AMGA Training in EGI TF 2011

- **Place : St Clair 4, Lyon Conference Centre**
- **Date: 2011.09.19, Monday**
- **Presenters**
  - SoonwooK Hwang, Taesang Huh, Geunchul Park, KISTI
- **Timetable**
  - Session 1 (14:00 ~ 15:30)
    - Overview of AMGA
    - Hands-on Exercise on AMGA CLI client tool
  - Coffee Break (15:30 ~ 16:00)
  - Session 2 (16:00 ~ 17:30)
    - Overview of AMGA GUI client called AMGA Manager
    - Hands-on Exercise on AMGA Manager

Overview of AMGA
(the EMI Metadata Software Product)

KISTI Supercomputing Center

Soonwook Hwang

2011. 09. 19

KiSTi
www.kisti.re.kr

EUROPEAN MIDDLEWARE INITIATIVE

Contents

- A Brief Overview of AMGA
- Main Features
- AMGA Client Tools
- AMGA Usage in Applications
- References

A Brief Overview of AMGA

- **The EMI metadata software product**
- **KISTI is responsible for the maintenance and evolution of AMGA in EMI**
- **EMI service to handle metadata on the Grid**
  - Access to Metadata for files distributed on the Grid
  - A simplified general access to relational data stored in database systems.

History of AMGA (1/2)

- **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

History 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 and EMI distribution**
- **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**

Metadata User Requirements

- **I want to**
  - store some metadata information, e.g., 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

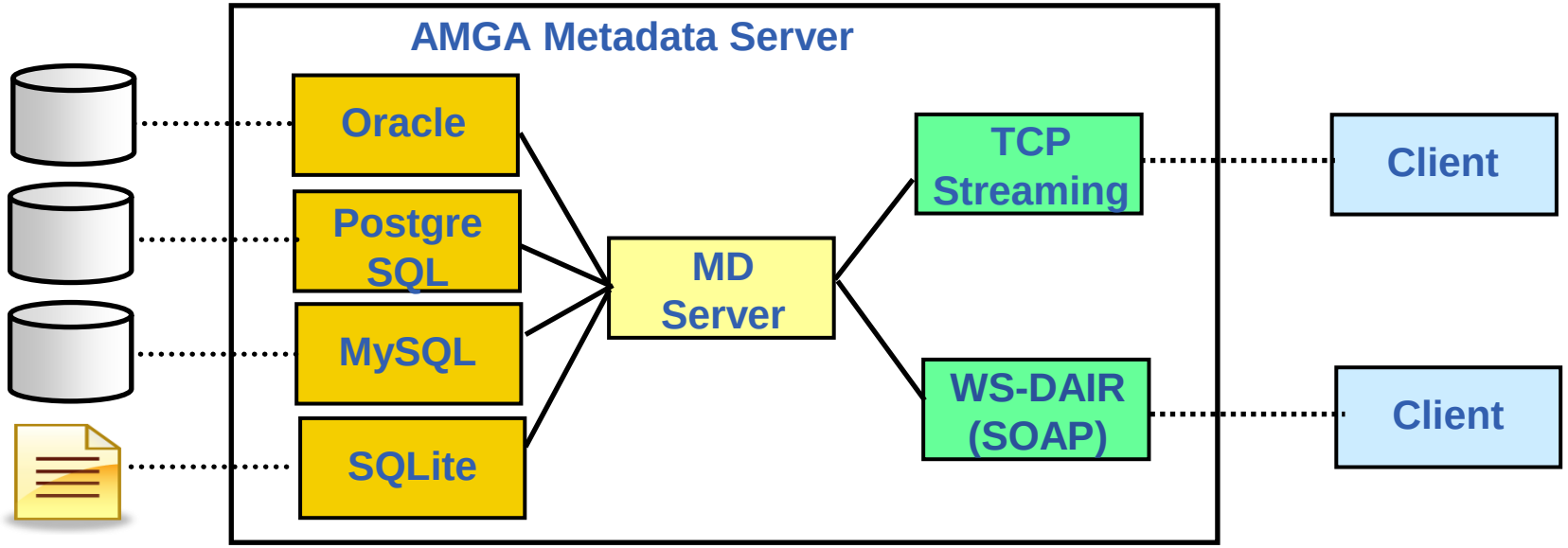
- **Entries**
  - Representation of real world entities, which we are attaching metadata to describe them
- **Attributes**
  - Type : Int, float, string
  - Name/Key : the name of the attribute
  - Value : Value of an entry's attributes
- **Schema**
  - **a set of attributes**
- **Collections**
  - **A set of entries associated with schema**
  - **Can be organized hierarchically containing sub collections**
- **Query**
  - **SELECT ... WHERE ... clause in SQL-like or SQL query language**

