Grid Scheduling Architectures with Globus GridWay and Sun Grid Engine

Sun Grid Engine Workshop 2007
Regensburg, Germany
September 11, 2007

Ignacio Martin Llorente
Javier Fontán Muiños
Distributed Systems Architecture Group
Universidad Complutense de Madrid
1. Computing Resources
   1.1. Parallel and Distributed Computing
   1.2. Types of Computing Platforms
   1.3. Local Resource Management Systems

2. Globus GridWay Infrastructures
   2.1. Integration of Different Administrative Domains
   2.2. The Globus Toolkit
   2.3. The GridWay Meta-scheduler
   2.4. Grid Scheduling Architectures

3. SGE Transfer Queues to Globus and GridWay
   3.1. Interfaces for Grid Infrastructures
   3.2. From the Cluster to the Grid

4. Demonstrations
   3.1. Enterprise Grid
   3.2. Transfer Queue to GridWay
1. Computing Resources

1.1. Parallel and Distributed Computing

Goal of Parallel and Distributed Computing

- **Efficient** execution of computational or data-intensive applications

Types of Computing Environments

**High Performance Computing (HPC) Environments**

- Reduce the execution time of a single distributed or shared memory parallel application (MPI, PVM, HPF, OpenMP…)
- Performance measured in floating point operations per second
- Sample areas: CFD, climate modeling…

**High Throughput Computing (HTC) Environments**

- Improve the number of executions per unit time
- Performance measured in number of jobs per second
- Sample areas: HEP, Bioinformatics, Financial models…
1. Computing Resources

1.2. Types of Computing Platforms

Centralized
Coupled
- Network Links
- Administration
- Homogeneity

Decentralized
Decoupled

**SMP** (Symmetric Multi-processors)

**MPP** (Massive Parallel Processors)

Clusters

Network Systems
Intranet/Internet

High Performance Computing

High Throughput Computing
1. Computing Resources

1.3. Local Resource Management Systems

Management of Computing Platforms

• Computing platforms are managed by **Local Resource Management (LRM) Systems**
  
1. Batch queuing systems for HPC servers
2. Resource management systems for dedicated clusters
3. Workload management systems for network systems

• There aim is to maximize the system *performance*

<table>
<thead>
<tr>
<th>Independent Suppliers</th>
<th>Open Source</th>
<th>OEM Proprietary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platform Computing</td>
<td>Altair</td>
<td>IBM</td>
</tr>
<tr>
<td>LSF</td>
<td>Open PBS</td>
<td>Load Leveler</td>
</tr>
<tr>
<td>Altair</td>
<td>University of Wisconsin</td>
<td>Cray</td>
</tr>
<tr>
<td>PBS Pro</td>
<td>Condor</td>
<td>NQE</td>
</tr>
<tr>
<td>Sun Microsystems</td>
<td>SGE</td>
<td></td>
</tr>
</tbody>
</table>
1. Computing Resources

1.3. Local Resource Management Systems

LRM Systems Limitations

- Do not provide a common interface or security framework
- Based on proprietary protocols
- **Non-interoperable computing vertical silos** within a single organization
  - Requires specialized administration skills
  - Increases operational costs
  - Generates over-provisioning and global load unbalance

Only a small fraction of the infrastructure is available to the user

Infrastructure is fragmented in non-interoperable computational silos
Contents

1. Computing Resources
   1.1. Parallel and Distributed Computing
   1.2. Types of Computing Platforms
   1.3. Local Resource Management Systems

2. Globus GridWay Infrastructures
   2.1. Integration of Different Administrative Domains
   2.2. The Globus Toolkit
   2.3. The GridWay Meta-scheduler
   2.4. Grid Scheduling Architectures

3. SGE Transfer Queues to Globus and GridWay
   3.1. Interfaces for Grid Infrastructures
   3.2. From the Cluster to the Grid

4. Demonstrations
   3.1. Enterprise Grid
   3.2. Transfer Queue to GridWay
2. Globus GridWay Infrastructures

2.1. Integration of Different Administrative Domains

"Any problem in computer science can be solved with another layer of indirection… But that usually will create another problem." David Wheeler

A New Abstraction Level

“A (computational) grid offers a common layer to (1) integrate heterogeneous computational platforms (vertical silos), that may belong to different administrative domains (systems managed by single administrative authority), by defining a consistent set of abstraction and interfaces for access to, and management of, shared resources”

Grid Middleware

Types of Resources: Computational, storage and network.

