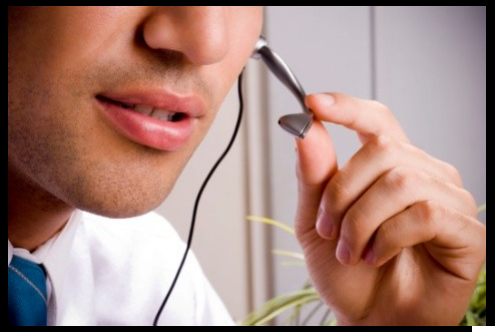
High Performance Computing @ Fermilab

Amitoj Singh

July 24th 2012

Computing @ Fermilab



Service Desk

Computer Security

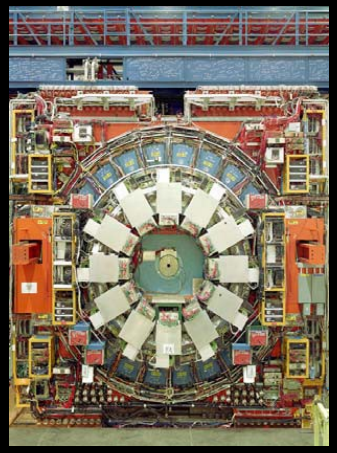
E Communication

Computing @ Fermilab

A major part of the Computing Sector's work is to support and improve the scientific programs at the lab. This includes computer support for experiment systems, design and implementation of the DAQ and control systems, accelerator and detector simulations, research and development of the physics analysis software.

The MINOS Experiment

The D0 Experiment



The CDF Experiment

CMS Experiment at LHC

What is HPC?

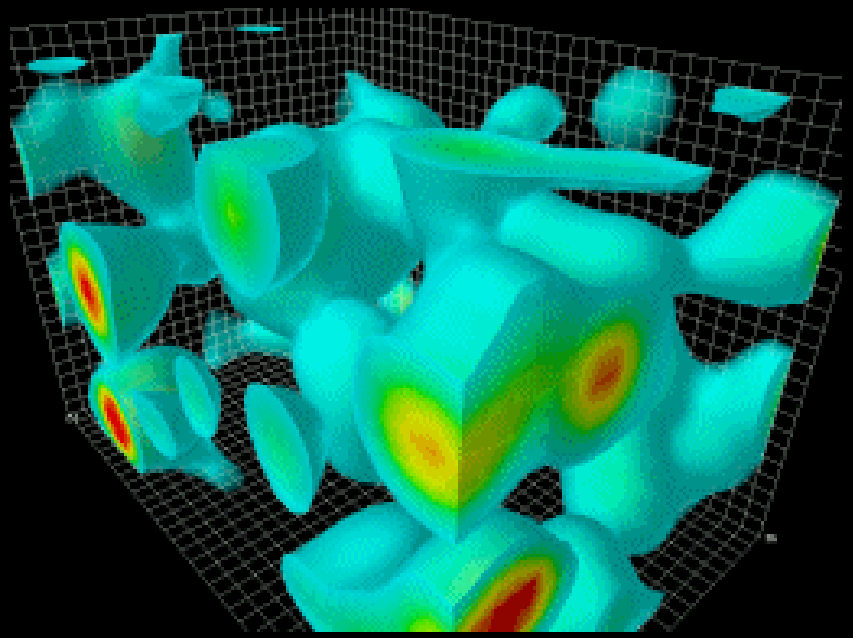
High Performance Computing (HPC) uses supercomputers and computer clusters to solve advanced computation problems.

Computer Cluster

IBM Blue Gene Supercomputer

Why we need HPC?

Discovered in the early 1970s, the theory of Quantum chromodynamics (QCD) consists of equations that describe the strong force that causes quarks to clump together to form protons and other constituents of matter. For a long time solving these equations was a struggle. But in the last decade using powerful supercomputers theorists are now able to finally solve the equations of QCD with high precision.



How do I measure the speed of a supercomputer?

