Advanced Technologies for Distributed Database Services Hyperinfrastructure

DPF Meeting

UC Riverside, USA

August 30, 2004

A. Vaniachine (Argonne National Laboratory)

M. Vranicar (PIOCON Technologies)

D. Malon (Argonne National Laboratory)*

**presenting on behalf of A. Vaniachine*

ARGONNE NATIONAL LABORATORY | THE UNIVERSITY OF CHICAGO | U.S Department of Energy | Office of Science Department of Energy | PIOCON TECHNOLOGIES

Outline

- Petascale data challenge, LHC and Grid
- Non-event data in RDBMS and the Grid
- RDBMS access limitations on the Grid
- Beyond the existing grid infrastructure we propose a hyperinfrastructure of distributed database services for efficient secure data access

Petascale Data Challenge

- The Large Hadron Collider (LHC) at the CERN Laboratory will become the largest scientific instrument in the world when it starts operations in 2007
- LHC experiments face an unprecedented multi-petabyte event data processing task. To address that challenge LHC computing is pioneering the deployment of emerging computing grid technologies
- Until recently LHC computing models' focus has been limited to the problem of managing the petabyte-scale file-based event store
- The grid security model and efficient data transport mechanisms are particularly well suited for handling file-based data
- File cataloging is reasonably well served by grid middleware infrastructure (RLS, etc.)
  - Emerging grid support for file-level metadata as well

