Clearing the Fog! Getting Started on Cumulus:
a Near to the Ground On Campus Cloud Solution

CU Research Computing
UNIVERSITY OF COLORADO BOULDER

Be Boulder.

Getting Started on CUmulus

Gerardo Hidalgo-Cuellar (RC User Support)
Chris Horen (RC Cloud Engineer)

www.rc.colorado.edu
rc-help@colorado.edu

Slides and tutorial:
https://github.com/ResearchComputing/CUmulus_tutorials

Survey at: http://tinyurl.com/curc-survey18

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Learning Objectives

- [Logging into CUmulus with Horizon](#) (the CUmulus web portal)
- [Creating your instance](#) (i.e. virtual machine)
- [Logging into your instance](#) via ssh

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Outline

- CURC Overview
- What is CUmulus?

- CUmulus Access
  - Access to CUmulus and the allocation process
  - Logging into Horizon (CUmulus web portal)
  - Creation of an instance
  - Logging into your instance

- Demo workflow one might use on CUmulus:
  - Web App & Database

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

CU Research Computing Overview

- Provides Computing and Data Beyond the Desktop resources:
  - High Performance Computing (HCP), clusters include:
    - Alpine (beta)
    - Summit
    - Blanca
  - Storage of Research Data
    - PetaLibrary
  - High-Speed Data Transfer
    - Globus Endpoint
  - Consulting in Computational Science and Data Management
  - **Cloud Resources**
    - **CUmulus**

What is CUmulus?

- **CUmulus** is CU Research Computing's free-to-use on-premise cloud service
- Supports cases not well-suited for HPC such as:
  - Research-Oriented Web Servers
  - Databases
  - Long-Running Services
- Provides users with persistent or ongoing availability by allocating logically isolated section of the cloud

What is CUmulus?

- You get your own virtual "world" <u>for experimentation</u> - an environment that can be easily created/tested/removed
  - Install Software
  - Administer your instance (you're in control!)
  - Run applications and jobs
  - Interface w/ other CURC services: Blanca, Alpine, PetaLibrary

- You can request specific resources (CPU, storage, memory) and can set up persistent storage

CUmulus Access

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

CUmulus Access and Allocation

Submit a proposal for your use case (email rc-help@colorado.edu)

- Describe your CUmulus workflow
- Describe why your workflow is appropriate for CUmulus
- Estimate the resources you require:
    - Operating System, CPU cores, Disk Space, Memory

This is an *iterative process* where we work with you to make sure the request for resources fits your (and our) needs

- Learn more about the allocation request process at https://www.colorado.edu/rc/userservices/allocations

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Log in to Horizon

There are 3 current authentication/login methods at cumulus.rc.colorado.edu/:

1. CU Boulder
2. CSU Fort Collins
3. XSEDE* (all other RMACC institutions)
   a. Create account: https://portal.xsede.org/#/guest
   b. Configure 2FA (Duo): https://portal.xsede.org/mfa

*Don't *need* an RC account with XSEDE for this class, but will make things *easier*

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

RC account check

- Who does **not** have a CU Research Computing account who would like to use a temporary account*?
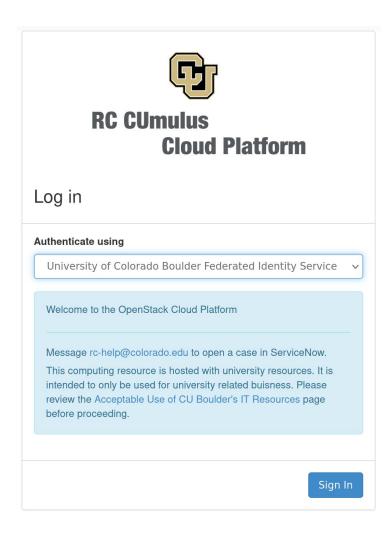
- If you don't have access to the CU VPN *or* don't have an RC account, you will *need* a temp account to connect to a VM

only available during session

Log in to Horizon

- Horizon is the CUmulus web portal
  - cumulus.rc.colorado.edu/
- Let's take a brief tour of Horizon
- Log in with your institution's credentials:

RC CUmulus Cloud Platform

Log in

Authenticate using

University of Colorado Boulder Federated Identity Service

Welcome to the OpenStack Cloud Platform

Message rc-help@colorado.edu to open a case in ServiceNow.

This computing resource is hosted with university resources. It is intended to only be used for university related buisness. Please review the Acceptable Use of CU Boulder's IT Resources page before proceeding.

