HEPiX and the LCG

HEPiX, Catania
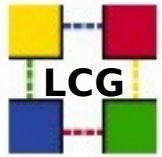## 19 April 2002

John Gordon
eSC, RAL

j.c.gordon@rl.ac.uk

Acknowledgements

- **Les Robertson, CERN – LCG Project Manager**
- **Various speakers at LCG Project Launch Workshop, March 2002**

The LHC Computing Grid Project

LCG

Goal –

Prepare and deploy the **LHC computing environment**

- applications - **tools, frameworks, environment, persistency**

- computing system → **services**
  - **cluster → automated fabric**
  - **collaborating computer centres → grid**
  - **CERN-centric analysis → global analysis environment**

- **foster collaboration, coherence of LHC** regional computing centres

- **central role of** data challenges

This is not yet another grid technology project –

it is a grid deployment project

The LHC Computing Grid Project

LCG

Two phases

Phase 1 – 2002-05

- **Development and prototyping**
- **Approved by CERN Council 20 September 2001**
- **Funded by special contributions from member and observer states**

Phase 2 –2006-08

- Installation and operation of the full world-wide initial production Grid
- Costs (materials + staff) included in the LHC cost to completion estimates

Funding

LCG

- **The LCG Project has two facets**
    - **management of the resources at CERN**
        - CERN base budget and special pledges
        - materials and people
    - **coordination of the overall activities**
        - important external resources – materials and personnel
            - regional centres
            - grid projects
            - experiment applications effort

Funding at CERN – preliminary planning

will evolve as the requirements and priorities are defined

Human Resources required by the LHC Computing Grid Project

preliminary planning numbers - FTEs

team	staff target 2002	CERN complement	special funding
Computer System			
Fabric Planning and Management	12	8	4
Physics Data Storage	7	2	5
Operation	6	2	4
Grid Data Management	6	4	2
LAN Management	3	1	2
Wide-area Networking	4	1	3
Security	4	0	4
Internet Services for Inter-working	6	0	6
Applications Support			
Application Software Infrastructure	5	1	4
Common Frameworks for Simulation and Analysis	13	7	6
Physics Data Management	9	5	4
Support for Physics Applications	9	0	9
Project management and coordination	10	6	4
Total Project Development Activities	**94**	**37**	**57**
LHC infrastructure services (including acquisition, installation, service operation, desktop, web, mail, basic network, ..)	53	53	0
Total resources at CERN	**147**	**90**	**57**

mountain

molehill

Special funding staff status

arrived – 3

contract – 8

in-process – 4-5

Recruitment under way

FZK - 8

PPARC – up to 15

EPSRC – up to 5

Italy - ~5

Funded by EU-Datagrid – 7

Materials funding at CERN

CERN Prototype Capacity & Cost
This is the capacity model agreed by the PEB on 12 February 2002

Capacity

year	2002	2003	2004	2005
processor farm				
no. of 2-cpu systems installed	400	400	600	800
estimated total capacity (SI95)	33,000	33,000	69,800	121,800
disk storage				
no. of disks installed	480	480	960	1,600
estimated total capacity (TB)	47	47	143	271
tape drives				
total capacity (achievable MB/sec)	350	450	600	800
automated media				
total capacity (TB)	100	200	400	600

Materials Costs

	2002	2003	2004	2005	total
Prototype	4,008	1,050	2,930	2,908	10,896
CC refurbishment	1,707	3,200	3,000	3,000	10,907
Physics WAN	10	200	200	2,000	2,410
Total prototype operation	5,725	4,450	6,130	7,908	24,213
Associates, subsistence, etc.	650	650	650	650	2,600
A: Total expenditure/estimate - special funding	**6.375**	**5.100**	**6.780**	**8.558**	**26.813**
B: Basic LHC support from CERN budget	**10.000**	**10.900**	**12.100**	**14.200**	**47.200**
C: Total cost of LHC support at CERN	**16.375**	**16.000**	**18.880**	**22.758**	**74.013**

Area Coordination

- **Applications – Torre Wenaus**
  - **common frameworks, libraries**
  - **general support for applications**
- **Computing Fabrics – Wolfgang von Rüden**
  - **basic computing systems and technology**
  - **CERN Tier 0+1**
  - **automated system management**
- **Grid Technology – Fabrizio Gagliardi**
  - **ensuring that the appropriate middleware is available**
- **Grid Deployment**
  - Regional Centre & Grid Deployment Policy – Mirco Mazzucato
    - authentication, authorisation, formal agreements, computing rules, sharing, reporting, accounting, ..
  - Data Challenge & Grid Operation – open post
    - stable, reliable, manageable Grid for – Data Challenges and regular production work

