Introduction to Grid Technology

Presented by Melvin KOH,
Grid Research Engineer, Sun APSTC

Simon SEE, Director
Sun APSTC – Asia Pacific Science & Technology Center
Agenda

• Overview of Grid Technology
  – DRM, Grid Middleware, Grid Portal Technology, Grid Service

• Sun's Grid Technology
  – N1 Grid Engine, Grid Engine Portal, Grid Solutions

• APSTC Grid Research
Overview on Grid Technology
Overview on Grid Technology

- Distributed Resource Manager (DRM)
- Grid Middleware
- Grid Portal Technology
- Grid Application Development Tools
Distributed Resource Manager

- Local cluster scheduler
- Usually deploy in single administrative domain
- Platform LSF
- Altair PBS
- Condor
- N1 Grid Engine
Grid Portal Technology

- Grid Portal Development Kit
  - JSP, Java bean, Tomcat
  - No longer supported

- GridSphere Portal Framework
  - JSR168 compliant, Grid Portlets

- GridPort Toolkit
  - CGI and Perl, Perl CoG kit

- Grid Engine Portal
Grid Middleware

• Globus Toolkit
  – *Resource mgnt, monitoring, discovery services*
  – *Open Grid Service Architecture (GT3)*
  – *Web Service Resource Framework (GT4)*

• Job Scheduling Hierachically (JOSH)

• Avaki
  – *Distributed data access through a single service*
  – *Hide complexities of individual data sources from applications and users*
Grid Application Development Tools

- Globus Commodity Grid Kits
- GridRPC API
  - NetSolve / GridSolve
  - Ninf-G
- Grid Application Toolkit (GAT)
- MPI-G2
- Distributed Resource Management Application API (DRMAA)
Sun's Grid Technology
N1GE 6 Features

- User Specifiable Resource
- Job accounting
- Failover capability
- Ticket-based scheduling policies
- Parallel support
- Checkpointing support
- Integrate with Globus Toolkit
N1GE 6 Features

• New Features
  – Increase scalability
  – More sophisticated scheduling algorithms
  – Resource reservation with backfilling
  – Cluster queues
  – DRMAA 1.0 compliant (C and Java bindings)
  – Web-based accounting and reporting console
ARCo: Accounting and Reporting Console

• Web-based tool for displaying data in reporting DB
• Ability to create simple and advanced (SQL-based) queries
• Generates tables, graphs, results exportable as CVS
• Not available in open source GE
### Error And Information Messages

**Select a view:** all    
**Select**

<table>
<thead>
<tr>
<th>Query Name</th>
<th>Timestamp</th>
<th>Type of view</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query_2003_001</td>
<td>yyyy:mm:ss:mm:ss</td>
<td>job</td>
<td>Delete</td>
</tr>
<tr>
<td>Query_2003_002</td>
<td>yyyy:mm:ss:mm:ss</td>
<td>job</td>
<td>Delete</td>
</tr>
<tr>
<td>Query_2003_003</td>
<td>yyyy:mm:ss:mm:ss</td>
<td>system</td>
<td>Delete</td>
</tr>
<tr>
<td>Query_2003_004</td>
<td>yyyy:mm:ss:mm:ss</td>
<td>host</td>
<td>Delete</td>
</tr>
<tr>
<td>Query_2003_005</td>
<td>yyyy:mm:ss:mm:ss</td>
<td>host</td>
<td>Delete</td>
</tr>
</tbody>
</table>
Runtime of completed jobs
Sorted ascending by host

<table>
<thead>
<tr>
<th>No.</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Egon</td>
</tr>
<tr>
<td>2</td>
<td>Max Holle</td>
</tr>
<tr>
<td>3</td>
<td>Asterisk Gaier</td>
</tr>
<tr>
<td>4</td>
<td>Gustav Klenner</td>
</tr>
<tr>
<td>5</td>
<td>Reiner Unhinn</td>
</tr>
</tbody>
</table>

save Resultset  save Query  export
Grid Engine Portal

- Previously “Technical Computing Portal”
- Opensourced by Sun
- Web-based GUI for easy to use interface
- Support exporting display of X-window applications using VNC
- Integrated with Sun Java System Portal Server
- Provides secure remote access
Grid Engine Portal

Sun ONE Portal Server 6.0 - Netscape

Job List
You have no running jobs.
Submit new job...