FLOPS

FLoating point **O**perations **P**er **S**econd

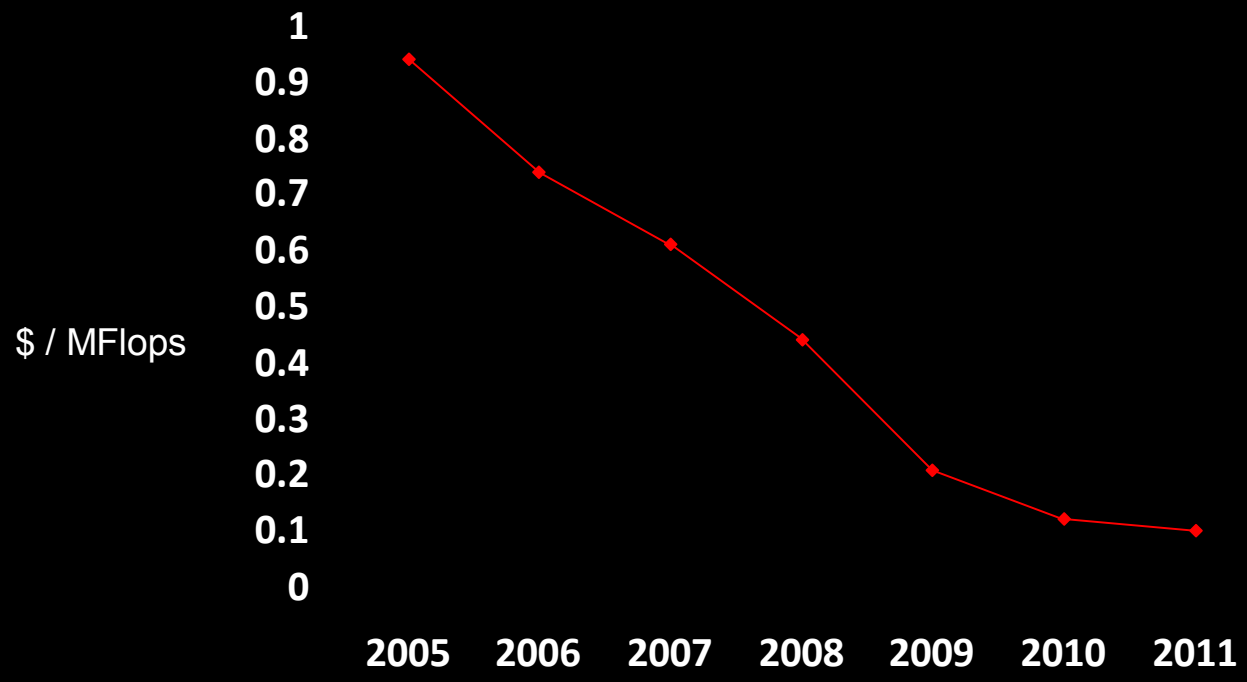
Examples of floating point numbers are

1.234567, 123456.7, 0.00001234567, 1234567000000000

LINPACK Benchmark

(http://www.top500.org/project/linpack/)

HPC cost trend

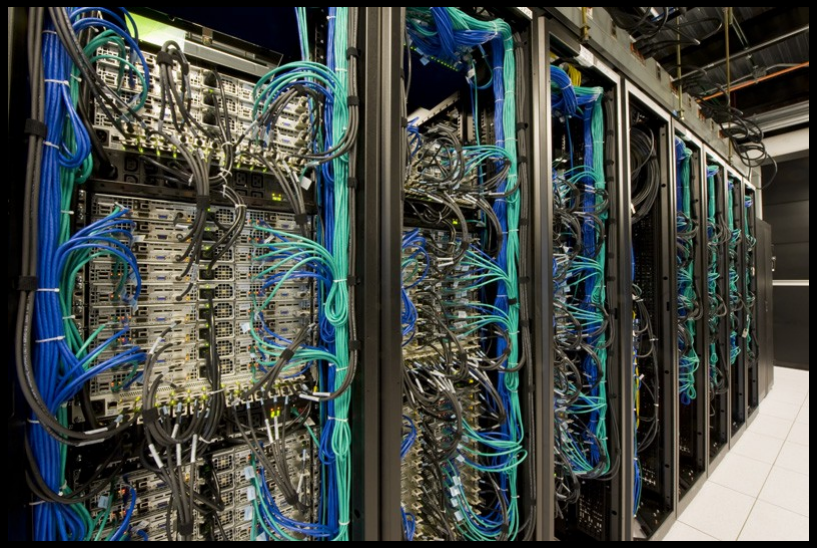


How much does 1 Million Flops cost?

Computer Cluster Architecture

The building blocks of our computer clusters are:

- Compute nodes.
- Network switches.
- Lots of disk storage.



