Grid Access Methods & Tools

Shamjith K V
(shamjithkv@cdacb.ernet.in)
System Software Development Group
CDAC, Bangalore
Overview

- Access Methods
- GARUDA Tools
Access Methods

• Two ways of Accessing Garuda Resources
  – Command Line Interface
  – Web Interface
Command Line Interface

• Provides Command Line Access to Garuda Resources
• Requirements
  – Valid IGCA Certificate
  – User is required to have Garuda User Account
• Facilities
  – Certificate and Proxy Management
  – Single Sign-On to all Garuda Resources
  – Job Submission, Monitoring and Management
  – Data transfer and Management
  – Querying Resource Information
  – Application Compilation Facility
  – Accomplishing Resource Reservation
• Useful for Advanced Users
Web Interface

• Provides Garuda Access using Web Browser
• Requirements
  – Valid IGCA Certificate
  – Garuda User Account
• Provides easy and intuitive interfaces for
  – Certificate and Proxy Management
  – Job Submission, Monitoring and Management
  – Data transfer and Management
  – Querying Resource Information
  – Carrying out Resource Reservation
• For Novice users
• Web Interfaces
  – Garuda Grid Portal
  – Problem Solving Environment
Garuda Grid Portal
Garuda Grid Portal - Features

- Basic Job Submission
- Advanced Job Submission
- Browse Available resources
- Provision to View Status, Output, Error Files
Garuda Job Submission Strategy 1: Basic Submission

- Supports sequential and parallel Jobs
- The jobs will run on a single cluster
- The user can specify the required environment variables
- Uses the scheduler for submission
- The files can be taken from Data Grid, the submit node, or the local machine
Garuda Job Submission Strategy 2: Advanced Submission

- Uses Globus
- Sequential and homogeneous parallel jobs can be submitted
- The parallel jobs should be compiled with MPI Libraries
The status of submitted jobs can be viewed at any point of time.

The output and error files can be viewed or downloaded.

The jobs submitted through the Basic Submission option can be cancelled.
Problem Solving Environment
Protein Structure Prediction

• Protein Structure Prediction (PSP) based on evolutionary computing (Genetic Algorithms) method.
• Functional modules of the PSP application implemented as grid services on Garuda
• Released PSE-PSP v1.0
• Accessible from Garuda Portal
GARUDA Tools
GARUDA Tools

- Monitoring & Management Tool (Paryavekshanam)
- Automatic Grid Service Generator
- Garuda Storage Resource Manager
Paryavekshananam
GARUDA Grid Monitoring Tool
Paryavekshanam

"GARUDA" – A National Grid Computing Initiative by the Centre for Development of Advanced Computing to setup a nationwide Grid that is distributed across 17 cities and 47 research institutions in India. GARUDA is an aggregation of heterogeneous resources and advanced capabilities to enable increasingly interdisciplinary scientific environments. Continuous monitoring of the Grid is crucial and challenging. Paryavekshanam ensures 24x7 monitoring that steers through the system hierarchically.

Grid Overview
Status As on 2010-07-07 08:00:09

Legend
- Resource Provider
- Non-Resource Provider

<table>
<thead>
<tr>
<th>No. Alerts</th>
<th>ID</th>
<th>Code</th>
<th>Priority</th>
<th>Status</th>
<th>Centre</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>EI124343</td>
<td>eHLK</td>
<td>Normal</td>
<td>Open</td>
<td>IIT Guwahati</td>
</tr>
<tr>
<td>2</td>
<td>EI123977</td>
<td>eHLK</td>
<td>Normal</td>
<td>Open</td>
<td>National Institute of Technology</td>
</tr>
<tr>
<td>3</td>
<td>EI123976</td>
<td>eHLK</td>
<td>Normal</td>
<td>Open</td>
<td>Raman-Chandra Research Institute</td>
</tr>
</tbody>
</table>

Legal Notice | Privacy Policy | © 2006 CDAC. All rights reserved.
Paryavekshanam – Features

• Hierarchically designed web pages
• Graphical representation of all parameters for easy comparison and understanding.
• Archiving data, 24 X 7 support
• Parya Dash Board. (Connectivity Map, Status Bar, Radar Graph)
• Separate pages for GMMC monitoring, Alerts
• Resource and Software Search.
Paryavekshanam - Parameters monitored

