Geospatial Web Services Gridification in enviroGRIDS

MINISTRY OF EDUCATION AND RESEARCH

TECHNICAL UNIVERSITY OF CLUJ-NAPOCA

CGIS Computer Graphics and Interactive Systems

Denisa Rodila, Dorian Gorgan, Victor Bacu

Computer Science Department

Technical University of Cluj-Napoca

denisa.rodila,dorian.gorgan,victor.bacu{@cs.utcluj.ro}

Outline

- Objectives
- Context
- Challenges/Problems
- OGC Web services
- Grid Environment
- Solutions and approaches
- Experimental results
- Conclusions

Objectives

- Scope
  - Design and implement an architecture to support the integration of geospatial domain (represented by the OGC Web services) with the Grid environment in the context of enviroGRIDS project
- Objectives
  - Discuss, analyze and propose different solutions and different approaches to bridge the gap between the Geospatial and the Grid infrastructures
  - Explore different “gridification” levels of the OGC Web services and solve the interoperability problem
  - Implement and exemplify the analyzed concepts on different OGC Web services
  - Propose new standards in the OGC and OGF collaboration

Interoperability Context

- Open Geospatial Consortium – OGC
- Open Grid Forum – OGF
- gLite OWS – G-OWS
- Collaboration between OGC and OGF should provide the necessary infrastructure for developing tools, software and services for multiple communities
- Research projects working on the interoperability between the two platforms: SAW-GEO, CYCLOPS, GDI-Grid, GEO-Grid, DEEGREE

enviroGRIDS Project

- **EnviroGRIDS** - Gridifying the Black Sea catchment to support its sustainable development
  http://www.envirogrids.net/

- Founded by the European Commission FP7 framework (Theme 6: environment), April 2009 – March 2013
  27 partners, 7.9 mil EUR.

- Coordinator
  - University of Geneva, Switzerland

OGC Web Services

- Layered on top of Internet standards: HTTP, URLs, MIME and XML World Wide Web standards
- OWS standards include:
  - Web Map Service – WMS
  - Web Feature Service – WFS
  - Web Coverage Service – WCS
  - Web Processing Service – WPS
  - Catalog Service for Web – CSW

Challenges

- Geospatial data come from multiple heterogeneous sources
- Geospatial data have to be accessed, integrated, analyzed and presented across a distributed computing environment
- Processing and storing resources in different formats
- Security and digital rights management
- User authentication and authorization

Grid Environment

- Why do we need the Grid technology in this context?
  - Data Management
  - Computational Power
  - Security
- Execute computational intensive calculations on very large amounts of data by using Web services, by standard approach
- Provide a distribution of calculations and datasets on Grid nodes with possibility of distributed high-speed data transfer

Grid Environment

- When do we need Grid technology?
  - The advantage introduced by the Grid infrastructure are visible only in certain cases:
  - For simple requests, the overhead introduced by the Grid is not compensated and the execution time takes longer
  - Approach: introduce the Grid resources only for the requests for which the overhead is compensated and the execution time is improved

Grid Environment

- To differentiate the requests for which the Grid can bring an improvement from those for which it introduces additional overhead, some analysis have to be made regarding:
  - The type of the requested service
  - The requested parameters
  - The type of functionality executed inside the service
- An estimation should be made regarding the boundary beyond which the advantages of the Grid are visible in the execution time

