LHC Computing Grid Project – LCG

LHCC Review November 2003

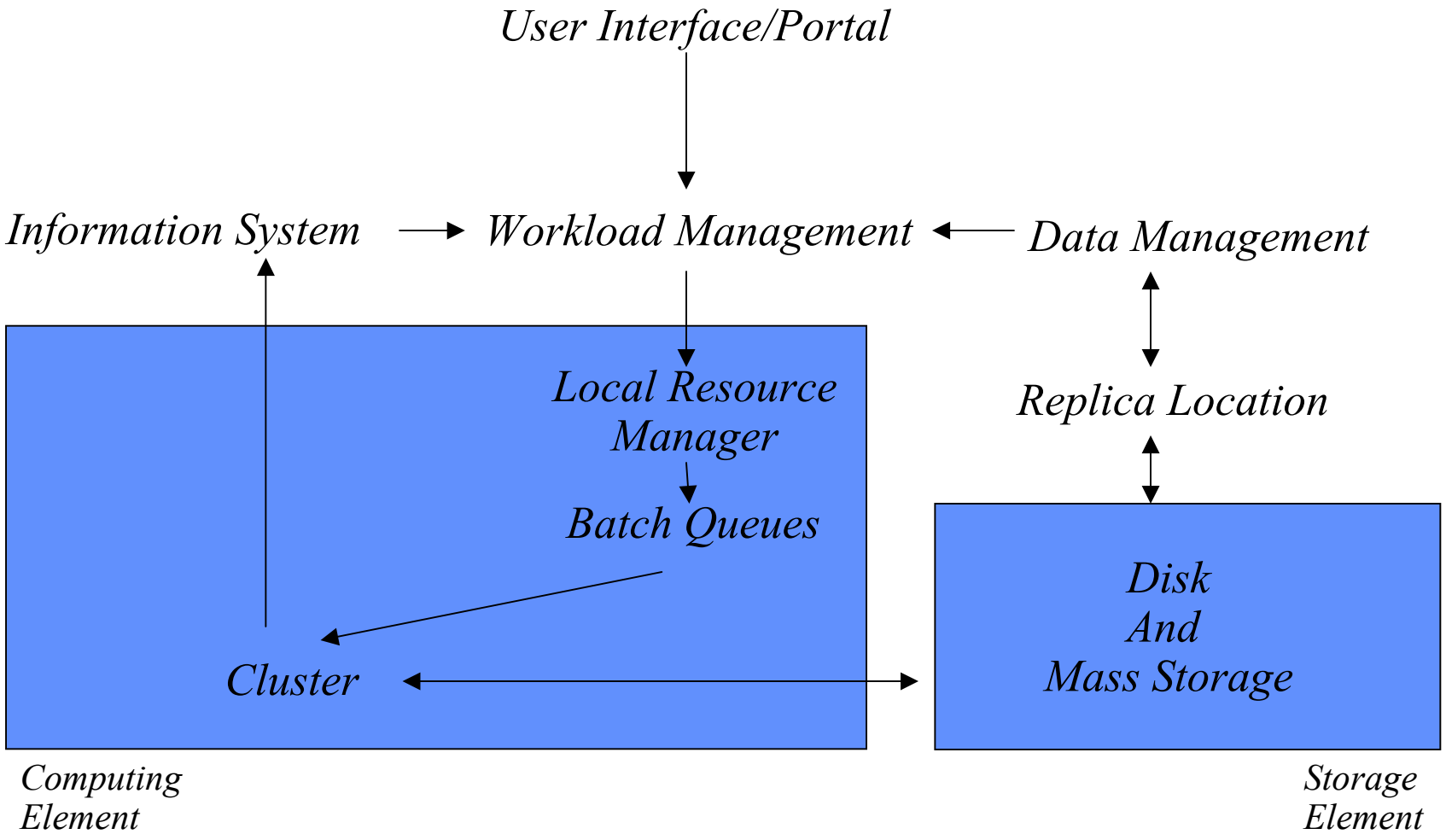
Grid Technology Area (GTA)

David Foster
david.foster@cern.ch

Review Structure

- Objectives
  - To put the GTA in the context of the overall LCG project
  - To explain the major issues facing the project in this area
  - To indicate the future strategy for the GTA area
- Agenda
  - Introduction to the Grid Technologies
  - GTA Achievements and Relationships
  - Future Challenges
  - EGEE

Functional Overview of a Grid



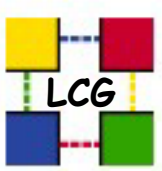
Some Fundamental Services

- Communication
  - Internet protocols: IP, DNS, routing, etc.
- Security: Grid Security Infrastructure (GSI)
  - Uniform authentication & authorization mechanisms in multi-institutional setting
  - Public key technology, SSL, X.509, GSS-API and supporting infrastructure: Certificate Authorities, key management, etc.

