Running Loci-Stream on Amazon Web Services (AWS) EC2 & Parallel Cluster

Chris Neal

Streamline Numerics, Inc.

December 1, 2023

Introduction

- The Amazon Web Services (AWS) platform is a great utility for consumers who either may not have access to large-scale parallel hardware needed to run simulations or may have intermittent usage patterns for CFD where an annual license is not desirable.

- This guide focuses on two aspects of the AWS platform: the standard Elastic Compute (EC2) framework and the Parallel Cluster framework.

  - For users of Loci-Stream, it is helpful to think of the EC2 platform as being a stand-in for computer hardware that can run small to moderately-sized problems on meshes from thousands to millions of cells.

  - The Parallel Cluster framework is useful if a user needs to run very large-scale simulations using an HPC-like job scheduler for problems ranging from tens to hundreds of millions of cells.

Getting Started with EC2 on AWS

Getting Started with EC2 on AWS

- Amazon has an Elastic Cloud tutorial that users can reference alongside the contents of this guide.

    - The following guide covers the steps to set up your Amazon EC2 account: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/get-set-up-for-amazon-ec2.html

    - This guide is focused on the specifics of using the EC2 framework: https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/EC2_GetStarted.html

- Your first step is to create an AWS account, which you can do by visiting either of the guides linked above and looking at the upper-right corner of the webpage and pressing the "Create an AWS Account" button.

    - Once you have created the account, you will typically log in using the "Root user" option from the AWS page, as you likely will be not be part of a larger organization of users.

- Once you are logged in, you will be taken to a dashboard that looks like the image shown on the next page.

STREAMLINE NUMERICS

Getting Started with EC2 on AWS

- This is the dashboard that you will see when logging into your AWS account.

Getting Started with EC2 on AWS

- From the dashboard shown on the previous page, you can access all the useful AWS pages that a user would want to view while using an EC2 instance to run Loci-Stream. At the top of the dashboard is a search bar, shown below, where you can search for different AWS elements. When looking at search results that pop up, there are little stars next to the entries. If you click a star, that entry will be saved to a shortcut bar located on your dashboard below the search bar. This makes accessing commonly used areas of the AWS dashboard easy.
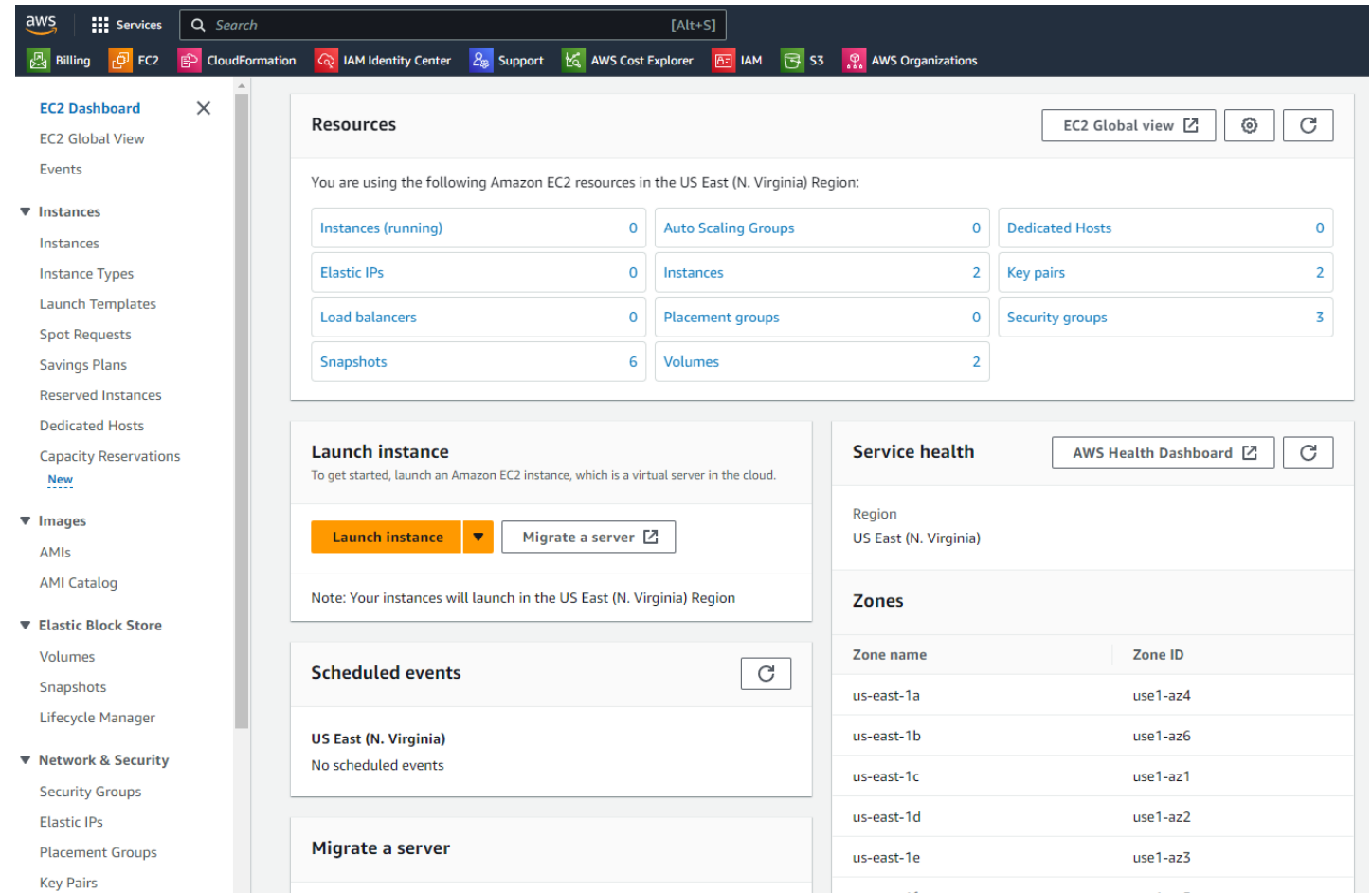

- In the search bar, search the following keywords and click the star next to the words to add them to your dashboard shortcut bar.

