University of Zurich[UZH]

GC3: Grid Computing Competence Center

# Elasticluster

## Automated provisioning of computational clusters in the cloud

Sergio Maffioletti
GC3: Grid Computing Competence Center
University of Zurich.

CHOSUG 06 December 2013

# GC3: the Grid Computing Competence Center

"The bridge between research
and computational infrastructure"

How ?

- Support scientists who need to run large-scale data processing.
- Develop tools to better integrate scientific usecases.
- Provide access to innovative infrastructures and technologies.

Want to know more ? http://www.gc3.uzh.ch

# Do you need to deploy...

a SGE cluster

. . . to cloud-enable your existing workload.

a Matblab cluster

. . . to run Matlab Distributed Computing Server.

an Hadoop cluster

. . . to scale your data processing.

an Ipython cluster

. . . parallelize the execution of your python code.

# What issues you may find

Manual deployment and configuration is cumbersome and error prone

Too many home made shell scripts with lot of assumptions on the local infrastructure

Need to migrate deployment from one provider to another

# What is elasticlluster

Elasticluster provides a user-friendly command line tool to create, manage and setup computing clusters hosted on cloud infrastructures like Amazon's Elastic Compute Cloud EC2, Google Compute Engine or a private OpenStack cloud).

Its main goal is to get your compute cluster up and running with just a few commands.

# How does elasticluster work?

Command line tool

1. creates virtual machines in a cloud
2. installs and configures the software you want
3. add and remove nodes if needed

customization is done by editing text files

# elasticluster demo

1. create 5 virtual machines on an OpenStack cloud.
2. install and configure Hadoop on them.
3. connect to the cluster.
4. Run an example.
5. destroy the cluster when done.

*show time!*

# elasticluster demo

1. create 5 virtual machines on an OpenStack cloud.
2. install and configure Hadoop on them.
3. connect to the cluster.
4. Run an example.
5. destroy the cluster when done.

*show time!*

# Configuration and management

We use **ansible** to deploy applications and perform configuration:

- software configuration is encoded in a text file
  - everything is on the client machine
  - changes are reproducible
- base OS images are used
  - independent from the infrastructure
  - Agentless: only python 2.4 or greater is required
- the same configuration works also on metal machines

# elasticluster features (1)

## Wide support for Batch cluster

- SLURM
- OpenGridEngine
- Torque+MAUI

## other type of computational clusters

- Hadoop
- Matlab Distributed Computing Servers

## multiple distributed filesystems

- GlusterFS
- Ceph
- HDFS

# elasticluster features (1)

## Wide support for Batch cluster

- SLURM
- OpenGridEngine
- Torque+MAUI

## other type of computational clusters

- Hadoop
- Matlab Distributed Computing Servers

## multiple distributed filesystems

- GlusterFS
- Ceph
- HDFS

# elasticluster features (2)

## Run on multiple clouds

- Amazon EC2
- OpenStack
- Google Compute Engine

## on multiple operating systems

- Ubuntu
- CentOS
- Scientific Linux

# elasticluster demo continued...

From a running Hadoop cluster . . .

1. add one more worker node.
2. re-run the example.
3. destroy the cluster when done.

*show time!*

# elasticluster demo continued...

From a running Hadoop cluster . . .

1. add one more worker node.
2. re-run the example.
3. destroy the cluster when done.

*show time!*

# Similar products

**StarCluster**

- Setup is bound to pre-configured image
- Not compatible with OpenStack or GCE (uses specific Amazon functionality to identify clusters)

**VirtualCluster**

- Setup is bound to pre-configured images
- Makes many assumptions about the underlying OpenStack setup
- Not sure about codebase maintenance

# References

- Elasticluster web page:
  http://gc3-uzh-ch.github.io/elasticluster/
- Elasticluster on PyPI:
  https://pypi.python.org/pypi/elasticluster

```
$ pip install elasticluster
```

- Elasticluster github page:
  https://github.com/gc3-uzh-ch/elasticluster/
- Elasticluster documentation:
  https://elasticluster.readthedocs.org
- GC3 home page: http://www.gc3.uzh.ch
- Ansible home page: http://www.ansibleworks.com