NRC·CNRC
Herzberg Institute of Astrophysics

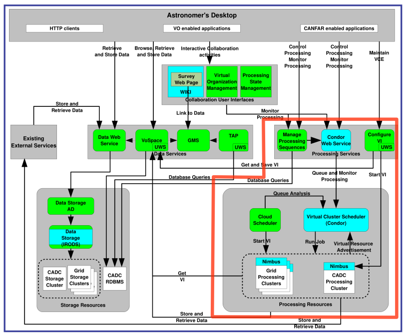
Virtualisation and Grid Utilisation within the CANFAR Project

CADC/CCDA

Séverin Gaudet (NRC), Patrick Armstrong (UVic), Nick Ball (NRC), Ed Chapin (UBC), Pat Dowler (NRC), Ian Gable (UVic), Sharon Goliath (NRC), Sébastien Fabbro (UVic), Laura Ferrarese (NRC), Stephen Gwyn (NRC), Norman Hill (NRC), Dustin Jenkins (NRC), JJ Kavelaars (NRC), Brian Major (UVic), John Ouellette (NRC), Mike Paterson (UVic), Michael Peddle (NRC), Chris Pritchet (UVic), Sarah Sadavoy (UVic), David Schade (NRC), Randall Sobie (UVic), David Woods (UBC), Kristen Woodley (UBC), Alinga Yeung (UVic)

Abstract

The Canadian Advanced Network For Astronomical Research (CANFAR) is an operational system for the delivery, processing, storage, analysis, and distribution of very large astronomical datasets. The goal of CANFAR is to support large Canadian astronomy projects. CANFAR is cyber-infrastructure combining the Canadian national research network (CANARIE), grid processing and storage resources (Compute Canada) and a data centre (CADC) into a unified storage and processing system. For processing, the project has combined the best features of the grid and cloud processing models by providing a self-configuring virtual cluster deployed on multiple cloud clusters. The CANFAR processing service makes use of many technologies from the grid, cloud and VO communities such as Condor, Nimbus (or OpenNebula, Eucalyptus or Amazon EC2), Xen, Cloud Scheduler, VOSpace, UWS, SSO, CDP and GMS.



CANFAR architecture with the processing system outlines in red

The Processing Context

There are existing models for providing CPU cycles to users, including well-established Grid systems and nascent cloud services. From anecdotal evidence, data intensive users have tried to use existing Grid infrastructure with unsatisfactory results. The problems reported by astronomers are:

- Environment customization and maintenance issues – Grid operators install and maintain the environment for their users. Astronomers tend to run complex software with many dependencies. Making their software run correctly in the environment provided by a cluster operator is very difficult. Cluster operators modifying their system to incorporate updates or support other users make the software maintenance issue an ongoing problem.
- Poor responsiveness – Astronomers complained that it would often take days for a job to start running. One common response to the poor responsiveness was to seek out other clusters that may be less busy. This exacerbated the environment configuration issues described above.

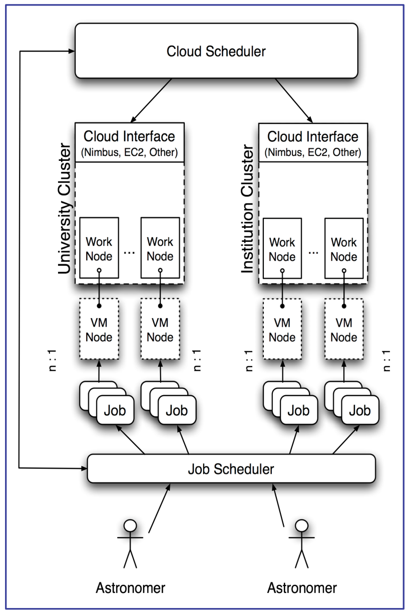
We are not aware of serious attempts by astronomers to use cloud infrastructure, however we anticipate several problems with the cloud:

- Until recently, only commercial clouds have been available. Paying the usage fees for significant CPU cycles will lead to funding and administrative issues.
- Clouds do not inherently provide job scheduling. A user can incorporate a job scheduling system into their virtual environments, however this is a non-trivial process.
- Clouds do not inherently share resources between multiple users. Existing clouds tend to allocate resources on a first-come-first-served basis.

