. 2012

IBM iDataPlex R&D Cluster CASTOR

System Description Overview (1)

- The IBM iDataPlex R & D Cluster CASTOR is a **new CSCS facility** which extends the current hybrid GPGPU/CPU resource portfolio

- During Q4 2011 it will be partially integrated into the CSCS Supercomputing ecosystem, and will be opened **exclusively** to the HP2C swiss scientific user community for **hybrid multicore/multi-GPU computing**, data analysis, and specific internal CSCS research activities.

- The CASTOR facility is a tightly coupled computing cluster system, running **RHEL 6.1 Server operating system** release and offers **32 IBM dx360 M3 nodes** based on the **dual-socket six-cores Intel Xeon 5650 processor** architecture running at 2.6 GHz, offering **24 GB of main system memory** per node, for a total of **384 cpu cores**, 768 HW threads and **768 GB** aggregate memory.

IBM iDataPlex R&D Cluster CASTOR

System description overview (2)

- The cluster nodes are all **homogeneous** in term of **HW configuration**, and come equipped with 2 NVIDIA Tesla GPUs per node :
- **NVIDIA Fermi M2090: http://www.nvidia.com/docs/IO/105880/DS-Tesla-M-Class-Aug11.pdf**
- Based on the **CUDA architecture** codenamed **Fermi** the Tesla M-class GPU Computing Modules are today (Summer 2011) the world's fastest parallel computing processors for HPC.(Claimed by NVIDIA)
- The highest performance Fermi-based GPGPU **M2090** has the following features :
  - 665 GFlops Peak DP
  - 6 GB memory size (ECC off)
  - 177 GB/sec memory bandwidth (ECC off)
  - 512 CUDA cores

IBM iDataPlex R&D Cluster CASTOR

System description overview (3)

- The IBM iDataPlex Cluster will be “partitioned” into 2 sub-clusters for specific project needs :
  - CASTOR partition : **Advanced R & D partition**
  - POLLUX partition : **CSCS Internal Test partition**

- Several **class of nodes** have been defined inside the cluster, covering special functionalities :
  - Class 0: Administration Node (1x)
  - Class 1: Login Node (2x)
  - Class 2: GPU Computing Nodes (32x)
  - Class 3: Storage Nodes (2x)

IBM iDataPlex R&D Cluster CASTOR

System description overview (4)

Cluster Networks (1):

- As an high speed network interconnect, the cluster CASTOR rely on a **dedicated** private internal **Infiniband QDR fabric** infrastructure, supporting parallel-MPI traffic, configured, due to specific research project needs, as a **2D mesh IB Fabric topology**, based on **16 IB QDR Mellanox switches**.

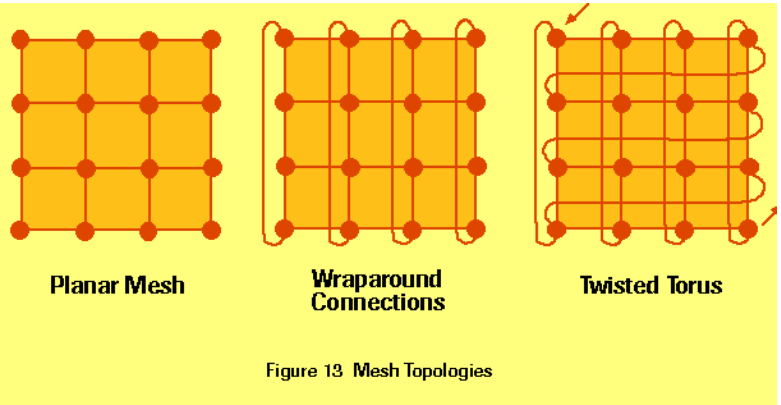
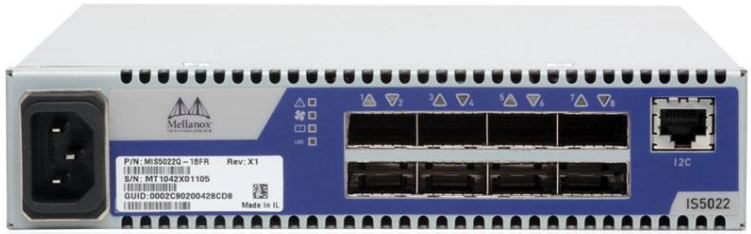


