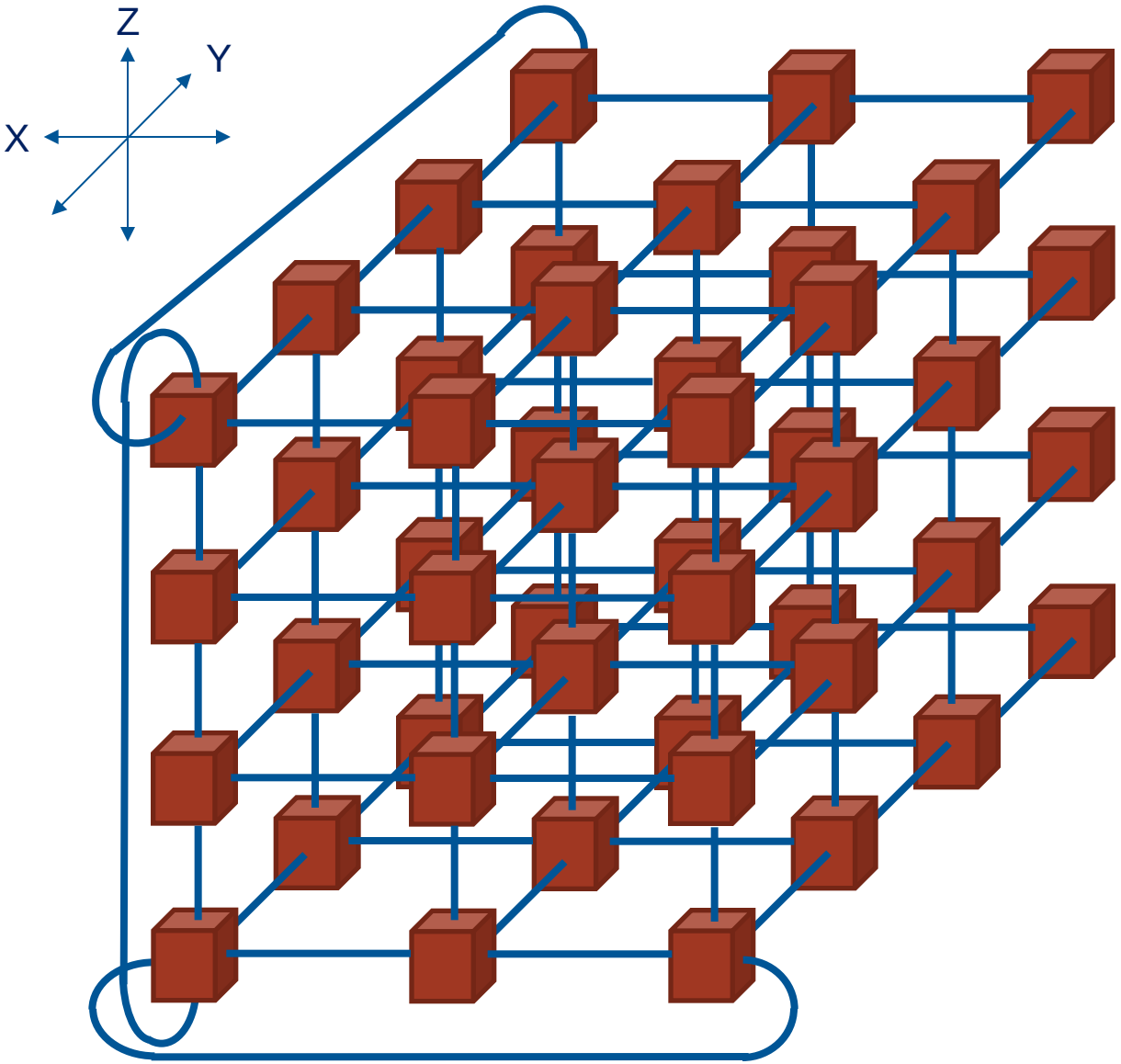
Cray XE6 Architecture Overview

Recipe for a good MPP

- **Select Best Microprocessor**
  - Function of time
- **Surround it with a balanced or "bandwidth rich" environment**
  - Interconnection network
  - Local memory
- **"Scale" the System**
  - Eliminate Operating System Interference (OS Jitter)
  - Design in Reliability and Resiliency
  - Provide Scaleable System Management
  - Provide Scaleable I/O
  - Provide Scaleable Programming and Performance Tools
  - System Service Life (provide an upgrade path)

