Overview of the NorduGrid Information System

Balázs Kónya

3rd NorduGrid Workshop
23 May, 2002, Helsinki
Simple Model of the Grid

Data Management

Resource & Job Management

Information System

+ security
The Information System

a) resource discovery

b) resource characterization / description

c) monitoring of services / resources

the nerve system of the Grid
information is a critical resource on the Grid
The challenge

- large number of resources
  => scalability
- diverse heterogeneous resources
  => characterization?
- decentralized, automatic maintenance
- efficient access to dynamic data
- quality and reliability of information
  => fake information can 'kill' the Grid
Grid users always want **prompt** access to all the information.

Inevitable compromise:

- load on the Grid $\iff$ up-to-dateness

- try to avoid continuous monitoring
- generate information on demand (pull model)
- apply elaborate caching and keep track of validity of the data
- organize information into some kind of hierarchy
Monitoring and Discovery Services (MDS 2.1) from the Globus Toolkit™ version 2.0

- a general framework for creating Grid Information systems
- OpenLDAP based
- provides the building blocks of an Information System
  - information model (resource characterization)
  - providers (to “create” information)
  - GRIS (Grid Resource Information Service) LDAP backend (to present the information of the providers)
  - GIIS (Grid Index Information Service) LDAP backend (link together GRISes, build hierarchy, caching)
Fully MDS-based:

- new information model (LDAP schema)
  - natural description of the Grid elements of our architecture
- inadequate existing models:
  - Globus schema:
    - single-machine orientated
  - EDG schema:
    - queue-specific
  - NorduGrid schema:
    - natural computing unit is a cluster
    - queue & job & user entries

Common Grid schema

-
NorduGrid Information System: Providers

- information providers
  - efficient scripts which collects status information about the resources
  - populates the entries of the NorduGRid schema
  - on-demand execution
  - the GRIS LDAP backend caches the infoprovider outputs in order to avoid system overloading
  - the script outputs are grouped into LDAP entries, and the entries are organized into the LDAP tree of the GRIS
NorduGrid Information System: Hierarchy

Hierarchy of GRISes/GIISeS
available information: cluster entry

• cluster-name, cluster-aliasname, cluster-contactstring
• cluster-architecture
• cluster-totalcpu, cluster-cpudistribution (1cpu:3 2cpu:4)
• cluster-homogeneity, cluster-nodecpu, cluster-nodememory
• cluster-opsys
• cluster-lrmstype, cluster-lrms-version, cluster-lrms-config
• cluster-gridarea, cluster-gridspace
• cluster-middleware, cluster-runtimeenvironment
• Cluster-support

Mds-validfrom: 20020522150633Z
Mds-validto: 20020522150703Z
available information:
queue entry

- assignedcpunumber, assignedcputype
- defaultcputime, mincputime, maxcputime
- maxqueuable, maxrunning, maxuserrun
- schedulingpolicy
- status, running, queued

dynamic queue information is obtained from qstat, static values are set by the local sysadmin
user based information is essential on the Grid:

- user is not really interested of the total number of cpus at Lund, but how many of those are available for her!
- number of queuing jobs are irrelevant if her submission gets immediately executed
- instead of total disk space she needs to know her quota
- **nordugrid-authuser** objectclass is designed to satisfy this requirement
  - freecpus
  - disk space
  - queue length
every Grid job in the NorduGrid system has a job entry:

- the job entry is generated on that cluster where the job is submitted
- job entries are generated by interfacing to the Grid-Manager and to the PBS information
- when the job is completed and the output is retrieved the job disappears from the MDS
- job status monitoring <=> Information System query

available information:

the job entry is generated on that cluster where the job is submitted.

job entries are generated by interfacing to the Grid-Manager and to the PBS information.

when the job is completed and the output is retrieved the job disappears from the MDS.

job status monitoring <=> Information System query
available information: job entry

attributes:

- globalowner, globalid, jobname
- execcluster, execqueue
- submissiontime, submissionui
- reqcput
- stdin, stdout, stderr
- usedcputime, usedwalltime, usedmem
- status, queue_rank, pbsjob-comment, errors
- sessiondir_ererasetime
available information:  
SE, RC

• Storage Element
  • se-baseurl
    gsiftp://bambi.quark.lu.se:2811/gamma/scratch
  • se-freespace
  • se-authuser

• Replica Catalog
  • rc-baseurl
    ldap://grid.uio.no:389/rc=NorduGrid,dc=nordugrid,dc=org
  • rc-authuser

these objectclasses are 'under construction' they will gain real importance with the NorduGrid Storage Manager
Security aspects

- MDS inherits OpenLDAP Version 3 security infrastructure
  - SASL (Simple Authentication and Security Layer) with GSS-API
  - secure authentication and message protection
  - GSI-GSSAPI is the Globus 'security layer' for the MDS
- In principle, user certificates can be used for accessing 'sensitive' or 'protected' information
- **However:**
  - only static ACL's (LDAP rules for who can access what)
  - the security is 'lost' between GRIS -- GIIS queries, queries initiated by GIISes are not authenticated
  - not much of use at the moment...
The NorduGrid providers are well-tested simple scripts

OpenLDAP is a robust server

A single GRIS backend seems to be stable

We are experiencing stability problems with the GIIS backend

- Stress tests were performed:
  - 3 sites were continuously registering in every 10 seconds
  - Continuous LDAP query loaded the same GIIS from two clients
  - The central GIIS was crashing regularly

- During normal daily operations, it occasionally dies

A hierarchic MDS 2.1 has never been widely deployed/tested

- Problems are continuously reported to Globus, some of them are fixed

- Always use the latest MDS release
interfaces to the infosys

NorduGrid Cluster Load Monitor

Denmark
- Copenhagen Grid Cluster
  - CPUs: 2
  - Cluster load: 2
  - Queue: 0
- Copenhagen LSCF Cluster
  - CPUs: 2
  - Cluster load: 3
  - Queue: 0

Norway
- Bergen Grid Cluster
  - CPUs: 1
  - Cluster load: 1
  - Queue: 0
- Oslo Grid Cluster
  - CPUs: 3
  - Cluster load: 3
  - Queue: 0

Terminal <2>

```
[konyab@grid.quark.lu.se /]$ ngrsh -d 2 -c grid.nbi.dk -f mandelbrot.xml -g ldap://grid.uio.no/mds-vo-name=norway,o=grid
```

Terminal <2>

```
[konyab@grid.quark.lu.se /]$ ldapsearch -h grid.nbi.dk -p 2135 -b "mds-vo-name=denmark,o=grid" -x
```
Summary

- NorduGrid operates an MDS based, hierarchically distributed Information System:
  - new information model for clusters, queues, jobs, users, SE, RC
  - efficient providers
  - all the job monitoring, resource discovery, status monitoring and brokering are exclusively built on top of the MDS
  - MDS hierarchy with dynamic site registrations
Future...

- hope for a common Grid schema
- add dynamic security
- solve of the GIIS instability problem
- fix the GIIS registration shortcomings (invitation, access control)
- add support for further Local Resource Management Systems
- restructure the GRIS DIT, enhance the providers
- develop the Storage Element, Replica Catalog information model
- introduce accounting
- add more sites for GIIS testing
  - scalability, fault tolerance
  - build more complex hierarchy
  - performance tests

... and during the DEMO