Sign In

Navigate Horizon

- Choose your project (top left)
  - Generally users only have 1 project

- 4 main sections
  - Compute
  - Volumes
  - Networks
  - Orchestration

Navigate Horizon: Overview

- Land on the Overview page under "Compute"
  - quick summary of your project

Navigate Horizon: Instances

- Navigate to:
    - Project->Compute->Instances

- Instances are **virtual machines** that run inside the cloud, more simply: an instance is just a digital version of a physical computer
    - Instances can perform almost all of the same functions as a computer, including running applications and operating systems
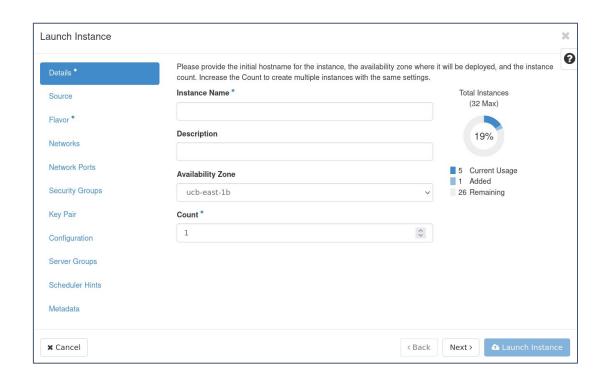
Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Instance Creation

Let's create a simple instance together

- From the instances page click on "Launch Instance"

- The Instance Creation Launcher will pop up giving us options to create our virtual machine
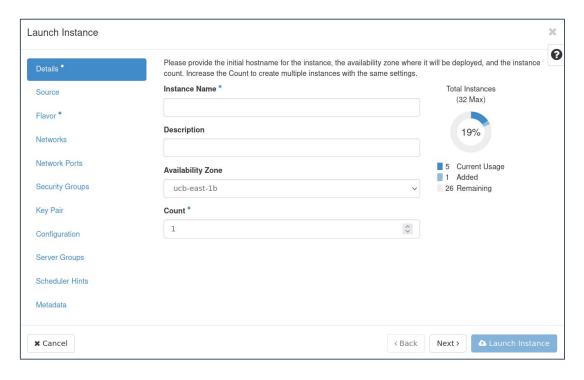
Details

- Fill out Instance details, including a name and description
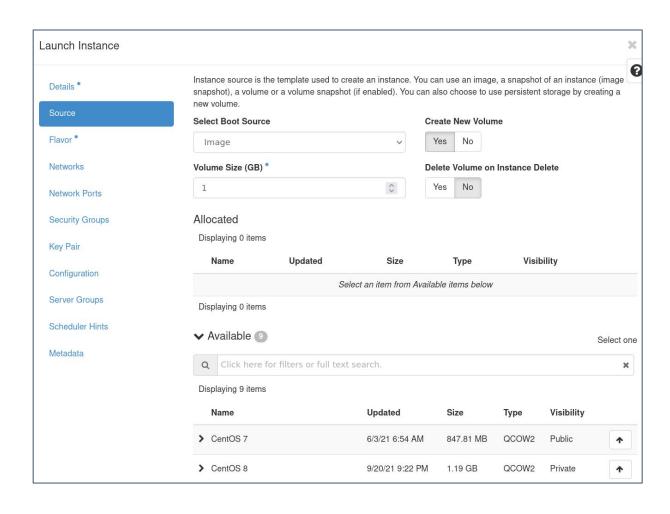 - availability zone and count can be left as defaults

Source

- Choose an image from the available list

- A virtual machine image ("image") is **a single file that contains a virtual disk that contains a bootable operating system**

- Choose to have your storage volume deleted on instance deletion
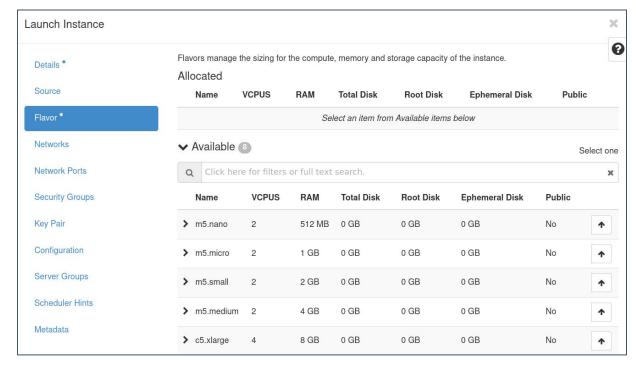 - If you select "no" be aware of "zombie" volumes
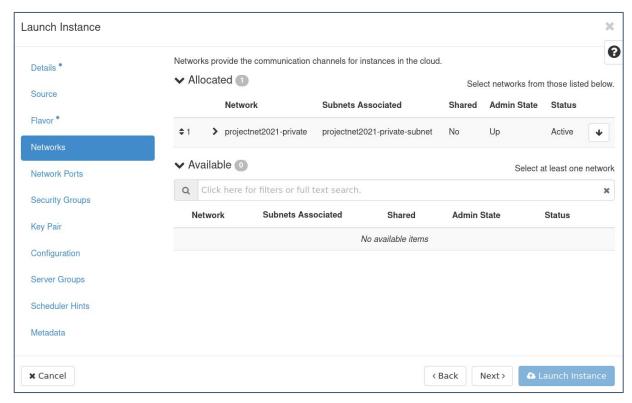
Flavor