  - Billing

  - EC2

  - S3

  - CloudFormation (If you need to use Parallel Cluster)

- Once you have done this, go ahead and click on the "EC2" button that is on your shortcut bar.
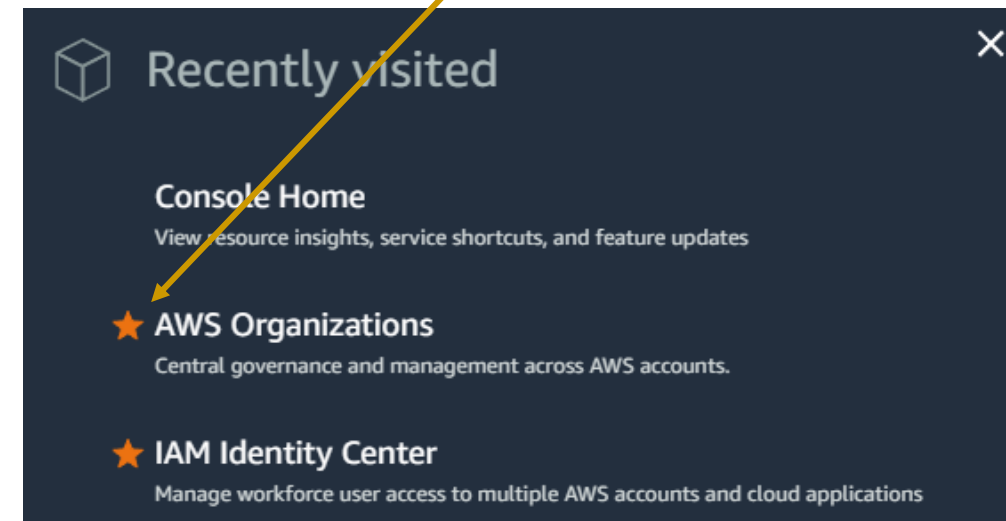
Getting Started with EC2 on AWS

- Clicking on the "EC2" button will take you to the EC2 dashboard shown on the right.

- The next step is to create a key pair. This is a file that stores a public/private key that is used for accessing instances instead of using passwords.
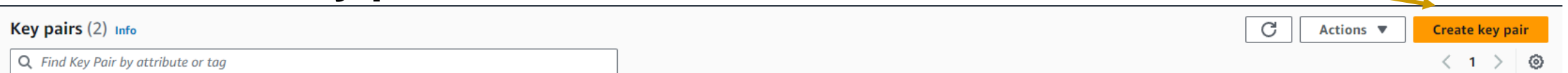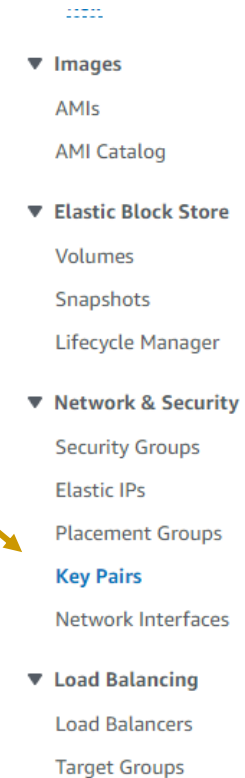
Getting Started with EC2 on AWS

- On the left side of your EC2 dashboard, scroll down to the **Network and Security** section and under that section click the **Key Pairs** option.

- The view that will open on the screen is the dashboard for generating key pairs. It will show any key pairs that you have previously generated, but it will be empty the first time you visit the page.

- On the key pairs page, at the top right, click the **Create key pair** button.

▼ **Images**
AMIs
AMI Catalog

▼ **Elastic Block Store**
Volumes
Snapshots
Lifecycle Manager

▼ **Network & Security**
Security Groups
Elastic IPs
Placement Groups
Key Pairs
Network Interfaces

▼ **Load Balancing**
Load Balancers
Target Groups

Key pairs (2) Info

Q *Find Key Pair by attribute or tag*

Actions ▼   **Create key pair**

‹ 1 ›

Getting Started with EC2 on AWS

- In the window that opens, give your key pair a name. Something like my_key or any name that is unique. Giving a numerical suffix helps if you need to generate multiple keys, like my_key_1, my_key_2, etc.

- The **RSA** option is the default for the **Key pair type**, but either option can be used. If you plan to connect to your instances using a terminal via SSH, select the **.pem** option for the **Private key file format**. Press **Create key pair** when finished.

- Note: You only need a single key pair for connecting to a Loci-Stream instance using EC2.

EC2 > Key pairs > Create key pair

Create key pair Info

Key pair
A key pair, consisting of a private key and a public key, is a set of security credentials that you use to prove your identity when connecting to an instance.

Name

Enter key pair name

The name can include up to 255 ASCII characters. It can't include leading or trailing spaces.

Key pair type Info

○ RSA   ○ ED25519

Private key file format
○ .pem
For use with OpenSSH
○ .ppk
For use with PuTTY

Tags - optional
No tags associated with the resource.

