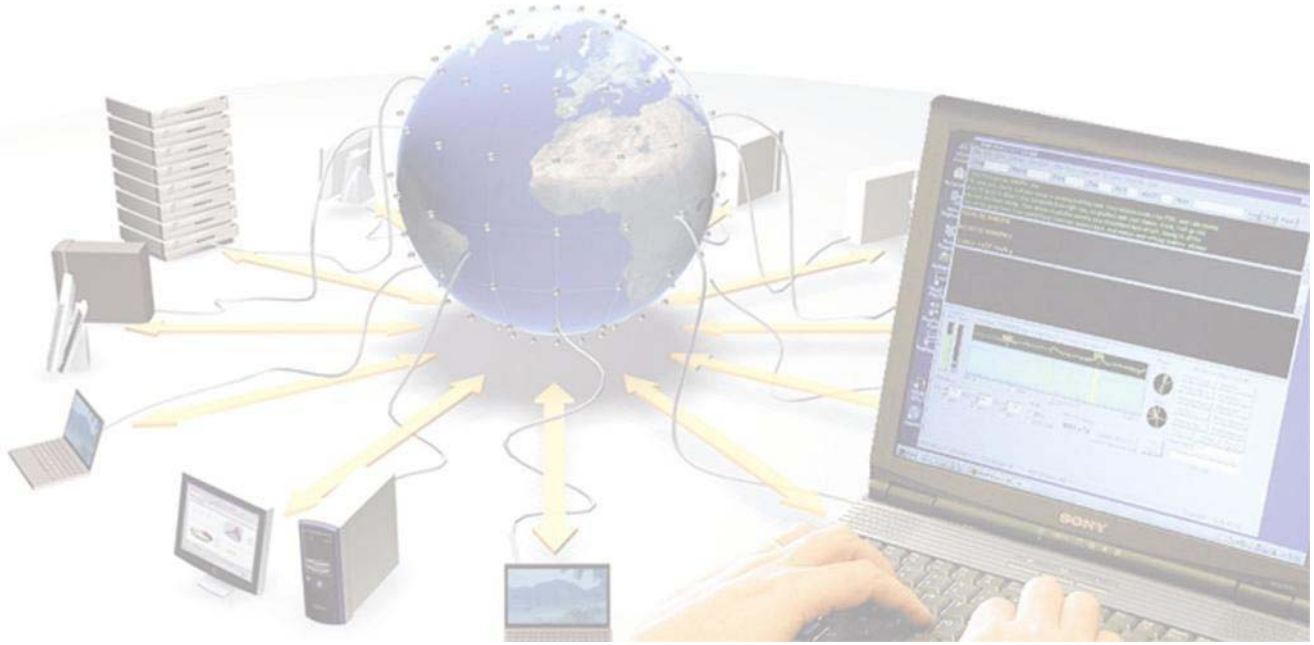
Enabling Grids for E-science in Europe

NA4 Open meeting
Catania
Jul 15-16

NA4/HEP work

F Harris (Oxford/CERN)

M.Lamanna(CERN)



EGEE is a project funded by the European Commission under contract IST-2003-508833

Contents

- Reminder of NA4/HEP aims in LCG/EGEE context
- Overview of LHC experiment data challenges and issues
- Overview of ARDA aims and status