Figure 13 Mesh Topologies

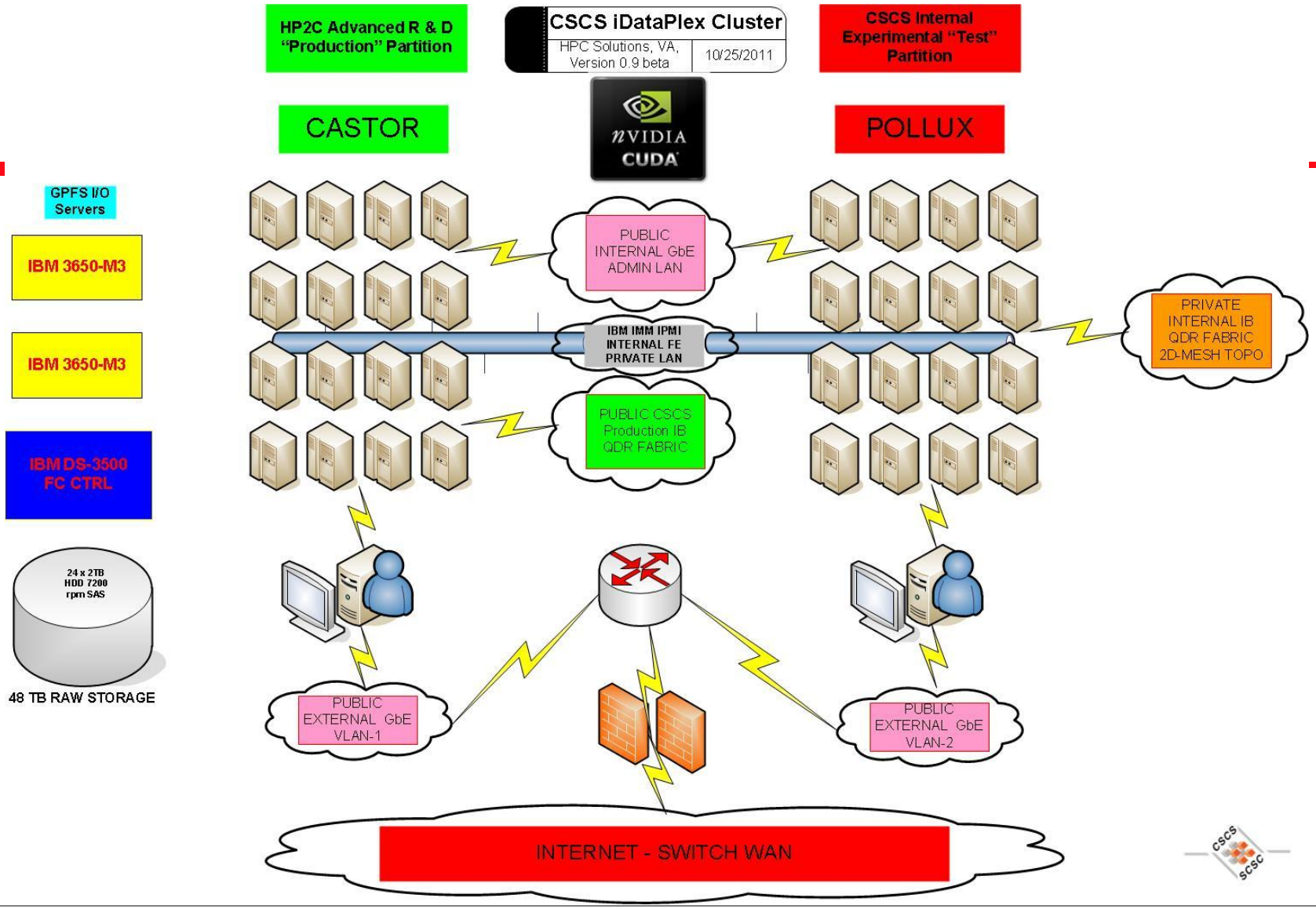


IBM iDataPlex R&D Cluster CASTOR

System description overview (5)

Cluster Networks (2) :

- Next to it a **standard IB QDR fabric** is also configured, connecting the whole cluster together using the additional IB ports provided by each node.
- A local **GPFS /scratch** cluster file system is running on top of the second IB fabric, in order to provide fast I/O performance to end user jobs.
- In addition, a **commodity 1 GbE LAN** ensures interactive login access, home, project and application NFS file sharing among the cluster nodes, and an additional standard 1 GbE administration network is also reserved for cluster management purposes.


