



Magda Distributed Data Manager

Torre Wenaus

BNL

October 2001



ATLAS PPDG Program

- ✧ Principal ATLAS Particle Physics Data Grid deliverables:
 - ❑ **Year 1: Production distributed data service deployed to users.** Will exist between CERN, BNL, and at least four US grid testbed sites (ANL, LBNL, Boston U, Indiana, Michigan, Oklahoma, Arlington)
 - ❑ **Year 2:** Production distributed **job management service**
 - ❑ **Year 3:** Create **'transparent' distributed processing** capability integrating distributed services into ATLAS software
- ✧ **Magda is focused on the principal PPDG year 1 deliverable.**
- ✧ Draws on grid middleware development while delivering immediately useful capability to ATLAS
 - ❑ This area – looking beyond data storage to the larger issue of data management – has not received much attention in ATLAS up to now
 - ❑ Now changing with the DCs approaching, and Magda is intended to help

Magda Background

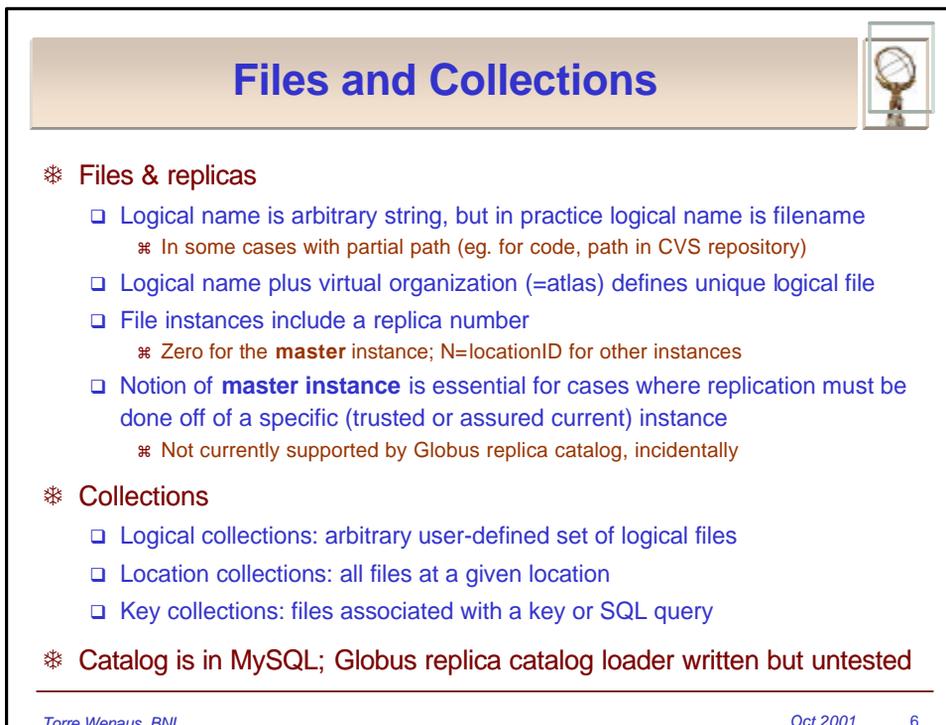
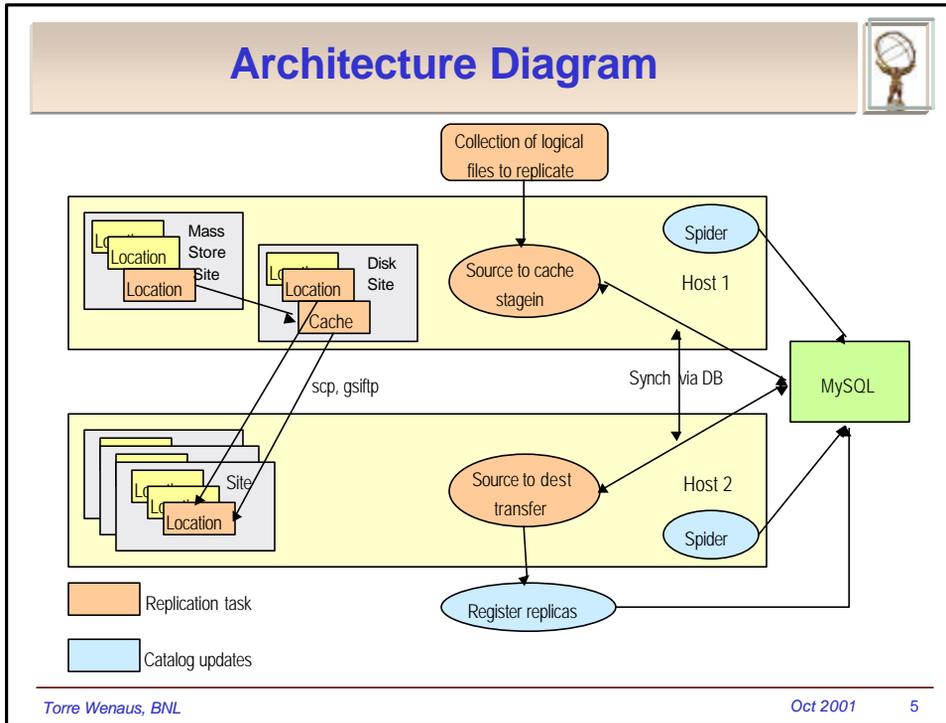


