

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

elasticsearch features (2)

Run on multiple clouds:

- Amazon EC2
- OpenStack
- Google Compute Engine

Works with multiple operating systems:

- Ubuntu
- CentOS
- Scientific Linux

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 idempotent
- smooth learning curve
- very well documented
- responsive community
- actively developed

website: www.ansibleworks.com

elasticcluster 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!

elasticcluster 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!

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>

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>