



Canadian Bioinformatics Workshops

www.bioinformatics.ca

bioinformaticsdotca.github.io

Creative Commons

This page is available in the following languages:

Afrikaans Български Català Dansk Deutsch Ελληνικά English (CA) English (US) Esperanto
Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE)
Euskara Suomiksi français français (CA) Galego മലബാറി ഹാവാസി Magyar Italiano 日本語 한국어 Macedonian Melayu
Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски (ћирилица) Sotho svenska
中文 華語 (台灣) isiZulu



Attribution-Share Alike 2.5 Canada

You are free:



to Share — to copy, distribute and transmit the work



to Remix — to adapt the work



Under the following conditions:



Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).



Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

Disclaimer

Your fair dealing and other rights are in no way affected by the above.

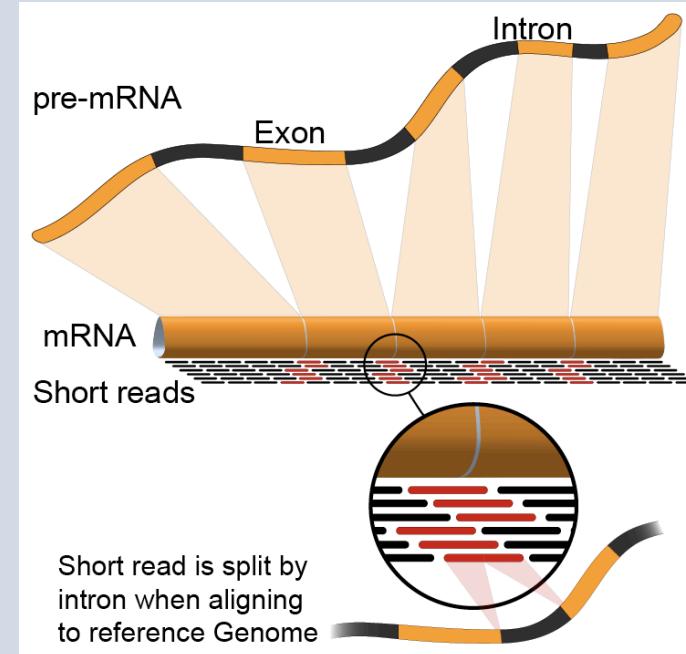
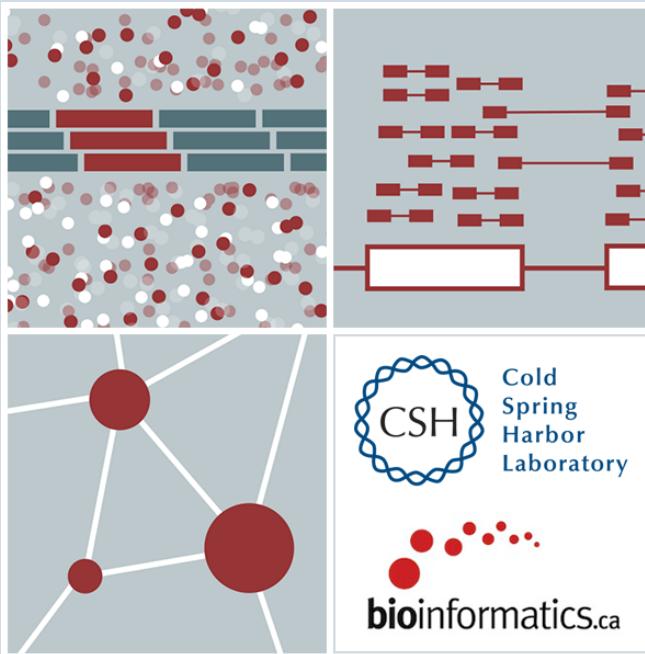
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:
[English](#) [French](#)

[Learn how to distribute your work using this licence](#)

Introduction to cloud computing



Obi Griffith and Malachi Griffith
RNA-seq Analysis 2023. July 17-19, 2023

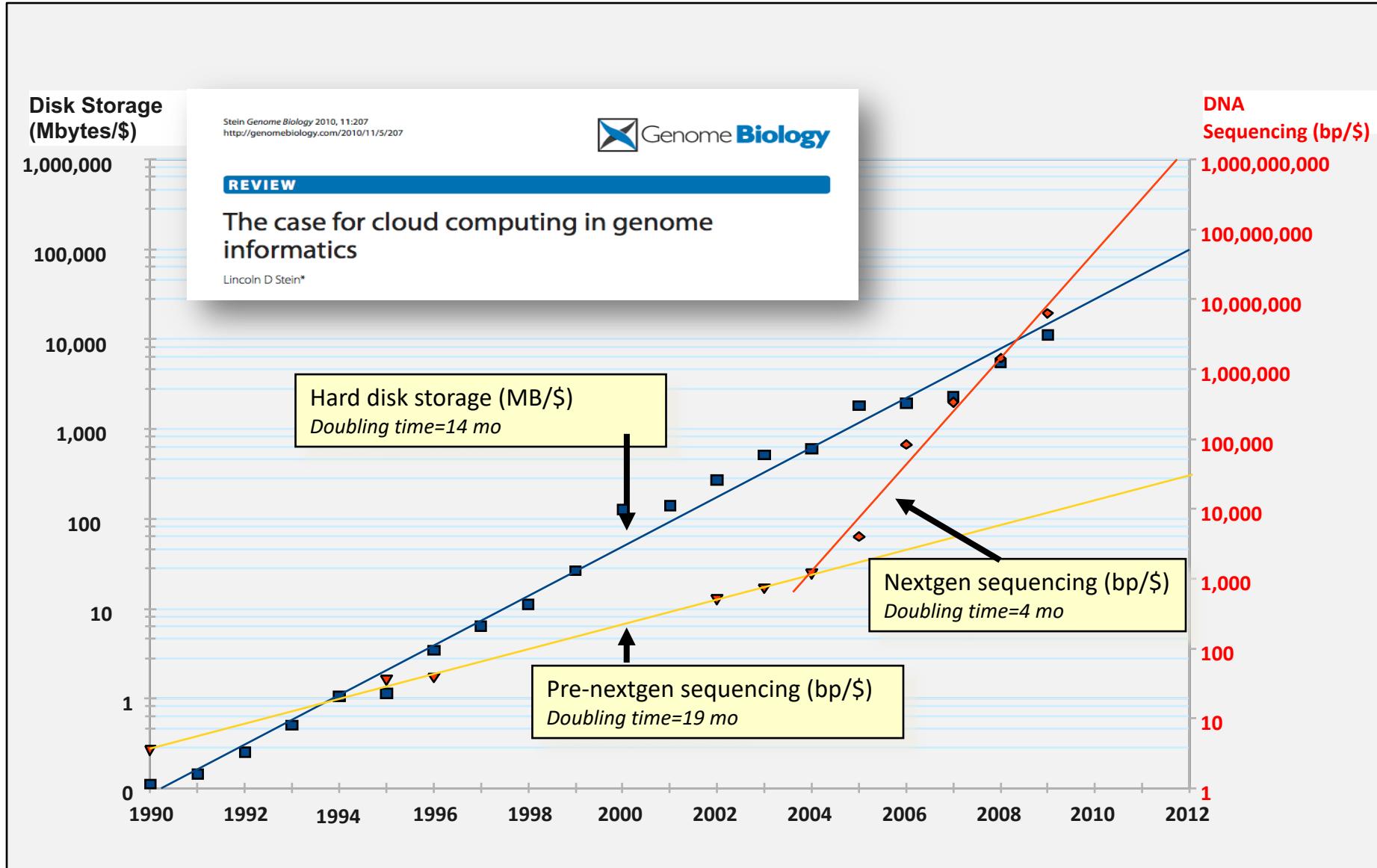


Washington University in St. Louis
SCHOOL OF MEDICINE

Learning Objectives

- Introduction to cloud computing concepts
- Introduction to cloud computing providers
- Use the Amazon EC2 console to create an instance for each student
 - Will be used for many hands-on tutorials throughout the course
- How to log into your cloud instance

Disk Capacity vs Sequencing Capacity, 1990-2012



About DNA and computers

- We hit the \$1000 genome* in ~2016
 - Need to think about the \$100 genome
- The doubling time of sequencing has been ~5-6 months.
- The doubling time of storage and network bandwidth is ~12 months.
- The doubling time of CPU speed is ~18 months.
- The cost of sequencing a base pair will eventually equal the cost of storing a base pair

What is the general biomedical scientist to do?

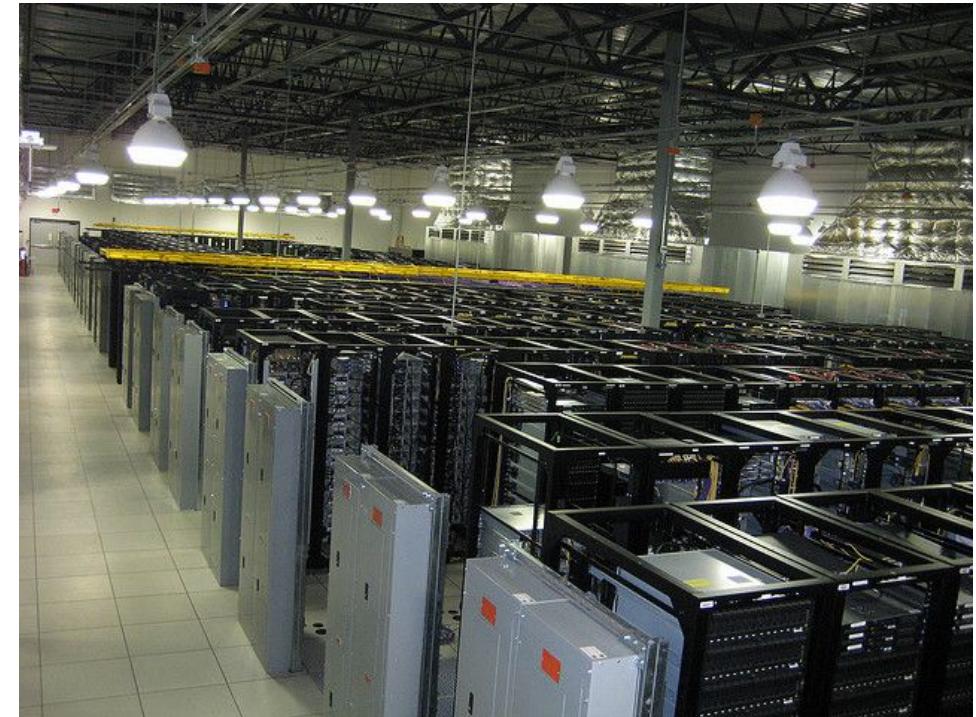
- Lots of data
- Poor IT infrastructure in many labs
- Where do they go?
- Get bigger hardware?
- Write more grants?

