Technical Aspects of AMGA

KISTI Supercomputing Centre
Soonwook Hwang
2010. 06. 17

Index

Overview of AMGA (1/2)

- **Metadata is data about data**
- **AMGA provides:**
  - Access to Metadata for files stored on the Grid
  - A simplified general access to relational data stored in database systems.

- **2004 – the ARDA project evaluated existing Metadata Services from HEP experiments**
  - AMI (ATLAS), RefDB (CMS), Alien Metadata Catalogue (ALICE)
  - Similar goals, similar concepts
  - Each designed for a particular application domain
    - Reuse outside intended domain difficult
  - Several technical limitations: large answers, scalability, speed, lack of flexibility

- **ARDA proposed an interface for Metadata access on the GRID**
  - Based on requirements of LHC experiments
  - But generic - not bound to a particular application domain
  - Designed jointly with the gLite/EGEE team

Overview of AMGA (2/2)

What is AMGA ? (ARDA Metadata Grid Application)

- **Began as prototype to evaluate the Metadata Interface**
  - Evaluated by community since the beginning:
  - Matured quickly thanks to users feedback
- **Now part of gLite middleware**
- **Requirements from HEP community**
  - Millions of files, 6000+ users, 200+ computing centres
  - Mainly (real-only) file metadata
  - **Main concerns : scalability, performance, fault-tolerance, Support for Hierarchical Collection**
- **Requirements from Biomed community**
  - Smaller scale than HEP
  - **Main concerns : Security**

ARDA Project (A Realisation of Distributed Analysis for LHC)

Metadata User Requirements

- **I want to**
  - store some metadata information about files
    - In a structured way
  - query a system about those information
  - keep information about jobs running on the Grid
    - I want my jobs to have read/write access to those information using the grid proxy certificate
  - NOT use a database

Metadata Concepts in AMGA

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

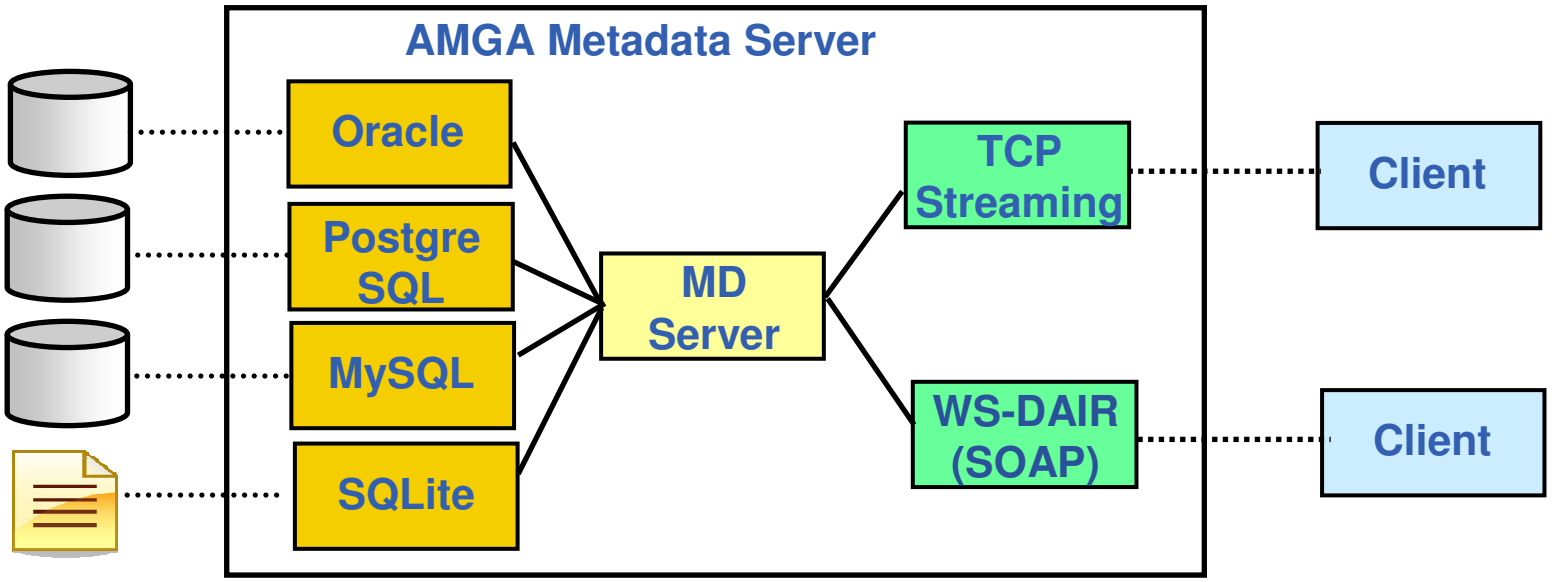
Examples

```
createdir /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")'
```

