Tools for data analysis

School on High Energy Physics,
Bucharest-Magurele
27 October 2008

Gabriel Stoicea
Particle Physics Department
IFIN-HH/Bucharest

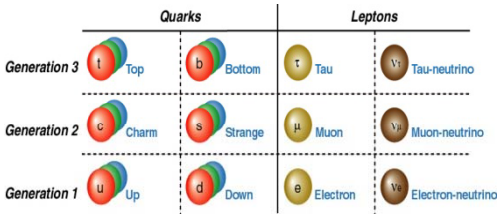
Outline

- Grid Computing
- OpenMP & GPU Computing

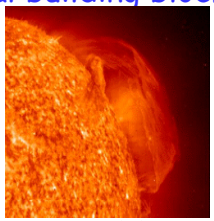
Grid Computing

- **Grid computing** is a form of distributed computing whereby a "super and virtual computer" is composed of a cluster of networked, loosely-coupled computers, acting in concert to perform very large tasks.
- **LCG (LHC Computing Grid)**

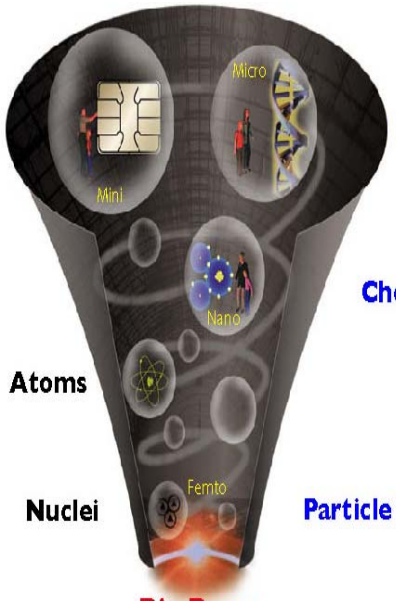
Particle Physics



Establish a periodic system of the fundamental building block

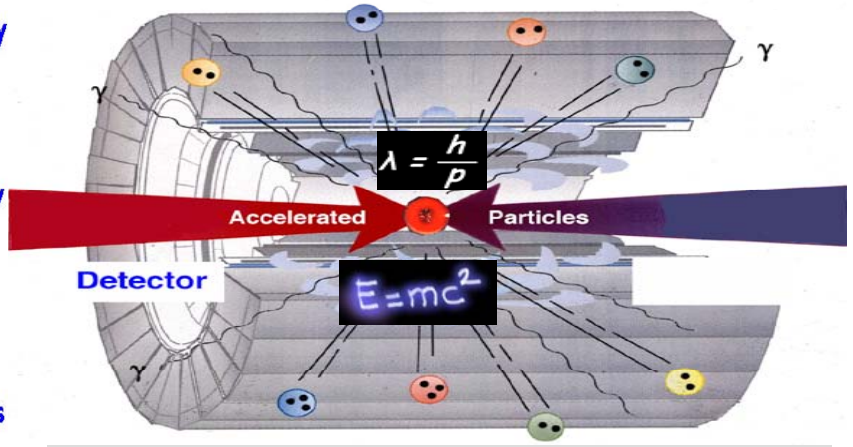


and understand forces



Methods of Particle Physics

The most powerful microscope

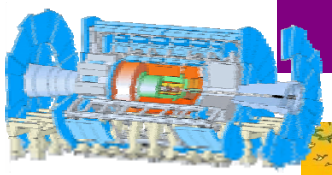


Creating conditions similar to the Big Bang

Particle Physics Challenges

Challenge 1: Large, distributed community

ATLAS

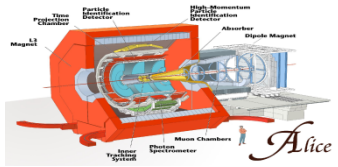


"Offline" software effort:
1000 person-years
per experiment

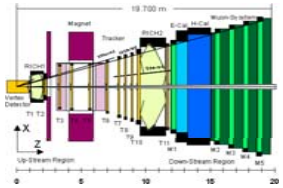
CMS

