HPC Cloud at SURFsara

- Oort is born -

SURFcursus August 5th 2015

Cloud computing recap

- … as a Service:
  - SaaS: Software - gmail, facebook
  - PaaS: Platform - Google App Engine
  - IaaS:
    - Examples: Amazon EC2, SURFsara HPC Cloud
    - No hardware to buy and maintain
    - Assemble your virtual machine (VM) from the ground up
    - Tailor VMs to your needs: cores, RAM, disks, OS (Linux or Windows)
    - Dynamic scalability for peak capacities - add when needed
    - Root access to your VM (user is also system admin)

What do we offer?

Services for scientists
…scientists ⇏ systems gurus

… complex users' problems

- **Data**: big, dirty, non-structured…
- **Computation**: complex (e.g.: modeling, simulation)
- Libraries nightmare
- 3rd party, incompatibility, maintenance…

… cooperate

… share

… trial and error

… show

… test

… scratch

… flexibility

… privacy

We offer

- A place to build a running system
- A place to build your own cluster
- A place to collaborate

Why SURFsara HPC Cloud

- 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
- 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

- No SLA, 24/7 service during office hours only
- You maintain everything in your VM
- You are responsible for all of your VM's behavior
- You are responsible for all abuse by hackers
- You must arrange your backups
- Pay for VM uptime, not just compute time (like gas, light)

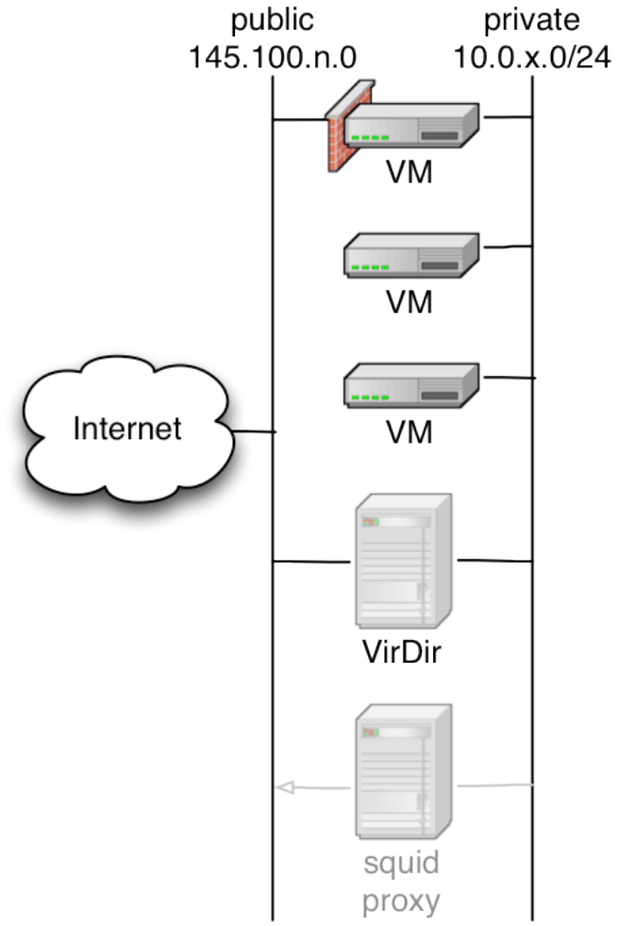
Who uses the SURFsara HPC Cloud

- Past:
  - 305 projects completed since January 2011
- On 2014-05-21, a slow day:
  - 120 active projects
  - 250 login accounts
  - 160 running VMs using 785 cores
    - Largest: 32 HPC cores, 245GB RAM
    - Smallest: 1⁄4 core, 2GB RAM