Main Features

Main Feature of AMGA

- **Modular back-end :** Oracle, PostgreSQL, MySQL, SQLite
- **Modular front-end :** TCP Streaming, WS-DAIR (SOAP)

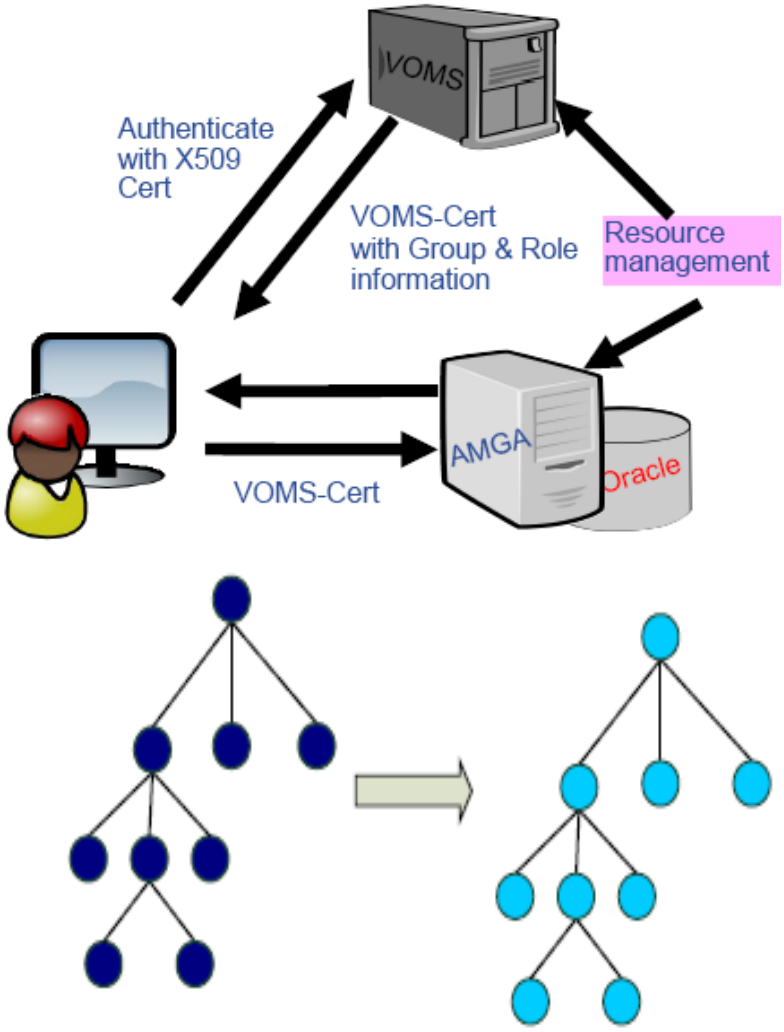


- **Streamed Bulk Operations**
- **Import existing databases**
- **Native SQL Query & AMGA Language Query**