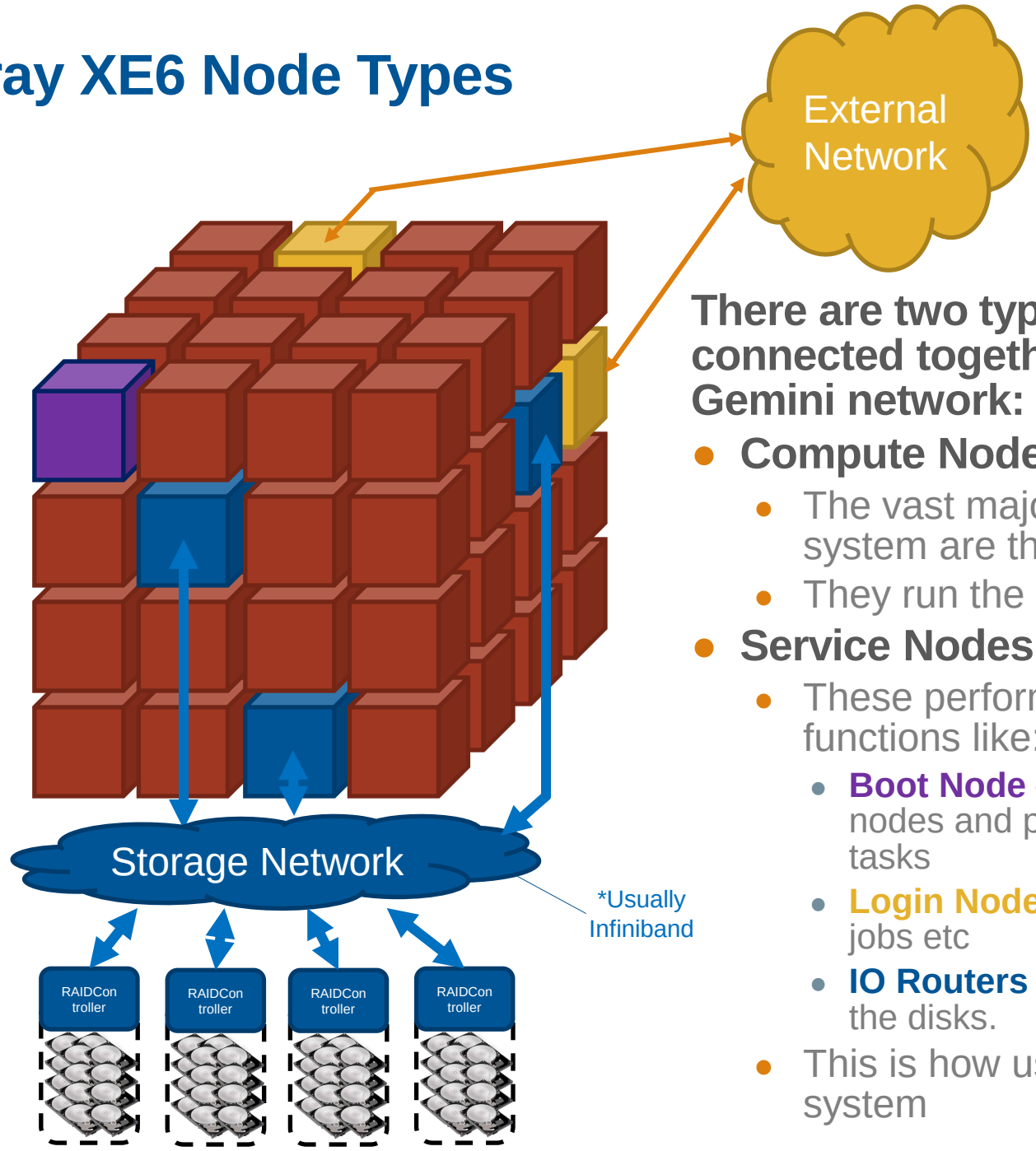
Cray XE6 Toroidal Topology



The Cray XE6 connects nodes together in a 3D Torus,

Each node is connected to other nodes in 6 directions via the Cray Gemini network interconnect.