A typical compute node

NVIDIA GK110 *Kepler* GPU

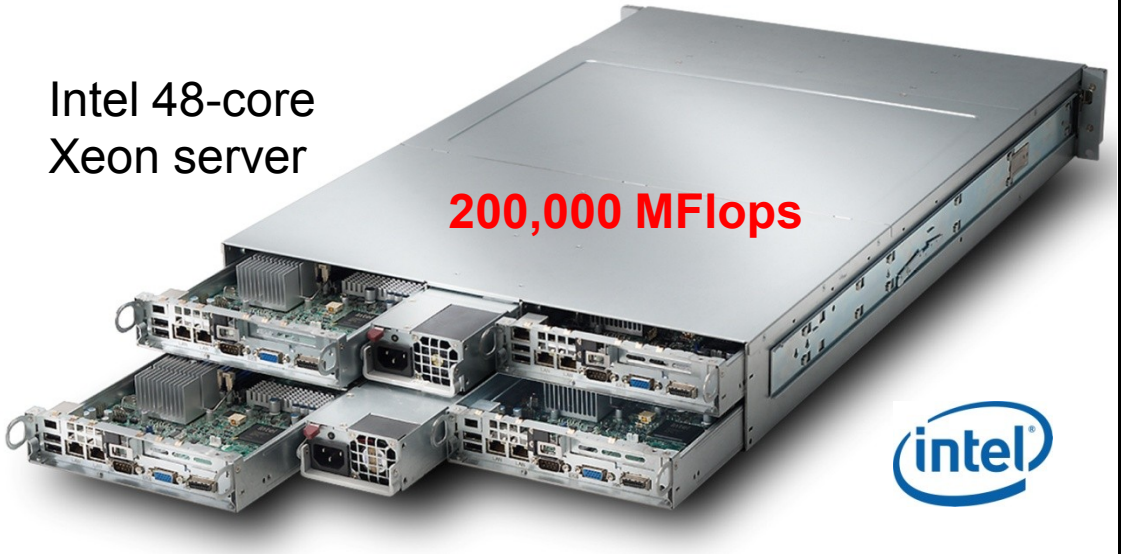
1,000,000 MFlops

NVIDIA

SONY

20,000 MFlops

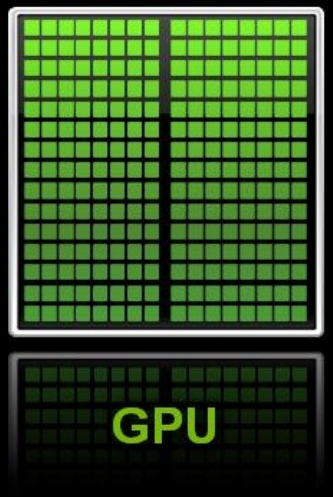
Intel 48-core
Xeon server



200,000 MFlops

intel

Difference between a CPU & GPU



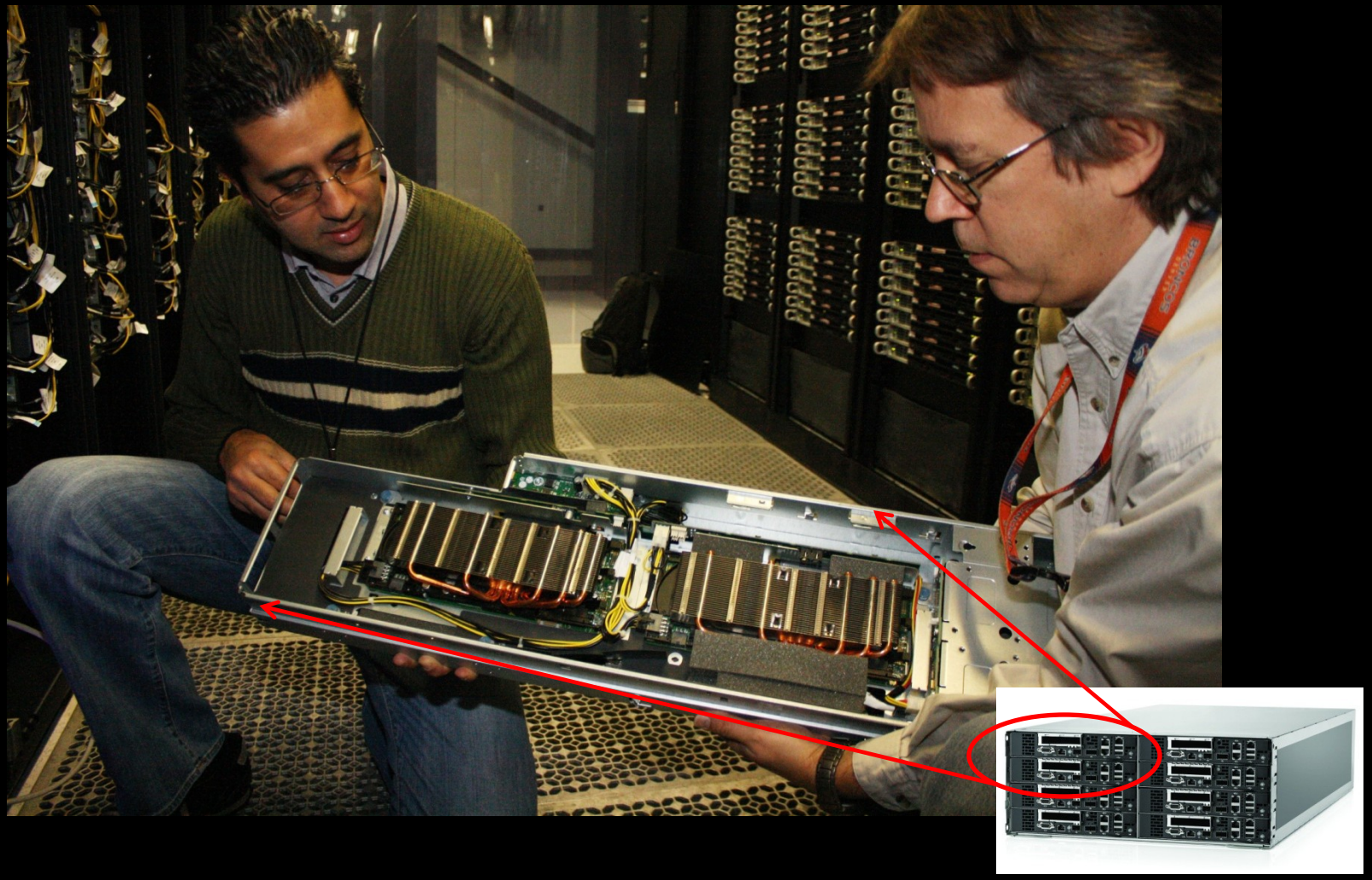
"GPUs are optimized for taking huge batches of data and performing the same operation over and over very quickly, unlike PC microprocessors, which tend to skip all over the place."

- Nathan Brookwood (Principal Analyst Insight64)

"The combination of a CPU with a GPU can deliver the best value of system performance, price, and power."

