The GReIC Project

Massimo Cafaro, Ph.D.
SPACI Consortium and University of Salento (Lecce), Italy
Outline

• Motivations
• GRelC Project
• GRelC DAS
  – Architecture
  – SDK
  – GUI
  – Queries
  – Experimental Results
  – Porting on gLite
  – On Line User Tutorial (GILDA)
• GRelC DAIS
• Conclusions
• Data Grids should provide a low level framework also for grid-database management (fine grained approach)

• No new DBMS or new query language

• Legacy systems/databases and standard SQL

• Need for more complex and efficient “query” in grid

• Integration with production grid environments (based on gLite, globus, …)

• Main requirements: security, transparency, interoperability, efficiency, robustness, scalability, etc.

• Main topics: Integration, access and management
Introducing the GReIC Project

- Grid Relational Catalog is a project which aims at designing and developing a set of efficient, secure and transparent Data Grid Services (Starting date, January 2001).

- **GReIC Data Access Service** aims at providing a large set of functionalities to access both relational and non relational DataBases in a grid environment.
GRelC Project: a bit of history
GReIC DAS Architecture (in the large)
GReIC DAS: Main Features

• Entirely based on the C programming language
• Multithreaded web service
• It exposes a GSI enabled and WS-I compliant web service interface
• **Mutual authentication** based on GSI (X.509v3 digital certificates)
• GReIC DAS Authorization based on ACL for local management
• **VOMS Support**, for global management
• Information System Support (**BDII** compliant)
• Wide set of data access control policies
• **Full GSI support**: data encryption, data integrity, protection against replay attacks and detection of out of sequence packets
GReIC DAS: Main Features

- SingleQuery and MultiQuery Support
- Support for **synchronous** and **asynchronous** queries
- Dynamic binding to **heterogeneous DBMSs**
- **Two-level logging** (users, connections, queries, etc.)
- GSI enabled remote administration tools and remote log
- **Compression, chunking, prefetching** and **streaming** to enhance performance on a WAN
- Wide **SDK** for developers (C, C++ and Java)
- **No dependencies** concerning other middleware (GSI is the only requirement)
GReIC DAS Data Mng Solution

GReIC DAS

- User Mng
- Host Mng
- VOMS Mng
- VOMS-enabled GReIC DAS front-end
- Grid-DB Mng
- Data Access

DBMS Client APIs

- FREETDS
- XML-RPC
- SOAP over GSI

Metadata Catalog

GReIC DAS

- UnixODBC
- PostgreSQL
- MySQL
- DB2
- Oracle
- SQLite
- Microsoft SQL Server
- Sybase

C Client

- HTML Client

C++ based application

Java-based application

SOAP over GSI

GReIC Portal URL

https://grelc.unile.it:8443/GReICPortal

Servlet Container (Apache Tomcat)

Servlet Container (Apache Tomcat)

C+G

GRelC/JProxy

GRelC Proxy

GReIC DAS

GRelC Qt GUI

XGReIC Qt GUI

GRelC++Proxy

GRelC++Proxy

C HTML Client

C Client

C Client
Standard Database Access Interface

Features:
- Standard access to data sources
- Types uniformity
- Error uniformity
- Plug-in architecture based on dynamic libraries

- Dynamic binding to:
  PostgreSQL  MySQL  SQLite  IBM/DB2, Oracle9i, MS-SQL Server, UnixODBC, Textual DBs, etc.
GReIC Queries

GReIC latest release supports the following query types:

- Single Query Online
- Single Query Memory (+ chunk management)
- Single Query File (+ chunk management)
- Single Query File + ZIP (+ chunk management)
- Single Query Prefetch (parallel chunk download/processing)
- Single Query Stream (resultset streaming)
- Web Single Query XHTML (+ chunk management / paging)
  - CSS v2.0, XHTML v1.0 Strict

Results displayed in the following formats:

- Tabular
- XML
- HTML
- RAW
Single Query File Approach (Zip)