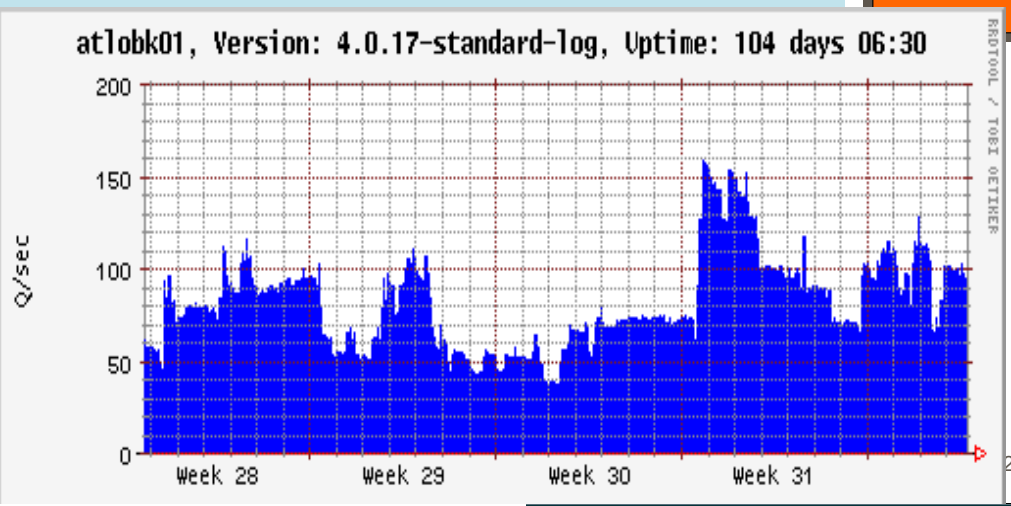
Databases and the Grid

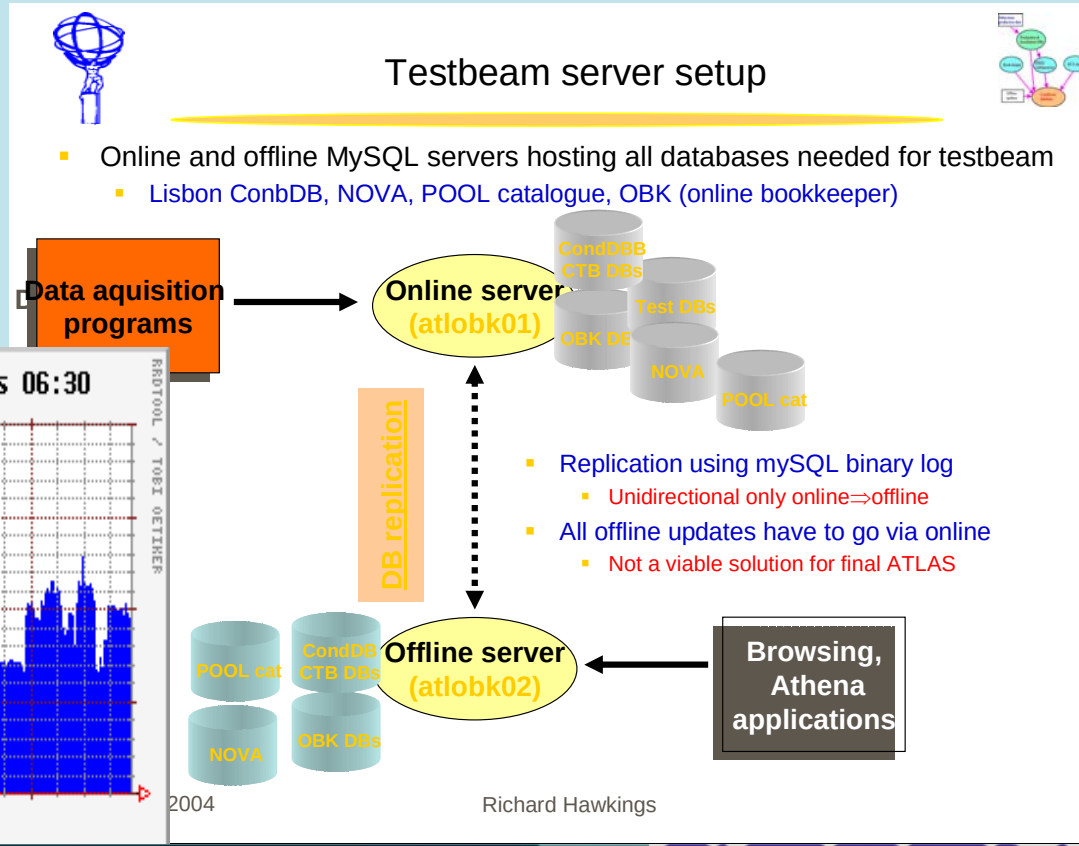
- In addition to file-based event data, LHC data processing applications traditionally require access to large amounts of valuable non-event data (detector conditions, calibrations, etc.) stored in relational databases
- In contrast to the file-based data, this database-resident data flow has to be detailed further
- ATLAS Data Challenges exercise the Computing Model, processing and managing data on three different grid flavors

Non-event Data Flow (1)