- Kevin Krewell (Senior editor Microprocessor Report)

GPU: Graphics Processing Units



Network Switches

This refrigerator size network switch built by Sun Microsystems consists of 3,456 ports and is capable of transferring 14 TBytes/second which is about 3000 DVDs worth of data in one second.

We use the smaller version of this switch on our Fermilab supercomputers.

Bandwidth v/s Latency

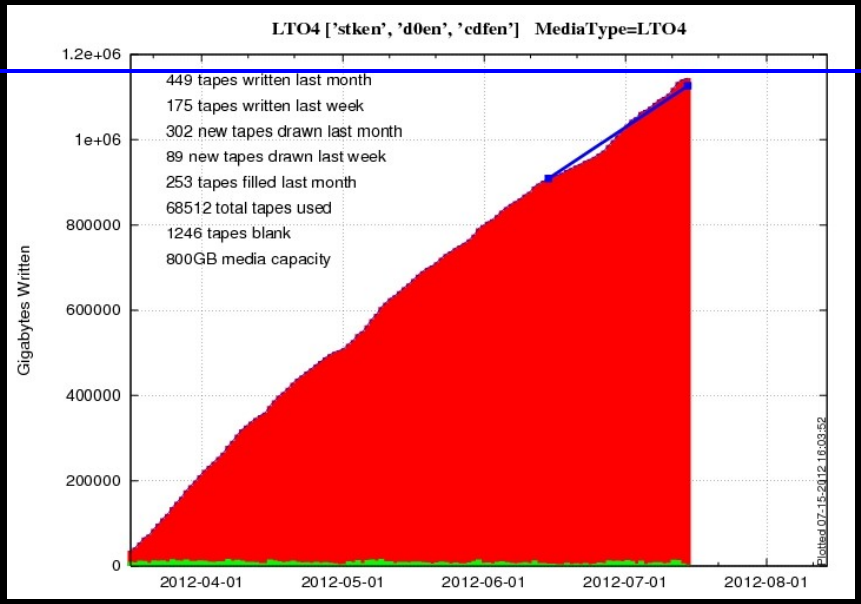
When selecting network switches for supercomputers we have to consider two key factors: Bandwidth and Latency . . . . and price at times since some high speed switches can be prohibitively expensive.