AMGA Analogies to RDBMS & File System

- **Analogy to the RDBMS**
  - Collection ⇔ table
  - Schema ⇔ table schema
  - Attribute ⇔ schema column
  - Entry ⇔ table row/record
- **Analogy to file system**
  - Collection ⇔ Directory
  - Entry ⇔ File

Main Features

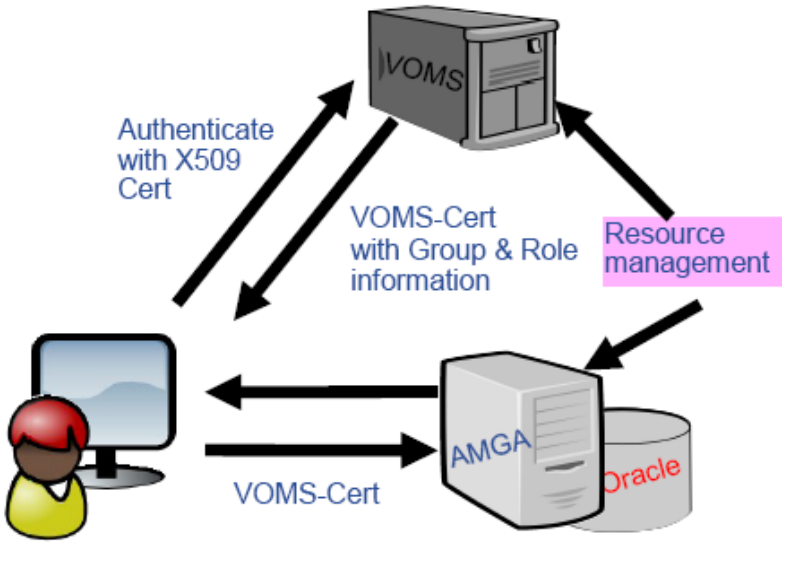
- **AMGA server implemented as a C++ multiprocess server**
  - back-end : Oracle, PostgreSQL, MySQL, SQLite
  - front-end : TCP Streaming, WS-DAIR (SOAP)



- **Streamed Bulk Operations**
- **Import existing databases**
- **Support for both AMGA SQL-like and Native SQL Query**