GSI: www.gridforum.org/security

The Local Resource Manager

- The Gatekeeper is the Globus Toolkit Implementation of the Grid Resource Allocation Mgmt (GRAM)
  - Remote allocation, reservation, monitoring, control of compute resources
  - Single point of entry Authenticates user, maps to local security environment, runs service
  - Job manager
    - A gatekeeper service, one instance per job
    - Layers on top of local resource management system (e.g., PBS, LSF, etc.)
    - Handles remote interaction with the job



Metacomputing Directory Service (MDS)

- Access information in a distributed directory
  - Directory stored in collection of LDAP servers
  - Each server optimized for particular function
- Directory can be updated by
  - Information providers and tools
  - Applications (i.e., users)
  - Backend tools which generate info on demand
- Information dynamically available to
  - Tools
  - Applications

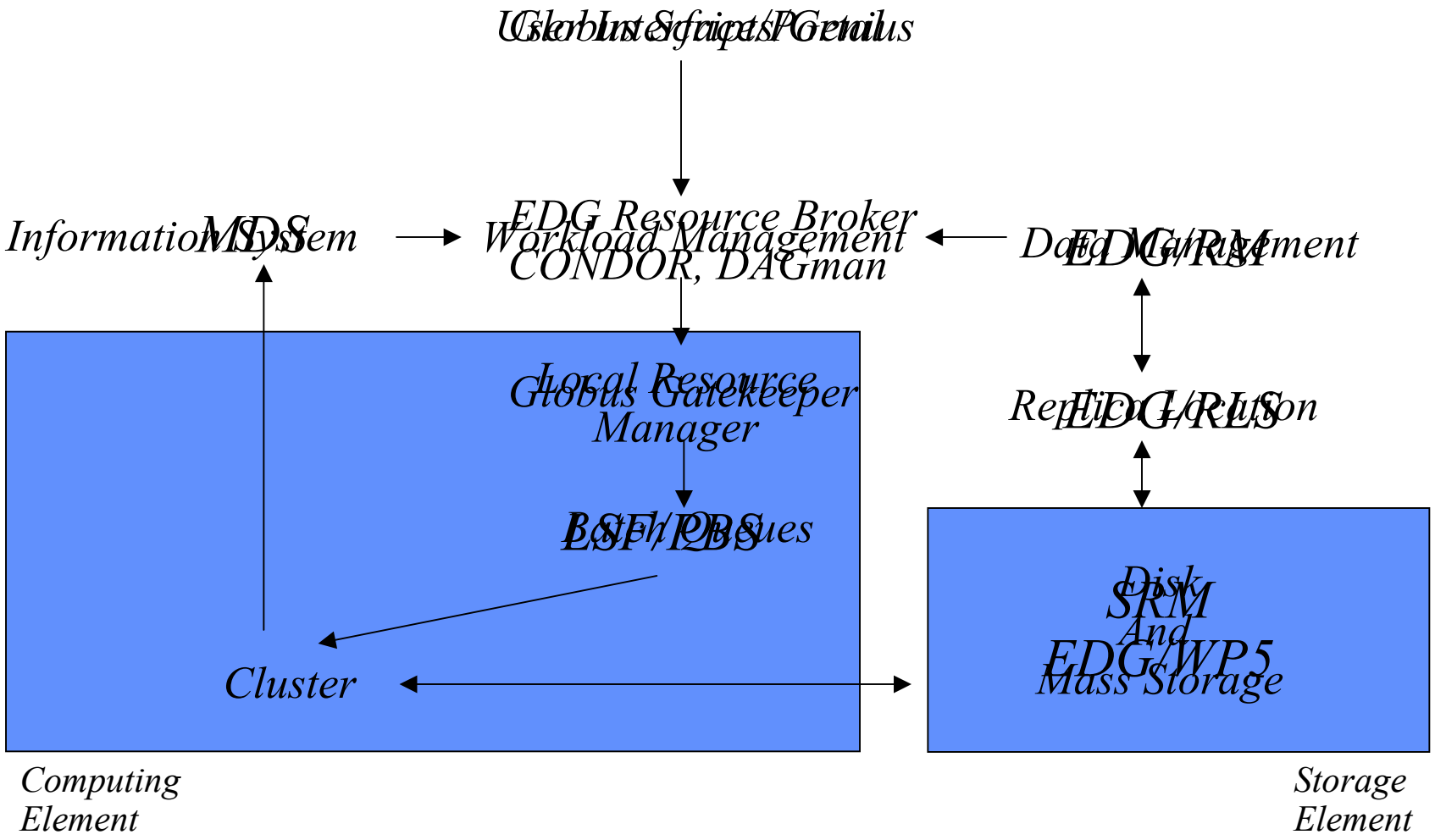
MDS Structure

- Provide access to static and dynamic information regarding system components
  - Grid Resource Information Service (GRIS)
    - Supplies information about a specific resource
    - Mostly "white pages" lookups
  - Grid Index Information Service (GIIS)
    - Supplies collection of information which was gathered from multiple GRIS servers
    - Supports efficient queries against information which is spread across multiple GRIS servers
    - Mostly "yellow pages" lookups

Basic Data Management Services

- Replica Location Services (RLS)
  - Database for registering/locating:
    - Multiple copies of files
- Replica Management (RM)
  - Creating/Deleting/Maintaining Replicas
- GridFTP: Data Transfer and Access
  - Common protocol for data movement
    - Secure, efficient, reliable, flexible, extensible, etc.
    - Grid Forum (Internet) Draft
  - Family of tools supporting this protocol
    - Wu-ftpd, ncftp, Globus Toolkit SDKs, etc.

Functional Overview of a Grid



Projects and Technologies

- Globus – provides the basic toolkit and infrastructure
- European Data Grid (EDG) – provides higer level services
