LCI workshop
on high performance clustered computing

Linux Clusters Institute: Introduction to High Performance Computing

University of Wyoming
May 22 – 26, 2017

Irfan Elahi, National Center for Atmospheric Research

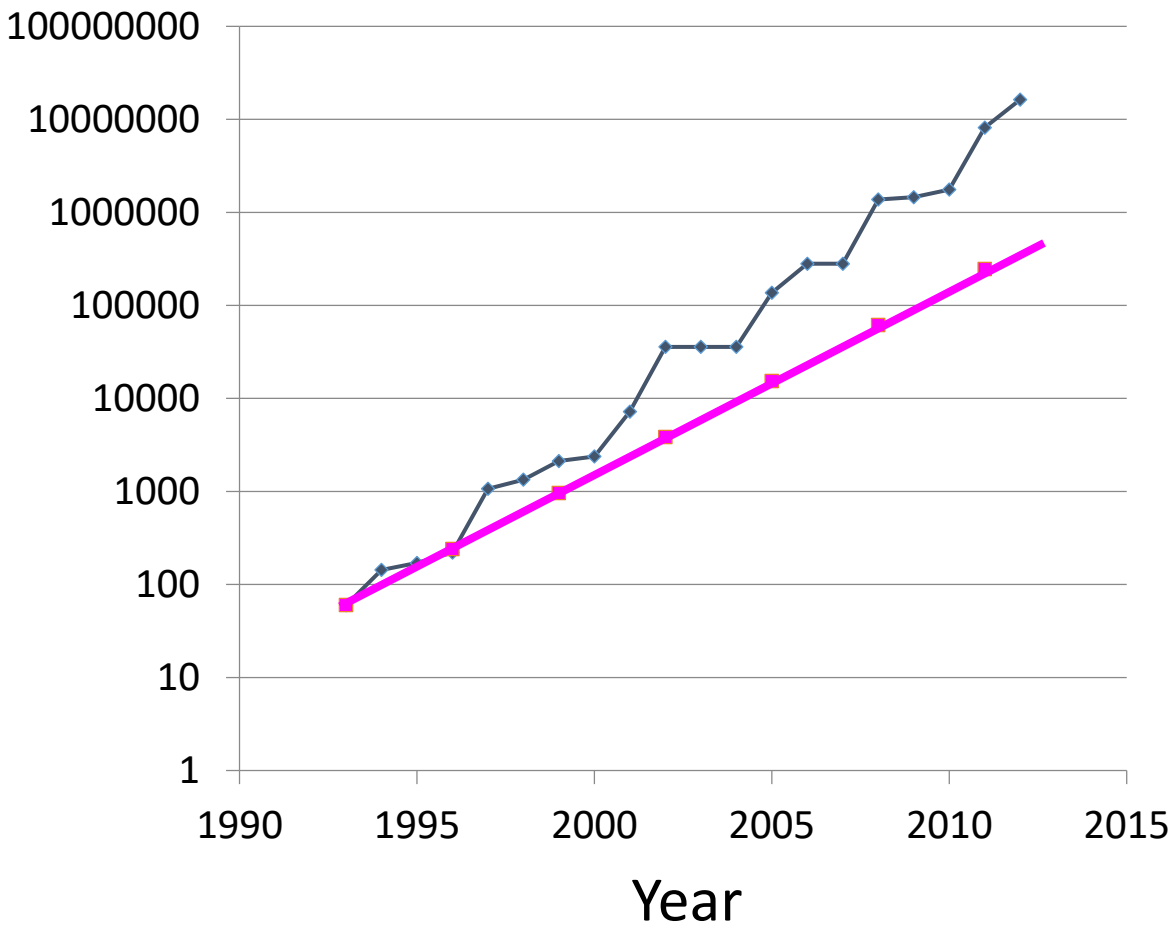
NCAR

What is Supercomputing or High Performance Computing?

LCI workshop
on high performance clustered computing

- The definition of supercomputing is **constantly changing**. Supercomputers can perform up to quadrillions of FLOPS or PFLOPS.
- **High-performance computing** (**HPC**) utilizes parallel processing for running large and advanced application programs efficiently. The term applies especially to systems that function above a hundred teraflops. The Top500 list has several multi-petaflop systems in the top 50.
- HPC aggregates *computing* power in a way that delivers much higher *performance* than one could get out of a typical desktop *computer* or workstation in order to solve large problems in science, engineering, or business.
- Supercomputers were introduced in the 1960s, and initially created by Seymour Cray at Control Data Corporation who led the HPC industry for decades.
- To me, personally, it is an eco-system that provides users with a high performance computational, networking, storage, and analysis platform, and the necessary software stack to stitch these resources.

