Web-based Portal for Discovery, Retrieval and Visualization of Earth Science Datasets in Grid Environment

Zhenping (Jane) Liu
Outline

- Challenges to share Earth Science Datasets
- Project Goals
- Proposed Solution
- Proposed System Architecture
- Demo System
Challenges to share Earth Science Datasets

- Difficulty caused by diverse data formats
- Difficulty to discover heterogeneous and distributed datasets
- Lack of data query and retrieval services
- Difficulty of data visualization and understanding
Project goals

- To provide a Web-based Portal for Discovery, Retrieval and Visualization of Earth Science datasets with extensibility, scalability, uniformity, transparency and heterogeneity in grid environments.
Specific Project goals

- For datasets sharing, implement
  - Dynamic Discovery
  - Heterogeneity Transparency
  - Location and Name Transparency
  - Distribution Transparency
  - Replication Transparency
- Remote and Interactive web-based visualization
- Thin Clients (Web browser)
Proposed solution

- Grid Technology
- Web Services Technology
- Java/J2EE
- Scientific Visualization Technology
- Four-tier Architecture
A Layered View of Our System

Web Clients

Application Services

Grid Middleware

Resources
Grid Technology

- Controlled and coordinated sharing of geographically distributed, dynamic and heterogeneous resources.

- Grid Middleware
  - Provide fundamental infrastructure for computing and data management.
  - Permits application services to interface with the resources in a uniform way.
Web Services Technology

- Web Service is a platform and implementation independent software component that can be:
  - Described
  - Published
  - Discovered
  - Invoked
  - Composed with other services
Benefits of Web Services

- Reducing complexity by encapsulation
- Promoting interoperability
  - Truly platform and language independent
- Enabling interoperability of legacy applications
Java/J2EE (JSP, Servlet, JavaBeans)

- Web portal development
Scientific Visualization

- Represent huge amount of data graphically to help better understanding of the data
- Remote and Interactive Scientific Visualization
Proposed System Architecture

- Four-tier Architecture
  - Data Sources tier
  - Grid Services tier
  - Application Web Services tier
  - Clients tier
Grid Services Tier

- Issues addressed
  - Resource Access and Management
  - GSI Security Services
  - High Performance Data Transport Services
  - Metadata Catalog and management
  - Replica Catalog and management
Grid Services Tier (Cont.)

- **Distributed metadata catalog**
  - Stores physical and conceptual information of datasets
  - Allows managing and accessing datasets intelligently and efficiently
  - Plays a key role in the areas of managing, discovery and sharing of datasets.
Grid Services Tier (Cont.)

- **Metadata management services**
  - Metadata query, search and discovery, extraction, conversion, aggregation, validation, registration, browsing, display, and metadata schema definition.
Grid Services Tier (Cont.)

- **Distributed Replica catalog**
  - Provides mappings between logical names for files and the storage locations of one or more replicas of these files.
- **Replication management services**
- **Replication selection services**
Application Services Tier

- Datasets discovery interface generation
- Data query interface generation
- Data Retrieval
- Data Viewer
- Scientific Visualization and Analysis
  - 2D plot, 2D/3D Transform, 3D Volume Visualization.
- Future applications
Clients Tier

- Web-based Data Portal
  - All application services are delivered with web-browser
  - Key advantages to thin clients
What’ve been done

- A demo system: Web-based data management, retrieval, analysis and visualization system
- Implementing authentication and authorization web service module by using Globus Grid Security Infrastructure (GSI).
- Implementing access control web service module.
- Implementing data transfer service module by using GridFTP.
Demo system – Features

- **Web-based portal**
  - All application services are delivered with web browser.
  - Thin clients

- **Several hundred of distributed earth science data sources are integrated into the system.**

- **Several common scientific data formats supported**
Data management based on metadata mechanism

- Metadata to describe logical category of datasets
- Metadata to customize the query GUI for a dataset
- Metadata to describe logical directories (with content and semantic information) within a dataset
- Metadata to describe format and structure of a data file.
- Metadata to define available analysis methods for a dataset or a data file.
Demo system – Features (Cont.)

- Dynamically generated dataset discovery
- Web interface based on metadata
  - Sample snapshots -- next two slides
Antarctic Cooperative Research Centre, Tasmanian Partnership for Advanced Computing (TPAC)
- Digital Library for Oceans and Climate
  - TPAC/CSIRO Climatologies
  - TPAC/CSIRO Satellite Altimetry
  - NCEP - DOE Reanalysis 2
  - WOCE Global Data Version 3.0
  - Australian Antarctic Automatic Weather Station Dataset

Carolinias Coastal Ocean Observing and Prediction System (Caro-COOPS)
- Storm Surge Data

Center for Ocean Land Atmosphere Studies (COLA)
- COLA AGCM Model Data
- COLA AVN Model Data
- COLA Eta Model Data
- COLA MRF Model Data

Columbia University/LDEO - International Research Institute (IRI/LDEO)
- ARCTIC
- Biosphere - A Global Change Laboratory

Selected Datasets
TPAC/CSIRO Climatologies
Please Choose a Source Type

- NOAA Distributed Model Intercomparison Project (DMIP)
- Univ of Washington
- Hydrologic Data Systems Branch (HDSB)
- National Climatic Data Center (NCDC)
- Global Precipitation Climatology Project (GPCP)
- Global Precipitation Climatology Centre (GPCC)
- Tropical Rainforest Measurement Mission (TRMM)
Demo system – Features (Cont.)