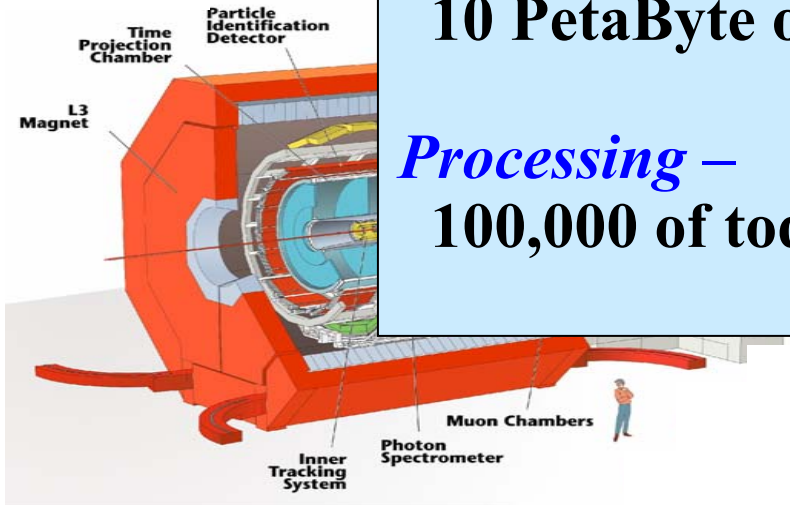
LHC Experiments

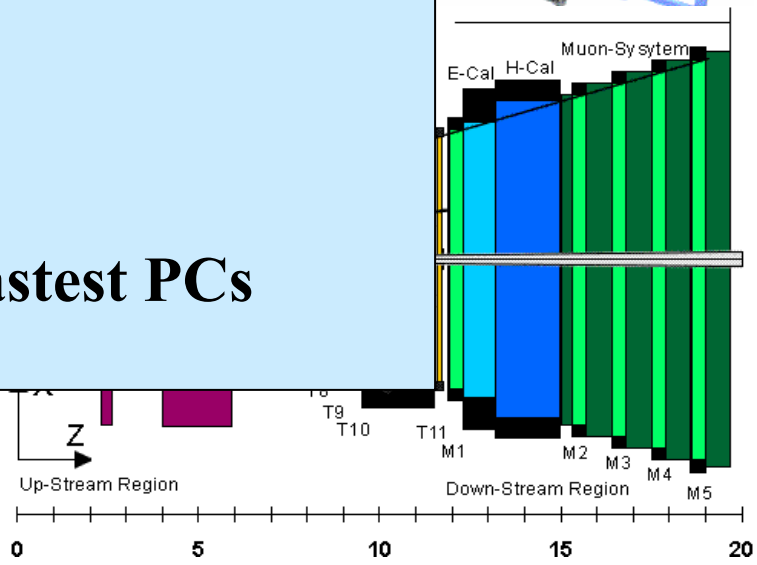
Enabling Grids for E-science in Europe

ATLAS

CMS

ALICE





Storage –

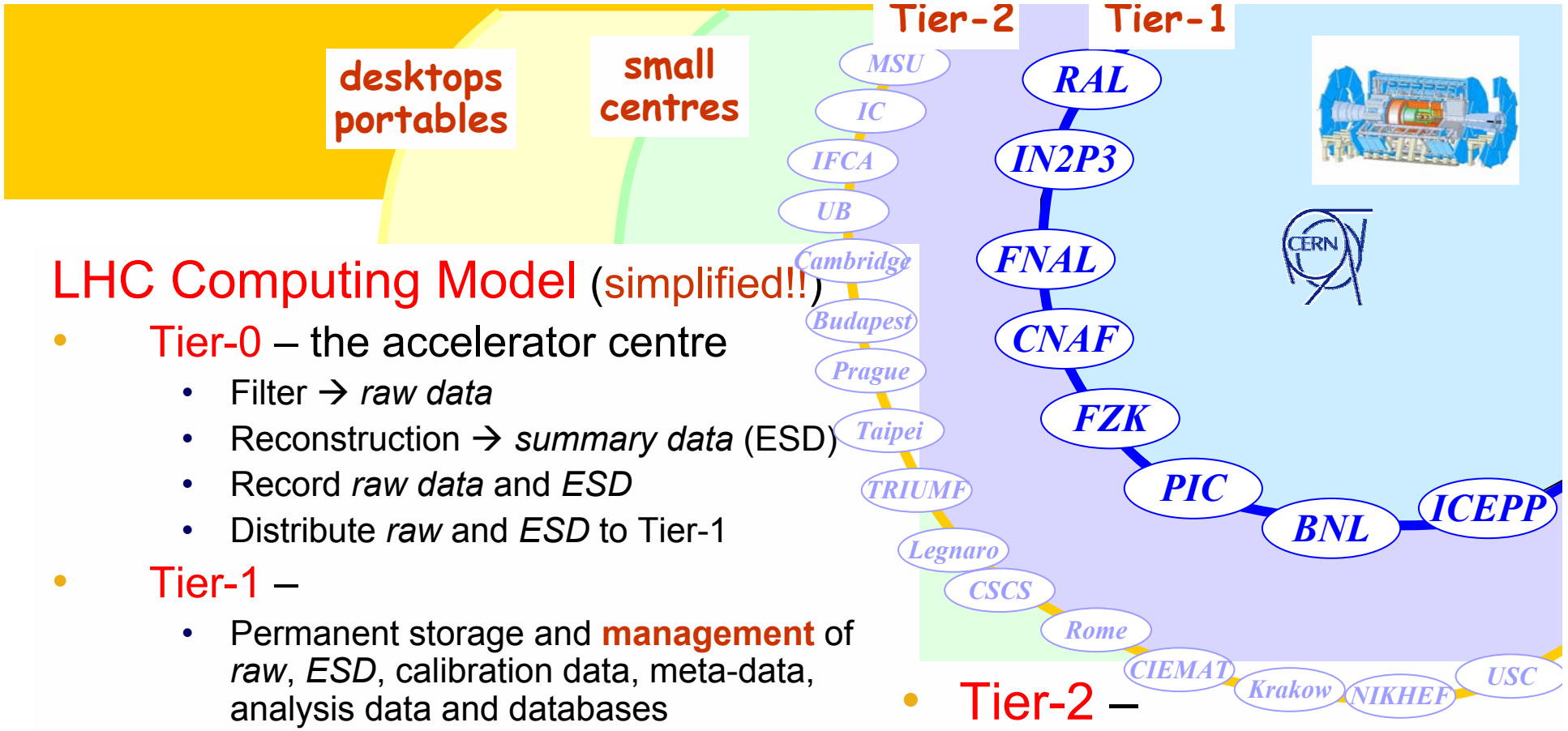
Raw recording rate 0.1 – 1 GByte/s

Accumulating at 5-8 PetaByte/year

10 PetaByte of disk

Processing –

100,000 of today's fastest PCs



LHC Computing Model (simplified!!)

