KISTI Grid R&D Activities in Korea

April 23, 2009

KISTI e-Science Division
Soonwook Hwang
Contents

- K*Grid Project
- National e-Science Project
- 4th Supercomputer/KREONET/GLORIAD
- Recent Grid R&D Developments in Collaboration with EGEE
## K*Grid Project

<table>
<thead>
<tr>
<th>Goal</th>
<th>Construction of a National Grid infrastructure in Korea</th>
</tr>
</thead>
</table>
| Period & Budget | • Period: 2002 ~ 2006 (5 Years) + 2007 ~ 2008 (2 Years)  
                • Total Budget: 32 (+3) M US dollars |
| Funding Agency | Ministry of Information and Communication (MIC)        |
| Coordinator   | Korea Institute of Science and Technology Information (KISTI) |
| Partners      | Various research partners selected from academia, industry, and government lab. through a public competition |
Major Achievements

Building TIGRIS (Tera-scale Infrastructure for K*Grid Service)

- To build a tera-scale Grid computing environment permanently available
  - Connection of sizable supercomputing resources (KISTI, Seoul National Univ., Busan National Univ.)
  - Development of web-based Grid service platform and providing services
  - Operation of Grid Certification Authority (CA)

Development of Grid Middleware Package: KMI-R1

- Development of Grid middleware for constructing the next generation IT infra
  - Support and expand Grid communities in Korea through KMI-R1 release
  - International collaboration with Globus, PRAGMA, etc

Discovery of Grid Applications linked with IT

- Applying Grid Technology to IT Applications
  - Discovery and Implementation of Prototypes with Industries (On-line Game Service, Telematics Service, Rendering Service)
  - Creation of Business Grid Market in IT
# National e-Science Project

<table>
<thead>
<tr>
<th><strong>Goal</strong></th>
<th>Establish a world-class e-Science environment by 2011, dramatically improving research productivity</th>
</tr>
</thead>
</table>
| **Period & Budget** | - Period: 2005 ~ 2007 (3 Years)  
- Total Budget: 7 M US dollars |
<p>| <strong>Funding Agency</strong> | Ministry of Education, Science and Technology |
| <strong>Coordinator</strong> | Korea Institute of Science and Technology Information (KISTI) |
| <strong>Partners</strong> | Universities, Industry and government lab. |</p>
<table>
<thead>
<tr>
<th>Project</th>
<th>Institute</th>
<th>Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>e-AIRS (Aerospace Integrated Research System)</td>
<td>SNU, SMWU</td>
<td>- CFD computational service by Fortran/Cactus solver</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portal interface</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Remote wind tunnel experiment management</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- CFD experimental data-comparison system</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Collaboration environment using Access Grid</td>
</tr>
<tr>
<td>Establishment of e-Science environment using HVEM (High Voltage Electron Microscope)</td>
<td>KBSI, SNU</td>
<td>- Change from Gt2.x + gridport2 to Gt4.x + Gridsphere</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Possible to storage file and other contents</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Manage DV data, not analog data</td>
</tr>
<tr>
<td>Data Grid for Meteorology e-Science</td>
<td>PKNU, KHU</td>
<td>- Establishing Gridsphere based Grid portal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Meta data cataloging</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Transmission, access service, system monitoring</td>
</tr>
<tr>
<td>HG2C Research Environment</td>
<td>SSU, Yonsei Uvi.</td>
<td>- Construct HG2C portal to seek active agents from the gene</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Develop BT Application S/W to execute HG2C</td>
</tr>
<tr>
<td>Integrated e-Science Portal for the Molecular Simulations of Glycoconjugates</td>
<td>KKU</td>
<td>- Construct simulation database for Bio/Nano applications</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Execute e-Quanta</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Develop molecular simulation research infrastructure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Develop the optimized MGrid on Bio research</td>
</tr>
</tbody>
</table>
KISTI’s 4th Supercomputing Facilities

- KISTI will be placed in the TOP 10 out of the TOP 500 list of the most powerful supercomputers
  - 300 TFLPOS of Computer nodes with 21,504 CPU cores
  - 1.2 PB of Disk Storage
  - 2.4 PB of Tape Storage

<table>
<thead>
<tr>
<th>Year</th>
<th>System</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1988</td>
<td>Cray C90 [2nd]</td>
<td>2GFlops</td>
</tr>
<tr>
<td>1993</td>
<td>16GFlops</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>NEC SX-5[3rd system]</td>
<td>131GFlops</td>
</tr>
<tr>
<td>2000</td>
<td>HP GS320 HPC160/320</td>
<td>242GFlops</td>
</tr>
<tr>
<td>2001</td>
<td>PC Cluster 128node</td>
<td>306GFlops</td>
</tr>
<tr>
<td>2002</td>
<td>IBM p690[3rd]</td>
<td>1,407GFlops</td>
</tr>
<tr>
<td>2003</td>
<td>IBM p690+[3rd]</td>
<td>8,000GFlops</td>
</tr>
<tr>
<td>2008</td>
<td>SUN [4th System 1]</td>
<td>30,000GFlops</td>
</tr>
<tr>
<td>2009</td>
<td>KISTI 2009</td>
<td></td>
</tr>
</tbody>
</table>