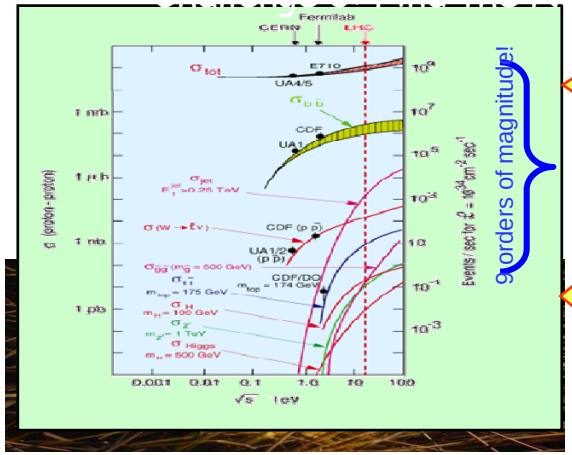
Software
life span:
20 years

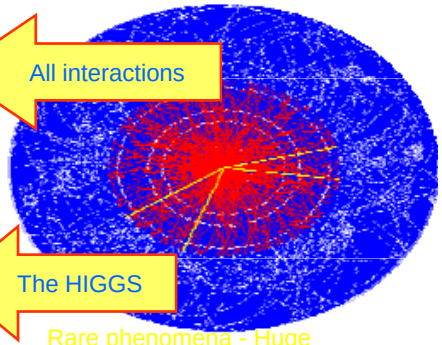
LHCb



~ 5000 Physicists
around the world
- around the clock





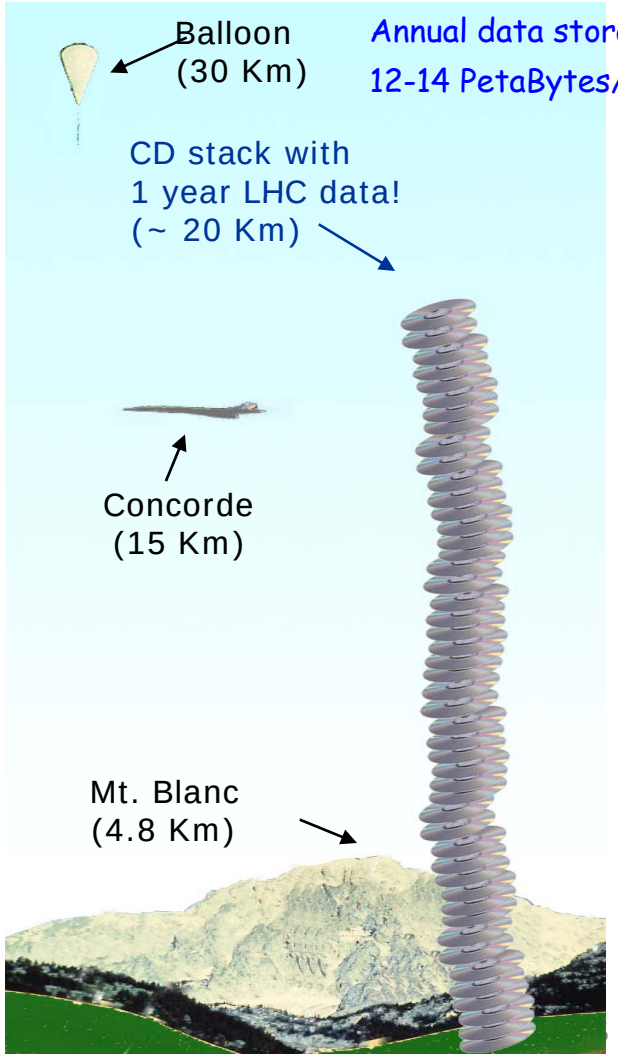


All interactions

The HIGGS

Rare phenomena - Huge background
Complex events

Challenge 2: Data Volume



What is the Grid? & How will it work?

- Resource Sharing
  - On a global scale, across the labs/universities
- Secure Access
  - Needs a high level of trust
- Resource Use
  - Load balancing, making most efficient use
- The "Death of Distance"
  - Requires excellent networking
- Open Standards
  - Allow constructive distributed development

- There is not (yet) a single Grid

The GRID middleware:

- Finds convenient places for the scientists "job" (computing task) to be run
- Optimises use of the widely dispersed resources
- Organises efficient access to scientific data
- Deals with authentication to the different sites that the scientists will be using
- Interfaces to local site authorisation and resource allocation policies
- Runs the jobs
- Monitors progress
- Recovers from problems

… and ….