Common Interface for Each Type of Resources: User can access a wide set of resources.
2. Globus GridWay Infrastructures

2.1. Integration of Different Administrative Domains

Grid Middleware (a computational view)

- **Services in the Grid Middleware layer:** Security, Information & Monitoring, Data Management, Execution and Meta-scheduling

- **Open Source Software Distributions**
  - gLite: glite.web.cern.ch
  - UNICORE: www.unicore.org
  - OMII: www.omii.ac.uk
  - GRIA: www.gria.org
  - VDT: vdt.cs.wisc.edu
  - Gridbus: www.gridbus.org

- **The Globus Toolkit**
  - Most widely used grid middleware
  - Software distribution that integrates a selected group of **Globus Alliance** technologies (Open Source Community)
2. Globus GridWay Infrastructures

2.2. The Globus Toolkit

Components for a Computational Grid
2. Globus GridWay Infrastructures

2.3. The GridWay Meta-scheduler

Global Architecture of a Computational Grid

- DRMAA
  - Application-Infrastructure decoupling
  - .C, .java

- GridWay
  - Grid Meta-Scheduler
  - open source
  - job execution management
  - resource brokering

- Globus
  - Grid Middleware
  - Globus services
  - Standard interfaces
  - end-to-end (e.g. TCP/IP)

- Infrastructure
  - highly dynamic & heterogeneous
  - high fault rate

- Applications
  - Standard API (OGF DRMAA)
  - Command Line Interface

- Results
  - CLI

- SGE
  - .C, .java

$> CLI
2. Globus GridWay Infrastructures

2.3. The GridWay Meta-scheduler

Benefits

Integration of computational platforms (Organization)

- Establishment of a uniform and flexible infrastructure
- Achievement of greater utilization of resources, which could be heterogeneous
- Higher application throughput

Support for the existing platforms and LRM Systems (Sys. Admin.)

- Allocation of grid resources according to management specified policies
- Analysis of trends in resource usage
- Monitoring of user behavior

Familiar CLI and standard APIs (End Users & Developers)

- High Throughput Computing Applications
- Workflows
2. Globus GridWay Infrastructures

2.3. The GridWay Meta-scheduler

Features

Workload Management

- Advanced (Grid-specific) scheduling policies
- Fault detection & recovery
- Accounting
- Array jobs and DAG workflows

User Interface

- OGF standards: JSDL & DRMAA (C and JAVA)
  - Your DRMAA application also runs on Globus infrastructures!
- Command line interface, similar to that found on local LRM Systems

Integration

- Straightforward deployment as new services are not required
- Interoperability between different infrastructures
2. Globus GridWay Infrastructures

2.3. The GridWay Meta-scheduler

Grid-specific 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)
2. Globus GridWay Infrastructures

2.3. The GridWay Meta-scheduler

The GridWay Project

GridWay is a Globus Project

- Released under Apache license v2.0
- Adhering to Globus philosophy and guidelines for collaborative development
- Welcoming code and support contributions from individuals and corporations around the world

History of the Project

- The project started in 2002
- Since January 2005,
  - 5 stable software releases
  - More than 1,000 downloads from 80 different countries (25% Industry and 75% Academia and Research)
- Best-effort support provided (contract support is also available)
- Widely used: Success stories at http://www.gridway.org
2. Globus GridWay Infrastructures

2.4. Deployment Alternatives

Centralized
Coupled

- Network Links
- Administration
- Homogeneity

Decentralized
Decoupled

SMP (Symmetric Multi-processors)

MPP (Massive Parallel Processors)

Clusters

Network Systems
Intranet/Internet

Grid Infrastructures

High Performance Computing

High Throughput Computing
2. Globus GridWay Infrastructures

2.4. Deployment Alternatives

Enterprise Grid Infrastructures

Characteristics

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

Goal & Benefits

• Integrate multiple systems, that could be heterogeneous, in an uniform/centralized infrastructure
• Decoupling of applications and resources
• Improve return of IT investment
• Performance/Usage maximization

