Early history of the LHC Computing Grid

from conception to Run 1

1998 – 2010

A distributed project to build a distributed data grid
A collaboration of the LHC community

Les Robertson, CERN retired

To put this talk in context …

Work on the LHC computing grid started 24 years ago – Remember 1998?

- Bill Clinton was impeached over his affair with Monica
- Google was launched in September
- The Nokia 6110 was the best mobile phone available – it would be another 9 years before the iPhone appeared
- A good home network connection was 64 kbits/sec
  A high-speed inter-site network was 622 Mbits/sec (if available)
- A big hard disk held 12GB in a hefty 5.25” package
  An IBM tape cartridge held 20GB
  – less than an Apple Watch today

- Amazon had been selling books online for three years but it would be a further 8 years before Amazon Web Services announced the S2 (storage) and EC2 (Elastic Computing) services – the first cloud

HEP Computing –

- PC clusters with Linux was the “standard”
  With remote job submission via the internet
- Major sites had mass storage management systems using tape robots
- There was very good TCP/IP network expertise for efficient and reliable data access

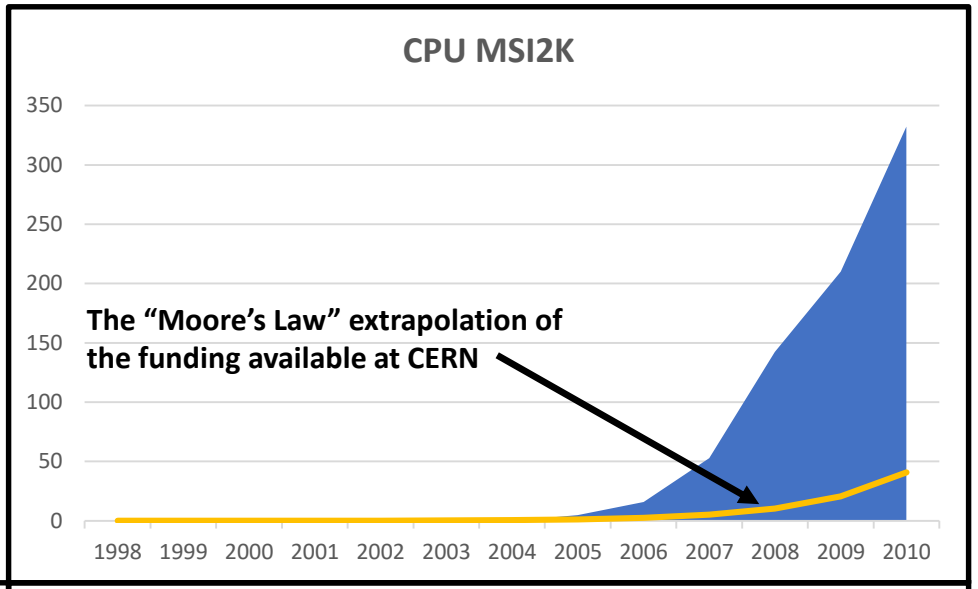
LHC –

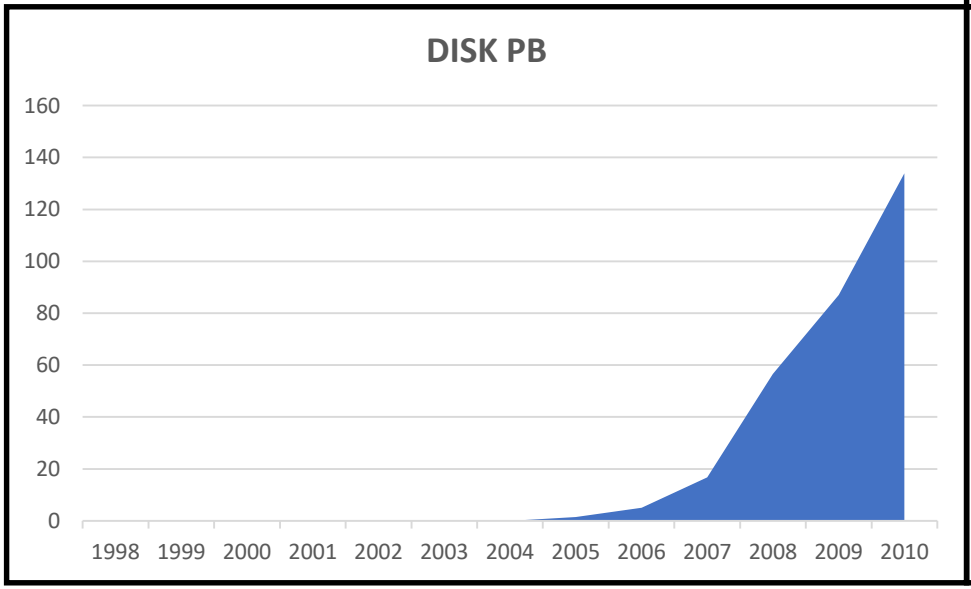
- Construction had been approved in 1995 with a target date for first beams of 2005
- The four experiment collaborations had already prepared initial estimates of the data rates, storage requirements and computing capacity that would be needed

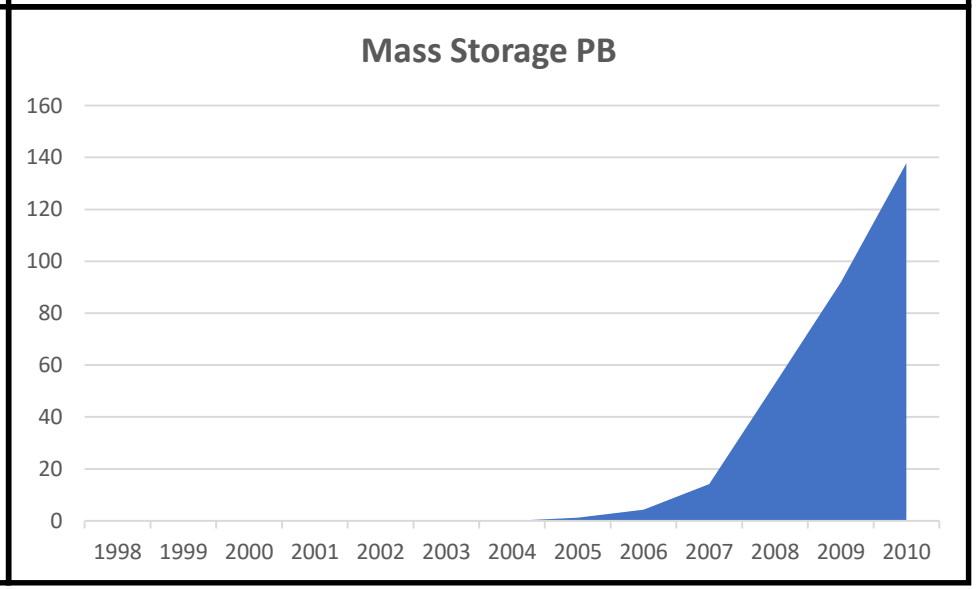
The Problem

Estimated** requirements for LHC offline computing

- Compared with LEP - LHC detectors
  -- would produce much more data
  -- with more complex events
- The LHC project budget had no line for computing at CERN
- Most of the computing capacity required would have to come from outside of CERN