This kind of query is suitable to retrieve medium/large resultsets.
This kind of query is suitable to retrieve medium/large resultsets.
GReIC Data Access: Clients

---

```xml
<RECORD>
  <id>1</id>
  <name>Johannes</name>
  <surname>Smith</surname>
  <address>1, Piccadilly Circus - London</address>
  <phone>441230876543</phone>
</RECORD>

<RECORD>
  <id>2</id>
  <name>Toby</name>
  <surname>Elmer</surname>
  <address>10, Downing Street - London</address>
  <phone>4412313667841</phone>
</RECORD>

<RECORD>
  <id>3</id>
  <name>Alessandro</name>
  <surname>Negro</surname>
  <address>Via Margherita,16</address>
  <phone>0300701201</phone>
</RECORD>

<RECORD>
  <id>4</id>
  <name>Giacomino</name>
  <surname>Fiore</surname>
  <address>Via Sperati,23</address>
  <phone>030709876543</phone>
</RECORD>
```

---

```bash
grelc@gandalf:/PRODUCTION/access/bin> ./grelc_service_submission_dml -t "select * from person" -d phonebook -x
```

---

```bash
[Sat Dec 16 00:21:11 2006] CSI plugin for gSOAP v2.6a; Established security context with: /UsGrid/00=smile.it/00:
```
This kind of query is suitable to retrieve VERY LARGE resultsets.
International Testbeds

Lecce (Italy) — Beijing (China)
Test Performance

SQ comparison (Beijing - Lecce) - large

- MEMORY
- DIME
- DIME+ZIP
- STREAM
- PSQL

The diagram shows the comparison of SQL query performance for different systems (MEMORY, DIME, DIME+ZIP, STREAM, PSQL) across varying numbers of tuples (10000, 50000, 100000). The y-axis represents the time in seconds, ranging from 0 to 450.
Porting of GRelC on gLite was straightforward
Porting on gLite of both client and server side
The middleware works fine both on LCG-2-7-0 and current gLite 3.x middleware
GRelC DAS runs also on several platforms:
  - Linux
  - MAC OS X
  - FreeBSD
Both IA64 and IA32 platforms are supported (we installed on SPACI-LECCE-IA64 (EGEE SA1 partner) the GRelC DAS)
GReIC on gLite: A New Service

- Straightforward integration within the EGEE farm model
- GReIC DAS provides fine grained data mng service
- This service can be used both as farm service and as VO service depending on the context, the database policies/constraints

Extended EGEE Farm Model
GReIC on gLite: BDII

- GLUE schema extension providing information about VOs and Databases (we plan to interact with OGF GLUE-WG)
- Local admin can set up the Information Provider Level parameter
  Min: 0 to publish just basic info (only the contact string)
  Max: 7 for all info (contact string, VOs, DBs, tables, fields, etc.)
• We provide global authorization by means of VOMS Extensions
• High level of scalability concerning DAPs related to VOs
• Two-level authorization framework: both local and global policies management can be provided (mixed mode)
Two-level authorization

- Global authorization (through VOMS extensions)
- Local authorization (by means of the local GRelC DAS authorization framework)
- The two masks obtained from global and local authorization are combined to infer the final User Privileges Mask (UPM)

- 3 scenarios
  - global mode, coarse grained approach
  - local mode, fine grained approach
  - combined mode
GRelC Website

URL: http://grelc.unile.it/

Mailing List

mail: grelc-user@sara.unile.it
GILDA Deployment
Functionalities:
- Login
- GReIC DAS Registration
- Host Management
- Instance Management
- User Management
- Query submission
- Deployment Map

Features:
- **Seamless** and **ubiquitous** access to GReIC DAS enabled resources
- **No additional software** installation / configuration is required
- **Complete** and **user-friendly** Grid Data Portal Interface (It entirely replaces CLI)
GReIC Portal: Some Snapshots
SDK for Developers

Grid Enabled Web Apps (GReIC Portal)

Grid Enabled Stand alone GUI (XGReIC)

High Performance Grid Services & Clients (CLI)
The GRelC DAS is now part of the INFNGRID release (ig_GRELC)
The ig_GRELC profile is available for gLite (full yaim compliant)

Server
ig_GRELC (GRelC DAS)

Clients
Installed by default on UI and WN of the INFNGRID

Documentation
Full support in terms of guides, tutorials, etc.