To address the limitations of the grid and cloud processing models, the CANFAR project has produced a hybrid processing system that makes use of scheduling while abstracting the grid processing resources as a cloud.

Features	Grid	Cloud	CANFAR
Ample CPU Cycles	✔	✔	✔
Job Scheduling	✔	✘	✔
User customized environment	✘	✔	✔
Resource Sharing	✔	✘	✔
Portability of environment	✘	✔	✔

Feature comparison of Grid, Cloud and CANFAR processing



Cloud Scheduler schematic showing multi-cluster use

```
# Job specifications
Universe     = vanilla
Executable   = sgw.sh
should_transfer_files = YES
when_to_transfer_output = ON_EXIT
getenv       = True
notification = Never
notify_user  = stephen.gwyn@nrc.ca

# Run-environment requirements
Requirements = VMType =?= "canfarbase_sgw" && Arch == "INTEL" &&
Memory >= 2048 && Cpus >=  1

# User requirements for Cloud Scheduler
+VMName          = "canfarbase_sgw"
+VMNetwork       = "private"
+VMCPUArch       = "x86"
+VMLoc           = https://www.cadc.hia.nrc.gc.ca/vospace/sgwyn/
vmdir/canfarbase_i386_sgw_AT.img.gz?view=data"
+VMMem           = "2048""
+VMCPUCores      = "1"
+VMStorage       = "10"

Arguments        = 708474
Output           = sex708474.sgw.out
Error            = sex708474.sgw.err
Log              = sex708474.sgw.log
Queue
```

Condor job submission file with Cloud Scheduler requirements highlighted

```
Name                OpSys     Arch   State     Activity LoadAv Mem   ActvtyTime

elephant21          LINUX     INTEL  Claimed   Busy     1.000  2048  4+18:09:16
elephant23          LINUX     INTEL  Claimed   Busy     1.000  2048  4+00:20:52
elephant25          LINUX     INTEL  Claimed   Busy     1.000  2048  4+18:00:07
elephant26          LINUX     INTEL  Claimed   Busy     1.000  2048  4+17:45:45
elephant27          LINUX     INTEL  Claimed   Busy     1.000  2048  4+19:13:57
…
elephant57          LINUX     INTEL  Claimed   Busy     1.000  2048  4+20:17:04
elephant58          LINUX     INTEL  Claimed   Busy     1.000  2048  0+01:43:53
elephant59          LINUX     INTEL  Claimed   Busy     1.000  2048  4+20:17:15
elephant60          LINUX     INTEL  Claimed   Busy     1.000  2048  4+12:13:57
elephant61          LINUX     INTEL  Claimed   Busy     1.000  2048  4+20:11:28
hermes-xen002       LINUX     INTEL  Claimed   Busy     1.000  2048  0+12:45:03
hermes-xen003       LINUX     INTEL  Claimed   Busy     1.000  2048  0+04:11:18
hermes-xen004       LINUX     INTEL  Claimed   Busy     1.000  2048  0+03:49:03
hermes-xen006       LINUX     INTEL  Claimed   Busy     1.000  2048  3+15:19:56
hermes-xen010       LINUX     INTEL  Claimed   Busy     1.000  2048  1+12:45:54
…
hermes-xen196       LINUX     INTEL  Claimed   Busy     1.000  2048  0+02:00:48
hermes-xen197       LINUX     INTEL  Claimed   Busy     1.000  2048  0+12:16:28
hermes-xen198       LINUX     INTEL  Claimed   Busy     0.960  2048  0+00:00:49
hermes-xen199       LINUX     INTEL  Claimed   Busy     1.000  2048  0+00:19:33
hermes-xen200       LINUX     INTEL  Claimed   Busy     1.010  2048  0+00:07:46
slot1@129.128.50.1  LINUX     X86_64 Unclaimed Idle     0.000  3894  0+00:06:04
slot1@129.128.50.1  LINUX     X86_64 Claimed   Busy     0.000  3894  0+00:00:03

                     Total Owner Claimed Unclaimed Matched Preempting Backfill

         INTEL/LINUX   170     0     170         0       0          0        0
        X86_64/LINUX     2     0       1         1       0          0        0

               Total   172     0     171         1       0          0        0
```

