SURFsara HPC Cloud Workshop

doc.hpccloud.surfsara.nl → UvA workshop 2016-01-25

UvA HPC Course Jan 2016
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)

SURFsara Computing

Pre-configured and maintained environments:

- Cartesius – National Supercomputer: 40,960 cores (1.559 PFlops), 117 TB RAM, 7.7 PB disk
- Lisa – National Compute Cluster (VU, UvA, SURF): 8960 cores (158 Tflops), 10 TB RAM
- Grid – International parallel batch processing: 11 sites, 15,000 cores, 7.1 PB disk 7.7 PB tape
- Hadoop – BigData-parallel processing framework: 700 cores, 1.2 PB disk

Self-service:

- HPC Cloud – Cloud computing: 1920 cores
  - 30 nodes with each 64 cores, 248 GB RAM
  - 1 High memory node, 40 cores, 2 TB RAM, 6.4 TB disk
  - 900 TB image storage
  - Nodes connected by fast network

Other:

- Beehub – Data storage, WebDAV access
- Visualization – render cluster and “Collaboratorium”
- Network

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 inside 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

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
- Your laptop is faster than a 1 core VM
- Interface to construct/start/stop VMs is not user friendly

Who uses the SURFsara HPC Cloud

Past:
- 226 projects completed since January 2011

January 2016:
- 122 active projects
- 400 login accounts
- 200 running VMs
- 160 TB used for disk images
- Largest possible VM: 64 cores, 248 GB RAM
- Introducing GPUs

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

SURFsara HPC Cloud workshop Jan 2015

Below decks

VM control: OpenNebula

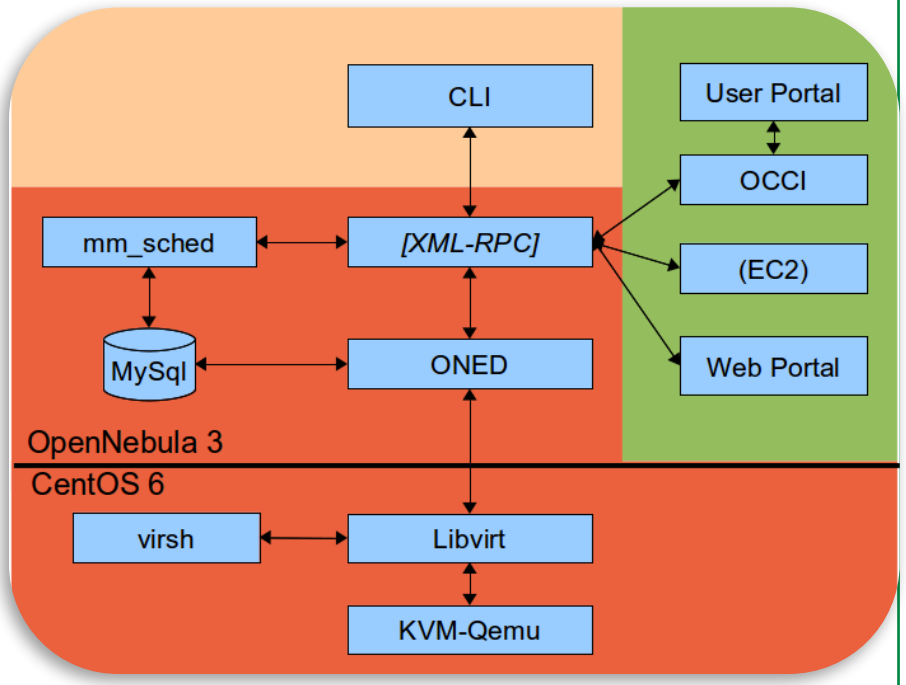
- Open source
- Adaptable to our needs
- Currently best practice for our situation

Virtualization: KVM, libvirt, CentOS

- Open source
- Low overhead
- Proven track record

Support:

- General Cloud wiki
- Helpdesk: email, phone, house calls



Cloud project networks

Network:

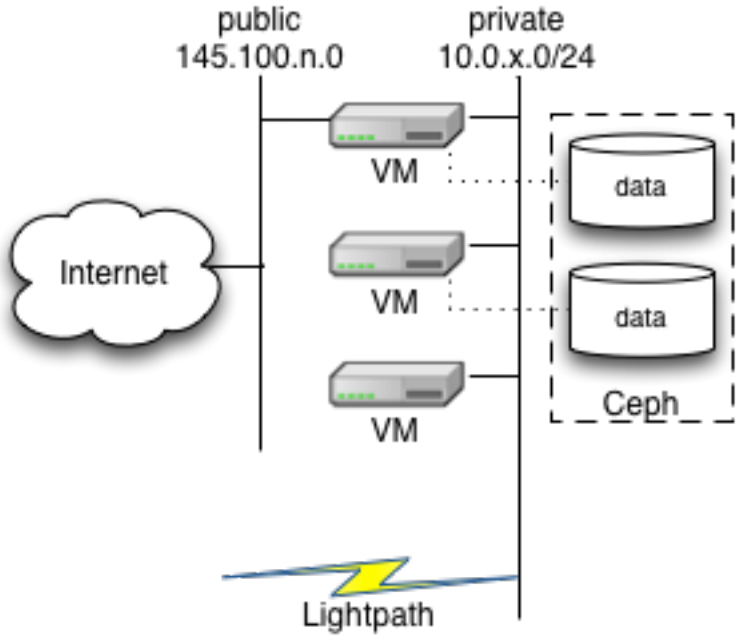
- Direct Internet access, no external firewall
- One private virtual network per project
- Fast interconnect between VMs
- IP and MAC addresses change every launch

Images:

- Host-local SSDs for OS images (limited size)
- On-line Ceph storage for data images (big data)
- No backups, but redundant storage

External data

- Beehub, NFS on VM, …
- Lightpath to your institute
- Community repository on Internet



Set-up assistance

Creating a VM from scratch can be a lot of work, we provide an “app market”.

You can “import” an “appliance”:

- CentOS 6.5 or 7
- Ubuntu 14.4 server or desktop.

Imported OS image:

- copied to node-local SSD
- persistent yes/no (run-time changes copied back)

Imported template (VM assembling instructions):

- # cores and RAM
- Internet and/or private lan
- additional data images on Ceph storage

Boot the VM and log in as root with SSH.

From then on, you must take care of firewalls, security, updates etc., importing an appliance 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
- Lisa, Cartesius

SURFsara HPC Cloud Workshop

HANDS ON

Support portal: https://doc.hpccloud.surfsara.nl/

UvA workshop 2016-01-25

Self-service portal: http://ui.hpccloud.surfsara.nl/

UvA HPC Course Jan 2015
Anatoli Danezi, Markus van Dijk
cloud-support@surfsara.nl