*** adapted from LCG Technical Design Report 2005 with start-up date of 2007*

Search for a Solution

- The MONARC** Project was started in 1997
- to assess requirements, technologies and costs for LHC computing.
- and propose a solution within the constraints of
  - the large data volumes
  - the need to incorporate resources of different capacity and functionality spread across the world
  - the bandwidth of wide area networking that could be expected in 2005
  - using an architecture that looked feasible to be implemented in the time available

** *CERN/LCB 2000-001 Models of Networked Analysis at Regional Centres for LHC Experiments – H.Newman (chair) et al*

Models of Networked Analysis at Regional Centres for LHC Experiments

(MONARC)

PHASE 2 REPORT

The Importance of Community Buy-in

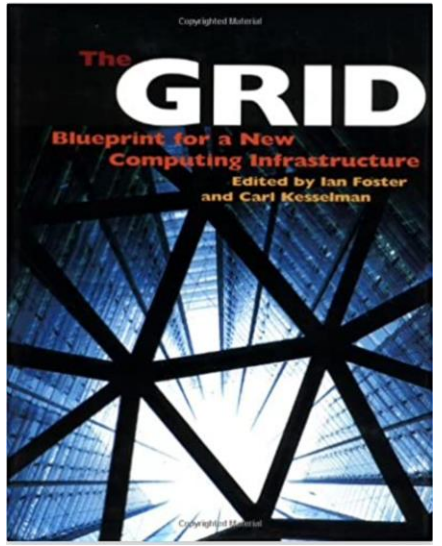
24th March 2000

MONARC Members

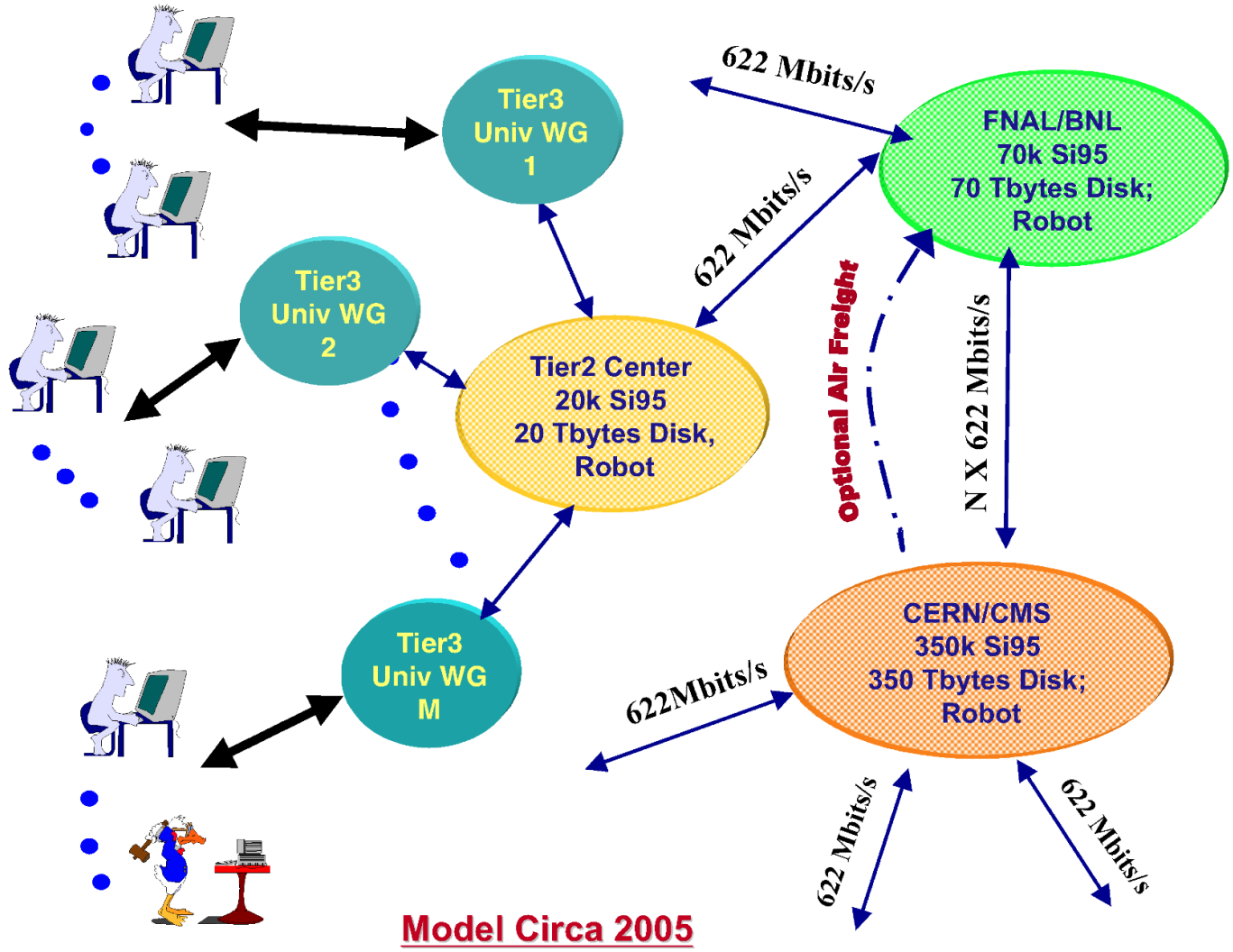
M. Aderholz (MPI), K. Amako (KEK), E. Auge (L.A.L/Orsay), G. Bagliesi (Pisa/INFN), L. Barone (Roma1/INFN), G. Battistoni (Milano/INFN), M. Bernardi (CINECA), M. Boschini (CILEA), A. Brunengo (Genova/INFN) J.J. Bunn (Caltech/CERN), J. Butler (FNAL), M. Campanella (Milano/INFN), P. Capiluppi (Bologna/INFN), F, Carminati (CERN), M. D'Amato (Bari/INFN), M. Dameri (Genova/INFN), A. di Mattia (Roma1/INFN), A. Dorokhov (CERN), G. Erbacci (CINECA), U. Gasparini (Padova/INFN), F. Gagliardi (CERN), I. Gaines (FNAL), P. Galvez (Caltech), A. Ghiselli (CNAF/INFN), J. Gordon (RAL), C. Grandi (Bologna/INFN), F. Harris (Oxford), K. Holtman (CERN), V. Karimäki (Helsinki), Y. Karita (KEK), J. Klem (Helsinki), I. Legrand (Caltech/CERN), M. Leltchouk (Columbia), D. Linglin (IN2P3/Lyon Computing Centre), P. Lubrano (Perugia/INFN), L. Luminari (Roma1/INFN), A. Maslennicov (CASPUR), A. Mattasoglio (CILEA), M. Michelotto (Padova/INFN), I. McArthur (Oxford), Y. Morita (KEK), A. Nazarenko (Tufts), H. Newman (Caltech), V. O'Dell (FNAL), S.W. O'Neale (Birmingham/CERN), B. Osculati (Genova/INFN), M. Pepe (Perugia/INFN), L. Perini (Milano/INFN), J. Pinfold (Alberta), R. Pordes (FNAL), F. Prelz (Milano/INFN), A. Putzer (Heidelberg), S. Resconi (Milano/INFN and CILEA), L. Robertson (CERN), S. Rolli (Tufts), T. Sasaki (KEK), H. Sato (KEK), L. Servoli (Perugia/INFN), R.D. Schaffer (Orsay), T. Schalk (BaBar), M. Sgaravatto (Padova/INFN), J. Shiers (CERN), L. Silvestris (Bari/INFN), G.P. Siroli (Bologna/INFN), K. Sliwa (Tufts), T. Smith (CERN), R. Somigliana (Tufts), C. Stanescu (Roma3), H. Stockinger (CERN), D. Ugolotti (Bologna/INFN), E. Valente (INFN), C. Vistoli (CNAF/INFN), I. Willers (CERN), R. Wilkinson (Caltech), D.O. Williams (CERN).

