LHC Computing Grid Project - LCG

Report of the GTA to the GDB

10th March 2003

David Foster – LCG Chief Technology Officer
Information Technology Division
CERN – European Organization for Nuclear Research
Geneva, Switzerland

david.foster@cern.ch

GTA Represented in …

- CERN OpenLab
- IBM grid activities
- HP grid activities
- European Grid Support Center
- EGEE
- HICB
- US ITR proposals
- GriPhyN and iVDGL (EAC)
- GGF Production Grids and OGSA activity
- PEB, GAG, GDB, GDB/WG1, GDB/GFA (Grid File Access)
- DataTag PMB
- UK eScience project definition (future)

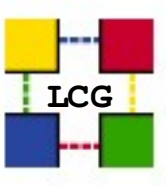
This Presentation

- Hybrid Report
  - Where the GTA development currently is
  - The current status of grid technology
  - The technology impact on LCG-1 this year

GTA Members

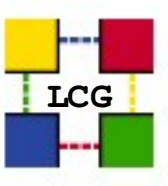
- Massimo Lamanna
  - Sub-contracted to EGEE until May
- Kathrin Paschen
  - Theoretical Computer Scientist (Karlsruhe)
  - Starts 1/04 or 1/05

- **In the context of the LCG, the current role of the GTA is directive and strategic rather than operational.**



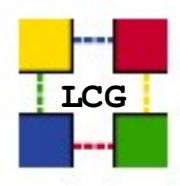
Current Situation

- LCG technology baseline established with WG1
  - Deployment will evolve the current recommendations
  - GAG to define future requirements
- GTA needs to help design overall LCG evolution
  - LCG Architecture
- GTA needs to track major technology changes
  - OGSA
- GTA needs to understand alternative solutions
  - Emergent Solutions
    - Products and Industrial Solutions
    - New Projects
- GTA needs to contribute requirements and Use Cases
  - Emergent Solutions



Architecture

- Single vision of the LCG system architecture is needed.
  - Need to refine what this is
- Several levels of modeling are possible
  - MONARC
    - Extended for the grid (Data distribution/replication policies)
    - Extended with the evolving understanding of physics requirements
  - Grid Infrastructure as a Service
    - Identification of grid services needed
    - Placement and interaction of grid services
  - LocalFabric
    - Worker – Storage Interaction
      - Data – Disk mapping
- Need to solve the most immediate problem first
  - Propose the basic services, how they are interconnected and distributed
    - GGF Proposed OGSA Services and Use Cases



Technology Changes – OGSA

- Globus "production" release still on-track for July
- Not clear how "backward compatible" the GT2/GT3 frameworks will be in practice.
- Basic definition of OGSA services and interfaces still being defined
  - GGF/OGSA working group
- GT3 will need to be proven in practice
  - New GRAM, wrapped GridFTP, CAS+RLS services
  - Integration of Condor-G not clear
  - Performance issues likely
  - New information services model
- Future Developments – but complex and a lot of work
  - Resource management and reservation
  - Database and replication, metadata, caching, access, storage services
  - Security
- At what point should LCG/GTA be involved ?
  - GT3 testbed with experiments from September ?

Technology Summary for 2003

- OGSA will come during 2003 but will be basically GT2 service functionality
- Higher level services are not being addressed (not Globus)
  - Resource Matching (apart from Condor)
  - Job Submission/Control
- Data Management is, and remains, an unsolved problem
  - A real storage element is still missing
  - Replication work is very basic
  - No data placement services
- No real attention being paid to large scale Grids
  - Large aggregated information systems
  - Large replica catalogs
  - High throughput resource brokering
- Basic functionality still being worked on, now in the OGSA context.
- A lot of R&D work still to be done to meet our objectives of a high capacity world wide grid.

