



**Hands-on with CUmulus:** a free Research Cloud platform for RMACC

# Getting Started on CUMulus

Andrew Wilson (RC User Support)

[www.rc.colorado.edu](http://www.rc.colorado.edu)

[rc-help@colorado.edu](mailto:rc-help@colorado.edu)

# Learning Objectives

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

# Account Setup for Demo

- What's needed for today:
  - *Access-ci Login information*
    - In the format: awilson1@access-ci.org
  - *CU Research Computing account*
    - Only needed to login to Instance
    - [curc.readthedocs.io](http://curc.readthedocs.io)

# Outline

- CU Research Computing 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*

# CU Research Computing Overview

- Provides Computing and Data Storage Beyond standard desktop resources:
  - *High Performance Computing (HCP), clusters include:*
    - Alpine
    - Blanca
  - *Storage of Research Data*
    - PetaLibrary
  - *High-Speed Data Transfer*
    - Globus Endpoint
  - ***Cloud Resources***
    - **CUmulus (On-premise Cloud)**
    - Cloud Foundations (Commercial Cloud)
  - *Secure Cloud Resources*
  - *Consulting in Computational Science and Data Management*



# CU Research Computing Overview

## CU RC Menu



### RESOURCES

#### ALPINE / \$0

Large Scale HPC resource. Free of charge to use, competitive use with queue times and allocations

#### BLANCA / COST VARIES

Computing system that Researchers can buy-into for exclusive access to purchased computing systems

#### PETALIBRARY / \$45/TB/YR

Large scale research data storage. Allocations are as large as you want to pay for

#### CUMULUS / \$0

On-premises cloud-like system that enables researchers to build custom VM's that support research

#### PUBILC CLOUD / CONTACT US

We assist getting researchers set up on public cloud vendors to support their workflows

### SERVICES

#### TRAINING

Collaborate with CU Libraries to provide a variety of classes

#### ONE-ON-ONE CONSULTATIONS

Feel free to email any time to set one up

#### SECURE RESEARCH

We provide guidance and advise on compliant/ secure research

#### THE PRESERVE/CONTACT US

Cloud-based CMMC compliant research computing environment

#### VISIT US

Rooms 667-679, 3100 Marine St,  
Boulder, CO 80303

<https://www.colorado.edu/rc/>  
[rc-help@colorado.edu](mailto:rc-help@colorado.edu)

# 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*
  - *Research Hubs (Jupyterhub, RStudio Server, etc)*
- Provides users with persistent or ongoing availability by allocating logically isolated section of the cloud



# What is CUMulus?

- You get your own virtual environment for experimentation - 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

# Example Use Case: Persistent Data Mining

- Researcher wanted to use online text database to train novel search algorithm
- Needs:
  - *Persistent Querying of an online text database*
  - *Persistent storage of data*
  - *Compute ability to continuously mine the data and advance their algorithm*
- They were able to:
  - *Setup a data stream that queries data from Reddit at specified times of the day*
  - *Have the resources for persistent storage of large datasets and compute resources*
  - *Test and develop their novel search mechanism*