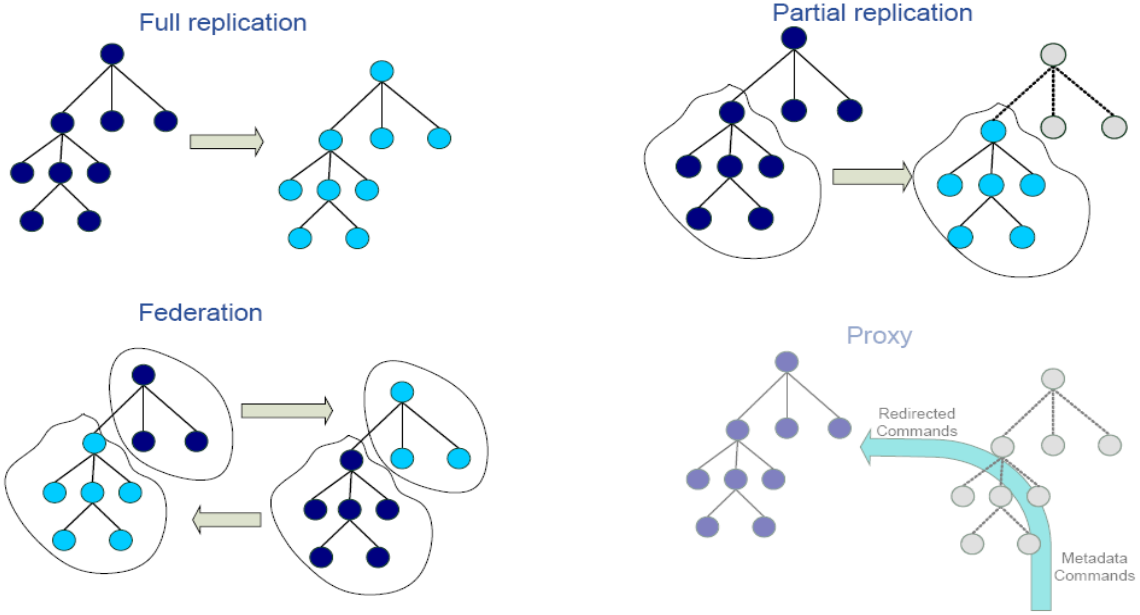
Security in AMGA

- **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



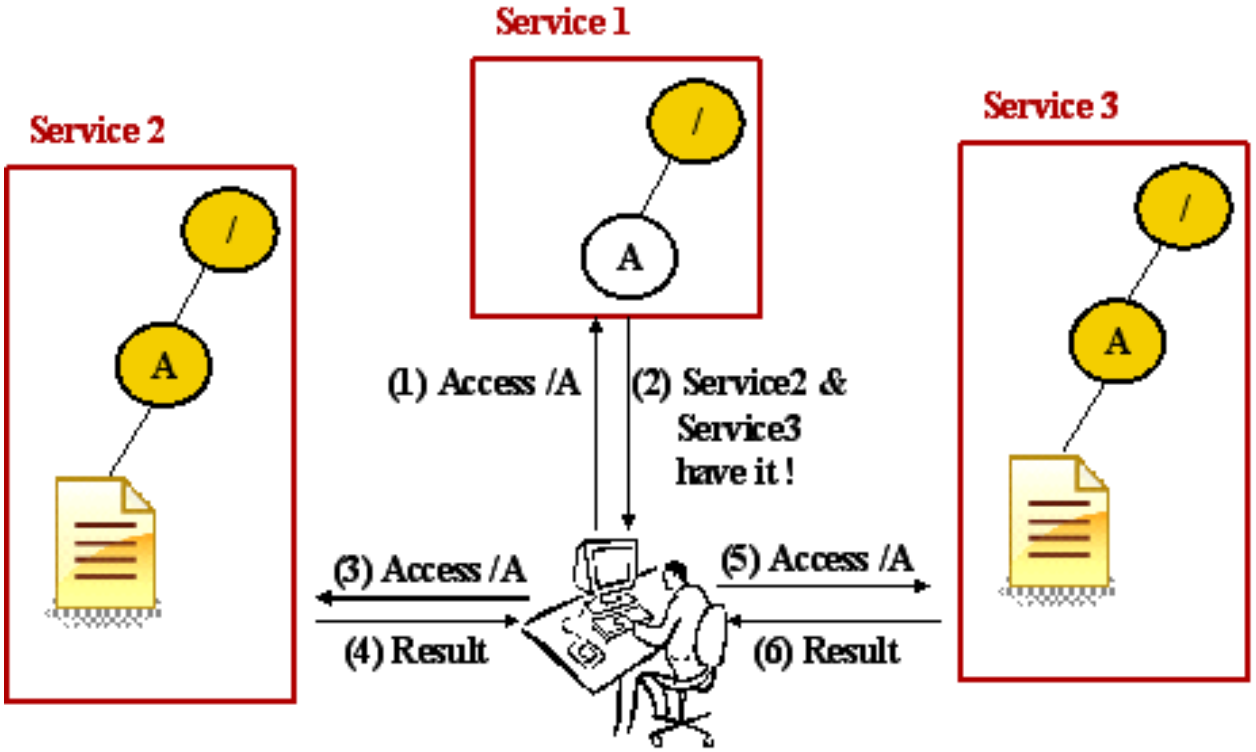
Replication in AMGA

- Metadata collections can be replicated to improve reliability, scalability and performance
  - Partial/full Replication:
  - Master-Slave & Asynchronous communication model



Federation in AMGA

- A mechanism to integrate distributed metadata
  - Provide a user with a virtualized view on metadata
- Two types: server side, client side