Add new tag

You can add up to 50 more tags.
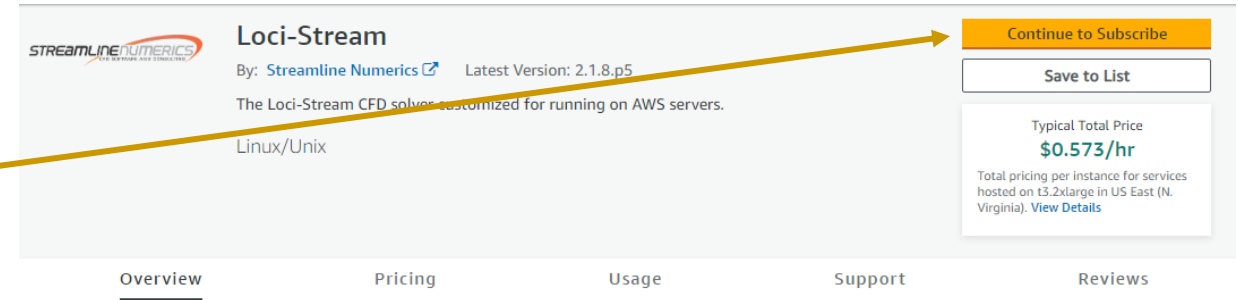
Cancel   Create key pair

Getting Started with EC2 on AWS

- Once the key pair has been generated, your web browser will auto-download a file with the same name as your key pair, such as my_key_1.pem. This is your private key file and you will use it to connect to instances. It serves as your password that authorizes you to connect to the instance. Save this file somewhere safe where it will not be accidentally deleted. If you lose the file, you can always regenerate a new key pair file and use that newer file to connect to instances.

Launching an Instance From the Marketplace Storefront

Launching an Instance (Marketplace)

- From our AWS Marketplace storefront, you will want to click the yellow Continue to Subscribe button on the upper right corner of the store page.

- You will be taken to a page where your request to subscribe to the Loci-Stream AMI will be process. To continue, you must click the **Accept Terms** button in the Terms and Conditions section. (This may take 1 or 2 minutes)

Launching an Instance (Marketplace)

- Once your request to subscribe has been process, the **Continue to Configuration** button at the top right of the page will turn yellow and can be pressed.

- Pressing the button will take you to a page where you will select the version of Loci-Stream that you wish to use. When you are done, click the **Continue to Launch** button to be routed to a simplified version of EC2 instance launch dashboard. See the following section for details about each option.
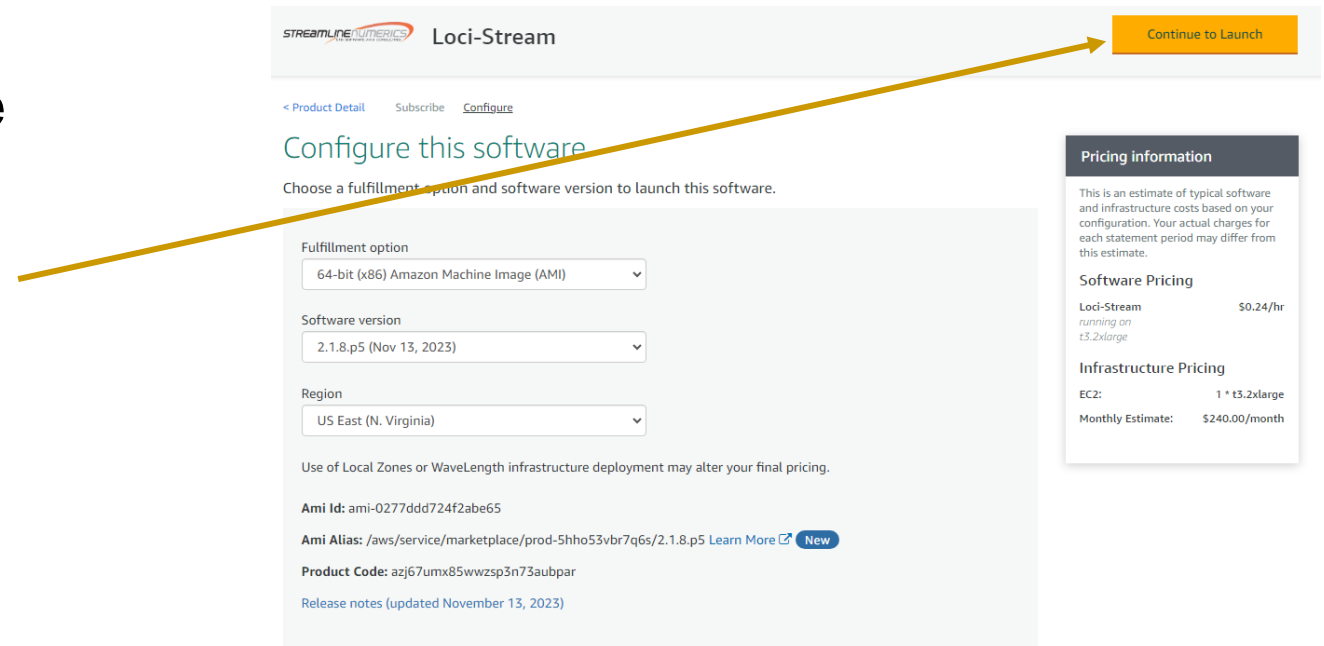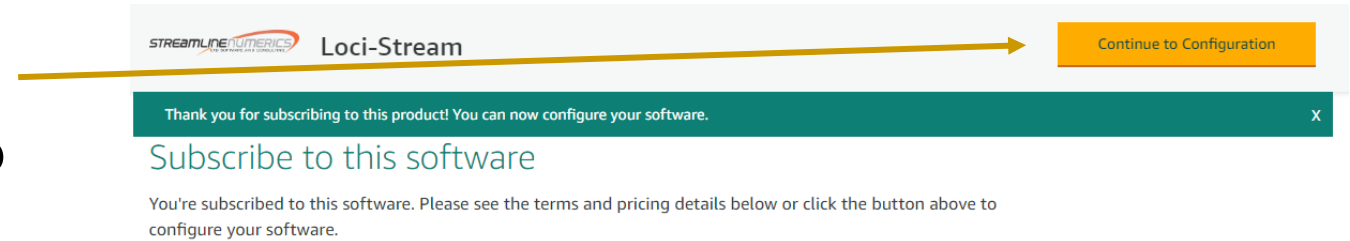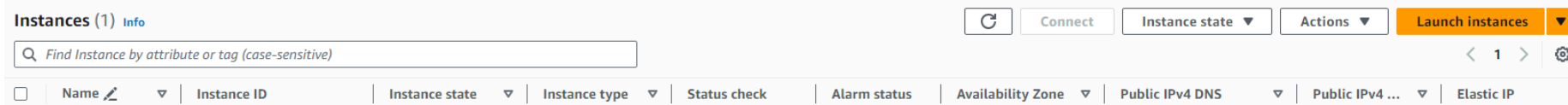
Launching an Instance from an AMI via EC2 Dashboard