Main Features

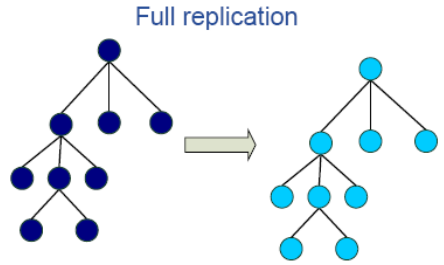
Main Feature of AMGA

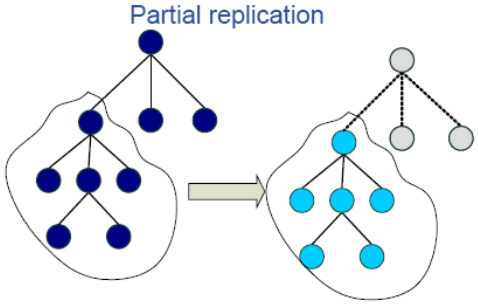
- **Integration with Grid Security :** Grid proxy authentication and VOMS authorization
- **Secure client connection using SSL**
- **Authorization using ACLS** with support for user and group management
- **Replication –** Metadata collections can be replicated to improve reliability, scalability and performance, considering security issues

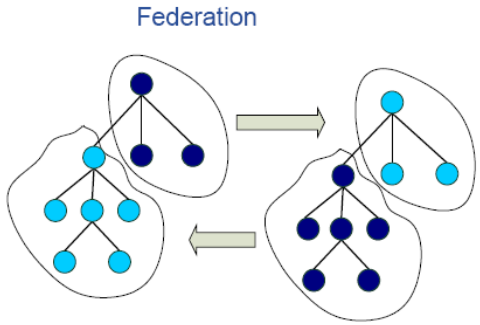


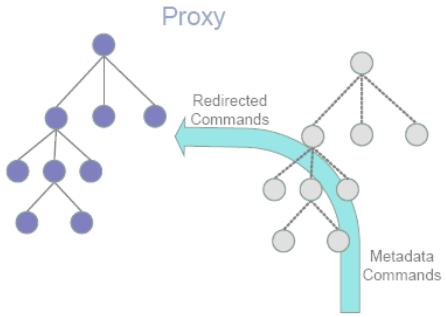
Replication Feature of AMGA for Belle-II

- **Replication for Performance, Reliability and Scalability**
  - Master-Slave & Asynchronous communication model
  - Full & Partial replications allowed