ATLAS Combined Testbeam of 2004

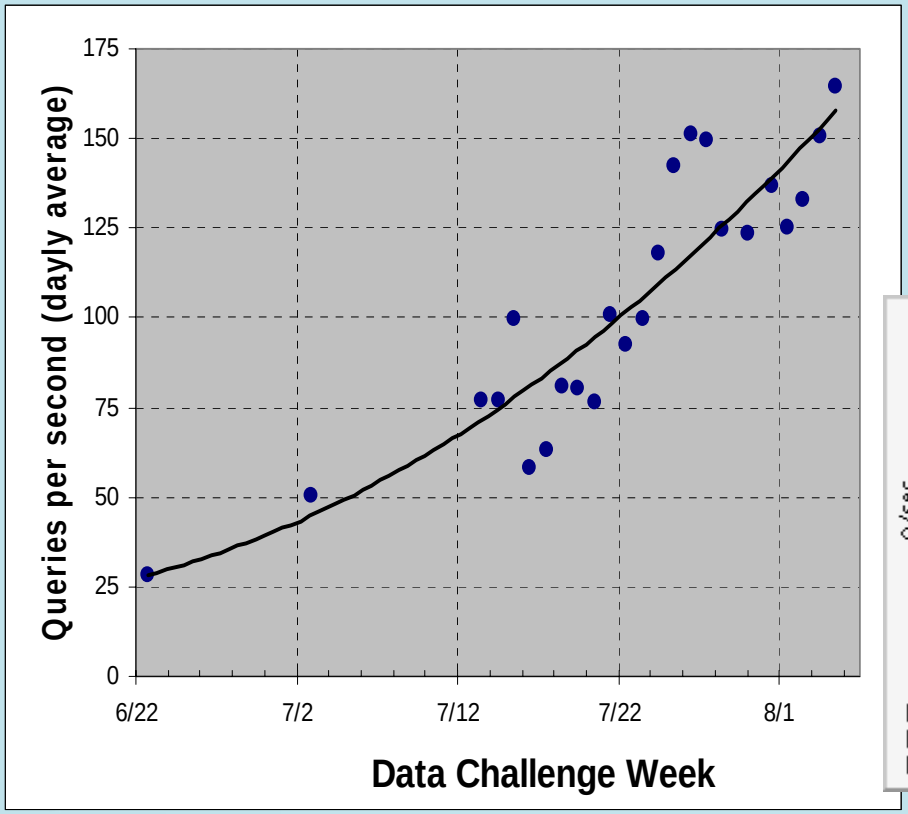
- Relatively secure LAN environment



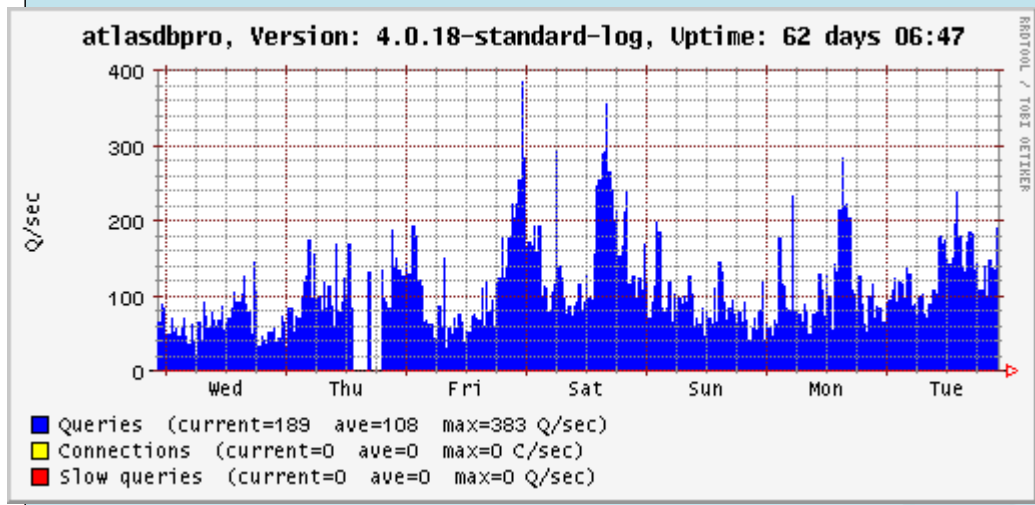


Non-event Data Flow (2)