- Network link status.
- Bandwidth utilization (includes available and used).
- Network Stability.
- Round Trip time (and hence latency).
- Cpu Load.
- Cluster status.
- Node availability.
- Memory utilization.
- Jobs (running, queued, pending and total).
- Ip address.
- Globus component status.
- Softwares available on clusters.
- Active users.
Automatic Grid Service Generator
Automatic Grid Service Generator (AGSG)

- The Automatic Grid Service Generator (AGSG) is a web tool to automatically generate Grid services.
- It has a friendly GUI. User needs to upload executable and provide command line arguments details, service name and details of any supported libraries.
- Supports executables of different languages – C, C++, Fortran, Java.
- Service details are sent to user’s emailid.
- Has a client interface (AGSGClient) to invoke the service generated. User can upload all input files at the time of service invocation.
- Released AGSG v1.0 and accessible from Garuda Portal.
Enter Service URI: https://gg-blr.tfg:8443/wsrf/services/Matrix

GARUDA Storage Resource Manager
Storage Resource Manager

- SRM is set of specifications to manage data on Grid.
- SRM services implement SRM specifications.
- Functionality of SRM middleware:
  - Dynamic space allocation
  - File management in spaces for shared storage components on the Grid.
- Storage Element (SE): Grid service that allow users to store and manage files on assigned Space.
- Files are referred by different name (in Grid):
  - Logical File Name (LFN): alias created by a user to refer to some item of data
    - \texttt{lfn://grid/<MyVO>/<MyDirs>/<MyFile>}
  - Grid Unique Identifier (GUID): A non-human-readable unique identifier for data entity
  - Site URL (SURL): Physical file name or the location of an actual piece of data on a storage system.
    - \texttt{<srm>://<SE_hostname>/<path>}
GARUDA Storage Resource Manager

• Based on Disk Pool Manager following SRM v 2.2 specifications.
• Adheres to OGF standards.
• Single Point access to GARUDA storage resources.
• Caters the basic GARUDA storage requirements like Data Availability, Security, Performance, Interoperability, Scalability.
• Supports advanced data management features:
  – Dynamic Space Management
  – Quota Allocation
  – Interfacing with High performance file system
  – Interoperability with other SRM implementations
GSRM Component-Feature Mapping

Data Availability
- GARUDA SRM User request
- Gsiftp server
- Rfio server

Performance
- GARUDA SRM Admin Request

GARUDA SRM

Space Mgmt
- Space Manager

Persistence
- User Request DB
- SRM server

Interoperability
- StoRM/Bestman Request

Gsiftp server

Rfio server

Global Namespace

Namespace Manager

Metadata (namespace DB)

Security
GSRM Storage

• Storage Systems supported:
  – Disk file systems
  – High performance file systems

• Storage Type :
  – Volatile
  – Permanent

• Storage Allocation
  – Default space :10GB
  – Allocated Storage type : depending on VO to which user belongs & Application domain
  – Storage can be dynamically allocate/ released as per the user space requirement

• Storage Access Control :
  – Can be done by the administrator by controlling the file & directory permissions for each VO /group
GSRM Integration with GARUDA

- Head Node
  - Portal
  - Grid way
  - GSRM Clients
  - VOMS/Myproxy clients

- Myproxy/VOMS server

- GARUDA Network

- Unified Disk based Storage
- GSRM Server
  - (DN-VO / DN-GID)
  - PFS Clients

- GARUDA Federated Information system

- PFS
- IO
- Servers

- GSRM Information server

- Internet
How to Get Access to GSRM

• Valid User certificate
  – https://ca.garudaindia.in/

• Garuda User Account
  – grid-help@cdacb.ernet.in

• Authorization to access SRM storage resources
  – rt-gds@cdacb.ernet.in
GSRM User Interfaces

- GSRM services are accessible through any of SRM client implementations like StoRM, Bestman, etc.
- Supported Functionalities:
  - Resource Discovery
  - File & Directory Management
  - Space Management
  - Space Reservation
- Supported User Interfaces:
  - Command Line (All functionalities available)
  - C & Java APIs
  - Web Interface (Basic Interface file browsing, upload, download)
- For details of Installation & Usage: Refer GSRM Installation & Usage manual - GARUDA wiki
Thank You
• The GridPort Toolkit (GridPort) enables the rapid development of highly functional grid portals that simplify the use of underlying grid services for the end-user.
• It comprises a set of portlet interfaces and services that provide access to grid technologies including the Globus Toolkit, the Grid Portal Information Repository (GPIR), and Condor.
• Latest version: **GridPort 4.0 (12-19-2005)**
• Official website: [http://gridport.net/main/](http://gridport.net/main/)
Projects using GridPort...