Scheduling

• Centralized meta-scheduler that allows the enforcement of Grid-wide policies (e.g. resource usage) and provides centralized accounting
2. Globus GridWay Infrastructures

2.4. Deployment Alternatives

Deploying Enterprise Grids with GridWay

- Could be heterogeneous and geographically distributed
- Services: MDS, GRAM, GridFTP
- One scheduling instance
- Grid-wide policies
- Could be **heterogeneous** and geographically distributed
- **DRMAA** interface
- **Portal** and/or CLI access

**Infrastructure**

**Middleware**

**Applications**

**Users**
2. Globus GridWay Infrastructures

2.4. Deployment Alternatives

Enterprise Grids: Examples

European Space Astronomy Center

- Data Analysis from space missions (DRMAA)
- Site-level meta-scheduler
- Several clusters
2. Globus GridWay Infrastructures

2.4. Deployment Alternatives

Enterprise Grids: Examples

UABGrid, University of Alabama at Birmingham

- Bioinformatics applications
- Campus-level meta-scheduler
- 3 resources (PBS, SGE and Condor)
2. Globus GridWay Infrastructures

2.4. Deployment Alternatives

Partner Grid Infrastructures

Characteristics

• “Large” scale infrastructures with one or several meta-scheduler instances providing access to resources that belong to different administrative domains (different organizations or partners)

Goal & Benefits

• Large-scale, secure and reliable sharing of resources between partners or supply-chain participants
• Support collaborative projects
• Access to higher computing power to satisfy peak demands

Scheduling

• Decentralized scheduling system that allows the enforcement of organization-wide policies
2. Globus GridWay Infrastructures

2.4. Deployment Alternatives

Deploying Partner Grids with GridWay

- **Infrastructure**
  - Multiple Admin. Domains
  - Multiple Organizations

- **Middleware**
  - Services: MDS, GRAM, GridFTP
  - Multiple scheduling instances
  - (V) Organization-wide policies

- **Applications**
  - DRMAA interface
  - Science Gateways

- **Users**
  - (Virtual) Organization
  - Users

- **GridWay**
  - ....
  - GridWay

- **Globus**
  - ....
  - Globus

- **SGE Cluster**, **PBS Cluster**, **LSF Cluster**
2. Globus GridWay Infrastructures

2.4. Deployment Alternatives

Partner Grids: Examples

AstroGrid-D, German Astronomy Community Grid

- Collaborative management of supercomputing resources & astronomy-specific resources
- Grid-level meta-scheduler (GRAM interface)
- 22 resources @ 5 sites, 800 CPUs
2. Globus GridWay Infrastructures

2.4. Deployment Alternatives

Partner Grids: Examples

**Massive Ray Tracing**

**CD-HIT workflow**

**Fusion**
- Users
- GridWay
- gLite
- SGE Cluster

**Biomed**
- Users
- GridWay
- gLite
- PBS Cluster

- **Services:** BDII, GRAM, GridFTP
- **VO Schedulers**
- **DRMAA interface**
- **EGEE Resource Broker**

**EGEE**
- Enabling Grids for E-sciencE
2. Globus GridWay Infrastructures

2.4. Deployment Alternatives

A Tool for Interoperability

- Different Middlewares (e.g. WS and pre-WS)
- Different Data/Execution architectures
- Different Information models
- Integration through adapters
- Global DN’s
- Demo in June 2007, TeraGrid07
Contents

1. Computing Resources
   1.1. Parallel and Distributed Computing
   1.2. Types of Computing Platforms
   1.3. Local Resource Management Systems

2. Globus GridWay Infrastructures
   2.1. Integration of Different Administrative Domains
   2.2. The Globus Toolkit
   2.3. The GridWay Meta-scheduler
   2.4. Grid Scheduling Architectures

3. SGE Transfer Queues to Globus and GridWay
   3.1. Interfaces for Grid Infrastructures
   3.2. From the Cluster to the Grid

4. Demonstrations
   3.1. Enterprise Grid
   3.2. Transfer Queue to GridWay
3. SGE Transfer Queues to Globus and GridWay

3.1. Interfaces for Grid Infrastructures

Interfaces Provided by Existing Grid Infrastructures

**Grid specific commands & API’s**
- Applications must be ported to the Grid
- Process (submission, monitoring…) must be adapted to the Grid
- New interfaces (e.g. portal) to simplify Grid use