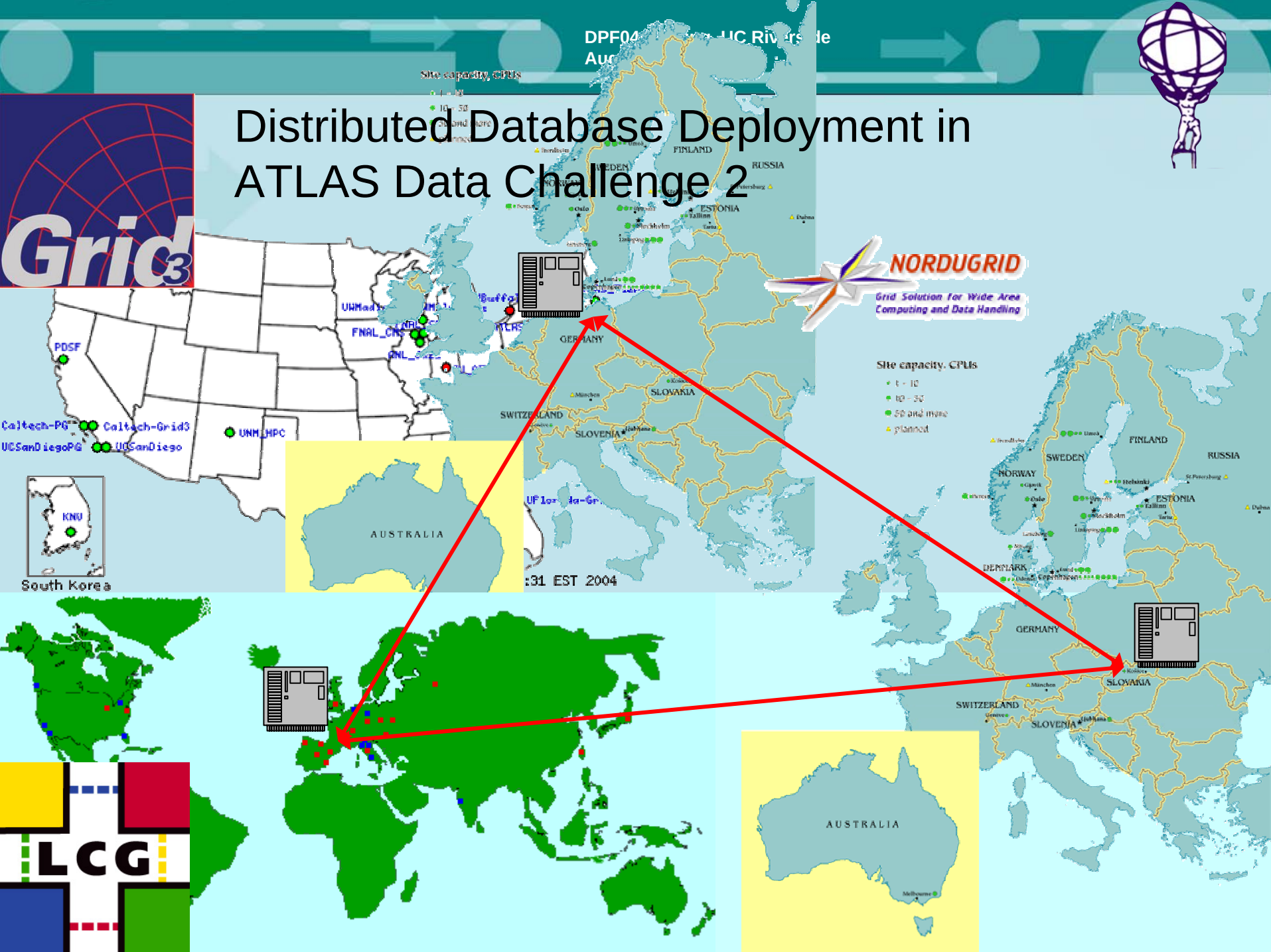
- Insecure WAN Environment of ATLAS Data Challenge 2