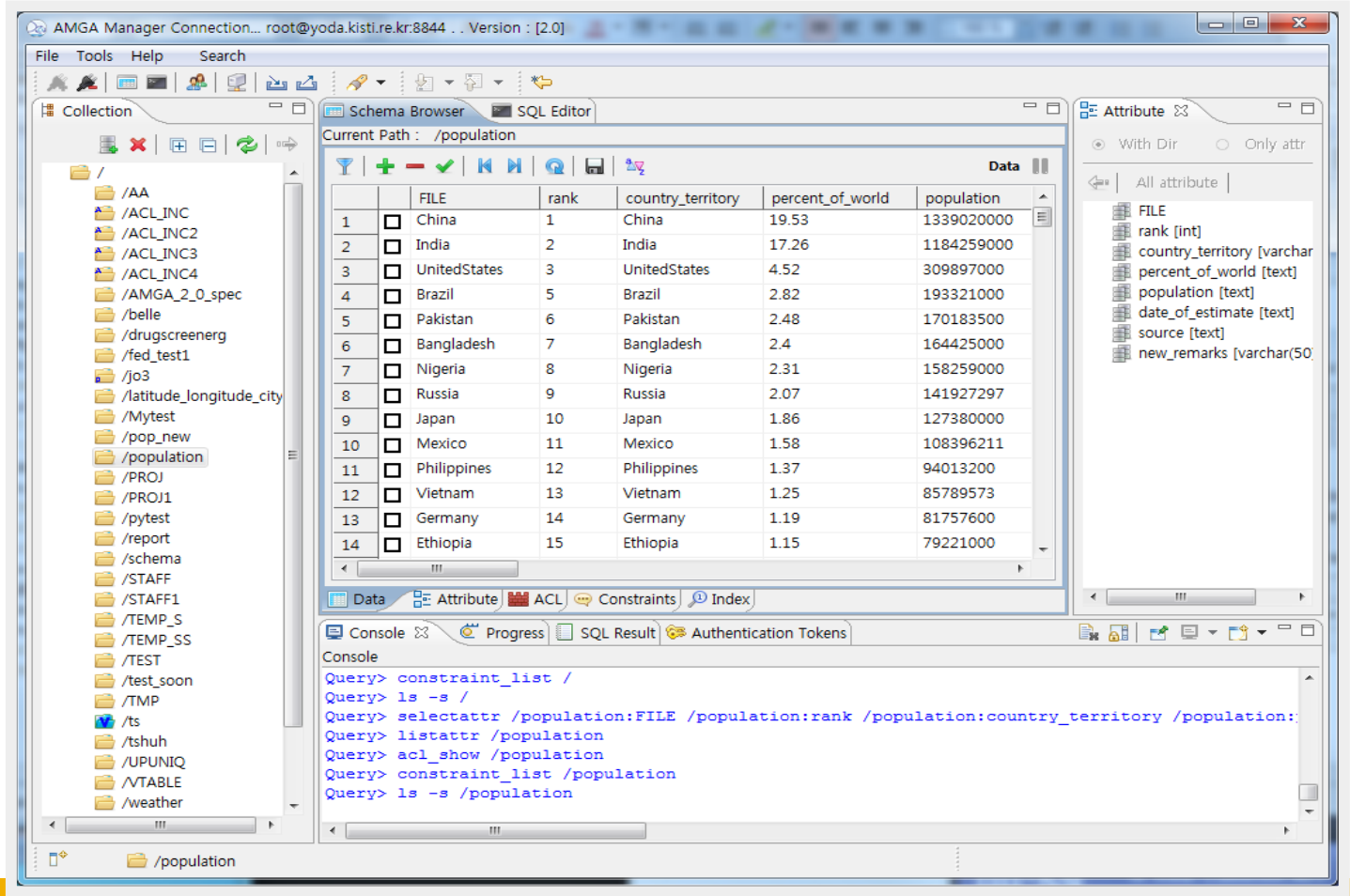


AMGA Client Tools

- **AMGA CLI (Command Line Interface) tools**
  - mdclient
    - Interactive CLI tool to interact with AMGA server
  - mdcli
    - Non-interactive, one-shoot command line tool
    - used in the shell script to access AMGA metadata
- **AMGA GUI Client**
  - AMGA manager
    - Easy-to-use, general purposed GUI interface to interact with AMGA
- **Programming API**
  - Support for C/C++, Java, Python, Perl, PHP API