Launching an Instance

- Once you have obtained the Loci-Stream AMI from our storefront, you can launch an instance using the **Instances** option on the left side of the **EC2** dashboard.

- Clicking the **Instances** button will open up the instances dashboard view, like the one shown below. From this view you can click the **Launch instances** button to begin the process of creating an instance to run Loci-Stream on.

EC2 Dashboard
EC2 Global View
Events

▼ Instances
  Instances
  Instance Types
  Launch Templates
  Spot Requests
  Savings Plans
  Reserved Instances
  Dedicated Hosts
  Capacity Reservations
  New

▼ Images
  AMIs
  AMI Catalog

▼ Elastic Block Store
  Volumes
  Snapshots
  Lifecycle Manager

Instances (1) Info | C | Connect | Instance state ▼ | Actions ▼ | **Launch instances** | ▼

Q Find Instance by attribute or tag (case-sensitive) | ‹ 1 › ⚙

| ☐ | Name ✎ ▽ | Instance ID | Instance state ▽ | Instance type ▽ | Status check | Alarm status | Availability Zone ▽ | Public IPv4 DNS ▽ | Public IPv4 ... ▽ | Elastic IP |

Launching an Instance
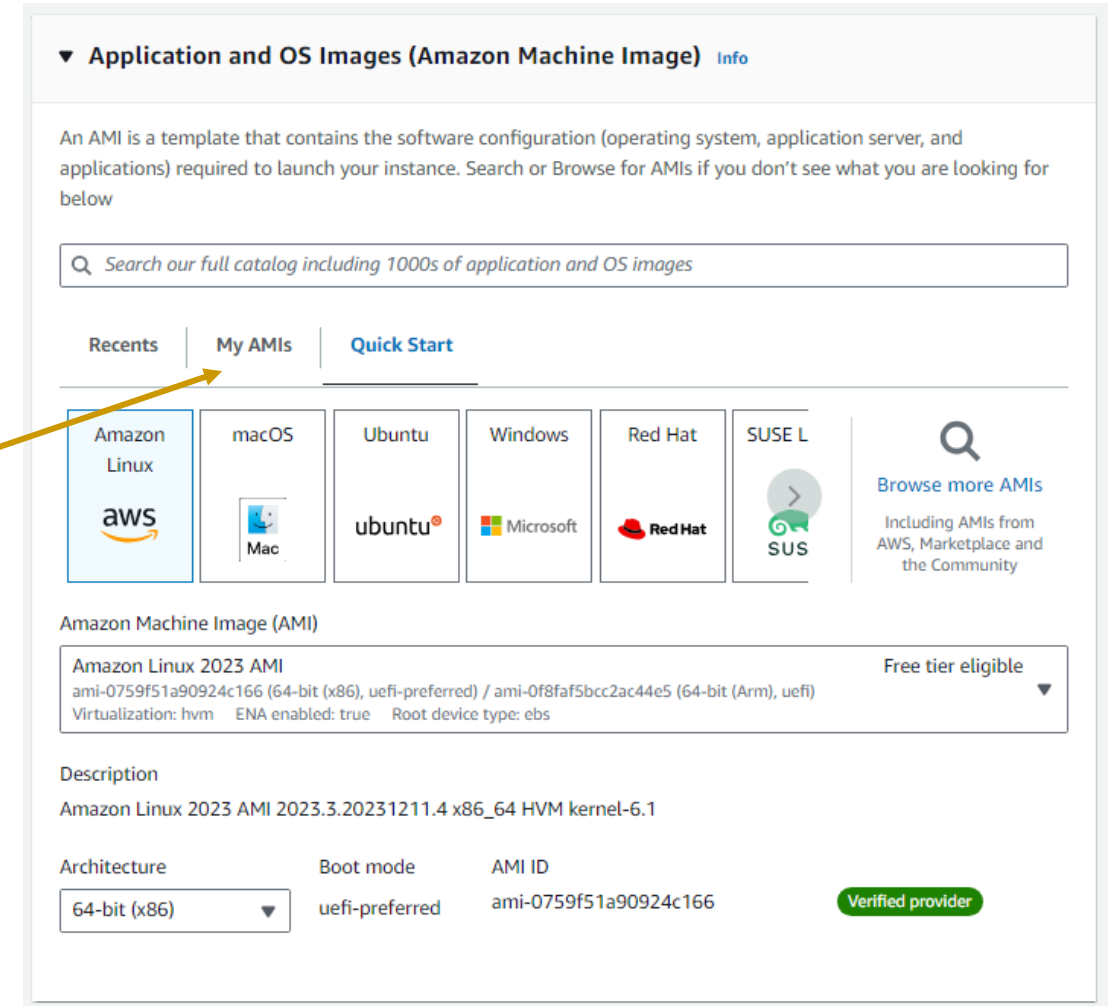
- One you have clicked the **Launch instance** button, you will be taken to a new screen. This is the screen where you specify the AMI, hardware, key-pair, security group, and disk volume size.