Cray XE6 Node Types



There are two types of nodes connected together on the Cray Gemini network:

- **Compute Nodes**
  - The vast majority of nodes in the system are the **compute nodes**
  - They run the user applications
- **Service Nodes**
  - These perform necessary system functions like:
    - **Boot Node** – Used to control the other nodes and perform system admin tasks
    - **Login Nodes** – Compiling, submitting jobs etc
    - **IO Routers** - Moving data to and from the disks.
  - This is how users interact with the system

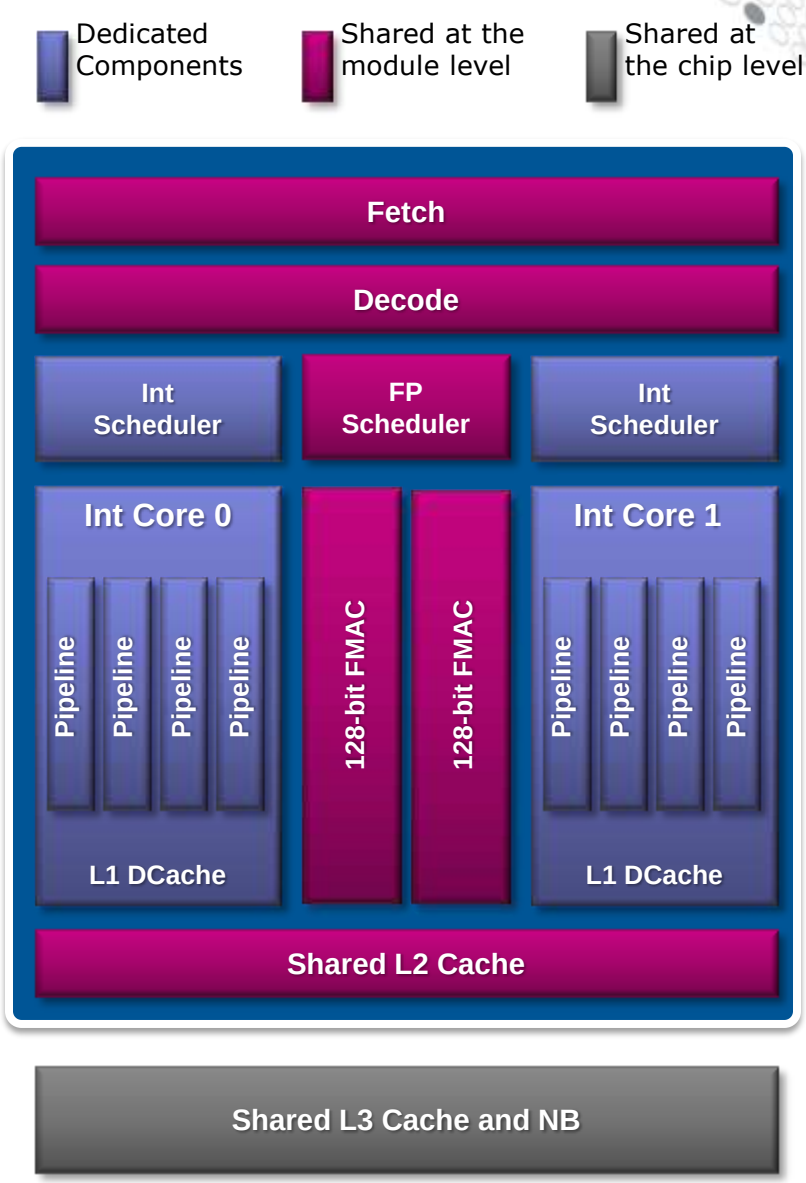
HECToR

- **HECToR is the Cray XE6 installed in Edinburgh and forms the UK's National Supercomputer Service.**
- **Each of HECToR's compute nodes has:**
  - Two AMD Opteron 6276 "Interlagos" Processors
    - 2 x 16 "Bulldozer" compute modules
    - 2.1 GHz CPU Frequency
    - 268 Gflops Theoretical peak performance
  - 32GB DDR3-1333 DRAM
    - 8 Channels per node
    - 86GB/s Theoretical peak memory bandwidth
- **There are 2816 Compute Nodes and 64 Service Nodes**
  - The torus is 15 x 12 x 16