- Tier-0 – the accelerator centre
  - Filter → *raw data*
  - Reconstruction → *summary data* (ESD)
  - Record *raw data* and *ESD*
  - Distribute *raw* and *ESD* to Tier-1
- Tier-1 –
  - Permanent storage and **management** of *raw*, *ESD*, calibration data, meta-data, analysis data and databases
    → **grid-enabled data service**
  - Data-heavy analysis
  - Re-processing raw → ESD
  - National, regional support

"online" to the data acquisition process high availability, long-term commitmen managed mass storage

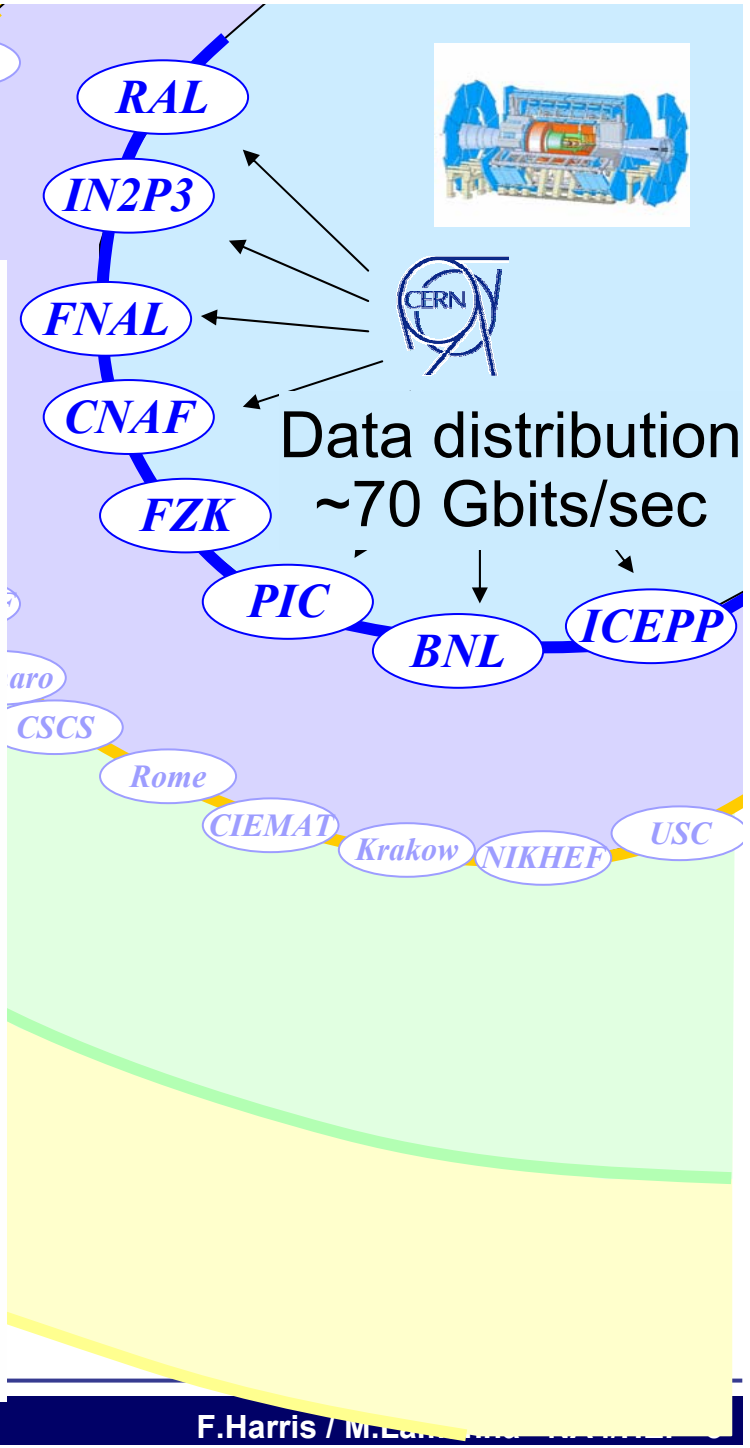
- Tier-2 –
  - Well-managed disk storage – grid-enabled
  - Simulation
  - End-user analysis – batch and interactive
  - High performance parallel analysis (PROOF)

Current estimates of Computing Resources needed at Major LHC Centres

First full year of data - 2008

	*Processing M SI2000***	*Disk PetaBytes*	*Mass Storage PetaBytes*
CERN	20	5	20
Major data handling centres (Tier 1)	45	20	18
Other large centres (Tier 2)	40	12	5
Totals	**105**	**37**	**43**

*** Current fast processor ~1K SI2000*



LCG service status Jul 9, 2004

HEP applications and data challenges using LCG-2

Enabling Grids for E-science in Europe

- **All have the same pattern of event simulation, reconstruction and analysis in production mode (with some limited 'user' analysis)**
- **All are testing their running models using Tier-0/1/2 with different levels of ambition, and doing physics with the data**
- **Larger scale user analysis to come with ARDA**
- **All have LCG-2 co-existing with use of other grids**

