Towards Sustainability: An Interoperability Outline for a Regional ARC based infrastructure into the EGEE infrastructure

Michael Grønager, PhD
NDGF Director

International Symposium on Grid Computing 09
Taipei, April 21st 2009 (Video from Copenhagen)
Outline

- History, Motivation and Goal
- Operation
- Interoperability of Services
  - Registration
  - Indexing
  - Monitoring
  - Accounting
  - Job submission
- Conclusion and Future
History

- NDGF
  - Nordic Production Grid since 2004
  - Based on ARC, dCache, and gLite components
  - Operates the biggest European ATLAS T1

- NorduGrid ARC
  - Initiated in 2001 as “Nordic EDG”
  - In production since 2002
  - Only grid in CERN ATLAS DC1
Nordic DataGrid Facility

Nordic Participation in Big Science:

- WLCG – the Worldwide Large Hadron Collider Grid
- Gene databases for bio-informatics sciences
- Screening of CO2-Sequestration suitable reservoirs
- Computational Chemistry
- Common Nordic User Administration, Authentication, Authorization and Accounting
- Other...
Motivation

- Nordic:
  - Resources are: Shared, distributed and heterogeneous (os'es, batch sys., setup)
  - Need for support of WLCG VOs
  - ARC optimal for this

- European:
  - More non-dedicated resources can be grid enabled
  - One of the UMD m/w

- Global:
  - HPC and HTC resources on grid
  - Optimal resource usage
Pave the way for the European Grid Initiative:

- Make a stable and production quality interoperability solution for Nordic and European grid users
  - Enable the CMS experiment to use the shared Finnish Tier-2 resources
- Get experiences with joint operation between the European wide and a larger regional grid initiative (NDGF)
- Ensure knowledge about which services are needed to gain full interoperability
- Understand the grids – pave the way for standards
Joint Operation

- Joint operation between the NDGF operation team and the North European ROC (NE-ROC):
  - Optimization of the Nordic Grid Operation
  - Merge the efforts for operation of single sites and smaller and larger virtual sites
    - The Nordic Tier-1: a virtual site build from 7 distributed sites in 4 countries
    - The Swedish and Norwegian Tier-2s: 2-3 sites in each country
- Nordic EGEE SA1 and NDGF takes shifts every other week
  - Close collaboration
  - Proposed mode of work also for post EGEE
Service Registration

- The GOCDB is the central registration point for grid services
- A service needs a service type in the GOCDB to be recognized on the grid
- The ARC-CE service type was added to the GOCDB

<table>
<thead>
<tr>
<th>Service Type</th>
<th>Location</th>
<th>NDGF 3D node</th>
<th>Front-end for Steno cluster</th>
<th>Ritsem ARC-CE</th>
<th>Classic-SE</th>
<th>SRM</th>
<th>VOMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site-BDII</td>
<td>site-bdii.ndgf.org</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARC-CE</td>
<td>arc-ce.smokemc.nsc.liu.se</td>
<td>yes</td>
<td></td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARC-CE</td>
<td>ce01.titan.uio.no</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARC-CE</td>
<td>db1tier1.ndgf.org</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FTS</td>
<td>fts001.nsc.liu.se</td>
<td></td>
<td></td>
<td></td>
<td>yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARC-CE</td>
<td>gateway01.dcsc.ku.dk</td>
<td></td>
<td></td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARC-CE</td>
<td>jeannedarc.hpc2n.umu.se</td>
<td></td>
<td></td>
<td>yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classic-SE</td>
<td>srm.ndgf.org</td>
<td></td>
<td></td>
<td></td>
<td>yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VOMS</td>
<td>voms.ndgf.org</td>
<td></td>
<td></td>
<td></td>
<td>yes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Service Indexing

