The NorduGrid Toolkit: Overview and architecture

The 4th NorduGrid Workshop
November 11th 2002, Uppsala

Anders Wäänänen <waananen@nbi.dk>
Project Overview

- Launched in spring of 2001, with the aim of creating a Grid infrastructure in the Nordic countries
- Partners from Denmark, Norway, Sweden, and Finland
- Initially meant to be the Nordic branch of the EU DataGrid (EDG) project
- Currently relies on very limited human resources (3 full-time researchers, few part-time ones) with funding from NorduNet2 - now NOS-N
Resources & organization

- 5 dedicated Linux *test-clusters* (3 to 5 CPUs each) + a couple of stand-alone machines. Several production clusters available and several sites showing interest in collaboration

- **NORDUNet** network

- The *steering group* and the *technical working group*, ca. 5 persons each

- *Most of the communications are done via the Internet or phone conferences; the technical working group convenes bi-monthly at different sites*

- *Plenary workshops twice a year*
**The development**

- NorduGrid **Authentication System** was put into operation in May 2001 - officially accepted by the EDG.
- The **fabric** of the NorduGrid was laid down by June 2001.
- The first middleware was deployed and the sites were **Grid-enabled** by July 2001.
- Further **Grid services** were put into operation (November-December 2001):
  - NorduGrid User Management System (Virtual Organization)
  - NorduGrid Information System
  - Grid Data Mirroring Package (GDMP)
  - Data replication catalog
- Deployment & evaluation of the first (Testbed 1) release of the **EDG Middleware** (December-January).
Facing Reality

- NorduGrid was only an 18 months project compared to 3 years for EU DataGrid

- Expected to run the ATLAS Data Challenge on a working Grid testbed in May 2002 in the Nordic countries

- Continuing problems with EDG testbed stability

- Architecture problems with bottlenecks and fragile system components

- The urgent need to have something stable and working resulted in the decision to create a new architecture not necessarily compatible with EDG
NorduGrid Philosophy 1

- No single point of failure
- Resource owners have full control over their resources
- Installation details should not be dictated
  - Method, OS version, configuration, etc...
- As little restriction on site configuration as possible
  - Compute nodes should not be required to be on the public network
  - Clusters need not be dedicated
- NorduGrid software should be able to use existing system software and Globus installation
  - Though Globus RPMs provided
NorduGrid Philosophy 2

- Portability
  - Support toolkit on variety of ia32 Linux platforms
  - Help people port the software to their own systems
- Always choose the best technical workable solution
  - Not afraid to throw things away
  - Not tied to solutions because of political reasons
  - Open for new ideas
- Start with something simple that works and proceed from there
- Solutions should be scalable and stable
NorduGrid components

- Grid manager
  - Complete replacement of Globus Resource management

- Information system
  - Based on Globus MDS, but with improved schema

- User interface with broker

- Globus Replica catalog for data management

- Load Monitor

- Use Globus libraries to build a working Grid
  - Globus RPM distribution based on a patched Globus 2.2.2
  - Can also use existing Globus installation
  - Integrate Globus transport protocols into applications
First Grid manager schematics
# Load Monitor - old

## NorduGrid Cluster Load Monitor

**Mon Aug 5 21:22:50 CEST 2002**

- **Processes:** Grid Other

### Denmark

<table>
<thead>
<tr>
<th>Cluster</th>
<th>CPUs</th>
<th>Load (processes: Grid+other)</th>
<th>Queueing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copenhagen Alpha Linux Cluster</td>
<td>1</td>
<td>3+6</td>
<td>0</td>
</tr>
<tr>
<td>Copenhagen Grid Cluster</td>
<td>6</td>
<td>3+6</td>
<td>0</td>
</tr>
<tr>
<td>Copenhagen LSCF Cluster</td>
<td>49</td>
<td>32+1</td>
<td>18</td>
</tr>
<tr>
<td>Copenhagen Nordita Cluster</td>
<td>14</td>
<td>3+6</td>
<td>0</td>
</tr>
</tbody>
</table>

### Norway

<table>
<thead>
<tr>
<th>Cluster</th>
<th>CPUs</th>
<th>Load (processes: Grid+other)</th>
<th>Queueing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bergen Grid Cluster</td>
<td>4</td>
<td>4+6</td>
<td>2</td>
</tr>
<tr>
<td>Oslo Grid Cluster</td>
<td>18</td>
<td>2+6</td>
<td>0</td>
</tr>
<tr>
<td>Parallab IBM Cluster</td>
<td>62</td>
<td>3+18</td>
<td>0</td>
</tr>
</tbody>
</table>