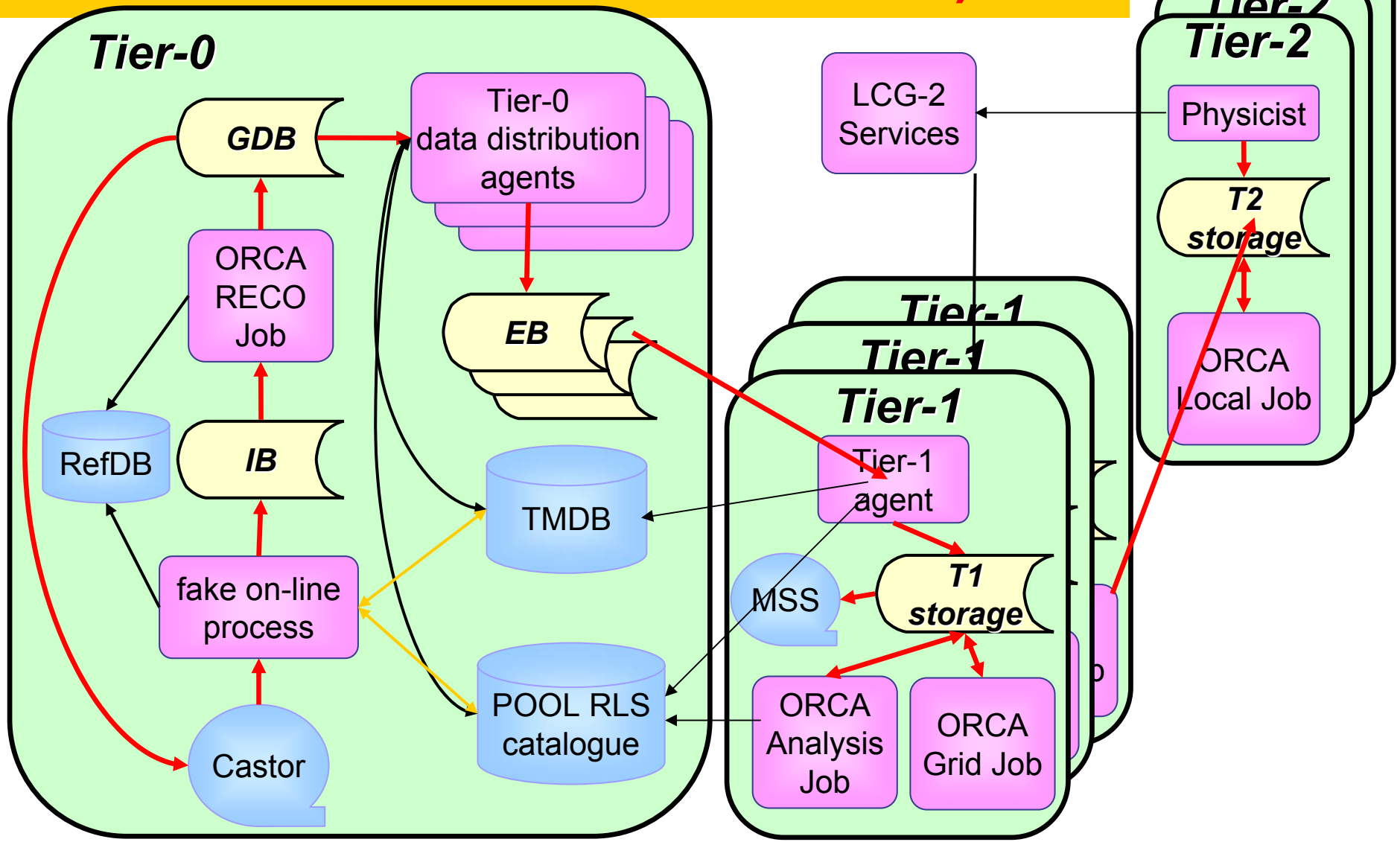
- ALICE and CMS started around February
- LHCb in May
- ATLAS just getting going
- D0 also making some use of LCG

- Next slides give a flavour of the work done so far by experiments
- Regular reports in LCG GDB and PEB

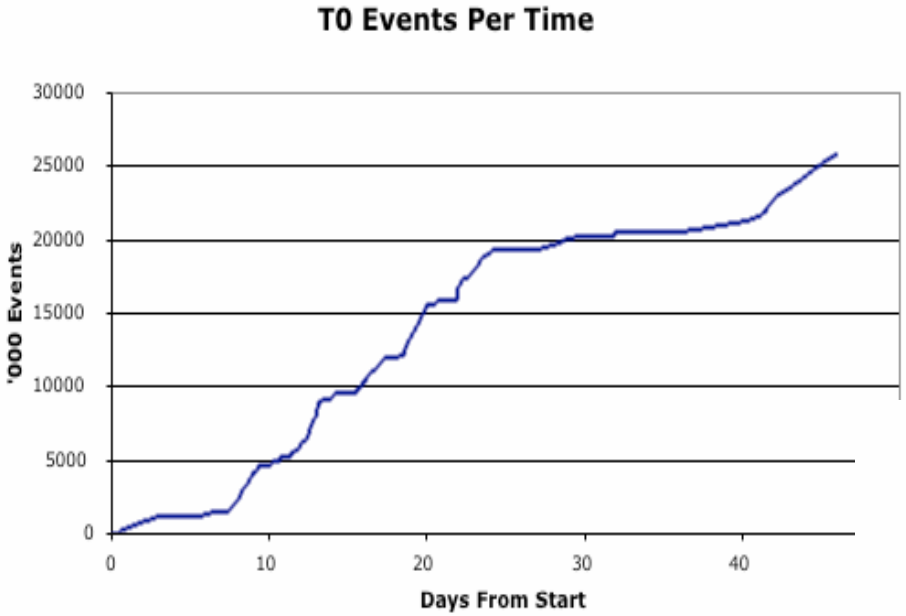
  see reports of June 14 at LCG/GDB

  http://agenda.cern.ch/fullAgenda.php?ida=a04114
- All very happy about LCG user-support 'attitude' – very cooperative