Tells you when the work is complete and transfers the result back!

The LHC Computing Grid Project - LCG

Collaboration

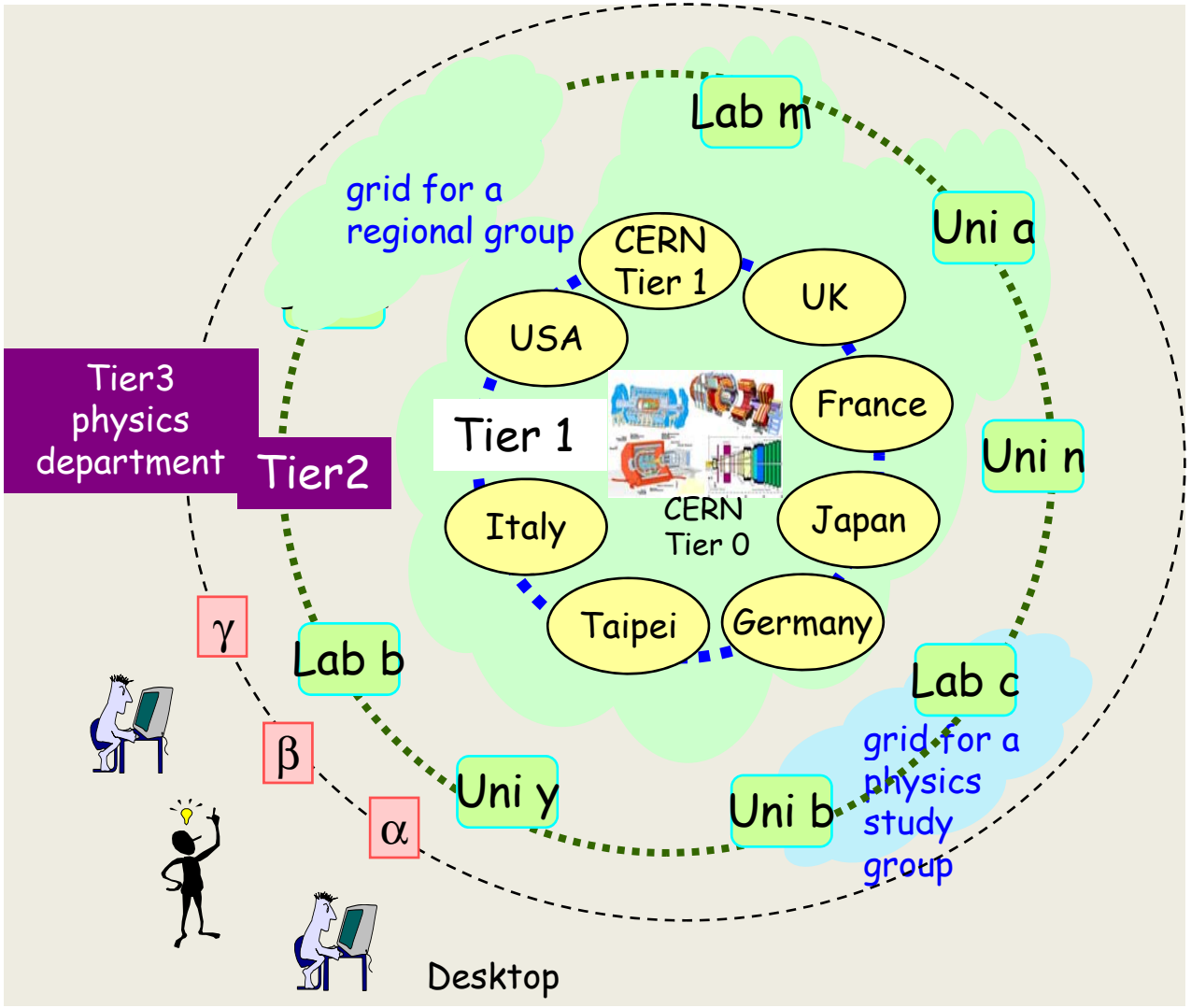
- LHC Experiments
- Grid projects: Europe, US
- Regional & national centres