### Parallel Computing Resources

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>System</th>
<th>CPUs</th>
<th>Peak GFlops</th>
<th>Memory GB</th>
<th>Disk GB</th>
<th>Status</th>
<th>Load</th>
<th>Jobs</th>
</tr>
</thead>
<tbody>
<tr>
<td>dlogin.sdsc.edu</td>
<td>San Diego Supercomputer Center</td>
<td>IBM Power4+</td>
<td>96</td>
<td>653</td>
<td>364</td>
<td>28864</td>
<td>Available</td>
<td>27R-0Q-1990</td>
<td></td>
</tr>
<tr>
<td>tpg-login.caitech.teragrid.org</td>
<td>California Institute of Technology</td>
<td>Intel Itanium2 IA-64 Cluster</td>
<td>144</td>
<td>5325</td>
<td>432</td>
<td>92160</td>
<td>Available</td>
<td>R-Q-Q</td>
<td></td>
</tr>
<tr>
<td>tpg-login.ncca.teragrid.org</td>
<td>The National Center for Supercomputing Applications</td>
<td>Intel Itanium2 IA-64 Cluster</td>
<td>1262</td>
<td>6144</td>
<td>2524</td>
<td>133120</td>
<td>Available</td>
<td>86R-63Q-10</td>
<td></td>
</tr>
<tr>
<td>tpg-login.sdsc.teragrid.org</td>
<td>San Diego Supercomputer Center</td>
<td>Intel Itanium2 IA-64 Cluster</td>
<td>512</td>
<td>3174</td>
<td>1024</td>
<td>61133</td>
<td>Available</td>
<td>7R-47Q-20</td>
<td></td>
</tr>
<tr>
<td>tpg-login.tacc.teragrid.org</td>
<td>Texas Advanced Computing Center</td>
<td>Intel Pentium4 IA-32 Cluster</td>
<td>1028</td>
<td>6291</td>
<td>1984</td>
<td>33736</td>
<td>Available</td>
<td>47R-0Q-70</td>
<td></td>
</tr>
<tr>
<td>tpg-login.uc.teragrid.org</td>
<td>University of Chicago/Argonne National Laboratory</td>
<td>Intel Itanium2 IA-64 Cluster</td>
<td>124</td>
<td>614</td>
<td>62</td>
<td>4096</td>
<td>Available</td>
<td>1R-0Q-00</td>
<td></td>
</tr>
</tbody>
</table>

**Total:** 3166 CPUs, 22201 Peak GFlops, 6410 Memory GB, 353109 Disk GB

### Visualization Resources

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution</th>
<th>System</th>
<th>CPUs</th>
<th>Peak GFlops</th>
<th>Memory GB</th>
<th>Disk GB</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>maverick.tacc.utexas.edu</td>
<td>Texas Advanced Computing Center</td>
<td>Sun E25K</td>
<td>128</td>
<td>3</td>
<td>512</td>
<td>562</td>
<td>Available</td>
</tr>
<tr>
<td>rachel.psc.edu</td>
<td>Pittsburgh Supercomputing Center</td>
<td>Alpha EV7</td>
<td>128</td>
<td>307</td>
<td>512</td>
<td>6144</td>
<td>Available</td>
</tr>
</tbody>
</table>

**TeraGrid User Portal**
Projects using GridPort…

SURA User Portal
(Southeastern Universities Research Association)
Projects using GridPort...
Information Repository - GPIR

- GPIR – Grid Portal Information Repository
- GPIR provides a place to store data about your grid that is readily accessible to a portal application.
  - Dynamic data
    - *GPIRIngester* - webservice updates the data
    - *GPIRQuery* - webservice for reading the data
  - Human-centric data
    - Managed via *GPIR Administration* client by the administrator.
- Official website: [http://gridport.net/services/gpir/](http://gridport.net/services/gpir/)
• **PURSE**: Portal-based User Registration Service developed by [http://www.grids-center.org/](http://www.grids-center.org/)

• PURSE is an "integrated solution" that provides an easy-to-use web interface for potential users of an application to "register" themselves and request sign-in credentials.

