

Scheduled = 15301 Running = 10525

## The LHC Computing Grid

Slides mostly by: Dr Ian Bird LCG Project Leader 18 March 2008





## Some precursors...



- Computing for HEP means data handling
  - Fixed-target experiments are <u>also</u> at the forefront
- COMPASS
  - >300 TB per year starting 2001 (still running)
  - Is used to investigating new computing technologies
    - One of the first reconstruction programmes entirely written using C++ and modern techniques
    - Raw data recorded at 35-70 MB/s (CMS expect ~200 MB/s)
    - Test of very-large database technologies
    - First user of CASTOR (transparent access of tape data)

## **Clusters of Inexpensive Processors**

#### **Requirements driven**

- We started this phase with a simple architecture that enables sharing of storage across CPU servers, that proved stable and has survived from RISC to Quad-core
- Parallel, high throughput
  Sustained price/perf
  improvement ~60% /yr
- Apollo DN10.000s
   1989 20 MIPS/proc
- 1990 SUN, SGI, IBM, HP, DEC, .... 5 each with its own flavour of Unix
- 1996 the first PC service
- 1998 COMPASS Computing farm
- 2008 dual quad core systems
  - $\rightarrow$  50K MIPS/chip  $\rightarrow$  ~20k cores available == ~20 MSI2K

### 5 orders of magnitude in 18 years



CERN**IT** Department

#### The LHC Data Challenge

- The accelerator will be completed in 2008 and run for 10-15 years
- Experiments will produce about
   **15 Million Gigabytes** of data
   each year (about 20 million CDs!)
- LHC data analysis requires a computing power equivalent to ~100,000 of today's fastest PC processors
- Requires many cooperating computer centres, as CERN can only provide ~20% of the capacity



Department

CERN IT Department CH-1211 Genève 23 Switzerland www.cern.ch/it

RC



Summary of Compu	iting Resou	irce Requi	rements	
All experiments - 2008	-	_		
From LCG TDR - June 2005				
	CERN	All Tier-1s	All Tier-2s	Total
CPU (MSPECint2000s)	25	56	61	142
Disk (PetaBytes)	7	31	19	57
Tape (PetaBytes)	18	35		53
CPU	Disk		Таре	
All Tier-2s 43% All Tier-1s 39%	CERN 12% 33% All Tier-1s 55%		All Tier-1s 66%	CERN 34%

## Solution: the Grid

 Use the Grid to unite computing resources of particle physics institutes around the world

The **World Wide Web** provides seamless access to information that is stored in many millions of different geographical locations

The **Grid** is an infrastructure that provides seamless access to computing power and data storage capacity distributed over the globe



Department

## How does the Grid work?

 It makes multiple computer centres look like a single system to the end-user

RC

CERN IT Department CH-1211 Genève 23

> Switzerland www.cern.ch/it

- Advanced software, called middleware, automatically finds the data the scientist needs, and the computing power to analyse it.
- Middleware balances the load on different resources. It also handles security, accounting, monitoring and much more.



Department

#### View of the ATLAS detector (under construction)



CERN

Department



## LHC Computing Grid project (LCG)

• More than 140 computing centres

RC

**CERN IT Department** 

CH-1211 Genève 23

Switzerland www.cern.ch/it

- 12 large centres for primary data management: CERN (Tier-0) and eleven Tier-1s
- 38 federations of smaller Tier-2 centres
- 35 countries involved



CERN

Department

## LCG Service Hierarchy

#### CERN**IT** Department

#### Tier-0: the accelerator centre

- Data acquisition & initial processing
- Long-term data safekeeping
- Distribution of data  $\rightarrow$  Tier-1 centres





Italy – CNAF (Bologna) Netherlands – NIKHEF/SARA (Amsterdam) Nordic countries – distributed Tier-1 Tier-1: "online" to the data acquisition process → high availability

- Managed Mass Storage –
   → grid-enabled data service
- Data-heavy analysis
- National, regional support

#### Tier-2: ~140 centres in ~35 countries

- Simulation
- End-user analysis batch and interactive



## WLCG Collaboration

#### The Collaboration

- 4 LHC experiments
- ~140 computing centres
- 12 large centres (Tier-0, Tier-1)
- 38 federations of smaller "Tier-2" centres
- ~35 countries
- Memorandum of Understanding
  - Agreed in October 2005, now being signed
- Resources

RC

CERN IT Department CH-1211 Genève 23

> Switzerland www.cern.ch/it

- Focuses on the needs of the four LHC experiments
- **Commits resources** 
  - each October for the coming year
  - 5-year forward look
- Agrees on standards and procedures
- Relies on EGEE and OSG (and other regional efforts)













ru-PNPI-LCG2

101 sites reporting

accounting data

#### CERN Grid activity

Department

- WLCG ran ~ 44 M jobs in 2007 workload has continued to increase – now at ~ 165k
- Distribution of work across Tier0/Tier1/Tier 2 really illustrates the importance of the grid system
  - Tier 2 contribution is around 50%; > 85% is

CERN

ASGC

CNAF FNAL

NDGF

PIC

RAL

TRIUMF

Tier-2s

CC-IN2P3

FZK-GridKA

NL-LHC-Tier-1

BNL

## Impact of the LHC Computing Grid in Europe Department

- LCG has been the driving force for the European multiscience Grid EGEE (Enabling Grids for E-sciencE)
- EGEE is now a global effort, and the largest Grid infrastructure worldwide
- Co-funded by the European Commission (Cost: ~130 M€ over 4 years, funded by EU ~70M€)
- EGEE already used for >100 applications, including...



**Bio-informatics** 



Education, Training

**Medical Imaging** 



## The EGEE project

#### • EGEE

- Started in April 2004, now in second phase with 91 partners in 32 countries
- 3<sup>rd</sup> phrase (2008-2010) in preparation



- Large-scale, production-quality grid infrastructure for e-Science
- Attracting new resources and users from industry as well as science
- Maintain and further improve "gLite" Grid middleware









# **egee**

## **Registered Collaborating Projects**

#### 25 projects have registered as of September 2007: web page





#### **Collaborating infrastructures**



Enabling Grids for E-sciencE

Archeology Astronomy Astrophysics Civil Protection Comp. Chemistry Earth Sciences Finance Fusion Geophysics High Energy Physics Life Sciences Multimedia Material Sciences

#### 21:13:50 UTC

Scheduled = 21539 Running = 25374

>250 sites
48 countries
>50,000 CPUs
>20 PetaBytes
>10,000 users
>150 VOs
>150,000 jobs/day



EGEE-II INFSO-RI-031688



## In silico drug discovery

- Diseases such as HIV/AIDS, SARS, Bird Flu etc. are a threat to public health due to world wide exchanges and circulation of persons
- Grids open new perspectives to *in silico* drug discovery
  - Reduced cost and adding an accelerating factor in the search for new drugs

# International collaboration is required for:

- Early detection
- Epidemiological watch
- Prevention
- Search for new drugs
- Search for vaccines







**Enabling Grids for E-sciencE** 

#### http://wisdom.healthgrid.org/



## For more information:





www.cern.ch/lcg



www.eu-egee.org



www.gridcafe.org

Thank you for your kind attention!

Ian Bird, CERN, IT Department



RC