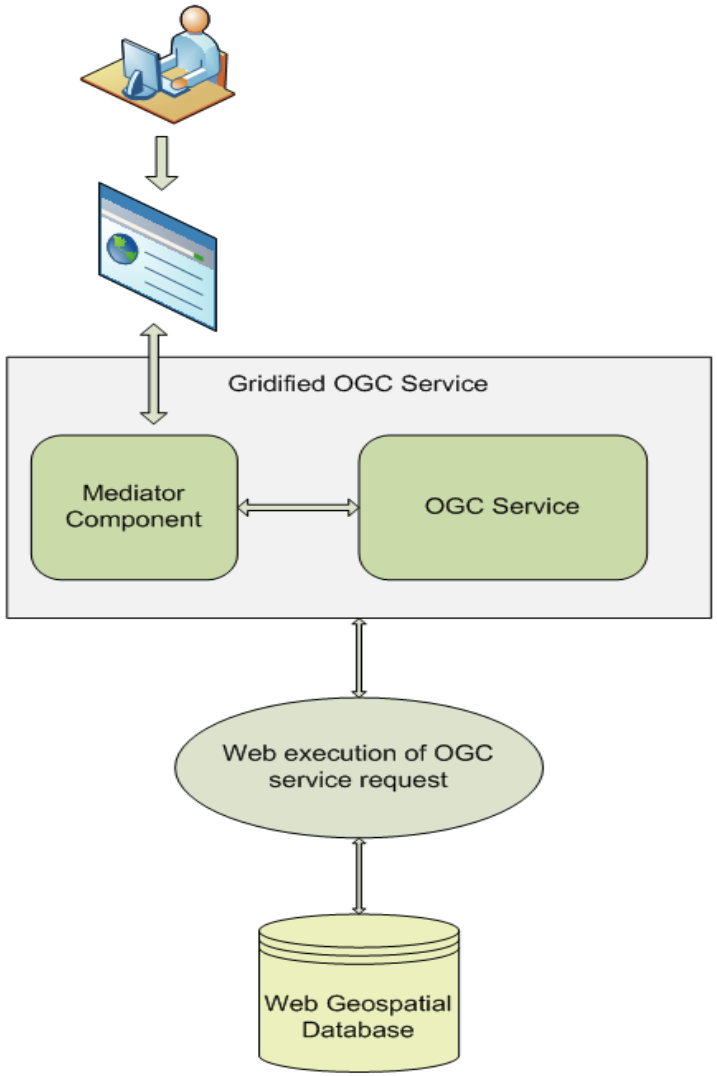
Solutions & Approaches

- A modified OGC Web services will be adapted to support different execution flows depending on the decision made regarding the complexity of the request and the necessity of the Grid as the execution environment
  - Execute the service directly
  - Split the initial request into several jobs and send them to execution on the Grid
- A Mediator component is introduced as a layer between Geospatial and Grid

Solutions & Approaches

- Case 1: The service uses Web Geospatial database.

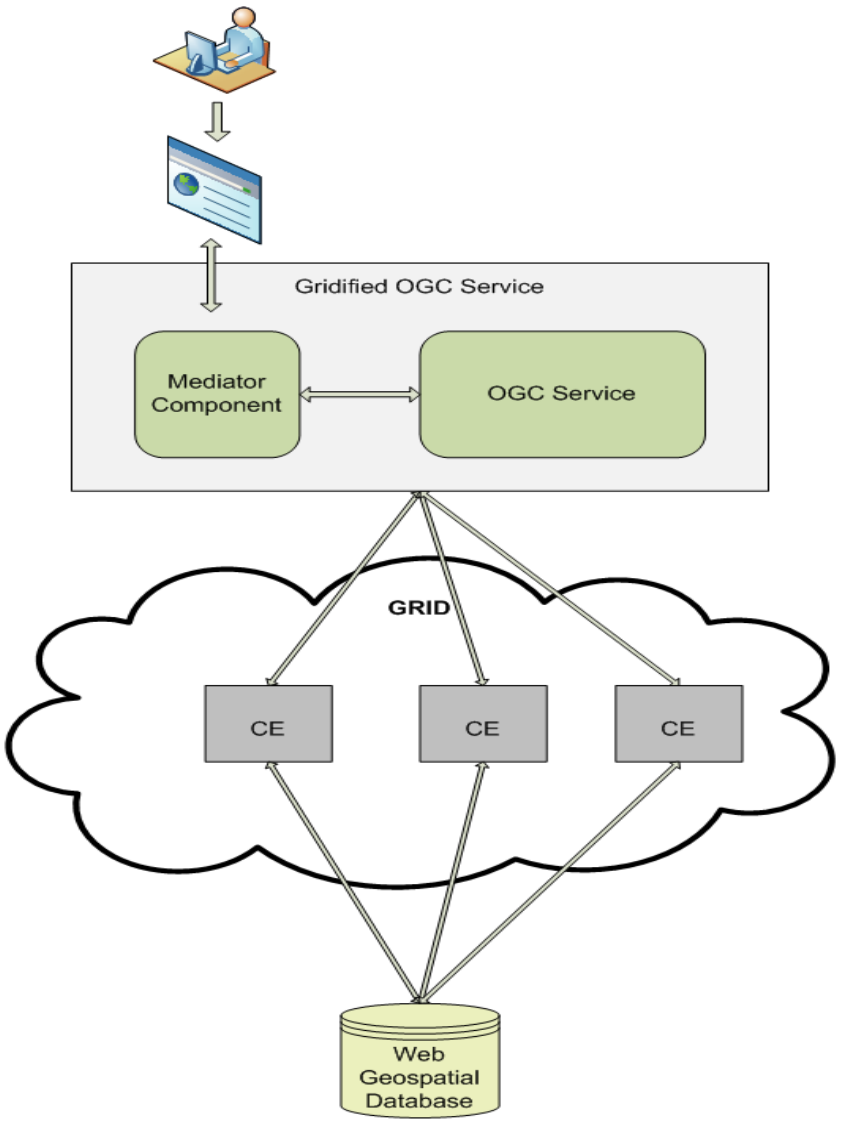
  The simple requests (those for which the execution time does not exceed the overhead introduced by Grid) are executed directly on the Web server and the Grid environment is no longer used