Choices

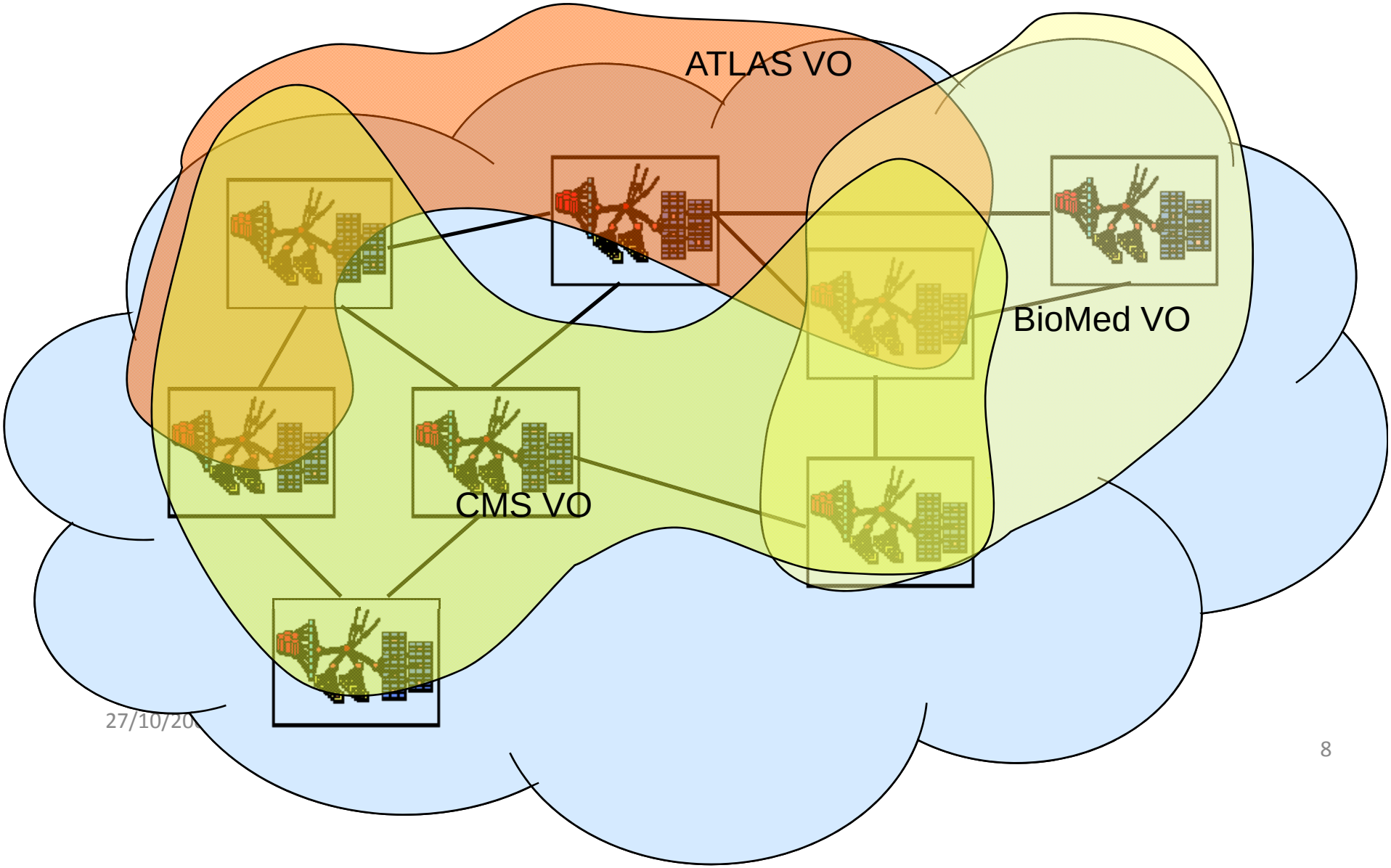
- Adopt Grid technology.
- Go for a “Tier” hierarchy.
- Use Intel CPUs in standard PCs
- Use LINUX operating system.