Cloud computing providers

- Amazon AWS
 - <https://aws.amazon.com/>
- Google cloud
 - <https://cloud.google.com/>
- Microsoft Azure
 - <https://azure.microsoft.com/en-us/>
- More...

Amazon Web Services (AWS)

- Infinite storage (scalable): S3 (simple storage service)
- Compute per hour: EC2 (elastic cloud computing)
- Ready when you are High Performance Computing
- Multiple football fields of HPC throughout the world



Some of the challenges of cloud computing:

- Not cheap
- Getting files to and from there
- Standardization can be a challenge if you don't control hardware
- PHI: personal health information & security concerns
 - In the USA: HIPAA act, PSQIA act, HITECH act, Patriot act, CLIA and CAP programs, etc.
 - <http://www.biostars.org/p/70204/>

Some of the advantages of cloud computing:

- There are better ways of transferring large files, and now AWS makes it free to upload files.
- A number of datasets exist on AWS (e.g. 1000 genome data, TCGA).
- Many useful bioinformatics AMI's (Amazon Machine Images) exist on AWS: e.g. cloudbiolinux & CloudMan (Galaxy) – now one for this course!
- Many flavors of cloud available, not just AWS

Key AWS concepts and terminology

- **AWS** - Amazon Web Services. A collection of cloud computing services provided by Amazon.
- **EC2** - Elastic Compute. An AWS service that allows you to configure and rent computers to meet your compute needs on an as needed basis.
- **EBS** - Elastic Block Storage. A data storage solution that allows you to rent disk storage and associate that storage with your compute resources. EBS volumes are generally backed by SSD devices.

Key AWS concepts and terminology

- **S3** - Simple storage service. Cheaper than EBS and allows for storage of larger amounts of data with some drawbacks compared to EBS. S3 volumes store data as objects that are accessed by an API or command line interface or other application designed to work with S3. EBS volumes on the other hand can be mounted as if they were a local disk drive associated with the Instance.
- **SSD** - Solid state drive. A particular type of storage hardware that is generally faster and more expensive than traditional hard drives.

What is difference between the 'Start', 'Stop', 'Reboot', and 'Terminate' (Instance States)?

- Start – turn on an EC2 instance that you have previously created
- Stop – turn off an EC2 instance that you have previously created
- Reboot – restart an EC2 instance
- Terminate – permanently stop and destroy an EC2 instance. Any associated EBS volumes may also be destroyed at this time depending on configuration

What is an AMI/snapshot?

- AMI (Amazon Machine Image) – a template that specifies how to launch EC2 instances
 - Root volume with operating system (OS), pre-installed applications, etc
 - Launch permissions determine who can use the AMI
 - Specification of (data) volumes to attach when launched
- You can create an AMI for any instance you have created/configured
- AMI can be made public for sharing (region-specific)
- Creating an AMI involves creating a snapshot of the root and any attached volumes. You will be charged to store this snapshot.

I can not log into my EC2 instance, what might have gone wrong?

- Is your instance running?
- Are you providing the correct path to your key file?
- Is it the correct key file?
- Have you set the permissions for your key file correctly?
- Did you specify a valid user for your AMI (e.g., ubuntu)?
- Did you specify the correct IP address?
- Does the Security Group for the instance allow access for your connection protocol (e.g., SSH) and location?

How much does it cost to use AWS EC2 resources?

Viewing 489 of 489 available instances						
Instance name	On-Demand hourly rate	vCPU	Memory	Storage	Network performance	
a1.medium	\$0.0255	1	2 GiB	EBS Only	Up to 10 Gigabit	
a1.large	\$0.051	2	4 GiB	EBS Only	Up to 10 Gigabit	
a1.xlarge	\$0.102	4	8 GiB	EBS Only	Up to 10 Gigabit	
a1.2xlarge	\$0.204	8	16 GiB	EBS Only	Up to 10 Gigabit	
a1.4xlarge	\$0.408	16	32 GiB	EBS Only	Up to 10 Gigabit	
a1.metal	\$0.408	16	32 GiB	EBS Only	Up to 10 Gigabit	
t4g.nano	\$0.0042	2	0.5 GiB	EBS Only	Up to 5 Gigabit	
t4g.micro	\$0.0084	2	1 GiB	EBS Only	Up to 5 Gigabit	
t4g.small	\$0.0168	2	2 GiB	EBS Only	Up to 5 Gigabit	
t4g.medium	\$0.0336	2	4 GiB	EBS Only	Up to 5 Gigabit	
t4g.large	\$0.0672	2	8 GiB	EBS Only	Up to 5 Gigabit	
t4g.xlarge	\$0.1344	4	16 GiB	EBS Only	Up to 5 Gigabit	

Data transfer (GB): In: free; Out: first 100 GB free, \$0.05-0.09 per GB depending on amount
EBS storage (GB/Month): \$0.08

S3 storage (GB/Month): \$0.023 standard, \$0.0125 infrequent access, or \$0.004 glacier

Why am I still getting a monthly bill?

- Generally you get an accounting of usage and cost on a 30 day cycle
 - Pricing is per instance-second consumed for each instance type.
 - Also charges for storage, transfers, etc
- Be aware of regions!
- Even when an instance is stopped, storage for root or other EBS volumes persist
- Creating AMIs/snapshots requires storage
- Explore the billing and cost management tools of AWS to track your spending, set warnings, etc

Amazon AWS documentation

https://rnabio.org/module-00-setup/0000/06/01/Intro_to_AWS/

<http://aws.amazon.com/console/>

In this workshop:

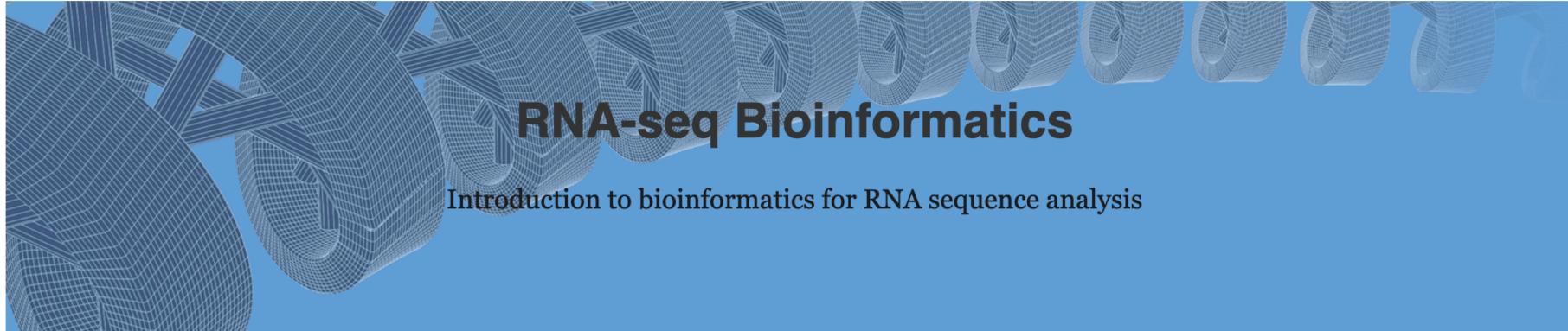
- Some tools (data) are
 - on your computer
 - on the web
 - on the cloud.
- You will become efficient at traversing these various spaces, and finding resources you need, and using what is best for you.
- There are different ways of using the cloud:
 1. Command line (like your own very powerful Unix box)
 2. With a web-browser (e.g. Galaxy): not in this workshop

Things we have set up:

- Loaded data files to a web server (genomedata.org)
- We started with a base Ubuntu 20.04 (Linux) instance and loaded a whole bunch of software for NGS analysis.
- We will copy this and create separate instances for everybody in the class.
- We've simplified the security: you basically all have the same login and file access, and opened ports. In your own world, you would be more secure to protect your data.

Logging into Amazon AWS

Go to course wiki, “Log into AWS” page



Log into AWS

[« Introduction to AWS](#)

[Course](#)

[Unix »](#)

Using cloud computing to complete this course involves two major components: (1) Launching an instance on the cloud (essentially renting a virtual computer by the hour and turning it on) and (2) logging into that instance).

Covered in this section: logging into AWS EC2 console, starting an instance from the course AMI, configuring it in the console (select instance AMI, instance type, instance details, storage volumes, tags, security group, and key pairs).



https://rnabio.org/module-00-setup/0000/07/01/Log_into_AWS/

Login to AWS console



Account ID or alias

cshlworkshops

IAM user name

cshl.student

Password

.....

Sign In



[Sign-in using root account credentials](#)

[Forgot password?](#)

<https://cshlworkshops.signin.aws.amazon.com/console>

Select "EC2" service

The screenshot shows the AWS search interface. A red arrow points to the search bar at the top left, which contains the text "EC2". Below the search bar, the results for "EC2" are displayed under the heading "Search results for 'EC2'". The first result is "EC2" with the subtext "Virtual Servers in the Cloud". To the right of the search results, there is a sidebar with the heading "Welcome to AWS" and sections for "Getting started", "Training and certification", and "What's new with AWS?". A red arrow points to the "N. Virginia" region selection in the top right corner of the sidebar.