Fastest Supercomputer vs. Moore



GFLOPs:

billions of calculations per second

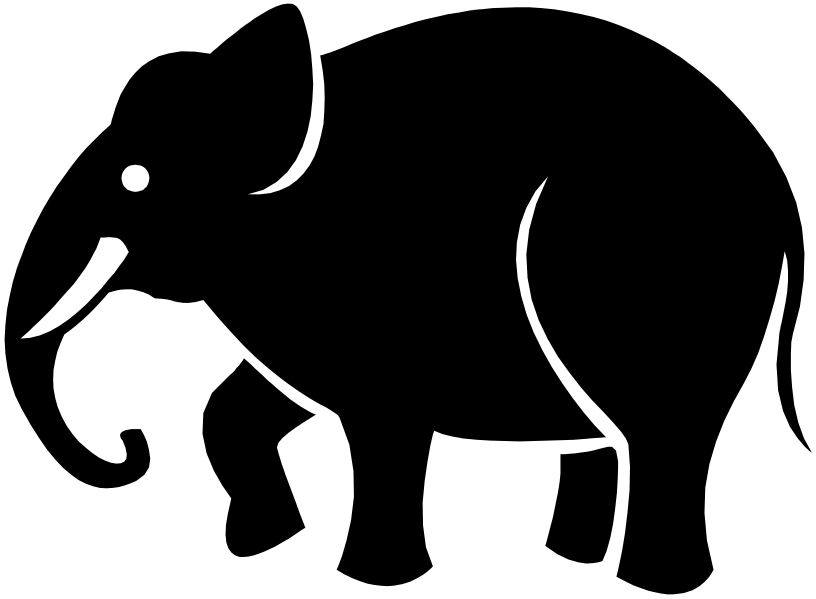
www.top500.org

http://www.mooreslaw.org/

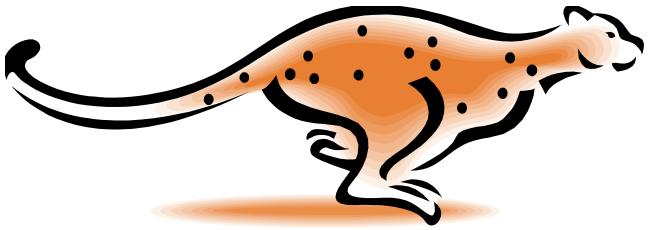
What is Supercomputing About?

LCI workshop
on high performance clustered computing

Size



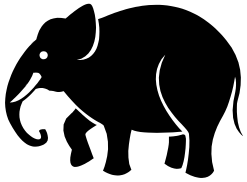
Speed



What is Supercomputing About?

LCI workshop
on high performance clustered computing

- ***Size***: Many problems that are interesting to scientists and engineers **can't fit on a single personal computer system** – usually because they need more RAM and/or more disk storage.



- ***Speed***: Many problems that are interesting to scientists and engineers would take a very, very long time to run on a single personal computer system - months or even years. But a problem that would take **a month on a PC** might take only **an hour on a supercomputer**.

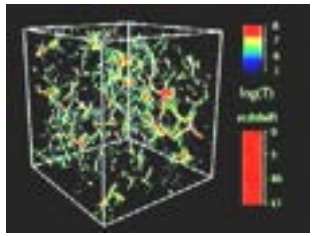
What is HPC Used For?