HP2C Advanced R & D "Production" Partition
CSCS iDataPlex Cluster
HPC Solutions, VA, Version 0.9 beta
10/25/2011
CSCS Internal Experimental "Test" Partition
CASTOR
nVIDIA CUDA
POLLUX
GPFS I/O Servers
IBM 3650-M3
IBM 3650-M3
IBM DS-3500 FC CTRL
24 x 2TB HDD 7200 rpm SAS
48 TB RAW STORAGE
PUBLIC INTERNAL GbE ADMIN LAN
IBM IMM IPMI INTERNAL FE PRIVATE LAN
PUBLIC CSCS Production IB QDR FABRIC
PRIVATE INTERNAL IB QDR FABRIC 2D-MESH TOPO
PUBLIC EXTERNAL GbE VLAN-1
PUBLIC EXTERNAL GbE VLAN-2
INTERNET - SWITCH WAN

IBM iDataPlex R&D Cluster CASTOR

Cluster SW infrastructure (1)

- The software middleware infrastructure planned for the IBM iDataPlex Cluster will be based on the **RHEL V 6.1 Server** distribution.
- The entire cluster will be provisioned and managed by **XCAT**, and will offer a comprehensive development environment for the end user.
- A set of additional **system management tools** will also be installed in order to support the daily monitoring, management and reporting activities of the cluster :
  - Parallel Distributed Shell
  - Cluster File Manager (Puppet)
  - GANGLIA Monitoring System
  - NAGIOS IT Infrastructure Monitoring Tool
  - NVIDIA *nvidia-smi* command tool

IBM iDataPlex R&D Cluster CASTOR

Cluster SW infrastructure (2)

- **Eclipse IDE C/C++ framework V 3.7.1**, with the integration of CUDA will be also provided, via the Eclipse CUDA plugin
- The following **compiler suites** will be offered :
  - PGI Accelerator C/C++/Fortran V 11.9
  - Intel C/C++/Fortran Composer XE 2011
  - Open64 4.2.4
  - PathScale EKOPat 4
  - GNU C/C++/Fortran V 4.6.1
- **Debugging tools** :
  - Allinea DDT with CUDA
  - NVIDIA *cuda-memcheck* and *cuda-gdb*
- **Profiling tools** :
  - Nvidia Visual Profiler (NVIDIA *cudaprof* and *openglprof)*

IBM iDataPlex R&D Cluster CASTOR

Cluster SW infrastructure (3)

- **NVIDIA CUDA V 4.x Toolkit**, together with the latest available NVIDIA CUDA SDK Kit and the latest NVIDIA GPU driver will ensure to deploy the most recent HW features of the NVIDIA Tesla FERMI GPUs available in the system.

- Among the several **GPU-native libraries** the following ones will be provided:
  - CUFFT, CUBLAS, CULA Tools
  - MAGMA, IMSL, CUSPARSE, CUSP, CURAND
  - Accelerys LibJacket, NPP

IBM iDataPlex R&D Cluster CASTOR

Cluster SW infrastructure (4)

- The Infiniband OpenFabrics Enterprise Distribution **(OFED™) V 1.5.3.2**, which is an open-source software for RDMA and kernel bypass applications, will be configured as the **primary IB software stack**.

- **MVAPICH2** and **OpenMPI** will be both available for Message Passing parallel jobs based on the MPI-2 standard. In particular the following versions will be provided :
  - MVAPICH2 V 1.7
  - OpenMPI V 1.5.4 beta release

IBM iDataPlex R&D Cluster CASTOR

Cluster SW infrastructure (5)

- **SLURM Resource Manager V 2.3.0-pre5** is the main batch queuing system installed and supported on the cluster which let the end-users access in a shared or reserved mode any available GPGPU computing resource.

- **IBM GPFS V 3.4.x** will be offered within the cluster, so a global parallel SCRATCH file system is available for parallel MPI jobs or for any other I/O intensive tasks.

IBM iDataPlex R&D Cluster CASTOR

Cluster SW infrastructure (6)

rCUDA Framework: (http://www.hpca.uji.es/rCUDA)

- The **rCUDA framework** enables the concurrent usage of CUDA-compatible devices remotely.
- rCUDA employs the **socket API for the communication** between clients and servers. Thus, it can be useful in three different environments:
  - *Clusters*: to **reduce the number of GPUs installed** in High Performance Clusters. this leads to increased GPU usage and therefore energy savings as well as other related savings like acquisition costs, maintenance, space, cooling, etc.
  - *Academia*. In commodity networks, to offer access to a few high performance **GPUs concurrently to many students**.
  - *Virtual Machines*. To enable the **access to the CUDA facilities** on the physical machine.