### Sweden

<table>
<thead>
<tr>
<th>Cluster</th>
<th>CPUs</th>
<th>Load (processes: Grid+other)</th>
<th>Queueing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lund Grid Cluster</td>
<td>4</td>
<td>3+6</td>
<td>2</td>
</tr>
<tr>
<td>Uppsala Grendel Cluster</td>
<td>16</td>
<td>11+8</td>
<td>0</td>
</tr>
</tbody>
</table>

*User base*
**Load Monitor - from today**

### NorduGrid Cluster Load Monitor

**2002-11-11 CET 11:53:24**

<table>
<thead>
<tr>
<th>Cluster</th>
<th>CPUs</th>
<th>Load (processes: Grid+other)</th>
<th>Queueing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Morpheus (iscf)</strong></td>
<td>49</td>
<td>0+2</td>
<td>0</td>
</tr>
<tr>
<td><strong>Sandman</strong></td>
<td>6</td>
<td>0+0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Nordita</strong></td>
<td>12</td>
<td>0+0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Preacher</strong></td>
<td>1</td>
<td>0+0</td>
<td>0</td>
</tr>
<tr>
<td><strong>pc30 (Kumpula test)</strong></td>
<td>1</td>
<td>0+0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Oslo Grid Cluster</strong></td>
<td>3</td>
<td>0+0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Bergen Grid Cluster</strong></td>
<td>4</td>
<td>0+0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Ingvar (NSC)</strong></td>
<td>32</td>
<td>0+21</td>
<td>17</td>
</tr>
<tr>
<td><strong>Seth (HPC2N)</strong></td>
<td>236</td>
<td>0+236</td>
<td>43</td>
</tr>
<tr>
<td><strong>Lund Quark Cluster</strong></td>
<td>6</td>
<td>1+0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Uppsala ISV Cluster</strong></td>
<td>4</td>
<td>0+0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Uppsala Grendel Cluster</strong></td>
<td>13</td>
<td>0+10</td>
<td>1</td>
</tr>
</tbody>
</table>

**User base:** NorduGrid, Guests, Tutorial
**NorduGrid Features**

- Dynamic Information System, Brokering, Monitoring
- Independence on Globus GASS cache (and it’s bug)
- Own GridFTP server, pluggable (and it submits jobs, too)
- Stable and tested Grid testbed
- Not Nordic specific - can be used anywhere
- Not ATLAS or HEP specific
- Tested on RedHat 6.2, 7.2 (also Alpha), Mandrake, Debian, Slackware
- Can share resources with non Grid applications
- Has been running ATLAS DC1 since May
- Can brew coffee + many other things
ATLAS data challenge

- NorduGrid ran 15 partitions of partition 2000 and all of 2003
  - 2000 - 125799 events
  - 2003 - 161497 events
- ATLAS software was recompiled on the different operating system using the Nordic RPMs (relocatable, can be recompiled from SRPMs)
- Each site had to undergo thorough validation
- Input retrieved from CASTOR using bbftp and distributed manually across various sites.
  - In total: 200 GB
- Output stored in Oslo and registered in Replica Catalog.
  - In total: 765 GB
- Processing time: ca. 530 CPU-days (700 MHz < speed < 1500 MHz)
- Everything was run using NorduGrid tools
- 8 different sites:
  - Bergen, Grendel, Ingvar, ISV, NBI, Oslo, Lund, LSCF
&

(executable="$ATLAS_ROOT/bin/athena.exe")
(arguments="HelloWorldOptions.txt")
(stdout="helloworld.log")
(join="yes")
(inputfiles=
  ("HelloWorldOptions.txt" "http://www.nbi.dk/~langgard/HelloWorldOptions.txt")
)
(outputFiles=
  ("helloworld.log" "helloworld.log")
)
(jobname="Athena-HelloWorld")
(runTimeEnvironment="ATLAS-4.0.1")
(middleware="nordugrid-0.3.10")
(maxCPUtime=1000)
(maxDisk=1200)
(ftpThreads=10)
(notify="e waananen@nbi.dk")
DC1 bad Zebra bank rerun