Search for EC2

EC2 Virtual Servers in the Cloud

EC2 Image Builder A managed service to automate build, customize and deploy OS images

AWS Firewall Manager Central management of firewall rules

GuardDuty Intelligent Threat Detection to Protect Your AWS Accounts and Workloads

See all 51 results ▶

Dashboard EC2 feature

Limits EC2 feature

N. Virginia kelscott @ 5772-5572-5291

Welcome to AWS

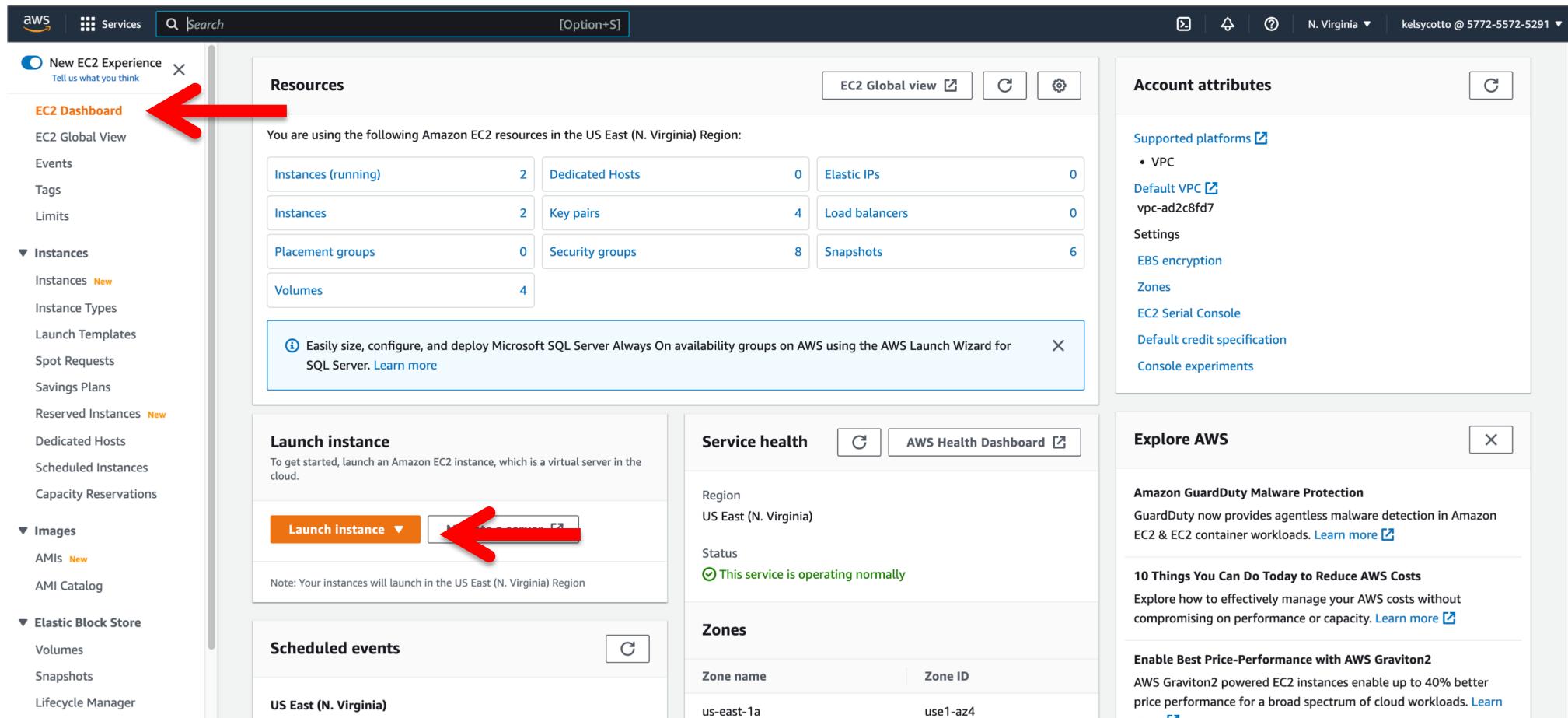
Getting started Learn the fur information 1

Training and certification Learn from AWS experts and advance your skills and knowledge.

What's new with AWS? Discover new AWS services, features, and Regions.

Make sure you are in N. Virginia region

From EC2 Dashboard, launch a new Instance



Name your instance “FirstnameLastname” (e.g. KelsyCotto)

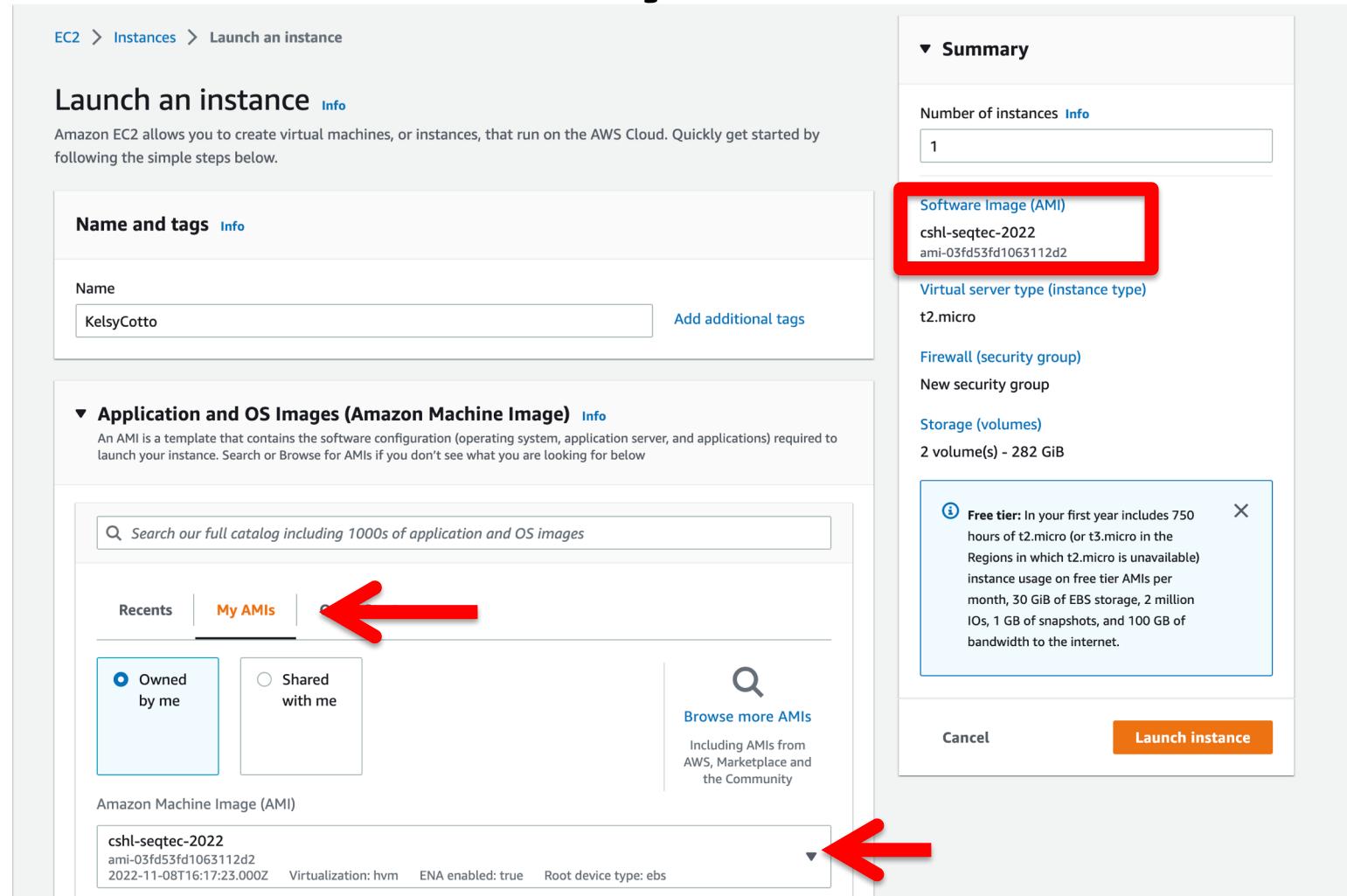
The screenshot shows the 'Launch an instance' wizard in the AWS EC2 console. The current step is 'Name and tags'. A red arrow points to the 'Name' input field, which contains the value 'KelsyCotto'. To the right of the input field is a link 'Add additional tags'. Below this section is a collapsed panel titled 'Application and OS Images (Amazon Machine Image)'. On the right side of the screen, there is a 'Summary' panel with the following details:

- Number of instances: 1
- Software Image (AMI): Amazon Linux 2 Kernel 5.10 AMI... (with a 'read more' link and AMI ID: ami-09d3b3274b6c5d4aa)
- Virtual server type (instance type): t1.micro
- Firewall (security group): New security group
- Storage (volumes): 1 volume(s) - 8 GiB

A tooltip for the free tier is displayed, stating: "Free tier: In your first year includes 750 hours of t2.micro (or t3.micro in the Regions in which t2.micro is unavailable) instance usage on free tier AMIs per month, 30 GiB of EBS storage, 2 million IOPS, 1 GB of snapshots, and 100 GB of bandwidth to the internet." At the bottom right are 'Cancel' and 'Launch instance' buttons.

Important: Don't forget to name your instance!
(FirstnameLastname)

Choose an AMI – Find the CSHL SEQTEC 2022 AMI in the My AMIs



Choose “m5.2xlarge” instance type, then “Next: Configure Instance Details”.

The screenshot shows the 'Configure Instance Details' step of the AWS CloudFormation 'Create New Stack' wizard. The left panel displays the 'Amazon Machine Image (AMI)' section, listing 'cshl-seqtec-2022' with details: ami-03fd53fd1063112d2, 2022-11-08T16:17:23.000Z, Virtualization: hvm, ENA enabled: true, Root device type: ebs. Below this, the 'Description' field is empty. The 'Architecture' field shows 'x86_64' and the 'AMI ID' is 'ami-03fd53fd1063112d2'. The right panel shows the 'Software Image (AMI)' section with the same details. A red box highlights the 'Virtual server type (instance type)' field, which is set to 'm5.2xlarge'. Other settings shown include a 'New security group' and '2 volume(s) - 282 GiB'. A callout box provides information about the 'Free tier' for t2.micro instances.

Amazon Machine Image (AMI)

cshl-seqtec-2022
ami-03fd53fd1063112d2
2022-11-08T16:17:23.000Z Virtualization: hvm ENA enabled: true Root device type: ebs

Description

-

Architecture AMI ID

x86_64 ami-03fd53fd1063112d2

▼ Instance type [Info](#)

Instance type

m5.2xlarge
Family: m5 8 vCPU 32 GiB Memory
On-Demand Linux pricing: 0.384 USD per Hour
On-Demand Windows pricing: 0.752 USD per Hour

[Compare instance types](#)

Software Image (AMI)

cshl-seqtec-2022
ami-03fd53fd1063112d2

Virtual server type (instance type)

m5.2xlarge

Firewall (security group)

New security group

Storage (volumes)

2 volume(s) - 282 GiB

i **Free tier:** In your first year includes 750 hours of t2.micro (or t3.micro in the Regions in which t2.micro is unavailable) instance usage on free tier AMIs per month, 30 GiB of EBS storage, 2 million I/Os, 1 GB of snapshots, and 100 GB of bandwidth to the internet.

Choose an existing key pair: "cshl_2022_student"

The screenshot shows the 'Choose an existing key pair' step in the AWS Lambda 'Create New Function' wizard. It displays two sections: 'Instance type' and 'Key pair (login)'. The 'Instance type' section shows an m5.2xlarge instance selected, with details about its family, vCPUs, memory, and pricing. The 'Key pair (login)' section shows the key pair 'cshl_2022_student' selected, with a red arrow pointing to it. A 'Create new key pair' button is also visible.

