Integration of Cloud and Grid Middleware at DGRZR

International Symposium on Grid Computing 2010

Stefan Freitag

Robotics Research Institute
Dortmund University of Technology

March 12, 2010
Overview

1. D-Grid Resource Center Ruhr

2. Clouds in the German Grid Initiative D-Grid
Introduction

D-Grid Resource Center Ruhr
Introduction

D-Grid Resource Center Ruhr (DGRZR)
- 256 Blades, Intel Xeon Dual CPU QuadCore, 16 GByte RAM
- Cluster runs SLES 10 SP3 with Xen 3.2 Kernel
- 100 TByte storage
- Since April 2008 in production as part of D-Grid infrastructure
- End of 2008: 25 TByte SFS (Lustre) storage extension
Site setup follows recommendations of D-Grid reference installation¹ (not 100%)

- Three compute middlewares
  - gLite 3.1 (lcg-CE) and 3.2 (CREAM-CE, BDII)
  - UNICORE 5 and 6
  - Globus Toolkit 4.0.8

- Two storage middlewares
  - dCache 1.9.x
  - OGSA-DAI 2.2

- Additional
  - LDAP for user management
  - DNS, DHCP
  - MySQL DB for OGSA-DAI

All services run in Xen virtual machines

¹http://dgiref.d-grid.de/wiki/Introduction
## D-Grid Services @ DGRZER

<table>
<thead>
<tr>
<th>Grid Middle-ware</th>
<th>Grid Compute Frontends</th>
<th>Grid Storage Frontends</th>
</tr>
</thead>
<tbody>
<tr>
<td>UNICORE VSite</td>
<td>gLite CE</td>
<td>WS GRAM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>OGSA DAI</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dCache SE</td>
</tr>
<tr>
<td></td>
<td></td>
<td>dCache Pool</td>
</tr>
<tr>
<td></td>
<td>Globus Toolkit</td>
<td></td>
</tr>
<tr>
<td>LRMS</td>
<td>Torque &amp; MAUI</td>
<td>Databases</td>
</tr>
<tr>
<td></td>
<td></td>
<td>File Systems</td>
</tr>
<tr>
<td>Fabric</td>
<td>Compute Cluster Worker Nodes</td>
<td>Online Storage</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Virtualization Layer</td>
<td></td>
</tr>
</tbody>
</table>

**Figure:** Pre-Cloud software stack
Extending DGRZR by Cloud Middleware

**Figure:** Current software stack including OpenNebula
OpenNebula at DGRZR

- Started with OpenNebula (ONE) 1.2
- Currently running: slightly adapted ONE 1.4 SVN snapshot (January 2010)
- Xen infrastructure and SSH transfer enabled
- Images/ templates for SL 4.8 and 5.4 (64bit) Grid workernodes
- In progress e.g. for gLite: lcg-CE, CREAM-CE, siteBDII
- Users interface with ONE via CLI
- one.grid.tu-dortmund.de supports OCCI via HTTP(S) (currently not in production use)
All blade servers are registered with OpenNebula

```
one:~ # onehost list
ID  NAME       RVM TCPU FCPU ACPU ... STAT
0   udo-bl1101 0 800 700 700   on
1   udo-bl1102 2 800 0   0   on
...  
246 udo-bl6307 0 800 98 98   on
247 udo-bl6308 0 800 99 99   on
```
OpenNebula at DGRZR

Created network definitions with MAC/IP mapping.

NAME = ”DGRZR Workernodes”
TYPE = FIXED
BRIDGE = eth0
LEASES=[ IP = 129.217.241.212, MAC=00:16:3e:6f:d2:09 ]
LEASES=[ IP = 129.217.241.213, MAC=00:16:3e:5b:09:c9 ]
LEASES=[ IP = 129.217.241.214, MAC=00:16:3e:14:ff:b1 ]
LEASES=[ IP = 129.217.241.215, MAC=00:16:3e:27:c6:04 ]
[... ]
OpenNebula at DGRZ

Created user accounts for D-Grid users

<table>
<thead>
<tr>
<th>ID</th>
<th>USER</th>
<th>PASSWORD</th>
<th>ENABLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>2</td>
<td>ad0001</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>ad0002</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>ad0003</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>ad0004</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ad0005</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>ad0006</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>ad0007</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>ad0008</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>ad0009</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- 100 accounts per Virtual Organization (D-Grid recommendation: 200)
- Supported VOs: at present 10, later 24
OpenNebula at DGRZR

Workernode Template (not using context)