- The BDII is the hierarchy for service indexing – The Information System
- All services need to publish information in GLUE
- Since 2006 an arc-bdii translating service has been running
- Today the ARC-CE supports direct publishing of GLUE

```
gateway01.dcsc.ku.dk:2811/nordugrid-LoadLeveler-tier1
GlueCEStateStatus: Production
GlueCEPolicyMaxRunningJobs: 880
GlueCEPolicyMaxWallClockTime: 5760
GlueHostOperatingSystemName: CentOS
GlueHostOperatingSystemRelease: 4.4
GlueHostOperatingSystemVersion: Final
GlueSubClusterPhysicalCPUs: 2284
GlueSubClusterLogicalCPUs: 2284
```
The SAM tests (Service Availability Monitoring) is the de facto standard for service monitoring

Important services needs SAM Sensor

An ARC-CE SAM test was developed, reviewed by a WLCG-MB review group and is since 2007 used in production
Accounting for all sites in EGEE are summed up in APEL
The ARC-CE sites uses SGAS for accounting
Data from sites are fed to an NDGF Accounting DB and
from here automatically fed further to APEL

NDGF was the biggest European ATLAS Tier-1 in 2007 and 2008, and the most efficient worldwide
Job submission

- Only thing left... Job Submission
- Modification of the gLite-WMS to accommodate for ARC-CEs
- However:
- Lets first understand the difference between the ARC and e.g. the CREAM CE...
Job submission: ARC-CE

- Philosophy for ARC: An automated Super Computer frontend
  - [Super Computer]: ...
  - [frontend]: ...

ISGC09. Taipei. April. 2009
Philosophy for ARC: An automated Super Computer frontend

- [Super Computer]:
  - Shared among multiple users and user groups
  - OS dictated by “others”, but not by the grid m/w
  - Jobs should behave “nice” - no heavy network usage by WNs. Need for throttling of network usage.
  - Optimized cluster file system for internal data management, or at least a shared file system
  - Not build for web hosting
  - Multi Core jobs preferred

- [frontend]
  - ...

Job submission: ARC-CE

- Philosophy for ARC: An automated Super Computer frontend
  - [Super Computer]:
    - ...
  - [frontend]
    - Compilation and optimization of the code for the SC
    - Installation of optimized code
    - Handling of job data – stage in/out from/to remote sources
    - Throttling of cluster usage
    - ... and the grid bits:
      - Unification of the interface
      - Data Caching and Transfer retries
Philosophy for ARC: An automated Super Computer frontend

- [Super Computer]:
  - ...
- [frontend]
  - ...

Job submission: ARC-CE

ARC

Other CEs
Job submission: Other-CEs

- Philosophy for Other-CEs: gateway to high throughput resources
  - [HTC]: ...
  - [gateway]: ...

Job submission: Other-CEs

- Philosophy for Other-CEs: gateway to high throughput resources
  - [throughput]:
    - Many individual machines
    - No cluster file system
    - Jobs handle their own workload
  - [gateway]
    - Unification of interface
- Only thing left... Job Submission
- Modification of the gLite-WMS to accommodate for ARC-CEs
- Interoperability based on WMS works and has been used in production in about a year
- It is the “easy fix” for smaller VO and to get things running
- Still a lot to gain by porting the applications / production tools
  - Resource Utilization is 10-15% higher with the pure ARC-CE setup
Future goals

- The European Grid Initiative is coming...
- We will see a Europe with:
  - Several different m/w
  - Different operation models
  - Different resources and resource models
- A need to bridge it all
- A need for a definition of a minimal set of infrastructure services – i.e. common:
  - Registration, Indexing, Monitoring, Accounting
- Unification of interfaces (standards?)
- Data have to be accessible throughout all domains!
Future goals

- The Universal Middleware Distribution (UMD)
  - ARC, gLite, and UNICORE
- Some components will be merged
- Process steered and monitored by:
  - DIESA, EGEE, and NDGF
  - And eventually EGI
- Still a lot of other services will exist
  - Interoperability of highly important
Acknowledgements

Thanks to:

Tord Ekelöf, Mattias Ellert, Laurence Field, Claudio Grandi, Daniel Johannson, Oliver Keeble, Josva Kleist, Balazs Konya, Erwin Laure, Francesco Preltz, Di Qing, Markus Schultz, Anders Selander, Oxana Smirnova, David Smith, Christian Søttrup, Mattias Wadenstein, Rod Walker, Anders Wäänänen and many others...
Thanks!

Questions