e-Science Infrastructure and Applications in Taiwan

Eric Yen and Simon C. Lin
ASGC, Taiwan
Apr. 2008
Outline

• Driving by WLCG -- Infrastructure, Reliability and Scalability
• Customized services and Application Extension
• Core Technology Building
• Interoperation
• Facilitating Regional and Global Collaboration
• Summary
ASGC and TWGrid

Worldwide Grid Infrastructure

Grid Application Platform

Large Hadron Collider (LHC)

Avian Flu Drug Discovery

Asia Pacific Regional Operation Center
ASGC Profile

• Operational from the deployment of LCG0 since 2002, and Takes the Tier-1 Center responsibility from 2005
• We support the ATLAS and CMS experiments at the same time in WLCG
  • ATLAS: Institute of Physics, Academia Sinica
  • CMS: National Taiwan University and National Central University
• Federated Taiwan Tier-2 center -- Taiwan Analysis Facility (TAF) is also collocated in ASGC
• Leader of EGEE e-Science Asia Federation
• ASGC is not just for WLCG, but also acting as the national center of Grid infrastructure and e-Science research and application in Taiwan
• Providing Asia Pacific Regional Operation Center (APROC) services to Asia Pacific WLCG/EGEE sites
What Do We Deliver?

- **e-Infrastructure Operation**
  - 21+ sites across 8+ countries in Asia Pacific Region
    - > 3,500 Cores and >2 PB storages
  - Continuous monitoring of grid services & automated site configuration management

- **Middleware R&D**
  - Production quality MW distributed under friendly open source license model
  - Application integration

- **User Support: Managed process from first contact to production usage**
  - Training
  - Expertise in grid-enabling applications
  - online helpdesk
  - Dissemination: attracting more collaborations

- **Interoperability: expanding geographical reach and interoperability with collaborating e-infrastructures**
Network Connectivity and Quality Monitoring

- Try to have real-time monitoring of site-to-site connectivity quality, including it’s latency, data throughput, etc.
- Optimize the site-to-site routing
Collaborating e-Infrastructures

Enabling Grids for E-sciencE

“Production” = Reliable, sustainable, with commitments to quality of service

Potential for linking ~80 countries
Reliability of WLCG Tier-1 Sites + CERN

September 2007 - February 2008

Data from SAM Monitoring. Plots show Reliability for last 3 Months.

Reliability is calculated as
\[
\text{time_site_is_available} / (\text{total_time} - \text{time_site_isScheduled_down})
\]

Target reliability for each site is 91% and Target for 8 best sites is 93% from June, 2007.

Target reliability for each site is 93% and Target for 8 best sites is 95% from December, 2007.
Atlas T0-T1 transfer

Throughput (MB/s)

Data Transferred (GBytes)

Completed File Transfers

Total Number Errors
## ASGC Resource Level

<table>
<thead>
<tr>
<th>Date</th>
<th>CPU (ksi2k)</th>
<th>Disk (TB)</th>
<th>Tape (TB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>636</td>
<td>360</td>
<td>800</td>
</tr>
<tr>
<td>2007</td>
<td>2300</td>
<td>1100</td>
<td>800</td>
</tr>
<tr>
<td>2008</td>
<td>3400</td>
<td>1500</td>
<td>1300</td>
</tr>
</tbody>
</table>
CPU Utilization Statistics

- max ~19K jobs/day, 41.5K CPU hours/day
- ~ 3K job slots available from late Feb. ’08
- 6 active VOs
- 179.6K jobs and 712.68K CPU hours in March 2008
Throughput of Data Storage System @ ASGC

- Figure 1: Inbound can reached 6.52 Gbps, Outbound can reached 3.7 Gbps in Mar. 2008.

- Figure 2: the average tape writing rate is ~200MB. (8 drives). Reading rate ~150MB/s.
Asia Pacific Regional Operations Center

- **Mission**
  - Provide deployment support facilitating Grid expansion
  - Maximize the availability of Grid services

- **Supports EGEE sites in Asia Pacific since April 2005**
  - 21 production sites in 8 countries
  - Over 3,500 CPU Cores and >2PB in 2008

- **Runs ASGCCA Certification Authority since 2003**

- **Middleware installation support**

- **Production resource center certification**

- **Operations Support**
  - Monitoring
  - Diagnosis and troubleshooting
  - Problem tracking
  - Security
TWGrid Introduction