- Server load increase during ramp-up of ATLAS data processing on the Grid



Distributed Database Deployment in ATLAS Data Challenge 2

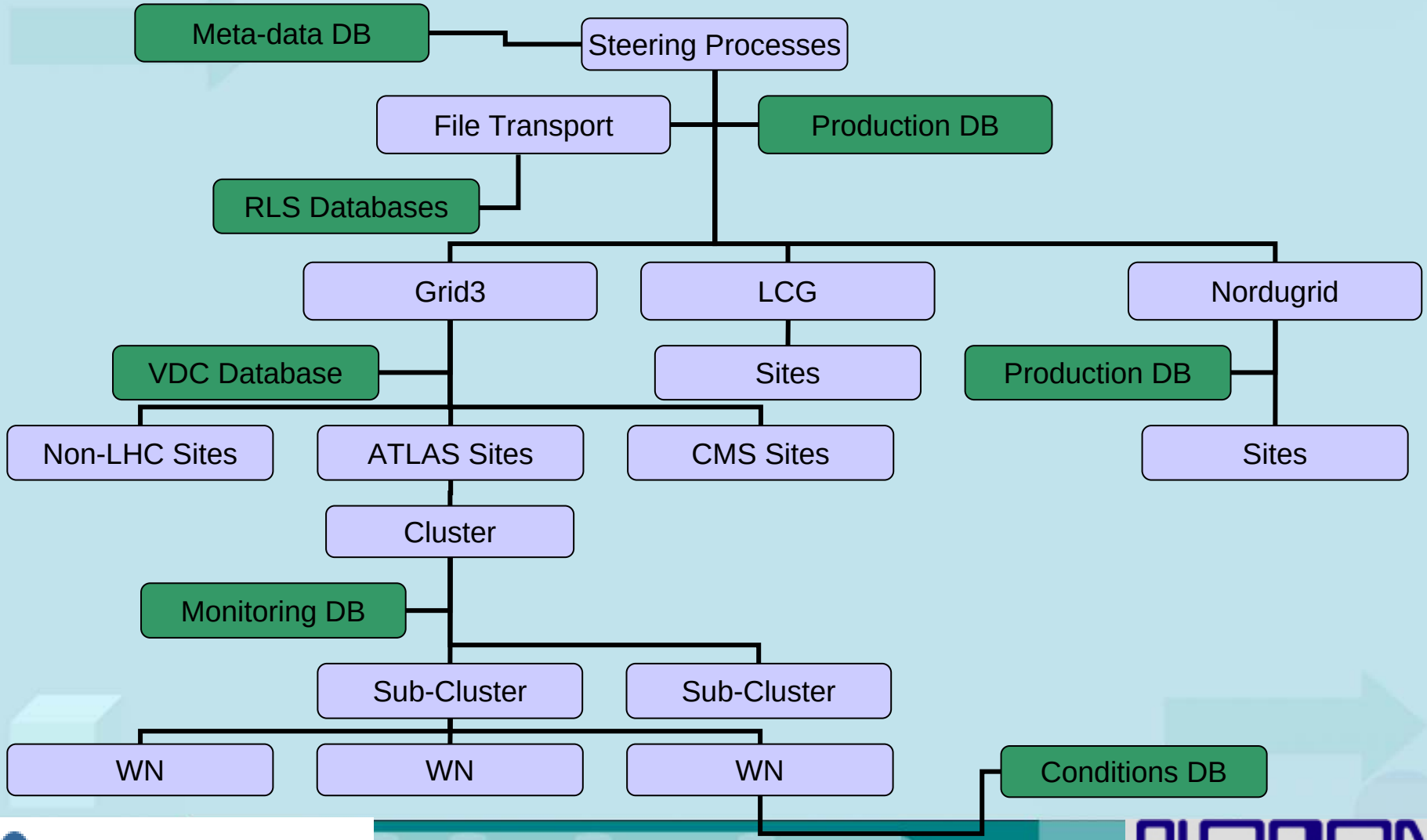


Database Access in Data Challenge 2

- Note: three inhomogeneous grids, not three sites
- All DC production is entirely grid-based: no use of non-grid batch queues
- Access to relational-database-resident data, while significant, would be greater still if ATLAS calibrations/conditions infrastructure were further developed