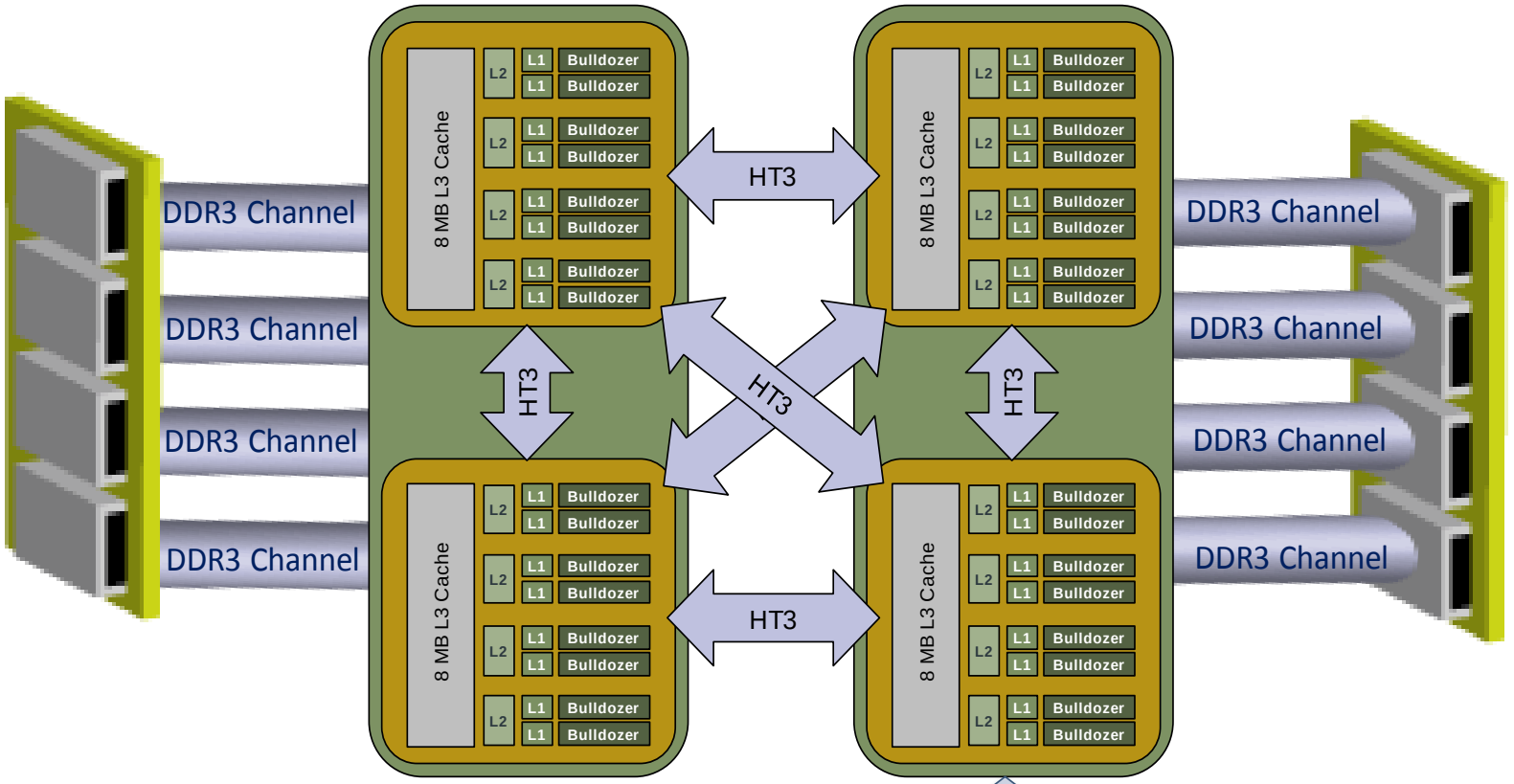
There are other XE6 models in PRACE which may have different processor , memory and network configurations.

