On-Demand Research Computing - Infrastructure as a Service -
Motivation

• **Research computing means many different things…**
  – Scientific workflows have different requirements at each step
  – Cloud is only part of the solution
  – Connecting to and from other CI resources is important

• **Nobody likes a bad surprise**
  – Transparency, no hidden costs
  – Need a way to bound financial risk

• **Economies of scale**
  – Sharing hardware and software where it makes sense
  – Pay for what you need, when you need it

• **Customized environments for various disciplines**
  – Collaboration tools
  – Data storage & analysis tools
  – Flexibility to support different computing models (e.g. Hadoop)
Red Cloud Provides:

**Predictable, Reproducible, Reliable Performance**
We publish hardware specifications (CPU, RAM, network) and do not oversubscribe.

**Convenient**
Need system up and running yesterday.
Need a big fast machine for only a few months, weeks or days.
Need a small server to run continuously.

**No Hidden Costs**
No cost for network traffic in or out of the cloud.

**Fast Access to Your Data**
Fast data transfers via 10Gb Ethernet in or out of the cloud at no additional charge.
Globus Online access

**Economies of scale**
IaaS: Infrastructure
SaaS: Software

**Expert Help**
System, application, and programming consulting are available.

**Easy Budgeting with Subscriptions**
No billing surprises!

**IaaS is Amazon API Compatible**
Migrate when your requirements outgrow Red Cloud.
Some Use Cases to Consider

• **Support for Scientific Workflows**
  – Pre & post-processing of data and results
  – Data analysis
  – Globus Online for fast reliable data transfer
    • https://www.globusonline.org/

• **Collaboration**
  – Wiki hosting
  – Customized data analysis & computational environments

• **Web Portals**
  – Science Gateways
  – Domain Specific Portals
  – Hub Zero
    • http://hubzero.org/pressroom
    • http://nanohub.org

• **Event Driven Science**
  – https://opensource.ncsa.illinois.edu/confluence/display/SGST/Semantic+Geostreaming+Toolkit

• **Education, Outreach & Training**
  – Pre-configured systems & software tools providing consistent training platform
  – Common laboratory computing environment

• **Bursting**
  – Additional software and hardware on demand
Infrastructure as a Service (IaaS)

red cloud provides on-demand:

• Virtual Servers

• Storage: Virtual disks for the virtual servers

Plus the necessary fast networking, power, space, and cooling to make everything work.

Configuration

• Total
  – 96 cores, 384GB of RAM in 8 servers:
• 8 Servers
  – 2, 2.7 GHz Xeon E5650 CPUs (2 x 6 cores)
  – 48GB of RAM (4GB/core)
  – Maximum virtual server size:
    • 12 cores, 48GB RAM
• 10 Gigabit Ethernet interconnect
• Storage
  – 7.5TB
  – Accessible by virtual servers via 10Gb iSCSI
• Software
  – Eucalyptus 2.0.3 (Open Source)
Subscription-based Recovery Model

Cornell University  $500/core year*

Other Academic Institutions  $750/core year

*A core year is equal to 8585 hours

Each subscription account includes 50GB of storage
<table>
<thead>
<tr>
<th>Storage</th>
<th>Consulting</th>
<th>Additional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cornell Users</td>
<td>$59.90/hour</td>
<td>$0.91/GB/year</td>
</tr>
<tr>
<td>Other Academic</td>
<td>$85.47/hour</td>
<td>$1.45/GB/year</td>
</tr>
<tr>
<td>Institutions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
About Cornell Center for Advanced Computing (CAC) Projects

- CAC operates on a cost-recovery model.
- Services are charged to the project’s CU account number.
- Published rates are subsidized by the Provost and revised annually.
- Fee-based services, e.g.:
  - Computing (consider pay as-you-go vs. lease)
  - Consulting
  - File storage
  - Cluster maintenance
  - Red Cloud (cloud computing subscription)
- No-cost services:
  - Exploratory accounts
  - General help
  - Documentation
  - Training workshops
  - Access to research networks
How to Set up a Project

http://www.cac.cornell.edu/services/

• PI: Complete the *Project Request Form*
  – project title and abstract
  – 7 digit CU account number
  – account point of contact: CU NetID

