SURFsara HPC Cloud Workshop

www.cloud.sara.nl → Tutorial 2014-06-11

UvA HPC and Big Data Course June 2014

Anatoli Danezi, Markus van Dijk

cloud-support@surfsara.nl

SURF SARA

Agenda

- Introduction and Overview (current presentation)

- Hands on

- Lunch

- Application design

- Hands on cont.

- Assignment (approx. 4 hours)

Cloud? What as a Service?

Wikipedia:

" Cloud Computing is a jargon term without a commonly accepted non-ambiguous scientific or technical definition.

" In science Cloud computing is a synonym for distributed computing over a network.

… as a Service:

- SaaS: Software – MS Office 360, gmail
- PaaS: Platform – Google App Engine
- IaaS: Infrastructure – Amazon EC2, SURFsara HPC Cloud
- …aaS

Known cloud-like services:

- Hosting
- Grid computing: massive parallel batch processing

Why Cloud?

Benefits

- No hardware to buy and maintain
- No software to buy and maintain (SaaS, PaaS)
- No maintenance downtime – live migration of virtual machines
- Dynamic scalability – add when needed

Drawbacks

- Control over data – privacy, business secrets, legal obligations (patient data, Patriot Act)
- Control over computing – availability, processing power (SLA, overcommitting)
- Different environment – virtualization layer, VM management

Types

- Private / Community / Public
- Dedicated / Overcommitted (CPU, network, disk space)

SURF SARA

SURFsara Computing

Pre-configured and maintained environments:

- Cartesius – National Supercomputer: 13984 cores (270TFlops), 41TB RAM, 2.4PB disk
- Lisa – National Compute Cluster (VU, UvA, SURF): 6528 cores (46Tflops), 17TB RAM
- Grid – International parallel batch processing: 11 sites, 5000 cores, 4PB disk 6PB tape
- Hadoop – BigData-parallel processing framework: 700 cores, 1.2PB disk

Self-service:

- HPC Cloud – Cloud computing: 960 HPC cores, 80 Light cores, 8GB RAM/core
  - Extra: high memory node with 40 cores, 2TB RAM, 6.4TB disk
  - 500TB shared storage
  - Nodes connected by fast network

Other

- Beehub – Data storage: WebDAV access
- Visualization – render cluster and "Collaboratorium"
- Network

SURF SARA

Who uses the SURFsara HPC Cloud

Past:
- 130 projects completed since January 2011

Current (2014-05-21, a slow day):
- 120 active projects
- 250 login accounts
- 160 running VMs using 785 cores
- Largest VM: 32 HPC cores, 245GB RAM
- Smallest VM: ¼ core, 2GB RAM
- 100TB used for disk images
- 70TB used in VirDir (project NASes)

Example techniques:
- Galaxy, RStudio, Matlab
- CFD: MPI on virtual cluster, multicore VMs
- De novo genome assembly: single machine, multicore VMs, high memory

Research fields:

- Biology
- Genetics
- Informatics
- Chemistry
- Ecology
- Linguistics
- Robotics
- Business
- Social sciences
- Engineering
- Humanities

SURF SARA

Why SURFsara HPC Cloud

SURFsara HPC Cloud is IAAS: Infrastructure As A Service, so
you assemble your virtual machine (VM) from the ground up.

General benefits:

- Data and computing in Amsterdam, backups in Almere
- No ties to US and its Homeland Security, Patriot Act
- Others cannot access data in your VM (including SARA personnel)
- Unrestricted Internet access (but fair use), including up/download of data

Technical benefits:

- No overcommitting, you alone use 100% of your core(s)
- Tailor VM to your needs: cores, RAM, disks
- Root access to your VM
- Free choice of OS, packages, versions
- Fast private network for all VMs in your project

SURF SARA

Why not SURFsara HPC Cloud

SURFsara HPC Cloud is IAAS: Infrastructure As A Service, so you assemble your virtual machine (VM) from the ground up.

Drawbacks:

- No SLA (yet), service during office hours
- You maintain everything in your VM
- You are responsible for all of your VM's behavior
- You must protect yourself against threats from the Internet (DDOS, virus)
- Pay for VM uptime, not just compute time (like gas, light)
- No automatic backups
- Slow disk I/O (designed for computing)
- Your laptop is faster than a 1 core VM
- Interface to construct/start/stop VMs is not user friendly

