Grids, Worldwide Computing, and ATLAS Software Development

ATLAS Software Workshop

Berkeley, California

12 May 2000

David Malon
malon@anl.gov

A worldwide computing model encompasses many things

- **Tier 0, Tier 1, Tier2, …, regional centers: number, sizing, scope, connectivity, roles**

which depend upon

- **What computing activities are expected to take place where (simulation, reconstruction, analysis, …), and what data are moved/replicated where**

which depend upon

- **What is POSSIBLE in the way of distributed computing (e.g., grid) and distributed data access**

National Computing Board has formed working groups (GRID, MONARC, REGIONAL CENTRES, WORLDWIDE COMPUTING MODEL) to address these issues.

What does the software development effort need to do to organize itself correspondingly?

Grids and the Computing Model

- ATLAS Computing Technical Proposal was written before the term “computational grid” was coined
- Grids are a technology--one component of a strategy used to implement an ATLAS computing model
- How do grids and other developments change the computing model?
  - Are these changes fundamental, or (merely) matters of implementation?
  - Are the implementation ramifications pervasive and extensive, or matters of detail?
  - Are the implications of grids equivalent, in terms of computing models, to the implications of better networking?

What is ATLAS saying?

- Sizing estimates presented to review panels are still largely “classical” (based upon initial reconstruction done entirely at CERN), although…
- one relatively early grid milestone involves distributed reconstruction
- one ATLAS grid workplan submission speaks of running a distributed trigger

This is okay in an R&D phase, but we need to move toward a coherent plan.

Timetable?

- **What do we need to learn from grid R&D, and on what timetable?**
- **Today, offline software development and grid projects are largely disjoint.**
- **When do these streams need to come together?**
  - **What does it mean for offline software, to be grid-aware, or grid-enabled, and what effort is required to make it so?**

We need to understand grids well enough to define an appropriate computing model in the time frame of the TDR, and to validate that model's viability in the time frame of the Mock Data Challenges.

Timetable

- Computing TDR is May-November 2002
- Mock Data Challenge 1 is “Start 2002” to provide input to Computing TDR.
  - “Comment from CERN review is that we need to strengthen this MDC to test computing model as much as possible.”
  - In latest response to CERN review panels, ATLAS has written “Test the world wide computing model (using GRID infrastructure and s/w)”
- …MDC2 (Start 2003) “Major test of Grid-enabled…”

MDC 1 and Grids

- **MDC 1 timing implies that enough grid software to allow ATLAS to elaborate an appropriate computing model for its TDR must be available BEFORE the start of 2002.**
  - **Implication is that grid software must be far enough advanced in 2001 to allow ATLAS to build implementations in 2001.**
  - **Grid experience to date suggests that making software grid-enabled takes a nontrivial amount of time and effort.**

This is why the ATLAS database work plan shows modest grid-related milestones (related to data replication and distribution) IN THE CURRENT YEAR.

Can we be EXPLICIT about what grid software (MUCH more than Globus) is required by ATLAS for MDC1, and on what timetable?

MDC 1 and Grid Timetables

How do ATLAS timetables compare to grid project timetables?
EU-Grid milestones show (from A. Putzer)

- Month 12 (December 2001) :Requirements studies, co-ordination with other WPs, interface with existing GRID services
- Month 18 (June 2002) : Run#0 : distributed MC production and reconstruction
- Month 24 (December 2002) : Run#1 : distributed analysis

Are these timetables in sync?

- EU-Grid Month 18 milestone may be
  - a bit too little (GRID as wide-area batch queue? Insufficient to evaluate grid's role in computing model?)
  - a bit too late,
- but
  - a more aggressive schedule may not be realistic given the state of grid R&D
  - ATLAS EU-Grid proposers have worked to keep the project milestones at least CLOSE to what ATLAS needs
- EU-Grid should not be singled out: other grid projects trigger the same timing concerns

Grids and ATLAS Software

What are the implications of a distributed computing model and grids for:

- The database domain?
  - **Extensive in almost any case**
- The control framework?
  - **Depends upon the model (e.g., distributed data sources versus distributing executables versus distributed execution)**
- Other ATLAS software infrastructure?

Answers are not entirely up to ATLAS

- Shared regional centers proposed in some countries will require common grid middleware
- LHC shared resources for Mock Data Challenges have already been proposed
- Grid projects are already LHC-wide or (in U.S.) involve ATLAS/CMS collaboration

Answers are not entirely up to high energy physics, either:

HEP does not control most grid R&D projects

What next?

How do we organize ourselves to connect

- **the ATLAS software development effort**

to

- **the NCB working group efforts**

to

- **the several grid initiatives in which ATLAS is involved?**

Can we be EXPLICIT about what grid software is required by ATLAS, on what timetable, and about how and when we will integrate grid middleware into ATLAS mainstream software development?

If ATLAS has a plan, we can have a strong voice in many of these grid projects.

The balance of this session

- **ATLAS participation in other grid projects**
  - **GriPhyN (Rob Gardner)**
  - **PPDG (Ed May)**
- **MONARC (Models Of Networked Architectures for Regional Centres) (Krzysztof Sliwa)**