How much can you carry?

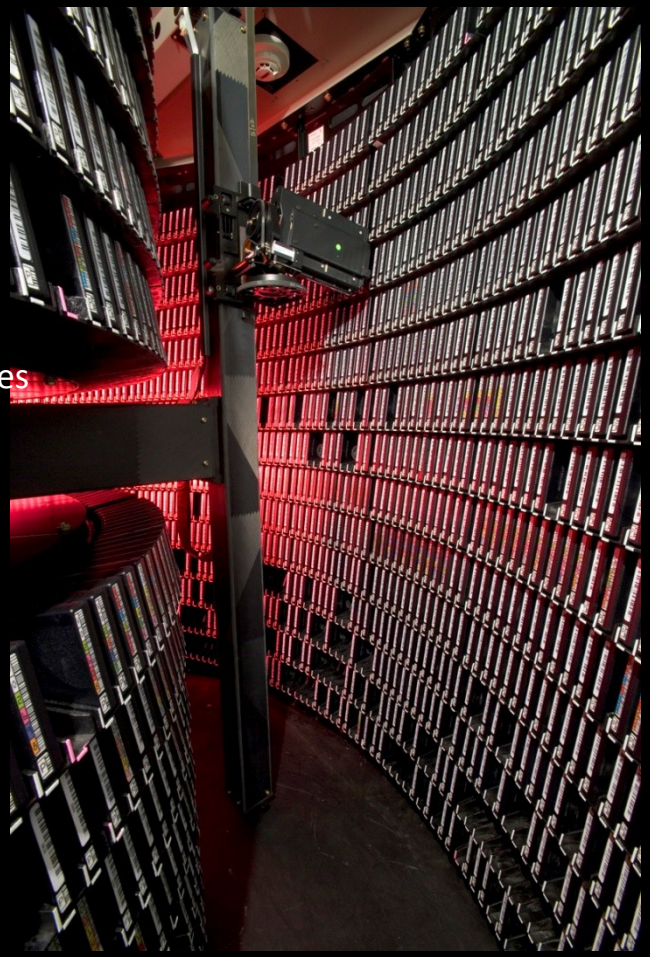
How fast can you carry it?

HPC Storage

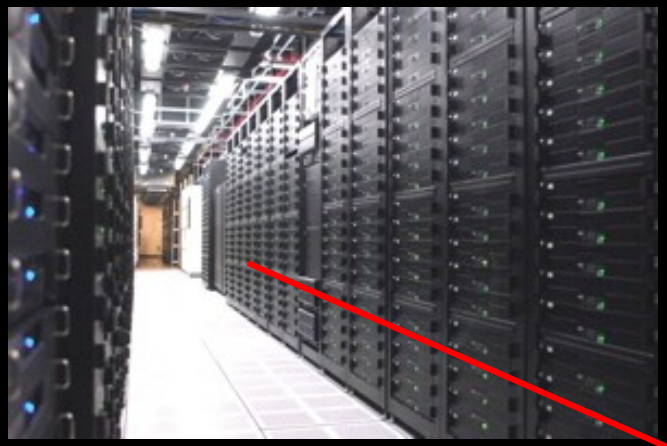
Valuable data is stored on tapes. FNAL has three tape robots with 44 Petabytes (44000 Terabytes) worth of data.

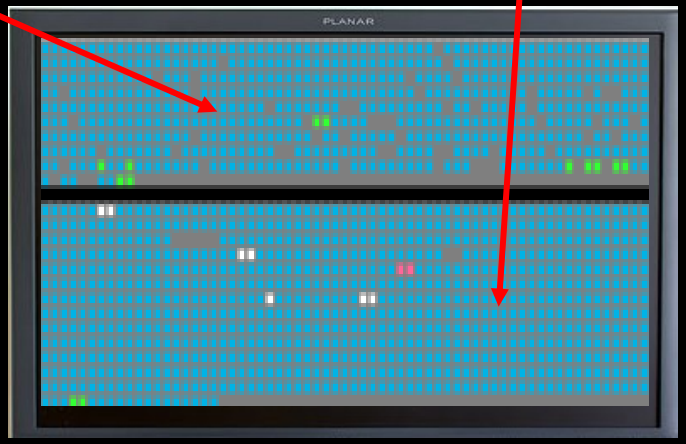


1.1 Petabytes



User interface to HPC





TORQUE is an open source resource manager providing control over batch jobs and distributed compute nodes. Users use TORQUE commands to submit jobs to the various HPC clusters.

