Contribution ID: **17**   Type: **Technical Session**

Elasticluster: provisioning computing clusters in the cloud with Python

Tuesday, 17 September 2013 11:05 (5 minutes)

Description of Work

Computational science has a long history of exploiting
batch-queueing compute clusters. Traditionally, this required buying and maintaining the compute cluster hardware, with the associated manpower and cost burden, often subtracted from research time and budget. The availability of powerful computing hardware in IaaS clouds is a game changer in this respect, in that it makes cloud computing attractive also for computational workloads that were up to now almost exclusively run on HPC clusters.
We present Elasticluster: a Python command line tool to create,
manage and setup compute clusters hosted on cloud infrastructures. Elasticluster can provision clusters on Amazon's Elastic Compute Cloud EC2 (and compatible ones), Google Compute Engine, or private clouds based on OpenStack.

More generally, Elasticluster allows "mix and match" of cluster
components, by leveraging the Ansible roll-out and configuration
engine (Ansible is written in Python and uses YAML as a
configuration/scripting language). New cluster configurations can be added by providing new Ansible play-books.

We would like to show how Elasticluster is used at the Grid
Computing Competence Center to enable self-service provisioning of compute infrastructures by research groups, and how it is used by systems administrators to set up test environments for new versions of the computational and systems software.

In the proposed talk will present how Elasticluster could be used
to enable an existing HPC workload on a cloud infrastructure; will also present the software architecture of Elasticluster and how it can be used from other Python programs to automate infrastructure provisioning.

URL for further information

https://github.com/gc3-uzh-ch/elasticluster

Wider Impact of this Work

What sets Elasticluster apart from similar attempts (e.g.,
STARcluster, Rocks'Virtual Cluster) is the fact that the entire
cluster configuration, including what software to install and how to set it up, is stored in text files on the client side. This allows a few very desirable features:
The same cluster set up can be executed on different cloud
infrastructures.
There is no dependency on pre-configured VM images: a cluster
can be installed on top of a basic Linux installation.
Different types of compute clusters can be installed:
traditional batch-queueing systems (e.g., SLURM, Grid Engine,
TORQUE+MAUI), Map/Reduce systems (Hadoop), etc.

Session, double-session

Session

Printable Summary

Elasticluster is a Python command line tool to create, manage and setup compute clusters hosted on cloud infrastructures. Elasticluster can provision clusters on Amazon's Elastic Compute Cloud EC2 (and compatible ones), Google Compute Engine, or private clouds based on OpenStack.

Primary author: BÄR, Nicolas (University of Zurich)

Co-authors: MURRI, Riccardo (UZH); MAFFIOLETTI, sergio (UZH)

Presenter: BÄR, Nicolas (University of Zurich)

Session Classification: Cloud computing: Lightning talks

Track Classification: EGI & Infrastructure as a Service (IaaS) Cloud Platforms (David Wallom/Michel Drescher)