In the wider world the Grid concept was attracting interest

- *The Grid : Blueprint For A New Computing Infrastructure, Morgan Kaufmann, July 1998 –* Carl Kesselman (USC) and Ian Foster (Chicago)
- The Globus Project supported an open-source toolkit that included the basic functions needed to operate a Grid.
- Early HEP pilots – the Particle Physics Data Grid in the US
  - a distributed simulation facility in INFN
  - …
- Gave some confidence that distributed computing services for HEP would be feasible on the timescale of LHC

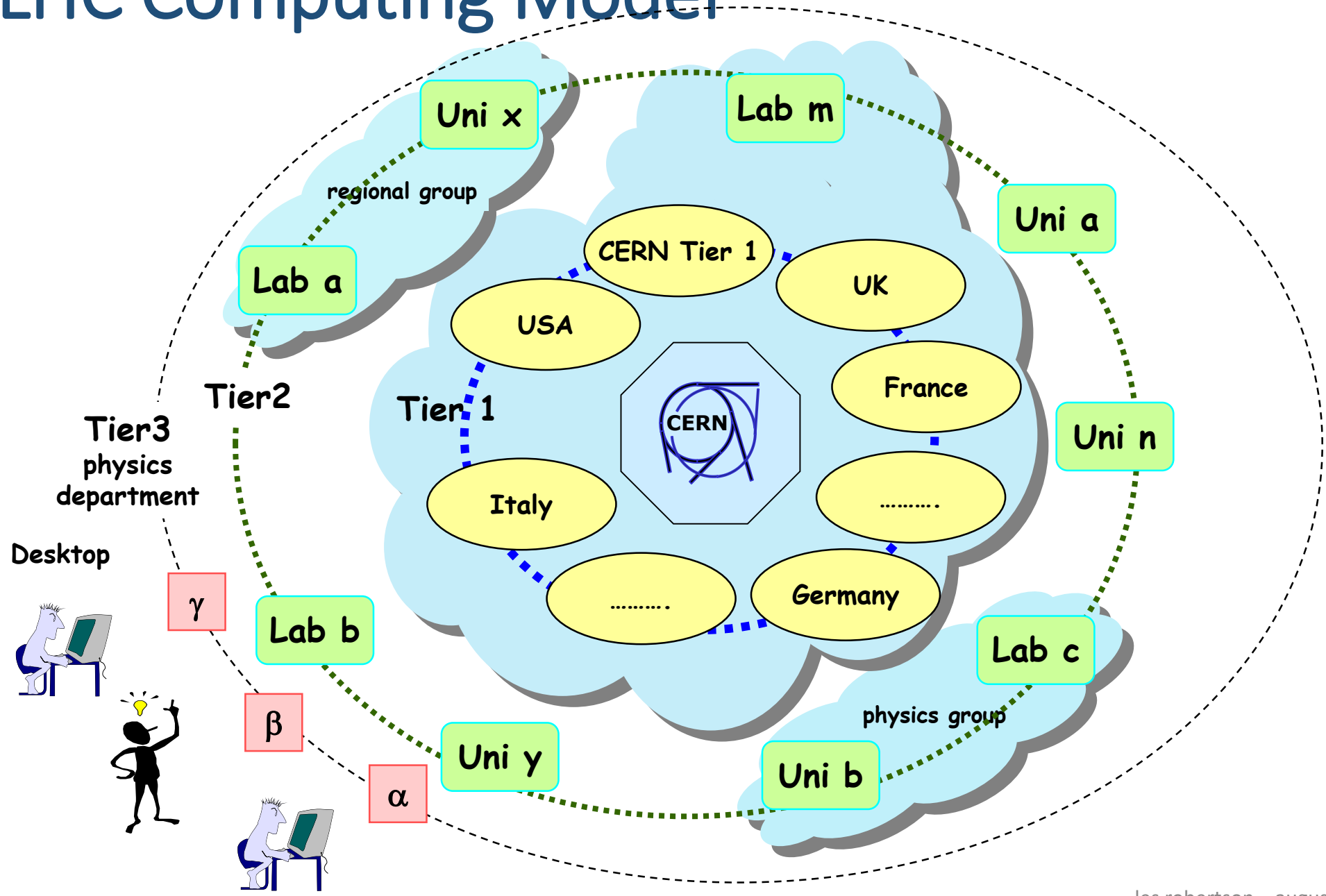


The Monarc distributed model



Three-tier distributed computing model for the early years of LHC operation – from *CERN/LCB 2000-001 Models of Networked Analysis at Regional Centres for LHC Experiments* (note the WAN data rates and the "optional air freight")

