Using Puppet to contextualize computing resources for ATLAS analysis on Google Compute Engine

UPPSALA UNIVERSITET

Henrik Öhman on behalf of the ATLAS Collaboration

ATLAS EXPERIMENT

Introduction

With new cloud technologies come new challenges, and one such is the contextualization of cloud resources with regard to requirements of the user and his experiment. We investigate the use of Puppet [1] to facilitate contextualization of cloud resources for ATLAS [2] analysis on Google's new cloud platform Google Compute Engine (GCE) [3].

ATLAS Analysis Clusters

An ATLAS analysis cluster consists of several elements. The software necessary for ATLAS analysis must be available on the computing elements; the computing elements must be connected to ATLAS storage to access input data and write analysis output; the cluster must have a job brokering system to distribute jobs between computing elements; and finally a service to request analysis jobs from the ATLAS distributed analysis system PanDA [4].

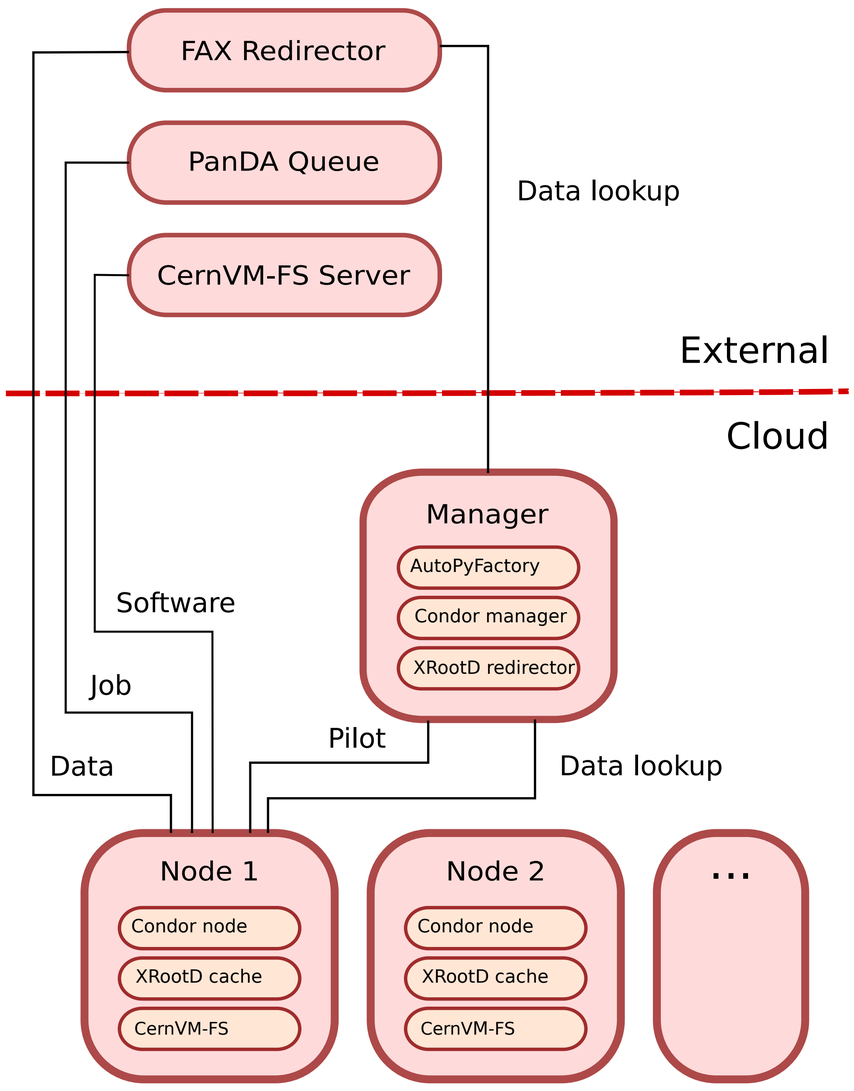

Figure: An ATLAS Analysis Cluster in the Cloud

Google Compute Engine

GCE is a relatively new Infrastructure-as-a-Service (IaaS) platform and it is unique in some important respects. The absence of a function to upload custom machine images is limiting, and prevents the use of readily available tools to create such images. The methods provided by GCE to customize images are not usable to create cloud independent configurations, and the process can be experienced as awkward and with high turn-around times.