▼ Instance type [Info](#)

Instance type

m5.2xlarge

Family: m5 8 vCPU 32 GiB Memory

On-Demand Linux pricing: 0.384 USD per Hour

On-Demand Windows pricing: 0.752 USD per Hour

▼ Compare instance types

▼ Key pair (login) [Info](#)

You can use a key pair to securely connect to your instance. Ensure that you have access to the selected key pair before you launch the instance.

Key pair name - *required*

cshl_2022_student

[Create new key pair](#)

Select an Existing Security Group, choose "SSH_HTTP". Then hit "Review and Launch".

Network settings [Info](#)

Edit

Network [Info](#)
vpc-ad2c8fd7

Subnet [Info](#)
No preference (Default subnet in any availability zone)

Auto-assign public IP [Info](#)
Enable

Firewall (security groups) [Info](#)
A security group is a set of firewall rules that control the traffic for your instance. Add rules to allow specific traffic to reach your instance.

Create security group Select existing security group SSH/HTTP/Jupyter sg-0fd7356e013ca81a3 X

Security groups [Info](#)
[Select security groups](#)

Compare security group rules

Software Image (AMI)
cshl-seqtec-2022
ami-03fd53fd1063112d2

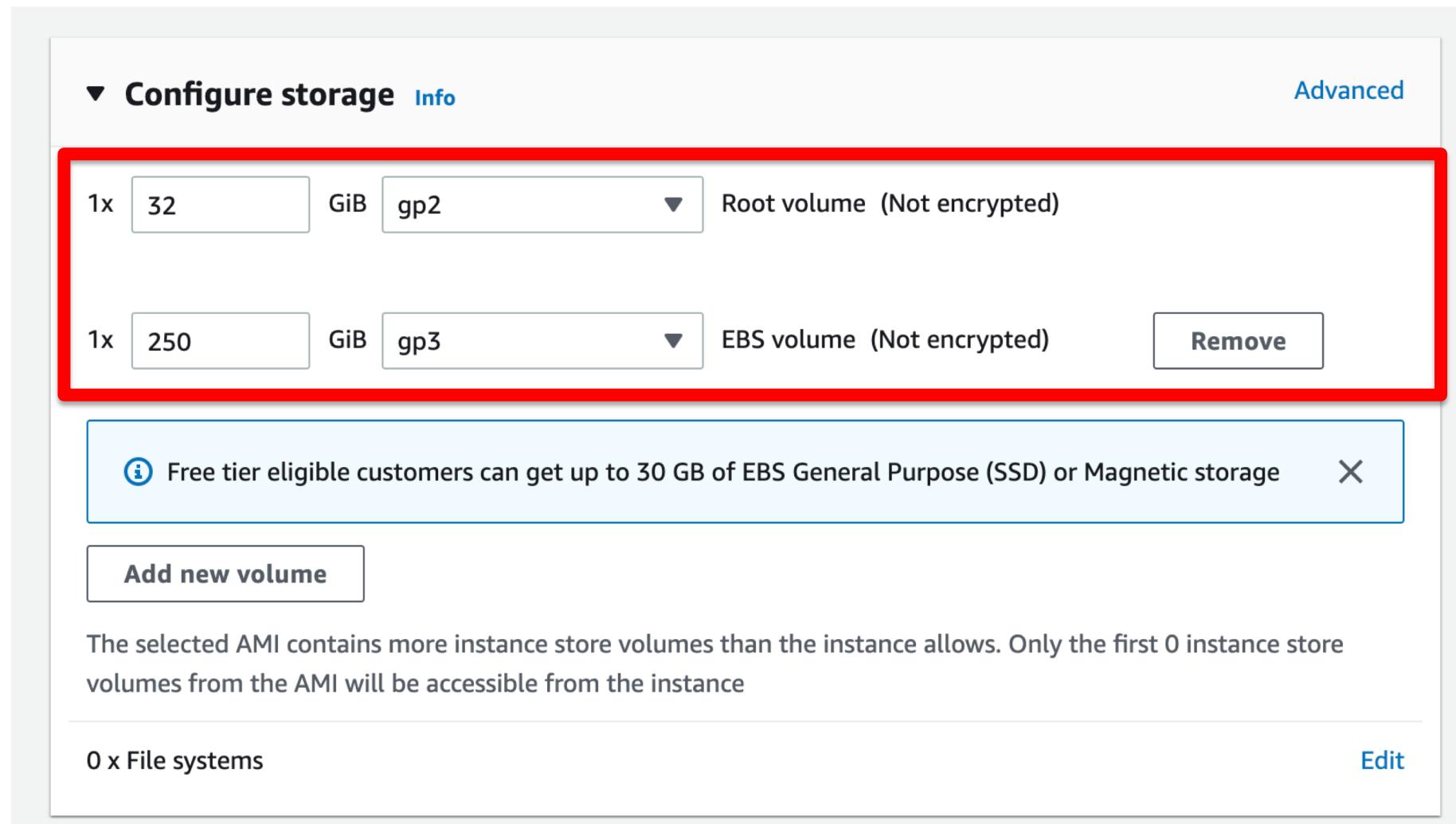
Virtual server type (instance type)
m5.2xlarge

Firewall (security group)
SSH/HTTP/Jupyter SSH/HTTP/Jupyter

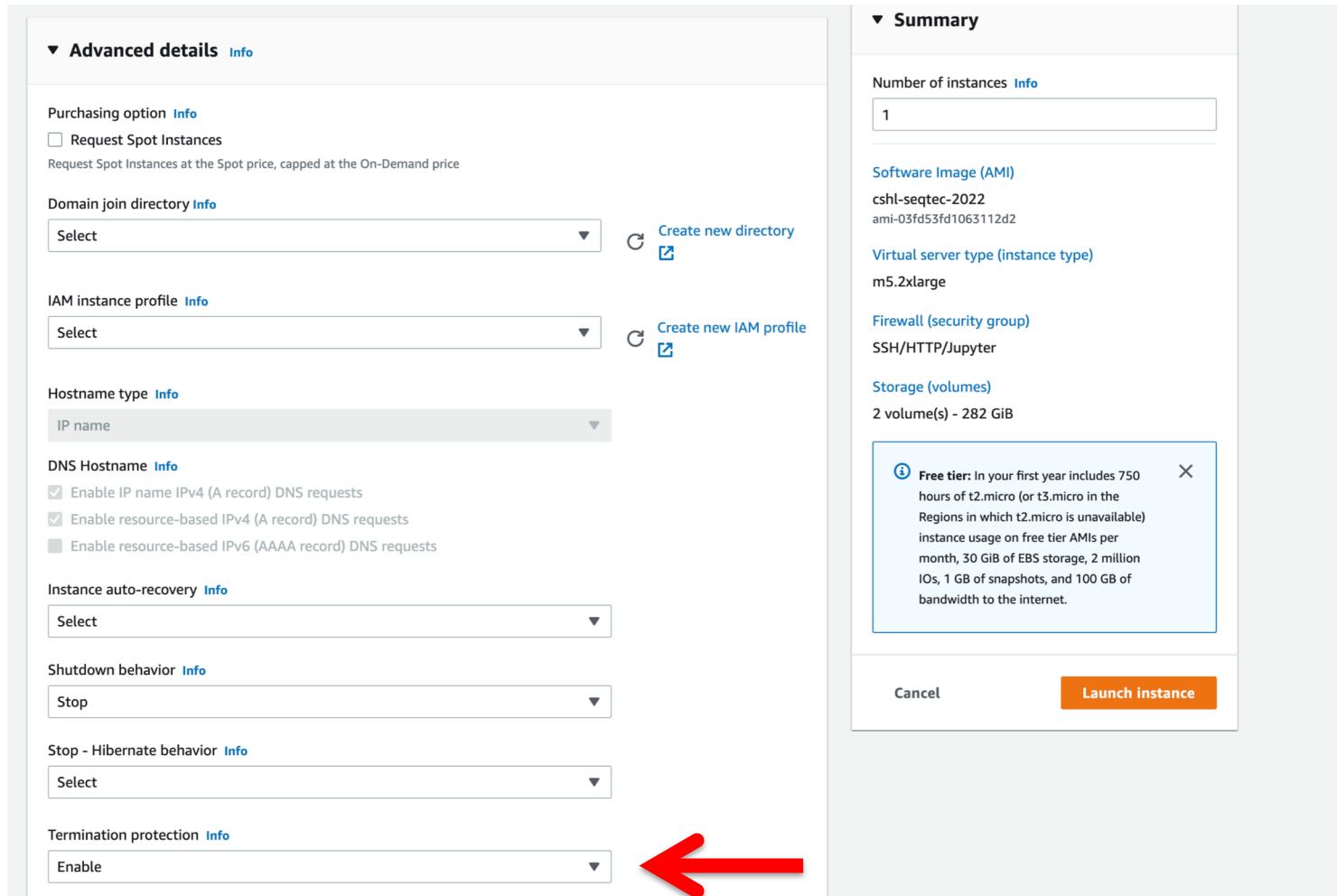
Storage (volumes)
2 volume(s) - 282 GiB

Free tier: In your first year includes 750 hours of t2.micro (or t3.micro in the Regions in which t2.micro is unavailable) instance usage on free tier AMIs per month, 30 GiB of EBS storage, 2 million IOs, 1 GB of snapshots, and 100 GB of bandwidth to the internet. X

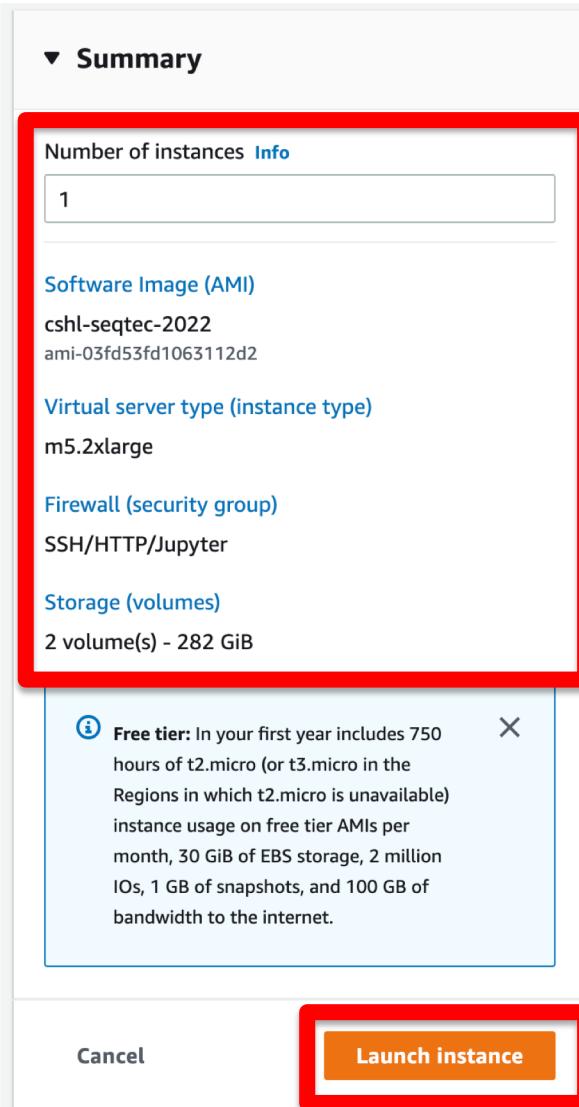
You should see 1 32 GiB root volume and 1 250 GiB EBS volume as the two storage volumes.



