Grids: “Eventually, users will be unaware they are using any computer but the one on their desk, because it will have the capabilities to reach out across the internet and obtain whatever computational resources are necessary”

(Larry Smarr and Charles Catlett, 1992)
~600 M€ yearly budget; ~2300 staff; >6000 scientists from 60 countries
we are here

Big Bang

Superstrings?

grand unification

inflation

forces separate

nucleosynthesis

atoms

first stars

today

Energy

10^{17} TeV 10^{19} TeV 1 TeV 0.3 MeV 1 eV 4 MeV 0.7 MeV

time

10^{-33} s 10^{-35} s 10^{-10} s 300 s 300 000 years 10^{9} years 15\cdot10^{9} years
Evidence for the Higgs at LEP at $M \sim 115$ GeV

The LEP Program has now ended
By-product of the LEP era:
WWW and “infocern”, the 1st web address ~1990
html (xml) open standards

A great achievement and a fantastic idea, at the right time, making the internet available to everybody

It proves something about the benefits of assembling together urgent needs, infrastructure and smart people, and letting them interact..

And why we should not always listen to wise people who tell us that industry will always do better than we will....
Scientific Instruments

Accelerator chain of CERN (operating or approved projects)

P-P, very high energy

Cold anti-P

P, high energy

Nuclear physics
Instantaneous power of one pp-collision: 200 Petawatt
10^9 events/s; 1 Petabyte/s at front-end electronics; 100MB/sec data rate;
150 million sensors, 10 million lines of code; selectivity 1/10^{12}; . . .
2000 collaborators from 150 institutes around the world
Detector Technologies
Simulation - GEANT4

Object-oriented toolkit

Worldwide collaboration

Geometry description of complex detectors

Detailed physics processes

Applications outside high-energy physics
CERN's Users and Collaborating Institutes

Europe: 267 institutes, 4306 users
Elsewhere: 208 institutes, 1632 users

538 70 4306 637 22 87 27 55 10
CERN's Network in the World

270 institutes in Europe, 4600 scientists
210 institutes elsewhere, 1650 scientists
1 LHC experiment: ~1/3 of this
LHC Collaborations - virtual, global Organisations

Clear common mission (~30 y for LHC), clear global and detailed objectives, competition

- Open, critical mass in all scientific and technological fields required, able to deal with all problems posed, therefore world-wide, free exchange of ideas, technologies, R&D, share IP, often “open IP”

- "Lean, bottom-up", democratic, self-organisation; success based on common goals, recognition of individual contributions

- Good record of achievements in terms of delivery to specs, schedules, budgets

- Peer Reviews: technical, scientific, organisational issues, ...

- Resources Reviews: regular reporting/reviews with all funding agencies involved
Increasing collaboration between experiments and focus on common work packages

DataGrid Proposal
1/2001-12/2003
HEP, EO, Biomed

DataGrid Project
EGEE proposal
EU-Integrated Infrastructure Initiative

LHC Computing Grid, LCG, Project
7/2001-12/2005
Phase I
2006-2008
Phase II, . . .

LCG: CERN part of EGEE
Integrate with Globus, . . .
( more work required)

the challenge:

- 4 experiments; 50-200 Hz data taking rate
- raw event size: 0.12 / 1 / 1-25 MB (LHCb / ATLAS-CMS / ALICE)
- total raw data storage: 7 PB/a (7 \cdot 10^{15} \text{ Bytes per year})
- total simulated Data storage: 3.2 PB/a
- world-wide* tape storage: 28.5 PB/a (40 million CD-Rom's)
- world-wide* disk storage: 10.4 PB/a (140k disks à 75 GB)
- world-wide* CPU capacity: 7350 k SI-95 (360k today's PCs)
- WAN bandwidth (Tier-0/-1): 1500 Mbps (1 experiment) (5000 Mbps when serving all 4 exp.'s)

* all Tier-0, Tier-1 and Tier-2 computing centres, excl. Tier-3 and -4
The LHC Computing Model

Experiments

Tier-0

Tier-1
Region 1
Region 2
Region 3
Region 4
CERN

Tier-2
Country 1
Country 2
Country 3
Country 4
Country 5

Tier-3
Lab 1
Uni 2
Uni 3
Lab 4
Uni 5
Lab 6

Tier-4

grid
middleware
The DataGrid Project

- 9.8 M Euros EU funding over 3 years, twice as much from partners
- 90% for middleware and applications (HEP, Earth Obs. and Bio Med.)
- Latest News: 2nd annual project review successfully passed in Feb 2003!
- Total of 21 partners
  - Research and Academic institutes as well as industrial companies
- Related projects and activities:
  - DataTAG (2002-2003)
  - GRIDSTART (2002-2004)
  - Grace (2002-2004)
Through links with sister projects, there is the potential for a truly global scientific applications grid