Contextualization and Puppet

An alternative way to contextualize machines that enables configuration from a bare state is desired. Puppet is software that automates many tasks that system administrators perform and it offers a comprehensive way to define and enforce the desired state of a machine. Through Puppet's mechanisms it is possible to configure machines for their different roles in ATLAS analysis clusters. Moreover the configuration can be made to be both OS and cloud independent, and can also be customized to suit different cluster configurations.

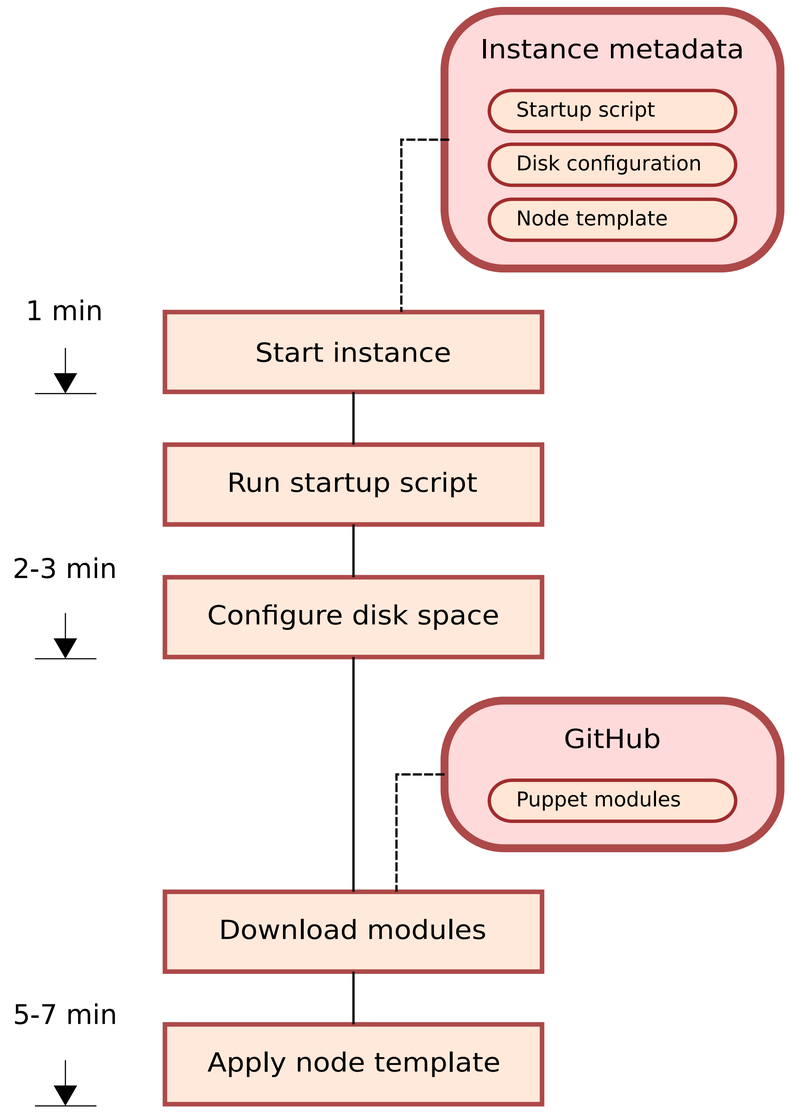

Figure: Schematic over the contextualization process. Times are approximate.

Results

We have used Puppet to contextualize machines for ATLAS analysis with Puppet on GCE. The Puppet configuration we have used [5] to contextualize these machines can be extended to any cloud service, grid site, or other system of machines. Contextualizing machines from a bare state is more time consuming than starting pre-configured machines, but the time it takes for machines to become fully contextualized still lies at a manageable 5-7 minutes. There are clear benefits from using Puppet: among others it offers a clear and comprehensive way to specify a machine configuration that can be made to work with multiple machine roles, on any operating system, and with any cloud provider.

References

[1] *Puppet Labs: IT Automation Software for System Administrators*, http://puppetlabs.com/ (Accessed: 2013-10-01)

[2] *The ATLAS Experiment at the CERN Large Hadron Collider*, ATLAS Collaboration. JINST 3 (2008) S08003

[3] *Google Compute Engine — Cloud Platform*, https://cloud.google.com/products/compute-engine (Accessed: 2013-10-01)

[4] *Overview of ATLAS PanDA Workload Management*, T. Maeno, Proc. of the 18th Int. Conf. on Computing in High Energy and Nuclear Physics (CHEP2010)

[5] *atlasgce-modules*, https://github.com/spiiph/atlasgce-modules (Accessed: 2013-10-10)