Select “Enable” for termination protection.



Review the details of your instance and hit Launch



View Instances to see your new instance spinning up!

EC2 > Instances > Launch an instance



Success

Successfully initiated launch of instance (i-00e1dc9cd6843478d)

▶ Launch log

Next Steps

Create billing and free tier usage alerts

To manage costs and avoid surprise bills, set up email notifications for billing and free tier usage thresholds.

[Create billing alerts](#)

Connect to your instance

Once your instance is running, log into it from your local computer.

[Connect to instance](#)

[Learn more](#)

Connect an RDS database

New

Configure the connection between an EC2 instance and a database to allow traffic flow between them.

[Connect an RDS database](#)

[Create a new RDS database](#) [Learn more](#)

[View all instances](#)

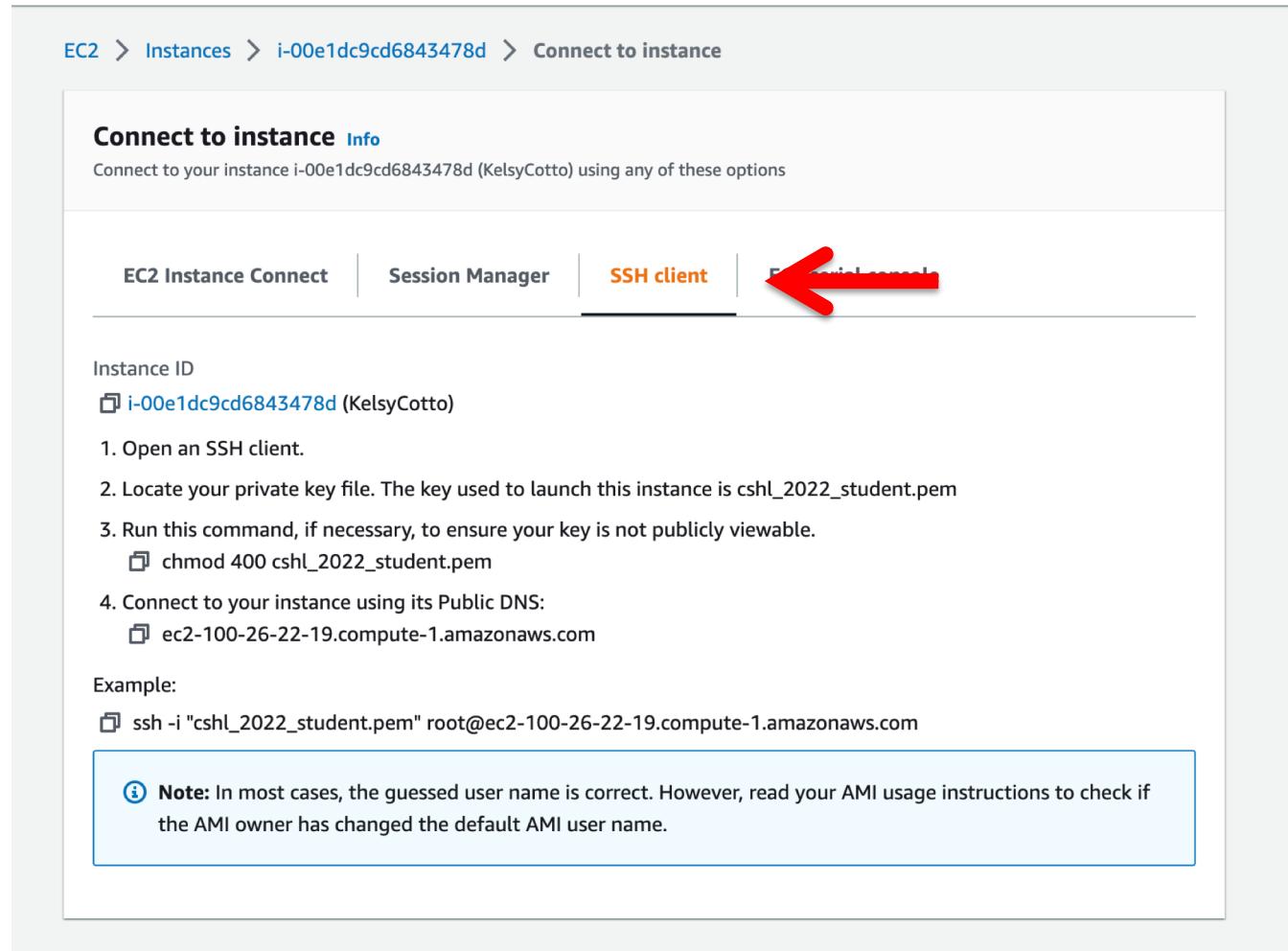
Find YOUR instance, select it, and then hit connect for instructions on how to connect (It may take some time for your instance to be ready)

The screenshot shows the AWS EC2 Instances page. On the left, there's a navigation sidebar with links like EC2 Dashboard, EC2 Global View, Events, Tags, Limits, Instances (selected), Images, Elastic Block Store, Network & Security, and more. The main area displays a table of instances:

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 IP	Elasti
Kcotto	i-0302a728fbdc4095f	Running	m5.2xlarge	2/2 checks passed	No alarms	us-east-1a	ec2-18-234-50-3.comp...	18.234.50.3	-
Instructor_test	i-088f1b7da8b8d0656	Running	m5.2xlarge	2/2 checks passed	No alarms	us-east-1a	ec2-18-212-77-248.co...	18.212.77.248	-
KelsyCotto	i-00e1dc9cd6843478d	Running	m5.2xlarge	Initializing	No alarms	us-east-1a	ec2-100-26-22-19.com...	100.26.22.19	-

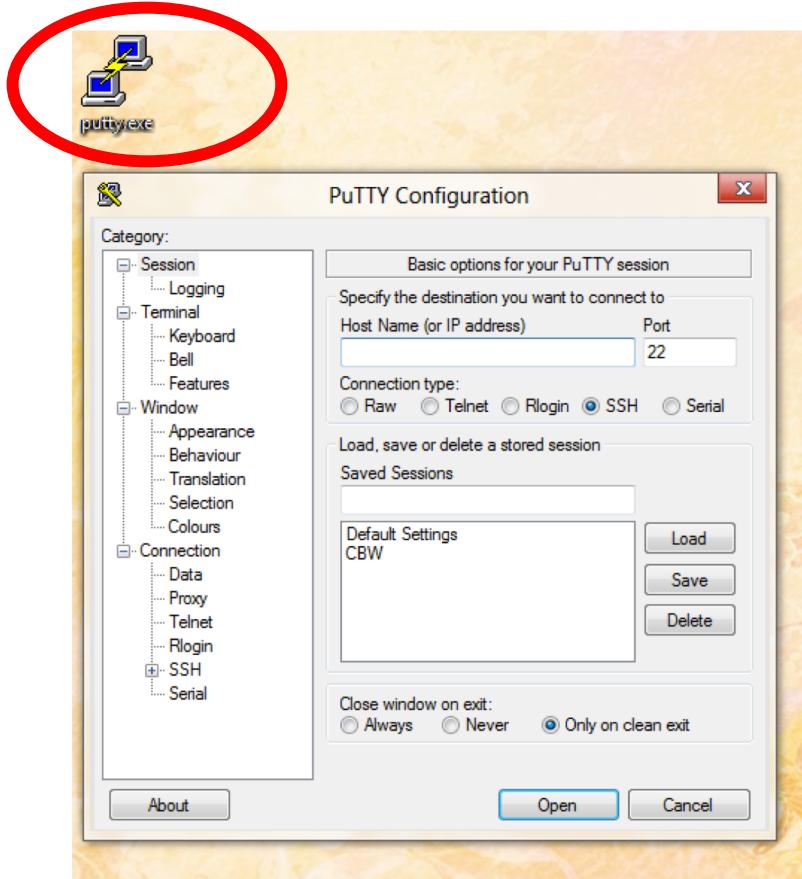
At the top right of the table, there's a 'Connect' button with a red box around it. Below the table, a specific instance is selected: 'Instance: i-00e1dc9cd6843478d (KelsyCotto)'. The 'Details' tab is selected in the sub-menu. A red arrow points to the Public IPv4 address '100.26.22.19 | open address'.

Take note of your Public DNS/IP and the instructions on changing permissions for the key file (Note, we will login as ubuntu NOT root)

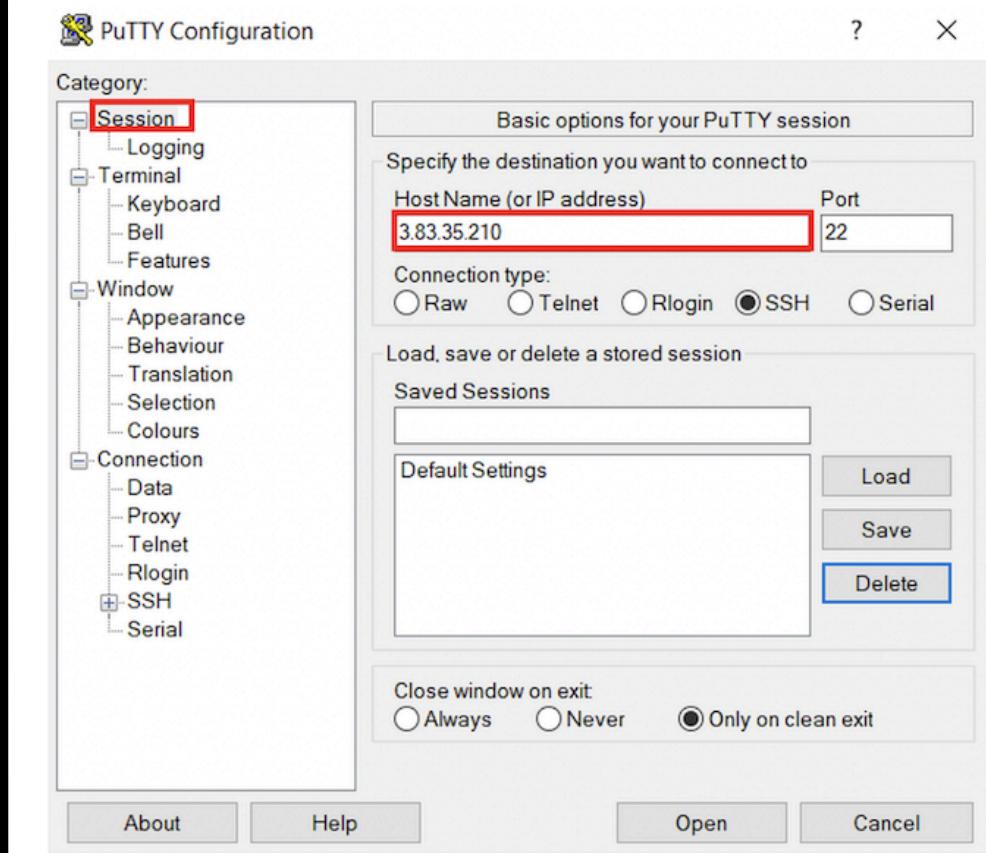


Logging into your instance (Windows)