# Example Use Case: Galaxy Server

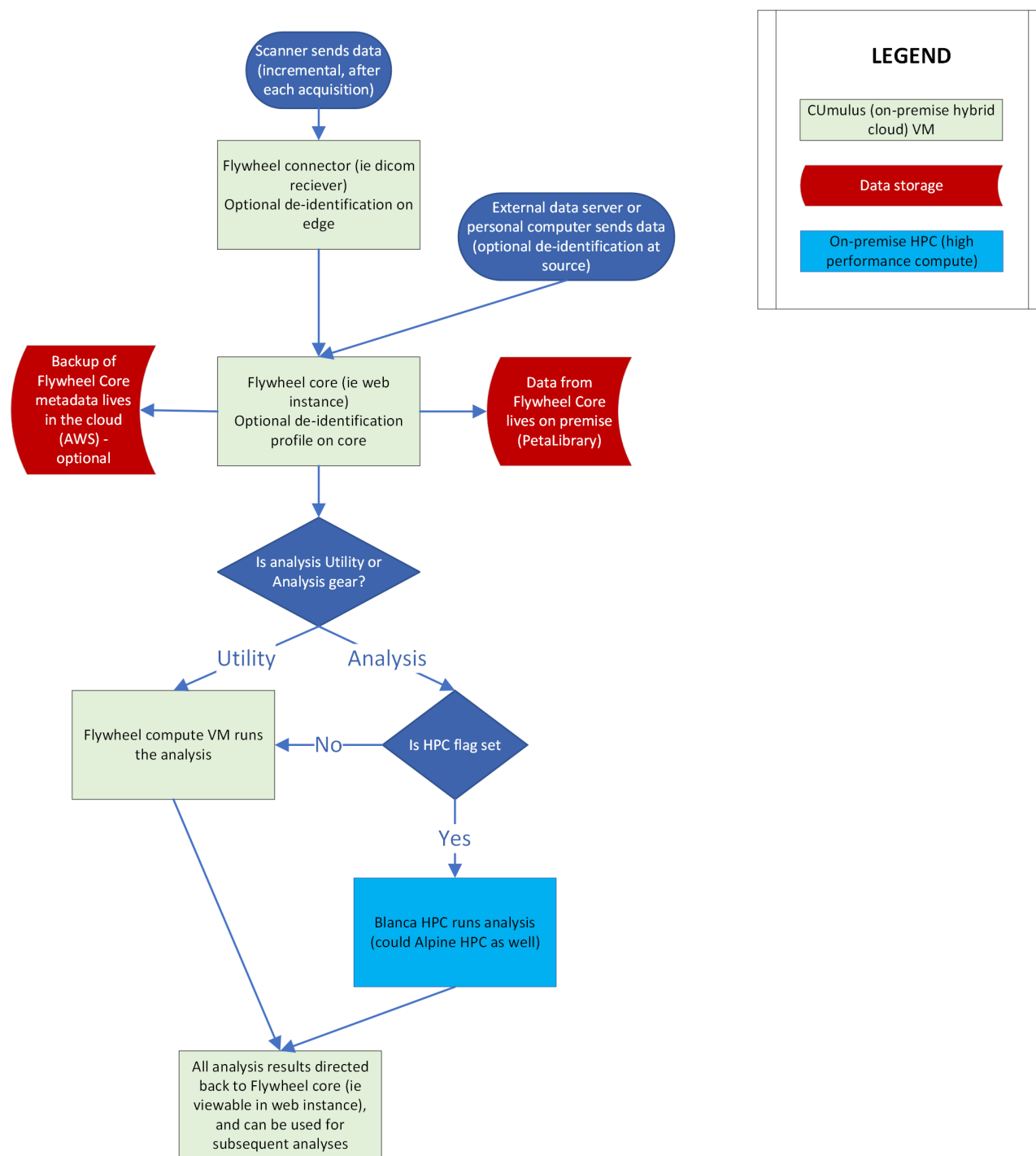
- Researcher wanted to create a custom Galaxy Server (web-based GUI for popular bioinformatics software)
- Needs:
  - *A way to host a website that the researchers can use to load and analyze data, create reports and schedule compute jobs*
  - *Persistent storage of data*
  - *The ability to create compute jobs on the virtual machine or send jobs to the Alpine Supercomputer*
  - *Run persistent jobs (multiple days long)*
- They were able to:
  - *Host the website, and control access of users who can access it*
  - *Have the resources for persistent storage of large datasets and compute resources*
  - *Create an interface to send jobs to the Alpine Supercomputer, all controlled via GUI*
  - *Test custom tools and data pipelines that they created*