- Dynamically generated data query web interface based on metadata
  - Sample snapshots -- next two slides
Please select the conditions for character fields:

Basin Centroids (Unit: ):

Please select the conditions for query fields:

Date (Format: "YYYY MM DD HH"): From To
Air temperature (Unit: K): From To
Surface Pressure (Unit: Pa): From To
Downward solar radiation (Unit: W/m**2): From To
Downward long-wave radiation (Unit: W/m**2): From To
Specific humidity (Unit: Kg/Kg): From To
U wind speed (positive means from west to east) (Unit: m/s): From To
V wind speed (positive means from south to north) (Unit: m/s): From To

Please customize the result format:

Date: Column
Air temperature: Column
Surface Pressure: Column
Downward solar radiation: Column
Demo system – Features (Cont.)

- Efficient data retrieval based on metadata
- Web-based data browser
  - Sample snapshots -- next two slides
Dataset: usc_MET1_met_latest.nc
latitude.time, 273886200, 273889800, 273891600, 273893400, 273895200, 273897000, 273898800,
latitude.latitude, 32.5231, 32.5231, 32.5231, 32.5231, 32.5231, 32.5231, 32.5231,
longitude.time, 273886200, 273889800, 273891600, 273893400, 273895200, 273897000, 273898800,
longitude.longitude, -77.5031, -77.5031, -77.5031, -77.5031, -77.5031, -77.5031, -77.5031,
wind_speed.time, 273886200, 273889800, 273891600, 273893400, 273895200, 273897000, 273898800,
wind_speed.wind_speed, 16.34, 14.69, 12.45, 8.16, 6.96, 8.05, 13.12, 9.12, 11.56, 19.62,
wind_from_direction.time, 273886200, 273889800, 273891600, 273893400, 273895200, 273897000,
winds_from_direction.wind_from_direction, 260.6, 279.7, 248.6, 209.4, 181.9, 143, 164.5,
<table>
<thead>
<tr>
<th>Year</th>
<th>Month</th>
<th>Day</th>
<th>Hour</th>
<th>Tair</th>
<th>Psfc</th>
<th>DSWR</th>
<th>DLWR</th>
<th>SPFH</th>
<th>U</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>1997</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>292.560</td>
<td>101391</td>
<td>0.00000</td>
<td>372.180</td>
<td>0.013000</td>
<td>12.3100</td>
<td>0.160000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>292.280</td>
<td>101451</td>
<td>0.00000</td>
<td>370.620</td>
<td>0.012500</td>
<td>10.4200</td>
<td>-2.23000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>291.440</td>
<td>101512</td>
<td>0.00000</td>
<td>369.070</td>
<td>0.012000</td>
<td>8.53000</td>
<td>-4.62000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>290.590</td>
<td>101572</td>
<td>0.00000</td>
<td>367.510</td>
<td>0.011500</td>
<td>6.64000</td>
<td>-7.00000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>289.750</td>
<td>101633</td>
<td>0.00000</td>
<td>358.460</td>
<td>0.011000</td>
<td>4.76000</td>
<td>-9.39000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>288.900</td>
<td>101693</td>
<td>0.00000</td>
<td>349.420</td>
<td>0.010600</td>
<td>2.87000</td>
<td>-11.7700</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>288.790</td>
<td>101753</td>
<td>0.00000</td>
<td>340.370</td>
<td>0.010100</td>
<td>0.980000</td>
<td>-14.1600</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>7</td>
<td>7</td>
<td>288.230</td>
<td>101867</td>
<td>0.00000</td>
<td>331.320</td>
<td>0.0099000</td>
<td>0.580000</td>
<td>-13.8300</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>287.660</td>
<td>101981</td>
<td>0.00000</td>
<td>322.270</td>
<td>0.0096000</td>
<td>0.180000</td>
<td>-13.5100</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>9</td>
<td>9</td>
<td>287.100</td>
<td>102094</td>
<td>0.00000</td>
<td>313.220</td>
<td>0.0094000</td>
<td>-0.23000</td>
<td>-13.1800</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>286.540</td>
<td>102208</td>
<td>0.00000</td>
<td>311.550</td>
<td>0.0092000</td>
<td>-0.63000</td>
<td>-12.8500</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>11</td>
<td>11</td>
<td>285.970</td>
<td>102322</td>
<td>54.27000</td>
<td>309.870</td>
<td>0.0090000</td>
<td>-1.03000</td>
<td>-12.5300</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>12</td>
<td>12</td>
<td>286.510</td>
<td>102435</td>
<td>157.730</td>
<td>308.200</td>
<td>0.0087000</td>
<td>-1.43000</td>
<td>-12.2000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>13</td>
<td>13</td>
<td>285.960</td>
<td>102438</td>
<td>250.590</td>
<td>306.530</td>
<td>0.0086000</td>
<td>-1.48000</td>
<td>-11.3300</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>14</td>
<td>14</td>
<td>285.410</td>
<td>102441</td>
<td>360.970</td>
<td>304.860</td>
<td>0.0084000</td>
<td>-1.53000</td>
<td>-10.4600</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>15</td>
<td>15</td>
<td>284.860</td>
<td>102444</td>
<td>420.450</td>
<td>303.180</td>
<td>0.0082000</td>
<td>-1.58000</td>
<td>-9.59000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>16</td>
<td>16</td>
<td>284.310</td>
<td>102446</td>
<td>451.440</td>
<td>305.100</td>
<td>0.0080000</td>
<td>-1.63000</td>
<td>-8.71000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>17</td>
<td>17</td>
<td>284.600</td>
<td>102449</td>
<td>443.090</td>
<td>307.020</td>
<td>0.0078000</td>
<td>-1.68000</td>
<td>-7.84000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>18</td>
<td>18</td>
<td>285.150</td>
<td>102452</td>
<td>413.420</td>
<td>308.930</td>
<td>0.0076000</td>
<td>-1.73000</td>
<td>-6.97000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>19</td>
<td>19</td>
<td>284.850</td>
<td>102458</td>
<td>355.720</td>
<td>310.850</td>
<td>0.0076000</td>
<td>-2.37000</td>
<td>-5.95000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>20</td>
<td>20</td>
<td>284.930</td>
<td>102464</td>
<td>239.920</td>
<td>312.760</td>
<td>0.0077000</td>
<td>-3.02000</td>
<td>-4.93000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>21</td>
<td>21</td>
<td>285.230</td>
<td>102471</td>
<td>152.080</td>
<td>314.680</td>
<td>0.0077000</td>
<td>-3.67000</td>
<td>-3.90000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>22</td>
<td>22</td>
<td>285.530</td>
<td>102477</td>
<td>54.0100</td>
<td>314.680</td>
<td>0.0078000</td>
<td>-4.31000</td>
<td>-2.88000</td>
</tr>
<tr>
<td>1997</td>
<td>1</td>
<td>23</td>
<td>23</td>
<td>285.830</td>
<td>102483</td>
<td>0.000000</td>
<td>314.680</td>
<td>0.0078000</td>
<td>4.960000</td>
<td>1.960000</td>
</tr>
</tbody>
</table>
Demo system – Features (Cont.)