- Choose from a list of pre-selected resources:
  - A flavor defines the compute, memory, and storage capacity of our instance.

Networks & Network Ports

- Select a project network, which determines **routability of either a public/internet or campus/internal floating IP.**
  - We'll choose an external network: **projectnet26140-private**

- Ports provide extra communication channels to your instances

- You can select ports instead of networks or a mix of both

Networks: An aside

Two network types are provided:

1. **172.18.XX.XX** - restricts your instance to CU **internal** network
2. **172.26.XX.XX** - allows certain ingress/egress from the **public** internet (for certain access, e.g http, https)

- Direct shell access (e.g. *ssh*, *powershell*) is restricted to the CU internal network due to multi-factor-authentication requirements
- This means that you will need to be on the CU network to access your instance (either VPN or through RC).

Research Computing
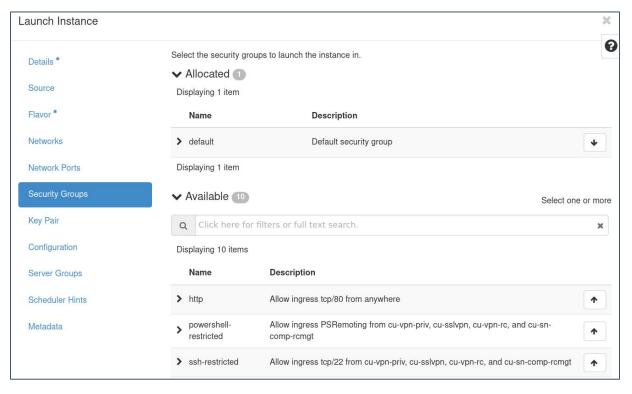UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Security Groups

- Security Groups act as a **virtual firewall** for your instance to control inbound and outbound traffic.

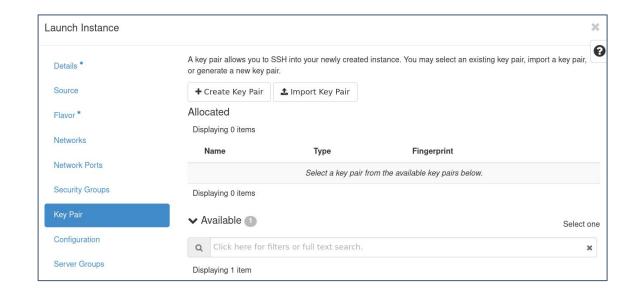- We'll choose ssh-curc, ssh-restricted, http, and https for our demo

Key Pairs

- A key pair allows you to SSH into your new instance

- You may select an existing key pair, import a key pair, or generate a new key pair

  - I find it easiest to create a keypair on my machine and import it
  - https://www.ssh.com/academy/ssh/public-key-authentication

Key Pairs: An aside

- You will need to create an SSH key *from* or copy it *to* a computer you have CU internal access to:

    - If you have an RC account already, login as follows from a terminal:

    ```
    $ ssh <username>@login.rc.colorado.edu
    # Where <username> is your identikey
    ```

    - If you don't have an RC account, log into a temporary account from a terminal:

    ```
    $ ssh user<XXXX>@tlogin1.rc.colorado.edu
    # Where user<XXXX> is your temporary username, RC will provide pw
    ```