User Interface to HPC

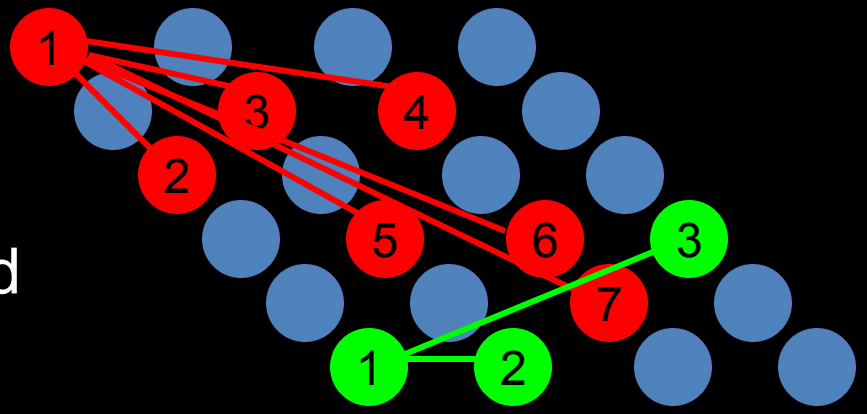
- The resource manager TORQUE maintains a *queue* of all such requests and assigns available and appropriate compute nodes to requests either FIFO (First In First Out) or depending on preset criteria.
- For example: TV screens that run commercials at gas stations or supermarket checkout lanes use the same concept. Frozen dinner entrée ads should only run after 6PM and cereal ads should run between 6 and 10AM. All other commercials are run FIFO (First In First Out).

Message Passing Interface

MPI is a language-independent communications protocol used to program parallel computers.

MPI's goals are high performance, scalability, and portability.

How do users figure out their resource (nodes, memory) requirements?



MPI

Argonne NATIONAL LABORATORY

Managing Supercomputers

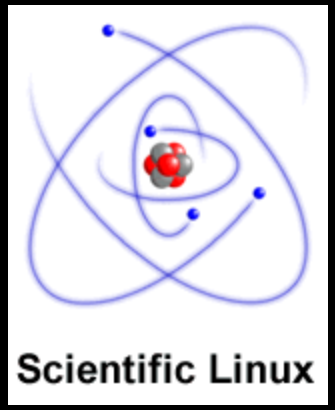
- <u>Biggest challenge:</u> A job on the supercomputer will run at the speed of the slowest component.

• 2.4GHz Intel Core 2 Duo processor with 3MB on-chip shared L2 cache	2,000 MFlops
• 32GB of 1066MHz DDR3 SDRAM	2,000 MTransfers/
•250GB 5400-rpm Serial ATA hard disk drive	50 MBytes/s

The disk which is 40 times slower is the slowest component!!