Grid Technology and the LCG

- **The LCG is a Grid** *deployment* **project**

- **So LCG is a** *consumer* **of Grid technology, rather than a developer**

- **There are many Grid technology projects, with different objectives, timescales and spheres of influence**

Collaboration with Grid Projects

- **Datagrid, iVDGL, PPDG, DataTAG, GriPhyN, CrossGrid, ...**
- **LCG must deploy a GLOBAL GRID**
  - **essential to have compatible middleware, grid infrastructure**
  - **better – have identical middleware**

- **Coordination**
  - **HICB, HIJTB for agreement on common approaches, interoperability (e.g. GLUE)**
  - **and a great deal of work by Fab**

- **The R&D grid projects cannot take commitments for implementation, support, operation of a grid for LHC**

Testbed Trident

LCG

iVDGL
Datagrid

LCG

developers' testbed

development testbed
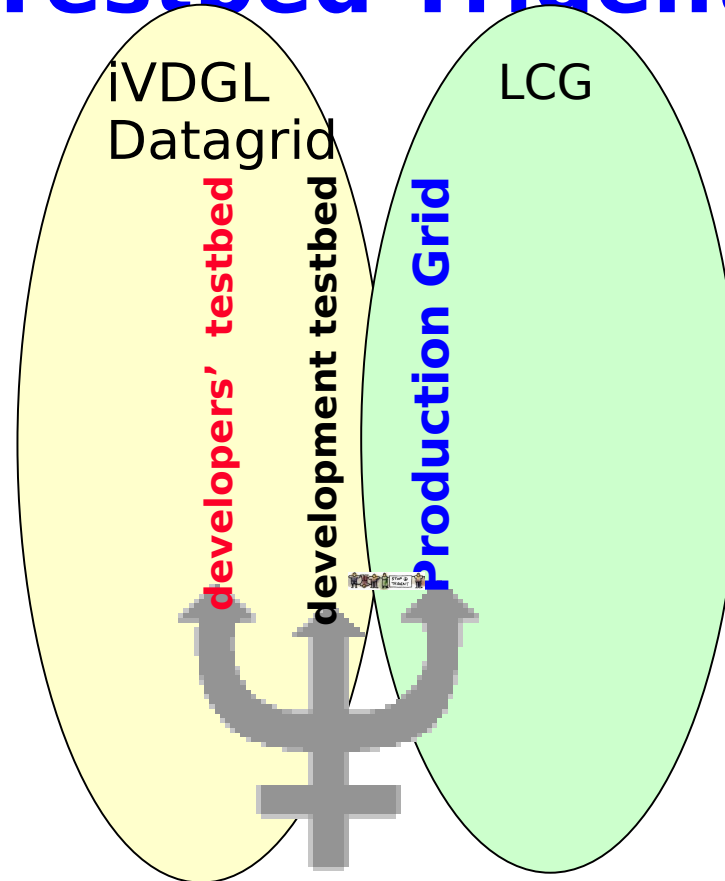
Production Grid

Grid projects -
• responsible for testbeds on which developers can build and test their software
• testbeds for demos, Beta testing, scheduled data challenges

LCG -
responsible for operating a production Grid –
  - 24 X 7 service
  - "as easy to use as LXBATCH"

LCG Grid Deployment DRAFT

- **~April 03 - LCG-1 Global Grid Service**
    - **deploy a sustained 24 X 7 service**
    - **based on "converged" toolkit emerging from GLUE**
    - **and delivered by Datagrid and iVDGL**
    - **~10 sites – including sites in Europe, Asia, North America**
    - **> 2 times CERN prototype capacity (= ~1,500 processors, ~1,500 disks)**
    - **permanent service for all experiments**
- **~Oct 03 –**
    - **reliability & performance targets**
    - **wider deployment**
- **This Grid services evolves slowly through 2005**
    - **new middleware – functionality, performance**
    - **grows with more sites and users**
    - **provides continuous service for the LHC experiments**