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Key Pairs: An aside

- SSH keys are an access credential that is used in the SSH protocol

- They can be tricky to set up however, so we'll go over a simple example here. From your terminal in a local machine use the ssh-keygen command to create a new ssh keypair:

```
$ ssh-keygen -t ed25519

Generating public/private ed25519 key pair.

Enter file in which to save the key (/home/username/.ssh/id_ed25519):
```
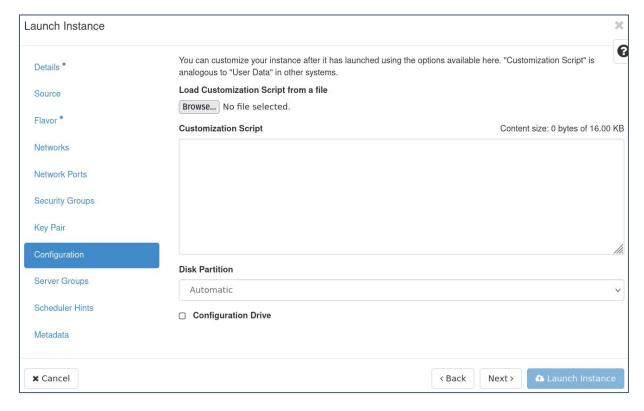
- Our new keypairs have been created at `/home/username/.ssh/` and are called `id_ed25519` and `id_ed25519.pub`. The public key (.pub) can be transferred to other remote servers (*this is the key we will import to our CUmulus instance*) but the private key (no suffix) should *never* leave the host machine.

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Config, Server Group, Scheduler Hints, and Metadata

- We'll leave these as defaults as they are extra configuration we can provide our instances, but not necessary

Launch Instance

Details *

Source

Flavor *

Networks

Network Ports

Security Groups

Key Pair

Configuration

Server Groups

Scheduler Hints

Metadata

You can customize your instance after it has launched using the options available here. "Customization Script" is analogous to "User Data" in other systems.

Load Customization Script from a file

Browse... No file selected.

Customization Script   Content size: 0 bytes of 16.00 KB

Disk Partition

Automatic

☐ **Configuration Drive**

✖ Cancel   ‹ Back   Next ›   ☁ Launch Instance

Research Computing
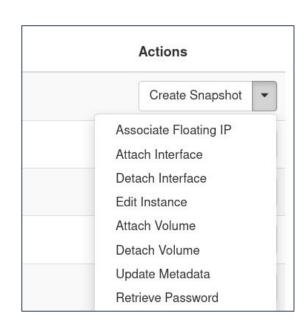UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Launch Instance and Associate IP

- Launch instance and wait for it to be set up

- In the meantime we can associate a Floating IP which will allow us to access the instance from outside of the CU network
  - On the right hand side of the newly created instance choose "Associate Floating IP" under the "actions" dropdown

Actions

Create Snapshot ▼

Associate Floating IP
Attach Interface
Detach Interface
Edit Instance
Attach Volume
Detach Volume
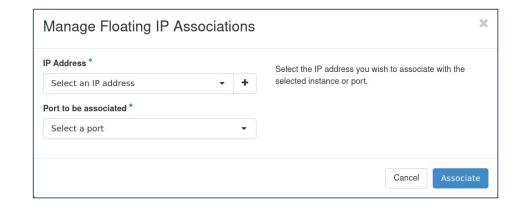Update Metadata
Retrieve Password

Associate IP

- Select from available IP addresses
  - If needed you can add a floating IP

- Select port to be associated
  - This should be pre-populated with the internal IP of your new instance

Manage Floating IP Associations

IP Address *

Select an IP address ▼ +

Port to be associated *

Select a port ▼

Select the IP address you wish to associate with the selected instance or port.

Cancel   Associate

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Logging into your Instance

Logging In

- You must be on the CU internal to connect via ssh (CURC restriction)

- Open up an ssh connection providing the identity (key) file:

```
$ ssh -i ~/.ssh/<private key> <hostname>@<external floating IP>
```

- For an ubuntu instance this will look something like:

```
$ ssh -i ~/.ssh/testkey ubuntu@123.456.789.123
```

Logged In

- Congratulations! You are now logged into your instance

- You can now:

  - Install Software

  - Administer your instance

  - Run applications and jobs

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Demo

Demo workflow: Twitter API with DB

- There are a (nearly) *infinite* number of workflows you could run on your CUmulus instance*

  As long as the workflow is related to research, personal projects are not permitted

- We'll demo a *potential* workflow: a web application which allows users to query using the Twitter API and store this data persistently to a mysql database

- This demo showcases a few important features of CUmulus not possible on HPC:
  - A persistent workflow not limited by wall clock times
  - User administration of compute resources (using root privileges for applications such as Docker)
  - Routable floating IPs available on the Public Internet

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Demos:

1. [Setting up an Instance](#)
2. [Twitter API with Database](#)
3. [CUmulus integration with CURC HPC](#)
4. [Mounting a remote filesystem from a CUmulus Virtual Machine](#)

Users that are *not* using tutorial accounts may access CUmulus and your instances for the next week to test and run through tutorials

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Thank you!

- Survey: http://tinyurl.com/curc-survey18
- Help Desk: rc-help@Colorado.edu

Research Computing
UNIVERSITY OF COLORADO BOULDER

Be Boulder.