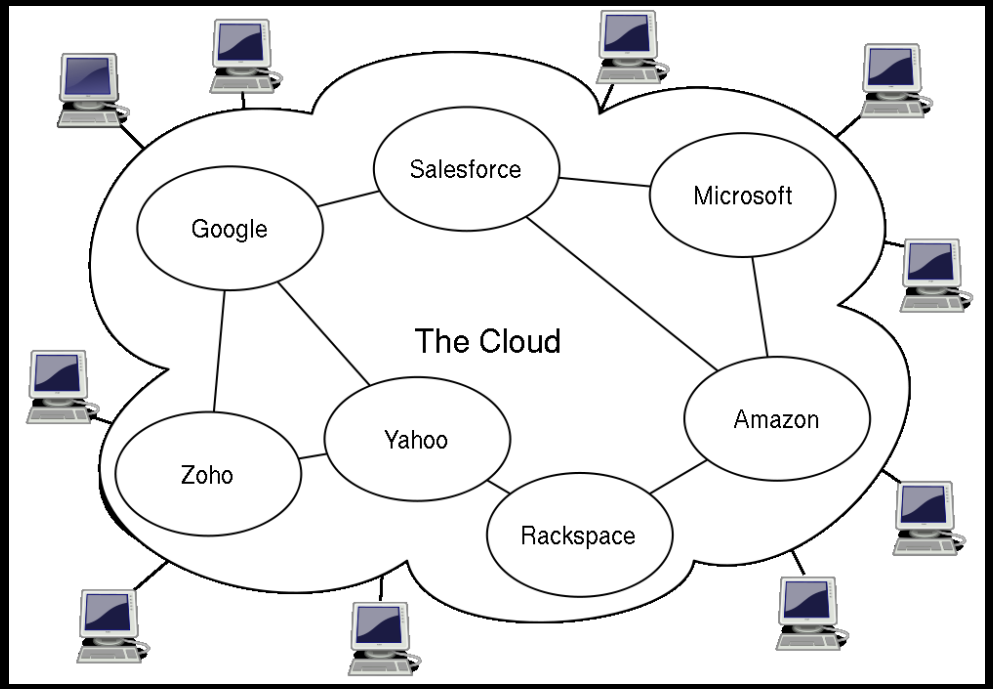
Managing Supercomputers

- Our secret weapon: Scientific Linux – Using built-in software tools, scripting languages such as Perl and Python we have automated the process of identifying slow or failing components that slow down the speed of the entire machine.
- I encourage all of you to use SL as much as possible while at Fermilab. linux-users@fnal.gov



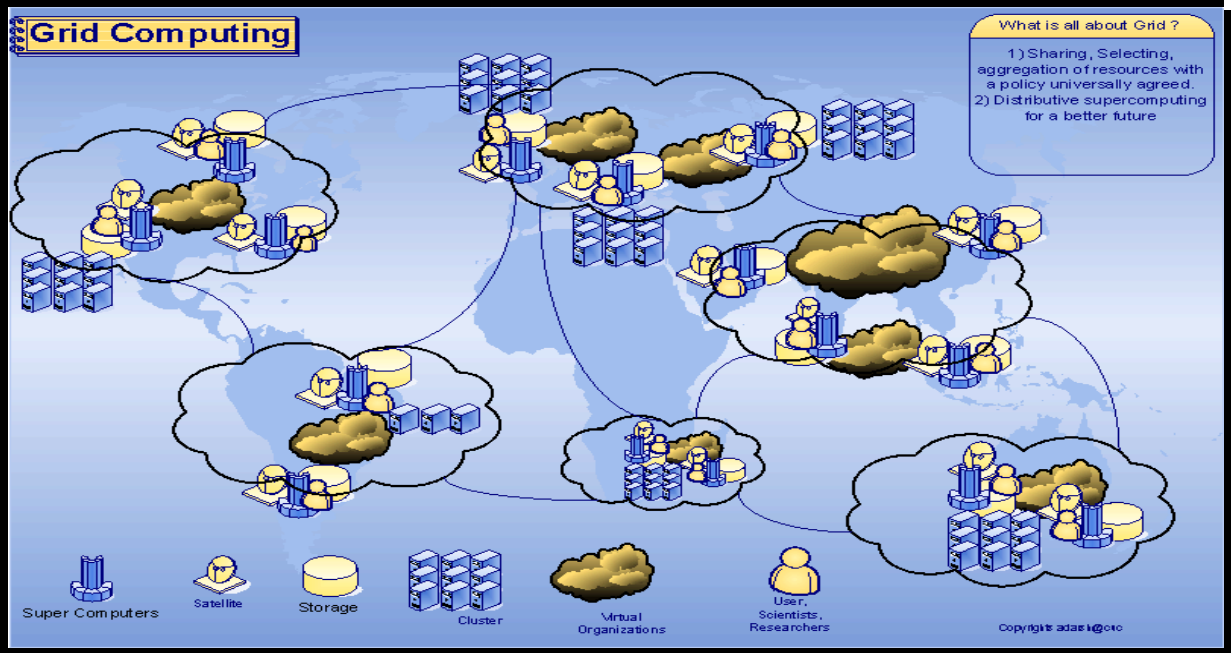
Is cloud computing HPC?

Cloud computing is Internet-based computing, whereby shared resources, software and information are provided to computers and other devices on-demand, like the electricity grid.



