

**The 2008 International Conference on High Performance Computing and Simulation
(HPCS'08)
June 3-6, 2008, Nicosia, Cyprus**

TUTORIAL II

Grid Database Access, Management and Integration

Sandro Fiore and Salvatore Vadacca
Euro-Mediterranean Centre for Climate Change (CMCC) and
SPACI Consortium
Lecce, Italy

TUTORIAL DESCRIPTION

Grids encourage and promote the publication, sharing and integration of scientific data, distributed across Virtual Organizations. Scientists and researchers (from bioinformatics, astrophysics, etc.) work on huge, complex and growing datasets. The complexity of data management within a grid environment comes from the distribution, heterogeneity and number of data sources. Along with coarse-grained services (such as grid storages, replica services and storage resource managers), there is a strong interest on fine-grained services concerning, for instance, grid-database access and management. This tutorial will explain in detail Grid-Database Management Systems, with topics including basics on DBMS & Grids, database virtualization, data access and integration, security issues, performance issues, and interoperability with existing middleware (Globus, gLite, etc.). We present and discuss the state of major projects in the area, with focus on emerging and consolidated grid standards and specifications as well as production grid middleware. A demo on the Grid Relational Catalog (GRelC) Project will show real scenarios and use cases related to data access and integration. Examples concern bioinformatics (Italian LIBI Project), climate changes (Euro-Mediterranean Centre for Climate Change Data Grid CMCC-DataGrid), virtual clinical folders, accounting, monitoring and others. Both relational and XML databases are referred to.

TUTORIAL OUTLINE

- Basics on Database Management Systems & Grids
- Database virtualization and Grid-DB concept
- Existing and novel approaches for data access and integration

- Security issues, e.g., ACL, VO Membership management systems
- Performance issues, e.g. advanced delivery mechanisms/protocols in grid, streaming, compression
- Interoperability with existing middleware (Globus, gLite, etc.)
- Scalability issues
- Data Grid Portals and short demo on the GReIC Portal.

TARGET AUDIENCE

The targeted audience includes people interested in concepts related to database access, management and integration (both relational and XML) and grid environments (both gLite and Globus-based). Participants may, for instance, have background in bioinformatics (molecules/protein DBs), astrophysics (astronomic DBs), or climate research (metadata DBs for Earth Science, CMCC scenario).

REQUIRED BACKGROUND

Basics on Database Management Systems and query languages (SQL for RDBMS and XPath for XML DBs).

TUTORIAL DURATION

Two hours:

- Hour 1 – Basic concepts on Grid data management systems (S. Vadacca)
- Hour 2 – Advanced concepts on Grid data management systems, state of the art and future roadmap (S. Fiore)

INSTRUCTORS BIOGRAPHIES

Sandro Fiore was born in Galatina (ITALY) in 1976. He received a summa cum laude Laurea degree in Computer Engineering from the University of Lecce (Italy) in 2001, as well as a PhD degree in Informatic Engineering on Innovative Materials and Technologies from the ISUFI-University of Lecce in 2004. Research activities focus on parallel and distributed computing, specifically on advanced grid data management. Since 2004, he is a member of the Center for Advanced Computational Technologies (CACT) of the University of Salento and technical staff member of the SPACI Consortium. Since 2001 he has been the Project Principal Investigator of the Grid Relational Catalog project (<http://greic.unile.it>). Dr. Fiore was involved in the EGEE project (Enabling Grids for E-science) and is currently involved in the EGEE-II project and other national projects (LIBI). Since June 2006, he leads the Data Grid group of the Euro-Mediterranean Centre for Climate Change (CMCC) in Lecce (Italy). He is author and co-author of more than 40 papers in refereed journals/proceedings on parallel & grid computing and holds a patent on advanced data management.

Salvatore Vadacca was born in Galatina (LE) in 1982. He received summa cum laude bachelor and master degrees in Computer Engineering from the University of Lecce, Italy in 2003 and 2006, respectively. His research interests include data management; distributed, peer-to-peer and grid computing; as well as web design and development. Since 2003, he has been a team member of the GReIC Project. In 2006 he joined the Euro-Mediterranean Centre for Climate Change (CMCC) in Lecce, Italy, where he works in the Data Grid group.