• Email is sent to the account rep (immediate)

• Account rep: Complete the *Full Project Account Information Form*
  – Check the 7 digit CU account number
  – Add Sub-account, Project, and Sub-object codes
  – Enter Account expiration date

• Email is sent to the PI (within 4 hours)
How to Manage a Project

http://www.cac.cornell.edu/services/

• PI and/or Proxy: Use the *Manage your Project Form*
  – Update description, account number, contact information
  – Add project members, specify a Proxy
  – Specify resource limits: computing, storage, consulting, cloud subscription

• Account rep: Use the *Update your Account Form*
  – Update account number, expiration date, contact information
  – Add project members, specify a Proxy
  – Specify resource limits: computing, storage, consulting, cloud subscription

• Monthly invoices are sent by email as PDFs

• Send questions to help@cac.cornell.edu
Infrastructure as a Service (IaaS) Cloud

Red Cloud provides on-demand:

- **Computing Cycles**: Virtual Servers in Cloud “Instances”
- **Storage**: Virtual Disks in Elastic Block Storage (“EBS”) Volumes
Red Cloud Documentation


- Have questions? E-mail [help@cac.cornell.edu](mailto:help@cac.cornell.edu)
Cloud Management Tools:

- euca2ools: Linux command line tool
- HybridFox: See Red Cloud user documentation for setup info
- Amazon EC2 API
First Time Logon

- Go to https://cloud.cac.cornell.edu:8443 in your web browser.
- Click on “Recover the Password” to set the initial password. This is **not** your CAC account password!
- After logging in, under the “Credentials” tab,
  - Click on the “Download Credentials” button to download X.509 credentials to use with euca2ools.
  - Click on the “Show Keys” button to obtain Amazon-style string-based keys (used by HybridFox)
Cloud Management Tool: euca2ools

- Linux command line tool


  - Download and install on your own Linux machine, or
  - If you have a CAC account, the tool is available on cloud-login.cac.cornell.edu. Log in using your CAC user credentials.
Set Up euca2ools

- Unpack the zip file containing X.509 user credentials downloaded from the “Credentials” tab:
  ```
  mkdir .euca
  cd .euca
  unzip euca2-<user name>-x509.zip
  ```
- If on a shared system, restrict permissions so the directory is readable to you.
  ```
  chmod 0700 ~/.euca
  chmod 0600 ~/.euca/*
  ```
- Setting up the environment. Do this before using euca2ools from a new login session.
  ```
  source .euca/eucarc
  ```
- Generate a keypair for logging into your instances:
  ```
  euca-add-keypair mykey | tee .euca/mykey.private
  chmod 0600 .euca/mykey.private
  ```
Start an Instance

- List available images for launching an instance:
  `euca-describe-images`
- Officially Supported Linux Distributions:
  - CentOS 5 (`emi-799C12DB & emi-17A11C17`)
  - CentOS 6 (`emi-296211C7`)
  - Ubuntu 10.04 (`emi-2EE41562`)
- Instance Types: we don’t oversubscribe CPU or RAM!

<table>
<thead>
<tr>
<th>Type</th>
<th>Cores</th>
<th>RAM</th>
<th>Disk</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1.small</td>
<td>1</td>
<td>4 GB</td>
<td>20 GB</td>
</tr>
<tr>
<td>C1.medium</td>
<td>2</td>
<td>8 GB</td>
<td>40 GB</td>
</tr>
<tr>
<td>M1.large</td>
<td>4</td>
<td>16 GB</td>
<td>200 GB</td>
</tr>
<tr>
<td>M1.xlarge</td>
<td>8</td>
<td>32 GB</td>
<td>400 GB</td>
</tr>
<tr>
<td>C1.xlarge</td>
<td>12</td>
<td>48 GB</td>
<td>1000 GB</td>
</tr>
</tbody>
</table>
Start an Instance Cont’d

• Start an instance: starts usage meter
  euca-run-instances -k mykey -t m1.small emi-17A11C17
  euca-describe-instances

• Once the instance is running, ssh into the instances as root user.
  ssh -i euca-shl1-x509/mykey.private

Tips

• My instance failed to respond on the network. What happened? Use the “euca-get-console-output <instance ID>” command to take a peek at the console.