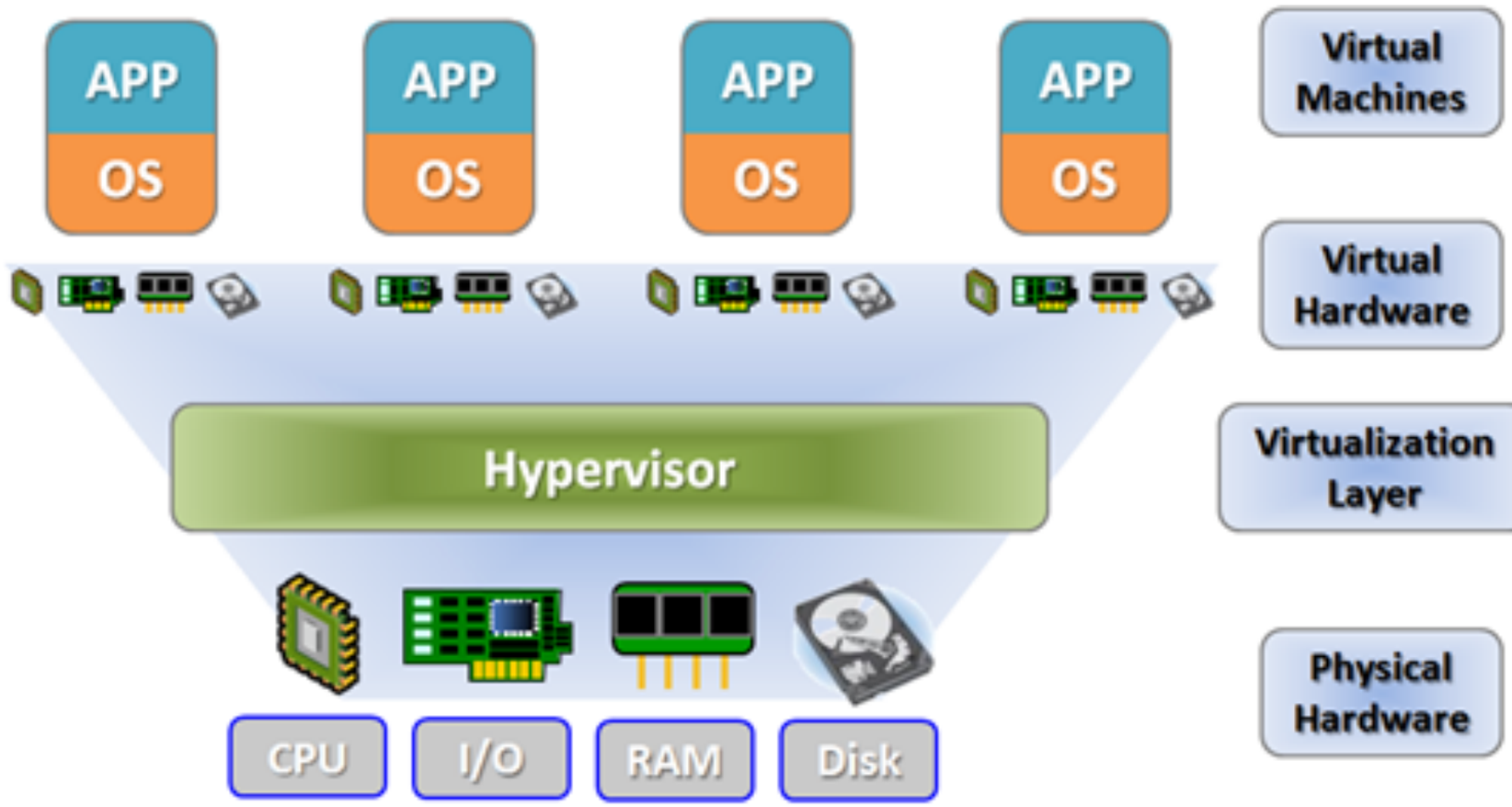
- Research fields:
  - Biology
  - Informatics
  - Chemistry
  - Ecology
  - Linguistics
  - Robotics
  - Business
  - Social sciences
  - Engineering

