USAGE STATISTICS AND USAGE PATTERNS ON THE NORDUGRID ANALYZING THE LOGGING INFORMATION COLLECTED ON ONE OF THE LARGEST PRODUCTION GRIDS IN THE WORLD.

K. Pajchel[*], Dr. U. Erkarslan, Prof. F. Ould-Saada, University of Oslo, Norway

Dr. M. Ellert, CERN, Geneva, Switzerland

Prof. P. Eerola, Dr. B. Kónya, Dr. O. Smirnova, Lund University, Sweden

Prof. J. R. Hansen, Dr. J. L. Nielsen, Dr. A. Wäänänen, Niels Bohr Institute, Copenhagen, Denmark

Prof. T. Ekelöf, Uppsala University, Sweden

Dr. A. Konstantinov, Vilnius University, Lithuania and University of Oslo, Norway

Abstract

The Nordic Grid facility (NorduGrid [1]) came into operation during summer 2002 when the Scandinavian ATLAS HEP group started to use the Grid for the ATLAS Data Challenges (DC) and was thus the first Grid ever contributing to an ATLAS production. Since then, the Grid facility has been in continuous 24/7 operation.

NorduGrid is being used by a growing set of active users from various scientific areas including physics, chemistry, biology and informatics. It has given major contributions to the ATLAS Data Challenge 1 [2] and the ongoing Data Challenge 2.

- The increasing number of resources has made NorduGrid one of the largest production Grids in the world, continuously running on more than 30 sites more than 3000 CPUs.
- The resources range from small test clusters at academic institutions to large farms at several supercomputer centers and the NorduGrid software runs on clusters with very different Linux distributions.

This presentation gives a short overview of the design and implementation of the NorduGrid middleware, logging and monitoring facilities. It will be followed by a description of a typical job on NorduGrid and the information about its parameters which are monitored on-line and persistified in the logging service.

MIDDLEWARE

The NorduGrid middleware (or Advanced Resource Connector, ARC [3]) is an open source software solution distributed under the GPL license, enabling production-quality computational and data Grids. It is a light-weight Grid solution, designed to support a dynamic, heterogeneous Grid facility, spanning different computing resources and user communities.

Emphasis is put on scalability, stability, reliability, robustness, and performance of the middleware. These goals of the architecture of the ARC imply concrete decentralization and functional independence of the components.

NorduGrid provides innovative solutions essential for a production quality middleware:

- the Grid Manager, gridftpd (the ARC/NorduGrid GridFTP server),
- the information model and providers (NorduGrid schema),
- User Interface and broker (a "personal" broker integrated into the user interface),
- extended Resource Specification Language (xRSL),
- logging service,
- and the monitoring system.

Most of these services are provided through the security layer of the GSI. The middleware builds upon standard open source solutions like the OpenLDAP, OpenSSL, SASL and Globus Toolkit 2 (GT2) libraries.

ARC main components are:

- Grid services running on the resources: the Grid Manager, the gridftpd and the information service.
- Indexing services for the resources and data.
- Client making intelligent use of the distributed information and data available on the Grid.

Distribution

The middleware is available as source and binary in tarballs and RPMS.
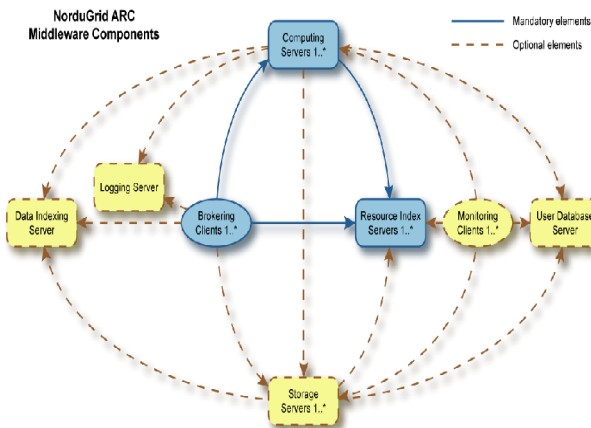
* katarina.pajchel@fys.uio.no

Figure 1: NorduGrid's ARC middleware components.

LOGGING

A Logger service is one of the Web services implemented by NorduGrid and based on gSOAP and Globus IO API. It provides a frontend to the underlying MySQL database to store and retrieve information about computing resources' usage (jobs).

This database can be accessed through a graphical Web interface implemented using PHP4, JavaScript and JPGraph based on GD library.

The main goals for the Logger:

- Provide information about the development and usage of NorduGrid over time.
- Provide statistics for different clusters, time intervals, applications and users.

```
columns:
start      2004-06-04 16:32:15
end        2004-06-05 00:07:16
cluster    /O=Grid/O=NorduGrid/CN=host/lscf.nbi.dk
user       Name of the user
id         46236674234684697
name       name of the job
failure    Job exit code is 265 != 0
lrms       pbs, condor
queue      fastq, slowq, gridlong
rsl        content of the .xrsl file
ui         129.240.86.35:57818;susy.uio.no
```

Figure 2: Content of the Logger database. The columns shown are temporary. The database is being extended.