- (1997 ~ 2003) 115GFlops
- (2001. 5. ~ 2008. 9) 80GFlops
- (2003. 2. ~) 160GFlops
- (2003. 12 ~ 2008. 9) 2,850GFlops
- (2008. 6. ~) 24,000GFlops
- (2007. 10. ~) 6,000GFlops

KISTI’s 4th Supercomputing Facilities

- KISTI will be placed in the TOP 10 out of the TOP 500 list of the most powerful supercomputers
  - 300 TFLPOS of Computer nodes with 21,504 CPU cores
  - 1.2 PB of Disk Storage
  - 2.4 PB of Tape Storage

<table>
<thead>
<tr>
<th>Year</th>
<th>System</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1988</td>
<td>Cray C90 [2nd]</td>
<td>2GFlops</td>
</tr>
<tr>
<td>1993</td>
<td>16GFlops</td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td>NEC SX-5[3rd system]</td>
<td>131GFlops</td>
</tr>
<tr>
<td>2000</td>
<td>HP GS320 HPC160/320</td>
<td>242GFlops</td>
</tr>
<tr>
<td>2001</td>
<td>PC Cluster 128node</td>
<td>306GFlops</td>
</tr>
<tr>
<td>2002</td>
<td>IBM p690[3rd]</td>
<td>1,407GFlops</td>
</tr>
<tr>
<td>2003</td>
<td>IBM p690+[3rd]</td>
<td>8,000GFlops</td>
</tr>
<tr>
<td>2008</td>
<td>SUN [4th System 1]</td>
<td>30,000GFlops</td>
</tr>
<tr>
<td>2009</td>
<td>KISTI 2009</td>
<td></td>
</tr>
</tbody>
</table>

- (1997 ~ 2003) 115GFlops
- (2001. 5. ~ 2008. 9) 80GFlops
- (2003. 2. ~) 160GFlops
- (2003. 12 ~ 2008. 9) 2,850GFlops
- (2008. 6. ~) 24,000GFlops
- (2007. 10. ~) 6,000GFlops
KREONET Backbone

Map of South Korea showing the network connections and bandwidth capacities among various cities. The map highlights the backbone network with SuperSIReN and its connections to major cities like Seoul, Incheon, Suwon, Cheonan, Daejeon, Daegu, Changwon, Busan, Jeonju, and Jeju. The network supports data rates of 20 Gbps, 10 Gbps, and 5-10 Gbps.
KREONET

- Korea’s national science & research network
  - Funded by MOST since 1988
  - Lambda-based gigabit backbone
- Around 200 linked organizations
  - Universities, Government lab, Industry lab
  - 5-20Gbps backbone
- Being upgraded to KREONet2
  - Hybrid optical and packet switching facility
  - Native IPv4, IPv6 and lightpath provisioning
    - Routed path and lightpath over a single link
    - Mbone
    - IPv6 Gigabit Network
- Connected to GLORIAD for international connection
  - The first round-the-world high-performance backbone network
GLORIAD (GLObal RIng Network for Advanced Applications Development): the first round-the-world high-performance network jointly established with 10 Gbps optical networking tools

- Seattle (USA) – Calgary, Toronto (Canada) – Chicago, NYC (USA) – Amsterdam (the Netherlands) – Moscow, Novosibirsk, Khavarovsk (Russia) – Beijing, Hong Kong (China) – Daejeon (Korea) – Seattle

- KISTI is in charge of Two 10Gs: KR to US and KR to CN (HK)
KISTI and EGEE

- an official partner since the 2nd phase of EGEE, April 1st 2006
- participating in SA1 and NA4 activities in EGEE-III
  - SA1 (Grid Operation and Management)
    - Grid Operation in close collaboration with ASGC
    - Operation of a ALICE Tier-2 center working with CERN’s ALICE computing team
  - NA4 (Grid Application Identification and Support)
    - Participating in the WISDOM project
      - International initiative for the deployment of large-scale in-silico docking on the grid
KISTI ALICE Tier2 Center

- KISTI is a certified EGEE site.
- Signing of WLCG MoU for the ALICE Tier-2 center (’07.10.23)
- Funded by KICOS (Korea Foundation for International Cooperation of Science & Technology)
  - 150,000 US Dollars/year
1.2% contribution to ALICE computing in the total job execution

