Setting up a cloud data processing facility

Frederik Orellana[1],...
October 2009

1 Niels Bohr Institute, University of Copenhagen

Introduction

The work described in the present document was motivated by a desire to streamline and simplify the steps a researcher has to go through in order to use a distributed infrastructure to process or simulate large amounts of data. Specifically, the aim is to prepare for analyzing the data that will now very soon arrive from the Large Hadron Collider at CERN.

The main idea is to strictly support the way most researchers do their computational work: 1) log in to a work station and compile, debug and test a given application, 2) run this same application multiple times[2], either with different parameters or with different input files. Typically 2 is done via some sort of batch or queuing system.

The crucial point here is that in order to painlessly go from step 1 to step 2, exactly the same environment (operating system, installed software and libraries) should be present both on the workstation where the development and initial testing was done *and* on the machine where the jobs are later placed by the mentioned batch or queuing system.

The way we have chosen to address this is via virtualization and a catalog of disk images. With disk images, we here refer to disk images suitable for booting a virtual machine. To manage jobs and interact with this catalog, we have developed a prototype GUI, coined GridPilot.

More detailed, the envisioned working procedure of a researcher on our facility is then:

1. Select, boot up and log in to a virtual machine – this is a one-step procedure, initiated by simply logging in via SSH to the front-end machine, where-after the researcher is prompted to choose a virtual machine.
2. Compile, debug, test.
3. If necessary, save the current state of the virtual machine as a new image and add it to the catalog.[3]
4. Prepare and run multiple jobs, each requiring to run in the same kind of virtual machine. This is done via the GUI.

Log-in

The front-end machine of the facility support log-in via standard SSH. On log-in, the front-end uses the WAYF and Confusa web services to verify the user's user name and password by redirecting him to his home institution for log-in. After that, the same web services are used for generating short lived X.509 credentials that then reside on the front-end. After log-in on the front-end is completed, the user is prompted if he wants to be logged in on any virtual machines he may have running or if he wants to choose and boot up a fresh virtual machine.

Virtualization, disk images and software provisioning

The shell script that is run when a user logs in on the front-end machine uses the OpenNebula CLI for finding out which machines the user already may have running, getting a list of available images and booting up a virtual machine using one of these images.

Some of the images in question are provided by the CernVM project. These images include the cvmfs/fuse kernel module, which is used for providing software. For more detail on this, see CERNVM.

2 referred to as running jobs.

3 This is not yet implemented.

Others simply have software pre-installed. All images are raw disk images and are used for booting up virtual machines via the Xen hypervisor.

Network storage

To allow mounting network storage inside the virtual machine on behalf of the user, the virtual machine must have the davfs2/fuse kernel module installed and loaded and the user must be member of a certain so-called virtual organization (see GRIDDK). All images in our catalog and all of our users satisfy these requirements.

The storage is provided over HTTPS from the dCache installations operated by the NDGF.

[TODO: NDGF, please fill in.]

Job and file management

Use cases

ATLAS data processing

Statistical epistasis analysis

Variance at risk simulation

Outlook

--- add VM functionality to grid middleware

Bibliography

GRIDDK, F. Orellana, J. Berthold, J. Bardino and B. Sedoc, “grid.dk – a compute infrastructure for Danish academia”, *in preparation*

CERNVM, P. Buncic, C. Aguado Sanchez, J. Bloomer, L. Franco, S. Klemer and P. Mato, “CernVM - a virtual appliance for LHC applications“, PoS(ACAT08)012, http://pos.sissa.it/cgi-bin/reader/conf.cgi?confid=70