



## **ATLAS, U.S. ATLAS, and Databases**

**David Malon**  
**Argonne National Laboratory**

DOE/NSF Review of U.S. LHC Software and Computing  
Fermi National Accelerator Laboratory  
27-30 November 2001

## **Outline**



- \* U.S. contributions since the May review
  - \* Near-term work: the Data Challenges
  - \* Issues and challenges
-

## U.S. contributions since the May review



- \* Joint responsibility (with Orsay) for global ATLAS database coordination  
(sole responsibility as of mid-October)
  - \* Database support for the Lund Program
  - \* Event store architecture proposal
  - \* ROOT conversion service
  - \* Database support for StoreGate (ongoing)
  - \* Package migration to CMT build/release tool (Orsay has principal  
responsibility for build/release infrastructure)
- 
- \* MySQL support for liquid argon test beam

## Database support for the Lund Program



- \* September 2001 ATLAS Physics Workshop in Lund, Sweden
    - See [David Quarrie's talk](#) for details of non-database work for Lund
  - \* Demonstration of persistence in two technologies
    - Objectivity/DB
    - ROOT
  - \* Demonstration of ability to switch storage technologies by changing  
only job options—no changes to user code
  - \* Persistence for output from generators (HepMC)
  - \* Persistence for fast simulation output (Atfast)
- 
- \* ~~User support and consulting~~

## Database support for the Lund Program



- \* All of these things were delivered successfully and on time, but at a cost—diversion of scarce core database developer resources
  - \* Seven new versions of transient Attfast between the time it was declared “ready” and the Lund Release 2.0.0! (then came the bug fixes, through 2.0.2)
  - \* This work remains a drain on core resources
    - HepMC is changing, Attfast is changing, ...;
    - Core group probably cannot decline to work further in this area, because Attfast is proposed to be integral to Data Challenge 0
- 

## Database architecture definition



- \* Database coordinators proposed last spring to prepare a database architecture document by the end of the summer
  - \* Architecture Review Committee strongly endorsed this plan
  - \* U.S.-led effort to forge an architectural proposal out of input from core developers and from other experiments
  - \* Event store architecture proposal presented at September 2001 Software Week
  - \* Broad collaboration input solicited at two meetings (one by phone, one during October ATLAS Week)
-

## Database architecture definition



- \* Proposed architecture is described in a storage-technology-independent way, and, thus far, makes few demands on underlying storage technology capabilities
  - \* Two-day meeting held with ROOT experts for architectural feedback, and to ensure that architecture has no unintended technology bias
  - \* Database architecture workshop held at CERN in October; consensus achieved in many areas, work remains in several others
  - \* Further iterations between now and December Software Week; review early in 2002
- 

## ROOT conversion service



- \* Based on work by ROOT experts for STAR, but independent of STAR code, and compliant with Gaudi/Athena conversion service approach
  - \* Essential element of technology comparisons planned for Phase II of Data Challenge 1
  - \* Work is in progress to connect this to ADL backend for automatic converter generation
-

## Database support for StoreGate



- \* Support for StoreGate DataLinks, and for “new” StoreGate backend
  - \* A few unresolved issues (e.g., StoreGate equivalent to Gaudi ItemList—mechanism for specifying which event data objects to save)
  - \* Delivered in most recent ATLAS release (2.4.0)
  - \* Package migrations to new infrastructure begin to appear in Release 2.4.1
- 

## Other database-related work



- \* Package migration to CMT build/release tool, to new version of Objectivity, to new compiler, ...
    - Orsay has principal responsibility for build infrastructure, but not for migration of specific packages
  - \* MySQL support for liquid argon test beam
  - \* Prototyping for grid-enabled data access in Athena
    - Cf. Malon et al, “*Grid-enabled data access in the ATLAS Athena framework*,” Proceedings of CHEP’01, Beijing, September 2001.
    - Expect that grid work will be described elsewhere
-

## Bookkeeping/cataloging for data challenges



- ✳ Database group sponsored a half-day bookkeeping/cataloging organizational meeting during ATLAS Week—well-attended (28 people)
  - ✳ Subgroups organized to address naming, run metadata, replica management
  - ✳ Good reasons for optimism about concrete coordination and complementarity among many groups (EU DataGrid WP2, BNL/PPDG Magda, Grenoble MySQL metadata, core database, ...)
  - ✳ Successful followup at 14 November Data Challenge meeting
- 