LHC Computing Model



les.robertson@cern.ch

At the CHEP meeting in Padova in 2000 the concept of LHC computing using regional centres managed as a *grid* was agreed

Not formally – it just seemed like the right way to go

Monarc really had done a very good job

- We had a lot of experience with the components of the nodes of the grid
  - large and small clusters, mass storage management, wide area networking, …
- We had starter sets of components from grid projects to get us off the ground
- But we needed to develop this into a grid that could handle the scale needed for LHC
  - volume of data
  - inter-site file transfer performance
  - large number of nodes, of all sizes
  - thousands of users
  - unpredictable usage patterns for analysis

The LHC Computing Grid Project

This was all going to be very expensive in money and manpower, so we needed a formal structure acceptable to the CERN Council and the agencies funding the experiments

- The project was set up in 2001 to
  - **coordinate** planning and provision of the computing resources for the LHC experiments
  - enable institutes and funding agencies to pledge formally resources and manpower
  - provide a collaboration style structure that ensured the wide representation of stake holders needed to maintain consensus
- The project reported to the LHCC – the high-level coordinating body for experiments using the LHC.

Developing the *middleware*

- The idea of the Grid had stimulated many R&D projects on both sides of the Atlantic
  - with HEP participation
  - concerned with developing middleware
  - and operating testbeds.
- This was excellent – lots of different ideas and highly motivated developers
- The difficulty was that in order to use the fruits of these developments we would need –
  - some form of homogeneity - otherwise the applications would have to handle a host of different interfaces.
  - scalability and performance to match LHC needs
  - Long-term support for the components we would use

and we had to be in production in time for LHC

- LHC planning was evolving and the startup date moved to 2007, so we had a breathing space for development before interfaces would have to be fixed
- The EU DataGrid project included participants from CERN** and many HEP institutes, and for the follow-on project, EGEE, it was agreed that institutes taking part in the LHC project could be associated no matter their home country – eventually involving 34 countries.
- In the US LHC participants migrated towards the Open Science Grid (OSG)
- As the LHC beams grew closer it was agreed that each participating centre would support a base set of functions

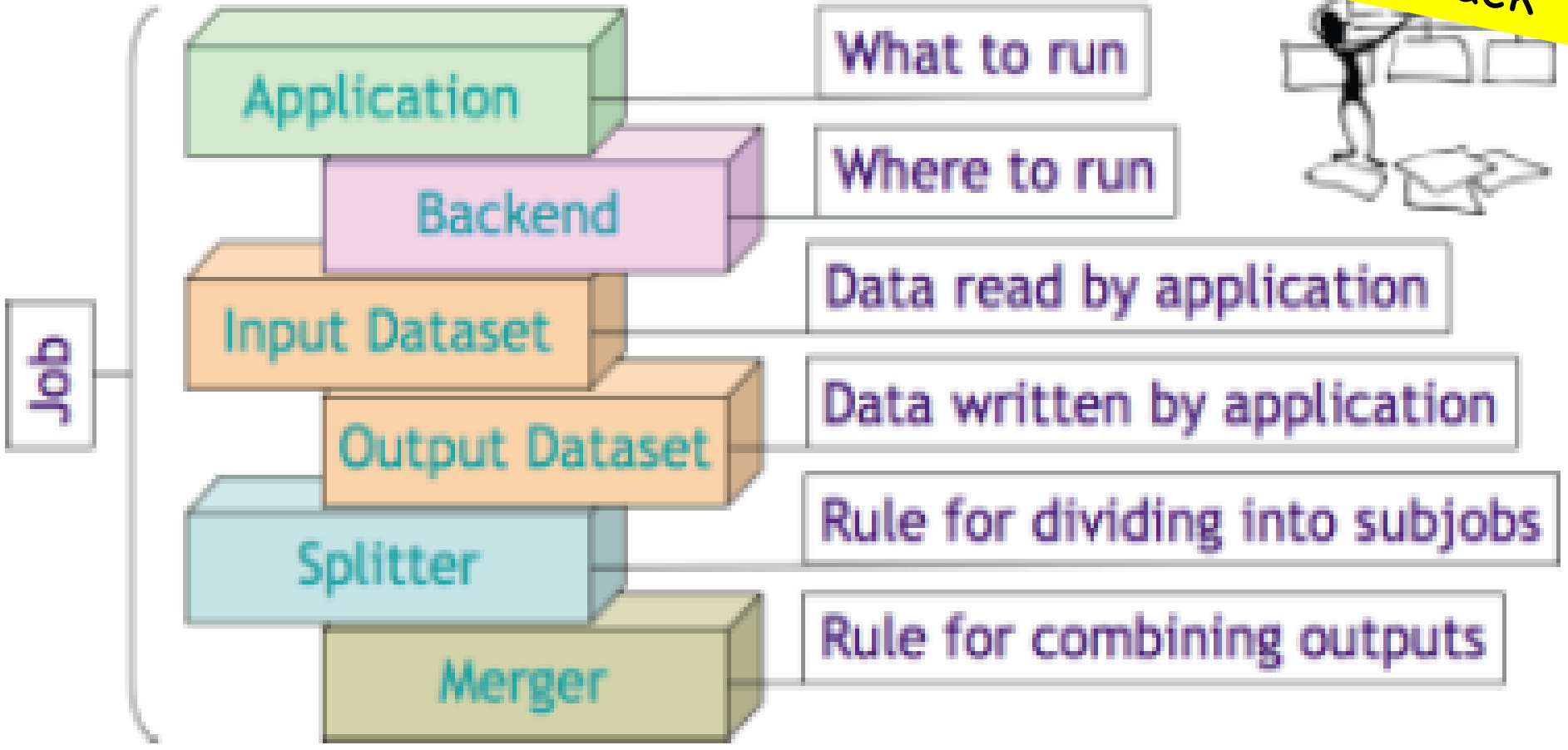
**The leader of EDG and EGEE was Fabrizio Gagliardi from CERN

The Middleware for the Baseline Services needed for the LHC Experiments

- Information system
- Security framework
- Storage Element
- SRM interface to Mass Storage
  dCache, DPM, CASTOR, STORM
- Basic data transfer tools –
  Gridftp, srmCopy.
- Reliable file transfer service –
  FTS
- Catalogue services –
  LFC, Globus RLS
- Catalogue and data management tools – lcg-utils
- Compute element –
  Globus/Condor-G based CE, Cream (web services)
- Reliable messaging service
- Virtual Organisation Management Services
- Database distribution services –
  ORACLE streams, SQUID
- POSIX-I/O interfaces to storage
- Workload Management –
  EGEE Resource Broker, VO-specific schedulers
- Job monitoring tools
- Grid monitoring tools
- Application software installation
- GUIs for analysis, production
  GANGA, CRAB, PANDA, ..