VCPU = 1  # usually 8
MEMORY = 512  # usually ~ 13 GByte
OS = [   bootloader = "/root/bin/domUloader.py" ]
RAW = [    type = "xen",
    data = "bootargs="--verbose --entry=xvda1""]
DISK = [
    source = "<some_path>/wn_sl54_x86_64.img",
    target = "xvda", readonly = "no" ]
DISK = [
    type = swap, size = 1024, target = "xvdb", 
    readonly = "no" ]
DISK = [
    type = "block", clone = "yes", target = "xvdc",
    source = "/dev/cciss/c0d0p4", readonly = "no" ]

NIC = [NETWORK="dgrzr", IP=129.217.241.215 ]
Scenario 1: Private Cloud (already tested in small scale)

Use ONE to deploy workernodes on-demand\textsuperscript{2}

- Assumption: VO software requirements are satisfied by VO specific workernodes VMs
- Interaction with LRMS of Grid middleware required
- 1:1 mapping of workernode type to LRMS queue
- A daemon checks the status of each queue
- Empty queue: reduce number of workernode VM assigned to this queue
- Re-assign freed resources to another (overcrowded) queue

\textsuperscript{2}B. Konrad: Dynamic management of VMs on HPC resources of TU Dortmund (diploma thesis, 2009)
Major difference to scenario 1

Allow users to deploy services/VM via a Cloud interface

Split physical resources into a Cloud and a Grid partition.

- Allow dynamic/workload-dependent changes in partition size
  - Cloud size=0: Grid resource
  - Grid size=0: Cloud resource
  - All other cases: hybrid resource

- Which VMs to suspend? → prioritization of VMs
  - Simple Grid batch jobs, MPI batch jobs, services
  - Normal, gold and platinum (paying?) customers
  - Talk of Johannes Watzl this afternoon
(One) Goal of D-Grid

Create sustainable & longterm Grid infrastructure in Germany

→ D-Grid is focused on Grid usage. What about Clouds?
  - Cloud interfaces offer a new and easier way to remote resources
  - Integration of Cloud middleware into D-Grid Software stack seems pretty obvious (→ increase sustainability)
  - Issues to be resolved for successful integration: user management, authorization, accounting/billing, monitoring, and information system

^3that's my personal view ;-)

D-Grid

- Central virtual organization membership service VOM(R)S
- Resources connect to VOM(R)S to query user information mapping to local user accounts
- User can have attributes & roles, belong to groups

OpenNebula

- Users stored in a local SQLite3 database

Open issues

- Connection between central VOM(R)S and ONE needed
- Support for groups, roles (First: Evaluation in D-Grid)
- Scalability
Authorization

D-Grid

- Based on X.509 certificates

OpenNebula

- At present: username/ password mechanism
- With ONE 1.6:\n  - Users are identified by abstract key/secret tokens. An underlying driver will then interface with the auth back-end (e.g. LDAP / X509 based / PAM / Policikit...) to authenticate the user.
  - General Authorization policies can be implemented, for example quotas or allow a user to submit VMs in a given time frame, user groups....

\textsuperscript{4}http://dev.opennebula.org/issues/203
Accounting

D-Grid
- Jobs pass through a Grid frontend and reach the Grid LRMS
- DGAS \(^5\) and OGF-UR format are used
- Cloud "jobs" do not reach LRMS, but start fabric level

OpenNebula
- Accounting information can be gathered by joining tables (history table, vm_attributes) in the SQLite3 database

Open issues
- Evaluation if equivalent metrics can be collected
- Design & implement prototype tool

\(^5\)Distributed Grid Accounting System
Information system

- Each Grid middleware runs an information system
- D-MON \(^6\) collects information from all these systems, aggregates and publishes it
- D-MON uses an adapters/ plug-ins
- New adapter for Cloud Middlewares must be developed

What information is needed?

Virtualization software (e.g. Xen, VMware), the available virtual appliances/ templates, limits concerning the maximum amount of cores and memory per virtual appliance

\(^6\)http://www.d-grid.de/index.php?id=401
Future plans

For D-Grid

- Close presented open issues and establish Cloud middleware as new pillar in the D-Grid software stack
- Project starts in summer 2010

In Dortmund

- Integrate more resources into the Cloud
  - Physics department (1000 Cores), in operation Mar 2010
  - Computer Sciences department (1000 Cores), around 2011

Extend the Cloud to the allieded universities Bochum, Essen/Duisburg

Thanks for your attention and for the great time being here!