**LRMS-like commands & API’s => GridWay**
- A familiar environment to interact with a computational platform
- SGE-like environment for Computational Grids
- Process still need to be adapted
- Applications would greatly benefit from standards (DRMAA)

*Transfer Queues: Seamless access to the Grid*
3. A Taxonomy for Grid Scheduling Architectures

3.2. From the Cluster to the Grid

From SGE to a Grid Infrastructure or a Cluster (the other way)
3. A Taxonomy for Grid Scheduling Architectures

3.2. From the Cluster to the Grid

Transfer Queues: Seamless access to the Grid

• Access to a grid infrastructure (or remote cluster) on demand driven by SGE scheduling policies
• End users keep the same SGE interface
• Applications running on SGE are able to access the Grid

Transfer Queues: Limitations

• Requirements of system configuration (software, data…) on remote resources for job execution
1. Computing Resources
   1.1. Parallel and Distributed Computing
   1.2. Types of Computing Platforms
   1.3. Local Resource Management Systems

2. Globus GridWay Infrastructures
   2.1. Integration of Different Administrative Domains
   2.2. The Globus Toolkit
   2.3. The GridWay Meta-scheduler
   2.4. Grid Scheduling Architectures

3. SGE Transfer Queues to Globus and GridWay
   3.1. Interfaces for Grid Infrastructures
   3.2. From the Cluster to the Grid

4. Demonstrations
   3.1. Enterprise Grid
   3.2. Transfer Queue to GridWay
4. Demonstrations

4.1. Enterprise Grid

Testbed Configuration

**Information Manager**: Static Discovery & Dynamic Monitoring (MDS2 & MDS4)

**Execution Manager**: Pre-WS and WS GRAM

**Transfer Manager**: GridFTP

Users

GridWay daemon

- Execution Manager
- Transfer Manager
- Information Manager
- Scheduling Module

SGE Cluster
- Aquila

PBS Cluster
- Hydrus

Fork
- Cygnus

SGE Cluster

EGEE

GridFTP

LCG 2 (based on pre-WS GT)

**Globus Toolkit 4.0.3 (WS)**

Globus Toolkit 4.0.3 (WS)
4. Demonstrations

4.2. Transfer Queue to GridWay

Testbed Configuration

Job submitted to the cluster but executed in the Grid
Globus GridWay and SGE

Conclusions

Globus GridWay for SGE Users

Benefits

• Integration of SGE clusters within the organization
• Sharing of SGE clusters between partner organizations
• Provision of computing services to other organizations
• Inter-operability with other LRMS

Deployment Alternatives

• Enterprise grid with a single meta-scheduling instance
• Partner grids with several meta-scheduling instances
• Utility grids to access on demand to remote grids or clusters

Interface Alternatives

• SGE-like CLI, DRMAA API and Portal
• Transfer queues
Thank you for your attention!
Backup Slides
2. Globus GridWay Infrastructures

2.4. Deployment Alternatives

Utility Grid Infrastructures

Characteristics

• Multiple meta-scheduler layers in a hierarchical structure
• Resource provision in a utility fashion (provider/consumer)

Goal & Benefits

• Supply resources on-demand, making resource provision more adaptive
• Access to *unlimited* computational capacity
• Transform IT costs from fixed to variable
• Seamless integration of different Grids (The Grid)

Scheduling

• Each Grid is handled as any other resource
• Characterization of a Grid as a single resource
• Use standard interfaces to virtualize a Grid infrastructure
2. Globus GridWay Infrastructures

2.4. Deployment Alternatives

Deploying Utility Grid Infrastructures with GridWay

globus-job-run, Condor/G, Nimrod/G ...

Access to Outsourced Resources

Virtualization of a Grid

GRID-GATEWAY

Globus

SGE Cluster

PBS Cluster

LSF Cluster

GridWay

Users

Globus

Users

Globus

GridWay

Utility Grid
2. Globus GridWay Infrastructures

2.4. Deployment Alternatives

Utility Grids: Example

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

- Delegate identity/ “VO” certificates
- In-house/provider gateway

- Regional infrastructure