Data Management

in Global Computing and Grids

Domenico Talia

CoreGRID / University of Calabria

www.coregrid.net

talia@deis.unical.it

BIGG Meeting – Sophia Antipolis – 28-29 November, 2006

European Commission

Information Society Technologies

AGENDA

- Introduction of the subject
- Objectives of the session
- Invited speakers (2)
- Moderated discussion on
  - Common challenges
  - Collaboration opportunities
  - Fostering Synergy

Introduction (1)

Distributed Data Management

Dealing with issues concerning representation, storing, querying, discovery, exchanging and integration of data (and resulting knowledge) in dynamic distributed environments.

- Those issues must be addressed by exploiting features offered by Grid/P2P/GC/UC Technologies.
- Multi-paradigm approaches can be envisioned.

Introduction (2)

- Data and knowledge are becoming key elements in GRIDs (and in Global Computing), as or more than high performance delivery.
- Many activities all over the word on
  - GRID/P2P databases and distributed repositories
  - Distributed metadata management
  - Pervasive information systems
  - GRID-based digital libraries
  - Distributed data streaming management
  - Distributed knowledge management
  - Data-oriented services
- A more important role is expected in the near future.

Objectives

- Discuss R&D issues in Data Management in Grids and Global Computing scenarios.

- Identify:
  - Missing Solutions in Distributed Data Management
  - Research Challenges in Global Data Management
  - Potential Overlaps and Gaps in current Research Activities
  - Common vision of Data Management and research interests
  - Industrial needs and transfer
  - Synergies and future common work

CoreGRID KDM Institute

- The KDM Institute is providing a collaborative environment for 11 research teams working on:
  - Distributed storage management on GRIDs
  - Data Access and Semantic GRID techniques and tools for supporting data intensive applications
  - Knowledge discovery and data mining in GRIDs.

- With focus also on
  - Service Level Agreement Negotiation and
  - Security Requirements for Data Management

Invited Speakers (1)

Prof. **Evaggelia Pitoura**

University of Ioannina, Greece

Talk: **Data Routes within Grids, through the Globe**

Invited Speakers (2)

Prof. **Vladimir Vlassov**

SICS & CoreGRID, Sweden

Talk: **Scalable Peer-Group Services in Grids**

DISCUSSION: Research Topics (1)

- Research issues:
  - Semantic technologies for Grid/GC data management
  - Decentralized Scheduling for data-intensive applications
  - Service Oriented models and architectures for data management (what we need?)
  - Data-intensive computing models for mobile/Grid environments
  - Dispersed Data Virtualization

DISCUSSION: Research Topics (2)

- Other research issues:
  - GRID Data Storage Access and Management Architecture
  - Distributed Data Integration Models and Architectures
  - Resource Description and Discovery in Large Collaborative Networks
  - GRID/GC Trust and Security Policies for Managing VOs
  - Distributed Data Mining in GRID/P2P/Pervasive Systems
  - Distributed Adaptive Query Processing in GRIDs and Mobile Environments

Cont.

THANKS

DATA MANAGEMENT Session

- Evaggelia Pitoura: Similarities and differences between GC and Grids in Data Management (and possible convergence).
- Vladimir Vlassov: P2P models for data self-management in Grids (replica management, P2P VO file system, P2P backup storage).
- Many common research issues
  - Data Resource Description and Discovery in Large Collaborative Networks
  - Service Oriented models and architectures for data management
  - Distributed metadata management
  - Distributed Adaptive Strategies and models for data management.
- As Grids are moving towards a more dynamic and service enabling infrastructure more common challenge/interests/activities with GC raise.