# Example Use Case: Open Access Research Tool

- Researcher wanted to publish and host trainings for a custom language processing tool that helps to build language databases and tools (think auto-correct) for endangered languages
- Needs:
  - *A web-based dashboard where any user can load and use the tools the research team developed*
  - *Persistent storage of data*
  - *Compute ability to continuously mine the data and advance algorithm*
- They were able to:
  - *Setup a Jupyterhub instance where they control access of users*
  - *Have an environment where researchers around the world can easily use their research tools*
  - *Collect the incoming test data and use that for future training*
  - *Have the resources for persistent storage of large datasets and compute resources*

# Example Use Case: Flywheel Project

- Researcher wanted to create a tool that interfaces with MRI machines in the area and pulls data to a persistent web application where they can create processing pipelines and send these to the Alpine Supercomputer
- Needs:
  - A web-based dashboard where users can develop processing pipelines, move data, and create compute jobs
  - Persistent storage of data (very large datasets)
  - Compute ability to continuously process the data
- They were able to:
  - Setup a Flywheel Dashboard, a web-based application that researchers can access from anywhere.
  - Store data on the VM or in PetaLibrary or AWS
  - Send jobs to the Alpine Supercomputer for faster processing



# CUmulus Access

# CUmulus Access and Allocation

Submit a proposal for your use case (email [rc-help@colorado.edu](mailto: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://curc.readthedocs.io/en/latest/cloud/cumulus.html>



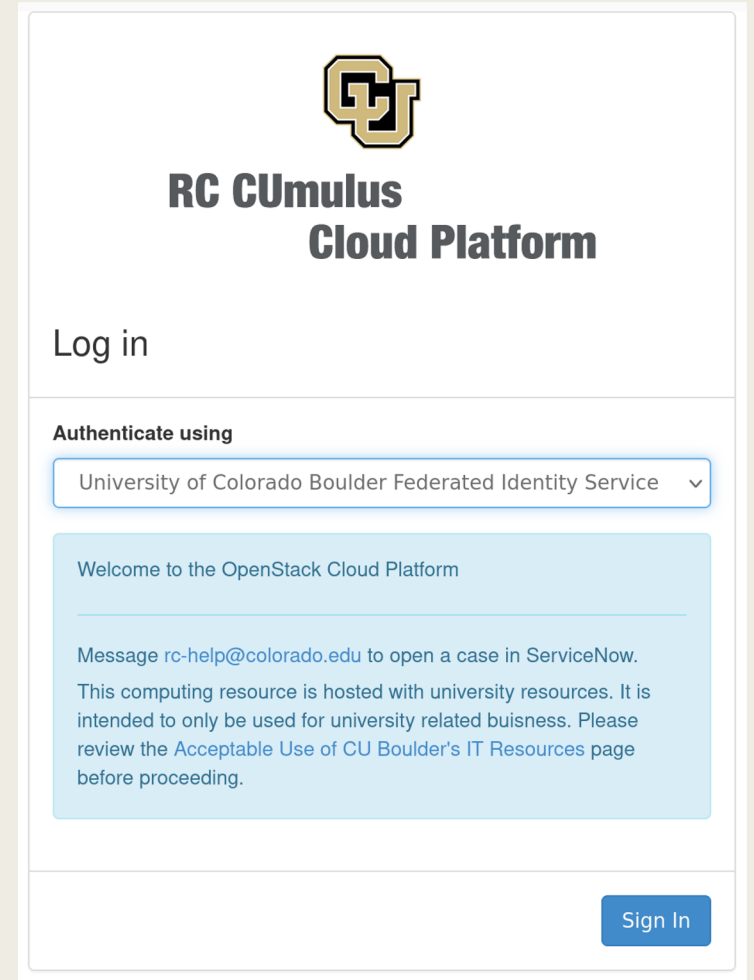
# Log in to Horizon

There are 3 current authentication/login methods at [cumulus.rc.colorado.edu/](https://cumulus.rc.colorado.edu/):


