

Grid Metadata with AMGA

B. Koblitz, CERN

with N. Santos and V.Pose

GridPP Metadata Workshop, Oxford

July 3rd, 2006

- **AMGA (ARDA Metadata Grid Application) is the Metadata catalogue of gLite 1.5**
- **AMGA started out as ARDA's tool to investigate metadata access on the GRID**
- **Metadata is relationally structured data for grid jobs (lives normally in databases)**
- **AMGA works in 2 modes:**
 - Side-by-Side a File Catalogue (LFC): File Metadata
 - Standalone: General relational data on Grid
- **AMGA has 2 front ends:**
 - SOAP with PTF standardised interface
 - Text-based TCP streaming protocol (proprietary, documented)
- <http://project-arda-dev.web.cern.ch/project-arda-dev/metadata/>

- **AMGA implements a common interface designed in close collaboration of gLite-DM and ARDA teams**
(P. Kunszt, R. Rocha, N. Santos, B.K.)
contains man ideas from UK GridPP Metadata group,
LHCb (Bookkeeping, GANGA), HEP...
- **EGEE standard for metadata access (AMI, FiReMan)**
- **Design Ideas:**
 - Versatile: Performance (HEP), Security (BioMed)
 - Provides simple relational data manipulation interface
 - Modular: Interface for Entry manipulation, schemes, security
 - Few requirements on back end, can be SQL-DB, XML...
 - Allows stateless & statefull implementations
- **Description of WSDL at**
<https://edms.cern.ch/document/573725>

- **Schema (directory)**
 - Has hierarchical name and list of attributes /prod/events
- **Attributes**
 - Have name and storage type
 - Interface handles all types as strings
- **Entry**
 - Live in a schema, assign values to attributes
- **Query**
 - SELECT ... WHERE ... clause in SQL-like query language

Examples

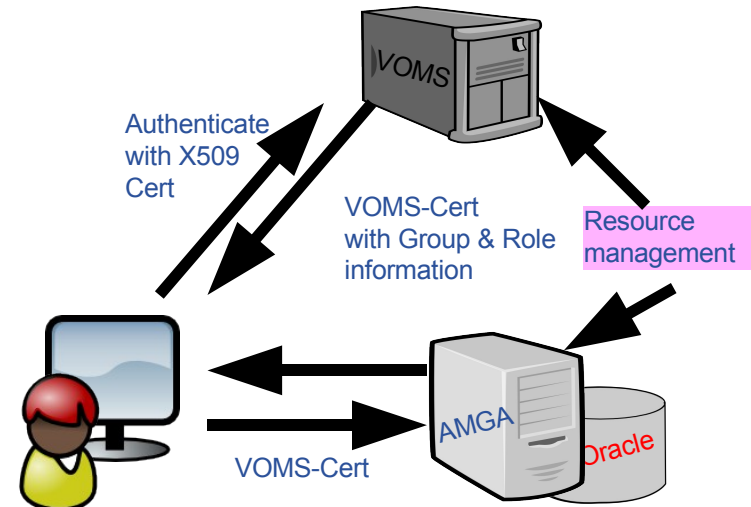
```
mkdir /jobs
addattr /jobs jobStatus int
addentry /jobs/job1 jobStatus 0
updateattr /jobs jobStatus 1 jobID>100
selectattr /DLibrary:FileName /DLAudio:Author /DLAudio:Album
'/DLibrary:FILE=/DLAudio:FILE and like(/DLibrary:FileName, "%.mp3")'
```

- Security very important for BioMed, not for HEP

Security ↔ Speed

- AMGA supports:

- SSL connections (Optional!)
- Authentication based on Password, X509 Cert, Grid Proxy
- Posix-ACLs and Unix permissions for entries and collections
- Built-in group-management like AFS or via VOMS

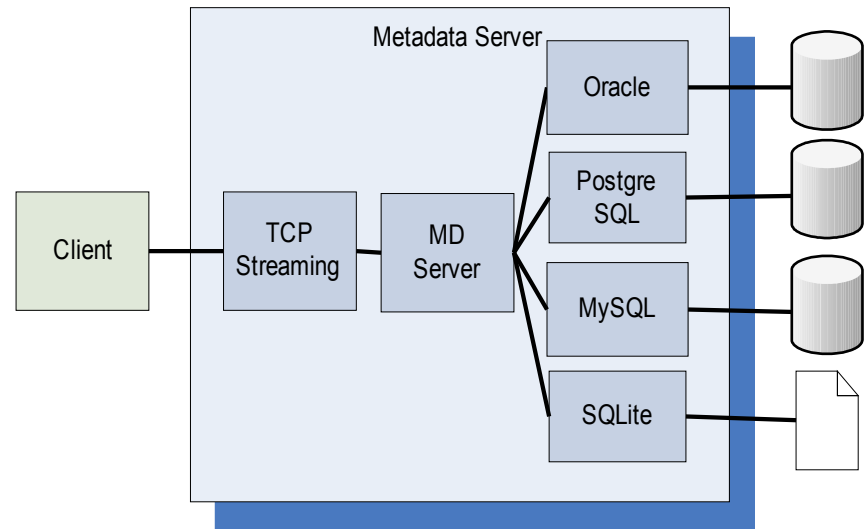


- Currently no security on attribute basis

- AMGA allows to create views: Safer, faster, similar to RDBMS

Security tested by GILDA and BioMed teams: Very positive feedback for built-in group management & ACLs

- **AMGA Implementation:**
 - SOAP and Text frontends
 - Streamed Bulk Operations
 - Supports single calls, sessions & connections
 - SSL security with grid certs (negotiated by client)
 - Own User & Group management + VOMS
 - PostgreSQL, Oracle, MySQL, SQLite backends
 - Works alongside LFC
 - C++, Java, Perl, Python clients



- ARDA metadata server **AMGA** in gLite 1.5, back in 3.1?
- Implements **EGEE Metadata Interface**
- **AMGA** currently in preproduction for
LHCb(GANGA, bookkeeping), MDM, DigitalLibrary, UnoSat...
- **Replication features available in AMGA**
- **Current work:**
 - Replication of AMGA (implemented, currently being tested)
 - User management in distributed, replicate catalogues
- **AMGA Web Site**
<http://project-arda-dev.web.cern.ch/project-arda-dev/metadata/>