&
(executable="ds2003.sh")
(arguments="364" "10000" "1000000")
(stdout="dc1.002003.simul.00364.hlt.pythia_jet_11.log")
(join="yes")
(rsl_substitution="RCINP" "rc://@grid.uio.no/lc=Dataset_2003,rc=Nordugrid,dc=nordugrid,dc=org")
(rsl_substitution="RCOUT" "rc://@grid.uio.no/lc=Output_2003,rc=Nordugrid,dc=nordugrid,dc=org/2003")
(inputfiles=
"ds2003.sh" "http://www.nbi.dk/~waananen/dc1.002003.simul.NG.sh")
(dc1.002003.evgen.0037.hlt.pythia_jet_11.root $(RCINP)/dc1.002003.evgen.0037.hlt.pythia_jet_11.root)
)
(outputFiles=
(dc1.002003.simul.00364.hlt.pythia_jet_11.log $(RCOUT)/log/dc1.002003.simul.00364.hlt.pythia_jet_11.log)
(atlas.00364.zebra $(RCOUT)/zebra/dc1.002003.simul.00364.hlt.pythia_jet_11.zebra)
(atlas.00364.his $(RCOUT)/his/dc1.002003.simul.00364.hlt.pythia_jet_11.his)
(dc1.002003.simul.00364.hlt.pythia_jet_11.AMI $(RCOUT)/ami/dc1.002003.simul.00364.hlt.pythia_jet_11.AMI)
(dc1.002003.simul.00364.hlt.pythia_jet_11.MAG $(RCOUT)/mag/dc1.002003.simul.00364.hlt.pythia_jet_11.MAG)
)
(jobname="dc1.002003.simul.00364.hlt.pythia_jet_11")
(runTimeEnvironment="ATLAS-3.2.1")
(runTimeEnvironment="DS2003")
(middleware="nordugrid-0.3.4")
(maxCPUPTime=1000)(maxDisk=1200)(ftpThreads=10)(notify="e waananen@nbi.dk")
NorduGrid job submission

Diagram showing the network connected components:
- RC
- SE
- MDS
- FTP

Connections and links are indicated by lines between the components.
Software

- All software accessible through ftp.nordugrid.org
  - Globus RPMs (3)
    - Configuration also used by EDG
  - NorduGrid RPMs (2)
  - Nightly builds from CVS
  - Certificate Authority RPMs
    - Shared by EDG
  - Applications: ATLAS
- All RPMs are accompanied with SRPMs for rebuilding
- All RPMs are relocatable
Quick server installation instructions

- Install 3 Globus RPMs (gpt, main, config)
- Install the NorduGrid server and client RPMs.
- Modify 2 configuration files (globus.conf, nordugrid.conf)
- Install pre-fetched host certificate
- Install appropriate Certificate Authority RPMs
- Create a mapping between authorized Grid users and local users
- Start the gridftp, grid-manager and Globus MDS daemons
- Run grid-proxy-init and submit a job

- A ~5 minutes exercise!
Quick client installation/job run

- As a normal user:
  - **retrieve** nordugrid-standalone-0.3.10.tar.gz
    `tar xvfz nordugrid-standalone-0.3.10.tar.gz`
  - `cd nordugrid-standalone-0.3.10`
  - `source ./setup.sh`
  - `grid-cert-request`
  - **install certificate per instructions**
    `grid-proxy-init`
    `ngsub '&(executable=/bin/echo)(arguments="Hello World")'`
Relationship to EDG

- Lot of overlap with EDG:
  - People (WP6, WP8)
  - Certificate Authorities
  - Similar authorization system using VO servers
  - Globus configuration
  - Contributed secure Replica Catalog
  - Patches to GDMP

- In the future:
  - Common information system - hopefully
Outside relationships

- We have not had the habit of issuing press releases when we succeed in “Hello World” type of jobs
Future work

- Authorization
- Accounting
- Optimize brokering
- Improve data management and replication service
- Parallel jobs (MPI, PVM, ...)
- Handle network requests from running jobs on “private” networks
The committee of Nordic natural science ministers NOS-N has decided to fund a new common Nordic Grid Project based on the work done by the NorduGrid project. This project should work on a proposal/recommendation for a Nordic DataGrid facility. The funding amounts to 8 MDKK (approx 1.1 M€).

Support for the toolkit in the future

This will be supported in each country by local Grid initiatives

Collaboration with the Nordic computing centers have already been initiated with the deployment of the toolkit on several large centers.

Use it for future ATLAS (maybe CMS) production in the Nordic countries

Further collaborate with other Grid projects especially EDG to increase interoperability
Resources

- Documentation and source code are available for download
- Main Web site:
  - http://www.nordugrid.org/
- Repository
The NorduGrid core group

- Александр Константинов
- Balázs Kónya
- Mattias Ellert
- Оксана Смирнова
- Jakob Langgaard Nielsen
- Anders Wäänänen
The usual suspects