- The first entry box is the **Name and tags**, which is where you can give the instance any name that you want. Typically, we name instances *locistream_<version>*, so for example for the 2.1.8-p5 version, the instances could be named *locistream_218p5.*

Launching an Instance

- The next entry box is the **Application and OS Images (Amazon Machine Image)**. This is where you specify which AMI to load onto the instance.

- There is a quick-select section under **My AMIs** in that section that will show any AMIs that you have. From here you would select the Loci-Stream AMI that you were granted access to use.

Launching an Instance

- The next entry box is the **Instance type.** This is a drop-down menu box where you select what type of hardware you want to run the AMI on. There are free options and paid options. The pricing for the hardware is shown for whichever option you select.
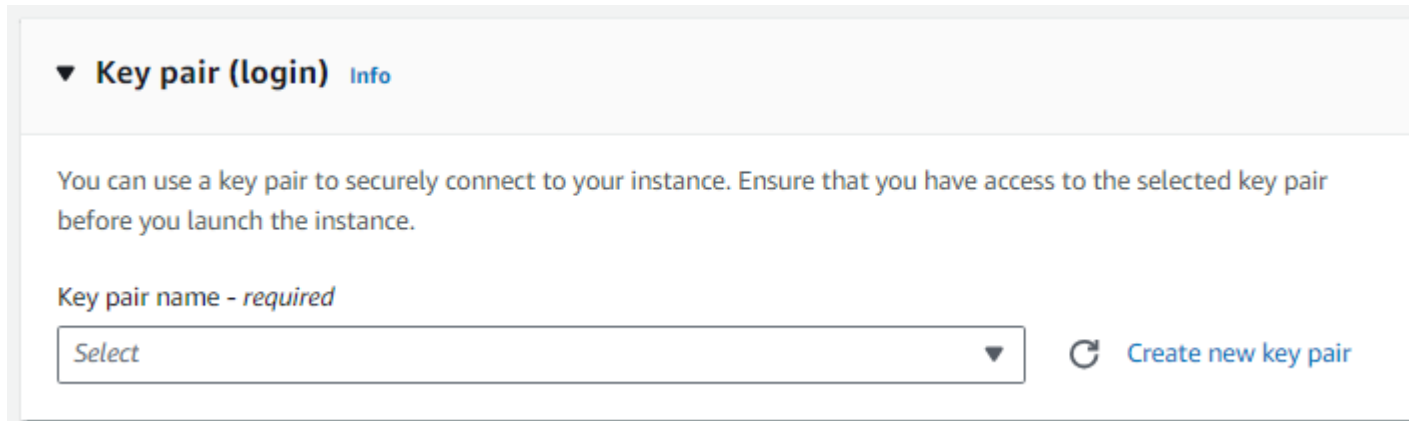
Launching an Instance