Open PuTTY

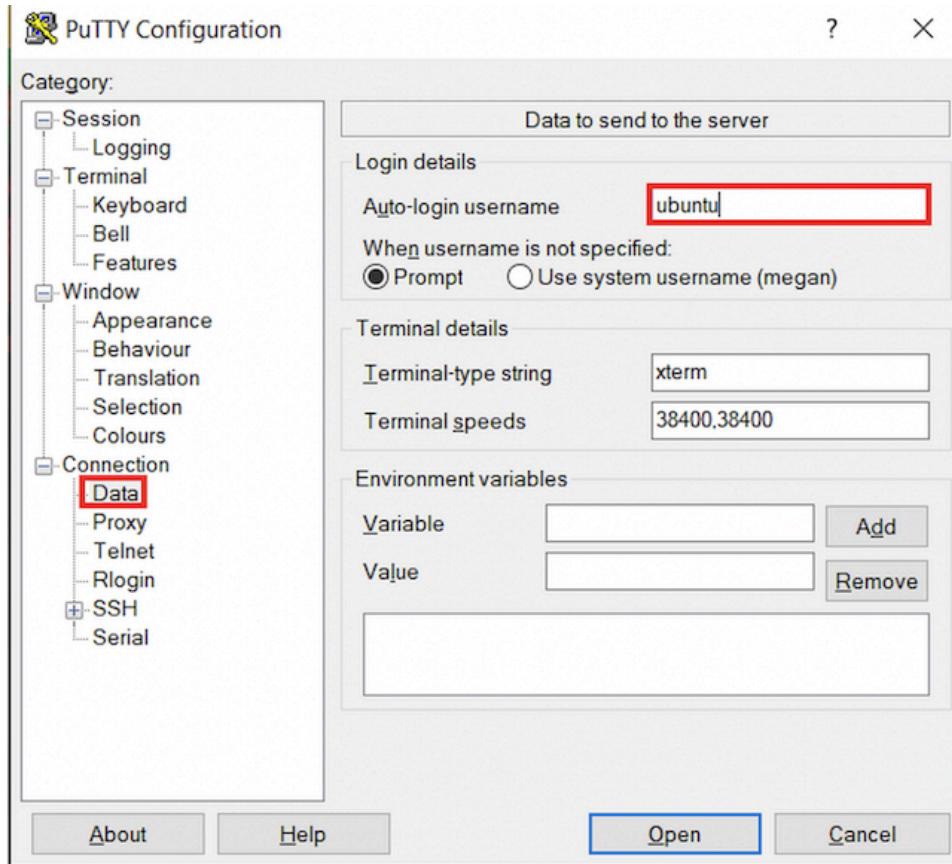


Enter the Host Name (IP address)

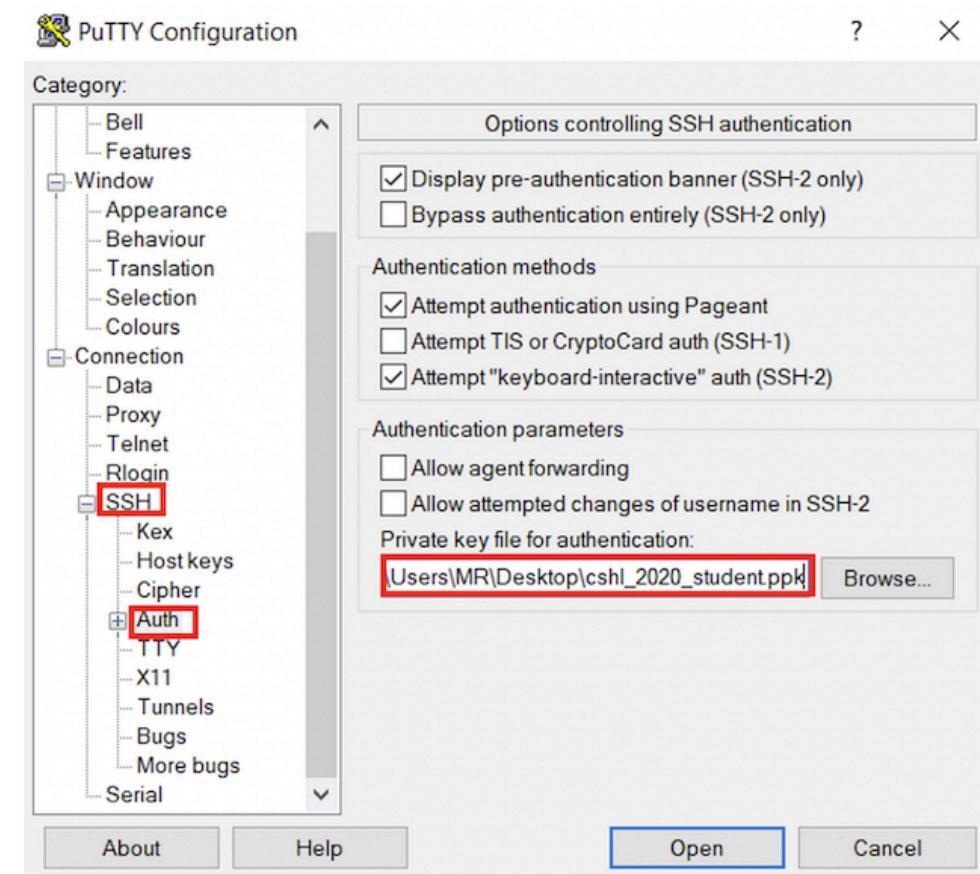


Logging into your instance (Windows)

Choose Connection -> Data
Enter the username 'ubuntu'



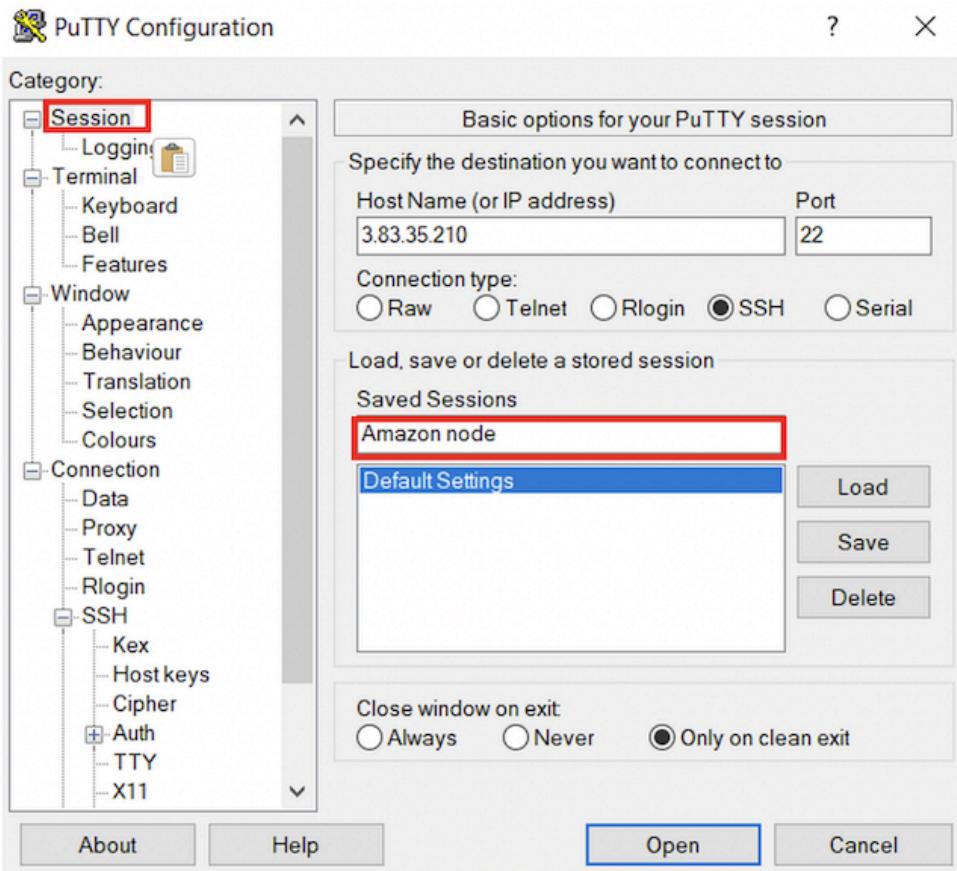
Choose SSH -> Auth
Browse to Private key (ppk) file



