D0 Grid: Architecture and Status

Igor Terekhov et al FNAL/CD/CCF, D0, CDF, Condor team, UTA, ICL
Plan of Attack

- Brief History, D0 and CDF computing, data handling
- Grid Jobs and Information Management
  - Architecture
  - Job management
  - Information management
  - JIM project status and plans
- Globally Distributed data handling in SAM and beyond
- Summary
Globally Distributed Computing and Grid

- D0 – 78 institutions, 18 countries. CDF – 60 institutions, 12 countries.
- Many institutions have computing (including storage) resources, dozens for each of D0, CDF
- Some of these are actually shared, regionally or experiment-wide
  - Sharing is good
  - A possible contribution by the institution into the collaboration while keeping it local
  - Recent Grid trend (and its funding) encourages it

Igor Terekhov, FNAL
Goals of Globally Distributed Computing in Run II

- To distribute data to processing centers – SAM is a way, see later slide
- To benefit from the pool of distributed resources – maximize job turnaround, yet keep single interface
- To facilitate and automate decision making on job/data placement.
  - Submit to the cyberspace, choose best resource
- To reliably execute jobs spread across multiple resources
- To provide an aggregate view of the system and its activities and keep track of what’s happening
- To maintain security
- Finally, to learn and prepare for the LHC computing
Data Distribution - SAM

- SAM is Sequential data Access via Meta-data.
  [http:// {d0,cdf}db.fnal.gov/ sam](http:// {d0,cdf}db.fnal.gov/ sam)
- Presented numerous times, prev CHEPS
- Core features: meta-data cataloguing, global data replication and routing, co-allocation of compute and data resources
- Global data distribution:
  - MC import from remote sites
  - Off-site analysis centers
  - Off-site reconstruction (D0)
Routing + Caching = Replication

Data Site
\[\text{WAN Data Flow}\]

Station Master
Mass Storage System

User

Igor Terekhov, FNAL
Now that the Data’s Distributed: JIM

- Grid Jobs and Information Management
- Owes to the D0 Grid funding – PPDG (the FNAL team), UK GridPP (Rod Walker, ICL)
- Very young – started 2001
- Actively explore, adopt, enhance, develop new Grid technologies
- Collaborate with the Condor team from The University of Wisconsin on Job management
- JIM with SAM is also called The SAMGrid

T<10min?

Igor Terekhov, FNAL
Job Management Strategies

• We distinguish grid-level (global) job scheduling (selection of a cluster to run) from local scheduling (distribution of the job within the cluster)

• We distinguish structured jobs from unstructured.
  - Structured jobs have their details known to Grid middleware.
  - Unstructured jobs are mapped as a whole onto a cluster

• In the first phase, we want reasonably intelligent scheduling and reliable execution of unstructured data-intensive jobs.
Job Management Highlights

• We seek to provide automated resource selection (brokering) at the global level with final scheduling done locally (environments like CDF CAF, Frank’s talk)

• Focus on data-intensive jobs:
  - Execution time is composed of:
    • Time to retrieve any missing input data
    • Time to process the data
    • Time to store output data
  - In the Leading Order, we rank sites by the amount of data cached at the site (minimize missing input data)
  - Scheduler is interfaced with the data handling system
Job Management – Distinct JIM

Features

• Decision making is based on both:
  - Information existing irrespective of jobs (resource description)
  - Functions of (jobs, resource)

• Decision making is interfaced with data handling middleware rather than individual SE’s or RC alone: this allows incorporation of DH considerations

• Decision making is entirely in the Condor framework (no own RB) – strong promotion of standards, interoperability
Condor Framework and Enhancements We Drove

• Initial Condor-G:
  - Personal Grid agent helping user run a job on a cluster of his/her choice

• JIM: True grid service for accepting and placing jobs from all users
  - Added MMS for Grid job brokering

• JIM: from 2-tier to 3-tier architecture
  - Decouple queing/spooling/scheduling machine from user machine
  - Security delegation, proper std* spooling, etc

• Will move into standard Condor

Igor Terekhov, FNAL
Condor Framework and Enhancements We Drove

- Classic Matchmaking service (MMS):
  - Clusters advertise their availability, jobs are matched with clusters
  - Cluster (Resource) description exists irrespective of jobs

- JIM: Ranking expressions contain functions that are evaluated at run-time
  - Helps rank a job by a function(job,resource)
  - Now: query participating sites for data cached. Future: estimates when data for the job can arrive etc
  - Feature now in standard Condor-G
Monitoring Highlights

• Sites (resources) and jobs
• Distributed knowledge about jobs etc
• Incremental knowledge building
• GMA for current state inquiries, Logging for recent history studies
• All Web based
Information Management – Implementation and Technology Choices

- XML for representation of site configuration and (almost) all other information
- Xquery and XSLT for information processing
- Xindice and other native XML databases for database semantics
JIM Project Status

• Delivered prototype for D0, Oct 10, 2002:
  - Remote job submission
  - Brokering based on data cached
  - Web-based monitoring
• SC-2002 demo – 11 sites (D0, CDF), big success
• April 2003 – production deployment of V1 (Grid analysis in production a reality as of April, 1)
• Post V1 – OGSA, Web services, logging service

Igor Terekhov, FNAL
Grid Data Handling

• We define GDH as a middleware service which:
  - Brokers storage requests
  - Maintains economical knowledge about costs of access to different SE’s
  - Replicates data as needed (not only as driven by admins)

• Generalizes or replaces some of the services of the Data Management part of SAM
Grid Data Handling, Initial Thoughts
The Necessary (Almost) Final Slide

• Run II experiments’ computing is highly distributed, Grid trend is very relevant
• The JIM (Jobs and Information Management) part of the SAMGrid addresses the needs for global and grid computing at Run II
• We use Condor and Globus middleware to schedule jobs globally (based on data), and provide Web-based monitoring
• Demo available – see me or Gabriele
Acks

• PPDG project, its management, for making it possible
• GridPP project in the UK, for its funding
• Jae Yu and others of UTA-Texas, FNAL CD mgmt for continuing support for student internship programs
• Other members of the Condor team for fruitful discussions
Backup Slides
Information Management

• In JIM’s view, this includes both:
  - resource description for job brokering
  - Infrastructure for monitoring (core project area)

• GT MDS is not sufficient:
  - Need (persistent) info representation that’s independent of LDIF or other such format
  - Need maximum flexibility in information structure
    - no fixed schema
  - Need configuration tools, push operation etc