For LCG, grid interoperability was required at the level of the baseline service
→ same software or standard interfaces or compatible functionality

The ATLAS/LHCb grid interface stack

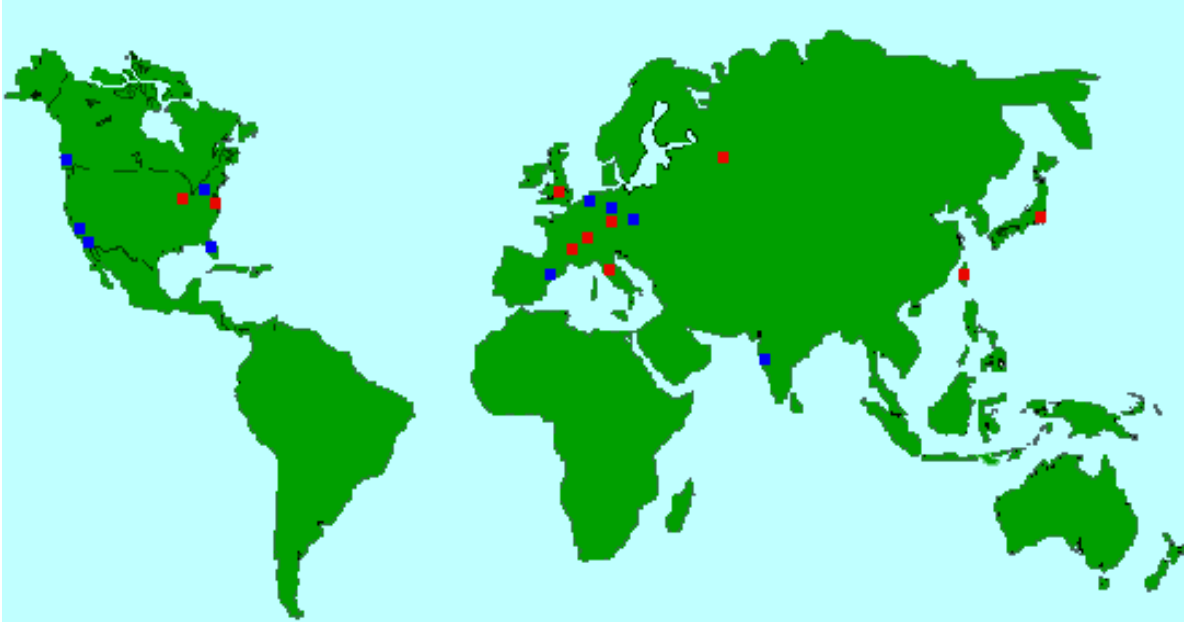


LCG-1 Service Opened September 2003

- Middleware package – components from –
  - European DataGrid (EDG)
  - US (Globus, Condor, PPDG, GriPhyN) → the Virtual Data Toolkit
- Agreement reached on principles for registration and security
- Certification and distribution process established and tested - June
- Rutherford Lab (UK) to provide the initial Grid Operations Centre
- FZK (Karlsruhe) to operate the Call Centre
- Pre-release middleware deployed to the initial 13 centres – July

- 1 September release :

Taiwan, Brookhaven, CERN, Bologna, Fermilab, Karlsruhe, Lyon, Budapest, Moscow, Prague, Barcelona, Rutherford UK, Tokyo



From testing to data:

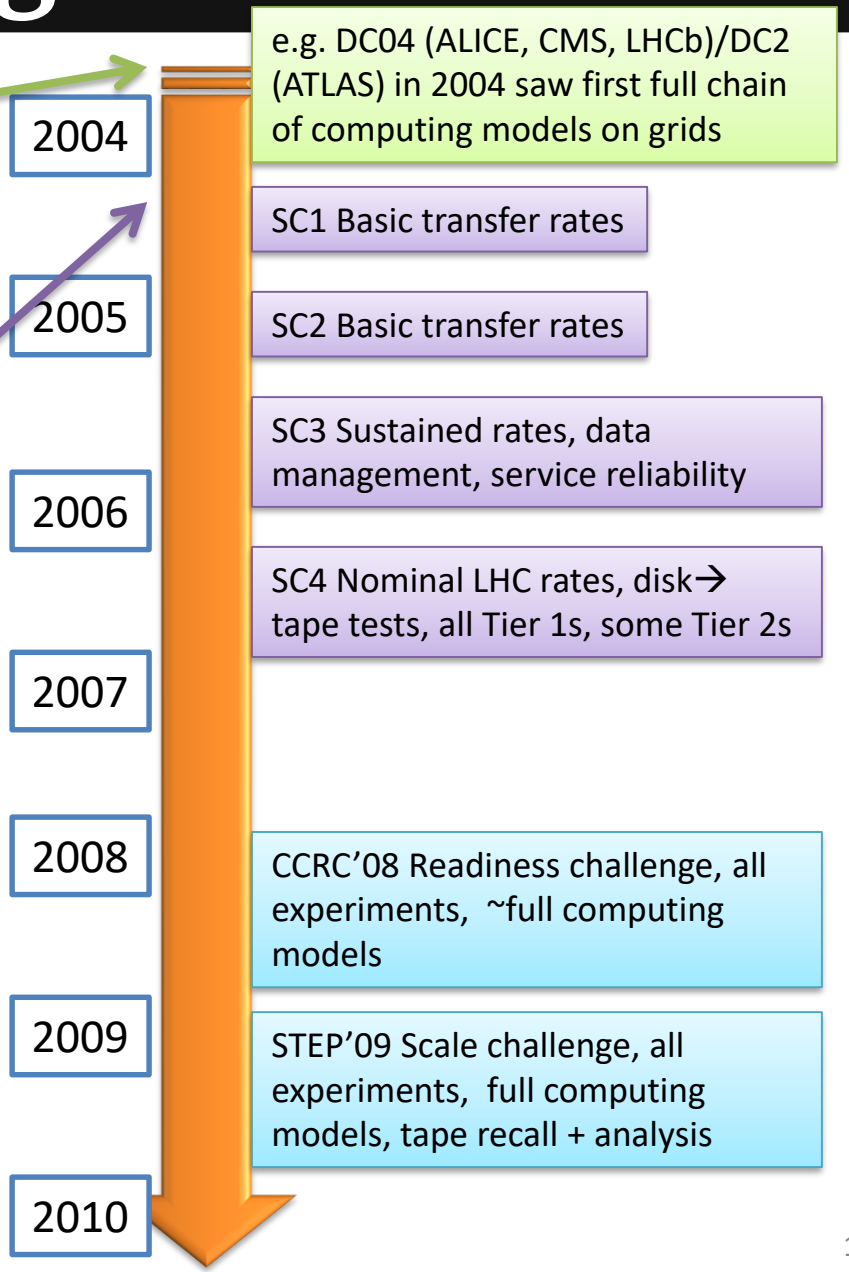
Independent Experiment Data Challenges

Service Challenges proposed in 2004

To demonstrate service aspects:

