GridWay: Open Source Meta-scheduling Technology for Grid Computing

Ruben S. Montero
dsa-research.org

Open Source Grid & Cluster
Oakland CA, May 2008
Contents

- Introduction
- What is GridWay?
- Architecture & Components
- Scheduling Policies
- Examples of Grid Deployments
- WSRF Interface for GridWay
- Sun Grid Engine Integration
Introduction

- **Resource selection:** Where do I execute my job?
- **Resource preparation:** What do I need?
- **Job submission:** How do I submit my job?
- **Job monitoring:** How is my job doing?
- **Job migration:** Is there any better resource?
- **Job termination:** How do I get my output?
Introduction

- **Meta-scheduler**: Job to resource (other schedulers) matching (*execution management*).

- **Goal**: Optimize the performance according to a given metric (performance model):
  - Global Throughput
  - Resource usage
  - Application (ALS) – Stand-alone, HPC, HTC and self-adaptive
  - User usage

- **Grid characteristics**
  - Heterogeneity (job requirements)
  - Dynamism (high fault rate, load, availability, price)
  - Site autonomy
What is GridWay?

The GridWay meta-scheduler is a scheduler virtualization layer on top of basic Globus services (GRAM, MDS & GridFTP)

For the user
• A LRM-like environment for submitting, monitoring, and controlling jobs

For the developer
• An standard-base development framework for Grid Applications

For the sysadmin
• A policy-driven job scheduler
• User-side Grid Accounting

For the Grid architect / solution provider
• A modular component to use different infrastructures
• A key component to deploy different Grids
What is GridWay?

Globus Projects
- GridWay

Globus Incubator Projects
- Gridshib
- DDM
- LRMA
- GRAADS
- Falkon

... and many many more!
Architecture

- DRMAA
  - .C, .java
  - Applications
    - LRM-like Command Line Interface
    - OGF DRMAA C & JAVA Bindings
    - JSDL (Posix & HTC profiles)
    - Array jobs, DAG workflows and MPI jobs

- Grid Way
  - Grid Meta-Scheduler
    - Advanced (Grid-aware) scheduling policies
    - Fault detection & recovery

- Globus Services
  - Grid Middleware
    - Straightforward deployment (basic services)
    - Globus-based infrastructures
    - Component to deploy different Grids

- Infrastructure
  - PBS, SGE
    - highly dynamic & heterogeneous
    - high fault rate
Scheduling Policies

Grid Scheduling = Job + Resource Policies

Resource Policies

- Rank Expressions
- Fixed Priority
- User Usage History
- Failure Rate

Job Policies

- Fixed Priority
- Urgent Jobs
- User Share
- Deadline
- Waiting Time

Matching Resources for each job (user)
Enterprise Grids

Characteristics

- “Small” scale infrastructures (campus/enterprise) with one meta-scheduler instance
- Resources within the same administration domain that may be running different LRMS and be geographically distributed

Goal & Benefits

- Integrate heterogeneous systems
- Improve return of IT investment
- Performance/Usage maximization
Enterprise Grids

Architecture

Applications
- Users
  - DRMAA interface
  - Portal
  - Command Line Interface
- GridWay
  - One meta-scheduler
  - Grid-wide policies

Middleware
- SGE Cluster
- PBS Cluster
- LSF Cluster

Examples

European Space Astronomy Center
- Data Analysis from space missions
- DRMAA

UABGrid, University of Alabama
- Bioinformatics applications
Partner Grids

Characteristics

◆ “Large” scale infrastructures with one or several meta-schedulers
◆ Resources belong to different administrative domains

Goal & Benefits

◆ Large-scale, secure and reliable sharing of resources
◆ Support collaborative projects
◆ Access to higher computing power to satisfy peak demands
Partner Grids

Architecture

Applications
- Users
- GridWay
- SGE Cluster
- PBS Cluster
- LSF Cluster

Infrastructure
- Multiple Admin. Domains
- Multiple Organizations

(Virtual) Organization
- Users
- Users
- GridWay
- GridWay
- Globus
- Globus
- Globus

Middleware
- Multiple metaschedulers
- (V)Organization-wide policies

Science Gateways
- DRMAA interface

Examples

EGEE-II
- gLite-LHC interoperability
- Virtual Organizations
  Fusion: Massive Ray Tracing
  Biomed: CD-HIT (Workflow)

AstroGrid-D, German Astronomy Community Grid
- Supercomputing resources
- Astronomy-specific resources
- GRAM interface
Globus Interoperability

- Different Middlewares (e.g. WS and pre-WS)
- Different Data/Execution architectures
- Different Information models
- Integration through adapters
- Global DN’s
WSRF Interface for Utility Computing

- **Applications**
  - GRAM
  - RFT
  - MDS

- **GridWay**
- **Middleware Drivers**

- **GLOBUS GRID**
  - Enterprise Grid
  - Partner Grid
  - Other utilities...

- **Standard functionality**
- **Abstraction of the underlying DRMS**
- **Implemented as GRAM jobmanager**
- **Virtualizes a whole Grid**

Consumer/Provider relationships
WSRF Interface for Utility Computing

- Access to different infrastructures with the same adapters
- EGEE managed as other resource

- Delegate identity/ “VO” certificates
- In-house/provider gateway
- Access through legacy applications, portals...

- Regional infrastructure
Transfer Queues

- Communicate LRM systems with meta-schedulers (the other way around)
- Users keep using the same interface, even applications (e.g. DRMAA, site scripts...)

Job submitted to the cluster but executed in the Grid
THANK YOU FOR YOUR ATTENTION !!!

Want to try it... TUTORIAL 11
(Thursday 13:30 – 15:00)

The GridWay Team

- Ignacio M. Llorente
- Ruben S. Montero
- Eduardo Huedo Cuesta
- Javier Fontan
- Jose Herrera
- Tino Vazquez
- Jose Luis Vazquez-Poletti