- The next entry box is the **Key pair (login)**. This is a drop-down menu where you select which key pair that you have generated will be used for logging into the instance once it is running. If you followed the AWS setup instructions earlier in the guide, your key pair should show up as an option in the drop-down menu.
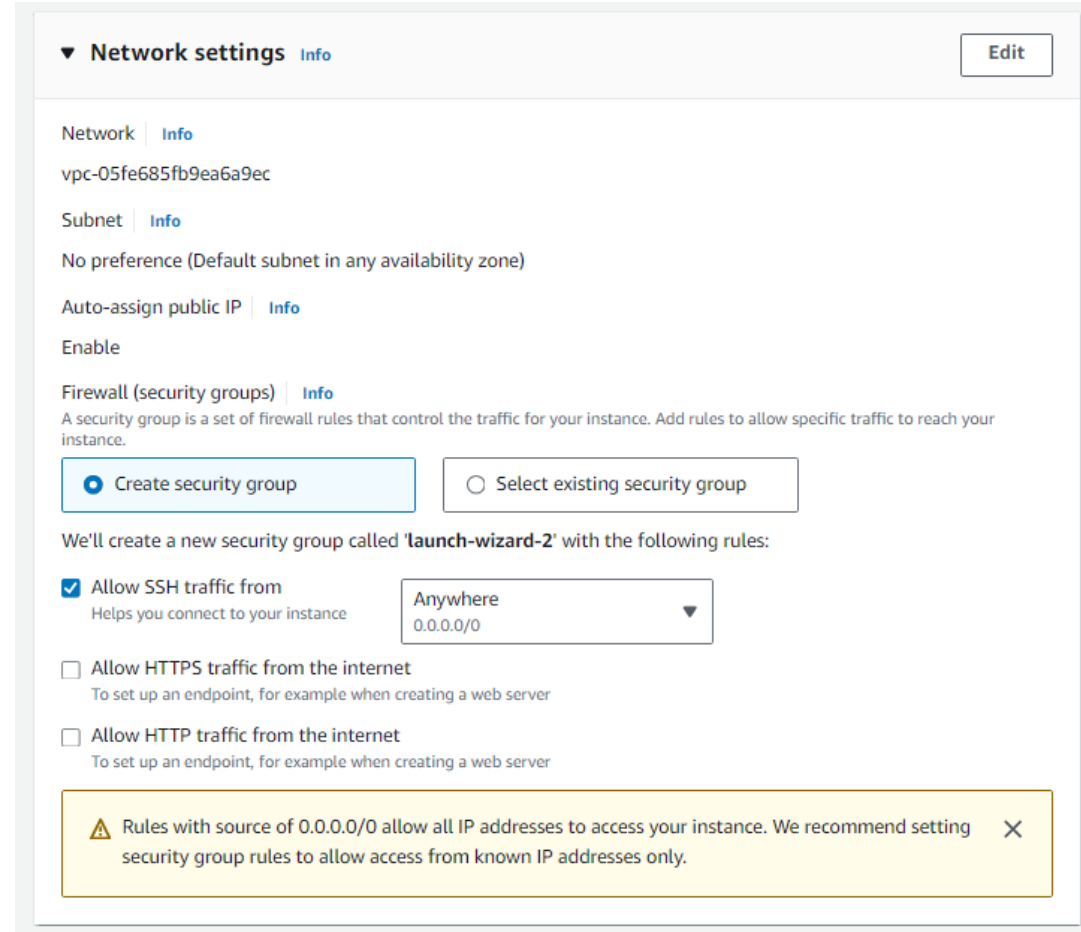
Launching an Instance
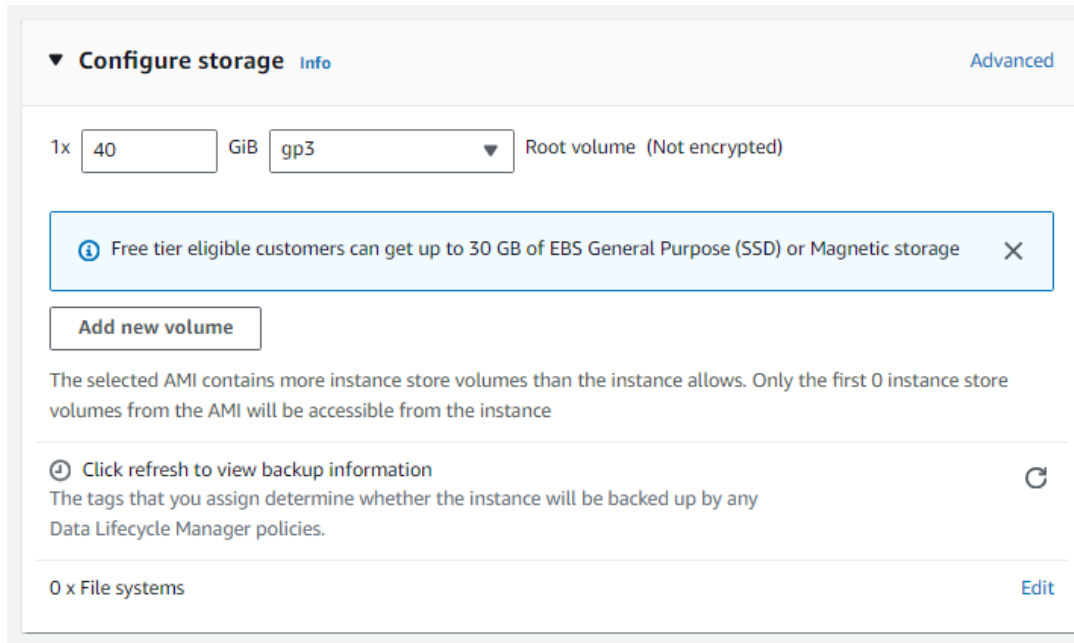
- The next entry box is the **Network settings**, which sets up the network permissions that are assigned to the instance. In the context of the Loci-Stream AMI, these permissions control which IP addresses can SSH into an instance.

- If this is your first instance ever, you can select the **Create security group**. Subsequent instances can then use the same security group.

- Make sure the **Allow SSH traffic from** box is checked and in the drop-down menu, select **My IP** to allow your machine to connect to the instance.

- If you have a security group already created, then look at the **Firewall (security groups)** section and check the **Select existing security group** option. From the dropdown you can select a security group that you have created.

Launching an Instance
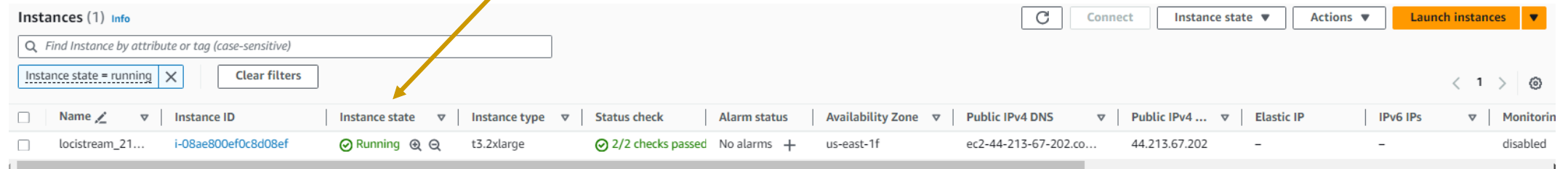
- The final required entry box is **Configure storage.** This is where you select the size and type of the storage volume that will be attached to your instance. You select the number of gigabytes that will be needed and the type of storage using the dropdown.

Launching an Instance

- Once you have filled out these options, press the **Launch instance** button on the right side of the screen to create your instance. If this is successful, you will now be able to see your running instance from the EC2 dashboard by clicking the **Instances** button on the left panel.

Connecting to an Instance

Connecting to an Instance

- One your instance is running, you will be able to see it in your EC2 dashboard by selecting the **Instances** option on the left sidebar. Clicking anywhere in the row that displays the information for your running instance, or selecting the left-most check box will select an instance.