1. CU Boulder
2. CSU Fort Collins
3. ACCESS-CI (all other RMACC institutions)
  - a. Create account: <https://identity.access-ci.org/new-user>
  - b. Configure 2FA (Duo)

# Log in to Horizon

- Horizon is the CUMulus web portal
  - [cumulus.rc.colorado.edu/](https://cumulus.rc.colorado.edu/)
- Let's take a brief tour of Horizon
- Log in with your institution's credentials:
- If you're using your ACCESS-CI Account
  - Choose: 'Access-CI Identity Provider'



The screenshot shows the login interface for the RC CUMulus Cloud Platform. At the top is the CU Boulder logo and the text "RC CUMulus Cloud Platform". Below this is a "Log in" section. Under the heading "Authenticate using", there is a dropdown menu currently set to "University of Colorado Boulder Federated Identity Service". A light blue informational box contains a welcome message and instructions to contact rc-help@colorado.edu for service. At the bottom right is a blue "Sign In" button.

  
**RC CUMulus  
Cloud Platform**

Log in

Authenticate using

University of Colorado Boulder Federated Identity Service ▾

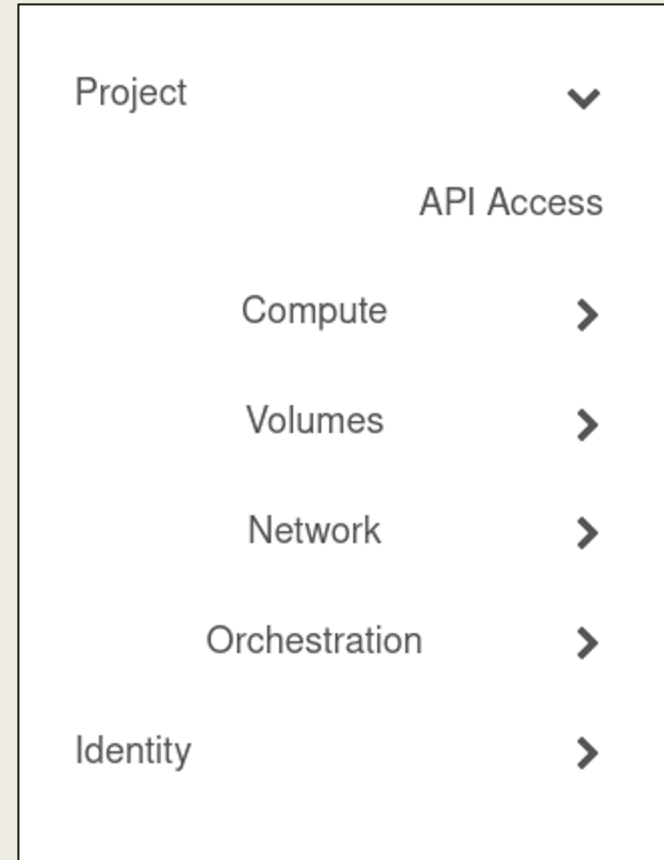
Welcome to the OpenStack Cloud Platform

Message [rc-help@colorado.edu](mailto: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 business. Please review the Acceptable Use of CU Boulder's IT Resources page before proceeding.