- Condor – an implementation of a distributed batch system that can interface with Globus (Condor-G)
- VDT – a bundling of Globus and Condor providing a number of different grid systems to be created.
- SRM – Storage Resource Manager is a specification of a uniform interface to mass storage.
- Enabling Grids for E-science in Europe (EGEE) – deployment project that will "productise" then deploy middleware.

Some I have not mentioned …

Grid Middleware (core services)

- Cosm P2P Toolkit
- Globus
- Gridbus
- Grid Datafarm
- Storage Resource Broker (SRB)
- GridSim: Toolkit for Grid Resource Modeling and Scheduling Simultation
- Simgrid
- Jxta Peer to Peer Network
- Legion: A Worldwide Virtual Computer
- PUNCH

DataGrid Initiatives

- Virtual Laboratory: Tools for Data Intensive Science on Grid
- EU DataGrid
- DIDC Data Grid work
- GriPhyN (Grid Physics Network)
- HEPGrid (High Energy Physics and Grid Networks)
- Particle Physics Data Grid (PPDG)
- Datacentric Grid
- GridPP

Grid Schedulers

- Nimrod/G Grid Resource Broker
- AppLeS
- SILVER Metascheduler
- ST-ORM
- Condor/G
- NetSolve
- DISCWorld
- Computing Centre Software (CCS)

Grid Systems

- Compute Power Market
- Global Operating Systems
- XtremWeb
- JAVELIN: Java-Based Global Computing
- MILAN: Metacomputing In Large Asynchronous Networks
- Harness Parallel Virtual Machine Project
- Management System for Heterogeneous Networks
- PUNCH - Network Computing Hub
- MOBIDICK
- MetaNEOS
- Amica
- MultiCluster
- Poland Metacomputing
- Echelon: Agent Based Grid Computing
- Bayanihan
- NeuroGrid
- GridLab
- DAMIEN
- CrossGrid
- DIET

Grid Programming Environments

- Nimrod - A tool for distributed parametric modeling
- Ninf
- Cactus Code
- MetaMPI - Flexible Coupling of Heterogenous MPI Systems
- Virtual Distributed Computing Environment
- GrADS: Grid Application Development Software Project
- Jave-based CoG Kit
- GAF3J - Grid Application Framework for Java
- ProActive PDC
- REDISE - Remote and Distributed Software Engineering
- Albatross: Wide Area Cluster Computing

