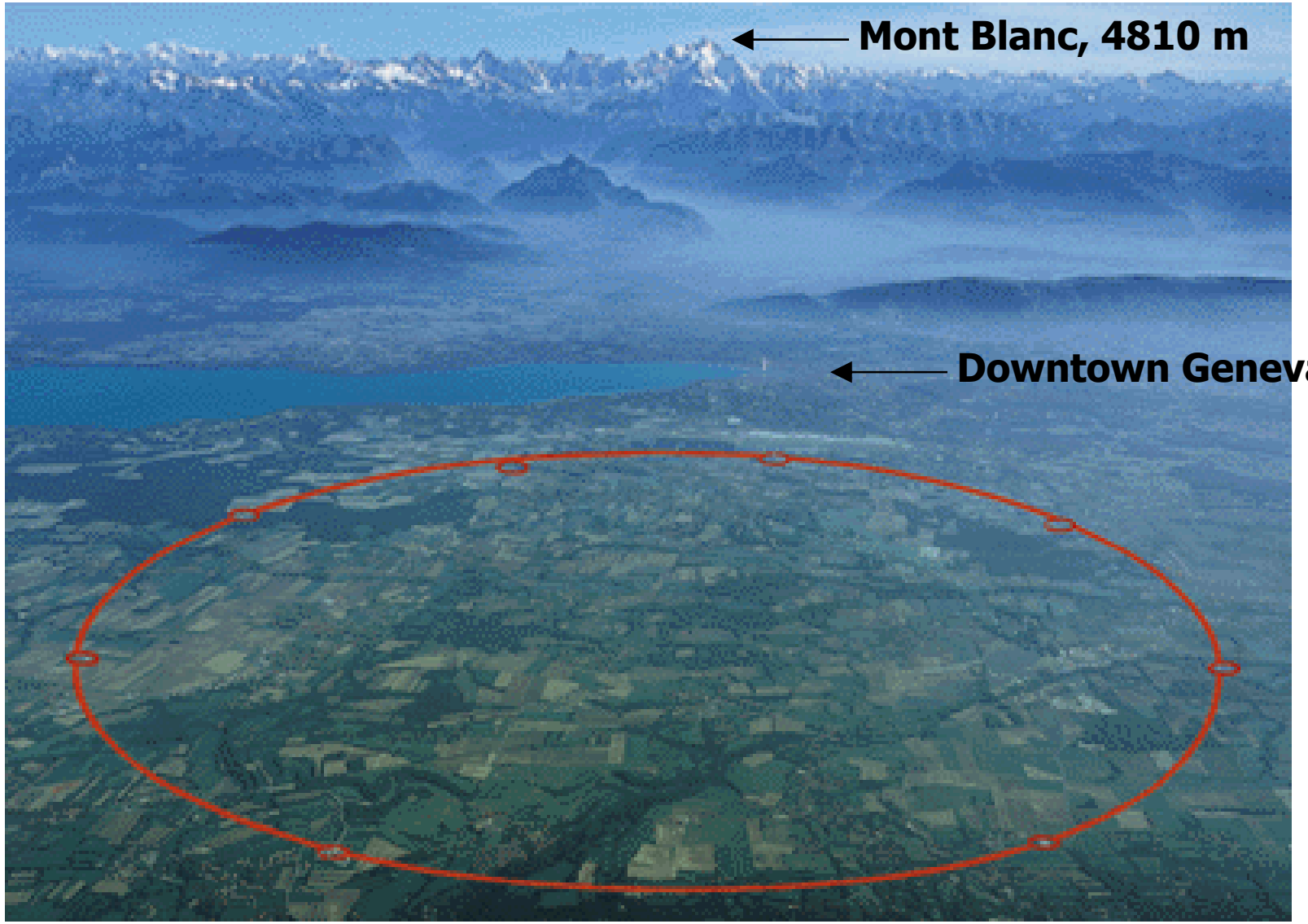
LHC Computing Grid Project (LCG): Status and Prospects

David Foster
Communications Systems Group Leader
Information Technology Department
CERN

Ian Bird
Grid Deployment Group Leader
Information Technology Department
CERN

Korea Digital Divide Workshop
May 2005

CERN site:
Next to Lake Geneva



LHC data (simplified)

Per experiment:

- 40 million collisions per second
- After filtering, 100 collisions of interest per second
- A Megabyte of digitised information for each collision = recording rate of 100 Megabytes/sec
- 1 billion collisions recorded = 1 Petabyte/year

With four experiments, processed data we will accumulate 15 PetaBytes of new data each year

1 Megabyte (1MB)
A digital photo

1 Gigabyte (1GB)
= 1000MB
A DVD movie

1 Terabyte (1TB)
= 1000GB
World annual book production

1 Petabyte (1PB)
= 1000TB
10% of the annual production by LHC experiments

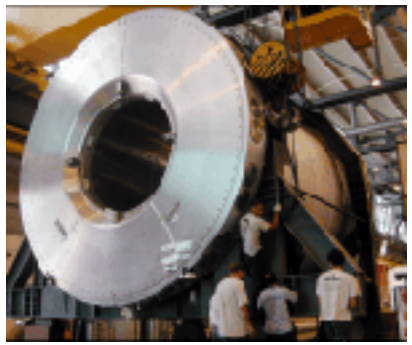
1 Exabyte (1EB)
= 1000 PB
World annual information production