## Data Challenge 0



- ✳ “Continuity test”—simulation→reconstruction→analysis; 100,000 events
  - ✳ Based on Geant3 simulation (“new” geometry), Athena-based reconstruction, limited analysis
  - ✳ Database group asked to support two additional paths
    - Physics TDR data→Athena-based analysis
    - Generator→Atlfast→analysis
  - ✳ Proposal for database work in support of DC0 sent to Data Challenge Coordinator on 18 October
-

## Data Challenge 1



- ✧ Nominally, February-July 2002, in two phases
  - ✧ Phase I is primarily event production for High-Level Trigger  
TDR
  - ✧ Phase II intended to exercise ATLAS software at terabyte  
scales, grid-enabled to some extent
  - ✧ Database plan for Phase II includes architecture evaluation,  
and datastore technology comparisons
- 

## Near-term issues



- ✧ Expect good progress on ROOT/MySQL infrastructure
  - ✧ Expect to be able to provide StoreGate and Data Challenge  
0 support
  - ✧ In question: sufficient progress on database architecture,  
and development and deployment of substantial database  
infrastructure for Data Challenge 1
-

## LHC-wide database collaborations



- \* Data management Technical Assessment Group identified as high priority by recent LHCC review
  - \* CMS is now beginning a serious effort to define and develop a hybrid “object streaming (ROOT) + relational database” approach jointly with the ROOT team
  - \* This is proposed to be an LHC common project
    - IT/DB and LHCb are participating
    - ATLAS sent an observer; I will join meetings next week
- 

## LHC-wide database collaborations-II



- \* ATLAS is disadvantaged here by lack of CERN database presence
  - \* ATLAS has experience to offer here (cf. BNL work)
  - \* We are fully committed to contributing to and taking advantage of this common project in our own hybrid technology deployment
-

## Near-term staffing concerns



- ✳ **Loss of Ed Frank seriously jeopardizes database architecture work**
    - ❑ Significant loss of expertise—probably the only person involved in ATLAS databases to whom database coordinators could have delegated the task of forging a coherent architecture proposal out of the many, many hours of intense, wide-ranging, conflicting discussions
    - ❑ Apparently, no funds to support replacement effort
  - ✳ **Likely decrease in ANL core database effort from a rate of 3.5 FTEs to something closer to 3.0**
    - ❑ 0.5 in plan comes from institutional support; had been hoping to use base funds, but will have trouble doing this within the current budget plan, and 0.2 FTEs must complement 0.8 FTEs of PPDG funding to support full-time grid hire
    - ❑ Still hoping budget will be better; meanwhile, looking around Lab for other sources of funds
- 

## Near-term staffing concerns - II



- ✳ **Internationally, loss of Helge Meinhard, who had been acting as database contact to CERN IT Division**
-

## Observations



- \* Much ATLAS database development has tended to be dilute and lateral (multiple converters in evolving versions, multiple storage technologies...), rather than concentrated and forward (new components and new capabilities, consonant with a more fully articulated architecture...)
  - \* Database coordinators have resisted this consistently, to no avail
- 

## Organization



- \* Shared database coordination responsibility until mid-October
    - David Malon (Argonne) and RD Schaffer (Orsay)
    - Now a U.S. responsibility
  - \* Database task leaders from each subsystem
    - Inner Detector: Stan Bentvelsen→David Calvet→??
    - Liquid Argon: Stefan Simion (Nevis)→Randy Sobie
    - Muon Spectrometer: Steve Goldfarb (Michigan)
    - Tile Calorimeter: Tom LeCompte (Argonne)
    - Trigger/DAQ: Antonio Amorim, Francois Touchard
- Organization has not proven to be a source of significant database development effort.*
-

## International staffing profile



- \* CERN: 1 FTE conditions databases (Goossens), new effort (Smirnov) to provide database platform support and IT liaison
- \* Orsay (~2.5 FTEs?): build/release infrastructure, access to TDR data (we need component development from Orsay)
- \* ANL (3.0+ FTEs): database coordination, software infrastructure and support for data challenges, architecture-related work, support for Objectivity conversion services, near-term support for HepMC and Attfast
- \* BNL (1.3 FTEs): ROOT and MySQL support and development

## Resource Requirements



- \* WBS estimates (see document from Torre Wenaus) show 14.6 FTEs in 2001, 15.8 in 2002, with a maximum of 18.5 in 2005
-

## Challenges for 2002



- \* Define and deploy significant components of a database architecture for Data Challenge 1
- \* Design an experiment to inform a database technology choice to be made in 2002
  - [Deploy in Phase II of Data Challenge 1](#)
- \* Find a way to participate actively in LHC-wide common database projects
  - [NSF-funded position at CERN would have helped us here](#)
- \* Transition from institutional responsibility for specific technologies to  
institutional responsibility for specific components