IBM iDataPlex R&D Cluster CASTOR

Cluster SW infrastructure (7)

MOSIX Virtual OpenCL : (http://www.mosix.org/txt_vcl.html)

- The VCL cluster platform allows OpenCL applications to **transparently utilize many GPU** devices in a cluster, as if all the devices are on the local computer.
- Can run unmodified **OpenCL 1.1** (and 1.0) applications.
- Applications can utilize **cluster-wide OpenCL devices**.
- SuperCL, an extension of OpenCL that under VCL allows micro-programs of OpenCL to run efficiently on devices of remote nodes.
- **Transparent selection** of devices.
- Applications can be **started on any hosting-computer**, including workstations without GPU devices.
- Supports **multiple applications** on the same cluster.
- **Runs on Linux clusters**, with or without MOSIX.

IBM iDataPlex R&D Cluster CASTOR

Cluster SW infrastructure (8)

GPUDirect technology :

- Mellanox OFED GPUDirect package is a Mellanox OFED package that also include Mellanox – NVIDIA GPUDirect technology. It provides the capability for **direct communication between the NVIDIA GPU and the Mellanox ConnectX InfiniBand adapter** through the host memory with **zero copy** and low latency RDMA (Remote Direct Memory Access).

- GPUDirect with Mellanox InfiniBand adapters accelerates local GPU to remote GPU communication by 30%, resulting in higher application performance.

IBM iDataPlex R&D Cluster CASTOR

Cluster SW infrastructure (9)

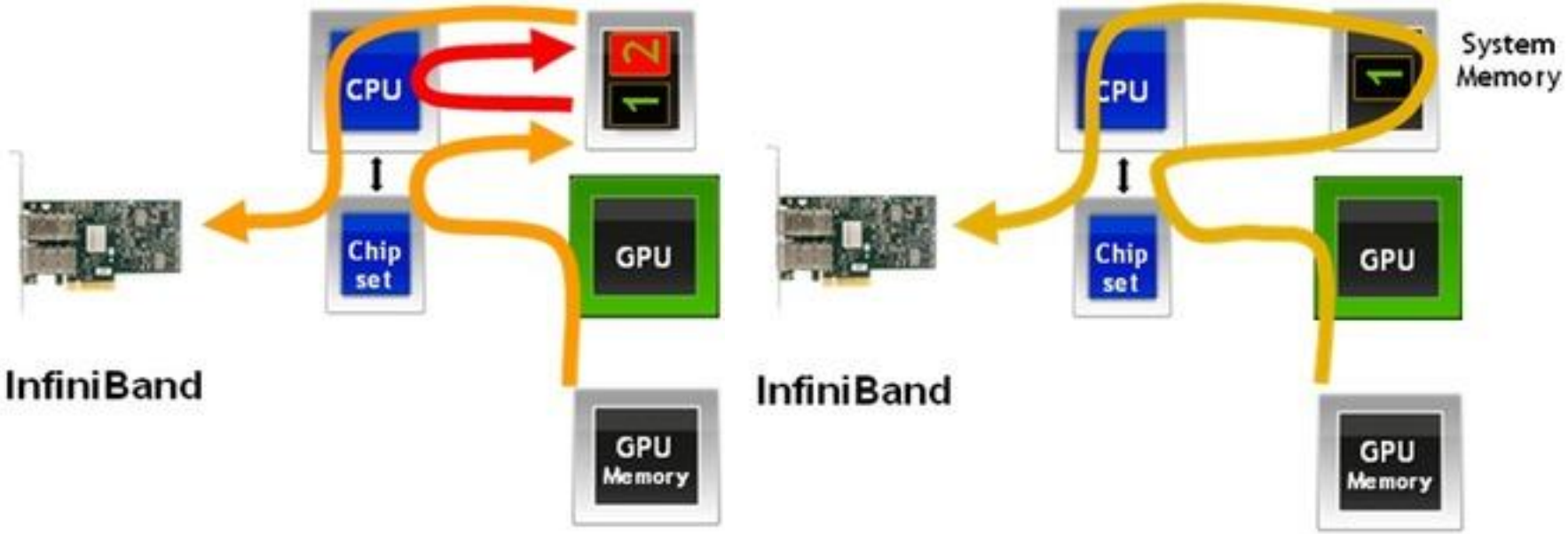
GPUDirect technology key features :

- Enables host memory sharing between NVIDIA GPU and Mellanox HCA
- Provides Zero Copy transaction for GPU-GPU communications over InfiniBand
- Reduces local to remote GPU latency by 30%
- Accelerates GPU-based applications performance
- Eliminates the host CPU involvement in the GPU-GPU data path

IBM iDataPlex R&D Cluster CASTOR

Cluster SW infrastructure (10)

GPUDirect technology:



Thank You.

Questions

&

Answers