= 1% of

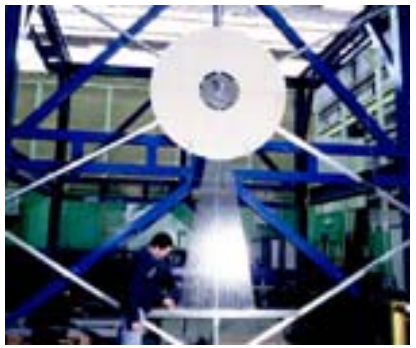
CMS

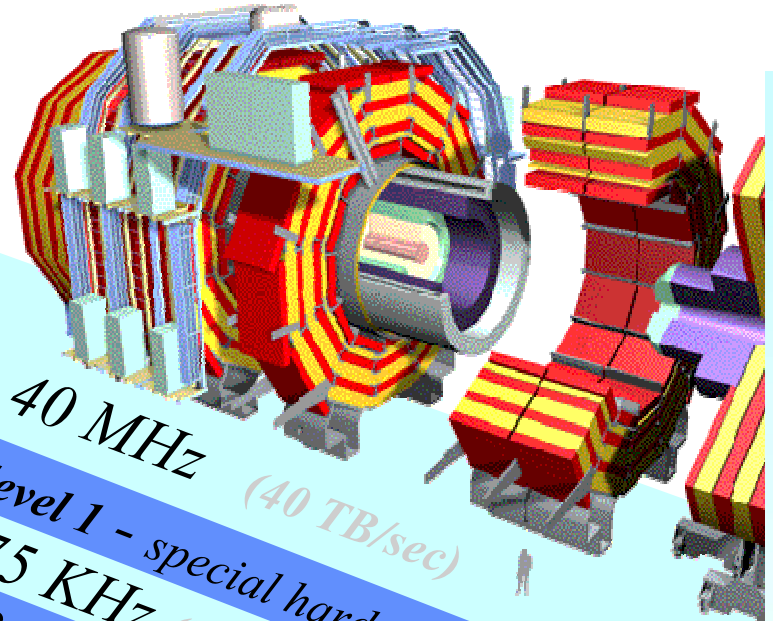
LHCb

ATLAS



ALICE





40 MHz (40 TB/sec)
level 1 - special hardware
75 KHz (75 GB/sec)
level 2 - embedded processors
5 KHz (5 GB/sec)
level 3 - PCs
100 Hz
(100 MB/sec)
data recording & offline analysis

~15 PetaBytes of data each year
Analysis will need the computing power of ~ 100,000 of today's fastest PC processors!

LCG – Goals

- The goal of the LCG project is to prototype and deploy the computing environment for the LHC experiments

- Two phases:

  - **Phase 1: 2002 – 2005**
  - Build a service prototype, based on existing grid middleware
  - Gain experience in running a production grid service
  - Produce the TDR for the final system

  - **Phase 2: 2006 – 2008**
  - Build and commission the initial LHC computing environment

LCG is not a development project – it relies on other grid projects for grid middleware development and support

LHC Computing Grid Project - a Collaboration

Building and operating the LHC Grid –
a global collaboration between

- The physicists and computing specialists from the LHC experiments
- The projects in Europe and the US that have been developing Grid middleware
- The regional and national computing centres that provide resources for LHC
- The research networks

Researchers

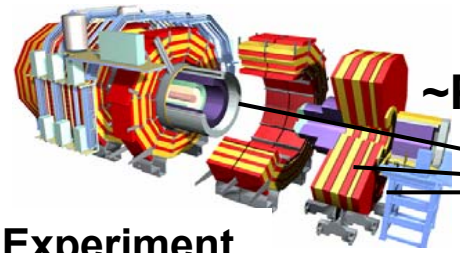
Computer Scientists & Software Engineers

Service Providers

Enabling Grids for E-science in Europe

Virtual Data Toolkit

LHC Computing Hierarchy



Experiment

~PByte/sec

Online System

~100-1500 MBytes/sec

CERN/Outside Resource Ratio ~1:2
Tier0/(Σ Tier1)/(Σ Tier2) ~1:1:1

Tier 0 +1

CERN Center
PBs of Disk;
Tape Robot

Tier 1

~2.5-10 Gbps

IN2P3 Center

RAL Center

INFN Center

FNAL Center

2.5-10 Gbps

Tier 2

~2.5-10 Gbps

Tier2 Center

Tier 3

Institute

Physics data cache

0.1 to 10 Gbps

Workstations

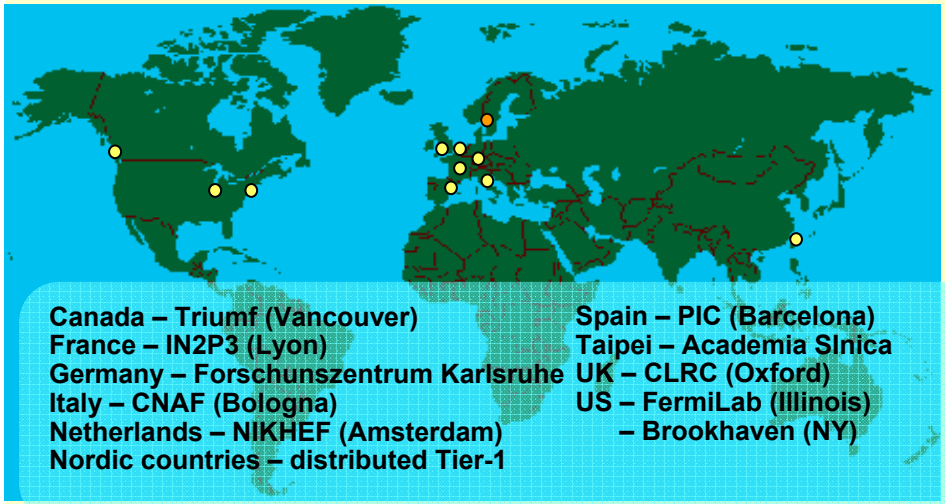
Tier 4

Tens of Petabytes by 2007-8.
An Exabyte ~5-7 Years later.

LCG Service Hierarchy

Tier-0 – the accelerator centre

- Data acquisition & initial processing
- Long-term data curation
- Distribution of data → Tier-1 centres



Canada – Triumf (Vancouver)
France – IN2P3 (Lyon)
Germany – Forschunszentrum Karlsruhe
Italy – CNAF (Bologna)
Netherlands – NIKHEF (Amsterdam)
Nordic countries – distributed Tier-1
Spain – PIC (Barcelona)
Taipei – Academia SInica
UK – CLRC (Oxford)
US – FermiLab (Illinois)
– Brookhaven (NY)

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 – ~100 centres in ~40 countries

- Simulation
- End-user analysis – batch and interactive

Project Areas & Management

Project Leader
Les Robertson
Resource Manager – Chris Eck
Planning Officer – Jürgen Knobloch
Administration – Fabienne Baud-Lavigne

Distributed Analysis - ARDA
Massimo Lamanna
Prototyping of distributed end-user analysis using grid technology

Applications Area
Pere Mato
Development environment
Joint projects, Data management
Distributed analysis

Middleware Area
Frédéric Hemmer
Provision of a base set of grid middleware (acquisition, development, integration)
Testing, maintenance, support

CERN Fabric Area
Bernd Panzer
Large cluster management
Data recording, Cluster technology
Networking, Computing service at CERN

Grid Deployment Area
Ian Bird
*Establishing and managing the **Grid Service** - Middleware, certification, security operations, registration, authorisation, accounting*

Joint with EGEE

Relation of LCG and EGEE

Enabling Grids for E-science in Europe

- **Goal**
  Create a European-wide production quality multi-science grid infrastructure on top of national & regional grid programs
- **Scale**
  70 partners in 27 countries
  Initial funding (€32M) for 2 years
- **Activities**
  Grid operations and support (joint LCG/EGEE operations team)
  Middleware re-engineering (close attention to LHC data analysis requirements)
  Training, support for applications groups (inc. contribution to the ARDA team)
- **Builds on**
  LCG grid deployment
  Experience gained in HEP
  LHC experiments → pilot applications