Condor status showing a virtual cluster of 172 VMs running on multiple clouds: elephant (UVictoria), hermes (WestGrid) and slot (UAlberta)

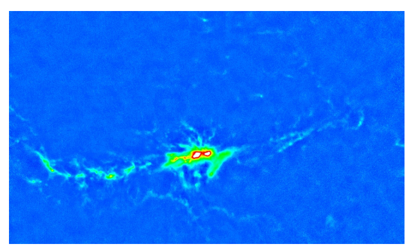
The Processing System

The CANFAR processing system presents a grid-like interface to users and creates a virtual cluster built from resources supplied by multiple cloud providers. The technologies used or supported by the CANFAR processing system are:

- **Virtual Image (VI) Management**: A service to allow users to boot and configure, save and share Virtual Images. The Virtual Images are made available to the Cloud Scheduler, where they are booted into the virtual cluster.
- **Virtualization**: Xen virtualization is being used. Both Xen and KVM were considered as potential virtualization technologies. Xen was selected because it was the most popular virtualization technology at the time. In addition, it was the only one utilized experimentally by facility operators.
- **Job Scheduler**: The CANFAR virtual cluster requires a batch job processing system to provide the functionality of a Grid cluster. Although both Condor and Grid Engine where considered, Condor was selected because it allows Virtual Machines to join the virtual cluster without modifying the Condor configuration. Grid Engine would have required the cluster configuration to be modified each time a Virtual Machine joins or leaves the virtual cluster.
- **Cloud Scheduler**: The glue between the cloud clusters provisioned by Nimbus, and the batch interface provided by Condor. The Cloud Scheduler examines the workload in the Condor queue, and uses the resources from multiple cloud clusters to create a virtual cluster suitable for the current workload.
- **Cloud functionality**: The primary cloud technology supported by the Cloud Scheduler is the Nimbus toolkit. Partial support was also developed for openNebula and Eucalyptus and Amazon EC2. Nimbus was selected as the primary development target because it is open source and allows the cloud workload to be intermixed with conventional batch jobs unlike the other systems. It is believed that this flexibility makes the deployment more attractive to facility operators.
- **Operating systems**: Both host and guest operating systems are assumed to be some flavour of Linux. The guest operating system currently in use is Scientific Linux 5.5.

Project Name	Project Lead	Institution	Telescope
SCUBA-2 Cosmology Legacy Survey	Mark Halpern	UBC	JCMT
SCUBA-2 All-Sky Survey	Douglas Scott	UBC	JCMT
Next Generation Virgo Survey	Laura Ferrarese	NRC-HIA	CFHT
Pan-Andromeda Archaeological Survey	Harvey Richer	UBC	CFHT
Time Variable Sky	Chris Pritchet	UVic	CFHT
Canada-France Ecliptic Plane Survey Simulator	JJ Kavelaars	NRC-HIA	CFHT
Shapes and Photometric Redshifts for Large Surveys	Ludo Van Waerbeke	UBC	CFHT

Projects currently using the CANFAR processing system



Dust emission from Orion Molecular Cloud at 850um processed using CANFAR (JCMT SCUBA-2 map thanks to Ed Chapin and Mark Halpern, UBC)

Status

- Since beta release in mid-June, the system has been dynamically using 3 clusters and used 40 core-years of processing in 21 weeks.
- Scientists are taking it up
- Once judged to be sufficiently robust (or at least the problems well understood and avoidable), and additional Compute Canada grid facilities are added to the processing pool, CANFAR services will be made available the whole community.
- The CANFAR project is scheduled to finish in 2011 at which point the Canadian Astronomy Data Centre will assume responsibility for operations.

University of Victoria · University of British Columbia · canarie · CANFAR Canadian Advanced Network for Astronomical Research · compute • calcul CANADA · University of Saskatchewan