DMOD: Distributed Management of Data Laboratory

http://www.dmod.cs.uoi.gr/

Data Routes

within Grids, through the Globe

Bridging Global Computing with Grid (BIGG) Meeting

Evaggelia Pitoura

http://www.cs.uoi.gr/~pitoura

Computer Science Department

University of Ioannina, Ioannina, Greece

In this short talk:

- Why “data” management?
- Our experience from participating in Global Computing Projects
- Data Management in Grids vs Data Management in Global Computers
- A couple of concrete applications

Why data on a global scale?

In the last decade:

online networks of information revolutionized the ways people obtain information and interact with one another

How they travel, meet, shop, learn, etc.

Underlying aspect of such interactions:

Information produced and shared collectively by a large number of individuals

Why data? A Couple of Success Stories

Google: management of Web pages

how to find information

Mapquest: management of maps - TripAdvisor

how to travel

Amazone: book etc catalogue

eBay: product catalogue

how to shop

Blogs: diaries

Flickr: picture database

how to communicate, share personal experiences

Napster (Bittorent, emule, bearshare, etc.): databases of music, movies etc

entertainment, production of art

Wikipedia: encyclopedia

how to learn

The Global Computing FET Initiative

Previous data-driven examples involve/produce

Computing systems that are large, autonomous, un-trusted, *mobile*, heterogeneous – exactly as defined by the GC

Data/information sharing is central

The Global Computing FET Initiative

Global computing projects are FET projects – more exploratory research

Focused not on specific technologies but rather on "abstractions"

(abstraction is an "abstract" term) meaning (for example):

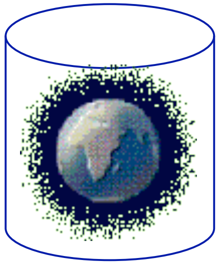
- Foundations (game theory, mechanisms design) *Theory-perspective*
- Data (metadata) Models and Languages (index, query processing) *DB-perspective*

among other things

Our participation in Global Computing

DBGlobe (Global Computing I) as coordinators

AEOLUS (Global Computing II) as partners



DBGlobe

Our Experience from DBGlobe (Global Computing Initiative I)

Data-centric and service-oriented approach to global computing [Sigmod Record, Sept 2003 for an overview]

Extend databases from small-scale distribution to a global scale, extend query languages (with discovery and computaion), continuous execution semantics (streams), etc

XML and web services

A couple of our results:

- [Routing] Multi-level Bloom filters for indexing XML [edbt 2004]
- [Computation] Active XML – a new language – that integrates service calls inside XML documents [sigmod 2003]

AEOLUS (Global Computing Initiative II)

Started in Sept 2005

Algorithmic Techniques for Building the OVERLAY COMPUTER

based on a set of basic functionalities

IP-Project

Combined theoretical + system approach

Examples from theory: Game theory

Examples from systems: Probabilistic replication, data routing and processing for advanced queries in a p2p scale

In this short talk:

- Why "data" management?
- Our experience from participating in Global Computing Projects
- Data Management in Grids vs Data Management in Global Computers
- A couple of concrete applications

Grids vs Global Computers

Grid computing original focus on large scientific applications running on distributed computational platforms

Global computing original focus on general computational tasks on small devices on the edge of the Internet

Different at both the intended applications and system coverage

FET on Global Computing was more on abstractions (models, algorithms) than middleware

Grids vs Global Computers

(a short list of specific differences …)

Grid (initially)	Global I
Deterministic	Probabilistic (best-effort semantics)
Efficient use of computational resource	Extended Functionality (data storage, discovery)
Share Computing resources	Model resources/Prove properties
Willing to cooperate	Selfish (incentives to cooperate)
Trusted	Malicious (security, trust)
Pragmatic (eg standards, stronger assumptions)	“Revolutionary”

Common themes

When we try to realize the global computer

When we extend data management from within grids to a larger deployment

Change of focus

Efficient Resource Management vs Discovery/Integration/Understanding Information and Interactions, Cleaning/Trusting data

Overlay (global computing) ≡ dynamic virtual organization (grid)

To share information

To store data

To share computation (grid)

In this short talk:

- Why "data" management?
- Our experience from participating in Global Computing Projects
- Data Management in Grids vs Data Management in Global Computers
- A couple of concrete applications

One Bridge:

Achieving High Quality of Data (GC) with Guaranteed Quality of Service (grids)

Data Quality

- Freshness
    - up-to-date
- Accuracy/Precision
    - how relevant – accurate (in case of sampling or approximations)
- Trust/Reputation
    - how trusted/secure/authorized/authentic vs copied
- Provenance
    - maintain the origin/history of data

Towards Merging Quality of Data and Quality of Service

Service Quality

- Performance

  Eg response time, resource consumption

- Fault-tolerance

- Load Balancing

Through scheduling, data redundancy techniques, etc

Some specific research problems

From global computing to the grids

Query language and search engines for grid resources

Data-driven workflows that take into account the data that they manipulate and their dependencies

Building “overlays”

Data cleaning tasks

Security/trust

Incentives for share

Probabilistic data quality

From grids to global computing

Platforms/Middleware for doing huge data manipulation – google on a grid?

Standards

Computational resource sharing

Conclusions

- There are commonalities and differences, thus

an interesting and potentially fruitful (research and application wise) integration of two initiatives

- BIGG a step towards this, also need for “incentives” (funding through an appropriate funding tool (strep, coordination activity, etc))

Thank you

Data Routes within Grids, through the Globe

Evaggelia Pitoura

http://www.cs.uoi.gr/~pitoura

Computer Science Department

University of Ioannina, Ioannina, Greece