NAREGI Middleware V.1.0
Production-Quality, Industry-Strength Grid Middleware for Petascale Supercomputing Grids in Japan.

Satoshi Matsuoka
Tokyo Institute of Technology
National Institute of Informatics
NAREGI Software Stack

Grid-Enabled Nano-Applications

Grid Visualization
Grid PSE
Grid Workflow Tool
Super Scheduler
Information Service

Data Grid

Grid Programming Libraries
- GridRPC
- GridMPI

WSRF (NAREGI implementation + Globus 4)

Grid VM

High-Performance & Secure Grid Networking, Certification

SuperSINET

Computing Resources

NII
IMS
Research Organizations
etc
1. Services
- Super Scheduler: Meta scheduler
- GridVM: Job manager on resources
- Info Service: Grid resource info and accounting
- Network Services: Network performance measurement & routing control

2. Features
- OGF/W3C/OASIS/DMTF etc. Standards-based
- Automatic resource brokering
- Advanced reservation based co-allocation (heterogeneous resources)
- Coexistence of non-reserve and reservation jobs
- Coexistence of grid and local batch jobs
- Bulk job submission (both native and non-native)
- Management tools
- Network measurement and control
- Interoperation with EGEE/gLite (prototype)
- Ease of Installation, Server/VO Operation packages
3. Supporting Standards

- DMTF/CIM
- OASIS/WSRF
- OGF/JSDL
- OGF/OGSA-EMS
- OGF/OGSA-RUS
- OGSA-DAI
The First OGSA-EMS incarnation

Open Grid Service Architecture – Execution Management Service
will be standardized by OGSA-EMS-WG

- NAREGI will contribute our middleware as a reference implementation of OGSA-EMS.
Reservation Based Co-Allocation

- Co-allocation for heterogeneous architectures and applications
- Used for advanced science applications, huge MPI jobs, realtime visualization on grid, etc...

Diagram:
- Client
  - Workflow
    - Abstract JSDL
  - Reservation based Co-Allocation
- Super Scheduler
  - Reservation, Submission, Query, Control...
- Distributed Information Service
  - CIM
    - Resource Info.
  - Accounting
- GridVM
  - Computing Resource
Bulk Job Submission

- Batch scheduler’s native bulk function support
- Accelerates parameter sweep jobs

**Non Native Bulk Functions on Batch Schedulers**

**Native Bulk Functions on Batch Schedulers**
1. Features

- File registration, distribution in grid environment.
- Shared file system in the grid environment accessible using global names.
- Import/export files to/from the shared file system.
- File staging to the computing environment in cooperation with the workflow-tool.
- Interoperation with EGEE/gLite (prototype).

2. Supporting Standards

- DMTF/CIM
- OASIS/WSRF
- OGF/GFS
Data Grid Environment Architecture

- Grid-wide data sharing services using extended Gfarm
- Data transfer service for Super Scheduler and WFT
Interoperation with EGEE/gLite (prototype)

- NAREGI and EGEE gLite clients can access to both data resources (e.g., bi-directional file copy) using SRM interface.
- GridFTP is used as its underlying file transfer protocol.
- File catalog (metadata) exchange is planned.
Security

1. Features
   - Production level CA configurable
   - VOMS, MyProxy based identity management
   - Access control through permission policy
   - Proxy certificate renewal service \textit{NEW}
   - Authorization service \textit{NEW}

2. Supporting Standards
   - GT4 GSI
   - X.509
   - XACML
Job, VO, Certificates

- Certificate and VO management
  - Production-level CA, APGrid PMA (V2.2: now available on NAREGI download site)
  - VOMS based VO user management

- Certificate Management Server
  - User Certificate
  - Private Key
  - Proxy Certificate with VO

- VOMS

- MyProxy
  - Proxy Certificate with VO

- Get VOMS Attribute
- Put ProxyCertificate with VO

- VOMS-myproxy-init
- ssh + voms-myproxy-init

- Client Environment
  - Portal
  - WFT
  - PSE
  - GVM

- SS client
  - Signed Job Description

- AuthZ service
  - GRAM/GridVM

- AuthZ Policy

- Batch sched.