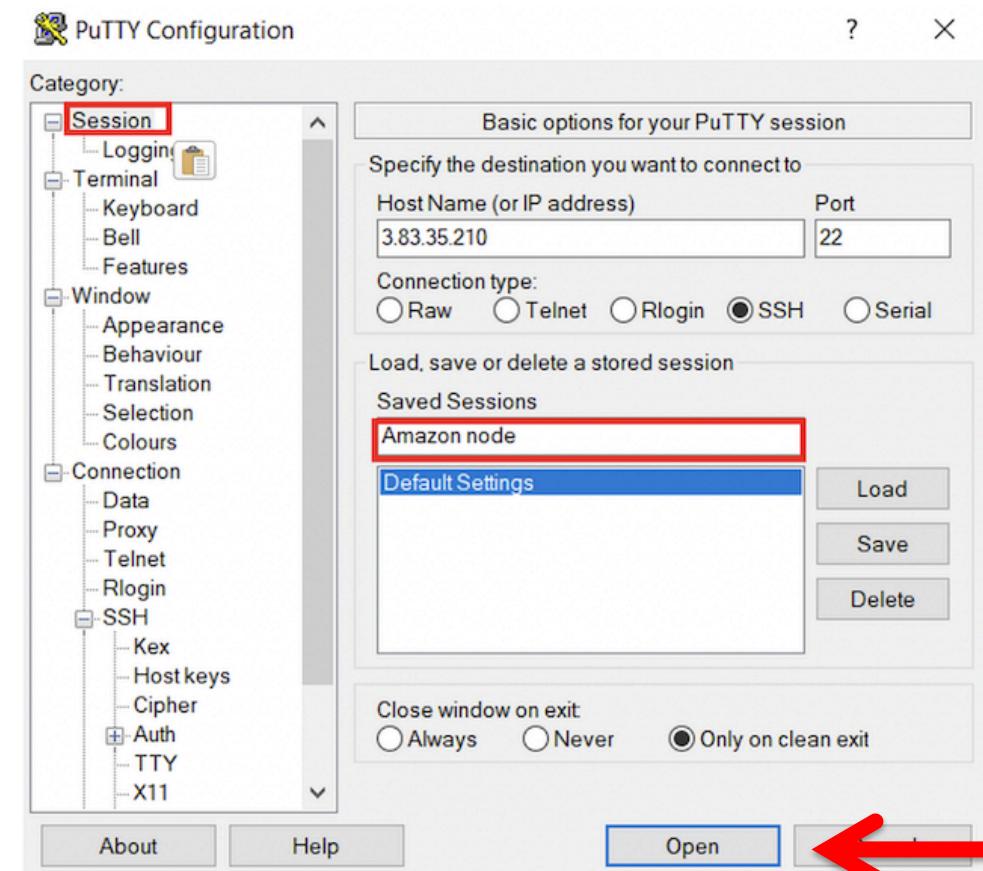
Logging into your instance (Windows)

Choose Session

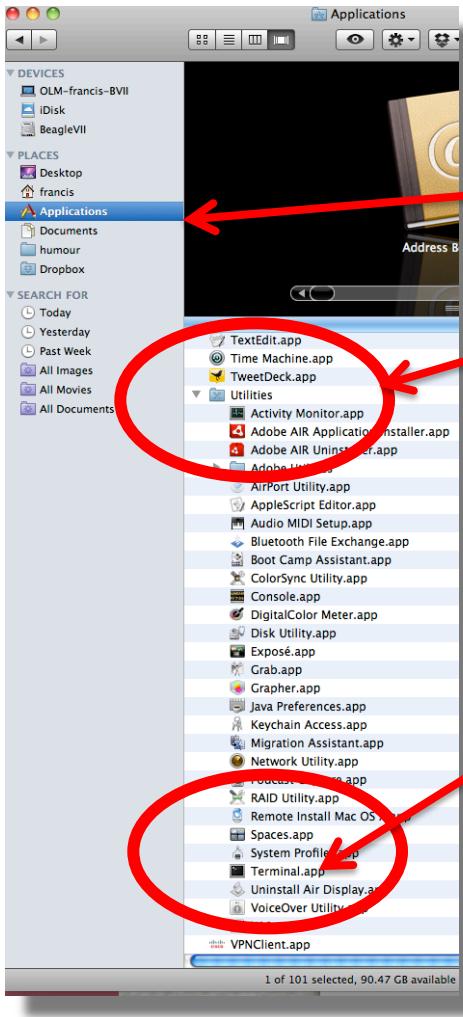
Save your session as "Amazon Node"



Double-click saved “Amazon Node” session OR
Select “Amazon Node” session and click Open



Logging into your instance (Mac)

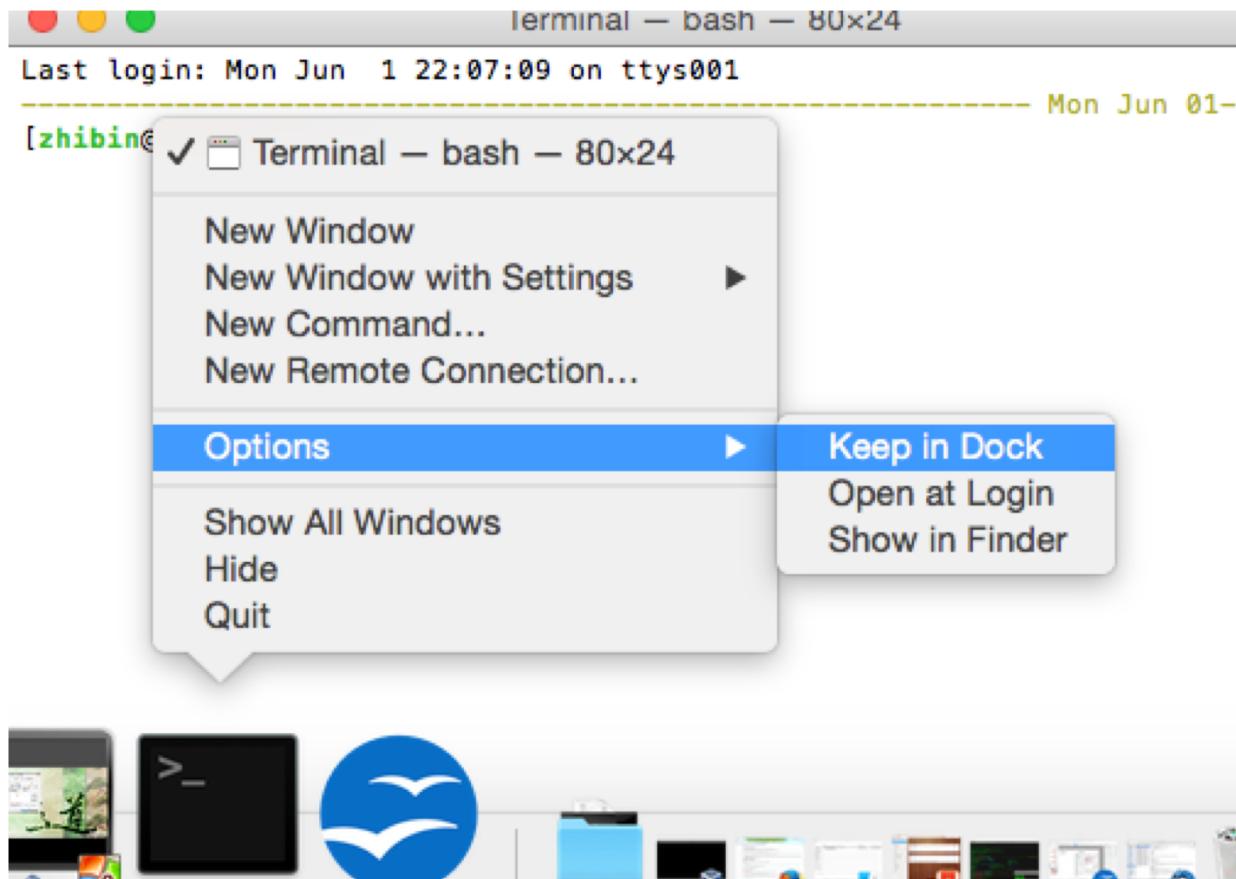


In a Finder window
'Applications' -> 'Utilities' -> 'Terminal'



Or on your dock

Add the terminal App to your dock



Creating a working directory on your Mac called ‘cshl’

```
obis-air:~ ogriffit$ pwd
/Users/ogriffit
obis-air:~ ogriffit$ ls
Applications   Desktop      Dropbox      Movies       Public      gittemp     temp
Attachments    Documents    Google Drive  Music        bin         igv
Box Sync       Downloads    Library      Pictures    git         ncbi
obis-air:~ ogriffit$ mkdir cshl
obis-air:~ ogriffit$ cd cshl
obis-air:cshl ogriffit$ ls -la
total 0
drwxr-xr-x  2 ogriffit  staff   68 Nov 13 22:18 .
drwxr-xr-x+ 58 ogriffit  staff  1972 Nov 13 22:18 ..
obis-air:cshl ogriffit$ █
```

mkdir cshl
cd cshl

Obtain the course SSH key file

- NOTE for Mac users. You will need to use a “.pem” file
- **NOTE for Windows Users.** You will need to use a “.ppk” file instead.
 - This is created from the “.pem” file.
 - <https://aws.amazon.com/premiumsupport/knowledge-center/convert-pem-file-into-ppk/>
- The SSH key file will be used to securely login to your student instance on the cloud

Save the pem/ppk file you received via email/slack to your new cshl folder

Viewing the ‘key’ file once downloaded

`cat cshl_2021_student.pem`

```
Kelsys-MacBook-Pro:keys kcotto$ cat cshl_2019_student.pem
-----BEGIN RSA PRIVATE KEY-----
MIIEpAIBAAKCAQEAgGtpnqERtEu/SCmeF2r1HMESMao0fEJiAwQwk2/SNXK8izr0IH0zTVvmE1D
VUwWq7pkvhjh05pDb+2U9HiZe3sxLv3S1NrkATYF/NsrpwB+q1vwqzGW9sQ6uj45RWrPkiZlsaj
TQZmyFRu+tLJTRU3hQDqA0MRWTx1Wxv0gFzuZy/qb+DALuFQsInrEKnijrwdLmd6usaBTvhc0gFS
B9oEelH0bZHJTZFW/wP+Z0uZq0Ujir7Qw0LTM45QH/L0dBdUl3k/mBeez00yvnKMwj8E4Xi0rQ0t
hHtQ7F9iSILK80W1rRH0qwxwt9ycEH1JtNMQmUTif0vE2XJ6l06chQIDAQABAoIBABG7P/FHu/Qp
WFgg+89myuqR6GvA2X55CFSzFyG0aQyrj5jDleFtdu2uXiISG8gUBZYvlzxx82aOC0P5j04SBq0
xD/qRlukY/jyXyPn77w/ExmaNoLJjl1W9RUSH0JYLIzVpFPGes3u5zGSGDTSDNh3sSdWhq1FX3l
7vY5b6UAQgahXf0dpGFxt6P6qb/BKFQFsThXk7GXMyS/kr4w7ZlhPWHRMSu2UBdS/a/beAVN76qU
E/10HR30oAuqghusZabpbX7MJl00VcviaQgeF8Z+xf2uugDEAKut1PxW0+yvGM4SpZZ0skFZz6YI
pBnX4ELWPhCeMzq4M18QY6ptR2UCgYEAc6YfNlymg24NJu8PANx8navTi50WYuXWXrj0mrLr3SR
/XY+w26cgipM+K5eQfxSr4Yb8BQKjRktMzBzf5nKdaX4pzYIquQH02B0HDhTooHAhkbTMTmKukv4
oEW06wcEE6RQi fw4xbMEnfQfHJB121am+jwI8Xb7idwMG4pU/nsCgYEAtD0x4bNC1X3A5by50uY0
WXrtgQszCXYcbkrpbjRET12f9hgz9MRMHY/xH/XGvMutZSFV2rCZRwd7lm+QWGadk/MQS0kouzW6
gSasyjFq+MKCkqYnS3/JTbx8yrLZmzl0LtX6pwmwg0Zy8aJjYHo9a2/EI8Tjh2d0SxeadIRVYP8C
gYAToiXww1Vdu+dj/7TDLqYCtdHOVAxJX/utI9Q3yoIryuh+bWmFvEIvAmIGXyyQZRyoZwgIS4A
PNH03+bEa+69wbzlhksiK5g8GKgISVdLC4rZZXB5ehgTmWV7IgJ89y/SF4G/Ityo30K0ohALh597
NcvNEzzqrutja1IIMvTKMwKBgQCkM+QP1Tqc0TbVlfvClviXuJBLsiJLCImYeZL0nZVmIMusbhxX
b8ZQYGSyUz09nulXau1G1QDvXvf089CzWL1SomxBoHlFJQvGwa9FfYQRIVPHuqut8rs4oPGn0QzC
h7M7QCJcr00oAcrSLLkQmgz+phIw7BzFr039J4HFiRInjQKBgQCtdEvcbtyk8Jh4WH3z0wpkc43f
U8DZhZwjRQpGWlD8CPj9RgRnE4+1PCH6s/RLQf7SiE1ZjX/0Ud0WPEvr0j5sVjy0IujohRbty0CM
oqWeSeUb1sLogRvMrTfCEpl/rz3GpoQ1SC/5s6XvjnnKK8RN8s7MseLuuJ63T/CRBpIs8A==
-----END RSA PRIVATE KEY-----Kelsys-MacBook-Pro:keys kcotto$
```