Grid Computing Challenges

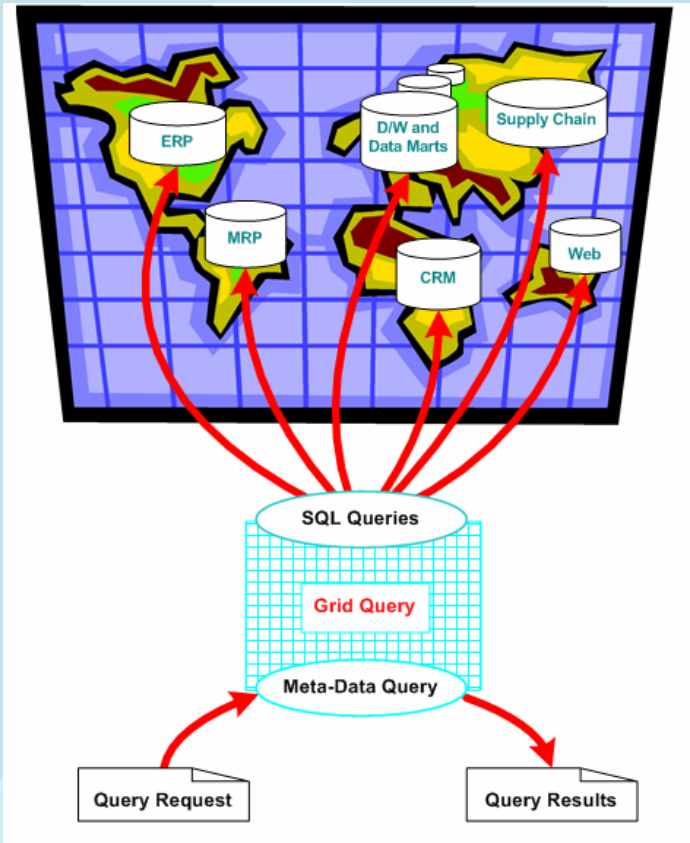
- Experience from deployment of grid technologies in a HEP production environment of D0 Run II and ATLAS Data Challenges demonstrated that a naïve view of the grid as a p2p system is inadequate
- Operations show that just a single database on top is not enough to run the production system effectively. A hyperinfrastructure of databases below raises the production system efficiency considerably.
- Thus a hyperinfrastructure of databases on the grid play the dual role: both as a built-in part of the middleware (monitoring, catalogs, etc.) and as a distributed infrastructure of the production system necessary to run the scatter-gather data processing applications on the grid. It is the latter role that supplies operations history that is providing data for the important 'post-mortem' analysis.
- The data-mining of the collected operations data reveals striking features of the computing hyperinfrastructure – a very high degree of correlations between the failures. I.e., if the job submitted to some cluster failed, there is a very high probability that a next job submitted to the same cluster would fail too. Alternatively, if the submit host failed, all the jobs scattered over different clusters will fail too. Taking these correlations into account is not yet automated by the grid middleware.
- That is why production databases and grid monitoring data that are providing immediate feedback on the Data Challenge operations to the production operators is very important for efficient utilization of grid capacities.


Meta-data DB
Steering Processes
File Transport
Production DB
RLS Databases
Grid3
LCG
Nordugrid
VDC Database
Sites
Production DB
Non-LHC Sites
ATLAS Sites
CMS Sites
Sites
Cluster
Monitoring DB
Sub-Cluster
Sub-Cluster
WN
WN
WN
Conditions DB

Securing Data on the Grid

- Secure grid query engine technologies federating heterogeneous databases on the grid
- Methods utilizing Grid Security Infrastructure data-transport channel for database services delivery to the grid clusters behind closed firewalls
- Grid certificate authorization technologies for database access control where the safety features are pushed into the database engine code

Grid Query Technology



Database Grid Solution

Technology enabling queries of distributed, heterogeneous databases, all through one simple, Grid enabled interface.

- Access to multiple, distributed, heterogeneous data sources
- Databases can house like content, such as in replicated environments
- Databases can house related, but non-identical content and non-identical data structures
- Meta-data layer masks details

Fermilab

Sam Queries

Database Grid Solution

- Prototype of the Database Grid Solution is in use in Fermilab's Run II Data Handling system, servicing millions of queries per day

Large, complex, distributed user base:	
Multiple experiments	D0, CDF, MINOS
Worldwide access	26 countries, 100+ institutions
Large user base	1500+ physicists
Large database	> 750M events, > 2.4M data files 1.5TB database size
Frequent queries	> 3M per day
Complex relational model	100+ tables
Hierarchical model (future)	Central, Regional, Institutional, and Desktop Centers/Stations