- Web-based remote and interactive 2-D/3-D data visualization toolkits
  - 2-D
    - Plots, Colormaps, Contours
  - 3-D
    - Surface
  - Animation
    - Plots, Colormap, Contours, 3-D Surface
- Sample snapshots (See next slides)
Sample – 2-D Plots
Sample – Surface

precip (mm)
Date: 1995-05-01-00

View Options
- 3D
- Color Bar
- Surface
- Shaded
- Skirt
- Flat
- Gray
- Contour Lines
- Off
- Color
- Off

Animation Action
- Play
- Stop

Animation Speed (Frames Interval, Unit: Sec)
- 1
- 6

Animation Frame
- 0
- 1
Sample – 2-D Colormap

precip (mm)
Date: 1995-05-01-00

View Options

- 3D
- Color Bar
- Surface
  - Shaded
- Skirt
- Flat
- Gray
- Contour Lines
  - Off
- Color
  - Off

Animation Action

- Play
- Stop

Animation Speed (Frames Interval, Unit :Sec)

- 1
- 6

Animation Frame

- 0
- 1
GSI Authentication and Authorization web service

- Primary security mechanism in the system.
- Data Portal retrieve a proxy certificate from a MyProxy server and act on users’ behalf.
Data Transfer web service

- Allows a file to be transferred between two locations using one of several transport protocols: filesystem I/O, HTTP, FTP, HTTPS, or GridFTP.

- In the case of GridFTP, a credential is first retrieved from a MyProxy server and used to authenticate to the GridFTP server.
Access control web service

- Gets a list of access privileges of the user after querying the access rights of the user from the database.
Project web site

- **Project Introduction**

  - [http://filebox.vt.edu/eng/ece/dmv/Grid/index.htm](http://filebox.vt.edu/eng/ece/dmv/Grid/index.htm)
Thank You