- Consortium Initiated and hosted by ASGC in 2002
- Objectives
  - Gateway to the Global e-Infrastructure & e-Science Applications
  - Providing Asia Pacific Regional Operation Services
  - Fostering e-Science Applications collaboratively in AP
  - Dissemination & Outreach
  - Taiwan Grid/e-Science portal
    - Providing the access point to the services and demonstrating the activities and achievements
    - Integration of Grid Resources of Taiwan
    - VO of general Grid applications in Taiwan
EGEE Asia Federation is

- Extending the gLite Infrastructure, currently led by ASGC
- Engaging more user communities to join worldwide e-Science collaboration
- Building regional e-Infrastructure and e-Science application
- Conducting and supporting a production e-Infrastructure
- Working together to provide better user support
- Conducting more business and industry cooperations for new business model and opportunity
Core Technology
Grid Application Platform (GAP)
A light-weight framework for developing problem solving applications on the Grid

Seamless access to Grid applications
Developing customizable problem solving applications
Distributed system architecture

Components
- Portable application package: light-weight client-side package for managing jobs and running applications
- Virtual Queuing System: high-level meta-schedule with application specific resource matching
- Local System Agent: uniform interface for adapting heterogeneous computing environments

Supported computing environments
- Single Server
- Computing Cluster: PBS,
- Grid: LCG, gLite

Features
- Service-oriented architecture
- Portable, intuitive and application specific user interface
- Integrated proxy delegation and automatic proxy renewal with MyProxy server
- Multi-user environment with historical job archiving and grid proxy management
- Uniform interface integrating a variety of computing environments ranged from single workstation to world-wide Grid
- Dynamic resource allocation based on application specification
- Full Java implementation
- Workflow support
SRM-SRB Development

- **Objectives:** make SRM the common interfaces for grid storages, and be interoperable among those storages.
- **Features**
  - Flexible file/space type supported: volatile, durable and permanent
  - Disk usage status checking is available
  - Space reservation functions
- **Progress**
  - Implementation of discovery, permission, directory, space functions are all finished.
  - Transfer function will be done in April.
  - Endpoint ready for testing

Testbed: http://fct01.grid.sinica.edu.tw:8443/axis/services/srm
Preproduction: http://tap02.grid.sinica.edu.tw:8443/axis/services/srm
e-Science Applications in Taiwan

- High Energy Physics: WLCG, CDF, Belle
- Bioinformatics: mpiBLAST-g2
- Biomedicine: Distributing AutoDock tasks on the Grid using DIANE, BioPortal
- Digital Archive: Data Grid for Digital Archive Long-term preservation
- Atmospheric Science
- Earth Sciences: SeisGrid, GeoGrid for data management and hazards mitigation
- Ecology Research and Monitoring: EcoGrid
- Humanity and Social Sciences
- General HPC Services
- Environment and Biodiversity Informatics
- Astronomy: ALMA, PanStar
- e-Science Application Development Platform
Bio-Portal and Virtual Screening Services

Avian Flu Drug Discovery

Interactive scoring Visualization

DIANE Master Process

Docking task pulling Docking complex returning.

Virtual Cluster (DIANE workers)

Grid Job Submission

Resource Broker

Grid Application Portal

EGEE Grid Resources
Virtual Screening Service with GAP
A standalone GUI Application

- One-click job submission
- View the best conformation of a simulation
- Visualize your job status
- Generate the histogram with a given energy threshold

Submit the docking job to the Grid with just one click

Submit the best conformation of a simulation
Submit the histogram with a given energy threshold
DataGrid for Long Term Preservation of NDAP
Grid for Earthquake Data Center

Outputs

Waveform Simulation

Quick Focal Mechanism Determination

TEC Community Library

Seismogram Retrieval

1999 Chi-Chi Taiwan Earthquake

Finite Source Inversion and 3D Wave Propagation
Field Server for Agriculture & Environment

Weather Condition Monitoring

The Farm

Type of Field Servers

Acadiaemia Sinica Grid Computing
Collaboration of NCeSS and ASGC on e-Social Sciences

• Comparative study of the current development and adoption of e-Infrastructure in e-(social) sciences in Taiwan and the UK, by mapping e-Social science in the areas of digital archives and geo-sciences.

