Korea e-Science Project

Dr. Sang Boem Lim
Supercomputing Center at KISTI
I  Introduction

II  e-AIRS (Aerospace Integrated Research Systems)

III  High Voltage Electron Microscope (HVEM) Grid System

IV  Meteorology e-Science

V  HG2C Project
   (Human Genome to Chemicals for Drug Discovery)

VI  e-Glycoconjugates: Integrated Grid Portal for the Molecular Simulations of Glycoconjugates

VII  e-Science Infrastructures
What is Grid?

Technologies and infrastructure that support sharing and coordinated use of diverse resources in dynamic, distributed virtual organizations (VO’s) – Ian Foster
Korea e-Science Project

Definition

National e-Science Project in Korea is about providing innovatively enhanced research infrastructure which enables researchers to use nationwide distributed R&D resources such as, high performance computing, databases, scientific instruments, and also human resources in research areas as a linkage system in cyberspace by using high-end information technology.

Purposes

- **Innovation of R&D environment and international competitive power in Science**
  - Rapid changes in paradigm of S&T into the global collaboration by e-Science
  - To participate in the standardization of global e-Science environment near future

- **Infrastructure of e-Science helps the improvement of the various fields of R&D activities in Korea**
  - Effective investment of R&D activities by sharing the nationwide R&D resources
  - A balanced development of national infrastructure for the S&T by the advancement of the traditional R&D activities by using e-Science

- **To achieve successful e-Science project by the use of the highly developed information technologies and network infrastructure in Korea and to do the good outreach to the benefits of common society**
Korea e-Science Project

Goals

- To advance research infrastructure and national competitiveness by effectively using distributed resources using information technology
  - Making the best use of Korea’s IT infrastructure
- To providing advanced collaborative environments to researchers distributed over the country
  - Helping the decentralization

Korean Situation

- Construction of Grid infrastructure is being constructed through “National Grid Project” of Ministry of Information and Communication
  - Limited to the construction of next generation internet
  - Support for application projects are limited, only small scale projects are supported
  - National scale application projects are not possible
- Sharing and remote use of expensive equipments in government labs and academic institutions are very limited
Korea e-Science Project

Goals

Improving productivity of R&D by e-Science

- 2010: To deploy the world-class e-Science infrastructure
- 2007: "Anytime, anywhere" accessible e-Science infrastructure
- 2006: Full-scale developing common SWs for e-Science
- 2005: Preparation for building the e-Science infrastructure

Summary

- Goal: To develop and deploy common SWs for e-Science
- Periods: 1st Phase (KFY 2005 ~ 2007)
  : 2nd Phase (KFY 2008 ~ 2010)
- Budget: 100 M USD for developing common softwares
- Supervisor: MOST (Developer: KISTI)
Currently, we selected 5 areas of application for e-Science project.

Each application will be developed based on common software and each group will contribute to the common software.

<table>
<thead>
<tr>
<th>Area</th>
<th>Title</th>
<th>Organization</th>
<th>Organization</th>
<th>Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>BT, NT</td>
<td>Development of Molecular Simulation E-Science Research Environment and e-Glycoconjugates</td>
<td>Konkuk University</td>
<td>BT, NT</td>
<td></td>
</tr>
<tr>
<td>BT</td>
<td>Development of e-Science Environment for HG2C (Human Gene to Chemical) based on Service-Oriented Architecture</td>
<td>Soongsil University</td>
<td>BT</td>
<td></td>
</tr>
<tr>
<td>Aerospace</td>
<td>Construction of Numerical Wind Tunnel on the e-Science Infrastructure</td>
<td>Seoul National University</td>
<td>Aerospace</td>
<td></td>
</tr>
<tr>
<td>Equipment</td>
<td>Establishment of e-Science environment using the high voltage electron microscope</td>
<td>KBSI(Korea Basics Science Institute)</td>
<td>Equipment control</td>
<td></td>
</tr>
<tr>
<td>Meteorology</td>
<td>Construction of e-Science Environment for Weather Information System</td>
<td>Pukyoung University</td>
<td>Meteorology</td>
<td></td>
</tr>
</tbody>
</table>
e-AIRS (Aerospace Integrated Research Systems)