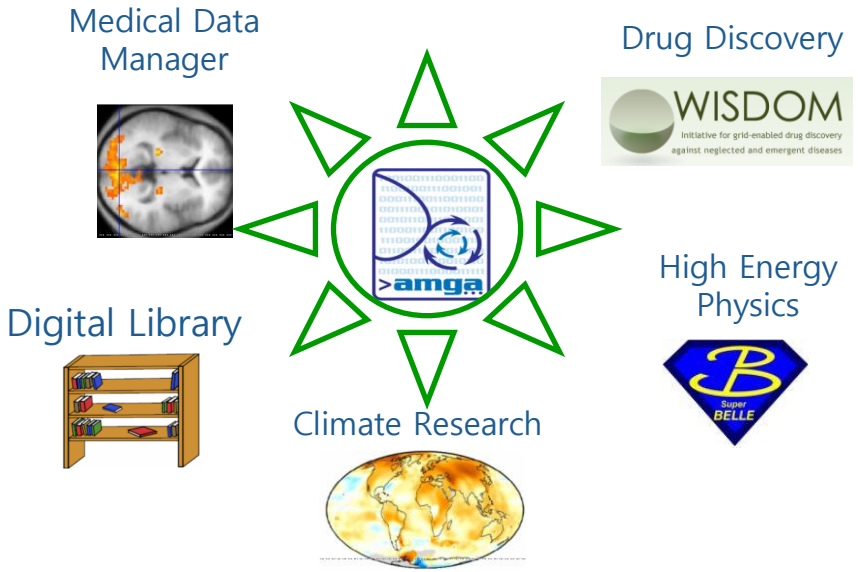
AMGA Manager

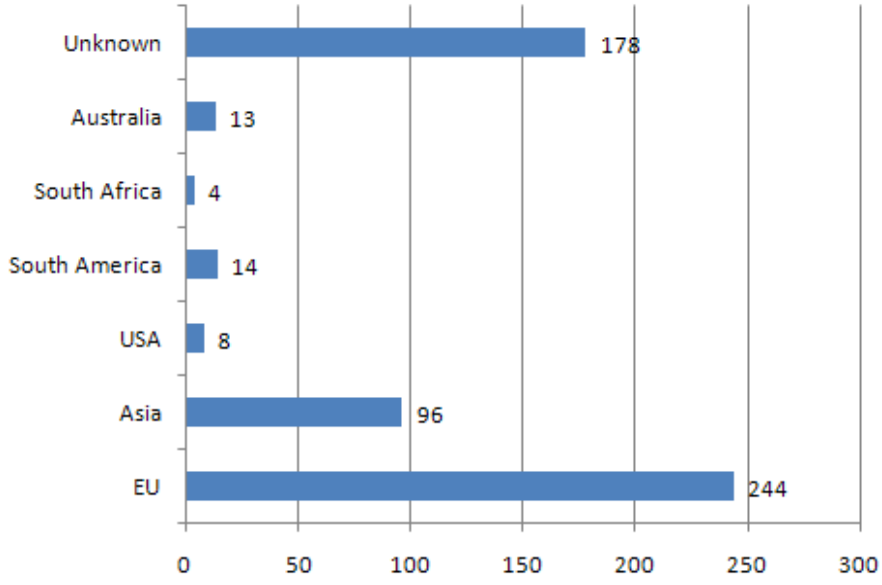
- **Easy-to-use, GUI client tool**
  - Interactive, easier and more intuitive metadata handling



AMGA Usage in Applications

- Has been adopted by many user communities for their metadata services including MDM, Belle II, WISDOM, e-Health-Child and so on
- In 2010, downloaded 607 times from ~110 different sites around the world