Grid Testbeds and Developments

- World Wide Grid (WWG)
- Polder Metacomputer
- NASA Information Power Grid (IPG)
- NPACI: Metasystems
- Asia Pacific Bioinformatics Network
- The Distributed ASCI Supercomputer (D
- G-WAAT
- Micro Grid
- Alliance Grid Technologies
- The Alliance Virtual Machine Room
- EuroGrid
- Internet Movie Project
- Nordic Grid
- ThaiGrid
- TeraGrid
- Irish Computational Grid (ICG)
- GrangeNet
- LHC Grid
- I-Grid
- OurGrid
- Kerala Education Grid
- N*Grid Korea

Grid Technology Area

- The structure of the LCG project has been well documented and discussed
  - http://www.cern.ch/lcg
- The GTA has its own web pages containing much information
  - http://www.cern.ch/lcg/peb/gta
- The GTA area went through a number of transitions
  - Until End 2002 was driven by the EDG project leader (Fabrizio Gagliardi).
  - From 2003 was driven by the current GTA manager (David Foster)
  - From April 2004 will be driven by the EGEE middleware manager (Frederic Hemmer)

GTA Personnel

- During 2002
  - Solely the EDG Project Leader
- To May 2003
  - Solely the current GTA Manager
- From May 2003
  - 1 FTE was added (Massimo Lamanna) to work on GTA issues in general.
- From June 2003
  - 1 FTE was added (Kathrin Paschen) to work on emerging modeling problems.
- In July 2003 the OGSA Engineering team was established under the GTA
  - Rotating people from Moscow State University, Dubna and the Academia Sinica Taipei contributed plus one student from the EDG.

The GTA has been only a small area within the overall project in terms of staffing

LCG Grid Technology

- LCG does not develop and therefore does not control the middleware technology evolution.
- But, within the project we do need to:
  - Identify the starting technologies (2003) to be deployed.
  - Identify the evolution strategy (2004-2006).
  - Identify the long term support strategies.
  - Work towards future middleware solutions that are coherent, acceptable and supportable.
- The GTA has worked on a number of well defined projects to enable the above objectives which will be described later
  - Technology Tracking
  - Technology Selection
  - Technology Evaluation
  - System Design

Technology Tracking

- PASTA III
  - During 2002 an updated report on "Processors, Architecture, Storage and Tapes" was created with the help of many external contributors.
  - This was completed in February 2003 and used to create the new costing model for LCG Phase I and Phase II.
  - An LCG Seminar on the results of PASTA III was held in 26 June 2003.
- UK e-Science
  - The UK e-Science core project was very active during 2003 and created a comprehensive review of grid technologies. A summary of this was created which was used as an LCG milestone.

Technology Selection

- GDB/WG1
  - Grid Deployment Board mandated a number working groups in November 2002. The Working groups completed their reports in January 2003.
  - The GTA took a leading role in the WG1 which was the technology selection for LCG-1. In particular it recommended:
    - A base supported level of VDT to be used.
    - A number of EDG components in the first release with others as they become ready.
    - That the issue of generalised grid file access to be studied.
- GFAL
  - The GDB recognised the file access problem and mandated a solution to be designed. The GTA completed this work in April 2003.
  - The GTA managed the implementation of the first prototype (May 2003) with manpower from the GDA.
  - The GFAL solution is to be in full production by the end of 2003.

Technology Evolution

- OGSA Engineering team
  - The GTA proposed to the LHCC referees in June 2003 that a serious effort be started to study the viability of the OGSA proposals and the Globus Toolkit 3 release in particular.
  - A team was put together in July 2003 with manpower from the GTA, EDG, MSU, Dubna and the Academia Sinica in Taipei.
  - The intention was to:
    - Report quickly (after 2 months)
    - Create an understanding of:
      - The effectiveness of GT3
      - The problems in creating new services
      - The opportunity to adapt existing software (AliEn)

Technology Evolution

- OGSA Engineering team
  - Many interesting results on performance and scalability issues were created and documented. (See GTA web pages)
  - The results were presented in an LCG seminar 24 Sep 2003.
  - The work was extremely well received by the Globus community and has resulted in a new and productive relationship.
    - Globus has changed some priorities based on our work (GRAM)
  - The EGEE middleware activity has supported the work and its continuation as a valuable pre-cursor to the starting of the EGEE project.
  - It will be a valuable resource for the emerging ARDA implementation.