- ***Simulation*** of physical phenomena by developing a model that represents the key characteristics of the selected physical or abstract system or process. Areas where simulation is heavy used:
  - Weather forecasting
  - Galaxy formation
  - Oil reservoir management

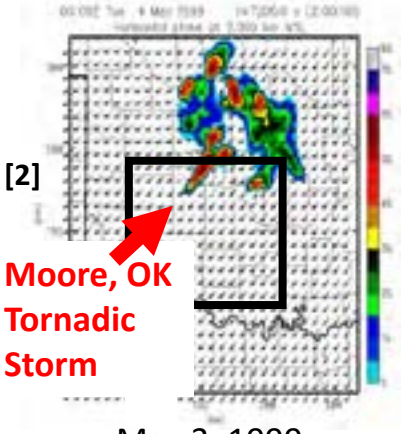
- ***Data mining***: finding **needle(s)** in a **haystack.** It is the process of analyzing *data* from different perspectives and summarizing it into useful and sometimes new information:
  - Gene sequencing
  - Signal processing
  - Detecting storms that might produce tornados

- ***Visualization***: turning a vast sea of **data** into **pictures** that a scientist can understand and analyze. Any technique for creating images, diagrams, or animations to communicate a message is called *Visualization.*

[1]

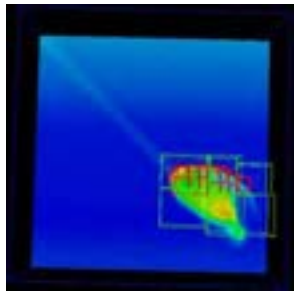


[2]



May 3, 1999

[3]



Supercomputing Issues

LCI workshop
on high performance clustered computing

- Storage hierarchy or storage tiers
- Parallelism: doing multiple things at the same time
- Scaling issues
- High-speed interconnect
- Software stack
- Facility

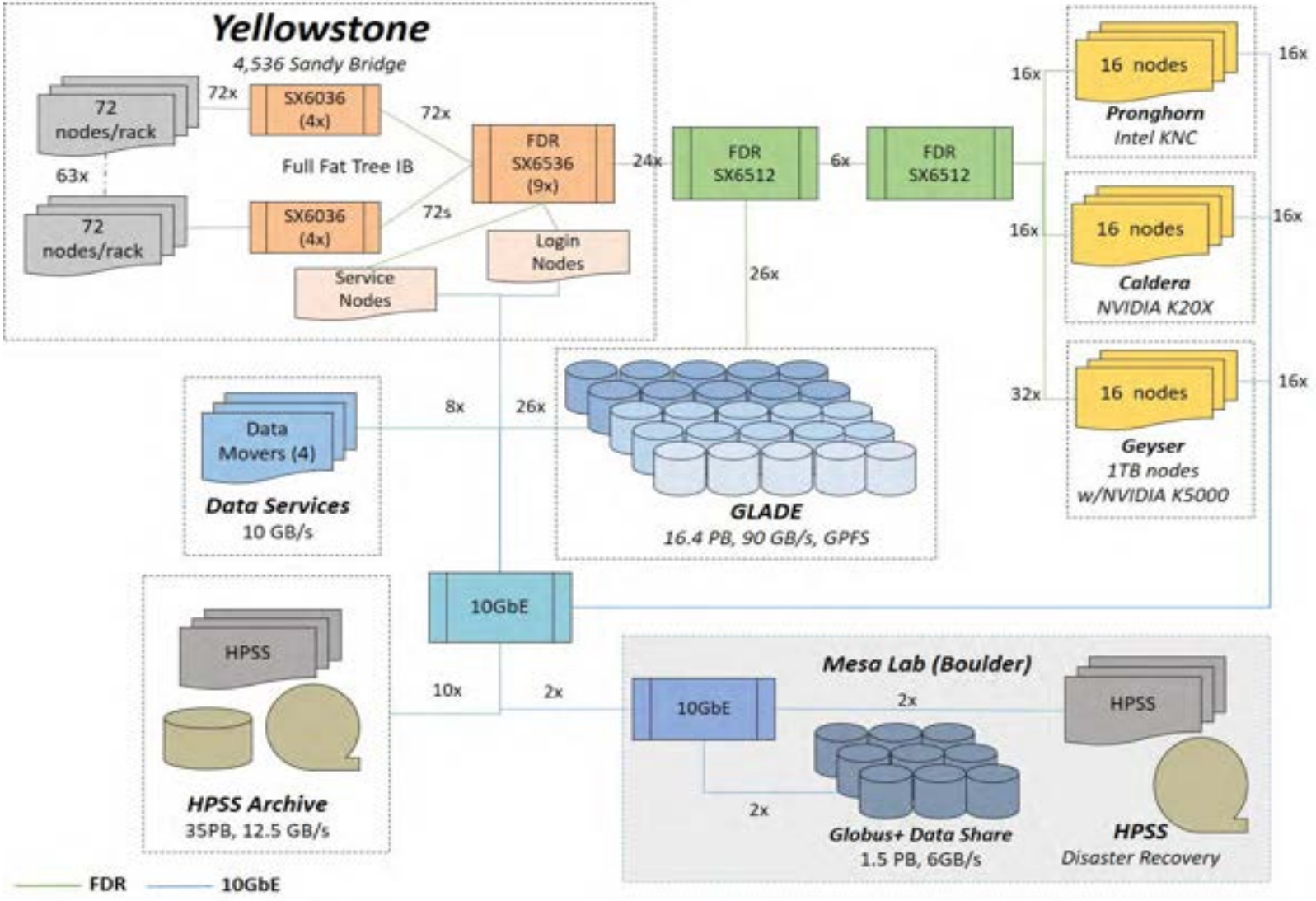
What is a “Cluster”?

