

# *Team 8*

## *ISSGC'05 Project*

**Claudia CORONNELLO**  
University of Palermo  
Italy

**Christos FILIPPIDIS**  
NCSR "Demokritos"  
Greece

**Dimitrios KORENTZELOS**  
Glasgow Caledonian University  
UK

**Homayoun POURHEIDARI**  
Serono  
Switzerland

**Lin YANG**  
University of Edinburgh  
UK

**Jun ZHAO**  
University of Manchester  
UK

# The “ONE” Problem

Dealing with Massive Amounts of  
Geographically Distributed Data Across  
Organizations

# We are a global organisation of 140 employees supporting 4'600 users in more than 60 locations and from 30 different countries in the world.

## North America

- Canada
- USA
- Puerto Rico

## Latin America

- Argentina
- Brazil
- Uruguay
- Venezuela
- Colombia
- Mexico



## Japan

## Asia-Pacific

- Singapore
- Korea
- Hong-Kong
- Taiwan
- Thailand
- China

## Oceania

- Australia
- New-Zealand

## Europe, Middle-East, Africa

- |               |              |             |                |                   |
|---------------|--------------|-------------|----------------|-------------------|
| ➤ Switzerland | ➤ Sweden     | ➤ Poland    | ➤ Portugal     | ➤ Algeria         |
| ➤ France      | ➤ Denmark    | ➤ Lithuania | ➤ Greece       | ➤ Morocco         |
| ➤ UK          | ➤ Finland    | ➤ Russia    | ➤ Israel       | ➤ Turkey          |
| ➤ Germany     | ➤ Norway     | ➤ Croatia   | ➤ South Africa | ➤ Jordan          |
| ➤ Austria     | ➤ Czech Rep. | ➤ Italy     | ➤ Egypt        | ➤ Saudi Arabia    |
| ➤ Netherlands | ➤ Slovakia   | ➤ Spain     | ➤ Tunisia      | ➤ United Arab Em. |

# Issues and Solutions

## Data Management and Manipulation

- Transfer
- Replication
- Coordination
- Collaboration

## Organize the Problem into 2 Tracks

- Compute Intensive (e.g. cycles to run the jobs, etc.)
- Data Intensive (e.g. replication, access, etc.)

# Compute Intensive Requirements

- Resource Management
  - Condor (Persistent Resource Pools) , GT4 (Indexing based on WSRF) , Unicore
- Job Management (Create and Deploy Jobs)
  - Condor, GT4, GLite/LCG, UniCore
- Schedule Jobs
  - Condor, GLite (WM)
- Prioritize Jobs
  - Condor (MatchMaker Community Policy),GT4 (Community Scheduler), Glite (WM)
- Security (Authn/Authz, Safe Execution)
  - GT4 (GIS), UniCore, Condor SandBox (via GIS)
- Monitor the jobs and their progress
  - Condor, GLite (BDII), UniCore (WSRF operations)
- Scalability of the System
  - Condor (Agents, Resource, and MatchMakers are independent), GT4, UniCore

# Data Intensive Requirements

- Replication Management
  - GT4 (Replica Mgmt. Services) provides
    - Creation
    - Registration
    - Location
    - Mgmt of dataset replicas
  
- Data Transfer (High Speed, Reliable, and Just in Time)
  - GT4 (GridFTP)
    - Parallel data transfer over TCP streams
    - Stripped and Partial file transfer
  
- Data Access (Discoverable, Reliable)
  - GT4 (OGSA-DAI)
  
- Security (Authentication, Authroization, and Secure Transfer)
  - GT4 (GSI, GSS, Kerberos, etc.)
  - HTTPS, X509

# Middleware Puzzle

## USER APPLICATIONS

Condor-G

**Globus Toolkit**

**gLite**

Community Authorization

Delegation Service

WS Authentication Authorization

Pre-WS Authentication Authorization

Credential Management

SECURITY

OGSA-DAI

RTF

GridFTP

RLS

DATA MANAGEMENT

MANAGEMENT

WS GRAM

Pre-WS GRAM

EXECUTION MANAGEMENT

INFORMATION SERVICES

INFORMATION SERVICES

Python WS Core

C WS Core

Java WS Core

C Common Libraries

MDS4

MDS2

MyProxy, PyGlobus, Monalisa, ...

Workload Manager

BDII

RGMA

Cataloguing

**ECONO Broker**

Service Portal

**UNICORE / GS**

**OGSA BES**

HLS

Condor

**VDT**

**GRID FABRIC**

# Progressive Exercise

- Task: to find the pillars on the surface and the texts that are embossed or etched on the top surface of each pillar
- Progress
  - Running scanner over the fifth and sixth points from the deepThoughtII.txt data file, with radius = 10.0 and step = 1.0
  - This returned points that are on the top of the pillars
  - Chose one of the points from the result set and defined a small scoped box to perform “Regular” function over this box
  - Repeatedly shrinking the scale of the box boundary in order to amplify the pillar and the text
- Results
  - Could not find any texts on these two points
  - Then we tried the first point in this text data file
- Future work
  - Write a script to automate this searching process in order to scan points from these two data files and find the knowledge



# Lessons Learned

- Grid technology
  - Web service
  - J2EE
  - XML, WSDL, SOAP
  - Distributed job scheduling, allocation and management
  - Distributed data management
  - Managed services
- “People Grid” worked
  - Computation intensive
  - Data intensive
  - **Team collaboration**: shared knowledge and human resource
  - **Cross-team collaboration**: shared workspace, resources
- Facing reality
  - Un-stability: broken server delays work progress
  - Security is not always there

Searching Knowledge Client

PillarsOfWisdom

Scanner

Regular

BigProbe

FileStore

FileDB1

FileDB80

# Middleware Puzzle

	<b>Team Leader</b>	<b>WS</b>	<b>Algorithm</b>	<b>Java Coding</b>	<b>Presentation</b>
<b>Claudia CORONNELLO</b>		<b>XX</b>		<b>XX</b>	<b>XX</b>
<b>Christos FILIPPIDIS</b>		<b>X</b>	<b>XXX</b>		<b>XX</b>
<b>Dimitris KORENTZELOS</b>		<b>X</b>	<b>XX</b>		<b>XXX</b>
<b>Homayoun POURHEIDARI</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>X</b>	<b>XX</b>
<b>Lin YANG</b>		<b>XXX</b>		<b>XX</b>	<b>X</b>
<b>Jun ZHAO</b>		<b>XXX</b>		<b>XX</b>	<b>X</b>

# Feedback

- Collaborative Environment
- Lectures (need improvement)
- Better Preparation/Organization
- Exposure & Training to Grid Technologies
- Good Facilities (Excellent Lab)
- Good Location
- Better Food (need better wine too :)