Oort
VM perspective

- Internet directly connected to VMs, no firewall
- Public IP addresses
- HOC Cloud project ≡ OpenNebula group
- One private trusted network for each project
- Your LightPath ends in project's private network
- Per project accounting
  - VM life time x # CPUs: beware of "oops, forgot to shut down"
  - Actual GBs of storage in use (Ceph/SSD/volatile storage)

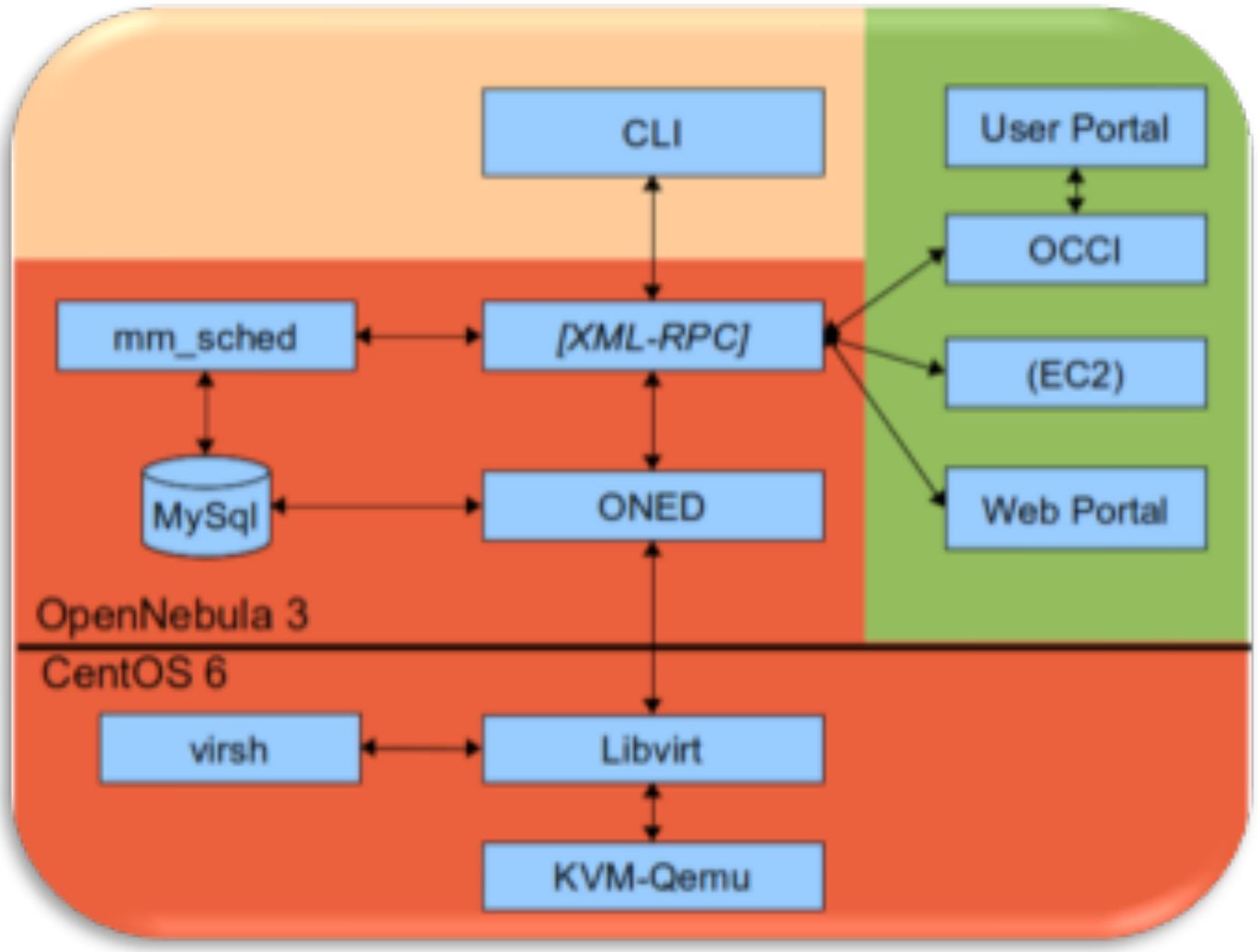


Oort
Server perspective



www.definethecloud.net

Oort
Server perspective



Oort dimensions

- Cluster with 2472 CPU cores:
  - 32 HPC compute nodes @ 64 cores, 256GB RAM, 3.2TB SSD
  - 12 GPU compute nodes @ 32 cores, 256GB RAM, 3.2TB SSD
- CEPH distributed object storage for VM disk images:
  - 800TB net storage (x3 redundancy, 2.4PB bruto)
  - 20Gbit replication network connecting the 50 storage nodes
  - 10Gbit network to compute nodes, with SSD write-caching
  - Efficient under high load, run-time expandable
- 1 single-user high-memory compute node
  - 40 cores, 2TB RAM, 3.2TB SSD
- Fast network between compute VMs

Before building a VM

Consider:

- Pipeline parallelization, independent parts, sequential parts
- Data access patterns and data locality
- Number of cores, RAM
- Wall-clock vs CPU time
- Disk storage space
- Operating system
- Backups
- Network interfaces (Internet, private LAN)
- Software type (licensing programs)

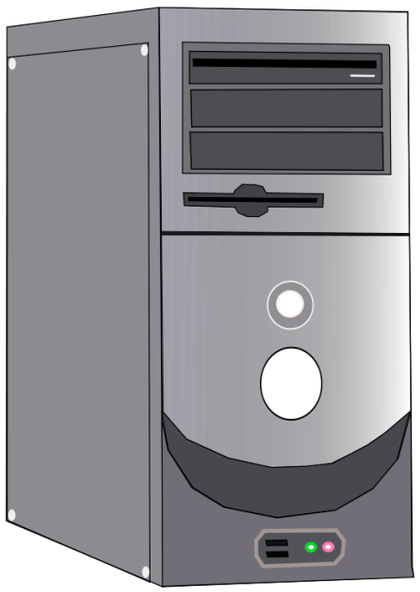
Components of a VM

- Disk image
  - Equivalent of a hard drive for OS or user data
  - a) import from AppMarket, b) install from CD
  - Persistent or use copy in VM
- Template