Project List
You have no project.
Create new project...

Admin Application List
- FastX
- LS-DYNA-MPP Using Form (c568D)
edit delete
edit delete
Add a new application...

Job Accounting
There are no running jobs.
Perform accounting...
APSTC Grid Research
APSTC Grid Research

• Introduction to APSTC
• Current research projects
  – NTU Campus Grid
  – BioBox Initiative
  – GriDE Project
  – Super Scheduler
• Other research
Introduction to APSTC

• To provide technical and scientific expertise for technical computing business
• To conduct Applied Research in Science and Engineering
• Current research focus on grid computing and performance analysis
NTU-Campus Grid

- Initiated by NTU professors and Sun
- Testbed for grid research
- Connecting four centres
  - APSTC, BIRC, NCSV, SCE
- Heterogeneous (OS, architecture)
- Portal Frontend (Clearing House Portal)
Introduction

The ClearingHouse Project at Nanyang Technological University, is an initiative to build a campus-wide grid infrastructure to provide a high-level architecture for pervasive access to heterogeneous computing resources. This project aims to achieve seamless access to computing resources across multiple administrative domains. A Clearinghouse System, built and administered by research staff from the Parallel and Distributed Processing Lab, in the School of Computer Engineering, provides a set of high level services to enable users to gain access to the resources through a web portal that can be accessed from this website.

This website also serves to provide information of our developments with regards to our strategic goals and R&D efforts to enable enterprise-wide access to computing and data resources. Our focus is on concepts and technologies deployed in Grid Computing and therefore, the Grid functions performed are resource management, automatic resource discovery and economy management of resources in a service-oriented environment. Our current aim is to build a middleware service infrastructure for ubiquitous access to heterogeneous computation and data entities registered to the clearinghouse. With a middleware infrastructure, we can provide a convenient means for users to access the campus grid to run massive parallel applications.
BioBox Initiative

• Easy-to-deploy installation package consisting of Solaris OS and most popular Bioinformatic applications
• Implemented using Sun's Web Start Flash technology
• Total 28 bioinformatic applications, 17 integrated with Grid Engine
BioBox Core Components

- Grid Engine Portal
- Sample databases
- Solaris operating environment
- Sun Grid Engine
- Ready-to-use Bioinformatics applications
GriDE Project

Q: What is GriDE?
A: GriDE is an integrated development environment that makes it straightforward for scientists and engineers to construct grid applications. It provides friendly tools to access grid resources and makes the development approach easy and fast.
GriDE Architecture

NetBeans Tools Platform

Flow Editor  Cross Compiler  Grid Performance Data Grid Project
   Apps   Debugger   Turning  Grid  Collaboration
   Templates  API

GriDE Layer

Java  C/C++  Fortran  CoGKit  MPI-G2  Grid Simulation

Programming languages/ Libraries Layer

Globus  SGE  NimRod

Middleware Layer

Computing Resources
GriDE Overview

• Extends Netbeans IDE
• Supports Globus Toolkit 2.x
• Runs on Solaris, Linux, Windows platform
• Current Features
  – Grid Job Submission
  – Resource Browser
  – Grid Job Monitoring
  – Grid Workflow Editor
Super Scheduler

- Brokering system over Globus Toolkit 3.x
- Supports grid services using OGSA
- Distributed and peer-to-peer model
- Infrastructure is based on JXTA
- Still in very preliminary stage
Other Research

• Grid performance evaluation and benchmarking
• Economic model in grid
• Interval arithmetic
Thank You

Melvin.Koh@Sun.Com

http://apstc.sun.com.sg