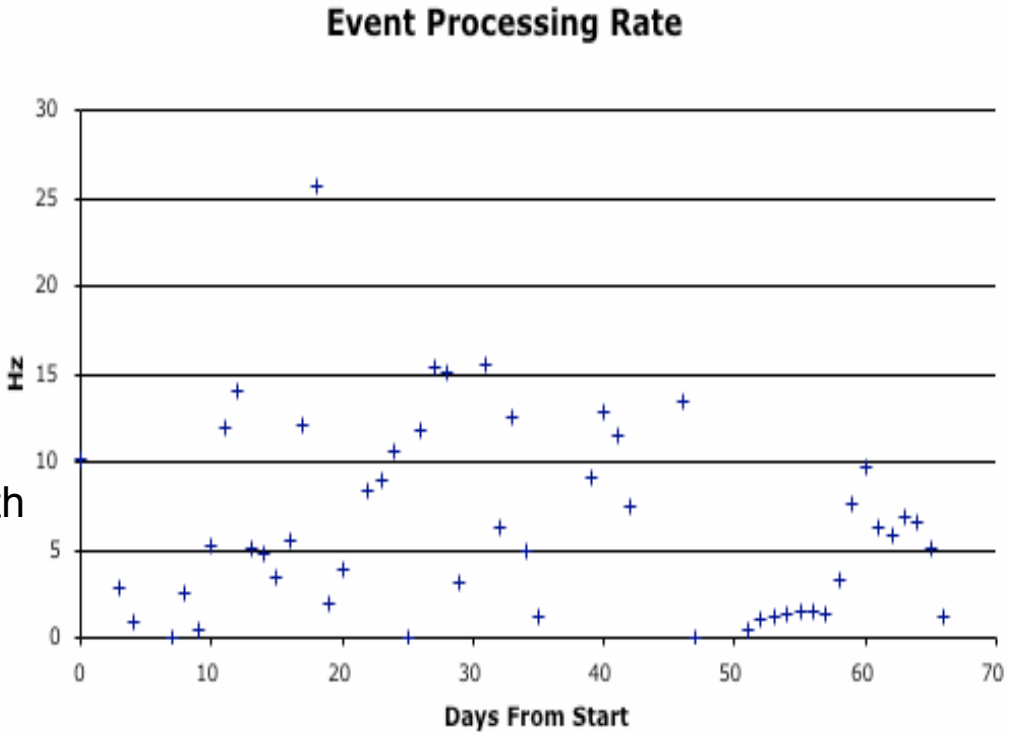
CMS DC04 layout (included US Grid3 and LCG-2 resources)



CMS DC04 Processing Rate



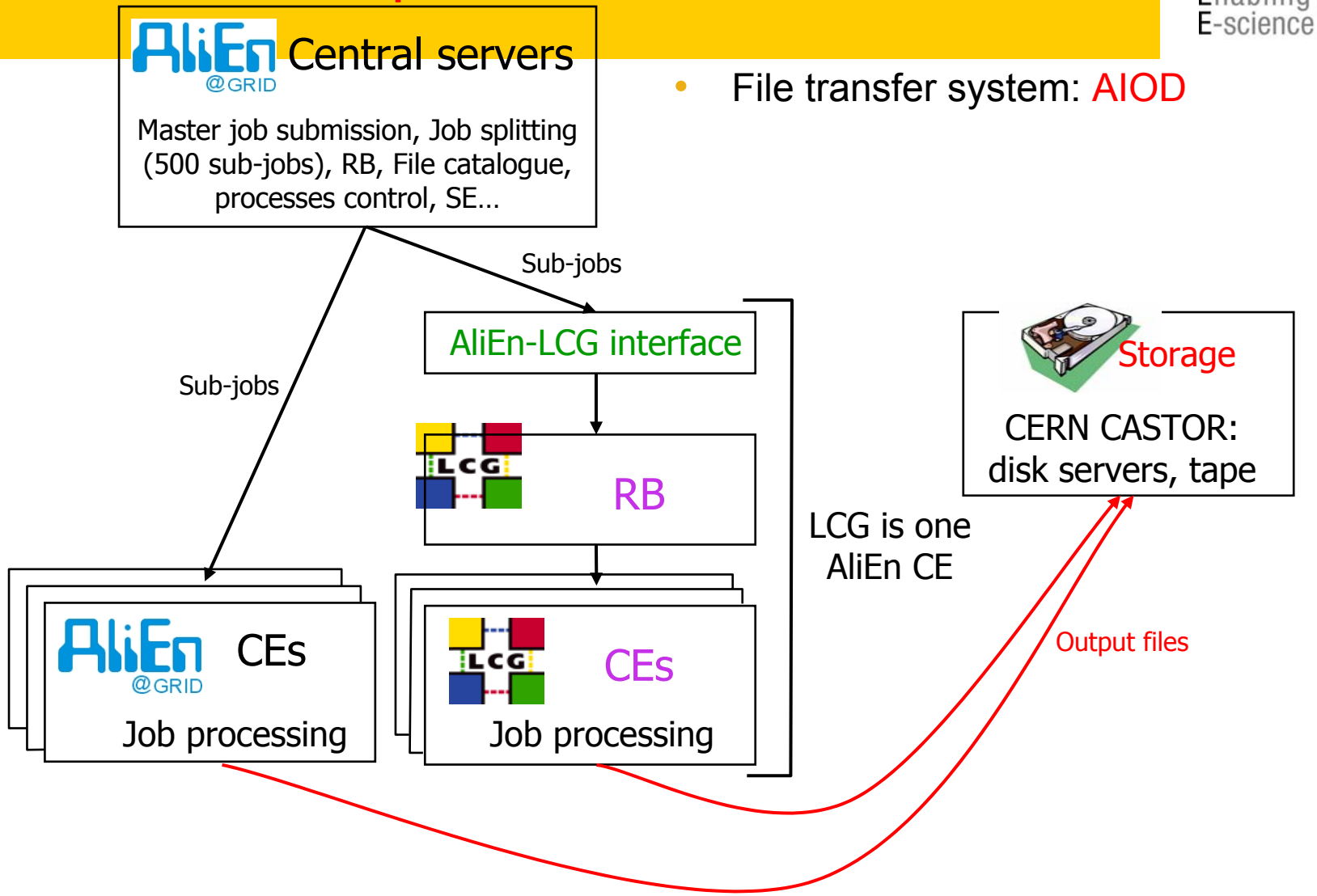
- Processed about 30M events
  - But DST “errors” make this pass not useful for analysis
- Generally kept up at T1’s in CNAF, FNAL, PIC



