BalticCloud: Leveraging Grid Infrastructure

Ilja Livenson, NICPB, Estonia
Technical coordinator of BalticCloud
ilja@kbfi.ee
Outline

- **Background**
  - BalticGrid & BalticCloud

- **Infrastructure optimization**
  - Resource usage
  - Security

- **“New” approach to HPC**
  - OLAP/OLTP applications
  - Stronger collaboration with SMEs

- **Educational activities**
The Baltic Grid Second Phase

- Started 1st of May, 2008, duration - 2 years
- 13 partners from Baltic States, Belarus, Poland, Sweden, Switzerland
- Continues the BalticGrid project (2005-2008)
- Financing – 3 MEUR
- Collaborating project for EGEE
- Supporting NGIs and their collaboration
General Structure

BalticGrid II

Planned activities

SA1-3, NA1-4, JRA

BGi - BalticGrid Innovation Lab

BalticCloud
Courses
SME connectivity

ISGC 2009, 23.04.2009, Taipei
Motivation

- Low resource usage
- Too steep learning curve
  - Low acceptance within industry
- Complicated to adapt new applications
  - Typically, need a “project” to do that
- Low security level of grid middleware
- Interoperability issues
  - There are standardization activities, but...
LAST YEAR WE RECOGNIZED THAT OUR PROCESSES WERE FAR TOO COMPLEX

SO WE PUT THEM INTO THE CLOUD

LET THE CLOUDS MAKE YOUR LIFE EASIER

from http://geekandpoke.typepad.com
Cloud model and BC

BalticCloud

- Infrastructure
- Platform
- Software

Users

… as a Service
Resource usage efficiency

- CPU usage
- RAM usage
- HDD usage

Even if users specify requirements of the job, the gap is still large!

- Priority inversion
  - low efficiency jobs occupy the whole cluster
Datacenter as IaaS

- Virtualize computational resources
  - There are drawbacks!
    - Virtualization overhead (CPU, IO), “noisy neighbours”, state preservation, etc

- On-demand resources
  - Creating virtual machine with a specified set of resources is possible
  - Programs consume as much resources as they actually need, overprovisioning is still there, but to a lesser extent

- “Cloud of clouds” approach
  - Open question
Expertise distribution

**OS Management**
- Data centers provide standard OS images
- Users can upload/modify their own
  - Depending on security policy

**Program debugging**
- Users get more control over execution
- Users can also test out application beforehand (try doing that with gLite …)
- Administrators can concentrate on their direct tasks
- Profit!
Virtualization toolkit

There are several projects out there offering cloud solutions (IaaS)

- OpenSource: Eucalyptus, Nimbus, AbiCloud, …
- Commercial: VMWare, Citrix, …

Our choice: Eucalyptus

- Great team!
- Integration with RightScale
- Integration with Ubuntu
  - Ubuntu 9.04 is coming out today!
  - 1.5 version
Storage virtualization

- **Eucalyptus offers basic implementation of Amazon S3 interfaces - Walrus**
  - Very simple bucket based filesystem with ACLs
- **Doesn’t include scalability/reliability mechanisms**
- **Plan: deploy Walrus on top of region level dCache**
  - Something like what NDGF does
  - Single access point
- **Performance overhead will (probably) be big!**
  - OK for educational and cloud research activities
Eucalyptus uses X.509 security infrastructure

- “Light” version
- No VOMS extensions, no OCSP, no CRLs, no SAML assertions, etc

Can we use BalticGrid/EGEE infrastructure?

- Well, yes, but…
- One of priorities for us
Security (process)

- **Job/Application isolation level**
  - Low (same users within VO)
  - what if if size(VO) > 1000?
  - Semi-low (different user groups)
  - Medium (e.g. jail/chroot, shared kernel)
  - High (separate VM for each job)

- **Lower level = lower security level, less customization options, (more efficient)**

- **Network security**
  - Not every switch supports VLANs, or does that well
  - Need that if we want to give local root to the users
    - Or just encrypt everything critical
Complicated problem

OGF has published a number of standards
- E.g. OGSA-BES for execution or GLUE for information system

Middleware of interest for our region:
- gLite, ARC and UNICORE
- Implementation of standards support is slow

Mid-term solution
- Provide core components for every system as virtualized images
- On-demand WN creation
Missing functionality

- **Missing**
  - Monitoring
  - Accounting
  - Reasonable OS image management
  - (Billing)

- **Can be solved by integration with RightScale**
  - Vendor lock-in
  - Not open-source

- **We are thinking of reinventing some wheels**
Cloud model and BC

BalticCloud

Software

Platform

Infrastructure

Users

... as a Service

ISGC 2009, 23.04.2009, Taipei
“New” application classes

- Grid was built for batch processing
  - There are some workarounds
    - Pull-mode execution
    - “VO-box” component of gLite
  - But no solution!

- OLAP/OLTP
  - Databases
  - Application servers, Web servers
  - BI
  - Load balancing

- Hadoop framework

- AppScale and others PaaS solutions
SME = Small and Medium Enterprises

So far so bad

- It has been very painful with gLite based solution
- It hasn’t been to easy with “pbs cluster access”

Reasons

- Shared WN (typical case for multicore) is not acceptable
- Interfaces are too complicated and components are too unstable
- Not enough control: “What do you mean by SL3/4? What do you mean by CLI? What do you mean by queuing? We already have J2EE application, just give us the cluster to deploy it on!”
- Accounting
- Interactive applications
Collaboration with SMEs (2)

- We hope to make it better with cloud approach
  - Industry driven, not HEP driven
  - Reasonable economical models
  - Existing success stories

- Courses for SMEs on best practices using clouds
  - In all countries involved
  - Building a network of adopters

- Free access to resources for academic startups/spinoffs

- Hybrid clouds
  - Selling unused cycles to SMEs?
Educational activities

- **Different types of courses**
  - 2-3 days crash courses
  - Semester long courses
  - Being part of e.g. “Software Project” course
  - Example applications
  - Labs are done on the BC infrastructure

- **The course is being developed at the moment**
  - First run: 18th of May, University of Tartu, Estonia (crash course)

- **We also plan to do both technical and more business oriented courses**
  - Economics of Cloud computing
Future work

BGi

BalticCloud

R&D

SMEs

Eucalyptus, AppScale

Custom tool development

Policies and SLAs

Cloud of clouds

Courses

Research

Networking and consultancy

ISGC 2009, 23.04.2009, Taipei
Summary

**Cloud computing is not a silver bullet**
- Many administration problems remain
- Some more appear
- “Cloudifying” is not trivial in case of legacy applications
- Too much hype

**But...**
- It does make many aspects better!
  - Security, resource usage, interactive applications
- Economically motivated
  - Startups and prototyping, one-off tasks, research at scale
  - Scaling solutions
Many thanks to people involved

- Aake Edlund - BalticGrid Project Director (Sweden)
- Mario Kadastik (Estonia)
- Janis Kulins (Latvia)
- Dalius Mazeika (Lithuania)
- Eduardas Kutka (Lithuania)
- Yuri Ziamtsou (Belarus)

http://cloud.balticgrid.eu
Happy end

Questions?

Many thanks to the ISGC organizers!