Grid Based Climate Data Analysis
Kerstin Ronneberger, Stephan Kindermann, Joachim Biercamp

Background

Climate research
Is largely based on analysis of large datasets;
A typical workflow basically consists of four steps:

- Find & Select
- Collect & Prepare
- Analyse
- Visualize

Addressed problems
- The data is stored in distributed data centers (e.g. world data centers)
- No central metadata catalog based on common metadata schema
- Different data access interfaces with different AAI policies
- A common platform with uniform access and standardized data description is needed

Grid solution

C3Grid & EGEE
- In the German C3Grid Project[1] a grid-enabled data interface has been developed to uniformly find and access data of the distributed climate data centres
- By using the C3-tools together with gLite software, climate analysis workflows can be ported to the EGEE infrastructure
- In this way EGEE can offer the common platform, needed for climate research to share, access and analyse climate data

Planned Extensions
- GUI/Portal components for AMGA interactions
- Tagging of EGEE-registered files in Amga
- Integrating the visualization in the EGEE workflow
- International cooperation with NDG[2] and ESG[3]

Next steps

Open issues
- Direct transfer of external files to, and registration in EGEE
- Interoperability of the AA infrastructures of EGEE and C3

References
[1] Collaborative Climate Community Data and Processing Grid: http://www.c3grid.de

Gridification-status of German climate-data providers

<table>
<thead>
<tr>
<th>Data Centers</th>
<th>Current Volume</th>
<th>Grid enabled</th>
</tr>
</thead>
<tbody>
<tr>
<td>DKRZ Archive (MPI-M/others)</td>
<td>~4 PB ~3 TB</td>
<td></td>
</tr>
<tr>
<td>WDCs (Climate/Mare)</td>
<td>~200 TB ~5 TB</td>
<td></td>
</tr>
<tr>
<td>IFM Geomar</td>
<td>~1 TB ~100 GB</td>
<td></td>
</tr>
<tr>
<td>DWD</td>
<td>~200 GB</td>
<td></td>
</tr>
<tr>
<td>FUB</td>
<td>~1 TB</td>
<td></td>
</tr>
<tr>
<td>PIK</td>
<td>~700 GB</td>
<td></td>
</tr>
<tr>
<td>AWI</td>
<td>~300 GB</td>
<td></td>
</tr>
<tr>
<td>DLR</td>
<td>~60 GB</td>
<td></td>
</tr>
</tbody>
</table>

Data description

1. Observation Data
2. Scenario data
3. Model Data

WN
WN
WN
WN
WN
WN

EGEE

Metadata

Data Resource

LFC Catalog

SE

(b) Harvest (OAI-PMH)

(f) Register & Store data (gLite)

(c) Request (jdbc or archive)

(d) Retrieve and Preprocess

(e) Transfer (gridftp)

(g) Process (cdo-tools)

(3) Analyse (jdl job)

(1) Find & Select (Amga Java API)

(2) Collect & Prepare (webservice request)

(4) Visualize (grads)

(1) Find & Select (Amga Java API)

AMGA Metadata Catalog

UI

SE

C3Grid data interface

Planned Extensions
- GUI/Portal components for AMGA interactions
- Tagging of EGEE-registered files in Amga
- Integrating the visualization in the EGEE workflow
- International cooperation with NDG[2] and ESG[3]