Workshop in a compute cloud: not that obvious

Markus van Dijk
SURFsara HPC Cloud advisor

SURFsara

SARA was founded in 1971 by the two universities and the mathematical institute in Amsterdam for their computational needs.
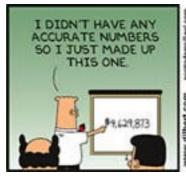
Today, SURFsara is the "Dutch national high-performance computing and e-Science support center" and hosts the national supercomputer.
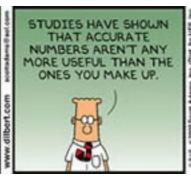
Academic research in the Netherlands can apply for free access to the resources.

Systems at SURFsara

	Cores	total RAM	RAM/core	storage	GPUs	
Cartesius	40960	117 TB	3 GB	7700 TB	132	1559 TFlops
Lisa	8960	10 TB	1 GB		-	158 TFlops
Grid	15000	***	8 GB	26000 TB	-	
Hadoop	1576	10 TB	6 GB	2300 TB	-	
HPC Cloud	1920	16 TB	8 GB	900 TB	20	
Visualization						
Archive				25000 TB		tape
SURFdrive				178 TB		max 100GB/user "dropbox" for academic use

SURFsara HPC Cloud

Created for High Performance Computing:
- Fast private network between VMs (MPI).
- Large, fast disk storage (900TB Ceph).
- No overcommit, wait if full.

Stability and MTBF less important:
- No compute redundancy.
- No backups, but redundant storage.

No SLA :-(

Demand for workshops

We had a growing demand for workshop support.

- By SURFsara: hands-on introduction to HPC Cloud as part of university curriculum.

- By institutes: as hands-on tooling training as part of their courses.

SURF SARA

Examples

- During 2015, 300 students from VU and AMC used 100.000 CoreHours during hands-on classes in bioinformatics and genomics.

- Visualization classes using Jupyter Notebooks in the cloud and Pandas, NetworkX, Folium (geological viz.).

- Hadoop training with Jupyter Notebooks in the Cloud using Spark to connect to SURFsara's Hadoop cluster.

- Hackathons

Organization

The course organizer:

- plans and requests resources,

- prepares and tests the VM images,

- launches VMs, creates and distributes student logins,

- cleans up afterwards.

The cloud provider:

- makes sure the resources are available.

SURF SARA

Choices

- Students work all at the same time and in the same room, or spread out in time and space?

- Use 1 big VM for everybody or 1 each?

- Where are the individual results stored?

- What is done locally, what in the cloud?

- Do I need a plan B?

SURF SARA

What can go wrong?

- Local testing OK, cloud image not OK.

- Performance problems on scale up/out.

- A VM dies, data is lost.

- Resources (partially) unavailable.

- Cloud down or network connection failure.

Your perspective

- How much can you trust the cloud provider to deliver the resources? What is your plan B?

- Do you have the skills to handle *many* VMs?

- Wouldn't you like the provider to handle VM startups and logins?

- Think about "high-availability" v.s. "prepare to fail".

- What problem does Docker solve?

Cloud provider perspective

- Capacity planning.

- Integration with Docker machine.

- Unpredictable usage by researchers → lock resources well in advance.

- SLA, Availability.

- High impact on current cloud, researchers suffer.

SURF SARA

Docker to the rescue?

Virtual Machine

- "safe" environment

- full blown boot

- "normal" OS

- multi-user

- hard to test locally

- full install and maint.

- familiar technology

Docker container

- not (yet) safe enough

- fast start/stop, small footprint

- single process

- multiple containers

- environment different

- reproducible build

- new (better!)

VM meets Docker

- Use VMs to host Docker containers.

- Login and IP/port access management remains.

- Embrace Docker philosophy: prepare to fail.

- Need remains for simple supporting tools.

- VM and container technology will develop towards each other.

SURF SARA

Future at SURFsara

- SURFsara wants a separate cloud for non-research activities.

- Resource planing.

- Docker containers with predictable environment.

- Managing end-users:
  - Acceptance of terms of use.
  - Login names, passwords, public keys.
  - IP address/port for end-user access.

- WaaS: Workshop as a Service?

- Dynamic scaling?

Thanks, any questions?