Interlagos Processor Architecture

- Interlagos is composed of a number of Bulldozer core “modules”
- A core module has shared and dedicated components
- There are two independent integer cores and a *shared*, 256-bit FP resource
- A single Integer Core can make use of the entire FP resource with 256-bit AVX instructions
- This architecture is very flexible, and can be applied effectively to a variety of workloads and problems
- **DL1 is 16 KB, L2 is 2 MB and L3 is 8MB**



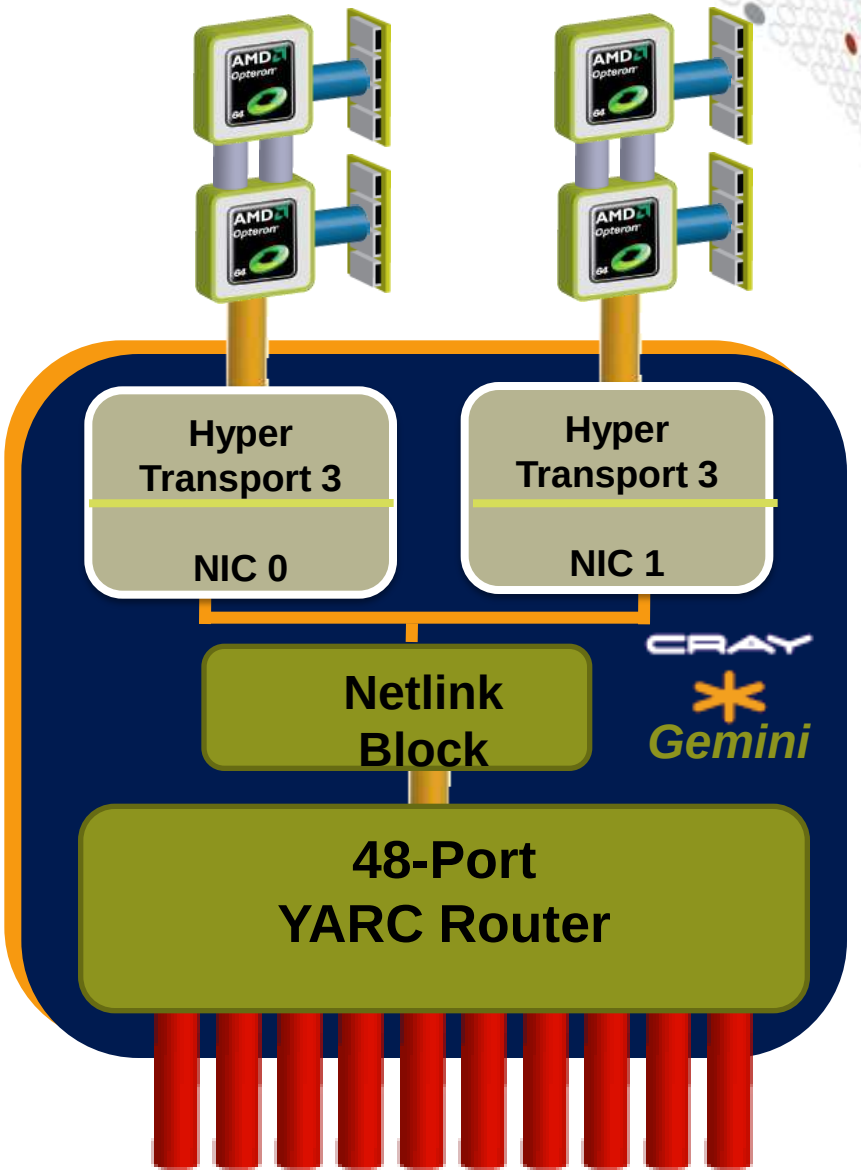
Cray XE6 Node Details – 32-core Interlagos



- 2 Multi-Chip Modules,
  4 Opteron Dies: ~300 Gflops
- 8 Channels of DDR3 Bandwidth to 8 DIMMs: ~105 GB/s
- 32 Computational Cores, 32 MB of L3 cache
- Dies are fully connected with HT3