• Format and use the rest of the local disk (/dev/vda2):
  mkfs -t ext3 /dev/vda2
  mount /dev/vda2 /mnt

  Warning: contents of both root disk (/dev/vda1) and local disk (/dev/vda2) will be lost when the instance is terminated!
Reboot and Terminate Instance

• Reboot an instance
  euca-reboot-instances <instance ID>

• Terminate an instance: stops usage meter
  euca-terminate-instances <instance ID>
Elastic Block Storage (EBS)

- Each EBS volume is a virtual disk that can be attached to or detached from a cloud instance on demand. Inside the cloud instance, each attached EBS volume is a block device (e.g. /dev/vdb).
- User can create EBS volumes up to 1 TB in increments of 1 GB.
- EBS volumes connect to cloud instances via iSCSI for best performance (300+ MB/sec observed in testing).

- Allocate a volume: usage meter starts
  euca-create-volume -s <size in GB> -z caccloud

- List all volumes:
  euca-describe-volumes

- Attach a volume to an instance:
  euca-attach-volume -i <instance ID> -d <device name> <volume ID>
Elastic Block Storage (EBS) Continued

- Detach a volume from an instance:
  - Make sure the file system is unmounted to flush any cached data to disk.
  - `euca-detach-volume <volume ID>`
- Destroy a volume: stops usage meter
  `euca-delete-volume <volume ID>`

Known Bugs

- The “-d” option in the `euca-attach-volume` command is required but ignored. The attached volume will be named as `/dev/vdX` where X is the next available letter in the list of device names.
- Detach all attached volumes before rebooting an instance. Otherwise the attached volume will remain “attached” but inaccessible until the instance is terminated.
Create a Customized Image

Method 1: Make a Copy of the Root Disk of a Running Instance

On a running instance,

• Customize the installation to your liking.
• Install euca2ools.
• Create a bundle of the root disk:
  euca-bundle-vol -v / -e <excluded directories> -r x86_64
  --fstab /etc/fstab --kernel <eki> --ramdisk <eri> -d
  <working directory> --no-inherit
Create a Customized Image Cont’d

- Copy the bundle elsewhere for safe-keeping, or
- Upload the customized image:
  - Upload the bundle:
    euca-upload-bundle -b <bucket name> -m <manifest>
  - Register image:
    euca-register <bucket>/<xml>
Create a Customized Image Cont’d

Method 2: start from a basic OS image

- Get a Basic OS Image here:
  https://cloud.cac.cornell.edu/images
- Mount the image: (root access required)
  mount -o loop <disk image file> <mount point>
- Add new software like this:
  rpm -root <mount point> -ivh ...
  yum -installroot <mount point> install ...
- chroot <mount point> could be useful.
- Unmount image when done:
  umount <mount point>
- Bundle the image: make sure the working directory has sufficient space!
  euca-bundle-image -i <image file> -d <working directory> --kernel <eki> --ramdisk <eri>
- Upload the bundle as method 1
Known Bug:

• Currently all uploaded images are readable by everyone. Eucalyptus ignores the image’s “launchPermission” setting (See `euca-describe-image-attributes` and `euca-modify-image-attribute commands`). Either:
  • Do not include confidential information (e.g. setting root password in `/etc/shadow`) in your uploaded images, or
  • Take down your image using `euca-deregister` and `euca-delete-bundle commands` as soon as you launch the desired instances. Keep a copy of the bundle somewhere safe and re-upload as needed.
Unsupported Amazon EC2 Features

• Saving instance root disk to EBS and booting from EBS: Coming in Eucalyptus 3!

• Elastic IP addresses / virtual private clouds / security groups

• Placement groups: not needed. All instances concentrate to a single 10 Gbit switch. Good network connectivity everywhere!