Short-term

- **Work closely with Datagrid and iVDGL during 2002**

- **Get experience, work on "productisation"**

- **Agree on which areas are the responsibility of the Grid projects, and which are the long-term responsibility of LCG**

2002 - Grow LCG Production Grid Services leveraging Grid Project Experience

- *stabilise middleware*
- *middleware maintenance and release process*
- *Datagrid testbed & data challenges*
- *iVDGL testbed & data challenges*
- *tests of converged testbed*
- *integrated distribution package middleware/apps support/expt environment*
- *infrastructure operation CA organisation, information services*
- *helpdesk, bug management*
- *licence management*
- *user rules*
- *installation guide*
- *operations manual*
- *user guide*
- *VO management*
- *LCG Global Grid operation*

)
) **Datagrid, iVDGL**
)
)
) **DataTAG, iVDGL**
)
) **share infrastructure and**
) **operation with Grid projects**
)
) **LCG – Regional Centres**
) **Institutes**
) **CERN**
) **….**
) **….**
)

Dependencies on Datagrid, DataTAG, iVDGL

- **Summer 2002 –**
  - **demonstrate that the testbed can do real work for LHC experiments – used by data challenges**

- **Sept 2002 –**
  - **GLUE recommendations available**

- **Oct 2002 –**
  - **agree on the middleware for LCG-1 (based on GLUE)**

- **December 2002 -**
  - **middleware for LCG-1 implemented & available**

So,

- **Where does HEPiX fit in?**

The LHC Computing Grid Project

Goal –

Prepare and deploy the **LHC computing environment**

- applications - **tools, frameworks, environment, persistency**

- computing system → **services**
    - **cluster → automated fabric**
    - **collaborating computer centres → grid**
    - **CERN-centric analysis → global analysis environment**

- **foster collaboration, coherence of LHC** regional computing centres

- **central role of** data challenges

This is not yet another grid technology project –

it is a grid deployment project

Area Coordination

LCG

- Applications – Torre Wenaus
  - **common frameworks, libraries**
  - **general support for applications**
- Computing Fabrics – Wolfgang von Rüden
  - **basic computing systems and technology**
  - **CERN Tier 0+1**
  - **automated system management**
- Grid Technology – Fabrizio Gagliardi
  - **ensuring that the appropriate middleware is available**
- Grid Deployment
  - Regional Centre & Grid Deployment Policy – Mirco Mazzucato
    - authentication, authorisation, formal agreements, computing rules, sharing, reporting, accounting, ..
  - Data Challenge & Grid Operation – open post
    - stable, reliable, manageable Grid for – Data Challenges and regular production work

Fabrics and Grid Deployment

- **Mostly involves the same people within the Regional Centres**

- **Most of the "fabric" decisions are the business of the regional centre**
  - **but should be taken with full knowledge of what is going on elsewhere**

- **While "grid deployment" needs common agreements on many issues - including**

Fabric and Grid deployment – decision making

LCG

- **Must be able to make practical, binding community agreements**
  - **experiments**
  - **regional centre management**

  resources, schedules, security, usage policies, ….

  software environment (op.sys., data services, grid services, ….)

  operating standards, procedures ….

  operation of common infrastructure

- **This needs a** working **group/committee/board**

  Membership
  - **experiment data & analysis challenge coordinators**
  - **regional centre delegates (but not one per regional centre!)**
  - **fabric, grid technology, grid deployment area coordinators**

  Reporting
  - **strategy and planning presented to SC2 for agreement**

Fabrics – information & technology selection

- **Fabrics – technology, practice, experience**
  - **HEPiX**
    - has been acting as an information exchange forum
    - is attended by many regional centre services people
    - organised a large cluster workshop at FNAL last year
    - experiments are NOT present
  - **Can HEPiX become the forum for information exchange and technical discussion of fabric issues important for LCG?**

- **New technology survey (PASTA)**
  - **organised during the next 6 months, convenor David Foster**
  - **experts from regional centres (including CERN)**

Target Audience?

- **Existing HEPiX Sites**
- **Tier1/2 Centres who do not attend – Brookhaven?**
- **Universities who host large resources for experiment**
- **All other institues**

- **Extend our membership**
- **2nd and 3rd are most important**

What to Do?

- **Workshops like the one at FNAL in May 2001**
- **Information Exchange between meetings**
- **Specific workgroups?**

- Discuss!