WebSite
www.grelc.unile.it

Data Grid Portal
https://grelc.unile.it:8443/GRelCPortal/
What’s next? ... GReIC DAIS 3.0

Architecture in the large
Requirements (I)

• **Transparency:**
  – Access: transparent access to heterogeneous DBMS
    ▪ RDBMS (Oracle, MySQL, Postgresql, IBM/DB2, etc.)
    ▪ XML (Xindice, eXist, libxml2 based documents, etc.)
  – Location: physical database location completely hidden
  – Naming: mapping between physical parameters and virtual/grid DB name

• **Efficiency:**
  – Multithreaded grid service
  – Prethreaded support to enhance performance
  – Advanced delivery mechanisms supporting:
    ▪ Compression
    ▪ GridFTP
    ▪ Streaming
    ▪ etc.
Requirements (II)

- **Scalability:**
  - Management of virtual organizations
  - Role-based management based on existing services:
    - VOMS (gLite - EGEE)
    - etc.
  - Decentralized solutions for data service interactions in grid
    - P2P data grid services
    - …

- **Interoperability:**
  - Standardized data request interface
    - Web Services Technologies
  - Wide set of standards for underlying technologies
    - OASIS
    - W3C
    - OGF
  - Working Groups
    - OGF, EGEE, etc.
Requirements (III)

• **Security:**
  – Full security support by means of Grid Security Infrastructure (GSI):
    – Mutual Authentication
    – Authorization based on local ACL and global VOMS support
    – Data Integrity
    – Data Confidentiality
    – …

• **Loosely Coupled System:**
  – Decoupling routing/access issues
  – Higher (second) level of Data virtualization
  – Based on P2P grid services & protocols
  – Integrated data will span:
    – Relational databases
    – XML Databases
    – Flat files
    – *Etc.*
P2P and Grid Protocols/Services

Key issues
1. UQI to identify the queries within the P2Pnet
2. HTL to provide a bound in terms of hops
3. TTL to provide a bound in terms of time
4. VO centric support (scalable and flexible)
5. Full GSI Support (delegation of credentials)
6. Synchronous and asynchronous support

Four phases
1. Query Forward
2. Querying local resources
3. Merging of partial results
4. Data retrieval
Data Grid Integration Scenario for Metadata Mng

1. Query Forward
2. Query Submission
3. Collecting resultset
- Database Log
- System Log
• Query Submission
Monitoring

GReIC Portal New Web Pages DAIS
Conclusions

- GReIC DAS/DAIS provides support in Grid for a wide range of DBMSs.
- It is currently tested on several grid environments (SPACI, SEPAC, GILDA, INFNGRID)
- A wide SDK (C, C++, Java) is available for developers
- CLI, XGReIC and Portal Interfaces to ease Grid-DB mng
- gLite compliant (porting on gLite 3.x/4.x and integration with VOMS framework, BDII, etc.)
- Support for several platforms (IA32 and IA64)
- Included in the INFNGrid Release
- Currently the software is included in the EGEE Respect Program
- Used in the Euro-Mediterranean Centre for Climate Change for climate metadata management
Supervisor: Prof. Giovanni Aloisio (giovanni.aloisio@unile.it)
Project P. I.: Sandro Fiore, Ph.D. (sandro.fiore@unile.it)
Team Members:
Massimo Cafaro, Ph.D.
Alessandro Negro, MSc
Salvatore Vadacca, MSc
GRelC WebSite: http://grelc.unile.it
Mailing lists: grelc-user@sara.unile.it