THE GRID MONITOR

The Grid Monitor allows the user and system administrators alike to monitor and check in an intuitive way the status of the ARC Grid facility and the information stored in the Information System. The structure of the Monitor is built upon the following objects:

- nordugrid-cluster: a cluster,
- nordugrid-queue: a queue at the cluster, accessible by the ARC users,
- nordugrid-job: a Grid job, associated with a queue,
- nordugrid-authuser: a user, authorized to submit jobs to a given queue

It also displays information about storage facilities, authorised users, and groups of users (Virtual Organisations), using the corresponding objects.

A SCHEMATIC EXAMPLE OF A TYPICAL NORDUGRID JOB

```
(&
(executable="myprog")
(inputFiles=
("myprog"
        "http://www.myserver.org/myfiles/myprog")
("myinputfile"
        "gsiftp://www.mystorage.org/data/file007.dat")
)
(outputFiles=
("myoutput"
        "gsiftp://www.mystorage.org/results/file007.res"
)
)
(disk=1000)
(notify="e myname@mydomain.org")
)
```

Once a proxy is obtained, jobs can be submitted to NorduGrid using simple commands and the user can keep track of the job through the whole process.

- Submitting:
  > ngsub -f myjob.xrsl
  The user can now follow the job on the Monitor
- Check the status of the job:
  >ngstat [options] [jobs]
- Retrieve the results of a finished job:
  >ngget [options] [jobs]

The job usage record gets logged in the logger database by the Grid Manager and can later be displayed using the Web based Logger.

GRAPHICAL INTERFACES

Nordugrid offers intuitive access to the system data through two different Web-based graphical interfaces. The users can easily follow their jobs and obtain information about the performance of the system and job statistics.

Grid Monitor

The Monitor is the starting point for browsing the system information. It shows the current state of the system. Grid jobs are not kept infinitely in the information system. Those whose lifetime has expired are removed from clusters and thus they disappear from the monitor too. All the elements in the Monitor are links to the information about the different objects in the system.

Figure 3: 1) Cluster alias 2) Cluster load showing both Grid- and non-Grid jobs 3) Queuing jobs

Figure 4: Example of cluster information
1) Attributes: dump of all the attributes of the nordugrid-cluster object. 2) Queues at a given cluster.

NGLogger

The NGLogger is a graphical interface to the underlying MySQL database. The logger service is the persistent database for job usage records. It provides information which is complementary to the Monitor, showing the history of the NorduGrid usage. The queries are based on different basic approaches to the database like cluster, time period, application or user.
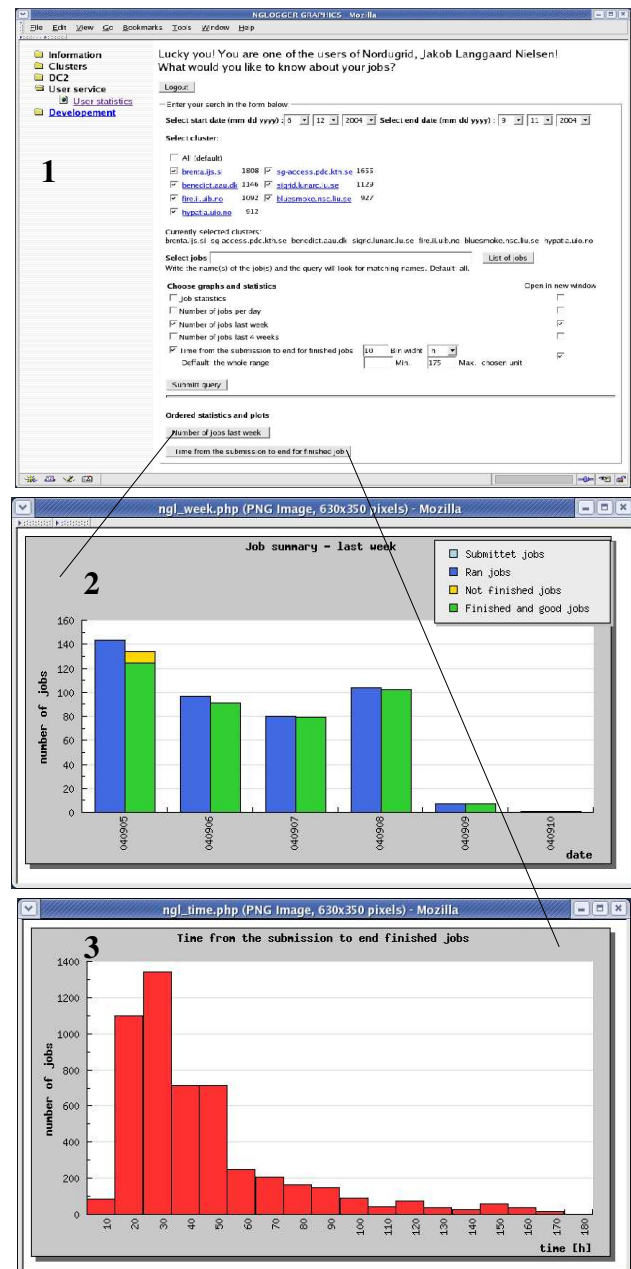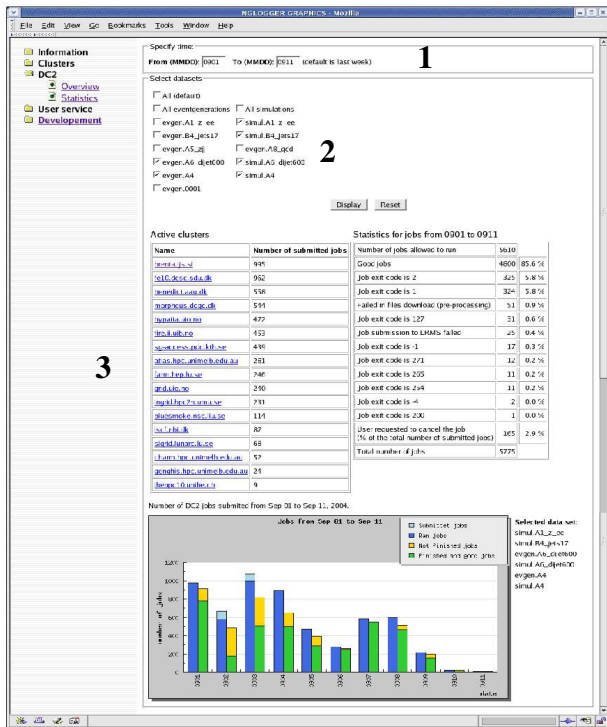