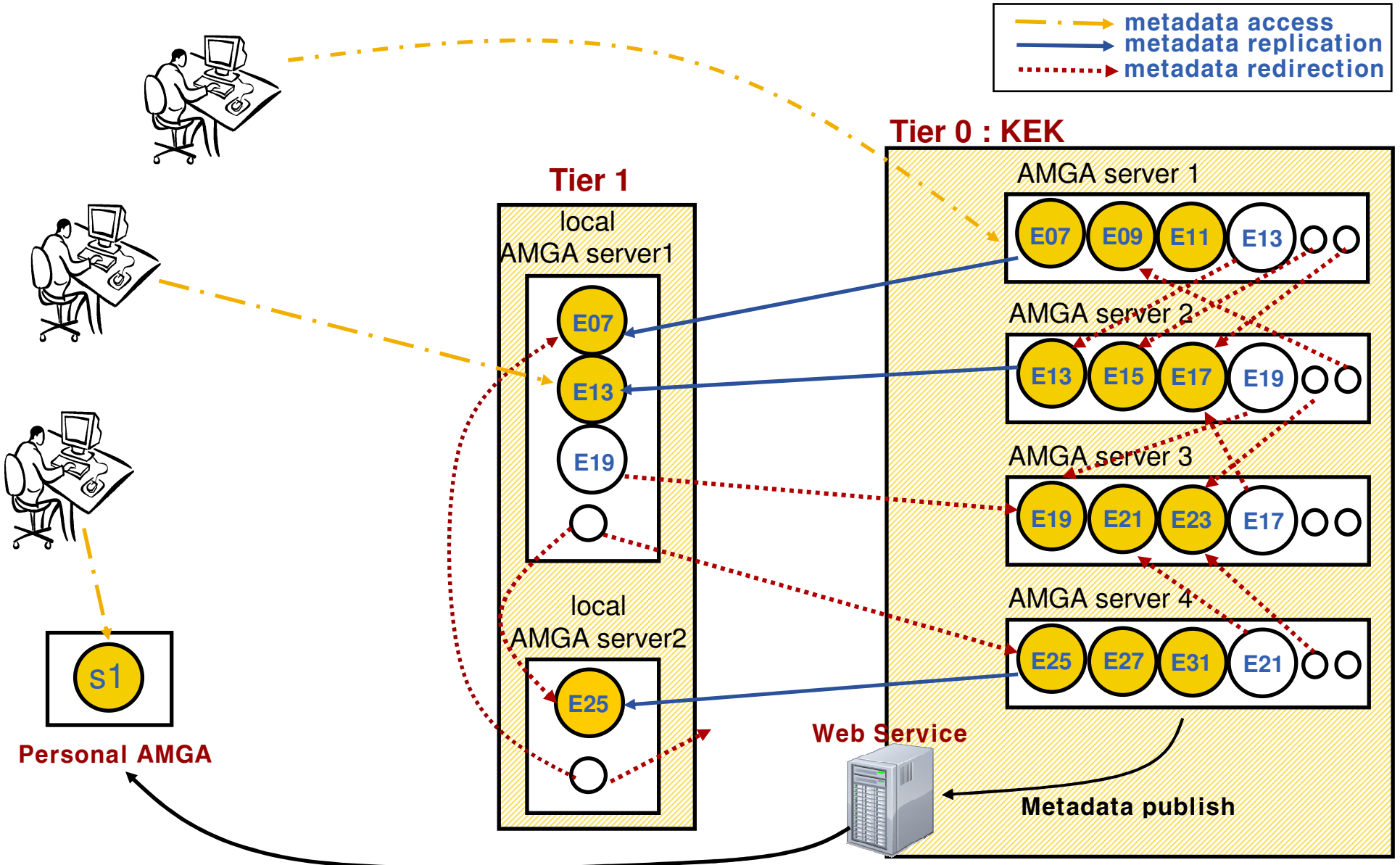






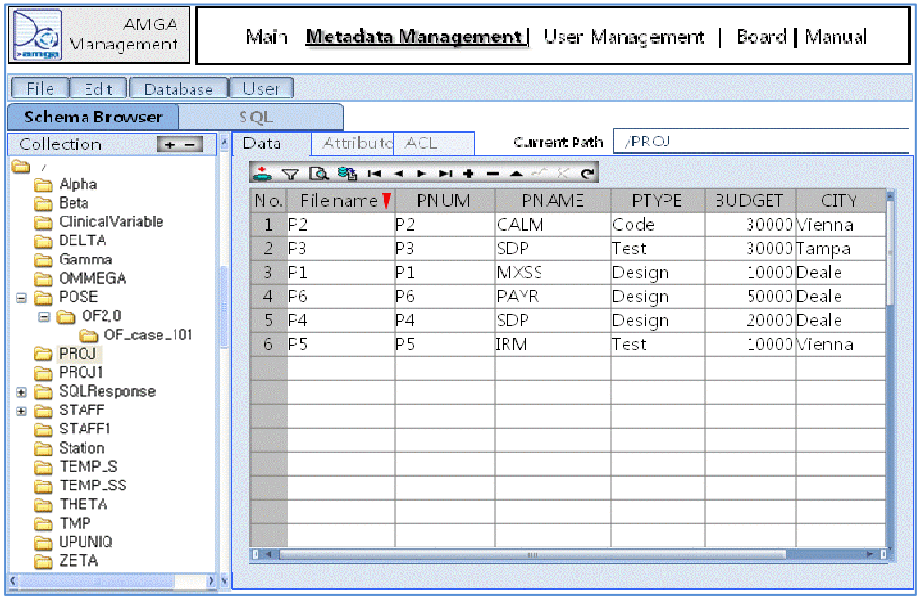
- **AMGA Testbed for Belle II**
  - Master at KISTI : 150.183.246.196
  - Slave at Melbourne : 192.231.127.47
  - File-level Metadata retrieved from Belle files

II. Technical Aspects of AMGA for Belle-II



III. Ongoing Works

- **AMGA 2.1 Release (3rd quarter 2010)**
  - Redirection & Federation
  - Easy to use GUI tool



- **Belle II metadata Scalability Tests**
  - Emulate 60 times as many data as Belle and evaluate performance

Thank you !!

hwang@kisti.re.kr