• Establish a long-running programme of collaboration internationally

• Idea is to understand and widen uptake of e-Infrastructure
  • Drawing on science and technology studies
    • Early adopters - followers - late adopters (Not character types)
    • Mutual shaping
    • Socio-technical alignment
    • Path dependencies - lock-in
    • Uneven distribution of costs & benefits
    • User-designer relations
  • Designing interventions: Based on understanding of drivers / barriers / enablers / alignment / beaten paths

• Social-Economic Applications will be another focus
EUAsiaGrid

- Identify and engage scientific communities which can benefit from the use of state-of-art Grid technologies;
- Disseminate EGEE middleware in Asian countries by means of public events and written and multimedia material;
- Provide training resources and organise training events for potential and actual Grid users;
- Support the scientific applications and create a human network of scientific communities by building on and leveraging the e-Science Grid infrastructure.
Dissemination & Outreach

• Training
  • Target Audience: Site Admin, User, Application Developer, and General Public

• Incubation Program
  • Grid Camp and Station Program
  • e-Science application and industrial program

• Symposium/Conference/Workshop
  • promoting EGEE values, ASGC services etc., and to engage more collaboration
  • Project coordination, learning, and sharing and interactions by hosting events in Taiwan.

• Evaluation & Follow-ups Program would be commenced
**Dissemination & Outreach (2)**

- International Symposium on Grid Computing (ISGC) from 2002
- TWGRID Web Portal ([http://www.twgrid.org](http://www.twgrid.org))
- Grid Tutorial, Workshop & User Training: ~700 participants in past 10 events
- Publication

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
<th>Attendant</th>
<th>Venue</th>
</tr>
</thead>
<tbody>
<tr>
<td>China Grid LCG Training</td>
<td>16-18 May 2004</td>
<td>40</td>
<td>Beijing, China</td>
</tr>
<tr>
<td>ISGC 2004 Tutorial</td>
<td>26 July 2004</td>
<td>50</td>
<td>AS, Taiwan</td>
</tr>
<tr>
<td>Grid Workshop</td>
<td>16-18 Aug. 2004</td>
<td>50</td>
<td>Shang-Dong, China</td>
</tr>
<tr>
<td>NTHU</td>
<td>22-23 Dec. 2004</td>
<td>110</td>
<td>Shin-Chu, Taiwan</td>
</tr>
<tr>
<td>NCKU</td>
<td>9-10 Mar. 2005</td>
<td>80</td>
<td>Tainan, Taiwan</td>
</tr>
<tr>
<td>ISGC 2005 Tutorial</td>
<td>25 Apr. 2005</td>
<td>80</td>
<td>AS, Taiwan</td>
</tr>
<tr>
<td>Tung-Hai Univ.</td>
<td>June 2005</td>
<td>100</td>
<td>Tai-chung, Taiwan</td>
</tr>
<tr>
<td>EGEE Workshop</td>
<td>Aug. 2005</td>
<td>80</td>
<td>20th APAN, Taiwan</td>
</tr>
<tr>
<td>EGEE Administrator Workshop</td>
<td>Mar. 2006</td>
<td>40</td>
<td>AS, Taiwan</td>
</tr>
<tr>
<td>EGEE Tutorial and ISGC’06</td>
<td>1 May, 2006</td>
<td>73</td>
<td>AS, Taiwan</td>
</tr>
<tr>
<td>EGEE Tutorial with APAN 23</td>
<td>26 Jan. 2007</td>
<td>30</td>
<td>Manila, Philippine</td>
</tr>
<tr>
<td>EGEE Tutorial with ISGC’07</td>
<td>26 Mar. 2007</td>
<td>90</td>
<td>AS, Taiwan</td>
</tr>
<tr>
<td>EGEE Tutorial with GridAsia2007</td>
<td>8 Jun. 2007</td>
<td>20</td>
<td>Singapore</td>
</tr>
<tr>
<td>GridCamp</td>
<td>28 Oct. 2007</td>
<td>97</td>
<td>AS, Taiwan</td>
</tr>
<tr>
<td>Do Son Grid School</td>
<td>Nov. 2007</td>
<td>80</td>
<td>HCMC, Vietnam</td>
</tr>
<tr>
<td>MIMOS Grid Tutorial</td>
<td>Dec. 2007</td>
<td>30</td>
<td>Malaysia</td>
</tr>
</tbody>
</table>
Summary

- Practitioner for the new generation infrastructure and international collaboration.
- Reliability is the first course. Scalability is totally another issue. Customized and advanced functionality upon user requirements is a long-term lesson.
- Asia Pacific Region is of virtuous potential to adopt the e-Infrastructure:
  - More and more Asia countries will deploy Grid system and take part in the e-Science world
  - From Asia Federation to EUAsiaGrid, we are widening the uptake of e-Science, by the close collaboration regionally and internationally
Data Federation for Biodiversity and Environment Informatics
engage the regional/domestic application communities to EGEE and the world

• --> EGEE, OSG, and others

•