CERN Fabric

- Fabric automation has seen very good progress
  - The new systems for managing large farms are in production at CERN since January
- New CASTOR Mass Storage System
  - Was deployed first on the high throughput cluster for the recent ALICE data recording computing challenge
- Agreement on collaboration with Fermilab on Linux distribution
  - Scientific Linux based on Red Hat Enterprise 3
  - Improves uniformity between the HEP sites serving LHC and Run 2 experiments

CERN Fabric

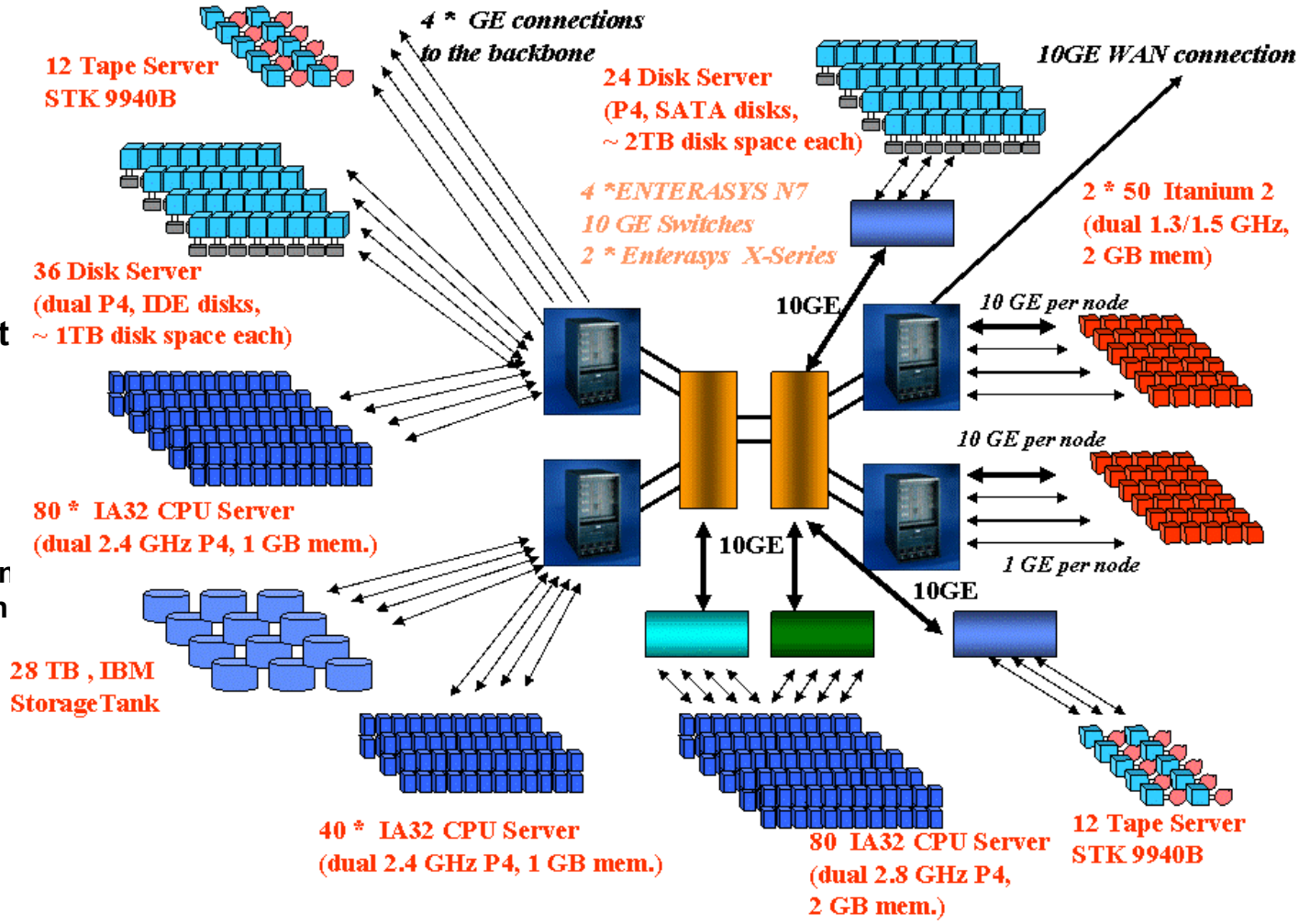
- CERN computer centre preparations
  - Power upgrade to 2.5 MW
  - Computer centre refurbishment well under way
  - Acquisition process started