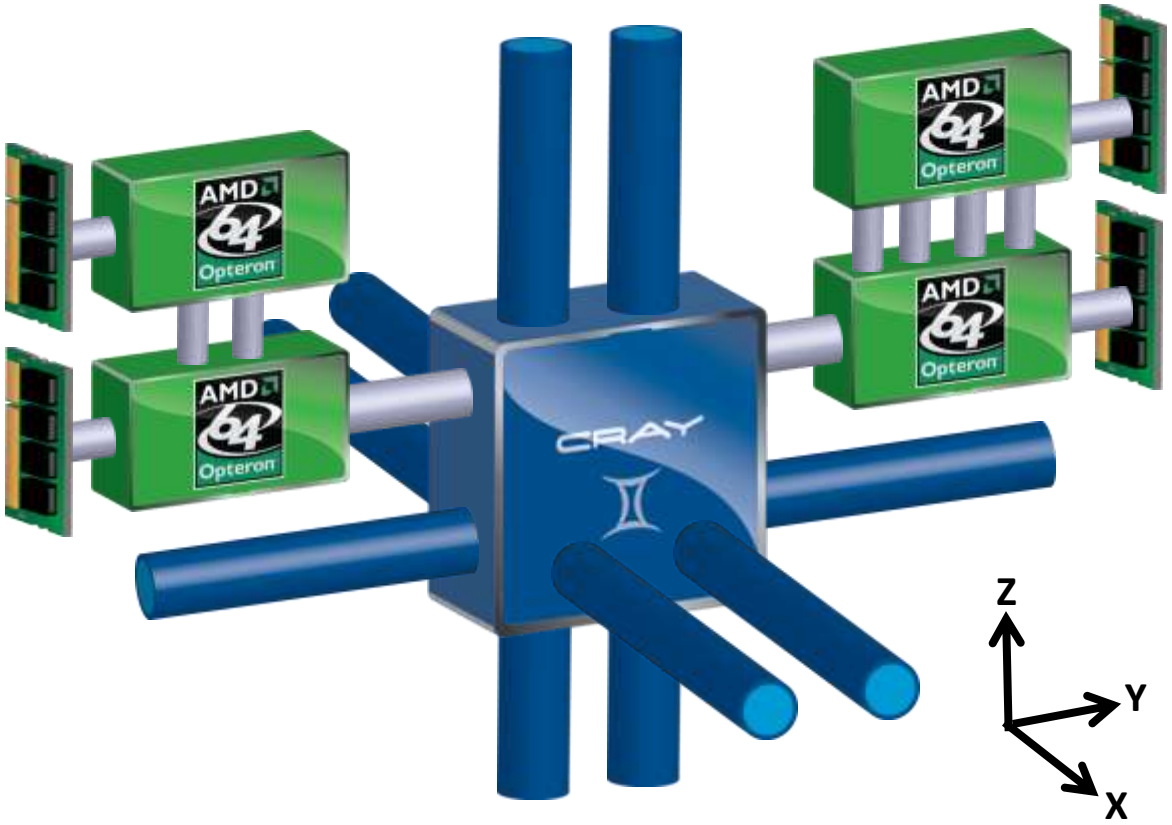
Cray Gemini ASIC

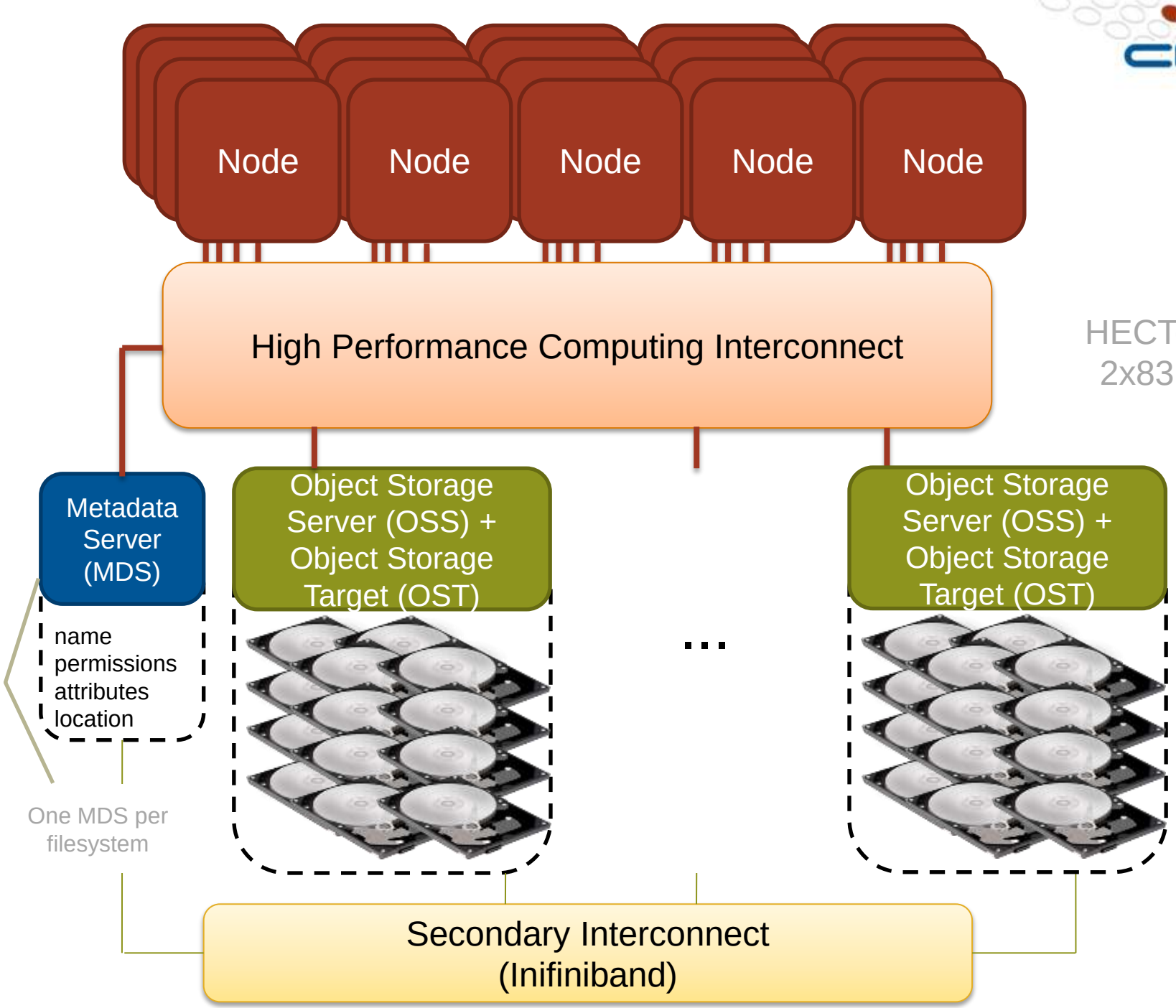
- **Supports 2 Nodes per ASIC**
- **3D Torus network**
  - XT5/XT6 systems field upgradable
- **Scales to over 100,000 network endpoints**
  - Link Level Reliability and Adaptive Routing
  - Advanced Resiliency Features
- **Advanced features**
  - MPI – millions of messages / second
  - One-sided MPI
  - UPC, Coarray FORTRAN, Shmem, Global Arrays
  - Atomic memory operations