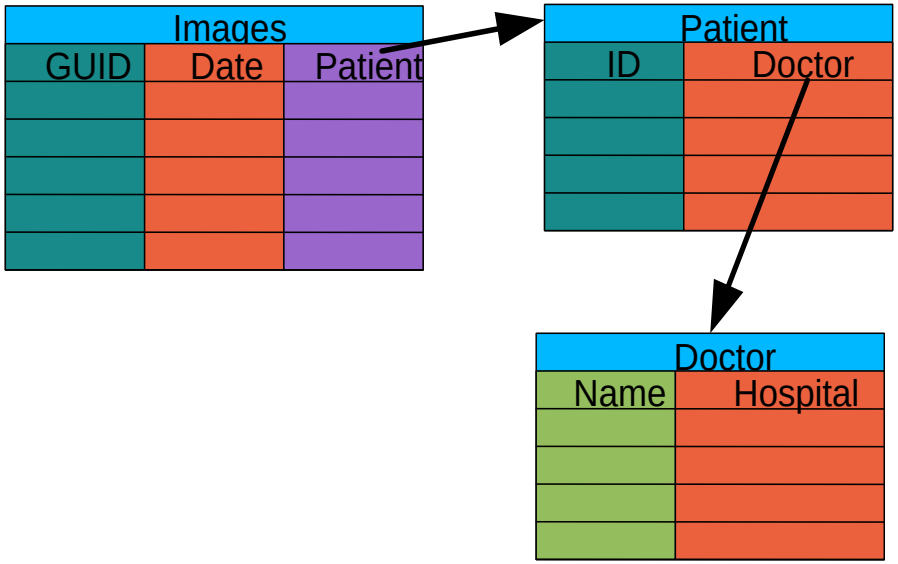




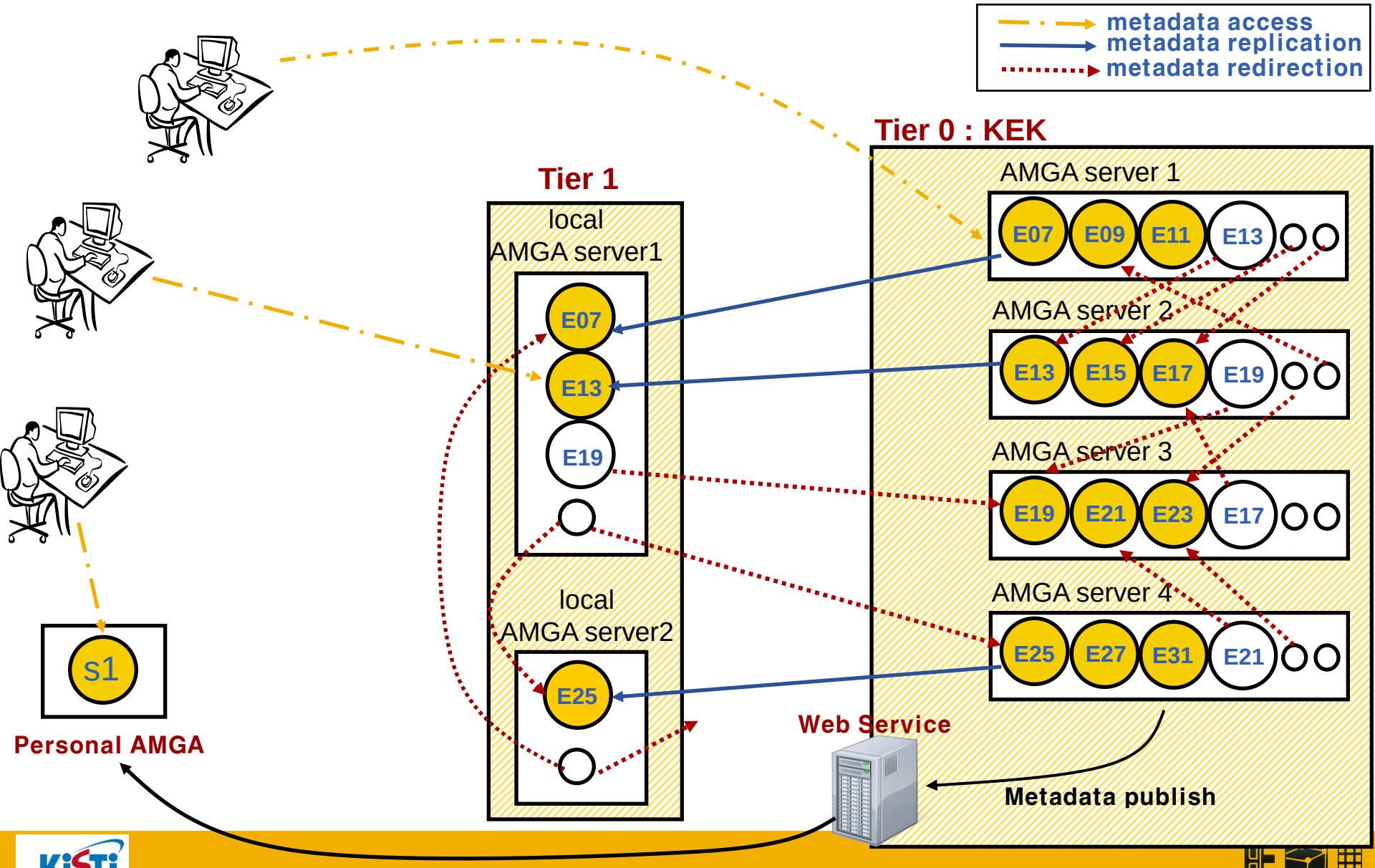
# of AMGA Download in 2010

AMGA Usage in Biomed

- **Medical Data Manager – MDM**
  - Store and access medical images and associated metadata on the Grid
- **Strong security requirements**
  - Patient data is sensitive
  - Data must be encrypted
  - Metadata access must be restricted to authorized users
- **AMGA used as metadata server**
  - Demonstrates authentication, fine-grained access control and encrypted access
  - Used as a simplified DB