- The main goal of e-AIRS is to establish the powerful and user-friendly collaboration environment to aerospace researchers.
- Through the user-friendly e-Science grid portal system, the e-AIRS provides remote CFD (Computational Fluid Dynamics) calculation and experiment management system.
- KISTI, SNU,
**e-AIRS (Aerospace Integrated Research Systems)**

<table>
<thead>
<tr>
<th>CFD Framework</th>
<th>Remote Wind Tunnel Exp. Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Fortran / Cactus CFD solvers</td>
<td>- Wind tunnel tests by accurate PIV (Particle Image Velocimetry) system</td>
</tr>
<tr>
<td>- Multi-block &amp; automatic parallel computation</td>
<td>- Portal interface: remote request and monitoring service on the portal service</td>
</tr>
<tr>
<td>- Accurate and efficient numerical schemes</td>
<td></td>
</tr>
<tr>
<td>- Euler / N-S / turbulence problems</td>
<td></td>
</tr>
<tr>
<td>- Design optimization</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collaboration System on the Access Grid</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>- Access Grid: Video and audio conference</td>
<td></td>
</tr>
<tr>
<td>- Data-sharing system between individual / group researchers</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Integrated Research Environment</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>- CFD computation service by fortran / Cactus solver</td>
<td></td>
</tr>
<tr>
<td>- Portal interface: independent of time &amp; place</td>
<td></td>
</tr>
<tr>
<td>- Remote management system of wind tunnel experiment</td>
<td></td>
</tr>
<tr>
<td>- CFD/experimental data-comparison system</td>
<td></td>
</tr>
<tr>
<td>- Collaboration environment using Access Grid</td>
<td></td>
</tr>
</tbody>
</table>
e-AIRS (Aerospace Integrated Research Systems)

- Portal Service
  - CFD calculation service – mesh generation and job monitoring
  - Remote experimental service – management of requests and monitoring service of the wind tunnel experiment
  - Simulated result correctness verification system
e-AIRS (Aerospace Integrated Research Systems)

**Expected Results**

- Because the expensive parallel computing resources and wind tunnel can be shared and any researcher can access these resources, the boundary of research topics will be enlarged.
- The R&D time and cost will be reduced and the efficiency will be increased.
- With the more powerful collaboration system, many individual and group researchers are able to establish large size projects and share their research products.
- Web portal service will provide the convenient research environment independent of time and space.
- Portal GUI will make it easy to access the research devices. This will make the non-experts produce their own research data more conveniently.
High Voltage Electron Microscope (HVEM) Grid System

- To improve research performance using the global research network
- To share and use the costly High Voltage Electron Microscope (HVEM) efficiently
- To maximize the synergy effect achieved by databases of experimental data
High Voltage Electron Microscope (HVEM) Grid System

- **Remote Control**: HVEM control service provides remote control of HVEM, goniometer, and CCD camera via encapsulated web service.

- **Image Handling**: Streaming service provides real-time streaming from CCD camera and capturing service enables the images streamed to be captured and stored.

---

Image from the HVEM Grid System:

- **Presentation Tier**
  - Client App.

- **Application Server Tier**
  - Grid Service Server
  - WebService for HVEM Control App.

- **Data Server Tier (Legacy App.)**

- **HVEM**

---

Data from the HVEM Grid System:

- **Client Program**
  - Stream Data to Server
  - Capture Images
  - Store Images
  - Present Images

---

Client App.

Application Server

WebService

HVEM

Presentation Tier

Tier

Multi Tier

Multi-Tier

Application Server Tier

Data Server Tier

Legacy App.