Preparing for 7,000 boxes in 2008

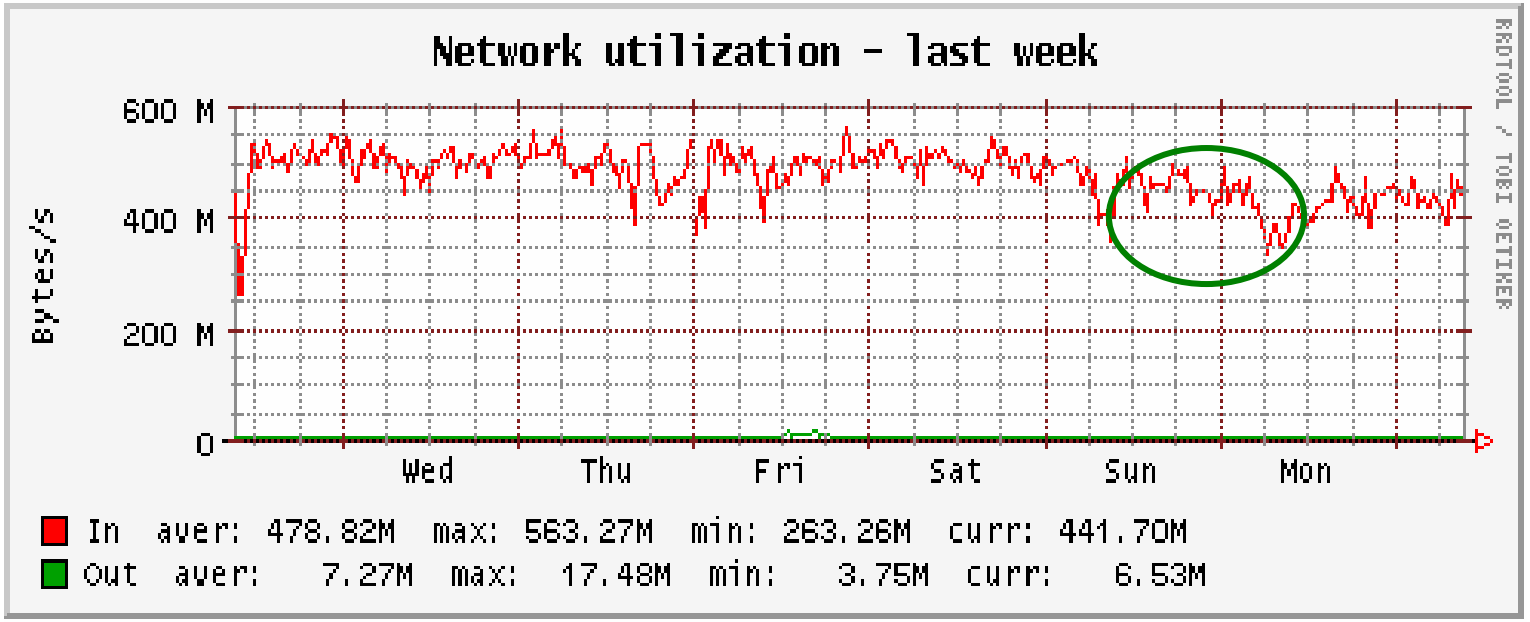
High Throughput Prototype openlab/LCG

CERN
openlab for DataGrid applications

- Experience with likely ingredients in LCG:
  - 64-bit programming
  - next generation I/O (10 Gb Ethernet, Infiniband, etc.)
- High performance cluster used for evaluations, and for dat challenges with experiments
- Flexible configuration
  - components moved in and out of production environment
- Co-funded by industry and CERN



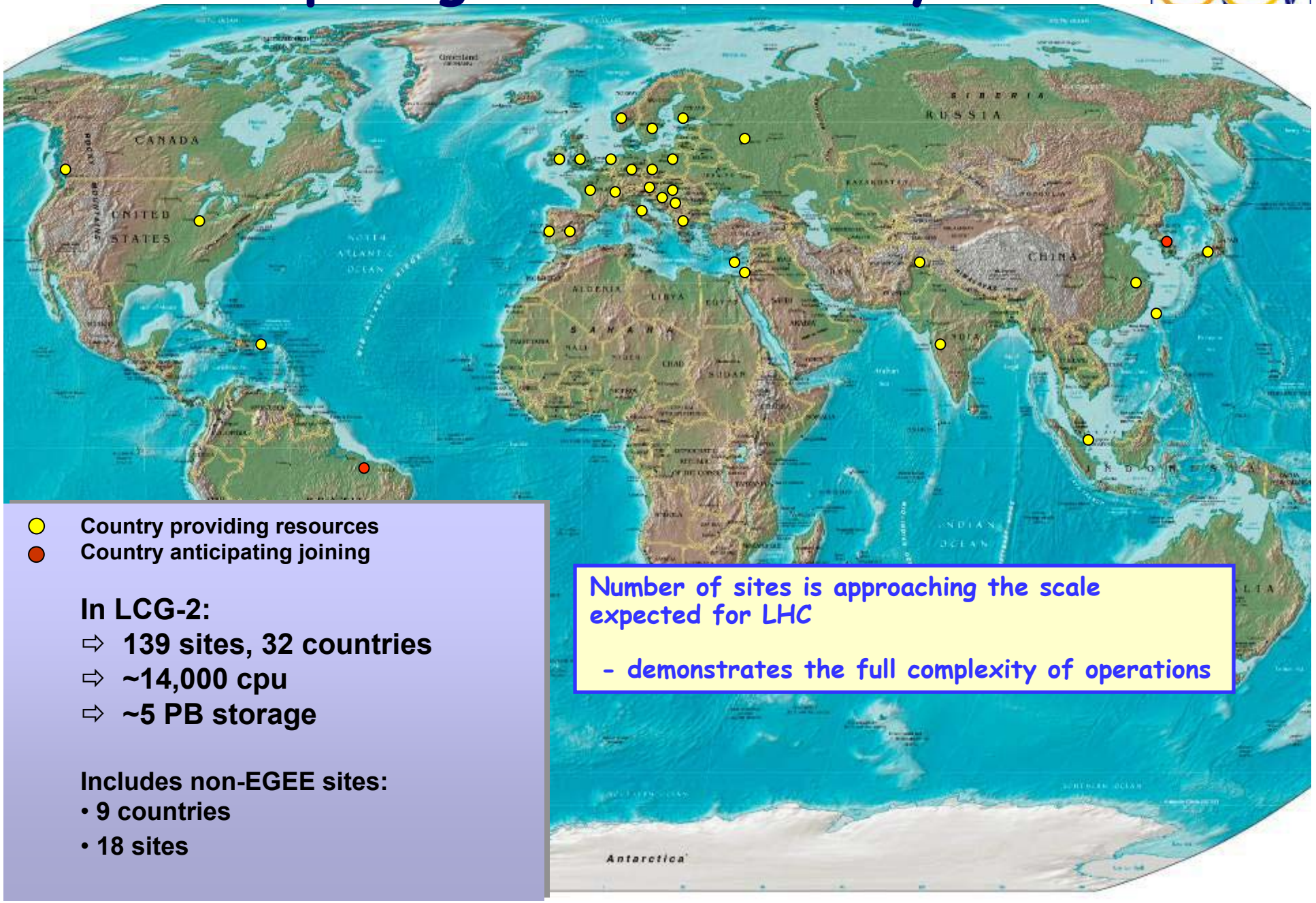
Alice Data Recording Challenge



- Target – one week sustained at 450 MB/sec
- Used the new version of Castor mass storage system
- Note smooth degradation and recovery after equipment failure

Deployment and Operations

Computing Resources: May 2005



- Country providing resources
- Country anticipating joining

In LCG-2:

- ⇨ 139 sites, 32 countries
- ⇨ ~14,000 cpu
- ⇨ ~5 PB storage

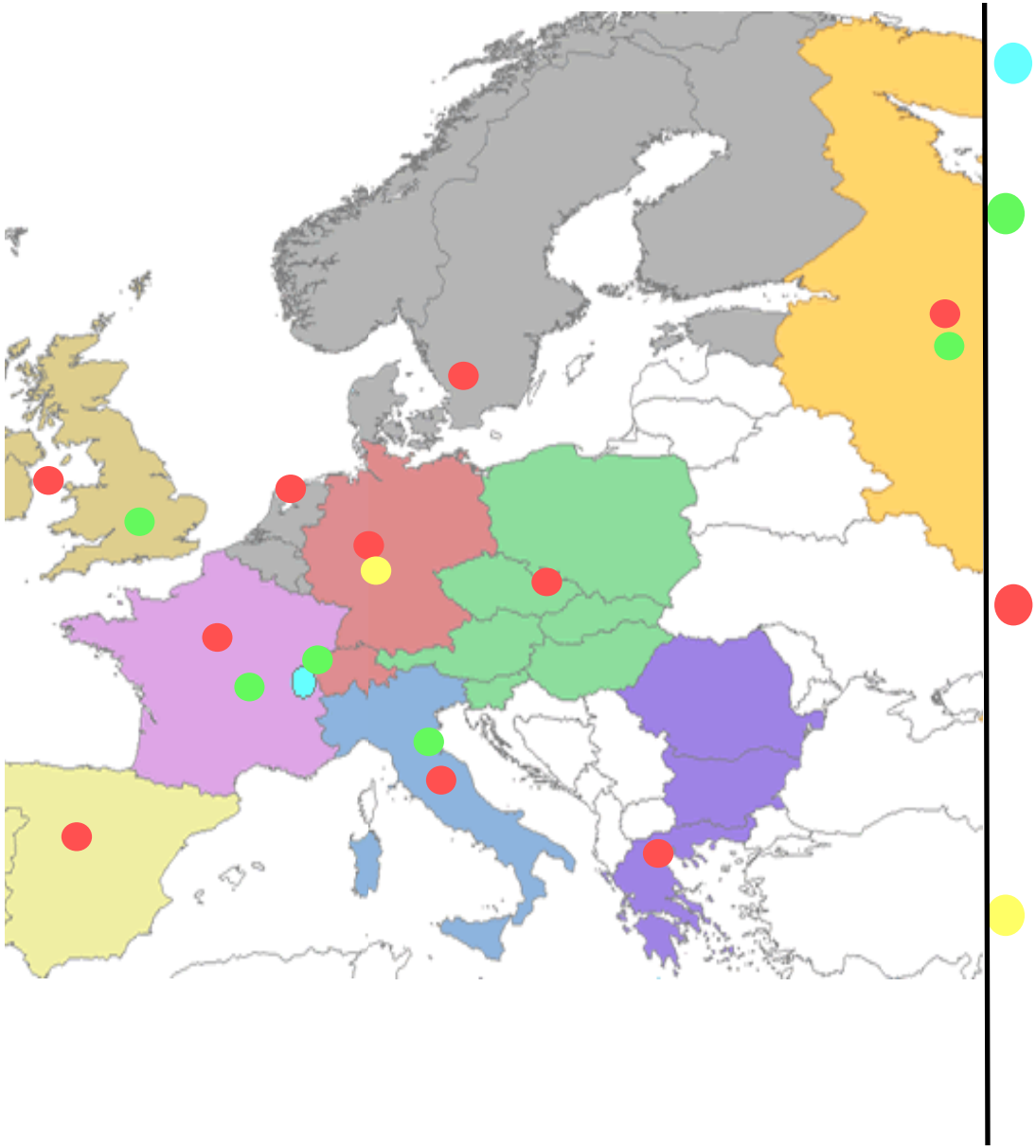
Includes non-EGEE sites:

- 9 countries
- 18 sites

Number of sites is approaching the scale expected for LHC

- demonstrates the full complexity of operations

Operations Structure



Operations Management Centre (OMC):

- At CERN – coordination etc

Core Infrastructure Centres (CIC)

- Manage daily grid operations – oversight, troubleshooting
- Run essential infrastructure services
- Provide 2nd level support to ROCs
- UK/I, Fr, It, CERN, + Russia (M12)
- Hope to get non-European centres

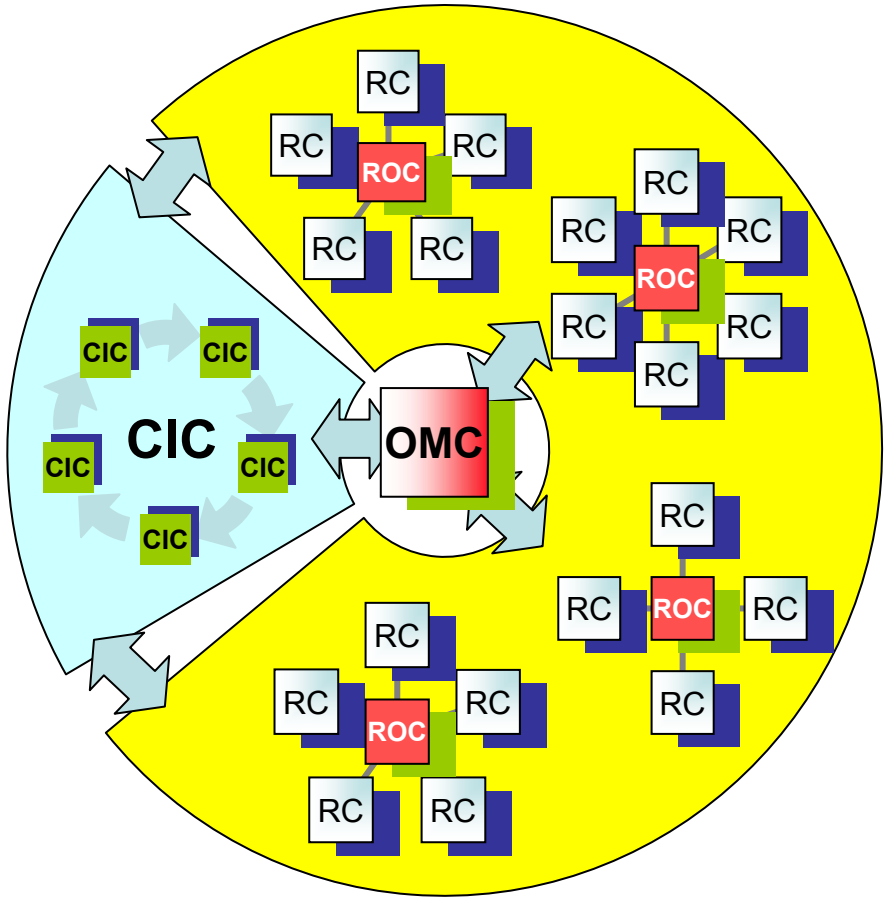
Regional Operations Centres (ROC)

- Act as front-line support for user and operations issues
- Provide local knowledge and adaptations
- One in each region – many distributed

User Support Centre (GGUS)

- In FZK – support portal – provide single point of contact (service desk)