Goal

- Prepare and deploy the computing environment to help the experiments analyse the data from the LHC detectors.



Virtual Organizations for LHC and others



Romanian Tier2 Federation

Members: NIHAM (ALICE)

RO-02-NIPNE (ATLAS & H1),

RO-07-NIPNE (ALICE, ATLAS, H1, LHCb),

RO-11-NIPNE (LHCb)

Partners: ISS (ALICE), ICI, ITIM (ATLAS)

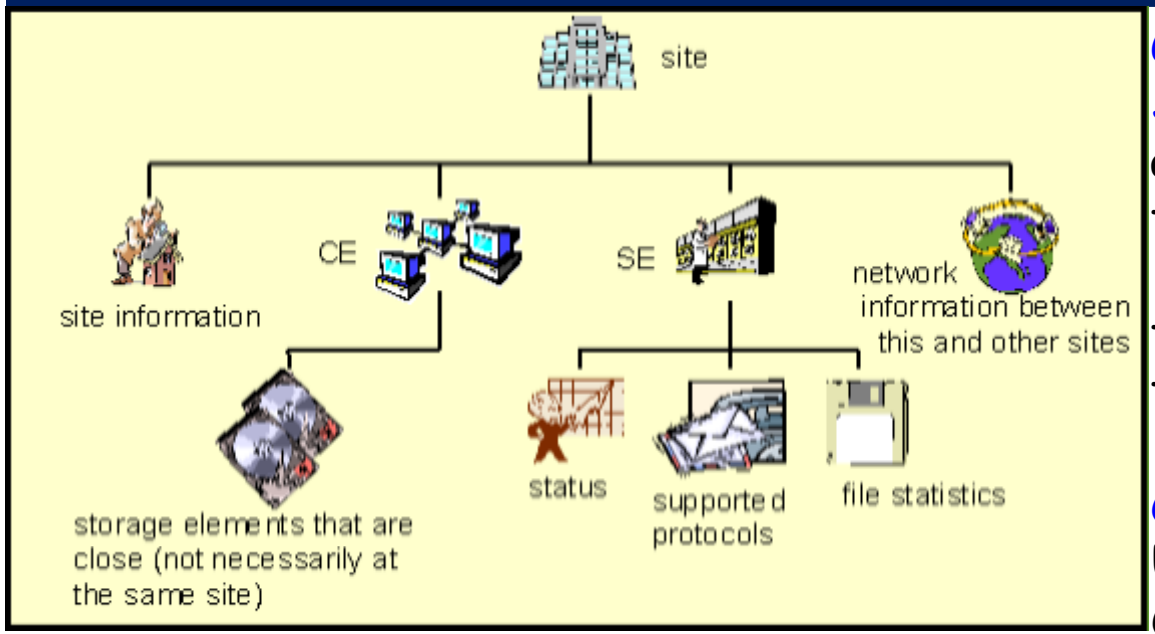
Agreed resources by NIPNE (kSI2k/TB)

	2007		2008		2009		2010		2011		2012		2013	
TOTAL	**463**	**89.86**	**1050**	**239**	**1700**	**424**	**2200**	**564**	**2650**	**705**	**3050**	**865**	**3450**	**1005**
NIHAM	**190**	**45**	**310**	**75**	**600**	**160**	**800**	**230**	**1000**	**300**	**1200**	**380**	**1400**	**450**
RO-02-NIPNE	**36**	**2.86**	**200**	**70**	**350**	**100**	**400**	**100**	**400**	**100**	**400**	**100**	400	100
RO-11-NIPNE	**50**	**2**	**140**	**4**	150	4	200	4	250	5	250	5	250	5
RO-07-NIPNE	**187**	**40**	**400**	**90**	**600**	**160**	**800**	**230**	**1000**	**300**	**1200**	**380**	**1400**	**450**
alice	70	20	150	45										
atlas	70	20	150	45										
lhcb	47	0	100	0										

