Distributed Data Broker

S. Vannarat, A. Panartkul,
D. Wishadakul, J. Pengsuwan,
P. Srichaikul, R. Chitradon
NECTEC

in Collaboration with
Matthew Laurenson
Prof. Seishi Ninomiya
NARO/Japan
Outline

- Grid concept
- What is data broker
- Benefits
- Distributed data broker
- Fit to Grid concept
- Current status
What are Grids?

Electric power grid metaphor

Grid makes
- interconnected resources
- more usable and controllable
- in a seamless manner

Infrastructure VS. Concepts

Network Infrastructure
- Rigid
- Uncontrollable

Solve this in 2 hr for $10...

Transfer this in 1 min for $1...
The challenges of the Grid

• Information services about the resources available on the Grid
• Resource Brokering
• Uniform access to resources: APIs
• Security
• Job scheduling
• Data Access
• Data Replication
## Data Grid Architecture

<table>
<thead>
<tr>
<th>Category</th>
<th>Functions and Components</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>App</strong></td>
<td>Discipline-Specific Data Grid Application</td>
</tr>
<tr>
<td><strong>Collective (App)</strong></td>
<td>Coherency control, replica selection, task management, virtual data catalog, virtual data code catalog, ...</td>
</tr>
<tr>
<td><strong>Collective (Generic)</strong></td>
<td>Replica catalog, replica management, co-allocation, certificate authorities, metadata catalogs, ...</td>
</tr>
<tr>
<td><strong>Resource</strong></td>
<td>Access to data, access to computers, access to network performance data, ...</td>
</tr>
<tr>
<td><strong>Connect</strong></td>
<td>Communication, service discovery (DNS), authentication, authorization, delegation</td>
</tr>
<tr>
<td><strong>Fabric</strong></td>
<td>Storage systems, clusters, networks, network caches, ...</td>
</tr>
</tbody>
</table>

From: Grid Architecture
Weather Data

- Agriculture applications: prediction & planning
- Data sources: Files, SQL databases
- Data organization:
  - by station
  - by element
  - by location
  - by each resolution (hourly, daily)
MetBroker

Number of code modules proportional to $n \times m$

Number of code modules proportional to $n + m$
Benefits of Mediated Architecture

• Single API
• No change on database server.
• Metadata
• Simple client
• Network security
Current status

• Implemented
  – using RMI & Web Services for communication
  – JDBC, HTTP etc for database access
• Currently running on Linux host
• http://www.agmodel.net/MetBroker/MetBroker.html
• Databases: Japan, New Zealand, Georgia, UK
Distributed Broker

Advantages

- High availability
- Optimize network utilization
- …
Distributed Broker

High availability
- Alternative paths

Optimize network
- Choose most efficient path
- Data re-routing and re-directing
Architecture

<table>
<thead>
<tr>
<th>Application</th>
<th>Application side AS API</th>
<th>Application side DS API</th>
</tr>
</thead>
</table>

AS

- AP Interface
- MB Interface
- MetBroker

DS

- Inter Broker Data Pipe
- DS Interface
- Meta data extractor
- Discoverer and API
- QoS Probe
- Meta data + QoS Exchange
- Meta data
- QoS

GridCA
Broker in Grid Architecture

- **App**
  - Collective (App)
    - Coherency control, replica selection, task management, virtual data catalog, virtual data code catalog...

- **Collective (Generic)**
  - Replica catalog, replica management, co-allocation, certificate authorities, metadata catalogs...

- **Resource**
  - Access to data, access to computers, access to network, performance data...

- **Connect**
  - Communication, service discovery (DNS), authentication, authorization, delegation...

- **Fabric**
  - Storage systems, clusters, network caches, network...

- **Network**
  - ...
Dev. Phases

Function

- Access Services
- Discovery Services
- Redirect
- Reroute
- QoS
- CA

Time line

Sept. 03

Sept. 05
Summaries

• Grids concept: seamlessly usable interconnected resources
• Network of MetBrokers
• Fit to Grid Architecture
  – Use Grid components
  – Use Grid infrastructure
• (Distributed) Data Broker -> Data Grid