Grid Operations



RC = Resource Centre

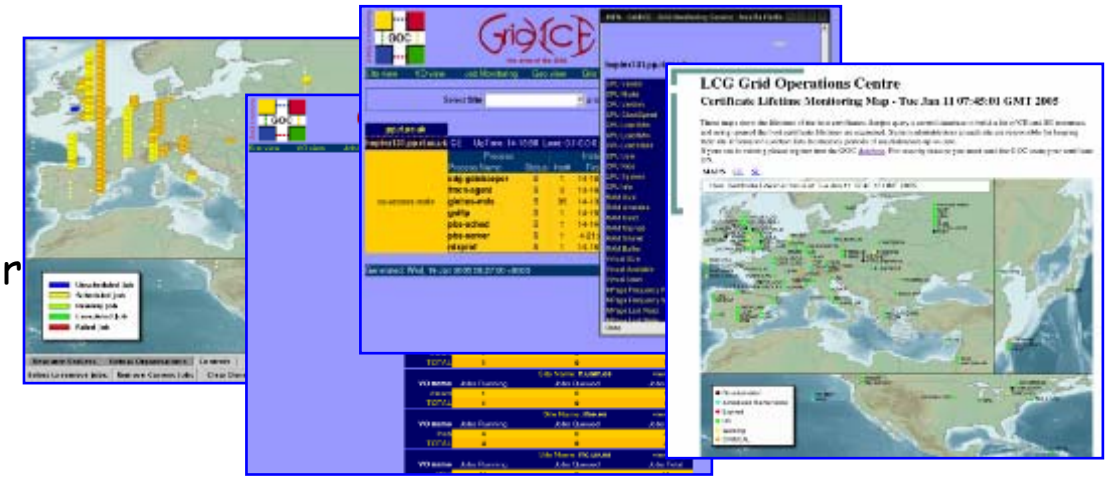
- The *grid* is flat, but
- *Hierarchy* of responsibility
  - Essential to scale the operation
- CICs act as a single Operations Centre
  - Operational oversight ***(grid operator)*** responsibility
  - rotates weekly between CICs
  - Report problems to ROC/RC
  - ROC is *responsible* for ensuring problem is resolved
  - ROC oversees regional RCs
- ROCs responsible for organising the operations in a region
  - Coordinate deployment of middleware, etc
- CERN coordinates sites not associated with a ROC

It is in setting up this operational infrastructure where we have really benefited from EGEE funding

Grid monitoring

- Operation of Production Service: real-time display of grid operations
- Accounting information
- Selection of Monitoring tools:

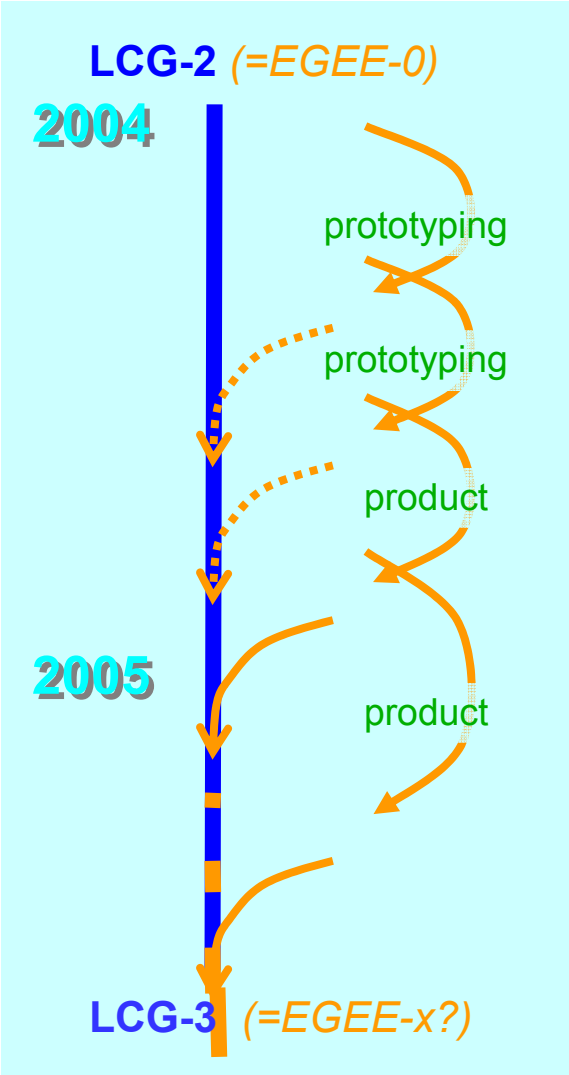
- GIIS Monitor + Monitor Graphs
- Sites Functional Tests
- GOC Data Base
- Scheduled Downtimes



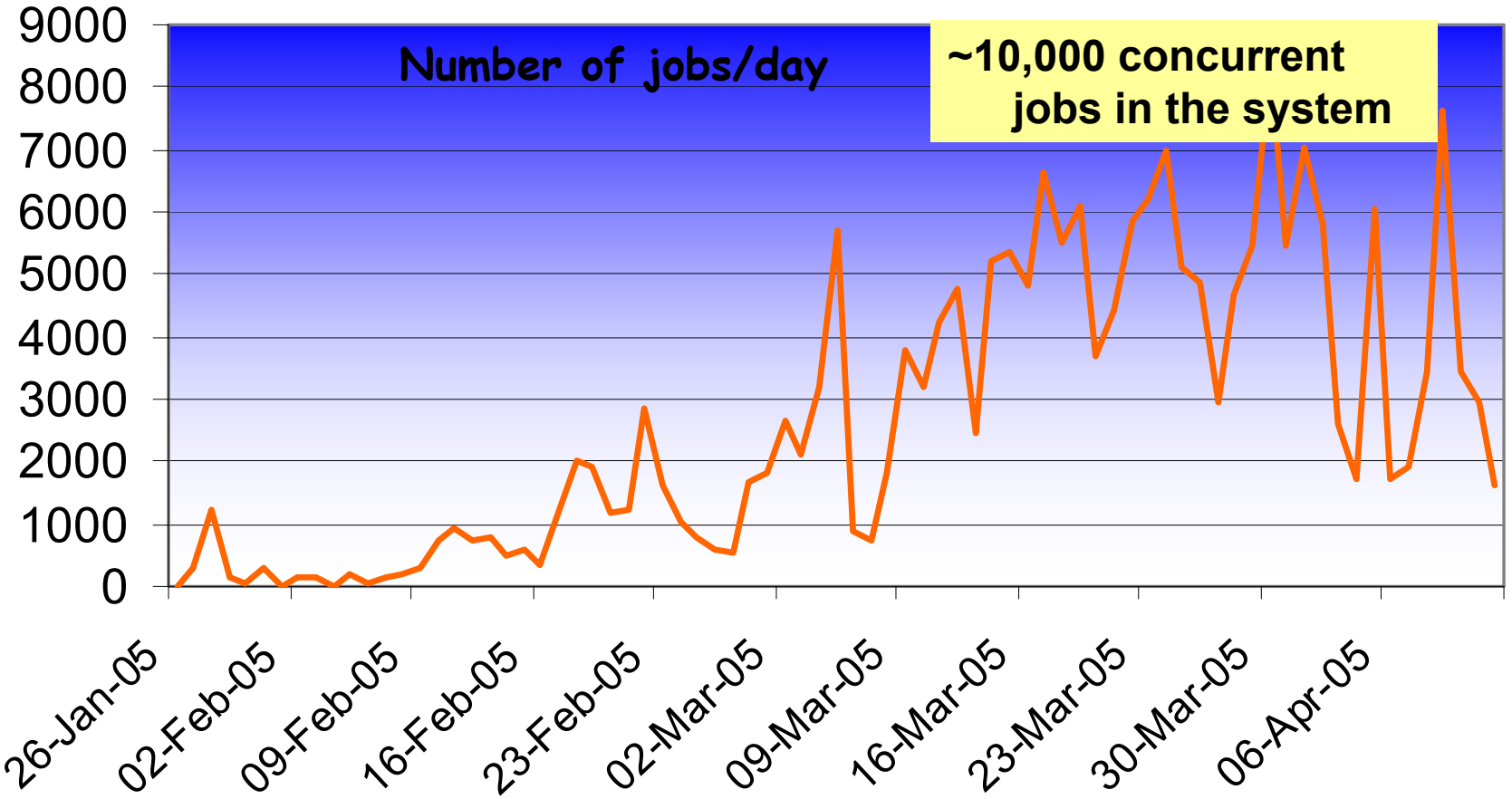
- Live Job Monitor
- GridIce – VO + fabric view
- Certificate Lifetime Monitor

