

Why ATLAS needs MySQL



- ❄ For software developed by the ATLAS offline group, policy is to avoid dependencies upon specific database products, but **ATLAS technology choices must support distributed deployment NOW**
- ❄ ATLAS uses a variety of relationally-hosted databases that are not today supported in Oracle
- ❄ ATLAS leverages externally developed software—we do not reinvent everything ourselves
 - ❑ Example: Chimera/Pegasus from GriPhyN uses MySQL or PostgreSQL
 - ❑ Source code is not under ATLAS control

MySQL and detector description



- ❄ NOVA, used for “primary numbers” that parameterize detector description, was developed under external auspices
 - ❑ Uses MySQL
 - ❑ Though some NOVA developers are now associated with ATLAS, they are currently engaged in other tasks
- ❄ Even our nightly software builds rely on MySQL: we generate detector description classes from database-resident data
 - ❑ Stable MySQL services are a must

MySQL for conditions and calibration data



- ❄ ATLAS offline community employs conditions database software developed by the Lisbon TDAQ group
 - Current implementation is in MySQL
- ❄ Slow start of the LCG conditions database common project means that the Oracle implementation does not yet have sufficient functionality
- ❄ These MySQL-hosted databases are needed NOW, for this year's combined test beam

MySQL event collections and POOL



- ❄ This year's ATLAS Data Challenge will make extensive use of POOL collections, the only current relational implementations of which are in MySQL
- ❄ LCG has consistently endorsed this line of development (following SC2's approval of persistence RTAG report, Applications Area planning, Architects Forum decisions, ...)
- ❄ LCG should take all reasonable measures to ensure successful adoption and deployment of such LCG-sponsored software, and should not tolerate tactics that impede this
 - ❑ Unthinkable that ATLAS should be able to use POOL everywhere but at CERN

Distributed Deployment



- ❄ TODAY ATLAS physicists can and do run simulations on their laptops, disconnected from networks, by using a MySQL embedded server for access to detector description
 - ❑ We don't know how to do this with Oracle
- ❄ In Data Challenge I, ATLAS successfully employed an extremely heterogeneous array of grids, with access to MySQL-resident data
 - ❑ Access to database-resident data (e.g., the liquid argon calorimeter noise database) from compute elements behind firewalls was accomplished by a variety of creative means (proxies, ...), including sometimes installing and launching a MySQL server as part of the submitted job
 - ⌘ We don't know how to do this with Oracle

Distributed deployment



- ❄ **ATLAS has replicated MySQL databases at several sites**
 - ❑ The sites at which we replicate data do not typically support Oracle, and we do not currently have the tools or know-how to do this with Oracle in any case
 - ❑ Our replication/extraction machinery is not very sophisticated, but we already need MySQL→MySQL tools for our external sites, and adding Oracle to the mix adds work for which we have insufficient available effort and expertise

- ❄ **Some of the things that we cannot today do with Oracle may be doable in principle, but until they are doable in practice, introducing Oracle adds more problems than we can solve this year**
 - ❑ ...and the problems are not ATLAS-specific in any case
 - ❑ Oracle experts and proponents must demonstrate distributed deployment solutions before asking ATLAS to move to Oracle, with sufficient lead time for a transition

A solution that works only at CERN is no solution at all



- ❄ Event collections, conditions and calibration databases, geometry databases—indeed, most offline databases—will all be replicated many times
 - Technology choices must reflect and support this TODAY
- ❄ There are a handful of exceptions—singleton databases—for these, ATLAS is already trying to use Oracle
- ❄ Many ATLAS people around the globe have MySQL experience, and even expertise
 - Very little equivalent Oracle expertise in ATLAS

Additional Obstacles



- ❄ ATLAS developers were surprised to learn that Oracle, as supported by CERN IT, actually has fewer administrative and support tools available than does MySQL
 - Apparently, nothing equivalent even to phpMyAdmin
- ❄ Smaller but nonetheless annoying technical issues, like lack of gcc-3.2 C++ bindings

ATLAS needs MySQL in 2004



- ❄ This year, there is no question that ATLAS requires MySQL servers and services for 2004 combined test beam and data challenges
 - ❑ Without them, combined test beam and tests of the ATLAS computing model will fail
 - ❑ Our current modest request to CERN-based resource providers is simply for machines on which MySQL servers can be run, not for MySQL services (details in accompanying document)
- ❄ ATLAS is already worried about meeting data challenge software development milestones because of a well-documented lack of human resources (cf. September 2003 LHCC “manpower” review)
 - ❑ It would be irresponsible to divert effort away from delivering required functionality and into porting working code to other technologies at this point in time
 - ❑ Support for two technologies, and cross-technology extraction/replication tools, would require additional effort that we lack in any case
- ❄ For most offline databases, it is too late to add technologies to the mix in 2004
 - ❑ Insufficient lead time, even if proposed alternative infrastructure were in place, which it clearly is not
 - ❑ Recall that ATLAS has been asking for machines on which to run MySQL servers since at least 2002

Moving forward



- ❄ LCG and CERN should not impede, implicitly or explicitly, successful adoption of LCG products like POOL collections, and should certainly not tolerate stances that put data challenges at risk
- ❄ ATLAS would like to see a distributed Oracle deployment model, with the infrastructure and tools and people and commitment to support it
 - ❑ Persistence RTAG report, endorsed by SC2, recommended specific steps in this direction. What steps have been taken?
- ❄ A **SUPPORTED** heterogeneous deployment infrastructure would be particularly interesting to ATLAS
 - ❑ Oracle at CERN (or CERN plus selected Tier 1s), MySQL elsewhere?

Oracle, MySQL, and LCG



- ❄ Heterogeneous (e.g., Oracle+MySQL) deployment is not an ATLAS-specific problem
 - ❑ Ditto for distributed deployment of an all-Oracle solution
- ❄ ATLAS does not have the resources to add these tasks to its workload, nor does it make sense to have every experiment solve these problems separately
- ❄ ATLAS eagerly awaits Oracle/heterogeneous distributed deployment solutions and documentation and training
- ❄ With sufficient lead time, ATLAS will strive to take advantage of such solutions when they become available
 - ❑ Are such solutions proposed to be available this year, in time for adoption in 2005?
- ❄ Until such solutions are available, asking “Why don’t you use Oracle?” is not constructive
- ❄ In the mean time, ATLAS will:
 - ❑ Continue to avoid building technology dependencies into new code
 - ❑ Continue to evaluate Oracle in the context of singleton (CERN-only) databases
 - ❑ (Gladly!) participate in an organized Oracle10g evaluation, at a level that does not jeopardize this year’s data challenge and combined test beam