- Select your running instance and then click on the **Actions** dropdown menu and select **Connect**. You will be taken to a new window with details about different ways to connect to the instance.
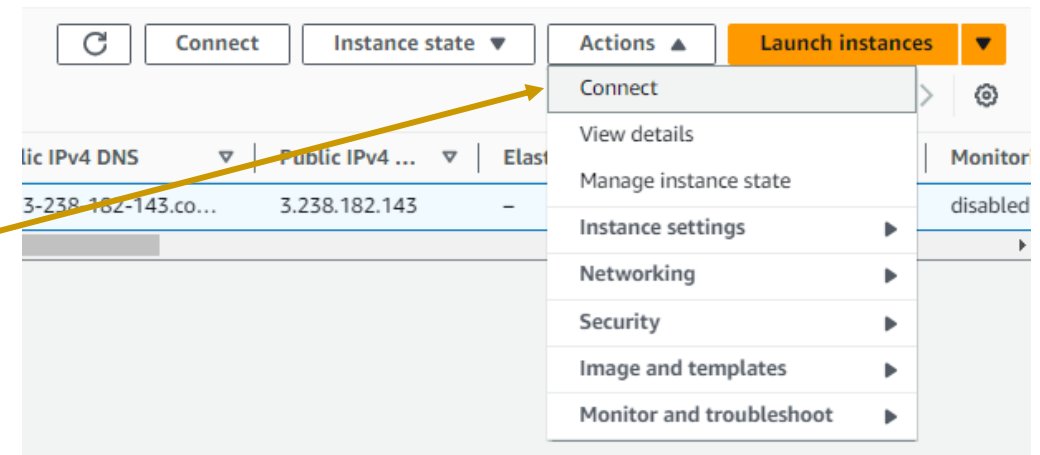
Connecting to an Instance

- Select the SSH client tab to view the details about connecting to the instance using a terminal and SSH.

- Numbers 1 and 2 in the list are important steps, and should only have to be done one time during your very first connection using the key pair file.

- For our Ubuntu-based AMI, the connection command that you will use will look like the following:

Connect to instance Info
Connect to your instance i-08ae800ef0c8d08ef (locistream_218p5_dev) using any of these options

EC2 Instance Connect | Session Manager | **SSH client** | EC2 serial console

Instance ID
i-08ae800ef0c8d08ef (locistream_218p5_dev)

1. Open an SSH client.
2. Locate your private key file. The key used to launch this instance is sni_key_1.pem
3. Run this command, if necessary, to ensure your key is not publicly viewable.
   chmod 400 sni_key_1.pem
4. Connect to your instance using its Public DNS:
   ec2-44-213-67-202.compute-1.amazonaws.com

Example:
ssh -i "sni_key_1.pem" root@ec2-44-213-67-202.compute-1.amazonaws.com

ⓘ **Note:** In most cases, the guessed user name is correct. However, read your AMI usage instructions to check if the AMI owner has changed the default AMI user name.

Cancel

ssh –i <path_to_keyPair_file> ubuntu@<contents_of_bullet_4>

Connecting to an Instance

- Upon using the SSH connection command you will be asked the usual question about adding the instance to the known hosts file, and you can hit the **y** key to accept.

- You should then be able to view the contents of the Loci-Stream AMI.

- The binary for the Loci-Stream executable is already in the system path to make execution easy. You can move case run files to the instance now and run. To learn more about how to run Loci-Stream, see our Users' Guide.

- The following section discusses how to move data to and from the instance that is running.

Moving Data To and From an Instance

Moving Data to an Instance

- Moving data such as required files that are needed for Loci-Stream to execute (vars file, vog file, etc.) requires you to use the SCP utility. The command to transfer data up into an instance that is running is very similar to how connections via SSH are made. Assuming that you are on a local machine and need to transfer a file named **myFiles.tar** up into an instance, the following command should be used:

  scp –i <path_to_keyPair_file> myFiles.tar ubuntu@<contents_of_bullet_4>:/home/ubuntu/

- The contents of bullet 4 are referring to the numbered bullets discussed previously, and the content is the specific DNS address of the instance. Often times it looks like: ec2-44-213-67-202.compute-1.amazonaws.com. You will have to go to **EC2->Instances->Actions->Connect->SSH client** to view the specific Public DNS to use. Each time you start up your instance, it will receive a new public DNS.

Moving Data from an Instance

- Once you have finished running Loci-Stream and you have all the output files that you wish to post-process, the next step is to tar all of those files up. Once that is done, you can retrieve them from the instance by logging onto your local machine and using the SCP command shown below.

scp –i <path_to_keyPair_file> ubuntu@<contents_of_bullet_4>:/home/ubuntu/<youRunDirectory>/OutputFiles.tar .

Notes on Billing

Notes on Billing

- One very important thing to always keep in mind is that Amazon charges for any instance that is running. You will accrue the software and hardware hourly charge if the instance is running. So it is very important that you shutdown an instance once you have finished using it. This is analogous to pressing the power button on a local machine to shut it down. The instance still exists, but it is not running on any AWS hardware, and thus you are not charged.
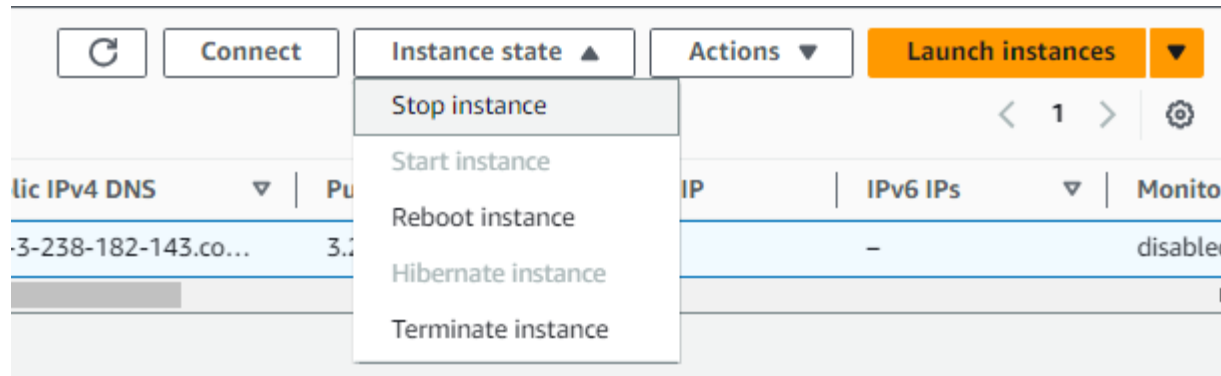
Notes on Billing