---

Data Management Server

WebService

Data Turbine (RISKE)

DM Library

Digital Micrograph

---

Grid Service

Web Service

HVEM

---

High Voltage Electron Microscope (HVEM) Grid System

HVEM Remote Access System GUI

HVEM Goniometer Remote Control GUI
High Voltage Electron Microscope (HVEM) Grid System

Expected Results

- Practical Virtual Laboratory
- Efficient Usage of HVEM
- Resource Sharing & Distributed Collaboration

Maximization of Research Performance!!
Meteorology e-Science

- Efficient and Standardized Data Access
- Intelligent Data Management of SAM file Pool
- Dynamic Cataloging
Meteorology e-Science

- **Data Management Service**
  - Compliant with Data Grid
  - On-Demand Cataloging on GDS (Grads Data Server)

- **MM5 Workflow**

- **Output Display Service**
  - GrADS Plug-in on Grid Sphere

- **Resource Management Service**
  - Computing Resource monitoring & discovery

- **Peta-Bytes Data Pool Handling**
  - Meaningful Data Representation from huge size of files
Meteorology e-Science

Expected Results

- Efficient Data Management (File & DB combination)
- Efficiency of Storage using Dynamic Cataloging
- Easy Access & Quick Searching to Peta-Bytes
- Weather Data
- Interoperability between LAS (Live Access Server) and GDS (Grads Data Server)
- Changing to Semantic Data Management
**HG2C Project**  
*Human Genome to Chemicals for Drug Discovery*

- Construct an HG2C Portal that integrates various BT applications for efficient drug discovery

![Diagram showing Bioinformatics and Cheminformatics](image)

- **Genome DB**  
  30,000 sequences (Human)

- **Chemome DB**  
  20,000,000 compounds

- **BT Application S/W**
  - **IDPro**
    - Predict protein structure
  - **IDPharmo (PharmoMap + PharmoScan)**
    - PharmoMap: Define active sites and generate multiple 3d feature maps
    - PharmoScan: Virtual screening
HG2C Project
(Human Genome to Chemicals for Drug Discovery)

HG2C Workflow

- Human Genome Sequence
  (3 billion base, 30,000 gene)
- PDB: 32,355 structure
- Scop, Cath, FSSP: protein fold database

- Sequence Alignment (Blast)
  - Using multiple alignment
  - Decide identity, and number of reference

- Protein Structure Prediction (IDPro)
  - Minimize level, MD level, rand level

- Comparative modeling
  - Threader
  - Ab initio etc

Identify Target Protein → Active Site Analysis → Compound Selections → Generate Pharmacophore

BLAST → IDPro → PharmoMap

PharmoScan
HG2C Project
(Human Genome to Chemicals for Drug Discovery)

- **Meta Services**
  - Define a part of a workflow as a new service
    - Workflow instance is declared as a workflow unit in the service description
    - By overriding some attributes of a workflow unit,
      - Pass parameters to the workflow unit
      - Setup service specific information
    - The new service can be wrapped to a Web service or a Grid service, and reuse it easily
  
  - Specify service specific information
    - Restrict resources to allocate a specific service (user preference)
    - Scheduling priority

- **MSF (Meta Services Framework)**
  - Meta services concept is combined with WfMS system
    - Provide more reusable and adaptable workflow management environment
    - Adapt Meta services to various service environments such as Web services, portal services, and Grid services
  
  - Workflow description is divided into
    - MSF_Service, MSF_Flow, and MSF_Task
  
  - Consist of five small agents
    - AM, RM, EM, OM, and SM
    - Increase flexibility and adaptability
  
  - Ontology concept is applied to describe services, flows, and tasks
    - Advanced searching and sharing of the description is possible
HG2C Project
(Human Genome to Chemicals for Drug Discovery)

ouses MAGE (Modular and Adaptive Grid Environment)

- Modular and Adaptive Grid Environment
  - Provide API for easy development of Grid application
  - Provide transparency to end-users and developers
    - Protocol transparency
    - Running location transparency
    - Implementation details transparency
  - Provide flexible and reliable layers

- Consist of 3 major components
  - Service Manager
  - Component Manager
  - Reconfiguration Manager
e-Glycoconjugates: Integrated Grid Portal for the Molecular Simulations of Glycoconjugates