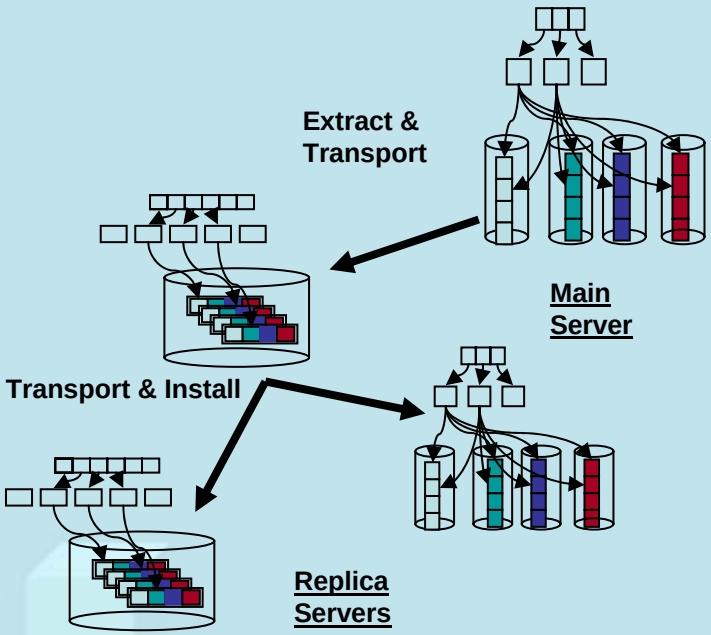
Database Access on the Grid

Two different security models (and schools)

- A separate server does the grid authorization:
  - Spitfire (EDG WP2) – SOAP/XML text-only data transport
  - DAI (IBM UK) – Spitfire technologies + XML binary extensions
  - Perl DBI database proxy (ALICE) – SQL data transport
  - Oracle 10g (separate authorization layer)
- Authorization is integrated in database server:
  - Instead of surrounding database with external secure layers the safety features are embedded inside of the code
  - By pushing secure authorization into the database engine the inefficient data transfer bottlenecks are eliminated

Secure GSI Transport Channel

Extract-Transport-Install



- MySQL simplified the delivery of the extract-transport-install components of ATLAS database architecture to provide database services needed for the Data Challenges for sites with Grid Compute Elements behind closed firewalls (some sites on Grid3 and NorduGrid)

Distributed Deployment of Databases

- LHC data processing will make use of a distributed infrastructure of relational databases for access to various types of non-event data and event metadata
- Distribution, replication, and synchronization of these databases, which are likely to employ more than one database technology:
  - Oracle in larger centers
  - MySQL in smaller settings and for smaller applications

  must be supported according to the needs of the various database service client applications
- We proposed and are involved in the new Distributed Deployment of Databases project at LCG to define and implement common LHC solutions
  - http://lcg3d.cern.ch

Why a LCG Database Deployment Project?

- Need for some standardisation as part of LCG
  - To allow applications to access data in a consistent, location independent way
  - To allow to connect existing db services via data replication mechanisms
  - To simplify a shared deployment and administration of this infrastructure during 24*7 operation
  - To increase the availability and scalability of the total LCG system
- Need to bring service providers (site technology experts) closer to database users/developers to define a LCG database service for the upcoming data challenges in 2005

Summary

- To overcome database access limitations one must to go beyond the existing grid infrastructure
- We proposed and tested the technologies laying a foundation of a new hyperinfrastructure:
  - Secure grid query engine technologies federating heterogeneous databases on the grid
  - Methods utilizing Grid Security Infrastructure data-transport channel for database services delivery to the grid clusters behind closed firewalls
  - Grid certificate authorization technologies for database access control where the safety features are pushed into the database engine code
- These technologies have been tested in the production environment of ATLAS Data Challenges and Run II experiments at Fermilab
- The LCG Distributed Deployment of Databases project provides a testbed to advance these technologies further for LHC Computing
  - A regional collaboration of ATLAS and CMS is being set up in the US for exercising the 3D slice involving Fermilab, Argonne and U Chicago