System Design

- Modeling
  - Very limited due to little manpower. Need to concentrate on specific small parts of the bigger problem.
  - Identified general issues at three different levels
    1. Experiment Data Models
    2. Middleware architecture and scalability
    3. Fabric solutions
  - Spent much time in understanding the modeling tools and their limitations.
  - Currently working with the Monarc tools to understand the scalability issues in the new GT3 information system architecture and the RLS file catalog.
- RTAGs
  - The GTA was a contributor to a number of RTAGs but more specifically HEPCAL II and ARDA

Middleware Relationships

- Globus
  - Enhanced relationship through the OGSA engineering activity based on joint problem solving and testing.
    - Much confidence building done
- VDT
  - Recent meetings held in Wisconsin and CERN to understand the role of VDT in future systems (e.g ARDA).
    - Much synergy and opportunity to combine the best of different developments.
- UK eScience
  - Tracking the activities of the "Core" e-science developments as part of the Software Engineering Advisory Group (SEAG).
    - Intention is to develop robust and "production" middleware.
- IBM
  - Evaluating the GT3 offering as part of the OGSA engineering effort.

EGEE

- The EGEE is a deployment project with a substantial middleware re-engineering activity to address the issues of the current middleware software.
- Decision was taken to combine the GTA and the EGEE middleware activity
  - Much mutual dependence and synergy of the two projects
  - Multiple middleware activities creates confusion
  - EGEE will have considerable manpower for technology developments
- But there are issues that will need working out as we go along
  - Does the world wide focus of the LCG and the European focus of the EGEE mix?
  - Will there be conflicts of priorities?
  - Can the technology of a long term project (LCG) come from a short term project (EGEE) and how will it be supported?
- There is cross representation of the EGEE and LCG projects through the respective management and execution boards.

Where we are

- We have seen that progress is being made, but slowly
  - The current middleware is not what we would have hoped but was the best available to be deployed at the beginning of this year.
  - Many complex problems remain in the grid technology area
    - Not much experience of the reliability of "grid systems".
    - Scalability is still an issue.
    - Manageability is still an issue.
  - In our problem domain many issues still exist
    - Configuration management.
    - Security and VO management.
    - Data management.
  - Much experience yet to be gained
    - Job throughput and splitting.
    - Efficient scheduling and matching.
    - Chaotic workloads.
- Many others too … seems clear we are at the very beginning of a very long process of technology development and evolution.
  - Concentrate on understanding and solving the basic problems by deploying a simple, but production, system based on the LCG-1 and LCG-2 technologies.

Networking

- Effective use of underlying network infrastructures still need much further work.
  - New EU proposals address some issues of middleware interaction (GRANDE).
  - New projects (*light) aim to create high speed (10Gbit) network infrastructures.
  - New protocol stack implementations required to exploit WAN of high latency.
    - Moving us towards an underlying circuit switched special purpose network in addition to a general purpose packet switch network.
- Tier 0/1/2/3 expected relationships and computing models will modify the networking requirements. These are starting to be debated.

Medium term opportunities

- 2003
  - Until now we have envisaged complex schemes and functionality.
    - Existing middleware is generally overly complex and under developed.
  - Politics and Marketing is both providing and consuming tremendous effort.
- 2004
  - **LCG-1 and LCG-2 production services will continue to be developed and reach a level of operational stability.**
  - **A service based approach (ARDA) will be developed in parallel but will need to be proven as a strategic direction.**
    - Milestones in March, June and October?
  - **Industry activity likely to grow slowly**
  - **Relationship of LCG/EGEE to US HEP activities critical to overall success**
    - Leverage the best technologies emerging in the coming years within the service based model.
- 2005
  - We need to be cleaning up the details of the LHC computing solution.

Conclusions

- Existing plans and planning assume a deterministic path to success but grid technologies are far from this at present…
  - Understanding of both requirements and technology is evolving
- Main middleware risks are lack of delivery of an effective middleware package due to:
  - Over ambition, complexity and too short timescales
  - "Mythical Man Month" problems.
  - Inertia generated through complex dependencies between projects.
- Managing complexity is the major challenge
  - **LCG-1 and LCG-2 will improve to address the main basic functionality.**
    - MSS to MSS copy across the wide area.
    - Local data access from worker nodes.
    - Simple job submission and control.
    - Excellent configuration management and deployment tools
  - **Realistic requirements needed through the GAG and HEPCAL-II like activities.**
  - **ARDA will prepare the "next generation" middleware in parallel by leveraging existing technologies and the basic functionality required.**
- We look forward to EGEE using its resources to move to the next phase in the LHC grid system (Analysis) but we have not yet mastered the existing phase (Production) fully.
  - We all (LCG, EGEE and the experiments) will need to put effort into building up a series of realistic analysis challenges.