Sign In

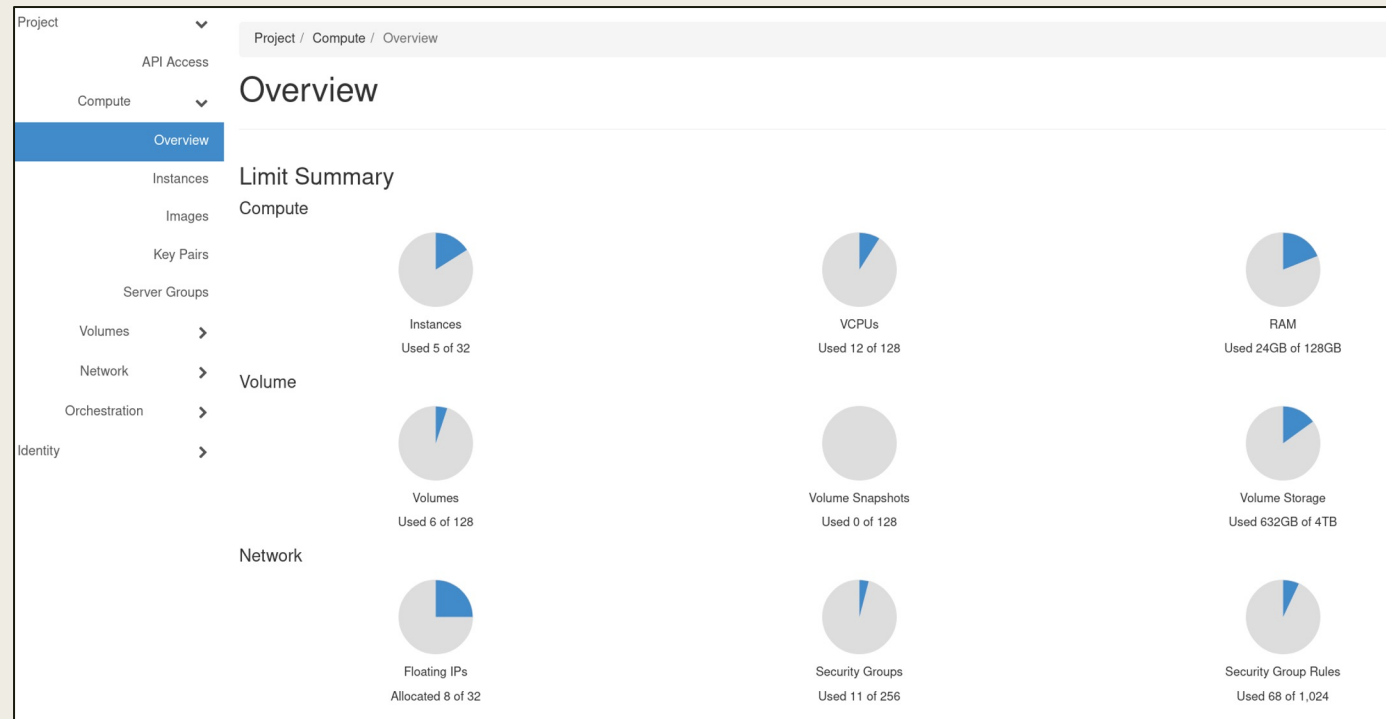
# Navigate Horizon

- 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 the same functions as a computer, including running applications and operating systems*

# 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

### Launch Instance

Details \*

Source

Flavor \*

Networks \*

Network Ports

Security Groups

Key Pair

Configuration

Server Groups

Metadata

Please provide the initial hostname for the instance, and the availability zone where it will be deployed.

Compute Availability Zones:  
ucb-spsc-g6      East Campus      SPSCv2 Datacenter

Instance hostnames greater than 15 characters will be truncated if launching a Windows instance.

Instance Name \*

Description

Availability Zone

ucb-spsc-g6

Count \*

1

Total Instances (16 Max)

38%

5

 Current Usage

1

 Added

10

 Remaining

✕ Cancel

< Back

Next >

Launch Instance



# Details

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

Launch Instance

Details \*

Source

Flavor \*

Networks

Network Ports

Security Groups

Key Pair

Configuration

Server Groups

Scheduler Hints

Metadata

Please provide the initial hostname for the instance, the availability zone where it will be deployed, and the instance count. Increase the Count to create multiple instances with the same settings.

Instance Name \*

Description

Availability Zone

ucb-east-1b

Count \*

1

Total Instances (32 Max)

19%

5

 Current Usage

1

 Added

26

 Remaining

✕ Cancel

< Back

Next >

Launch Instance

# 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*

Launch Instance

Details \*

Source

Flavor \*

Networks

Network Ports

Security Groups

Key Pair

Configuration

Server Groups

Scheduler Hints

Metadata

Instance source is the template used to create an instance. You can use an image, a snapshot of an instance (image snapshot), a volume or a volume snapshot (if enabled). You can also choose to use persistent storage by creating a new volume.