Agreed resources by partners

TOTAL			**259**	**22.5**	**404.4**	**41**	**528**	**59.5**	**661.6**	**77**	**708**	**92**
ISS			**200**	**20**	**300**	**30**	**400**	**40**	**500**	**50**	**500**	**50**
ICI			**9**	**0.5**	**14.4**	**1**	**18**	**1.5**	**21.6**	**2**	**28**	**2**
ITIM			**50**	**2**	**90**	**10**	**110**	**18**	**140**	**25**	**180**	**40**

RO-02-NIPNE Grid Site



Grid middleware: gLite 3.1

Services:

atlasgw.nipne.ro: GW

tbat01.nipne.ro: CE, Site-BDII

tbat05.nipne.ro: SE

tbat02.nipne.ro: MONBOX

Cluster Configuration:

Using NAT

OS: Scientific Linux 4

Batch system: TORQUE/MAUI (OpenPBS)

WNs: 250 cores(x86_64) Xeon
RAM 2 GB/core
VMEM 2.4+ GB/core

Network:

internal: 1 Gbs

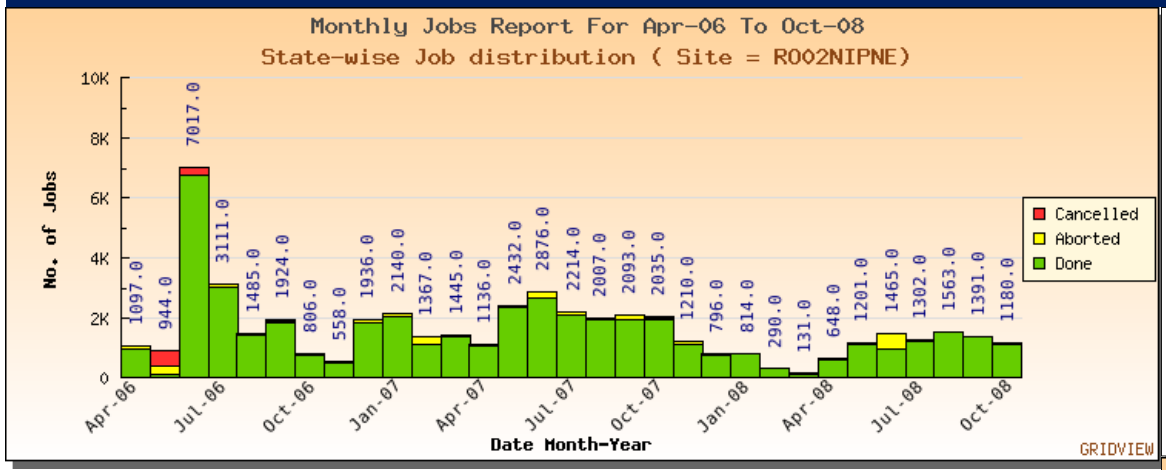
RoEduNet up-link: 10 Gbs

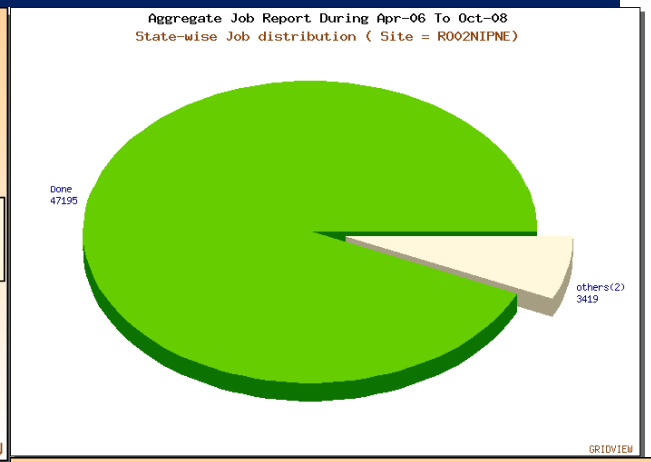
Storage:

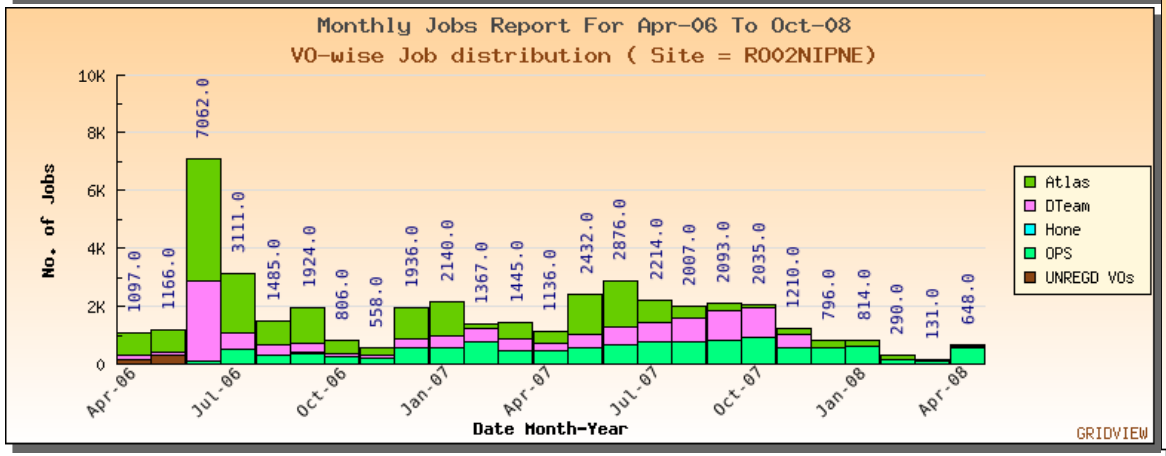
DPM (Disk Pool Manager) type

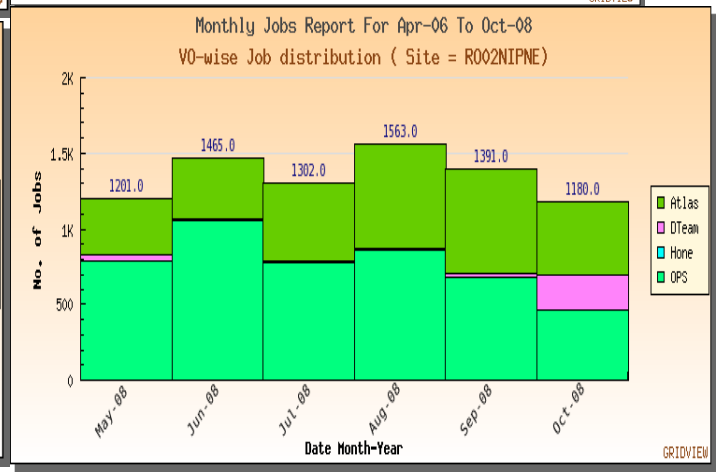
Raw Capacity ~ 75 TB

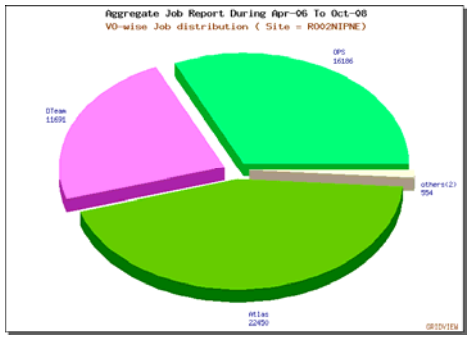
RO-02-NIPNE Statistics











Overall Efficiency: ~ 98.1 %

New Ways

OpenMP

an application programming interface (API) that supports multi-platform shared memory multiprocessing programming in C/C++ and Fortran on many architectures, including Unix and Microsoft Windows platforms. It consists of a set of compiler directives, library routines, and environment variables that influence run-time behavior.

Supported by Open Source Tools:

GCC >= 4.2

GPU Computing

With the increasing programmability of commodity graphics processing units (GPUs), these chips are capable of performing more than the specific graphics computations for which they were designed. They are now capable coprocessors, and their high speed makes them useful for a variety of applications.

Free of charge tools:

CUDA Programming Environment from NVIDIA

Summary

- Grids offer a way to solve Grand Challenge problems for the new era of HEP Experiments.
- Grids offer a way of using the information technology resources optimally inside an organization. They also provide a means for offering information technology as a utility for commercial and non-commercial clients, with those clients paying only for what they use, as with electricity or water.
- In the light of new CPU and GPU architectures new ways of parallel computing could be developed for HEP community.