NBIC Galaxy@HPC Cloud

Mattias de Hollander
David van Enckevort
Leon Mei
Marc van Driel
Rob Hooft
SURFsara HPC Cloud

- National computing and networking center
- 19 nodes, 32 cores and 256 GB RAM each
  - Intel 2.13 GHz (Xeon-E7 "Westmere-EX")
- 400 TB storage in total
- Maximum 40 Gb access to storage (per node)

- Lightpath connection (1Gb->10Gb) between all Dutch universities & medical centers, BGI?

https://www.surfsara.nl/systems/hpc-cloud
Migration to Cloud

- Galaxy2.nbiceng.net
  - 4 CPU, 24G RAM, 1.5T HD, 100Mbit Internet
- Migration project July-September 2012
- Supported by BiGGrid, SURFsara, NBIC, NIOO
- Used as the base for other project specific Galaxy servers in the HPC cloud
  - CTMM-TraIT, CGD
Architecture

- **Persistent Master Node**: 4core, 32G
- **Dynamic Worker Node**: Cloudman auto-scaling
- **NFS mounts**: Tools, User data, Genomes, 50TB
- **Backup**

Internet connection to the cloud.
Cluster Status

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Job ID</th>
<th>Status</th>
<th>Progress</th>
<th>Runtime (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>7121414</td>
<td>Running</td>
<td>23%</td>
<td>00:14</td>
</tr>
</tbody>
</table>

Job status (1 running, 0 queued, 0 finished, 1 total) (data is refreshed every 90 seconds)
Some Hurdles

• Installation using Cloudman
  • Fabric scripts use hardcoded versions, some of them are outdated

• Too many layers so sometimes hard to find the error
  • Tool, Galaxy, VM or Cloud

• I/O Performance issue

• MySQL to PostgreSQL: py-mysql2pgsql