  - Describes your VM. It is just a **recipe**, not the machine itself.
  - # virtual CPUs, # real CPUs, RAM, network, boot and data images
- Virtual Machine
  - Constructed from template, disassembled on shutdown
  - Single or Multiple instances

How to set up your VM

- Creating a VM from scratch can be a lot of work.
- We provide AppMarket images
- Networks: Internet and/or local
- A new VM is created and launched
- Connect with SSH
- Tailor OS and software
- Add data disks

Scale up vs. Scale out



- Scaling up: use a single. larger VM
  - Easy to setup
  - Limited
- Scaling out: use multiple small VMs
  - Break in smaller work units
  - Add multiple worker nodes
  - Traditional clusters: Master / Workers

Scale up

Scale out

Choose the right scaling

- Scaling up: use a single VM with more RAM and/or more CPUs
  - Easy to setup but expensive
  - Limited to dimensions of compute node
  - Hard to find available free space
- Scaling out: using multiple ‘small’ VMs
  - Break in smaller work units
  - Add multiple worker nodes
  - Traditional clusters:
    - The **Master**: initializes the execution and splits the work according to the number of available workers
    - The **Workers**: receive work from the Master and perform calculations

General Challenges

- Scaling up limitation - Amdahl's law
- Scaling out - More machines, more problems
- Machines fail
- Networks fail
- Latencies
- Data locality / Persistence
- Backup images/datablocks
- VM installation and maintenance underestimated

Where can I find it?

- User Interface: https://ui.hpccloud.surfsara.nl
- Documentation wiki: https://doc.hpccloud.surfsara.nl/