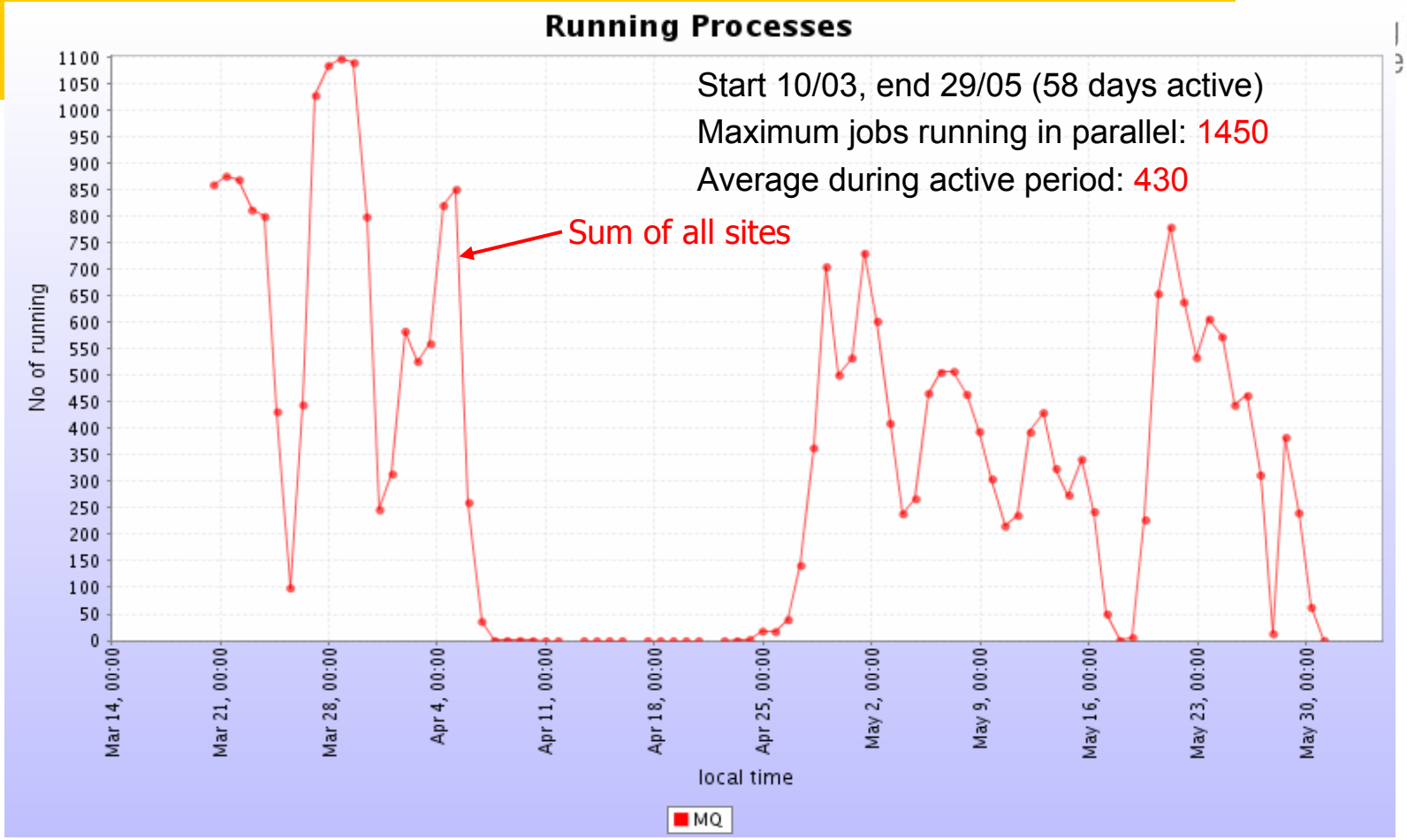
- Got above 25Hz on many short occasions
  - But only one full day above 25Hz with full system
- Working now to document the many different problems

