Pacific Rim Application and Grid Middleware Assembly: PRAGMA

A NSF Supported community building collaborations and advancing grid-based applications

William Y. B. Chang, wychang@nsf.gov
Office of International Science and Engineering
National Science Foundation
U. S. A.

May 28, 2004

http://www.pragma-grid.org
Founding Motivations

- Science is an intrinsically global activity
- The grid is transforming computing and collaboration
- The problem remains that the grid is too hard to use
- Middleware software needs to interoperate
Overarching Goals

Establish sustained collaborations and

Advance the use of the grid technologies for applications among a community of investigators working with leading institutions around the Pacific Rim

Working closely with established activities that promote grid activities or the underlying infrastructure, both in the Pacific Rim and globally.
Participating Institutions

- Australia Partnership for Advanced Computing and its partners
- Bioinformatics Institute of Singapore, part of Agency for Science and Technology and Research
- Computer Network Information Center, Chinese Academy of Sciences
- Global Scientific Information and Computing Center, Tokyo Institute of Technology
- Grid Technology Research Center, National Institute of Advanced Industrial Science and Technology
- Korea Institute of Science and Technology Information
- National Center for High Performance Computing
- Research Center for Ultra-High Voltage Electron Microscopy and the Cybermedia Center, Osaka University
- STAR TAP/StarLight initiative, supported by NSF and organized by the University of Illinois at Chicago, Northwestern University and Argonne National Laboratory
- Thai Social/Scientific Academic and Research Network (ThaiSARN-3), National Electronics and Computer Technology Center
- TransPAC initiative, supported by NSF at Indiana University
- Universiti Sains Malaysia
- University of California, San Diego and SDSC, CalIT², CRBS, NLANR
- University of Hyderabad

Accepted Operating Principles and Procedures 25 Feb 2003
## Resource End Points

<table>
<thead>
<tr>
<th>Organization</th>
<th>Resource</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monash University</td>
<td>24 Node cluster</td>
</tr>
<tr>
<td>CNIC Chinese Academy of Sciences</td>
<td>8 Node AIX (Power 3) cluster</td>
</tr>
<tr>
<td>AIST</td>
<td>66+10 Linux</td>
</tr>
<tr>
<td>Titech</td>
<td>16 Linux</td>
</tr>
<tr>
<td>Osaka University</td>
<td>12 node Linux (P3) cluster</td>
</tr>
<tr>
<td>KISTI Supercomputer Center</td>
<td>80 node Linux</td>
</tr>
<tr>
<td>USM</td>
<td>16+5+10 Linux</td>
</tr>
<tr>
<td>BioInformatics Institute</td>
<td>32 nodes Linux</td>
</tr>
<tr>
<td>NCHC</td>
<td>32 node cluster</td>
</tr>
<tr>
<td>NECTEC</td>
<td>24 node cluster</td>
</tr>
<tr>
<td>SDSC</td>
<td>16 node Linux (P4) cluster</td>
</tr>
<tr>
<td>TransPAC</td>
<td>15 node cluster</td>
</tr>
</tbody>
</table>

Compiled Nov 02 – Jan 03 Source: Kishore Sakharkar (BII)
Activities

- Encourage and conduct joint (multilateral) projects that promote development of grid facilities and technologies
- Share resources to ensure project success
- Conduct multi-site training
- Exchange researchers
- Meet and communicate regularly
- Collaborate with and participate in major regional and international activities such as APAN, APGrid, APECTel, GGF
- Disseminate and promote knowledge of using the grid among domain experts and scientists
- Disseminate proceedings and summaries of events
- Provide resource for PRAGMA members to raise level of awareness and funding for grid activities
Schedule of Meetings

• PRAGMA 4: 4-5 June 2003, Melbourne, Australia
  – ICCS2003: 3-4 June
  – David Abramson (APAC): Chair; Co-chair: Fang-Pang Lin (NCHC)

• PRAGMA 5: 22-23 October 2003, Hsinchu/Fushan, Taiwan
  – Fang-Pang Lin (NCHC): Chair; Co-chair: Kai Nan (CNIC)

• PRAGMA 6: 16 – 18 May 2004, Beijing, China
  – Baoping Yan (CNIC): Chair; Co-chairs: Mason Katz (UCSD), Jim Williams (TransPAC)

• PRAGMA 7: Mid September 2004, San Diego, USA
  – Chairs: Mason Katz (UCSD), Jim Williams (TransPAC)
Expected Outcomes