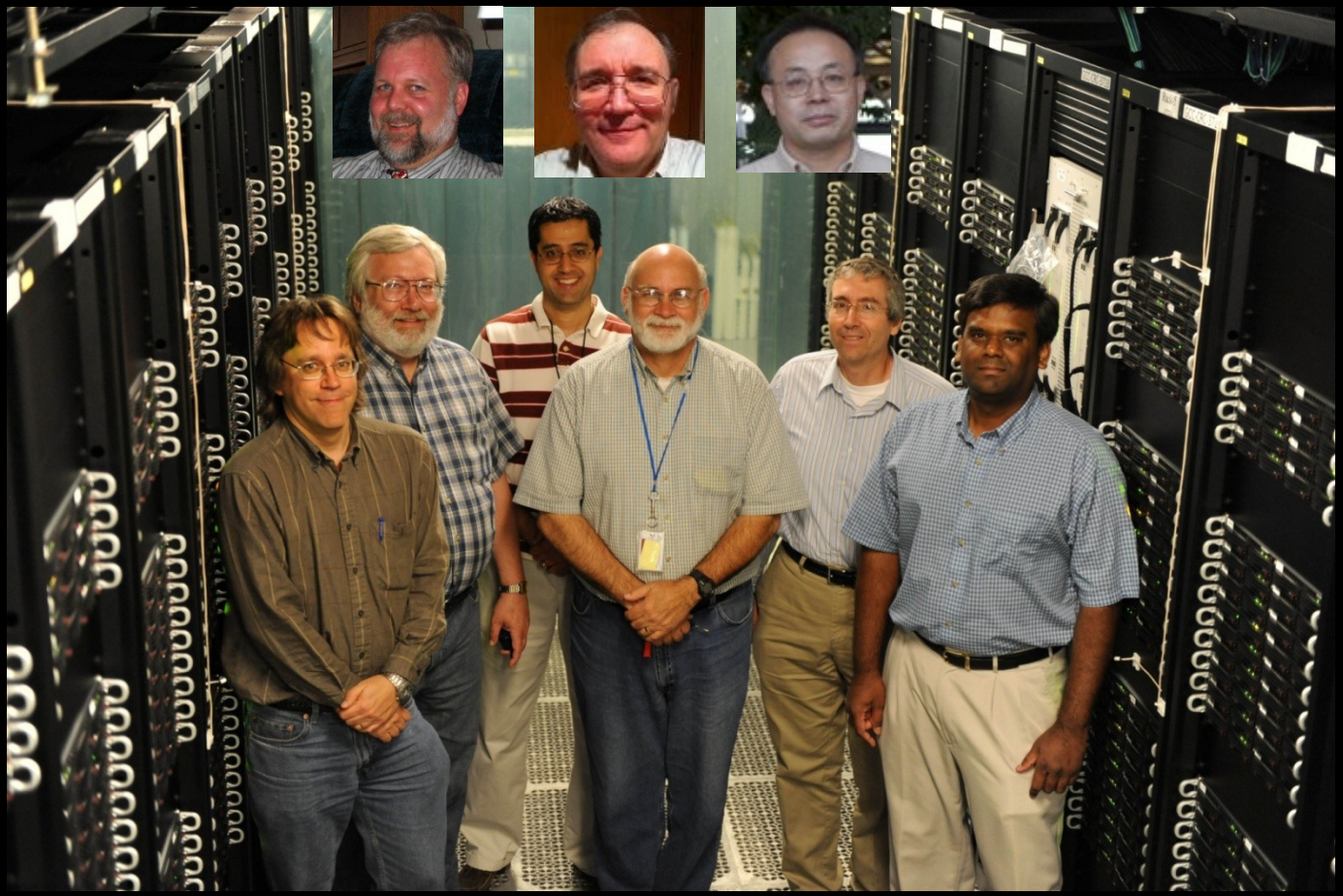
What is Grid Computing?

Grid computing is the combination of computer resources from multiple administrative domains for a common goal. Think SETI@home that uses internet connected computers for the Search for Extraterrestial Intelligence.



Computing Facilities

HPC @ Fermilab Crew



Also of HPPC; Bill Boroski, (Contractor Project Manager); Bakul Banerjee, (Associate Contractor Project Manager).
Paul Mackenzie of PPD is the (Chair of the USQCD Executive Committee).

Our esteemed users

Conclusion

It is an exciting time to be in the field of High Performance Computing which is at it’s peak in terms of potential, available hardware options and the variety of research that can be conducted using the supercomputing power provided by the world’s fastest super computers.