Structure of ALICE event production

- File transfer system: AIOD



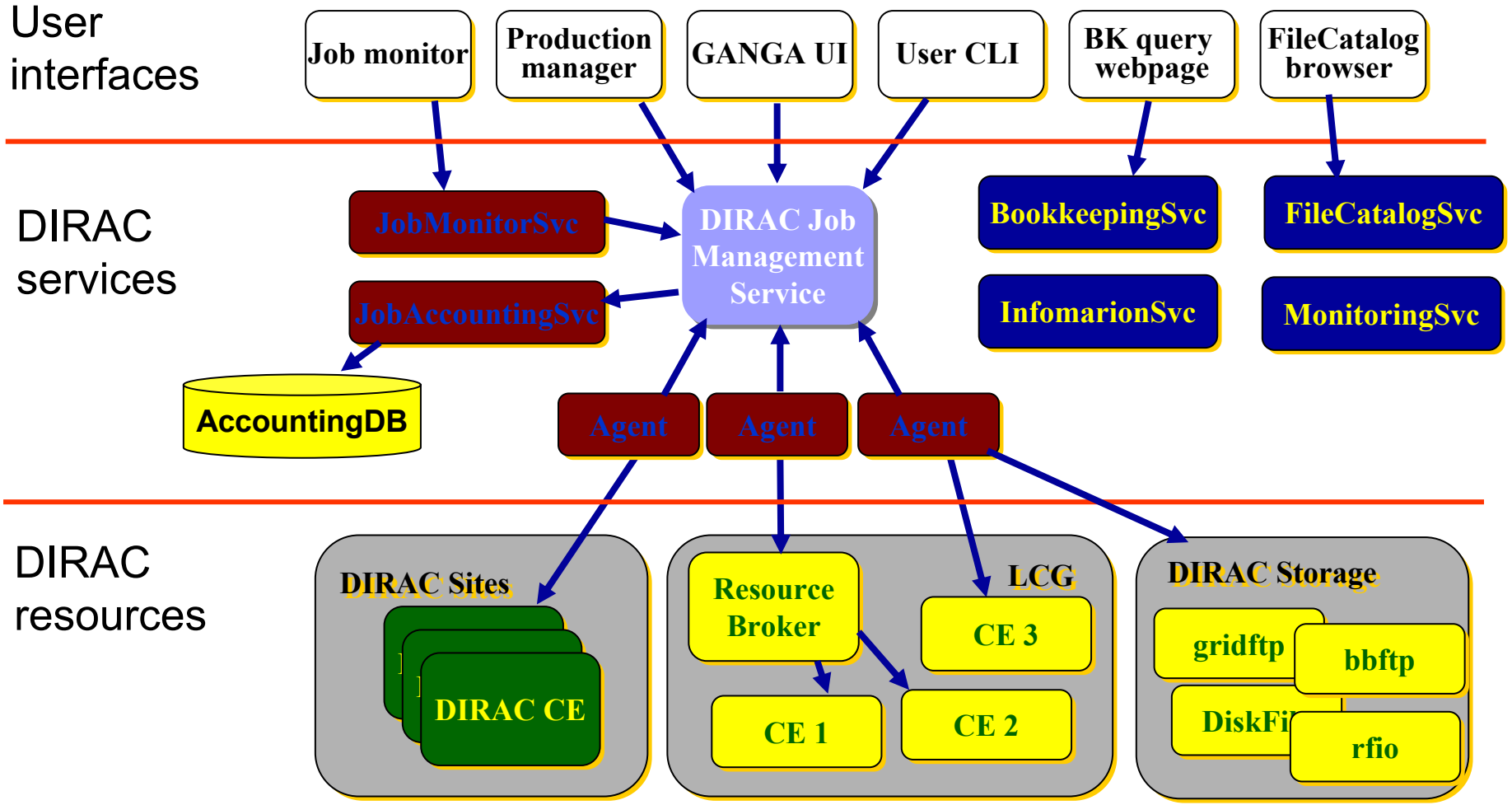
ALICE DATA CHALLENGE HISTORY



- CEs: Bari, Catania, Lyon, CERN-LCG, CNAF, Karlsruhe, Houston (Itanium), IHEP, ITEP, JINR, LBL, OSC, Nantes, Torino, Torino-LCG (grid.it), Pakistan, Valencia (+12 others)

LHCb Grid Production (driven by DIRAC)

Enabling Grids for E-science in Europe



List of running jobs - Mozilla <2>

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop http://fpegaes1.usc.es/dmon/joblist.html Search Print

Home Bookmarks Hotmail - The World's... Google [6e sens, le portail d... MMS, Texto, Email... ... Gestion en ligne/Acc... Gaudi Data Production

Accounting

Overview

Details

Production ID: All

Site:
- All
- ANY
- DIRAC.Barcelona.es
- DIRAC.Bologna.it
- DIRAC.CERN.ch

Job Status: All

App Status:
- All
- Application FAILED
- Application finished successfully
- Boole execution for step 4