- The e-Science Grid portal system of e-Glycoconjugates was motivated to study whole range of glycoconjugates with web-interface.
- It employs an innovative architecture to execute the molecular simulation, to analyze the simulation trajectory and to share the simulation data with all other world users.
**e-Glycoconjugates: Integrated Grid Portal for the Molecular Simulations of Glycoconjugates**

**Process Workflow**
- Our problem solving process can be classified into three steps as follows:
  1. Input script programming & Job submit
  2. Multiple job simulations on the Grid computing system
  3. Trajectory analysis & Text data processing

**Trajectory Processing**
- The analysis facility of e-Glycoconjugates will support the automatic trajectory processing to obtain valuable information on the hydrogen bonding, hydration number, translational diffusion, rotational motion and radial distribution of water molecules around carbohydrates or glycoconjugate molecules.

---

**Input Script**
- OPEN Parameter File
- OPEN Molecular Topology File
- Molecular Structure Build
- Energy Minimization
- DYNABIO MD
- WRITE Trajectory

**Computing**
- Grid-Based MD Simulations
- Multiple Job Submission
- Web-based Job Management

**Data Output**
- Trajectory Analysis
- Graphic & Text Data Processing

**Process workflow. Human check points were minimized to run a automatic job simulation**

**Data Analysis Grid**

**Trajectory output and E,R-diagram for DB construction**
**e-Glycoconjugates: Integrated Grid Portal for the Molecular Simulations of Glycoconjugates**

- A web portal for e-Glycoconjugates has a graphic user interface (GUI)
  - create a job,
  - submit a job to the computing grid,
  - save and retrieve data to the database on the data grid
- Supports various search methods on the molecular simulations,
  - queries by keyword, structure, classification or by combining or all these factors.
- Full support for the name searching and automatic analyses are now in progress.
e-Glycoconjugates: Integrated Grid Portal for the Molecular Simulations of Glycoconjugates

Bio-Information Production

- We have been carrying out conformational mapping of more than 1,000 carbohydrate chains to serve dynamic structure DB containing a wide range of glycans of which three-dimensional structures are known or unknown.
- The effects of glycan chain structure on the protein folding and stability are also going in a progress.
- The e-Glycoconjugates performs the molecular simulations on the more than 100 biologically important glycoprotein such as prion, collagen, antifreeze peptide or ribonuclease.
- Understanding for the structural dynamics of glycoconjugates is a prerequisite for the opening in personalized medicine. We are convinced that valuable systematic information from e-Glycoconjugates is able to offer theoretical background to the structure-based design of novel glycoconjugates.

Bio-simulation Job - Grid Computing Engines - Information

Bio-information production on the biological glycoconjugates from the Grid computing data
Korea e-Science Project

**Domestic Research Network**

- **Area**: High-speed Research Network (KREONET) Giga backbone deployment and services
- **Characteristic**: Seoul-Daejeon(10Gbps), 12 local network centers (5 and 2.5Gbps)
- **Integrated monitoring system service (24 hours)**, Apply next-generation technology (IPv6, QoS, Multicast)
- **Results**: High-end application (e-Science, Grid) needs based Infrastructure
Korea e-Science Project

Global Ring Network for Advanced Application Development

- Definition: The world's first global science & technology network connecting the continent with 10Gbps ring-type lambda network (Funded by Korea, US, Russia, and China)

- Application: Serve as international collaborative network for high-end science and technology areas such as high-energy physics, bioscience and nuclear fusion that require real-time large scale data transfers
Korea e-Science Project

Cyber R&D infrastructure

- e-Science
- e-Science Infrastructure
- National Grid infrastructure
- High-end S&T Networks
- GLORIAD
  - S&T Information Systems
  - Supercomputing Infrastructure
  - Collaborative Utilization of Equipments
  - Global Collaboration
Yes KiSTi

“과학·기술·산업정보”
Korea Institute of Science and Technology Information
한국과학기술정보연구원