

Introduction to Globus

Shamjith K V

System Software Development Group,
CDAC, Bangalore.

Presentation Overview

- **What is Globus Toolkit ?**
- **Components of Globus Toolkit**
- **Functions of Globus Components**
- **Job submission interfaces**
- **Advantages of Globus**

What is Grid Middleware ?

- **Grid Middleware** refers to the **Software Infrastructure** that facilitates **security, resource management, data access, instrumentation, policy, accounting, and other services required for applications, users, and resource providers to operate effectively in a Grid Environment.**

Popular Middleware

- **Globus** – Globus Alliance
- **GridBus** – University of Melbourne
- **UNICORE** - Uniform Interface to Computing Resource
- **gLite** - CERN

What is Globus Toolkit (GT)

- **GT is an Open Source software toolkit used for building Grid Systems and Applications.**
- **By Globus Alliance.**
- **GT allows people to Share Computing Power, Databases, and other tools Securely online across Institutions, and across Geographic boundaries.**
- **GT includes Software Services and Libraries for Resource Monitoring, Discovery, Security and File Management.**

Components of GT



Security:
Grid Security
Infrastructure
(GSI)



**Resource
Management:**
Grid Resource
Allocation
Management
(GRAM)



**Information
Services:**
Monitoring and
Discovery
Services
(MDS)



Data Management:
Grid File Transfer
Protocol **(GridFTP)**,
and **Replica
Management**
services.

GSI - Grid Security Infrastructure

- **Motivations**
 - ∅ Secure Communication
 - ∅ Security across Organizational boundaries
 - ∅ Support Single Sign-On
- **GSI : Tools, Libraries and Protocols used in Globus to allow users and applications to securely access resources.**

GSI Continued...

- **Based on PKI**
 - ∅ Private and Public Keys
 - ∅ Certificates
- **SSL for authentication and message protection**
 - ∅ Encryption
 - ∅ Signature
- **Adds features needed for Single-Sign on**
 - ∅ Proxy Credentials
 - ∅ Delegation

Resource Management

- **GRAM** - Grid Resource Allocation Management
 - ∅ GRAM protocol and client API allows programs to be started on remote resources, despite local heterogeneity
 - ∅ Resource Specification Language (RSL) is used to communicate requirements
- **DUROC** - Dynamically-Updated Request Online Co-allocator
 - ∅ used to coordinate transactions with each of the RMs and bring up the distributed pieces of the job.
- **GASS** - Global Access to Secondary Storage
 - ∅ IO redirection, File & executable staging with GRAM
 - ∅ Uses GSI-enabled HTTP

GIS - Grid Information Service

- **Provide access to static and dynamic information regarding system components**
- **Useful Information**
 - ∅ Characteristics of a compute resource
 - ∅ Characteristics of a network
 - ∅ Characteristics of the Globus infrastructure
- **Features of GIS**
 - ∅ Uniform, flexible access to information
 - ∅ Scalable, efficient access to dynamic data
 - ∅ Access to multiple information sources

MDS - Monitoring and Discovery Service

- **GRIS- Grid Resource Information Service** supplies information about a specific resource
 - ∅ Configurable to support multiple information providers
 - ∅ LDAP as inquiry protocol
- **GIIS - Grid Index Information Service**
 - ∅ Supplies collection of information which was gathered from multiple GRIS servers
 - ∅ Supports efficient queries against information which is spread across multiple GRIS server
 - ∅ LDAP as inquiry protocol

Data Management

- **Data Transfer Management**
 - ∅ GridFTP – Grid File Transfer Protocol
 - Secure, efficient, flexible data transfer
 - Based on FTP
- **Replica Management**
 - ∅ RLS – Replica Location Service
 - Maintain a mapping between logical names for files and collections and one or more physical locations
 - multiple copies of files

Job Submission Interfaces

- **GT includes several command line programs for job submission**
 - ∅ **globus-job-run**: Interactive jobs
 - ∅ **globus-job-submit**: Batch/offline jobs
 - ∅ **globusrun**: Flexible scripting infrastructure
 - ∅ Submission through APIs
- **Other High Level Interfaces**
 - ∅ General purpose
 - Nimrod-G, Condor-G, PBS etc
 - ∅ Application specific
 - Cactus, Web portals

Job Submission Interfaces - globus-job-run

- **Run interactive jobs**
- **Additional functionality beyond rsh**
 - ∅ Ex: Run 2 process job w/ executable staging
\$ globus-job-run -: host -np 2 -s myprog arg1 arg2
 - ∅ Ex: Run 5 processes across 2 hosts
*\$ globus-job-run \
-: host1 -np 2 -s myprog.linux arg1 \
-: host2 -np 3 -s myprog.aix arg2*
 - ∅ For list of arguments run:
\$ globus-job-run -help

Job Submission Interfaces - globus-job-submit

- **Running of offline jobs**

- ∅ ***\$ globus-job-submit***

- Same interface as globus-job-run
 - Returns immediately

Submit job

- **Job control and status**

- ∅ ***\$ globus-job-status***

Check job status

- ∅ ***\$ globus-job-cancel***

Cancel job

- ∅ ***\$ globus-job-get-output***

Get job stdout/err

- ∅ ***\$ globus-job-clean***

Cleanup after job

Job Submission Interfaces - globusrun

- **Flexible job submission for scripting**
 - ∅ RSL string to specify job request
 - ∅ Contains an embedded globus-gass-server
 - Defines GASS URL prefix in RSL substitution variable:
(stdout=\$(GLOBUSRUN_GASS_URL)/stdout)
 - ∅ Supports both interactive and offline jobs

Advantages of Globus Toolkit

- **RSL to express the required constraints**
- **Services:**
 - ∅ Resource Co-allocation,
 - ∅ Executable Staging,
 - ∅ Remote Data Access
 - ∅ I/O Streaming
- **Integrates with:**
 - ∅ **MPICH-G**: Grid-enabled MPI
 - ∅ **Nimrod-G**: Grid Resource Broker
 - ∅ **Gridbus**: Grid Business Infrastructure
 - ∅ **Condor-G**: High-throughput Broker
 - ∅ **PBS-Pro**: Meta Schedulers



Thank you!

অর্থাৎ অর্থাৎ অর্থাৎ অর্থাৎ

Advanced Computing For Human Advancement