Select Boot Source

Image

Create New Volume

Yes

No

Volume Size (GB) \*

1

Delete Volume on Instance Delete

Yes

No

Allocated

Displaying 0 items

Name	Updated	Size	Type	Visibility
Select an item from Available items below				
Displaying 0 items				

▼ Available 9

Select one

Q

Click here for filters or full text search.

×

Displaying 9 items

Name	Updated	Size	Type	Visibility	
▶ CentOS 7	6/3/21 6:54 AM	847.81 MB	QCOW2	Public	↑
▶ CentOS 8	9/20/21 9:22 PM	1.19 GB	QCOW2	Private	↑

# Source

- A brief note about Source Images
  - *We have custom images for things like The Littlest JupyterHub, Rstudio Server, Galaxy Server, and LAMP servers.*
  - *Makes setup simple for these tools, with only the need to setup SSL.*
- Anything else you would find useful?

Launch Instance

Details \*

Source

Flavor \*

Networks

Network Ports

Security Groups

Key Pair

Configuration

Server Groups

Scheduler Hints

Metadata

Instance source is the template used to create an instance. You can use an image, a snapshot of an instance (image snapshot), a volume or a volume snapshot (if enabled). You can also choose to use persistent storage by creating a new volume.

Select Boot Source

Image

Create New Volume

Yes

No

Volume Size (GB) \*

1

Delete Volume on Instance Delete

Yes

No

Allocated

Displaying 0 items

Name	Updated	Size	Type	Visibility
Select an item from Available items below				
Displaying 0 items				

▼ Available 9

Select one

Q

Click here for filters or full text search.

×

Displaying 9 items

Name	Updated	Size	Type	Visibility	
▶ CentOS 7	6/3/21 6:54 AM	847.81 MB	QCOW2	Public	↑
▶ CentOS 8	9/20/21 9:22 PM	1.19 GB	QCOW2	Private	↑

# Flavor

- Choose from a list of pre-selected resources:
  - *A flavor defines the compute and memory capacity of our instance.*
- One CUMulus, we have a maximum allowed flavor of:
  - *24 vCPUs*
  - *96GB RAM*

Launch Instance

Details \*

Source

Flavor \*

Networks

Network Ports

Security Groups

Key Pair

Configuration

Server Groups

Scheduler Hints

Metadata

Flavors manage the sizing for the compute, memory and storage capacity of the instance.

Allocated

Name	VCPUS	RAM	Total Disk	Root Disk	Ephemeral Disk	Public
Select an item from Available items below						

▼ Available 8

Select one

Click here for filters or full text search.

Name	VCPUS	RAM	Total Disk	Root Disk	Ephemeral Disk	Public	
> m5.nano	2	512 MB	0 GB	0 GB	0 GB	No	↑
> m5.micro	2	1 GB	0 GB	0 GB	0 GB	No	↑
> m5.small	2	2 GB	0 GB	0 GB	0 GB	No	↑
> m5.medium	2	4 GB	0 GB	0 GB	0 GB	No	↑
> c5.xlarge	4	8 GB	0 GB	0 GB	0 GB	No	↑

# 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

Launch Instance

Details \*

Source

Flavor \*

**Networks \***

Network Ports

Security Groups

Key Pair

Configuration

Server Groups

Metadata

Networks provide the communication channels for instances in the cloud.

A floating IP must be attached after the instance is created to be reachable outside of the cloud environment.

If attaching a UCB public/internet floating IP is desired, select the network ending in **pub-vpc**.  
If attaching a UCB internal floating IP is desired, select select the network ending in **int-vpc**.

▼ Allocated

Select networks from those listed below.