Solutions & Approaches

- Case2: The service uses Web Geospatial database but it is using the Grid environment for the execution.

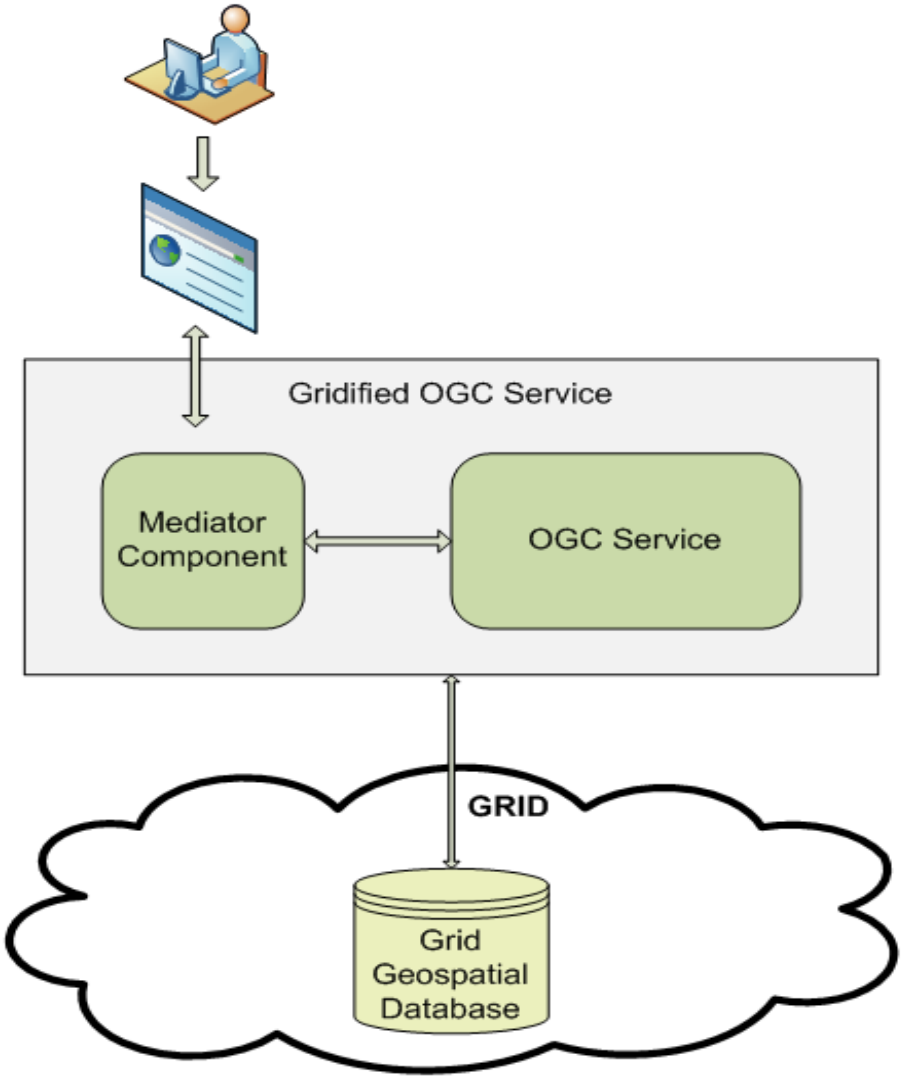
  The request is split into several sub-requests which are executed on individual workers. The workers connect to the Web Geospatial database and retrieve the necessary data.