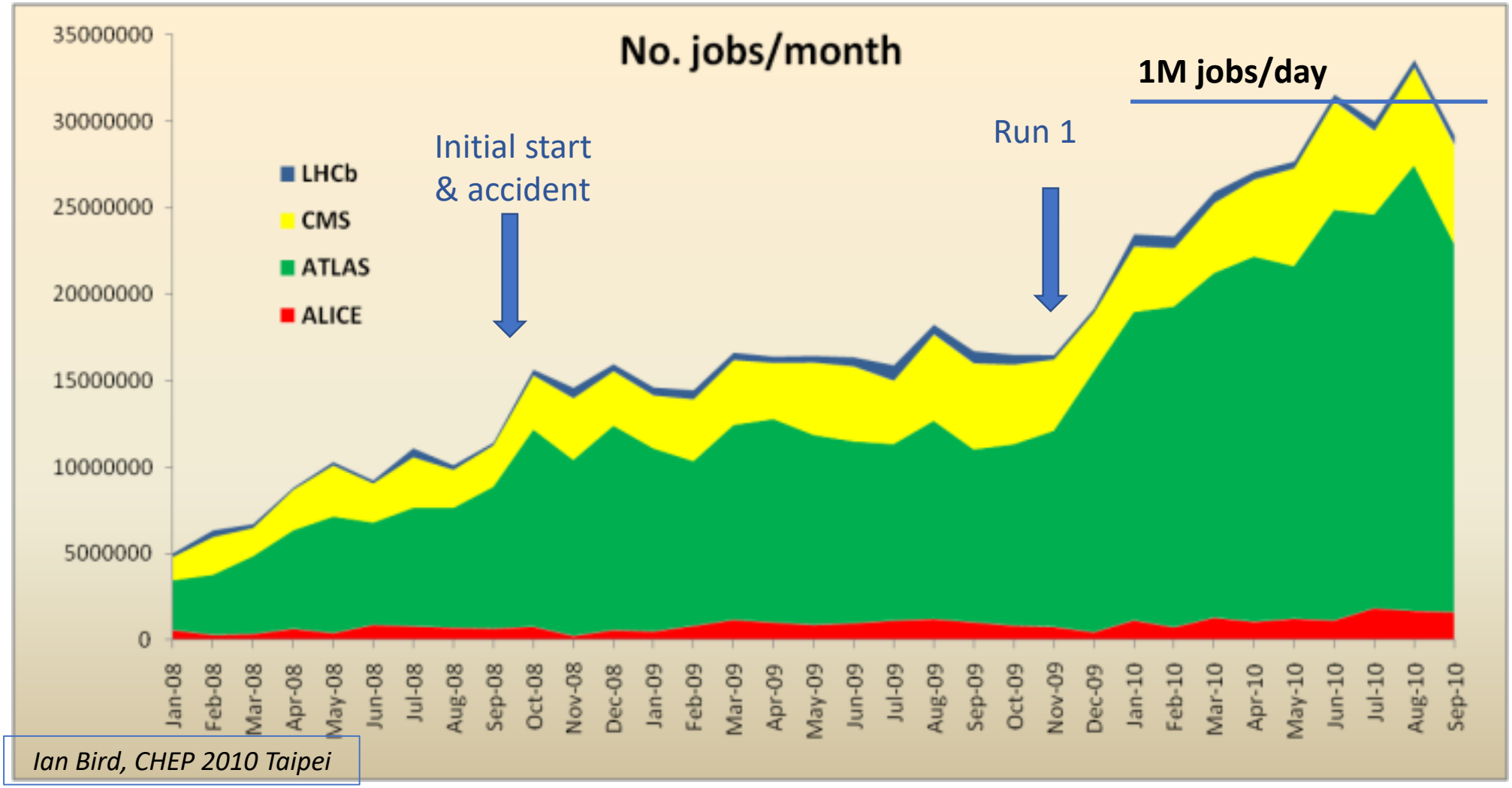
- -Data transfers for weeks on end
- -Data management
- -Scaling of job workloads
- -Security incidents ("fire drills")
- -Interoperability
- -Support processes

- Focus on real and continuous production use of the service over several years (simulations since 2003, cosmic ray data, etc.)
- Data and Service challenges to exercise all aspects of the service – not just for data transfers, but workloads, support structures etc.



How did it work for first data taking?