• PURSE combines the **Simple CA** and **MyProxy** components with a back-end database and a web portal to automate user registration requests.
Grid Monitoring
MONALISA

- MONitoring Agents using a Large Integrated Services Architecture
- Monitoring all aspects of complex systems:
  - System information for computer nodes and clusters.
  - Network information (traffic, connectivity, topology) for WAN and LAN.
  - Monitoring the performance of Applications, Jobs or services.
- Official website:
  - [http://monalisa.caltech.edu/monalisa.htm](http://monalisa.caltech.edu/monalisa.htm)
GridICE

- The **GridICE** monitoring tool developed from INFN, Italy, is a distributed monitoring tool designed for Grid systems.
- Its design is based on requirements given by different type of users, each of them dealing with a different abstraction level of a Grid: the Virtual Organization level, the Grid Operation Center level, the Site Administration level and the End-User level.
- Official website: [http://gridiceforge.cnaf.infn.it/](http://gridiceforge.cnaf.infn.it/)
Debuggers
Debuggers – Mercury Monitoring & Debugging

• The Mercury Monitoring System which is developed as part of the GridLab project is a generic Grid monitoring framework, also allowing remote debugging.

• It employs GDB and GDBServer

• Debugging is performed by sending a message to the Mercury monitoring library that is compiled into the application.
Debuggers - Harness

• Harness is a metacomputing system that defines a simple but powerful architectural model to overcome the limited flexibility of traditional distributed software frameworks.

• Consists of a debugger kernel and plug-ins that provide various services for users.

• It is implemented in Java to leverage the homogenous architecture, the JVM, over heterogeneous computer platforms.

• The debugger is closely intertwined with the framework and it cannot be used with other Grid middleware.
Gridhra
Need for a parallel debugger for the Grid

- Debugging is an important aspect of any development environment, helping users to locate the defects and take corrective action for rectifying them.
- Nevertheless, debugging might prove to be a tedious, time consuming task.
- With parallel programming, the importance of debugging becomes predominant.
Need for a parallel debugger for the Grid …

- With the **multiplicity of processes**, debugging becomes extremely difficult in parallel environments.
- And hence with **Grid applications**, the difficulty increases many fold, owing to the **heterogeneity of resources** on which the applications run.
- So a parallel debugger in Grid Environment can help user to reduce the application development time by providing an easy debugging framework.
Gridhra Debugger

• Provides a simplified and unified framework for debugging parallel programs on the grid with a web user interface.

• Caters to heterogeneity of resources in the grid.
Gridhra Visualization

- Profiling framework for MPICH-G2 applications.
- User has to compile the application with the script provided in $GRIDHRA_ROOT/bin for visualization.
- Graphical representation of the Call Graph sequence and Communication patterns in the MPI application.
- Helps the user to fine tune applications.
Call Graph

- Process wise Call Graph
Communication and Computation

Graphs

- Communication v/s Computation breakup, both at the process level and application level.
Grid IDE
Grid Integrated Development Environment (GIDE)

- GIDE is the comprehensive GRID Program Developing tool with inbuilt components for:
  - Project Development and Management
  - Inbuilt Compile/Build
  - Source Level debugger
  - Profiler
GIDE - Advantages

• GIDE - Provides a **single** convenient environment with various developing tools.
• Enable users to carry out **entire program development** life cycle for the **Grid**
• Greatly improves the efficiency of the application development and **productivity of the developer**
• Provides **friendly** tools to access **grid resources** and makes the development approach easy and fast.
Components of Grid IDE (GIDE)

• Project Development and Management

• Resource Management

• Job Management

• Source level Debugger

• Profiler

• Help
GSRM Usage in GARUDA

- **Myproxy / VOMS server**
- **HN Portal**
- **Gridway**
- **GSRM Clients**
- **VOMS/Myproxy clients**
- **Proxy init**
- **Input file**
- **Output file**
- **Unified Disk based Storage**
- **Gsiftp/ rfio Data transfer**
- **High Performance IO**
- **PFS Clients**
- **Lcg-dm file (DN-VO / DN-GID)**
- **PFS IO Servers**

**GSRM Server**

**GSRM Usage in GARUDA**
AGSG – Service Deployment

Dear Admin,

You are requested to deploy the GAR file for the Service, in the Samba container of the specified cluster according to details below:

1) Service Name : Blast
2) Location : Bangalore
3) Head Node Name : grid-blcl
4) Source Path : /home/agsg/Blast
5) Destination Path : /home/bioinfo/janaki/Blast

6) Expected Completion Date: Fri May 28 18:23:48 IST 2010

I am also providing the GAR file as attachment with this mail.

Thanks,

/home/agsg/Blast/service