Solutions & Approaches

- Case 3: The service connects directly to the Grid database, using grid certificates and some special libraries.

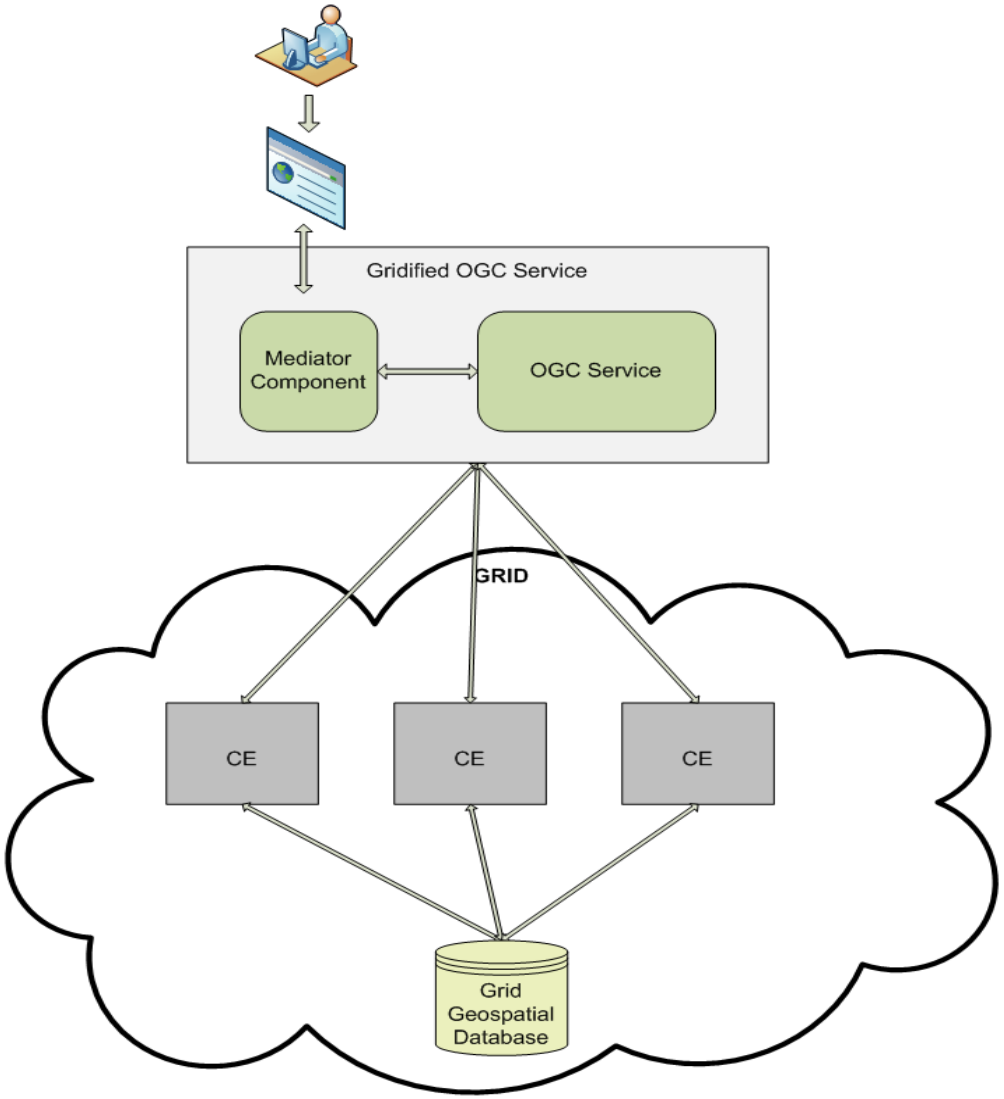
  The connection is established using HTTPS and the data is copied using dedicated scripts.