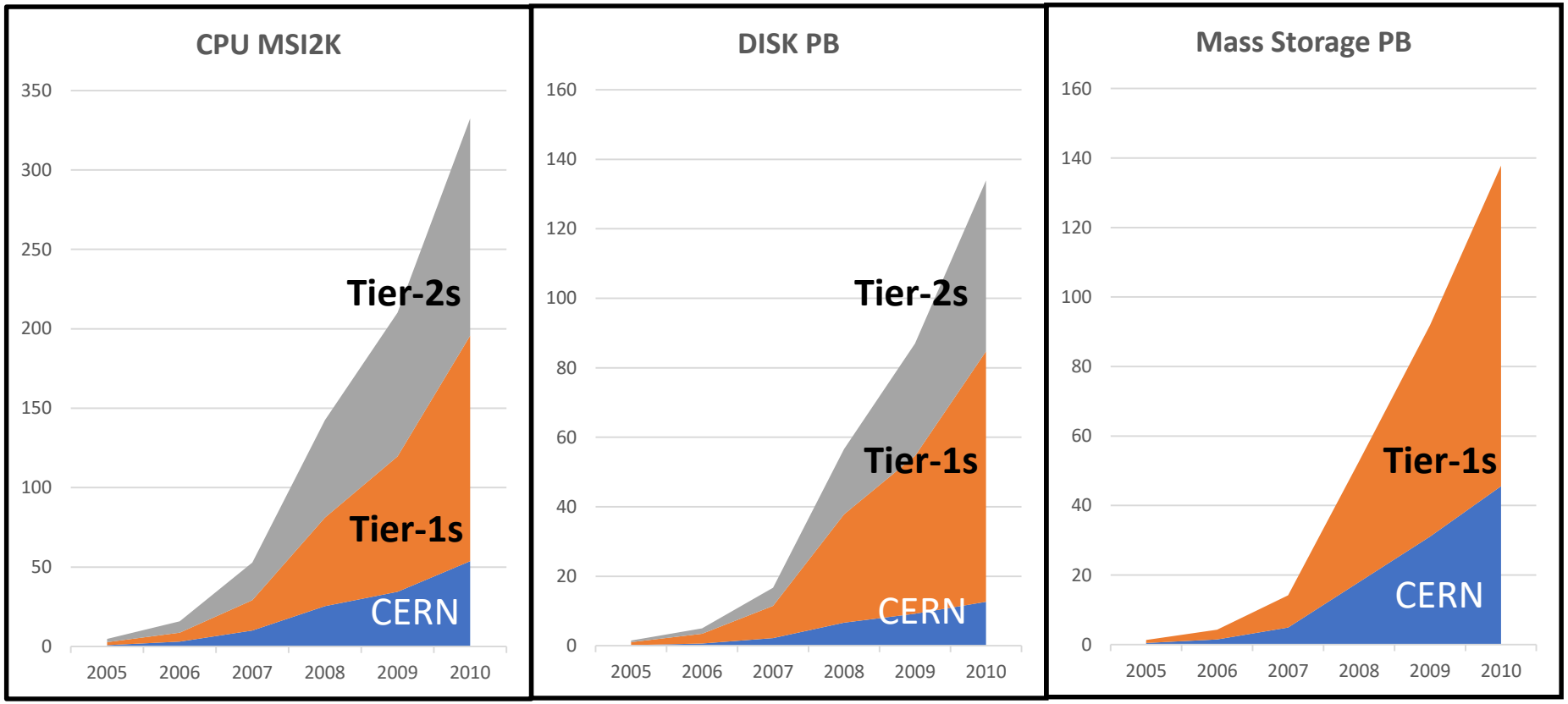
>140 sites, ~250K CPU cores, ~100 PB disk



Ian Bird, CHEP 2010 Taipei

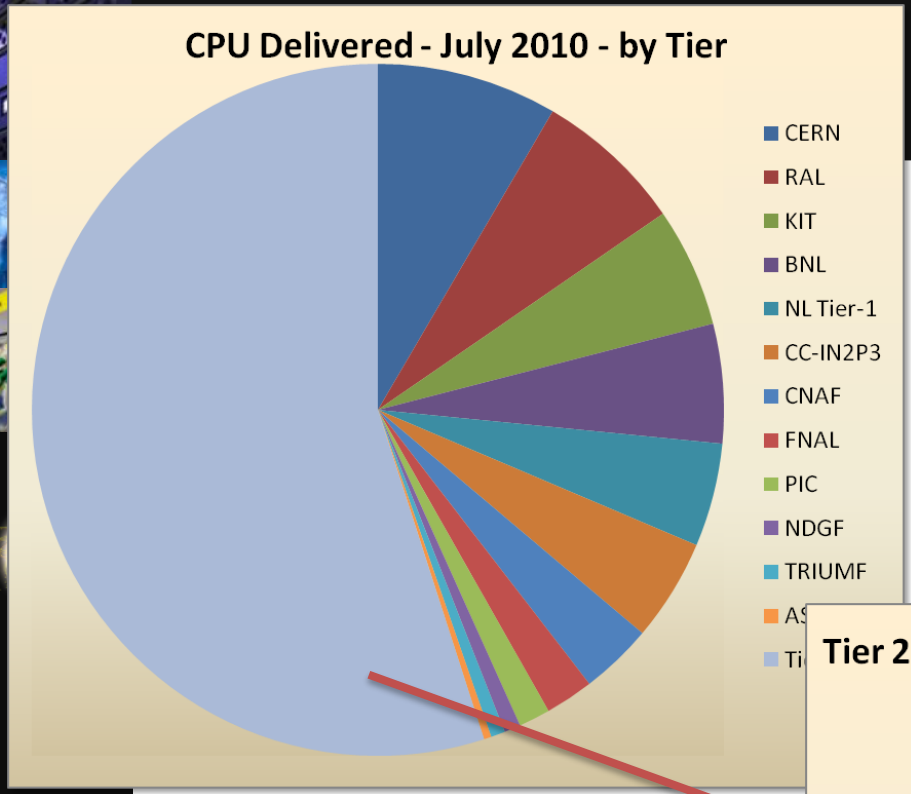
Planned distribution of Resources across the Grid

