



# Elasticcluster

Provisioning computational clusters in the cloud with Python

**Antonio Messina** <antonio.s.messina@gmail.com>

Nicolas Bär <nicolas.baer@gmail.com>

## Who we are

The GC3 group supports scientists who need to run large-scale data processing.

*large-scale* means that one or just a few computers are not enough

so you need a *cluster*, a group of computers acting like just one system.

but if you don't have one, or it does not fit your needs, what can you do?

## Who we are

The GC3 group supports scientists who need to run large-scale data processing.

*large-scale* means that one or just a few computers are not enough

so you need a *cluster*, a group of computers acting like just one system.

but if you don't have one, or it does not fit your needs, what can you do?

## Who we are

The GC3 group supports scientists who need to run large-scale data processing.

*large-scale* means that one or just a few computers are not enough

so you need a *cluster*, a group of computers acting like just one system.

but if you don't have one, or it does not fit your needs, what can you do?

## Who we are

The GC3 group supports scientists who need to run large-scale data processing.

*large-scale* means that one or just a few computers are not enough

so you need a *cluster*, a group of computers acting like just one system.

but if you don't have one, or it does not fit your needs, what can you do?

# Three solutions

## 1. Buy a cluster

- buy the machines
- find a room
- setup air conditioning and ensure you have enough power
- hire a system administrator

## 2. Run on someone else's cluster

- it may not have all the software you need
- need to negotiate policies
- resource usage conflicts

## 3. Use **elasticcluster** to create a cluster of virtual machines *in the Cloud*

- you choose the software and the configuration
- as soon as you need it

# Three solutions

## 1. Buy a cluster

- buy the machines
- find a room
- setup air conditioning and ensure you have enough power
- hire a system administrator

## 2. Run on someone else's cluster

- it may not have all the software you need
- need to negotiate policies
- resource usage conflicts

## 3. Use **elasticsearch** to create a cluster of virtual machines *in the Cloud*

- you choose the software and the configuration
- as soon as you need it

# Three solutions

## 1. Buy a cluster

- buy the machines
- find a room
- setup air conditioning and ensure you have enough power
- hire a system administrator

## 2. Run on someone else's cluster

- it may not have all the software you need
- need to negotiate policies
- resource usage conflicts

## 3. Use **elasticsearch** to create a cluster of virtual machines *in the Cloud*

- you choose the software and the configuration
- as soon as you need it



# How does elasticcluster 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

## elasticsearch demo

1. create 3 virtual machines on an OpenStack cloud.
2. install and configure the SLURM queue system on them.
3. connect to the cluster.
4. submit a simple job.
5. check that it is actually running :)
6. add one more worker node.
7. destroy the cluster.

*show time!*

## elasticcluster demo

1. create 3 virtual machines on an OpenStack cloud.
2. install and configure the SLURM queue system on them.
3. connect to the cluster.
4. submit a simple job.
5. check that it is actually running :)
6. add one more worker node.
7. destroy the cluster.

*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
- the same configuration works also on *real* machines

## **elasticcluster features (1)**

Different kind of computational clusters are supported:

- Batch systems:
  - SLURM
  - OpenGridEngine
  - Torque+MAUI
- Hadoop
- Matlab Distributed Computing Servers

Multiple distributed filesystems:

- OrangeFS/PVFS
- GlusterFS
- Ceph
- HDFS

## **elasticcluster features (1)**

Different kind of computational clusters are supported:

- Batch systems:
  - SLURM
  - OpenGridEngine
  - Torque+MAUI
- Hadoop
- Matlab Distributed Computing Servers

Multiple distributed filesystems:

- OrangeFS/PVFS
- GlusterFS
- Ceph
- HDFS

## **elasticcluster features (2)**

Run on multiple clouds:

- Amazon EC2
- OpenStack
- Google Compute Engine

Works with multiple operating systems:

- Ubuntu
- CentOS
- Scientific Linux

## References

- Elasticcluster on PyPI:

<https://pypi.python.org/pypi/elasticcluster>

```
$ pip install elasticcluster
```

- Elasticcluster github page:

<https://github.com/gc3-uzh-ch/elasticcluster/>

- Elasticcluster web page:

<http://gc3-uzh-ch.github.io/elasticcluster/>

- Elasticcluster documentation:

<https://elasticcluster.readthedocs.org>

- GC3 home page: <http://www.gc3.uzh.ch>

- Ansible home page: <http://www.ansibleworks.com>



# Thank you

## elasticsearch feature summary

- works on Amazon EC2, OpenStack and Google GCE
- Creates the cluster you need, when you need it, starting from vanilla images
- Typical use cases:
  - On demand computational cluster provisioning
  - Testing of new infrastructures or configurations
- All the configuration is on your laptop.
- easy to modify the setup of the virtual machines.
- makes your results *reproducible*

# Ansible

## Configuration and management system

- Goal oriented, not scripted
- Agentless (only python 2.4 or greater is required in the managed machine)
- changes are reproducible and idempotents
- smooth learning curve
- very well documented
- responsive community
- actively developed

website: [www.ansibleworks.com](http://www.ansibleworks.com)

## 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