- Site

- Set AuthZ Policy
- Admin

- National Research Grid Initiative
User Environments

1. Modules
   - Portal: Web based portal for NAREGI Environment
   - Workflow Tool: GUI/CUI based job workflow management tool
   - PSE: Problem Solving Environment
   - GVS: Grid Visualization Service

2. Features
   - Single Sign On
   - GUI and Command based workflow job management
   - GUI based job workflow document management
   - Compile, Test, Deployment service
   - Application repository
   - Visualization using grid resources

3. Supporting Standards
   - OASIS/WSRF
   - OGF/JSDL

National Research Grid Initiative
Registration & Deployment of Applications

Application sharing in Research communities

- Register Application
- Select Compiling Host
- Select Deployment Host
- Compile
- Send back Compiled Application Environment
- Deploy
- Register Deployment Info.

ACS (Application Contents Service)

PSE Server

- Application Summary
- Program Source Files
- Input Files
- Resource Requirements etc.

Application Developer

Server#1
- Compiling: OK!
- Test Run: OK!

Server#2
- Test Run: NG!

Server#3
- Test Run: OK!

National Research Grid Initiative
Description of Workflow and Job Submission Requirements

Workflow Description
By NAREGI-WFML

NAREGI JM I/F module

BPEL
<invoke name=EPS-jobA>
  □ JSDL -A
  □ JSDL -A

BPEL+JSDL

Super Scheduler

GridFTP (Stdout Stderr)

Information Service

DataGrid

Applet-A
JSDL

Applet-B
JSDL

Data icon
Program icon

http(s)

Applet

Web server (apache)

tomcat

Workflow Servlet

Server

NATIONAL RESEARCH INITIATIVE
Use case: RISM-FMO Coupled Simulation

Electronic structure of Nano-scale molecules in solvent is calculated self-consistent by exchanging solvent charge distribution and partial charge of solute molecules.

*Original RISM and FMO codes are developed by Institute of Molecular Science and National Institute of Advanced Industrial Science and Technology, respectively.*
Communication Libraries and Tools

1. Modules
   - GridMPI: MPI-1 and 2 compliant grid ready MPI library
   - GridRPC: OGF/GridRPC compliant GridRPC library
   - Mediator: Communication tool for heterogeneous applications
   - SBC: Storage based communication tool

2. Features
   - GridMPI
     - MPI for a collection of geographically distributed resources
     - High performance optimized for high bandwidth network
   - GridMPI
     - Task parallel simple seamless programming
   - Mediator
     - Communication library for heterogeneous applications
     - Data format conversion
   - SBC
     - Storage based communication for heterogeneous applications

3. Supporting Standards
   - MPI-1 and 2
   - OGF/GridRPC
Grid Ready Programming Libraries

- Standards compliant GridMPI and GridRPC
Communication Tools for Co-Allocation Jobs

- Mediator
  - GridMPI (↔)

- SBC (Storage Based Communication)
  - SBC protocol (↔)
Three layered package installation and management over distributed resources

Layer 1 (multiple-nodes-layer)
All setups are executed from the central node, where configuration information on servers is defined.

Layer 2 (node-layer)
APT: Advanced Packaging Tool
http://apt-rpm.org/

Layer 3 (package-layer)
RPM: Redhat Package Manager
http://www.rpm.org/
NAREGI installation

- Done by the use of APT-RPM packages and tools
- All the operations are invoked from the central node
Summary

• NAREGI middleware realizes to built a virtual single computing environment on geographically distributed computing and storage resources

• NAREGI middleware covered grid computation environment from infrastructure level to application programming libraries level

• NAREGI services and modules are designed and developed using grid standards

• NAREGI middleware will interoperate with other grid environment

• NAREGI Version1 will be distributed as an open source on 2008
Thank you!!