Operations focus

- Main focus of activities now:
  - Improving the operational reliability and application efficiency:
    - Automating monitoring → alarms
    - Ensuring a 24x7 service
    - Removing sites that fail functional tests
    - Operations interoperability with OSG and others
  - Improving user support:
    - Demonstrate to users a reliable and trusted support infrastructure
  - Deployment of gLite components:
    - Testing, certification → pre-production service
    - Migration planning and deployment – while maintaining/growing interoperability
  - ☞ Further developments now have to be driven by experience in real use



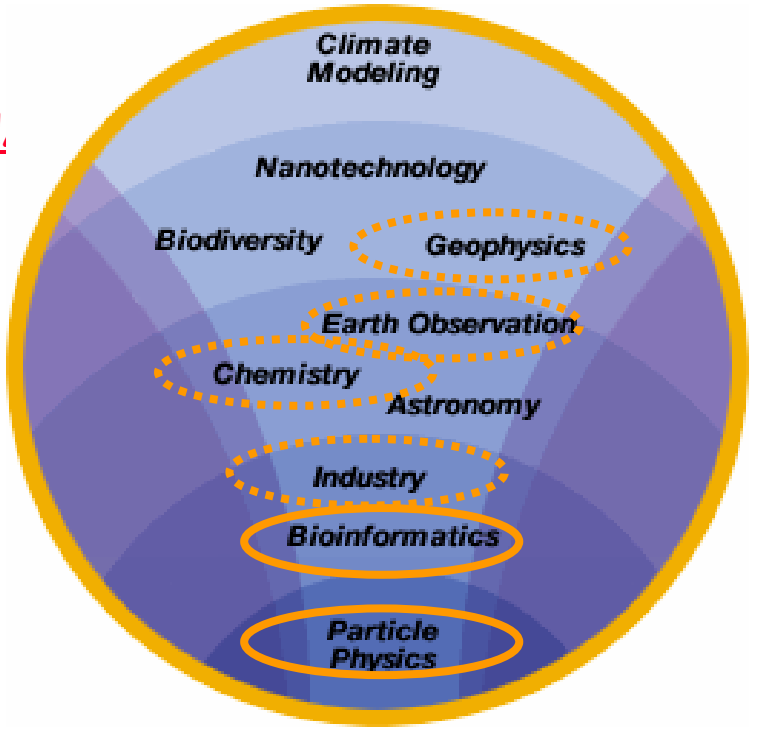
Recent ATLAS work



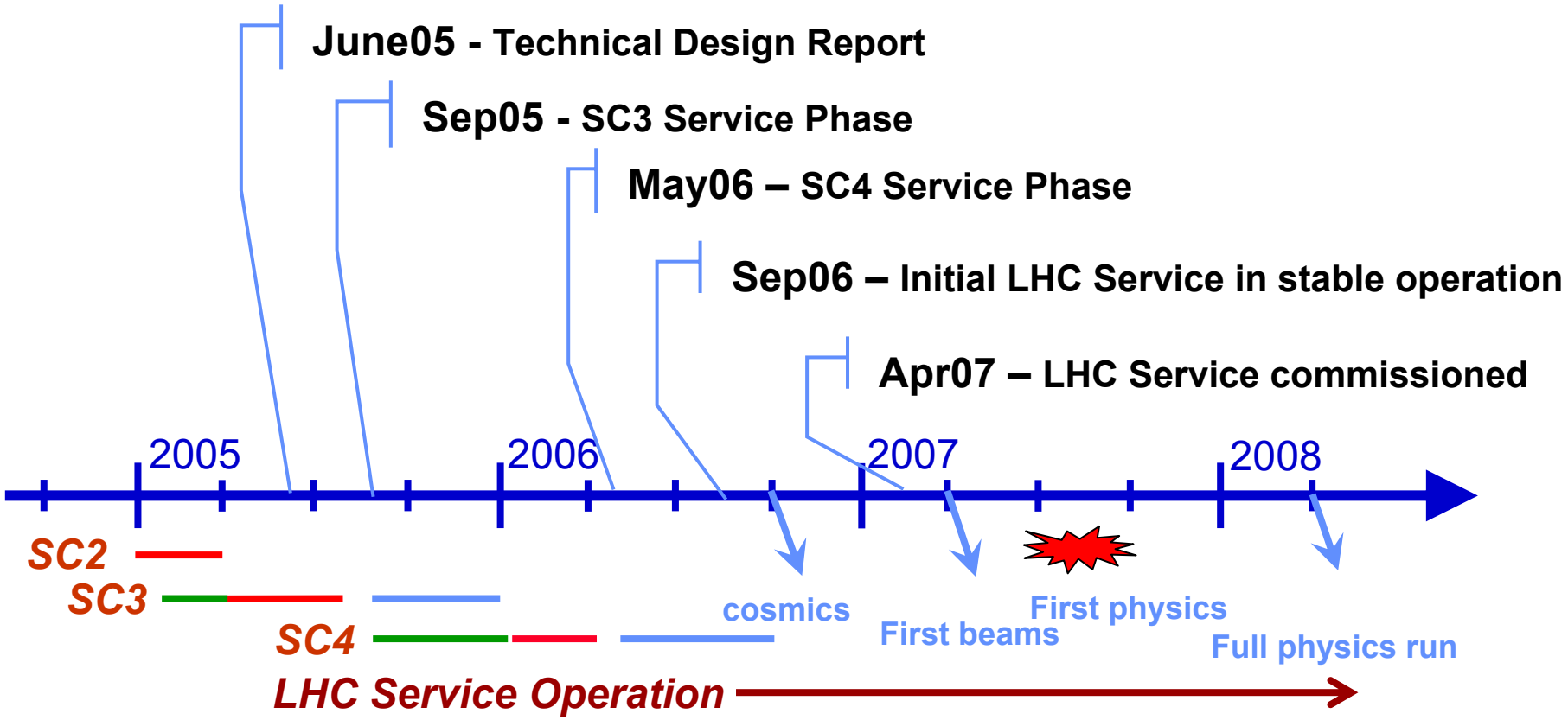
- ATLAS jobs in EGEE/LCG-2 in 2005
  - In latest period up to 8K jobs/day
- Several times the current capacity for ATLAS at CERN alone – shows the reality of the grid solution