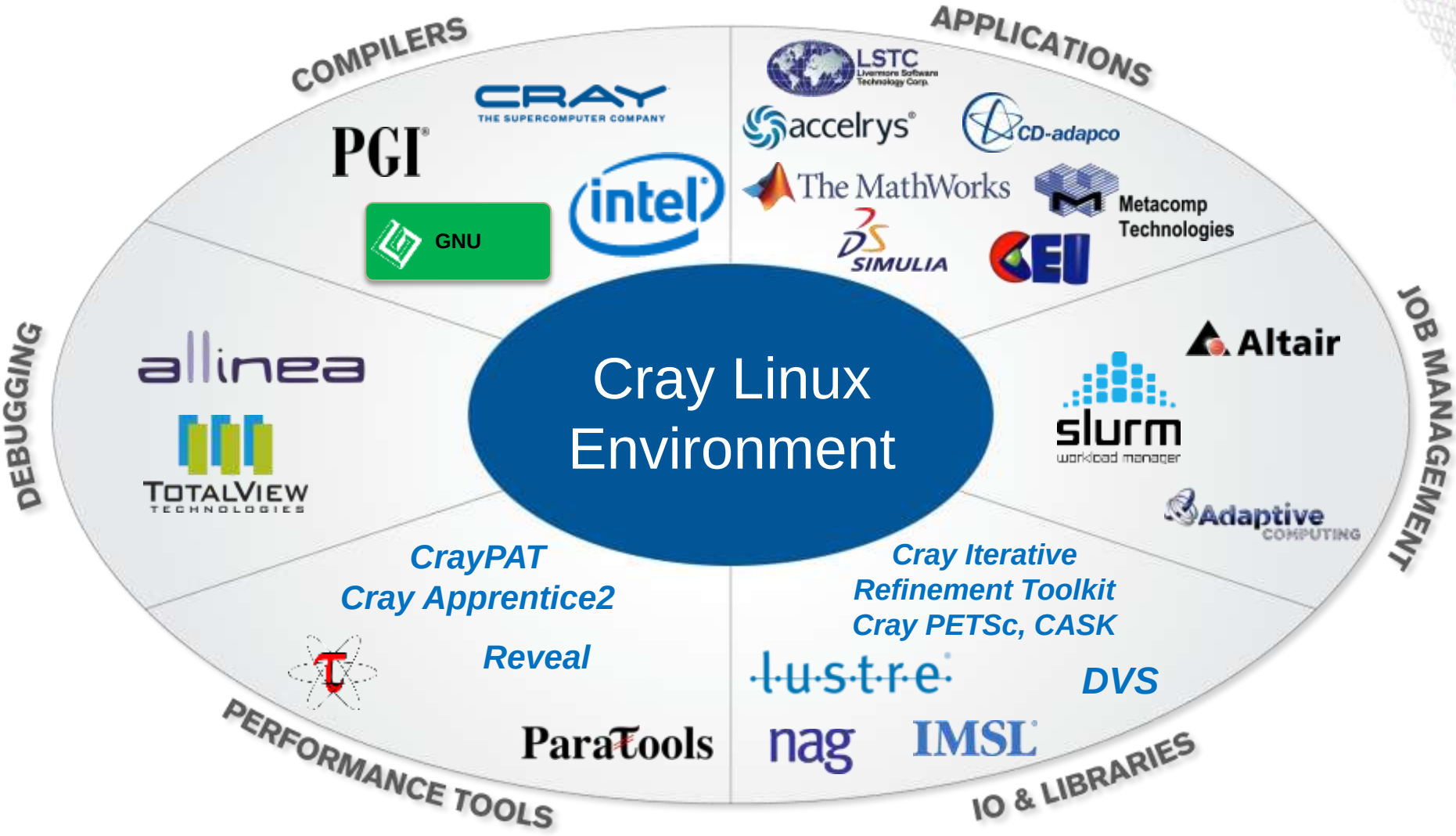
A Complete Cray XE6 Compute Node

- Built around the Gemini Interconnect
- Each Gemini ASIC provides 2 NICs enabling it to connect 2 dual-socket nodes




Node
Node
Node
Node
Node
High Performance Computing Interconnect
HECToR has 2x83 OSTs
Metadata Server (MDS)
name
permissions
attributes
location
One MDS per filesystem
Object Storage Server (OSS) + Object Storage Target (OST)
...
Object Storage Server (OSS) + Object Storage Target (OST)
Secondary Interconnect (Inifiniband)

Cray Software Ecosystem



Cray Programming Environment Distribution Focus on Performance and Productivity

Programming Languages

- Fortran
- C
- C++

Programming models

- Distributed Memory (Cray MPT)
  - MPI
  - SHMEM
- Shared Memory
  - OpenMP 3.0
  - OpenACC
- PGAS & Global View
  - UPC (CCE)
  - CAF (CCE)
  - Chapel

Compilers

- Cray Compiling Environment (CCE)
- GNU
- 3rd Party Compilers

Tools

- Environment setup
  - Modules
- Debuggers
  - TotalView
  - Allinea (DDT)
  - lgdb
- Debugging Support Tools
  - Abnormal Termination Processing
  - STAT
- Performance Analysis
  - CrayPat
  - Cray Apprentice²
- Scoping Analysis
  - Reveal

Optimized Scientific Libraries

- LAPACK
- ScaLAPACK
- BLAS (libgoto)
- Iterative Refinement Toolkit
- Cray Adaptive FFTs (CRAFFT)
- FFTW
- Cray PETSc (with CASK)
- Cray Trilinos (with CASK)

I/O Libraries

- NetCDF
- HDF5

Cray developed
Licensed ISV SW
3rd party packaging
Cray added value to 3rd party

Summary

- **Cray XE6: a dedicated MPP machine**
  - High Performance CPUs
  - Gemini interconnect featuring high message throughput and bandwidth
  - Scalable, tailored, jitter-free programming environment
- **This architecture is designed for massively parallel jobs solving grand challenges**
- **In this workshop, we will learn**
  - Basic usage of the programming environment and tools
  - How to porting, analyse and optimize your applications using the Cray Advanced Tools.