• Advance scientific applications
• Contribute to the international grid development efforts
• Increase interoperability of grid middleware in Pacific Rim and throughout the world
• Increase productive and effective use of the grid by researchers and scientists in the Pacific Rim
• Increase multi-lateral scientific collaboration on the grid in the Pacific Rim
• Increase grid activities within Pacific Rim
• Create grid testbeds for regional e-science projects
Summary of PRAGMA SC’03 Demos

• Total 22 Abstracts: Accomplishments or Promotions
• Most involve 3 or more partners
• Every members involved in at least one demo.
• 13 institutions present, 9 have booths
• Technologies: Ninf-G, Nimrod, Grid Datafarm, MGrid, Grid Server Broker, Optimization Algorithms, Lattice Data Grids, Telescience
• Applications: Quantum Chemistry, Molecular Energy Calculations, Astronomy, Climate, Molecular Biology, Structural Biology, Ecology and Environment, SARS Grid, Neuroscience
PRAGMA Success Stories

• Grid Community Pulls together to Battle SARS

• Merging Grid Technology and Computational Chemistry

• Telescience Marshals Rich Network of Technologies at iGRID2002

• Grid Demo Sets US to Japan Data Speed Records

• EcoGrid

• Encyclopedia of Life
Steering Committee
Came into effect 25 Feb 2003

- John O’Callahan, Bernard Pailthorpe, David Abramson: APAC
- Larry Ang: BII
- Baoping Yan, Kai Nan: CAS/CNIC
- Satoshi Masuoka: TITech/GSICC
- Satoshi Sekiguchi, Yoshio Tanaka: AIST
- Sangsan Lee, Jysoo Lee: KISTI
- Whey-Fone Tsai, Fang-Pang Lin: NCHC
- Shinji Shimojo: Osaka University/CMC
- Royol Chitradon, Piyawut Srichaikul: NECTEC
- Maxine Brown: StarTap
- Rick McMullen, Jim Williams:
- Habibah Wahab: U Sains Malaysia
- Philip Papadopoulos, Peter Arzberger: UCSD/SDSC/Cal-(IT)$^2$/CRBS

PACIFIC RIM APPLICATIONS AND GRID MIDDLEWARE ASSEMBLY
Challenges

• Knowledge

• Speed

• Accuracy
The Challenge of Genomic Sequencing

*Homo sapiens*
(humans)

*Haemophilus influenzae*
The Information Tsunami
--An Example

- **Terabyte** [1,000,000,000,000 bytes OR \(10^{12}\) bytes]
  - 1 Terabyte: An automated tape robot OR all the X-ray films in a large technological hospital OR 50,000 trees made into paper and printed OR daily rate of EOS data (1998)
  - 2 Terabytes: An academic research library OR a cabinet full of Exabyte tapes
  - 10 Terabytes: The printed collection of the US Library of Congress
  - 50 Terabytes: The contents of a large Mass Storage System
  - 400 Terabytes: National Climactic Data Center (NOAA) database
- **Petabyte** [1,000,000,000,000,000 bytes OR \(10^{15}\) bytes]
  - 1 Petabyte: 3 years of EOS data (2001), OR 1 sec of CMS data collection
  - 2 Petabytes: All US academic research libraries
  - 8 Petabytes: All information available on the Web
  - 20 Petabytes: Production of hard-disk drives in 1995
  - 200 Petabytes: All printed material OR production of digital magnetic tape in 1995
- **Exabyte** [1,000,000,000,000,000,000 bytes OR \(10^{18}\) bytes]
  - 2 Exabytes: Total volume of information generated worldwide annually
  - 5 Exabytes: All words ever spoken by human beings
- **Zettabyte** [1,000,000,000,000,000,000,000 bytes OR \(10^{21}\) bytes]
- **Yottabyte** [1,000,000,000,000,000,000,000,000,000 bytes OR \(10^{24}\) bytes]
Extended TeraGrid Facility
Challenges

• Institutional & Infrastructural Ecology
  – Technological change more rapid than institutional change
  – Country and Cultural Uniqueness

• Broadening Participation

• Community-Building

• Seamless Integration of New and Old
  – Balancing upgrades of existing and creation of new resources/diverse levels of resources
  – Legacy data/model
New Modes of Interaction with Resources

W. Feiereisen
Enabling the nation’s future through discovery, learning and innovation