- When you allocate disk space for using with your instances, you will be charged for the size of the volume that you allocate. This charge is a monthly charge and is based on the number of gigabytes used per month. So if you start up a Loci-Stream instance and decide that your simulation needs dictate that a disk size of 200GB is necessary to store the outputs, you will receive a monthly charge for this 200GB that is allocate.

- Thee volume(hard drive) fee, unlike the hardware fee that is charged when an instance is actively running, is charged regardless of the state of the instance. This is simply because that hard disk space is preserved for the entire life of the instance, until the instance is terminated. When an instance is terminated, the disk volume that is attached to the instance is also deleted and you are no longer charged for that volume.

Using Parallel Cluster

Using Parallel Cluster

- The AWS team has created a utility called Parallel Cluster that encapsulates multiple features of the AWS platform and automates the creation of a compute cluster that users can then log into and submit jobs to.

- The Parallel Cluster essentially creates a typical high performance computing(HPC) environment with several computer nodes linked together with a network fabric that allows for cross-node parallelism.

- Jobs (simulations) are submitted to the created cluster using the SLURM scheduler. Users log into a "head node" or "login node" and have their storage volumes attached to this node, and from this node they submit their simulation to the cluster via standard SLURM commands.

- Simulations that are running on the cluster can be seen via standard SLURM commands.

Using Parallel Cluster

- The detailed instructions for getting started with using the AWS Parallel Cluster command line can be found below for more experienced users:

  - https://docs.aws.amazon.com/parallelcluster/latest/ug/install-v3.html

  - You want to use the ed25519 for the key-pair type, as the head login node created by the parallel cluster utility will not allow SSH connections from an RSA key-pair (RSA is a weaker type of encryption).

- 1.) First create a single key-pair. This is used for authenticating your account so that you can 'log in' using the command line interface and issue commands from the terminal to create your cluster to run Loci-Stream on.

  - https://docs.aws.amazon.com/AWSEC2/latest/UserGuide/get-set-up-for-amazon-ec2.html#create-a-key-pair

  - Save this file somewhere secure and not easily accessible to anyone but yourself.

Installing the Parallel Cluster Command Line Interface

- **Install the AWS Parallel Cluster command line interface (CLI)**
    - For this you need to have a version of Python that is 3.6 or greater

- **To install the Parallel Cluster CLI in a Python virtual environment (recommended) follow the following instructions.**

    - https://docs.aws.amazon.com/parallelcluster/latest/ug/install-v3-virtual-environment.html

    - The name of the virtual environment (apc-ve) isn't required, and you can name it, and it can be located wherever you want it (~/ isn't required)

Installing the AWS Command Line Interface

- Prior to installing the Parallel Cluster command line interface, you need to install the Amazon Web Services (AWS) command line interface. The detailed instructions are here:

  ❑ https://docs.aws.amazon.com/cli/latest/userguide/getting-started-install.html

- Once the AWS CLI is installed and you have generated the key-pair from earlier, you can run the aws configure command and provide the information from the key-pair file that is downloaded when you create the key-pair. Instructions are at the top of the following webpage:

  ❑ https://docs.aws.amazon.com/parallelcluster/latest/ug/install-v3-configuring.html

  ❑ The red values are specific to your information in your key-pair file and your particular region.

Configuring a Cluster

- The Parallel Cluster command line interface allows users to configure a cluster themselves, which then saves the configuration data into a YAML formatted file that can be saved and used later to create clusters based on that file. Instructions for manually creating a cluster can be found below, but for using Loci-Stream we already have a YAML cluster file that can be used as a template for new users.

  - https://docs.aws.amazon.com/parallelcluster/latest/ug/install-v3-configuring.html

  - The cluster can then be launched using:

    - pcluster create-cluster --cluster-name test-cluster --cluster-configuration cluster-config.yaml

Configuring a Cluster

- It takes about 15 minutes for the cluster to be setup, after which you can connect to the cluster using:

  - pcluster ssh --cluster-name test-cluster --i ~/path/to/keyfile.pem

- When you are done with the cluster you can delete the resources using:

  - pcluster   delete-cluster --region us-east-1 --cluster-name test-cluster

Configuring a Cluster

- A cluster is composed of machines and network hardware, when deleting a cluster, that only moved the machines, but the network infrastructure is left as a "stack". This stack must also be deleted, because you will be charged for a stack even if it is idle.

- Find the stack name using:

  - aws --region us-east-1 cloudformation list-stacks \
    --stack-status-filter "CREATE_COMPLETE" \
    --query "StackSummaries[].StackName" | \
    grep -e "parallelclusternetworking-"

  - You will get a name like: "parallelclusternetworking-pubpriv-20191029205804"

  - Then use aws –region us-east-1 cloudformation delete-stack –stack-name parallelclusternetworking-pubpriv-20191029205804