- ✳ **Initiated (as 'DBYA') in 3/01 for rapid prototyping of distributed data management. Approach:**
 - A flexible infrastructure allowing quick development of components to support users quickly,
 - with components later substituted easily by external (eg grid toolkit) tools for evaluation and long term use
- ✳ **Stable operation cataloging ATLAS files since 5/01**
- ✳ **Replication incorporated 7/01**
- ✳ **Deployed at CERN, BNL, ANL, LBNL**
- ⇒ Developers are currently T. Wenaus, W. Deng (BNL)
 - Info:** <http://www.usatlas.bnl.gov/magda/info>
 - The system:** <http://www.usatlas.bnl.gov/magda/dyShowMain.pl>

Architecture & Schema



- ✳ MySQL database at the core of the system
- ✳ DB interaction via perl, C++, java, cgi (perl) scripts
 - C++ and Java APIs autogenerated off the MySQL DB schema
- ✳ User interaction via web interface and command line
- ✳ Principal components:
 - **File catalog** covering arbitrary range of file types
 - **Data repositories** organized into *sites* and *locations*
 - Computers with repository access: a *host* can access a set of *sites*
 - Logical files can optionally be organized into **collections**
 - **Replication, file access** operations organized into *tasks*
- ✳ To serve environments from production (DCs) to personal (laptops)



Distributed Catalog



- ✳ Catalog of ATLAS data at CERN, BNL (plus ANL, LBNL recently added)
 - Supported data stores: CERN Castor, CERN stage, BNL HPSS (rftp service), AFS and NFS disk, code repositories, web sites
 - Current content: TDR data, test beam data, ntuples, code, ATLAS and US ATLAS web content, ...
 - About 150k files currently cataloged representing >2TB data
 - ✳ Has run without problems with ~1.5M files cataloged
- ✳ 'Spiders' crawl data stores to populate and validate catalogs
 - Catalog entries can also be added or modified directly
- ✳ 'MySQL accelerator' provides good catalog loading performance between CERN and BNL; 2k files cataloged in <1sec (W. Deng)

Data (distinct from file) Metadata



- ✳ Keys
 - User-defined attributes (strings) associated with a logical file
 - Used to tag physics channels, for example
- ✳ Logical file versions
 - Version string associated with logical file distinguishes updated versions
 - Eg. for source code, version is the CVS version number of the file
- ✳ 'Application metadata' (run/event # etc.) not included
 - Separate (Grenoble) catalog to be interfaced via logical file
- ✳ To come: Integration as metadata layer into 'hybrid' (ROOT+RDBMS) implementation of ATLAS DB architecture
- ✳ To come: Data signature ('object histories'), object cataloging
 - Longer term R&D

File Replication



- ✳ Supports multiple replication tools as needed and available
- ✳ Automated CERN-BNL replication incorporated 7/01
 - CERN stage ⇒ cache ⇒ **scp** ⇒ cache ⇒ BNL HPSS
 - *stagein, transfer, archive* scripts coordinated via database
 - Transfers user-defined collections keyed by (e.g.) physics channel
- ✳ Recently extended to US ATLAS testbed using Globus **gsiftp**
 - Currently supported testbed sites are ANL, LBNL, Boston U
 - BNL HPSS ⇔ cache ⇔ **gsiftp** ⇔ testbed disk
 - BNL or testbed disk ⇔ **gsiftp** ⇔ testbed disk
 - **gsiftp** not usable to CERN; no grid link until CA issues resolved
- ✳ Plan to try other data movers as well
 - GDMP (flat file version), bbcp (BaBar), ...

Data Access and Production Support



- ✳ Command line tools usable in production jobs
 - **getfile** – under test
 - ✳ Retrieve file via catalog lookup and (as necessary) staging or (still to come) remote replication
 - ✳ Local soft link to cataloged file instance in a cache or location
 - ✳ Usage count maintained in catalog to manage deletion
 - **releasefile** – under development
 - ✳ Removes local soft link, decrements usage count in catalog, deletes instance (optionally) if usage count goes to zero
 - **putfile** – under development
 - ✳ Archive output files (eg. in Castor or HPSS) and register them in catalog
- ✳ Adaptation to support simu production environment
 - Working with Pavel on simu production scenario
- ✳ Callable APIs for catalog usage and update to come
 - Collaboration with David Malon on Athena integration

Near Term Schedule



- * Deploy data access/registration command line tools
- * Create prototype (simu) production scenario using Magda
- * Incorporate into Saul Youssef's pacman package manager
- * Interface to other DC tools (Grenoble etc.)
- * Apply to cataloging/replication/user access of DC0 data
- * Adaptation/implementation for DC1
- * Integration into hybrid DB architecture implementation
- * Evaluation/use of other grid tools: Globus RC, GDMP, ...
- * Athena integration