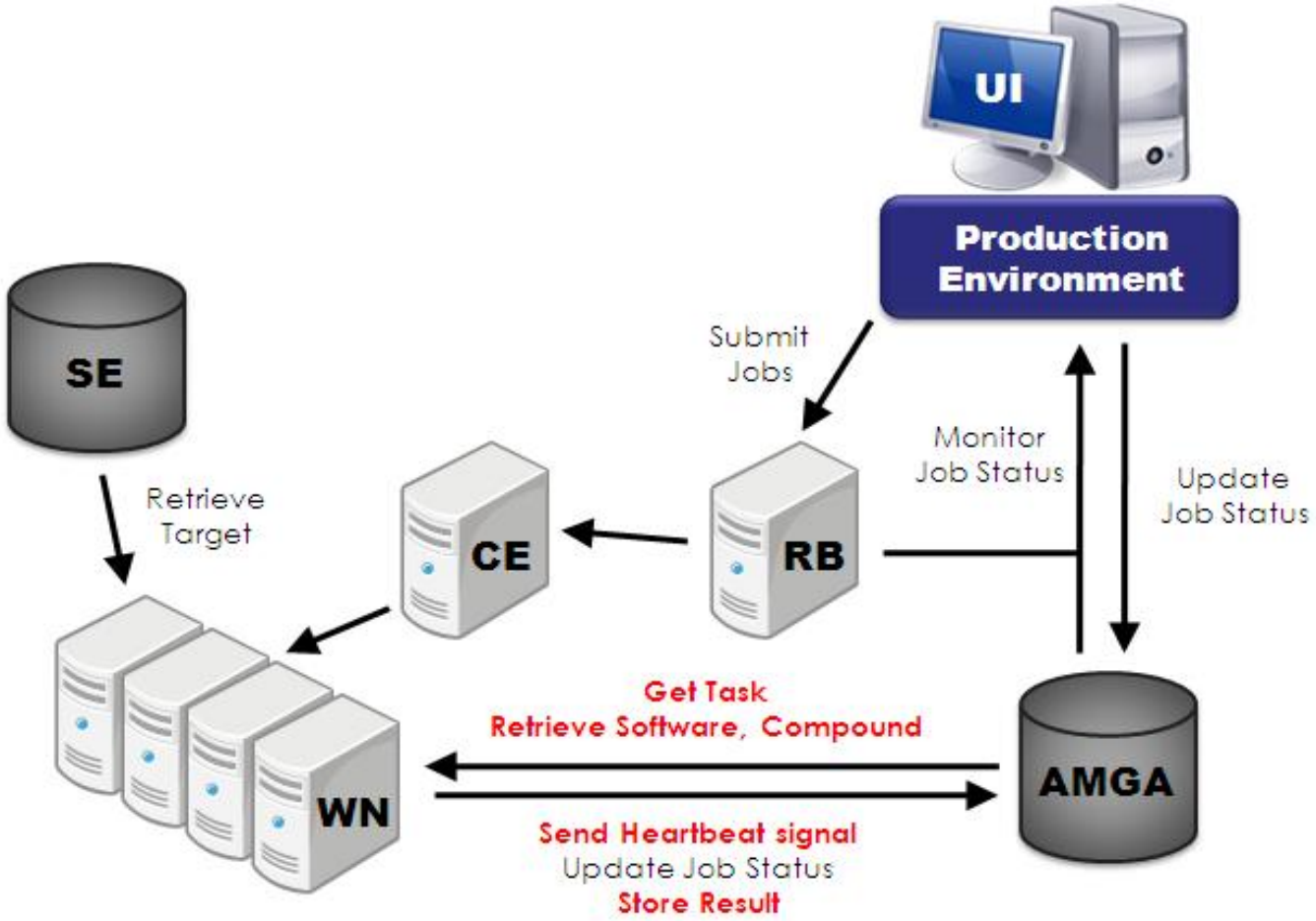


AMGA Use Case in Belle-II



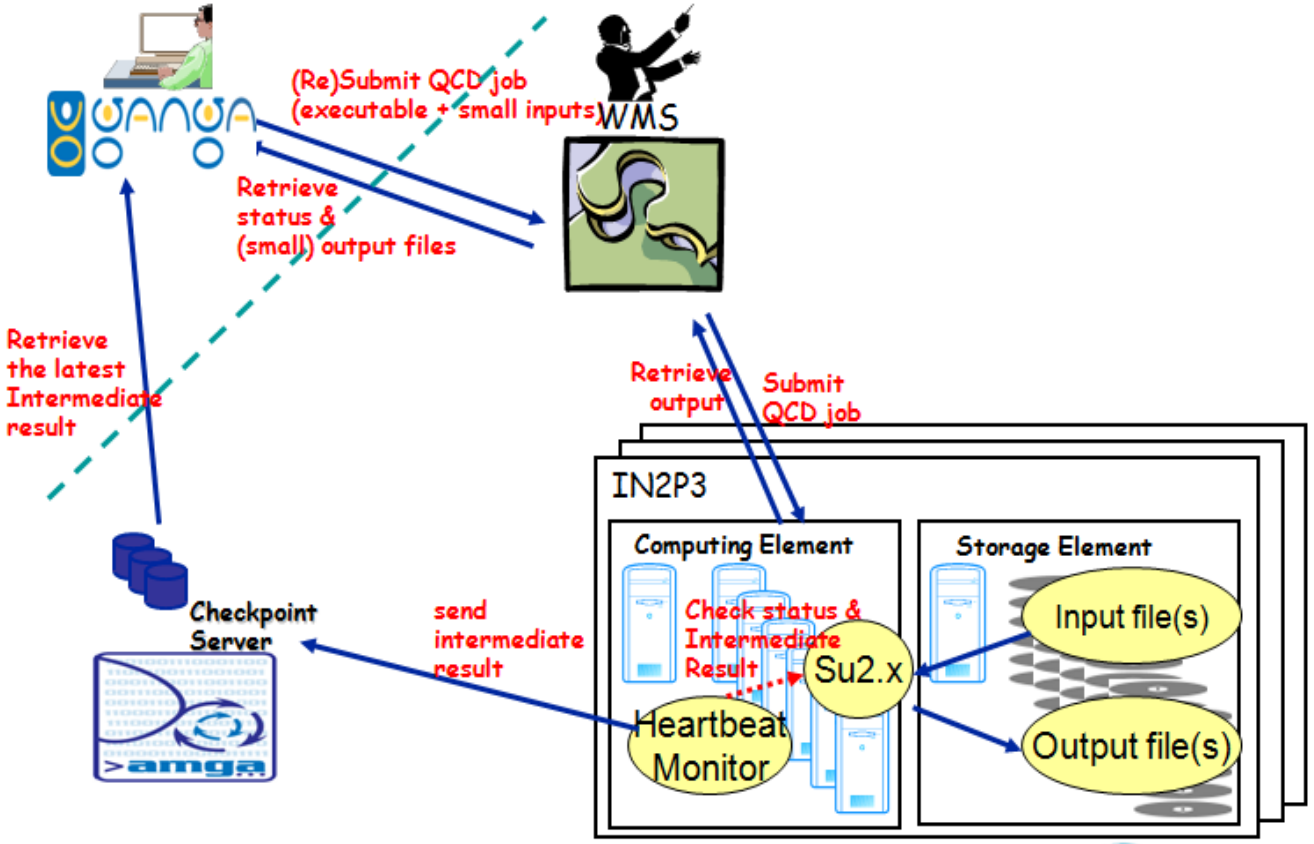
AMGA Usage as Task Manager in WISDOM

- **WISDOM was International initiative to deploy large-scale *in-silico* docking jobs on the grid**



AMGA Usage as Checkpoint Server in QCD

- **QCD simulation is a long-running job lasting ~10 days to complete**
  - Application-level checkpointing is required to successfully deploy and run such long-duration jobs on the grid



References

- **AMGA Web Site**
  - http://amga.web.cern.ch/amga/
- **EMI Software Repository**
  - http://emisoft.web.cern.ch/emisoft/index.html

hwang@kisti.re.kr