Network	Subnets Associated	Shared	Admin State	Status
Select an item from Available items below				

▼ Available 2

Select at least one network

Q

Click here for filters or full text search.

×

Network	Subnets Associated	Shared	Admin State	Status	
> projectnet1840-private	projectnet1840-private-subnet	No	Up	Active	↑
> projectnet26140-private	projectnet26140-private-subnet	No	Up	Active	↑

✕ Cancel

< Back

Next >

Launch Instance

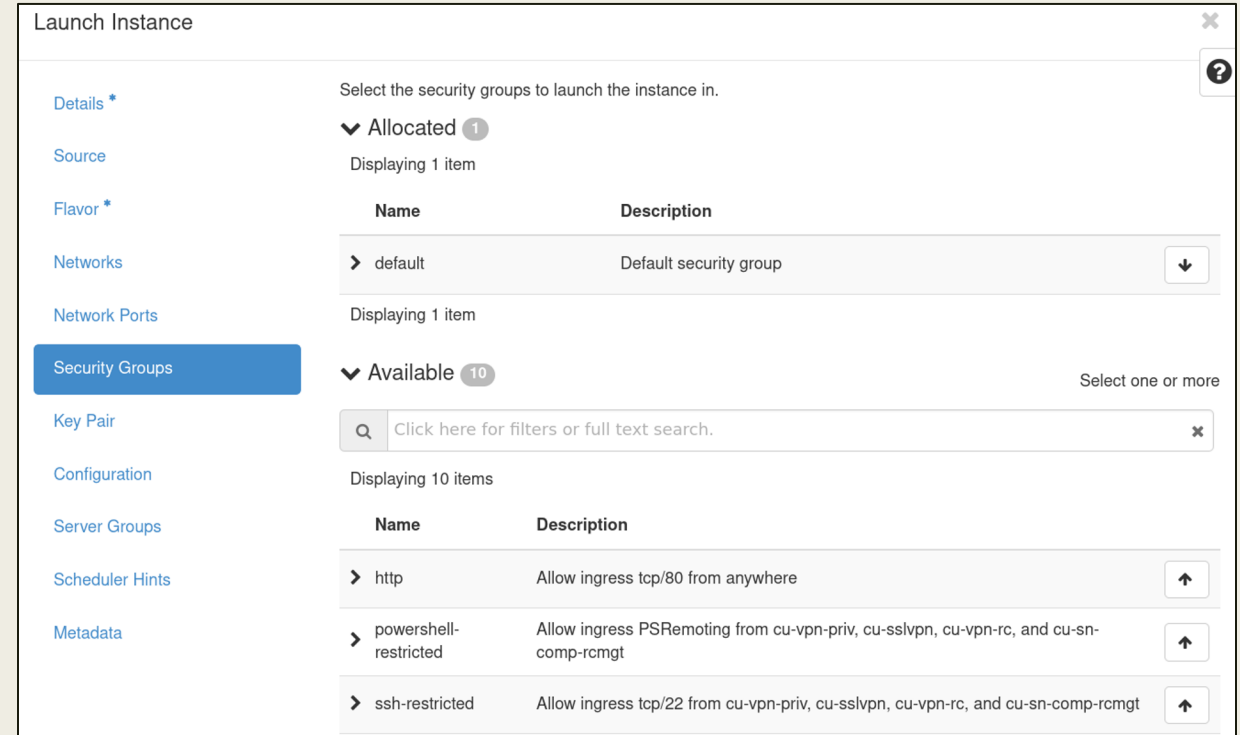
# Networks: An aside

Two network types are provided:

1. **Private Network**- restricts your instance to CU **internal** network
  2. **Public Network**- 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).