“The Grid empire is expanding -- it might strike back” (user’s view)
Connections of the different nodes of the EDG test bed are made possible by the EU-funded **GEANT project**

- connecting more than 30 countries across Europe
- speeds of up to 10 Gbit/s
- high data throughput
- Quality of Service

**EDG and GEANT**: the first major production quality tests of the network

- speed
- reliability
- monitoring capabilities
DataGrid in Numbers

**People**
- >350 registered users
- 12 Virtual Organisations
- 16 Certificate Authorities
- >200 people trained
- 278 man-years of effort
  (100 years funded)

**Software**
- 50 use cases
- 18 software releases
- >300K lines of code

**Testbeds**
- >15 regular sites
- >10’000s jobs submitted
- >1000 CPUs
- >5 TeraBytes disk
- 3 Mass Storage Systems

**Scientific applications**
- 5 Earth Obs institutes
- 9 bio-informatics apps
- 6 HEP experiments
DataGrid Applications

Genomic Exploration
Earth Observation
High Energy Physics

more and more scientists begin to use the EDG middleware and testbed, relying on Grid technology to solve huge data challenges
EGEE vision
Enabling Grids for E-science in Europe

Goal
Create a general European Grid production quality infrastructure on top of present and future EU RN infrastructure

Build on
EU and EU member states major investment in Grid Technology
International connections (US and AP)
Several pioneering prototype results
Large Grid development team
Goal can be achieved for about €100m/4 years on top of the national and regional initiatives

Approach
Leverage current and planned national and regional Grid programmes (e.g. LCG)
Work closely with relevant industrial Grid developers, NRNs and US-AP projects
Activities of the partners

Most of partners continued to build national and regional Grid consortia to participate in EGEE

Condition to participate in EGEE is to have already an established Grid activity or to be an established Grid technology centre

EGEE overall project (100 M Euros requested for 4 years) will need to submit staggered proposals to respond to several separate EU calls

The first EU call opened on December 17th and closes May 6th
GÉANT and GRIDs: The model

- GRIDs use GÉANT infrastructure
- GÉANT profits from technological innovation
- GRIDs empowered GÉANT

Application areas

GRIDs platforms

GÉANT network

International dimension

R&D on GRIDs
The LHC Computing Challenge

- **Geographical Dispersion** of people and resources
  - Communication and collaboration at a distance
  - Distributed computing resources
  - Remote software development and physics analysis

- **Complexity** of the detectors and LHC environment
  - 1,000,000,000 highly selected events/year

- **Scale** of the data to be treated
  - Approx. 8 PetaBytes of data/year (10 million CD-ROMS)
  - Need 200,000 of today’s PCs to process the data
Meeting the challenge: the LCG project

LHC Computing project creating a Global Virtual Computing Centre for Particle Physics

Goal - prepare and deploy the LHC Computing environment
applications - tools, frameworks, environment, persistency
computing system → global grid service
cluster → automated fabric
collaborating computer centres → grid
CERN-centric analysis → global analysis environment

acquire and organise support for robust, maintainable middleware
set up and learn how to operate a service
Experiments needs of Grid technology and infrastructure

HEP common application layer HEPcal

Grid Application Group (GAG)

operating through LCG
Timeline for the LCG computing service

**LCG-1**
- Used for simulated event productions
- Principal service for LHC data challenges
- Batch analysis and simulation
- Validation of computing models
- **Stable 1st generation middleware**
- Developing management, operations tools

**LCG-2**
- More stable 2nd generation middleware

**LCG-3**
- Very stable full function middleware
- Acquisition, installation, commissioning of Phase 2 service (for LHC startup)
- Validation of computing service

- **2003**
- **2004**
- **2005**
- **2006**

- **2003**
- VDT, EDG tools building up to basic functionality
- **Stable 1st generation middleware**
- Developing management, operations tools

- **2004**
- **2005**
- **2006**

- **Computing model TDRs**
- Validation of computing models
- **Phase 2 TDR**
- Validation of computing service
- **Phase 2 service in production**