LCG 2003

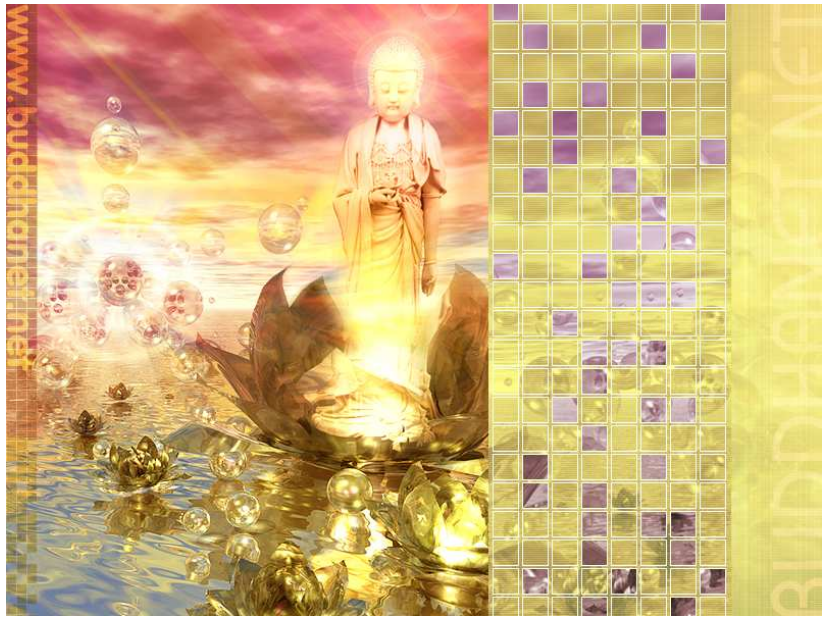
- LCG-1 (mid 2003) = GT2 Base (VDT+EDG+Globus)
  - Can expect slowly diminishing support for GT2.
- Release of GT3 in mid-2003
  - But probably not stable before Q 4
- VDT will be tested with GT3 starting mid-2003.
- EDG components will need to be tested.
- By end 2003 we will have a GT2 system that will do what it will do and a preliminary GT3 system that will be just starting to be tested seriously.
  - LCG-1 may be too ambitious in terms of components to be used based on real experience
- No serious help from industry in this timeframe.
- System for DC04 will be very primitive but may do something sensible if we cut out all elegance and complexity.
  - Scaling may be an inherent problem
  - Will not be the model of the grid for the future
  - "Successful" grids of today are the really simple ones
- OGSA is the platform that will catch the imagination of the developers but this will probably not start to develop until mid-2004.
  - If awareness, education and momentum continue
  - If there is a reason to develop grid components (opportunity)
  - New middleware projects may start in this timeframe (UK)

Things that need to be done for the future TDR

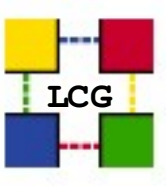
- LCG Grid Systems Architecture
  - Services, Interconnects and Placement
    - Analysis Throughput, System Manageability
- CERN Tier0/1 Systems Architecture
  - Services, Interconnects and Placement
    - Job Throughput, Fabric Manageability
- CERN Tier0/1 Storage Architecture
  - Services, Interconnects and Placement
    - Data Throughput, Data Manageability
- Physics Analysis Data Management Architecture
  - Services, Interconnects and Placement
    - Access Patterns, Data Relationships and Replication Policies

- Need to start working towards these objectives

Things to Watch



- Industrial Activities
  - IBM
  - Avaki
  - Platform
  - Veritas
  - HP
  - GridSystems
- New Projects
  - EGEE
  - New UK eScience projects
  - OMII? = UK Core + EGSC + NMI + (part)EGEE +?
- Other Projects
  - UNICORE (EUROGRID)



Some Open Questions

- EDG software
  - No support path beyond 2003.
  - No clear direction for evolution.
    - OGSA?
    - Missing Functionality?
    - Scalability?
- US Projects have a finite life (VDT, NMI, Globus) but tend to have a rolling future.
- We need to evolve as well as a technical architecture, a grid management architecture for LCG.
  - Communication flows and interactions
  - Escalation procedures
  - Necessary and sufficient formalism (documentation, meetings, responsibility)
- LCG currently has a big exposure to non-delivery of functionality by the various middleware activities.
  - No grid solution we can "live with"
  - Specific Middleware activities driven by funding
  - LCG the most complex of projects from an organisational perspective so simplification is needed wherever possible.

Current Activities

- Baseline activities
  - WG1 Completed
  - Grid File Access being worked on
    - Foster, Ernst, Petravik
- Need to start some architecture activities
  - Foster, Lamanna, Paschen
- Globus OGSA Technology evaluation
  - Experiments ?
- IBM OGSA Solution
  - In the context of OpenLab ?
- Others ?