# Security Groups

- Security Groups act as a **virtual firewall** for your instance to control inbound and outbound traffic.
- Note: We can create custom security groups if there are specific ports you need opened
- We'll choose ssh-curc, ssh-restricted, http, and https for our demo



Launch Instance

Details \*  
Source  
Flavor \*  
Networks  
Network Ports  
**Security Groups**  
Key Pair  
Configuration  
Server Groups  
Scheduler Hints  
Metadata

Select the security groups to launch the instance in.

▼ Allocated <sup>1</sup>  
Displaying 1 item

Name	Description
> default	Default security group

Displaying 1 item

▼ Available <sup>10</sup> Select one or more

Click here for filters or full text search.

Displaying 10 items

Name	Description
> http	Allow ingress tcp/80 from anywhere
> powershell-restricted	Allow ingress PSRemoting from cu-vpn-priv, cu-sslvpn, cu-vpn-rc, and cu-sn-comp-rcmgt
> ssh-restricted	Allow ingress tcp/22 from cu-vpn-priv, cu-sslvpn, cu-vpn-rc, and cu-sn-comp-rcmgt



# Key Pairs

- A key pair allows you to SSH into your new instance
- You may select an existing key pair, or import a key pair find it easiest to create a keypair on my machine and import it
  - <https://www.ssh.com/academy/ssh/public-key-authentication>

Launch Instance

Details \*

Source

Flavor \*

Networks

Network Ports

Security Groups

Key Pair

Configuration

Server Groups

A key pair allows you to SSH into your newly created instance. You may select an existing key pair, import a key pair, or generate a new key pair.

+ Create Key Pair

📁 Import Key Pair

Allocated

Displaying 0 items

Name	Type	Fingerprint
Select a key pair from the available key pairs below.		

Displaying 0 items

▼ Available 1 Select one

🔍

Click here for filters or full text search.

✕

Displaying 1 item

# 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
```

# 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.

# 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**  
 No file selected.

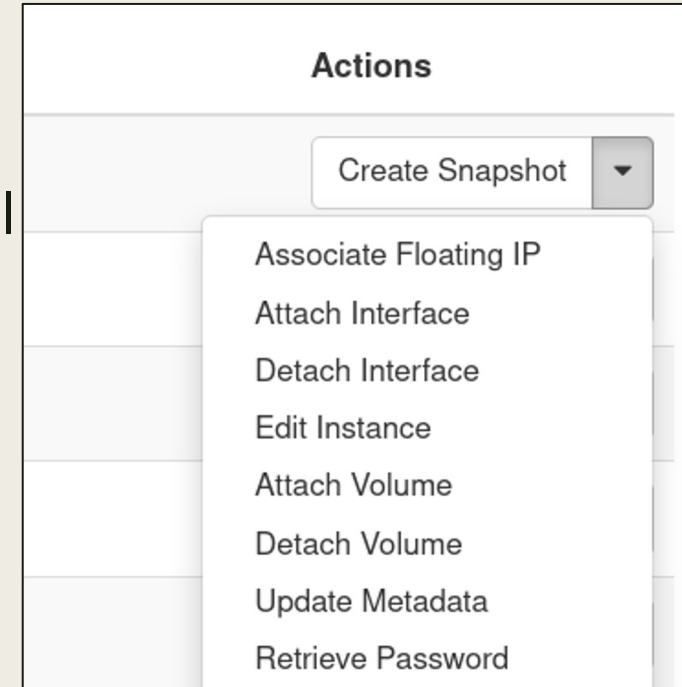
**Customization Script** Content size: 0 bytes of 16.00 KB

**Disk Partition**  
Automatic

☐ Configuration Drive

# 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*



# 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

# 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*

# Demo

# Demos:

1. [Setting up an Instance](#)
2. [CUmulus integration with CURC HPC](#)
3. [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

# A few questions for you:

1. What types of applications could you see your users building on CUMulus?
2. What demand do you have for these types of persistent services?

# Thank you!

- Contact: [anwi7603@colorado.edu](mailto:anwi7603@colorado.edu)
- Help Desk: [rc-help@colorado.edu](mailto:rc-help@colorado.edu)