Owner: All

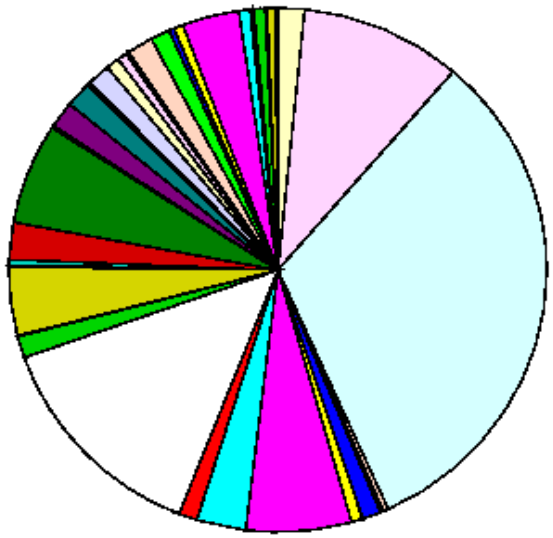
Max results: 40

Job Ids:

Submit Reset

LHCb DC'04 Monitoring

Output Data: 23.634 TB



@2004-07-03 Between 2004-05-03 – 2004-07-02

ANY	0.001%
CracowAgu	0.008%
DIRAC.Barcelona.es	1.582%
DIRAC.Bologna.it	9.716%
DIRAC.CERN.ch	31.95%
DIRAC.CracowAgu.pl	0.320%
DIRAC.IF-UFRJ.br	0.185%
DIRAC.IHEP-Protvino.ru	1.124%
DIRAC.IHEP2-Protvino.ru	0.682%
DIRAC.ITEP-Moscow.ru	6.334%
DIRAC.Imperial.uk	2.966%
DIRAC.JINR-Dubna.ru	1.108%
DIRAC.Karlsruhe.de	13.59%
DIRAC.LHCBONLINE.ch	1.334%
DIRAC.Liverpool.uk	0.057%
DIRAC.Lyon.fr	4.228%
DIRAC.Manno.ch	0.114%
DIRAC.Oxford.uk	0.331%
DIRAC.Santiago.es	2.282%
DIRAC.ScotGrid.uk	6.208%
DIRAC.Zurich.ch	0.134%
LCG.BHAM-HEP.uk	0.010%
LCG.CERN.ch	1.627%
LCG.CNAF.it	1.555%
LCG.Cambridge.uk	0.174%
LCG.FNAL.us	0.051%
LCG.FZK.de	1.380%
LCG.Imperial.uk	0.807%
LCG.KFKI.hu	0.564%
LCG.Krakow.pl	0.192%
LCG.Legnaro.it	1.564%
LCG.Milano.it	1.169%
LCG.NCU.tw	0.309%
LCG.NIKHEF.nl	0.566%
LCG.PIC.es	3.431%
LCG.RAL.uk	0.676%
LCG.Roma.it	0.132%
LCG.Sheffield.uk	0.084%
LCG.Torino.it	0.674%
LCG.Toronto.ca	0.116%
LCG.Triumf.ca	0.498%
LCG.USC.es	0.134%
Oxford	0.005%

Total | by Site (A) | by Site (#) | Running Jobs | by ProductionID

ATLAS DC2: goals

Enabling Grids for E-science in Europe

- The goals include:
  - **Full use of Geant4; POOL; LCG applications**
  - **Pile-up and digitization in Athena**
  - **Deployment of the complete Event Data Model and the Detector Description**
  - **Simulation of full ATLAS and 2004 combined Testbeam**
  - **Test the calibration and alignment procedures**
  - **Large scale physics analysis**
  - **Computing model studies (document end 2004)**
  - **Use widely the GRID middleware and tools**
  - **Run as much as possible of the production on Grids**
  - **Demonstrate use of multiple grids (LCG-2,Nordugrid,USGrid3)**

"Tiers" in ATLAS DC2 (rough estimate)

Enabling Grids for E-science in Europe

Country	"Tier-1"	Sites	Grid	kSI2k
Australia			NG	12
Austria			LCG	7
Canada	TRIUMF	7	LCG	331
CERN	CERN	1	LCG	700
China			LCG	30
Czech Republic			LCG	25
France	CCIN2P3	1	LCG	~ 140
Germany	GridKa	3	LCG	90
Greece			LCG	10
Israel		2	LCG	23
Italy	CNAF	5	LCG	200
Japan	Tokyo	1	LCG	127
Netherlands	NIKHEF	1	LCG	75
NorduGrid	NG	~30	NG	380
Poland			LCG	80
Russia			LCG	~ 70
Slovakia			LCG	
Slovenia			NG	
Spain	PIC	4	LCG	50
Switzerland			LCG	18
Taiwan	ASTW	1	LCG	78
UK	RAL	8	LCG	~ 1000
US	BNL	28	Grid3/LCG	~ 1000
Total				~ 4500

General comments on data challenges

Enabling Grids for E-science in Europe

- All experiments making production use of LCG-2 – stability has steadily improved. LCG have learned from feedback of experiments. This will be ongoing till 2005

- Some issues being followed with LCG in a co-operative manner (Following input from CMS,ALICE and LHCb)

  - ***Mass Storage (SRM) support for variety of devices (not just CASTOR)***
  - ***Debugging is hard when problems arise (develop monitoring)***
  - ***Flexible s/w installation for analysis still being developed***
  - ***Site Certification being developed***
  - ***Disk Storage reservation fro jobsa***
  - ***Developing use of RB(batches of jobs,job distribution to sites…)***
  - ***File transfer stability***
  - ***RLS performance issues (use of metadata and catalogues)***

- We are learning…data challenges continuing
- Experiments using multi-grids
- Looking to ARDA for user prototype analysis

ARDA (Architectural Roadmap for Distributed Analysis)

Enabling Grids for E-science in Europe