A startup in the Grid: Why, What, where

Florido Paganelli
Lund University, Sweden

Open Source Days 2012
Copenhagen, Denmark
Why: Goal 1
Why: Goal 2
What is the grid?

- 172,494 CPUs
- 445,432,651 GB
- 1,359,902 Jobs

Federates Computing Resources across the world
What is the grid?
What, Where: Grid Middleware

- Open Science Grid (USA)
- ... more
Technologies

- **Languages**: C++, Java, Phyton, Perl, shell scripting
- **Information/control**: LDAP, Web Services (i.e. RESTFul)
- **Security**: Openssl, Globus toolkit GSI, SAML, XACML
- **Cluster systems**: PBS/Torque, LL, SGE, SLURM, LSF, Condor, GridSite, DG
It's Free SW and Open Source!

- Official repos of:
  - Red Hat
  - EPEL
  - Fedora
  - Debian

- Self-maintained:
  - Scientific Linux
  - Ubuntu
What is it good for?

- Focused on batch processing
- Modelling, simulations (e.g. Monte Carlo)
- Other stuff:
  - Code compiling/building
  - Movie encoding
  - Data Mining
  - … ?
How to access it?

- PKI/X509 of Trust among universities
- Problem: startups don't belong to academic world
Why: How academic world can benefit from a startup

- Resources are running low in the academic world. Need to find other ways of funding
- Researchers demand storage for their experiments: Cloud computing too expensive and unreliable for huge data needs
- Moderate need for computational power
- Drive to develop new features requested by markets
Why: How a startup can benefit being in the Grid

- Computing power spread across the world, high availability offers services for Startups involved in modelling, statistics, data analysis
- In principle, access to huge resources for an affordable price: Cloud computing is more expensive on intensive computing
- Use that computing power experiments don't use
What has been tried

- A small company in Lund requested allocatable resources for their customers
- Simulation and modelling – engineering field – the company was selling modelling software and eventually computing power
What didn't work

- Universities don't own resources, they administer them
- National Grid Infrastructures only have a diplomatic role
- Research centers and grid initiatives are just part of academia

=> Nobody really has right to sell/provide these services!
What couldn't work

• Hard to see any SLA between universities and companies: management is done on best-effort
• Downtimes are predictable, but not always advertised in real time
Some ideas

- A new mechanism for researchers and companies to work together in a R&D framework
- Maintenance shared between companies and universities to achieve a real SLA
- Virtuous circle where private sector helps public sector to improve
- Resources shared in a fair way – a startup might repay and entire dedicated cluster/storage after a reasonable time
How to achieve?

- Joint European project with a group of startups and involved academic resource owners
- National projects with involvement of local incubators as mediators
- Other ideas?
## Grid VS Cloud

<table>
<thead>
<tr>
<th></th>
<th>Grid</th>
<th>Cloud</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nature of services</strong></td>
<td>Persistent resources</td>
<td>On-Demand services</td>
</tr>
<tr>
<td><strong>Interaction</strong></td>
<td>Asynchronous batch processing</td>
<td>Real Time interaction</td>
</tr>
<tr>
<td><strong>Resource availability</strong></td>
<td>High Availability</td>
<td>Depends on SLA</td>
</tr>
<tr>
<td><strong>Ownership</strong></td>
<td>Managed by academic world</td>
<td>Managed by private companies</td>
</tr>
<tr>
<td><strong>Security</strong></td>
<td>PKI/X509, web of trust, SAML</td>
<td>... and other technologies</td>
</tr>
<tr>
<td><strong>Business Model</strong></td>
<td>None</td>
<td>SLA-based, accounting for resources used.</td>
</tr>
</tbody>
</table>
# Grid VS Cloud

## Grid
- Persistent resources
- Async batch processing
- High availability
- Managed by academic world
- PKI/X509 and web of trust, SAML

## Cloud
- On-demand services
- Real time interaction
- SLA availability
- Managed by private companies
- Other technologies
What is the grid?

Internet
What is the grid?

Internet

client tools

INDEX

INDEX

INDEX

proxy

Computing Element

Computing Element

Computing Element

proxy

Storage Element

Storage Element

Storage Element