Changing file permissions of your ‘key’ file (Mac/Linux)

ls -l (long listing)

```
-rw-r--r--@ 1 kcotto staff 1696 Nov 9 09:19 cshl_2022_student.pem
```

 rwx : owner

 rwx : group

 rwx: world

 r read (4)

 w write (2)

 x execute (1)

Which ever way you add these 3 numbers, you know which integers were used (6 is always 4+2, 5 is 4+1, 4 is by itself, 0 is none of them etc ...)

So, when you have:

chmod 400 <file name>

It is “r” for the the file owner **only**

Logging into your instance

Mac/Linux

```
cd ~/cshl  
chmod 400 cshl_2022_student.pem  
ssh -i cshl_2022_student.pem ubuntu@[YOUR PUBLIC IP]
```

Copying files from AWS to your computer (using a web browser)

Index of /

<u>Name</u>	<u>Last modified</u>	<u>Size</u>	<u>Description</u>
 anaconda3/	2022-11-07 16:56	-	
 bin/	2022-11-08 10:39	-	
 src/	2022-11-08 10:25	-	
 tmp	2022-07-04 07:24	736K	
 workspace/	2022-11-08 10:58	-	

Apache/2.4.41 (Ubuntu) Server at 100.26.22.19 Port 80

[http://\[YOUR PUBLIC DNS OR IP\]/](http://[YOUR PUBLIC DNS OR IP]/)

Logging out of your instance

Mac/Linux – simply type exit

exit

Note, this disconnects the terminal session (ssh connection) to your cloud instance. But, your cloud instance is still running! See next slide for how to stop your instance.

When you are done for the day you can “Stop” your instance – Don’t Terminate!

The screenshot shows the AWS EC2 Instances page. On the left, there's a navigation sidebar with sections like EC2 Dashboard, EC2 Global View, Events, Tags, Limits, Instances (selected), and Images. Under Instances, 'Instances' is also selected. A red arrow points from the 'Instances' section in the sidebar to the 'Instances' tab in the main header. Another red arrow points from the 'Instances' tab to the table below. In the table, three instances are listed: 'Kcotto', 'Instructor_test', and 'KelsyCotto'. The 'KelsyCotto' row has a red box around it and a red arrow pointing to it. The 'Actions' column for 'KelsyCotto' has a context menu open, with 'Stop instance' highlighted by a red arrow. The table columns include Name, Instance ID, Instance state, Instance type, Status check, Alarm status, Availability Zone, Public IPv4 DNS, and Public IPv4 ... (partially visible).

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...
Kcotto	i-0302a728fbdc4095f	Running	m5.2xlarge	2/2 checks passed	No alarms	us-east-1a	ec2-18-234-50-3.comp...	18.234.50.3
Instructor_test	i-088f1b7da8b8d0656	Running	m5.2xlarge	2/2 checks passed	No alarms	us-east-1a	ec2-18-212-77-248.co...	18.212.77.248
<input checked="" type="checkbox"/> KelsyCotto	i-00e1dc9cd6843478d	Running	m5.2xlarge	2/2 checks passed	No alarms	us-east-1a	ec2-54-221-110-174.co...	54.221.110.174

Instance: i-00e1dc9cd6843478d (KelsyCotto)

Details Security Networking Storage Status checks Monitoring Tags

Instance summary Info

Instance ID

Go to AWS EC2 Dashboard, select “Instances” tab, then find your instance. Right-click and chose ‘Stop instance’

Next morning, you can “Start” your instance again

The screenshot shows the AWS EC2 Instances page. On the left, there's a sidebar with various navigation links. A red arrow points from the 'Instances' link in the sidebar to the 'Instances' tab in the main header. Another red arrow points from the 'Instances' tab in the header to the 'Actions' dropdown menu. Inside the dropdown, a red arrow points to the 'Start instance' option, which is highlighted. The main table lists three instances:

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 ...	Elast...
Kcotto	i-0302a728fbdc4095f	Running	m5.2xlarge	2/2 checks passed	No alarms	us-east-1a	ec2-18-234-50-3.comp...	18.234.50.3	-
Instructor_test	i-088f1b7da8b8d0656	Running	m5.2xlarge	2/2 checks passed	No alarms	us-east-1a	ec2-18-212-77-248.co...	18.212.77.248	-
KelsyCotto	i-00e1dc9cd6843478d	Stopped	m5.2xlarge	-	-	t-1a	-	-	-

At the bottom of the page, a red box contains the following text:

Go to AWS EC2 Dashboard, select “Instances” tab, then find your instance. Right-click and chose ‘Start Instance’

When you restart your instance you will need to find your new Public DNS or IP address. Select your instance and “Connect” or look in Description tab. Then go back to instructions for “Logging into your instance”

The screenshot shows the AWS EC2 Instances page. In the top left, there's a sidebar with various navigation links like EC2 Dashboard, Events, Tags, Limits, Instances, Images, Elastic Block Store, and Network & Security. The main area displays a table of instances with columns for Name, Instance ID, Instance state, Instance type, Status check, Alarm status, Availability Zone, Public IPv4 DNS, and Public IPv4 IP. Three instances are listed: Kcotto, Instructor_test, and KelsyCotto. The KelsyCotto instance is selected, indicated by a checked checkbox in the first column. A red arrow points to the 'Connect' button in the top right corner of the 'Instances (1/3) Info' header. Below the table, a detailed view for the KelsyCotto instance is shown. The 'Details' tab is active, displaying tabs for Security, Networking, Storage, Status checks, Monitoring, and Tags. Under the 'Instance summary' section, the Public IPv4 address is listed as 100.26.22.19, with a link to open the address. A red arrow points to this link. Other details shown include Instance ID (i-00e1dc9cd6843478d), Instance state (Running), Private IP DNS name (ip-172-31-28-82.ec2.internal), Instance type (m5.2xlarge), and Elastic IP addresses (none).

Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IPv4 DNS	Public IPv4 IP	Elast
Kcotto	i-0302a728fbdc4095f	Running	m5.2xlarge	2/2 checks passed	No alarms	us-east-1a	ec2-18-234-50-3.compute...	18.234.50.3	-
Instructor_test	i-088f1b7da8b8d0656	Running	m5.2xlarge	2/2 checks passed	No alarms	us-east-1a	ec2-18-212-77-248.compute...	18.212.77.248	-
KelsyCotto	i-00e1dc9cd6843478d	Running	m5.2xlarge	Initializing	No alarms	us-east-1a	ec2-100-26-22-19.compute...	100.26.22.19	-

So, at this point:

- Your laptop/pc is ready for the workshop
- If it is not, you know where to get the information you need
- You know how to login to AWS
- The next step is to login to your linux machine on AWS and learn the basics of a linux command line

We are on a Coffee Break & Networking Session

Workshop Sponsors:



Canadian Centre for
Computational
Genomics



HPC4Health



GenomeCanada

Key AWS concepts and terminology

- **HDD** - Hard disk drive. A particular type of storage hardware that is generally cheaper and larger but slower than SSD. HDD drives are traditional hard drives that access data on a spinning magnetic disk.
- **Ephemeral storage** - Also known as Instance Store storage. Data storage associated with an EC2 instance that is local to the host computer. This storage does not persist when the instance is stopped or terminated. In other words, anything you store in this way will be lost if the system is stopped or terminated. Instance store volumes may be backed by SSD or HDD devices.

What is a Region?

- An AWS Region is set of compute resources that Amazon maintains (like the Data Center image shown before)
- Each Region corresponds to a physical warehouse of compute hardware (computers, storage, networking, etc.).
- At the time of writing there are 27 regions with more planned to come online soon: (US East (N.Virginia), US East (Ohio), US West (Oregon), US West (N. California), GovCloud (US-West), GovCloud (US-East), Canada (Central), EU (Ireland), EU (Frankfurt), EU (London), EU (Paris), EU (Milan), EU (Stockholm), Middle East (Bahrain), Middle East (UAE), Africa (Cape Town), Asia Pacific (Singapore), Asia Pacific (Sydney), Asia Pacific (Seoul), Asia Pacific (Tokyo), Asia Pacific (Mumbai), Asia Pacific (Hong Kong), Asia Pacific (Beijing), Asia Pacific (Osaka), Asia Pacific (Jakarta), Asia Pacific (Ningxia), and South America (Sao Paulo)).
- When you are logged into the AWS EC2 console, you are always operating in one of these regions.

What is a Region?

- Current region shown in the upper right corner of console
- It is important to pay attention to what region you are using for several reasons.
 - When you create an EC2 instance (EBS volume, etc) in one region you won't see it in another region.
 - The cost to use many AWS resources varies by region.
 - The region may influence network performance when you are accessing the instance, especially if you need to transfer large amounts of data in or out.
 - Billing is tracked separately for each region
 - Generally you should choose a region that is close to you or your users. But cost is also a consideration.