LCI workshop
on high performance clustered computing

- A cluster **needs** a collection of small computers, called ***nodes***, hooked together by an ***interconnection network*** (or ***interconnect*** for short).
- It also **needs** software that allows the nodes to communicate over the interconnect.
- A cluster is **all** of these components working together as if they’re one big computer ... a **super** computer.

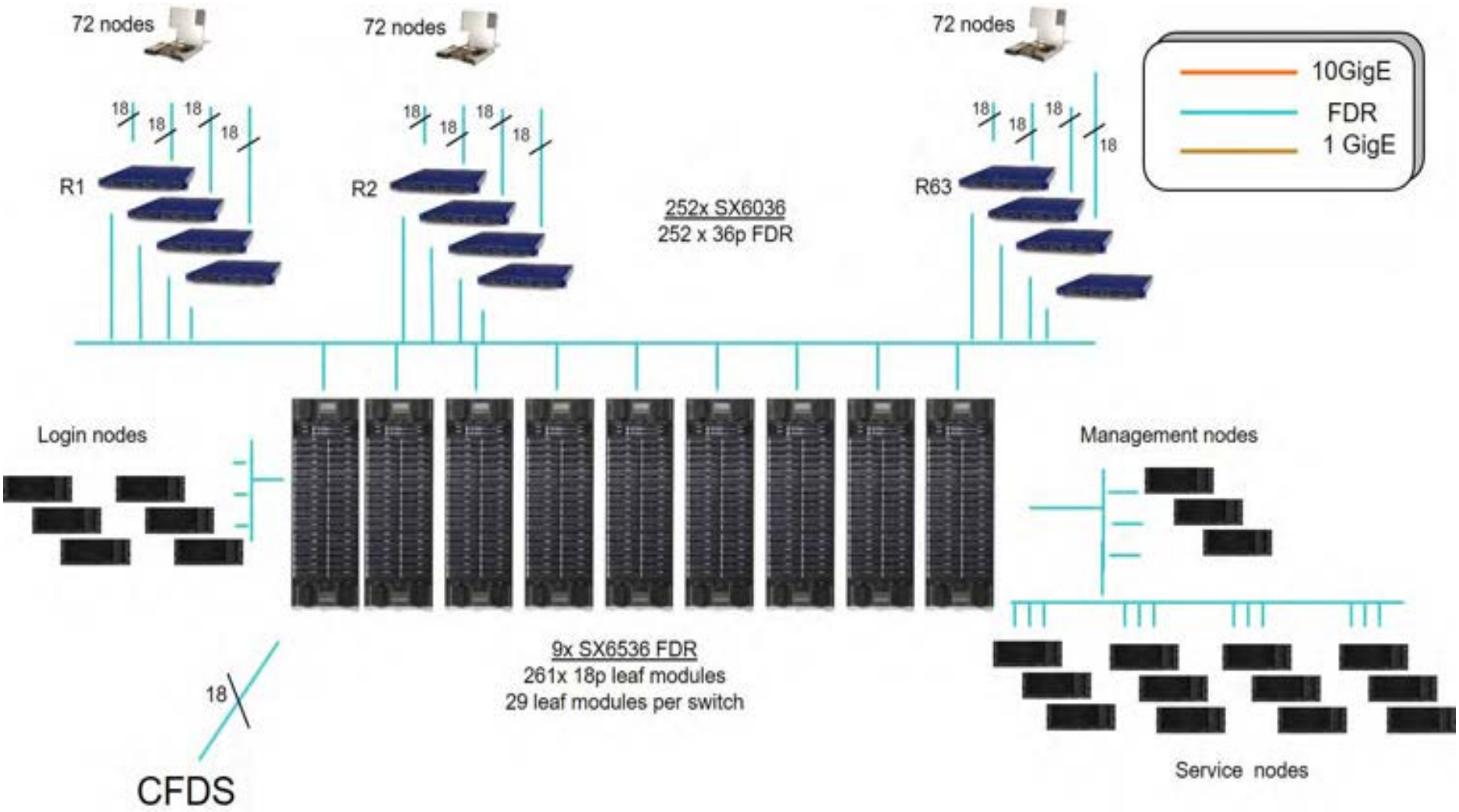
What Does a Cluster Look Like?