from LCG Technical Design Report – LHCC2005-0024 - start-up in 2007 -

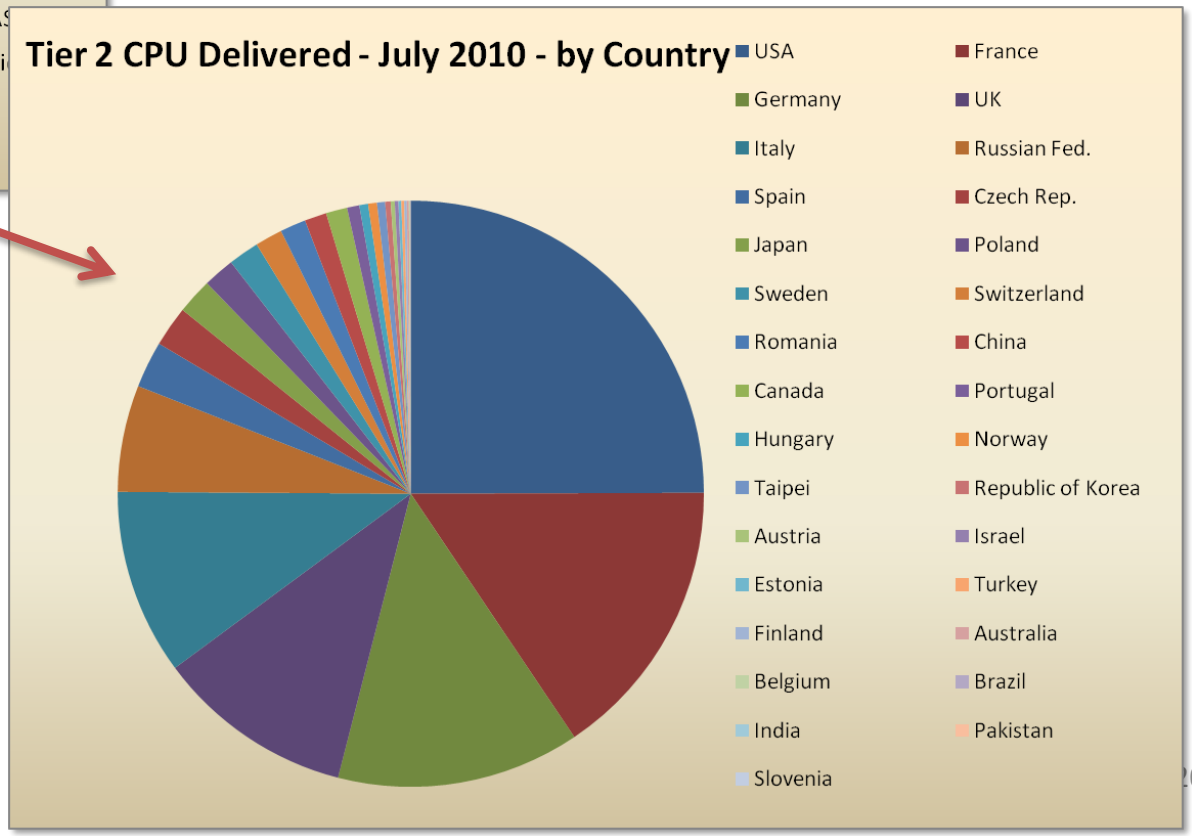


CERN share : 16% *9%* *33%*

CPU – July 2010

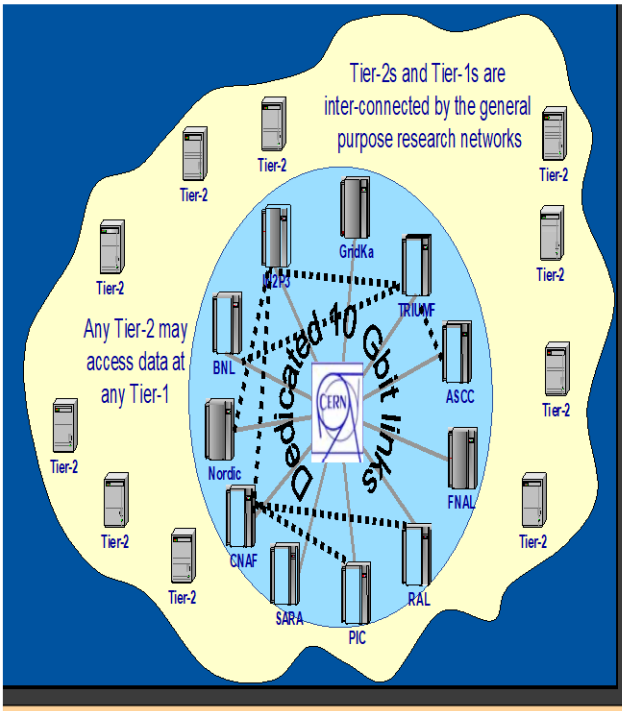


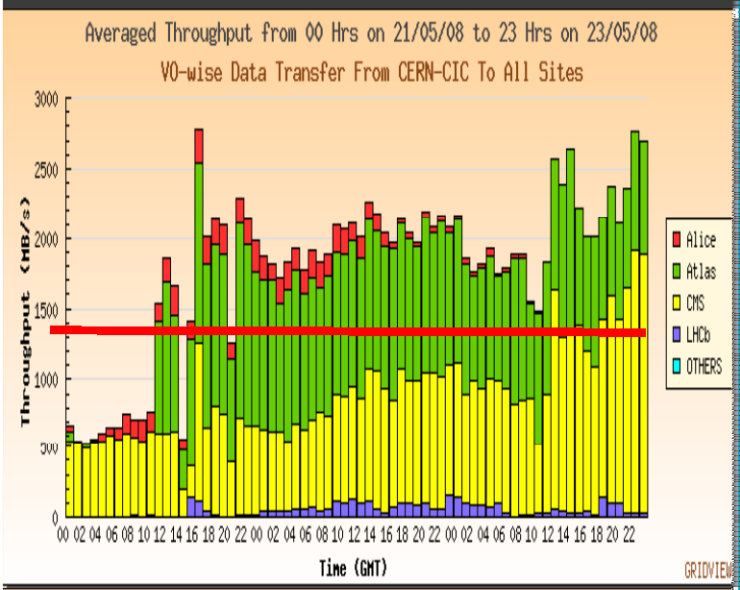
- Significant use of Tier 2s for analysis
  - frequently-expressed concern that too much analysis would be done at CERN is not reflected

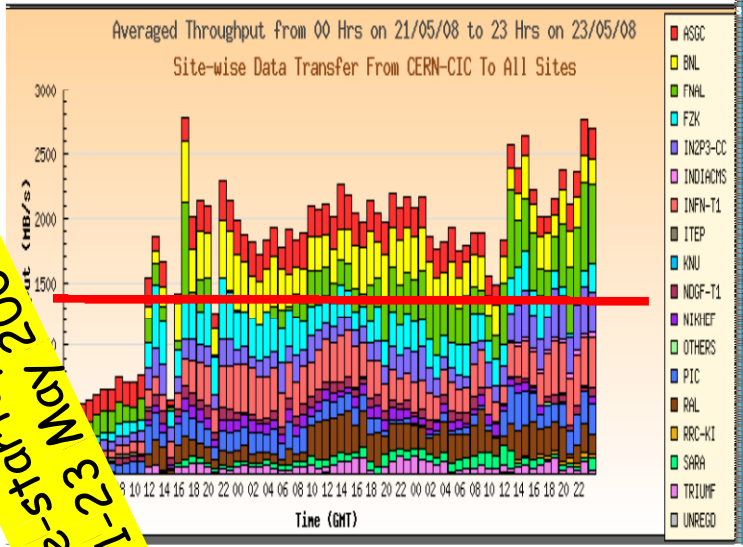


- Tier 0 capacity underused in general
  - But this is expected to change as luminosity increases

Data transfer out of Tier 0







- Full experiment rate needed is 650 MB/s
- Desire capability to sustain twice that to allow for Tier 1 sites to shutdown and recover
- Have demonstrated far in excess of that
- All experiments exceeded required rates for extended periods, & simultaneously
- All Tier 1s achieved (or exceeded) their target acceptance rates

Pre-startup test 21-23 May 2008

Final comments

- **The Worldwide LHC Computing Grid** was set on the right track by the Monarc project
  - as a community-wide endeavour
  - aimed at exploiting a novel type of distributed service
- Close collaboration with and funding from EU, NSF and many national grid middleware projects
- The significant and continuing participation of the regional centres → large and small centres all have a stake in the service
- The service challenges and data challenges were essential to keeping the project on track
- The accelerator deadline was always there to keep our feet on the ground

The Worldwide LHC Computing Grid

A collaboration of the LHC community

Very many people made important contributions unfortunately too many to mention by name