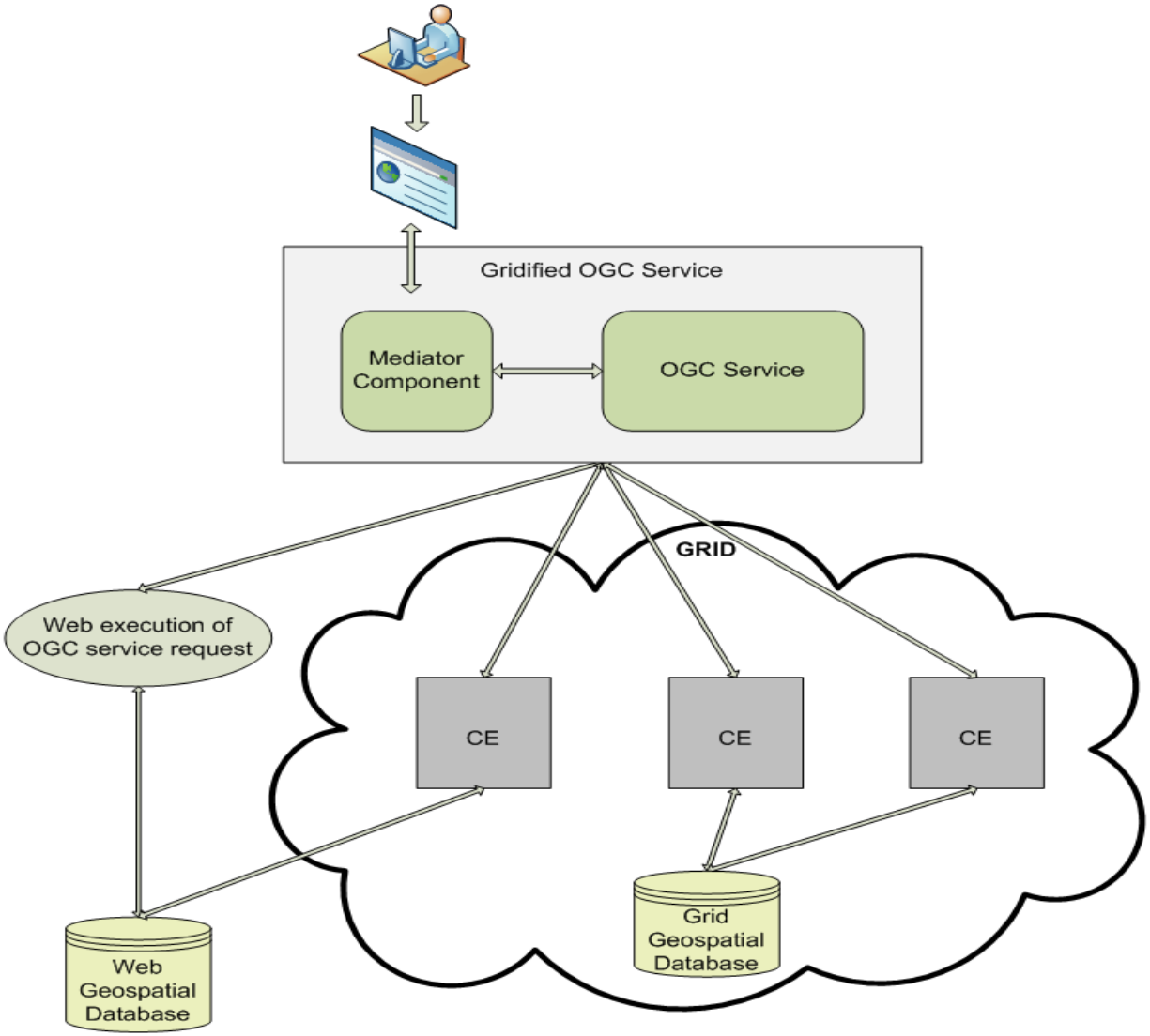
Solutions & Approaches

- Case 4: The service request is split into several jobs which will be executed on different workers (Computation Elements).

  The workers are each responsible for connecting to the SEs (Storage Elements) and obtaining the necessary data.