Network View


Yellowstone
4,536 Sandy Bridge
72 nodes/rack
72x
SX6036 (4x)
63x
72 nodes/rack
SX6036 (4x)
Full Fat Tree IB
72x
72s
FDR SX6536 (9x)
24x
FDR SX6512
6x
FDR SX6512
Login Nodes
Service Nodes
26x
16x
16 nodes
Pronghorn
Intel KNC
16x
16x
16 nodes
Caldera
NVIDIA K20X
16x
32x
16 nodes
Geyser
1TB nodes
w/NVIDIA K5000
16x
Data Movers (4)
Data Services
10 GB/s
8x
26x
GLADE
16.4 PB, 90 GB/s, GPFS
10GbE
HPSS
HPSS Archive
35PB, 12.5 GB/s
10x
2x
Mesa Lab (Boulder)
10GbE
2x
HPSS
2x
Globus+ Data Share
1.5 PB, 6GB/s
HPSS
Disaster Recovery
FDR
10GbE

What Does a Cluster Look Like?

LCI workshop
on high performance clustered computing



NCAR

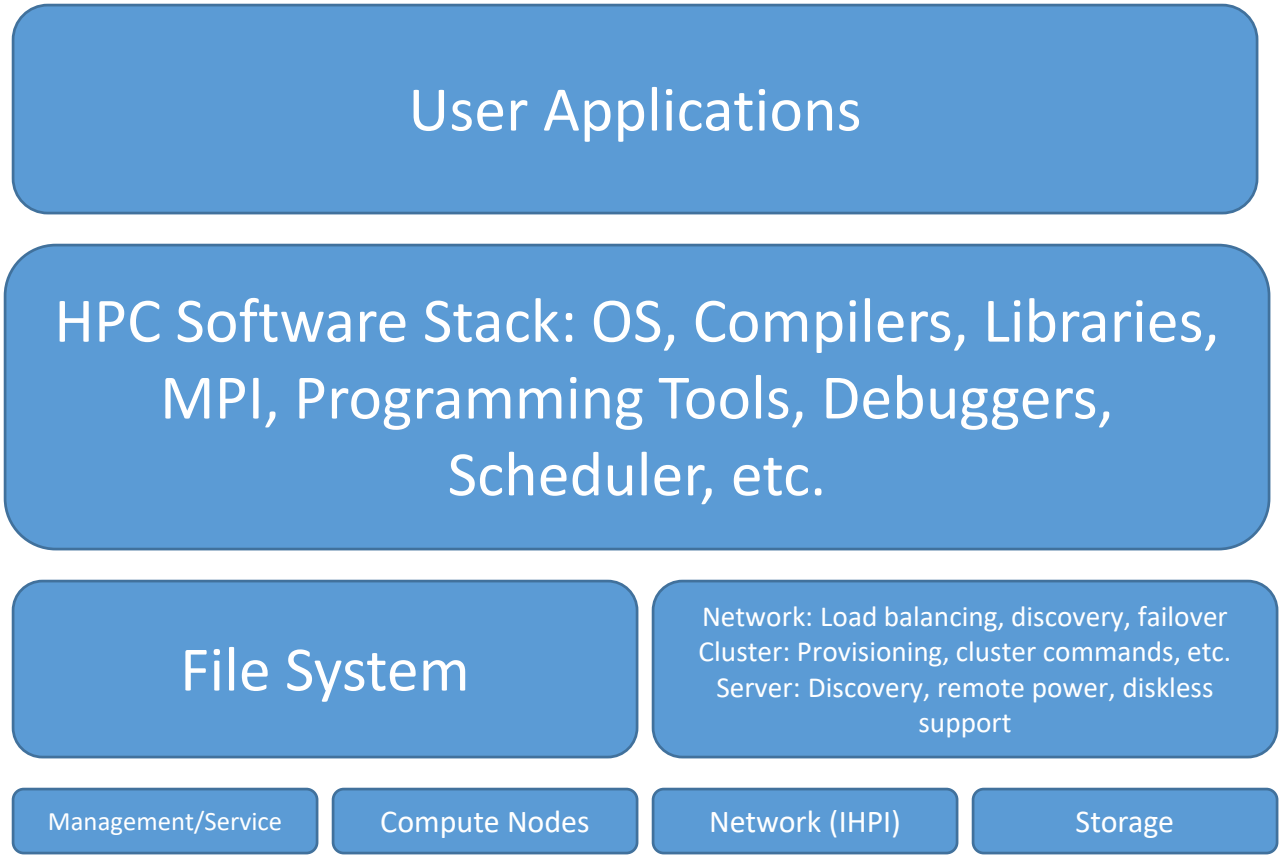
Cluster Components

All Components Working Together

- Computational resources
- Storage and file system
- Management infrastructure
- High-speed interconnect
- HPC software stack
- HPC applications and workflow

Cluster Components

All Components Working Together



Computational Resources

- Compute nodes and software
- Compute node has:
  - Processor (CPU)
  - Memory
  - Networking
  - Software
  - Access to file system for long or short term retention

Processor Types Examples

- **X-86 Architecture:**
  - The instruction set architecture (ISA) is Intel's most successful line of processors.
  - Xeon & Xeon Phi - Many-Core (Intel) and Opteron (AMD)
- **GPGPU or GPU:**
  - General-purpose computing on graphics processing units is the use of a graphics processing unit (GPU), which typically handles computation only for computer graphics, to perform computation in applications traditionally handled by the central processing unit (CPU).
  - NVIDIA, AMD, ASUS, etc., manufacture GPGPU/GPU.
- **POWER: (Power Optimization with Enhanced RISC)**
  - IBM has a series of high performance microprocessors called POWER.
  - IBM launched **OpenPOWER Foundation** for collaboration on their Power Architecture in 2013. Google, Tyan, Nvidia, and Mellanox are founding members.