Deployment of other applications

- Pilot applications
  - High Energy Physics
  - Biomed applications
    http://egee-na4.ct.infn.it/biomed/
- Generic applications – Deployment under way
  - Computational Chemistry
  - Earth science research
  - EGEODE: first industrial application
  - Astrophysics
- With interest from
  - Hydrology
  - Seismology
  - Grid search engines
  - Stock market simulators
  - Digital video etc.
  - Industry (provider, user, supplier)
- Many users
  - broad range of needs
  - different communities with different background and internal organization



Service Challenges – ramp up to LHC start-up service



SC2 – Reliable data transfer (disk-network-disk) – 5 Tier-1s, aggregate 500 MB/sec sustained at CERN
SC3 – Reliable base service – most Tier-1s, some Tier-2s – basic experiment software chain – grid data throughput 500 MB/sec, including mass storage (~25% of the nominal final throughput for the proton period)
SC4 – All Tier-1s, major Tier-2s – capable of supporting full experiment software chain inc. analysis – sustain nominal final grid data throughput
LHC Service in Operation – September 2006 – ramp up to full operational capacity by April 2007 – capable of handling twice the nominal data throughput

Why Service Challenges?

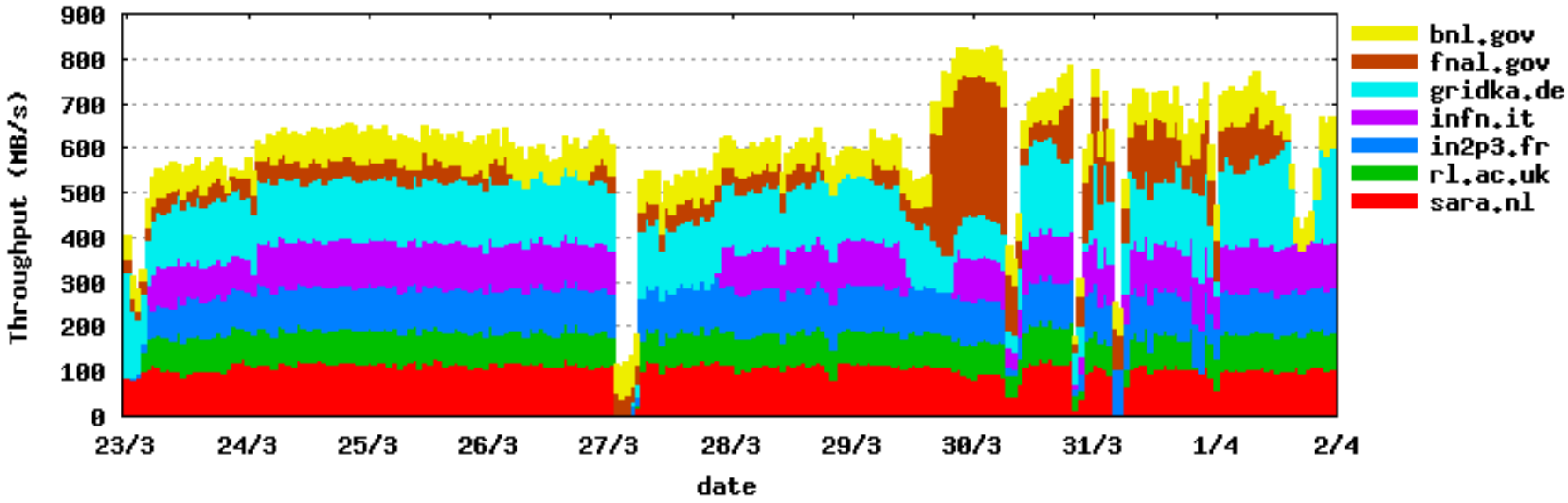
To test Tier-0 ←→ Tier-1 ←→Tier-2 services

- Network service
  - Sufficient bandwidth: ~10 Gbit/sec
  - Backup path
  - Quality of service: security, help desk, error reporting, bug fixing, ..
- Robust file transfer service
  - File servers
  - File Transfer Software (GridFTP)
  - Data Management software (SRM, dCache)
  - Archiving service: tapeservers,taperobots, tapes, tapedrives, ..
- Sustainability
  - Weeks in a row un-interrupted 24/7 operation
  - Manpower implications: ~7 fte/site
  - Quality of service: helpdesk, error reporting, bug fixing, ..

➢ **Towards a stable production environment for experiments**

SC2 met its throughput targets

- >600MB/s daily average for 10 days was achieved - Midday 23rd March to Midday 2nd April
  - Not without outages, but system showed it could recover rate again from outages
  - Load reasonable evenly divided over sites (give network bandwidth constraints of Tier-1 sites)



Ramping up to the LHC Service

- The services for Phase 2 will be ramped-up through two **Service Challenges** SC3 this year and SC4 next year
- These will include CERN, the Tier-1s and the major Tier-2s
- Each service Challenge includes –

  -- a set-up period

  - check out the infrastructure/service to iron out the problems *before the experiments get fully involved*
  - schedule allows *time to provide permanent fixes* for problems encountered
  - A throughput test

  -- followed by a long stable period for experiments to check
  out their computing model and software chain

Summary

- The LCG project – with EGEE partnership – has built an international computing grid
  - Used for real production work for LHC and other HEP experiments
  - Is also now being used by many other application domains – bio-medical, physics, chemistry, earth science
- The scale of this grid is approaching the level needed for LHC in terms of number of sites
  - We already see the full scale of complexity in operations
- There is a managed operation in place
  - But – there is still a lot to do to make this into a reliable service for LHC
- Service challenge program – ramp up to 2007 – is an extremely aggressive plan
- Many opportunities for collaboration in all aspects of the project