Solutions & Approaches – general case



Solutions & Approaches

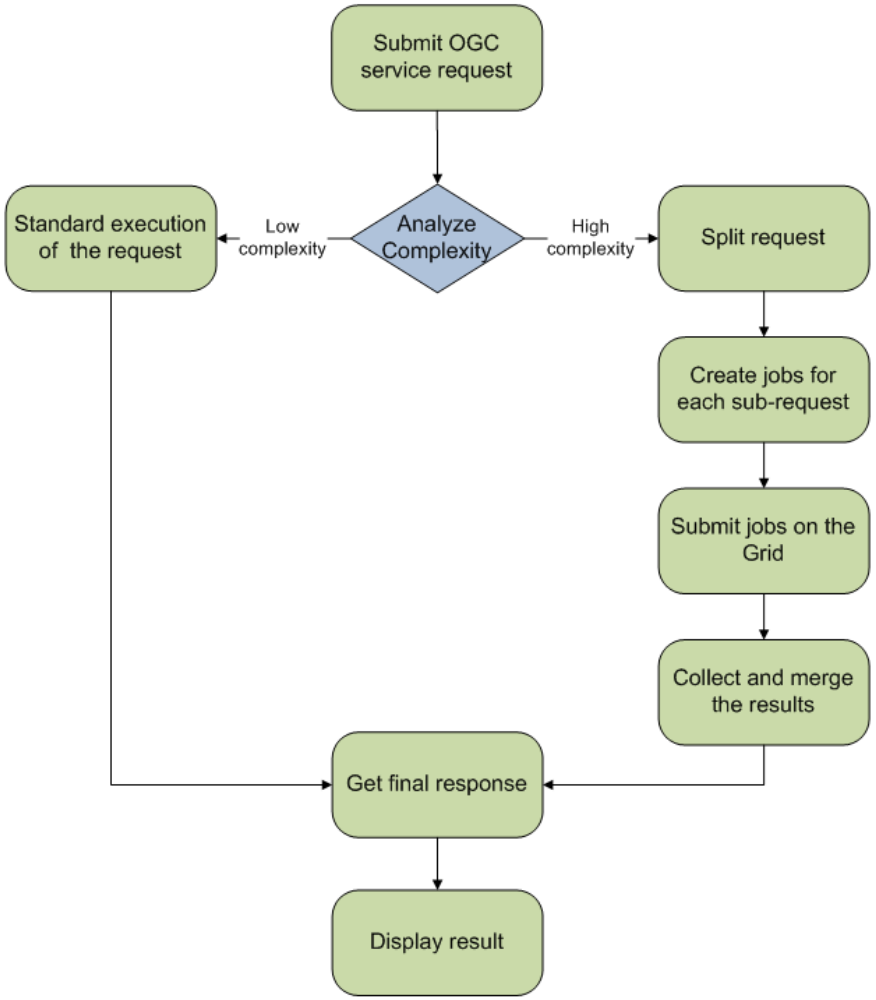
- The gratification of OGC Web services must be done while:
  - Maintain the functionality and the interface in the geospatial context
  - Take advantage of the Grid architecture in executing OGC workflow services
- The parallelism offered by the Grid technology can be applied at different levels:
  - Data parallelism
  - Computing parallelism

Experimental Results

- The development of the gridified services started from the standard implementation of some of the most common OGC Web services developed inside the **deegree project.**
- The implementation of the gridified OGC services is included in the enviroGRIDS project.
- The tests were performed on WFS and WCS services using PostgreSQL database enabled with the PostGIS functionalities on small test data

Experimental Results

- We have tested the first level of gridification for OGC services (deployment of an OGC service instance to different CE nodes inside a Tomcat/Jetty container)



Experimental Results

- The tests were performes using **GANGA** and **DIANE** Grid tools
- The executions were performend on two VOs, available in the enviroGRIDS project: **vo.gear.cern.ch** and **envirogrids.vo.eu-egee.org**
- Experimental executions:
- Simple request (one query) executed locally: ~17s
- Simple request (one query) executed on the Grid: ~2 min 50s
- Medium request (20 queries) executed locally: ~ 6 min 15s
- Medium request (20 queries) executed on the Grid: ~ 5 min 20s

Conclusions

- The integration of OGC Web services into the Grid environment is a complex process and has the following target points:
  - Data parallelism management;
  - Data security;
  - Complex execution parallelism;
- We have identified the main challenges and we proposed different gridification levels of OGC Web services; some of them have already been implemented and the other have been planned to be achieved
- The Grid – OGC service interoperability has practical applications in enviroGRIDS project

Thanks, Questions

Denisa Rodila, Dorian Gorgan, Victor Bacu
Computer Science Department
Technical University of Cluj-Napoca
denisa.rodila,dorian.gorgan,victor.bacu{@cs.utcluj.ro}