- **ARM (Advanced RISC Machines)**
  - CPUs based on the RISC (reduced instruction set computer) architecture developed by Advanced RISC Machines (ARM).
  - Companies using ARM cores on their chips are Qualcomm, Samsung Electronics, Texas Instruments, and Cavium among others.

Storage and File System

- Used to control how data is stored and retrieved
- Short and long term retention
- HPC or Cluster's require shared FS:
  - Storage (Disk)
  - Software (Lustre/GPFS)
  - Transport/Networking

Management Infrastructure

- Cluster Management
- Cluster Network Management
- Service Nodes
- Head/Login Nodes
- Facility

Cluster Management

LCI workshop
on high performance clustered computing

- Node provisioning, hardware/power control, discovery, and OS disk-based/disk-less deployment
  - Monitoring and log management
- Fabric management
- Cluster startup and shutdown
- Parallel shell
  - One ring to rule them all, one ring to find them, one ring to bring them all, and in the darkness bind them.

Facility

- Power
  - 120 V, 280 V, 440V, etc.
  - 3-Phase, DC
  - N + N or N+1 Redundancy
  - UPS, Generators
- Cooling
  - Water cooled or air cooled

HPC Software Stack

LCI workshop
on high performance clustered computing

- Cluster Management Software
- Operating System
- Compilers, Programming Tool, and Libraries
- File System
- Scheduler/Resource Manager

High-Speed Interconnect

- Low latency, high-bandwidth
- Library support, FCA, MPI offload, RDMA, etc.
- Fabric management
- Examples:
  - Ethernet: performance is expected to hit 400 Gbps soon
  - Infiniband: EDR = ~100 Gbps, HDR = ~200 Gbps and NDR = ~400 Gbps
  - OPA: Gen1 =~100 Gbps

HPC Applications and Workflow

LCI workshop
on high performance clustered computing

- Parallelism:
  - Speedup is not linear
- Dependencies
- Tuning:
  - Race conditions
  - Mutual exclusions
  - Synchronization