SURF SARA

Below decks
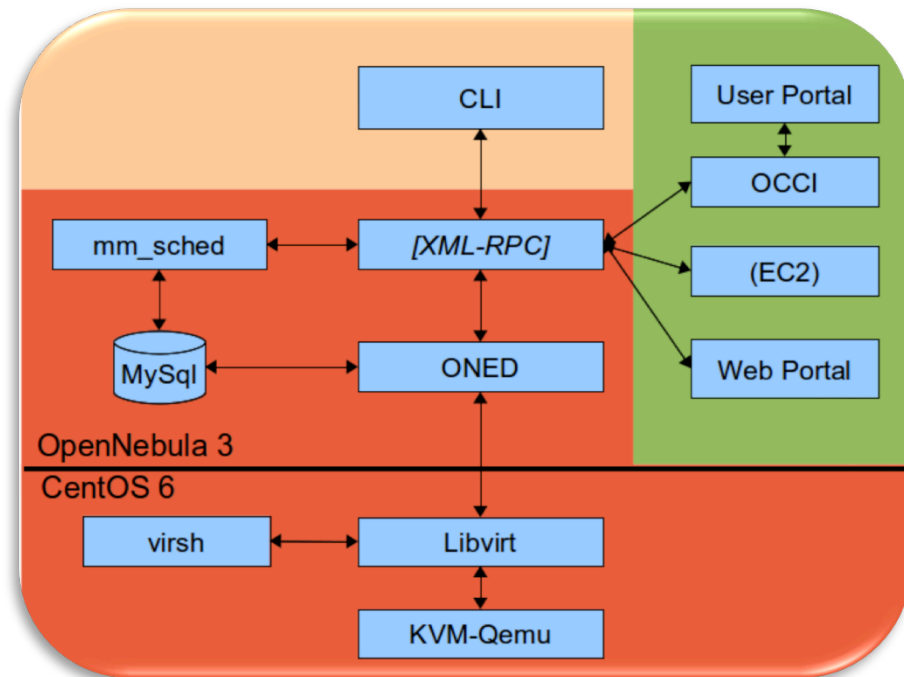
VM control: OpenNebula

- Open source
- Adaptable to our needs
- Currently best practice for our situation

Virtualization: KVM, libvirt, CentOS

- Open source
- Low overhead
- Proven track record

Support portal: Redmine

- Open source
- General Cloud wiki
- Per project: wiki and issue tracking
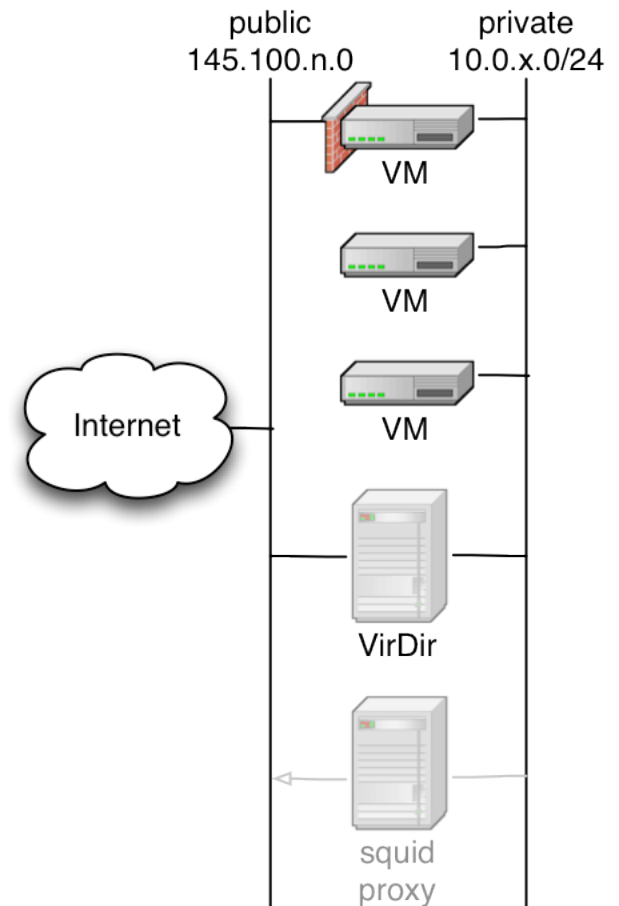
Cloud project networks

- Direct Internet access
- One private virtual network per project
- Fast interconnect between VMs
- Squid proxy for non-Internet VMs
- IP and MAC addresses change every launch
- Dynamic DNS

VirDir: shared storage per project
- **/Scratch**, **/Backup** daily to tape
- Mount VirDir in VM: NFS, but with limitations
- Access VirDir via SFTP with private key
- Used for up/download of disk images

Extra:
- MySQL server
- NoSQLserver

public
145.100.n.0

private
10.0.x.0/24

VM

VM

VM

Internet

VirDir

squid
proxy

SURF SARA

Build a Virtual Machine

A "template" is a VM recipe:

- Node type: Large / Medium / Small
- # virtual cores, # real CPUs, RAM
- Private project network
- Internet: needs packet filter
- Your disk images (IDE, virtio), boot from hd / cdrom

Create template: "Create VM" wizard or Templates→New

NODE		VCPU / CPU	RAM / CPU	RAM / VCPU
L	HPC	1/1	8GB	8GB
M	light	2/1	8GB	4GB
S	light	4/1	8GB	2GB

Set-up assistance

Creating a VM from scratch can be a lot of work.

We provide a set-up wizard: you choose
- CentOS / Ubuntu, Desktop / Server
- # cores and RAM
- Disk size
- Networks: Internet and/or local, webserver reachable from Internet?
- A new VM is created and launched
- Connect to console
- Set up root password and first user
- Connect with SSH

From then on, you must take care of updates etc., the wizard is just a set-up help.

Alternatives

Well-known international Clouds:

Note: "US-EU Safe Harbor" worthless

- Amazon Elastic Compute Cloud (EC2)
- Google Compute Engine
- Rackspace

Other:

- Hosting
- Grid Computing
- Hadoop

SURFsara HPC Cloud Workshop

HANDS ON

Support portal: https://www.cloud.sara.nl

Search: *Tutorial 2014-06-11*

Self-service portal: https://ui.cloud.sara.nl

Make sure you use Google Chrome browser.

UvA HPC and Big Data Course June 2014

Anatoli Danezi, Markus van Dijk

cloud-support@surfsara.nl

SURF SARA