10-March 2003
Hans Hoffmann, CERN
Situation Today

Still solving basic reliability & functionality problems
  Still a long way to go to get to a solid service
  A solid service in mid-2003 looks ambitious
Not yet addressed system level issues
  How to manage and maintain the Grid as a system providing a high-quality reliable service.
  Few tools and treatment in current developments of problem determination, error recovery, fault tolerance etc.
Some of the advanced functionality needed is only being thought about now
  Comprehensive data management, reservation schemes, interactive use, . .
Many many initiatives are underway and more coming

How to manage the complexity of all this?
LCG Guidelines

Focus on a Service for Physics

simulation first

then batch analysis

and later interactive analysis

Keeping everything else

as simple as possible
Deploying the LCG Service

Middleware:
- Testing and certification
- Packaging, configuration, distribution and site validation
- Support - problem determination and resolution; feedback to middleware developers

Operations:
- Grid infrastructure services
- Site fabrics run as production services
- Operations centres - trouble and performance monitoring, problem resolution - 24x7 globally

Support:
- Experiment integration - ensure optimal use of system
- User support - call centres/helpdesk - global coverage; documentation; training
Centres taking part in the LCG prototype service (2003-05)

around the world ➔ around the clock
LCG Grid Technology Organisation

STAG strategic technical advisory group

recommendations

grid technology manager

GAG grid applications group

consultation

requirements consultation

negotiation deliverables

negotiation deliverables

Associated national, regional Grid projects -- GridPP(UK), INFN-grid(I), NorduGrid, Dutch Grid, ...

US projects
LCG Grid Middleware Challenges

Have identified the starting technologies to be deployed
  Driven pragmatically through the GDB/WG1
  Initial suppliers - VDT and EDG
Identify the medium term supply & support strategies
  Requirements from GAG
  Short life-time projects (EDG, Trillium, NMI ....) with unclear continuations
  We need to see credible supplier projects that focus on product quality, maintainability, support, end-user service
  What about industrial products
Work towards future middleware solutions that are coherent, acceptable and supportable
  Inter-working, standards, ...
    -- LCG must be mainline, not HEP-special
OGSA helps - but will standards emerge on the LCG timescale?
EGEE as a Solution for LCG

Middleware
EGEE could provide the critical mass
- to fund hardened & supported middleware serving a wider community
- and form a global middleware partnership

LCG cannot do this on its own

Operation
funds the operation of a core grid in European countries
assuming that -
- the EGEE infrastructure is integrated with national infrastructures
- and with US, Asian infrastructures
- the LCG regional centres are smoothly integrated from the start
- the aim is a long-term infrastructure

Simplification
LCG (and HEP) can concentrate on physics services and leave middleware and grid operation to someone else

The long term model is provision of a core Grid service
GEANT + NRENs provide a model for this
but they had a more tangible deliverable
and the model took some time to mature
LCG as a Catalyst for EGEE

LCG has a real need
and a reasonable scale
and has mature global collaborations of scientists

LCG must acquire basic middleware for a global science grid

Solutions for LCG are quite general - readily applicable to other sciences

LCG will deploy a core grid infrastructure in Europe, America, Asia

EGEE can build on this
to learn how to provide a “production quality” service during LHC preparation (data challenges)
exploit the LHC core infrastructure as the foundation for a general science grid

Timing critical
EU approach: Grid Research and Deployment (FP6)

Grids for Complex Problem Solving
- Architecture, design and development of the next generation Grid
- Enabling application technologies

DG IST - F2

Application-oriented Strategic Objectives
e.g. eBusiness, eGov, eWork, eHealth, risks management

Research Infrastructure
- Deployment of specific high performance Grids
- GEANT Upgrade
- Research networking testbeds

DG IST - F3

Research & Development
125 M€ (IST)

Technology-oriented strategic objectives
e.g. semantic web, software and services

R&D

Deployment
200 M€ RI
**Conclusions**

*Grids need Computer Science and Engineering and Industries as well as ample Prototyping to develop the new, common, initial infrastructure (common approach with NSF Cyberinfrastructure, others?)*

*Grids as “open” infrastructure require global collaboration*

Share middleware development work in a “global, virtual middleware institute” (P.Messina)

- well defined work packages, responsibilities, deliverables, resources and agreed leadership

*Computer-literate application communities are essential to create a useful infrastructure with open standards:*

- generate regional e-Science communities in contact with other regions
- involve applications with the infrastructure teams

Les, Fab, Davids, Bob, Federico, . . . and many have contributed with their work, their transparencies . . . Thanks!