Figure 5: NGLogger
The user can get access to information about his or her usage of NorduGrid through a search form.
1) Search form.
2) Number of jobs.
3) Time from submission of the job until it is finished.

Figure 6: Data Challenge 2 on NorduGrid
Parameters to be set by the user:

1) time period
2) dataset
3) cluster

CONCLUSION

NorduGrid offers a user-friendly monitoring service showing the present status of the system as well as a complementary tool, the logger service, which provides the history of the activity on the Grid. Both services are web based and offer graphical interfaces to the usage data.

REFERENCES

[1] NorduGrid live: **http://www.nordugrid.org** - Grid Monitor and NGLogger

[2] Eerola, P; Ekelof, T; Ellert, M; Hansen, J R; Hellman, S; Konstantinov, A; Konya, B; Myklebust, T; Nielsen, J L; Ould-Saada, F; Smirnova, O; Wäänänen, A, *Atlas Data-Challenge 1 on NorduGrid* [arXiv:physics/0306013]

[3] The NorduGrid/ARC User Guide, Available: http://www.nordugrid.org/documents/userguide.pdf
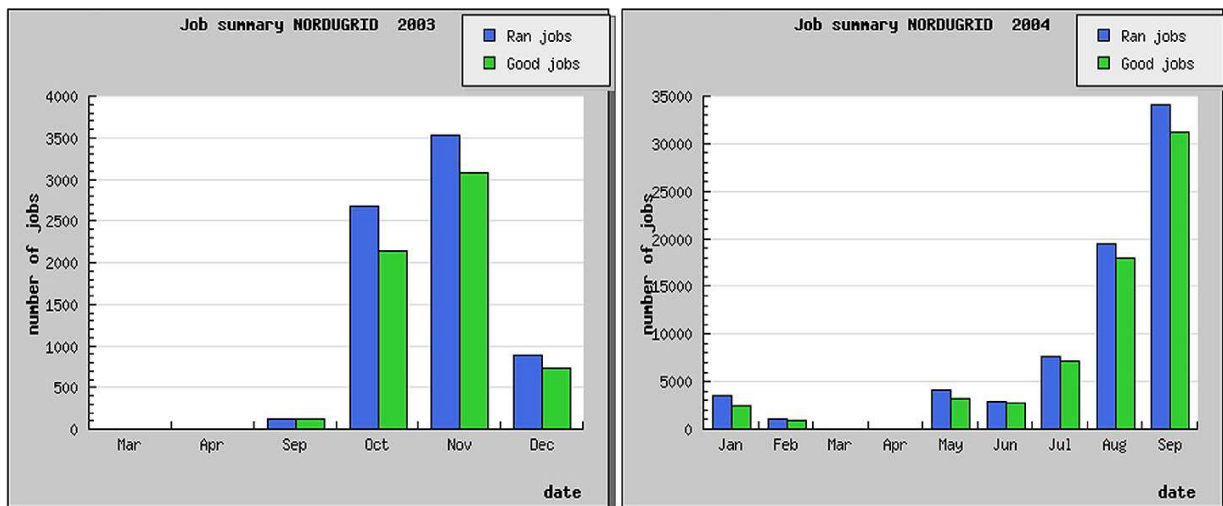
Figure 7: Number of jobs recorded in the logger. Over 30 000 jobs were run successfully in September 2004. Notice that it is not the Grid, but the experimental logger service which was down in March and April. The early information may be incomplete.