Currently, 128 CPU cores and 30 TB storage provided for ALICE computing

Processing near 8000 jobs per month in average
Recent Developments

- **In WISDOM**
  - Performed a data challenge for human diabetes type II with a significant performance improvement in WISDOM PE
  - Prototyped a WISDOM GUI client, DrugScreener-G

- **In AMGA**
  - Native SQL support in AMGA 1.9
  - WS-DAIR support in AMGA 2.0
The WISDOM Project

- International initiative to deploy a large-scale in-silico docking on a public grid infrastructure
- An attempt to find potential drugs against neglected or emerging diseases
  - e.g., Malaria, Avian Flu
- Previous deployments were either limited to
  - a single cluster, or
  - grids of clusters in the tightly protected environment of a pharmaceutical laboratory
- Since 2005, four large-scale deployments (data challenge) done on the EGEE infrastructure within the framework of WISDOM collaboration

![Diagram showing deployments 2005 to 2008 with targets and diseases]
Data challenge against Diabetes Type II with the improved WISDOM PE

- Data challenge to Human pancreatic amylase inhibitor (1u2y), a target protein of diabetes type 2, with 308310 chemical compounds
- Achieved significant improvement in efficiency and throughput

<table>
<thead>
<tr>
<th></th>
<th>WISDOM-II</th>
<th>DIANE</th>
<th>KISTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of dockings</td>
<td>156,407,400</td>
<td>308,585</td>
<td>308,310</td>
</tr>
<tr>
<td>Estimated duration on 1 CPU</td>
<td>413 years</td>
<td>16.7 years</td>
<td>39.0 years</td>
</tr>
<tr>
<td>Duration of the experiment</td>
<td>76 days</td>
<td>30 days</td>
<td><strong>2.4 days</strong></td>
</tr>
<tr>
<td>Cumulative number of Grid jobs</td>
<td>77,504</td>
<td>2,580</td>
<td>103,583</td>
</tr>
<tr>
<td>Maximum number of concurrent</td>
<td>5000</td>
<td>240</td>
<td>7,370</td>
</tr>
<tr>
<td>CPUs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of used Computing</td>
<td>98</td>
<td>36</td>
<td>127</td>
</tr>
<tr>
<td>Elements</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Crunching Factor</td>
<td>1983</td>
<td>203</td>
<td><strong>5937</strong></td>
</tr>
<tr>
<td>Distribution Efficiency</td>
<td>39%</td>
<td>84%</td>
<td><strong>81%</strong></td>
</tr>
</tbody>
</table>
DrugScreener-G

- DrugScreener-G: Integrated Environment for Grid-enabled Large-scale Virtual Screening
  - Aiming at providing a **user-friendly integrated environment for Grid-enabled large-scale virtual screening** for **users without much knowledge of Grid computing** to exploit full power of Grid computing infrastructure for drug discovery
  - Target users: Bioinformaticians, Biologists, Drug Chemists
AMGA

- An official EGEE gLite middleware component for a metadata catalogue service on the grid
- AMGA provides:
  - Access to metadata for files distributed on the grid
  - A simplified general access to relational data stored in database systems
- AMGA has been used in many different user communities
KISTI contributions to recent developments in AMGA

- AMGA 1.9 released on Oct. 6, 2008
  - Native SQL support
  - Multi-threaded Server with connection pooling

- Currently preparing AMGA 2.0
  - To be released in this Summer
  - WS-DAIR (Web Service Data Access and Integration on Relational Database)
    - OGF Standard proposed by the OGF DAIS-WG
    - Stability of WS-DAIR server needs to be improved before release
KISTI’s contribution to AMGA development – (1)

Native SQL support

- **Background**
  - Biomedical community asks the support for native SQL queries in AMGA
    - To ease the work needed to port existing SQL-based database applications to the Grid using AMGA
    - To complement the existing metadata query language in AMGA

- **AMGA allows queries now in native SQL**
  - Support for entry level SQL 92 done
    - SELECT, UPDATE, INSERT, DELETE
  - Some SQL 92 intermediate level supported
    - OUTER JOIN

- **All queries are subject to AMGA access restrictions**
  - All queries are parsed by AMGA, AMGA “understands” security implications
  - ACLs for tables
KISTI’s contribution to AMGA development – (2)

WS-DAIR support

- WS-DAIR
  - OGF proposed recommendation specification for data access to relational DB's on the Grid
- Background
  - Enable interoperability to facilitates sharing data
    - AMGA uses its own protocol and message patterns
  - Integration of WS-DAIR in AMGA will make AMGA a relational data source in a WS-based environment!
- AMGA-DAIR Integration
  - Implementation of direct access and indirect access done.
    - Data returned in Sun's WebRowSet specification.
  - All queries use SQL.
  - Clients written in C++ (gSOAP)
Thank you for your attention!

謝謝