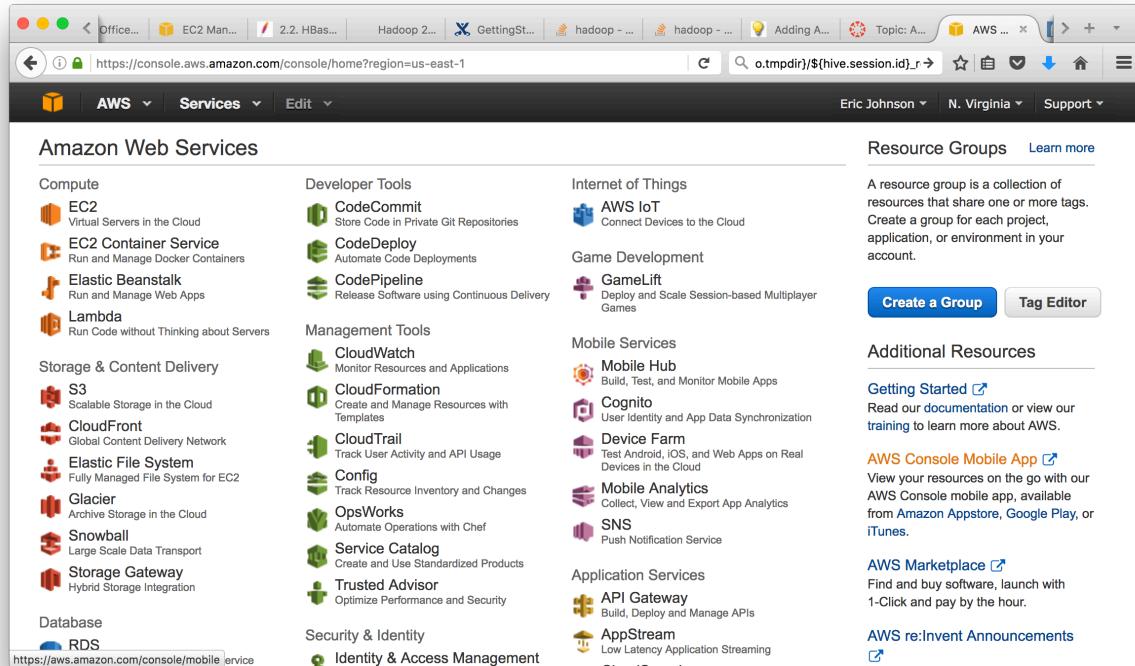
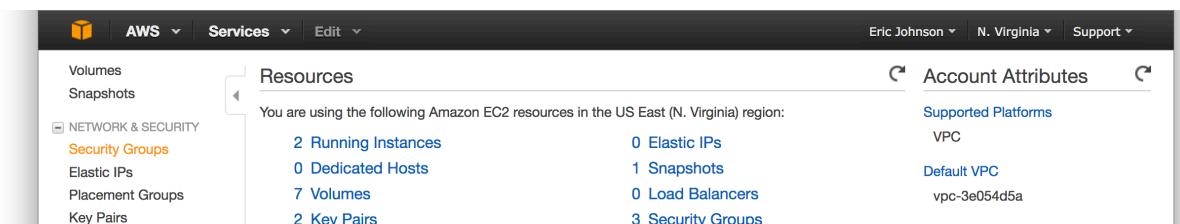


1) Visit AWS EC2 Console Logon Page and select "EC2"



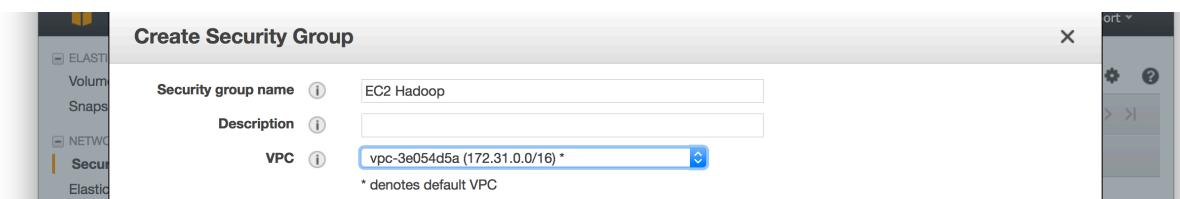
The screenshot shows the AWS Home page with the 'Services' menu expanded. Under the 'Compute' section, 'EC2' is selected, which is highlighted in orange. Other services listed under Compute include EC2 Container Service, Elastic Beanstalk, Lambda, and RDS. The 'Storage & Content Delivery' section includes S3, CloudFront, Elastic File System, Glacier, Snowball, and Storage Gateway. The 'Database' section includes RDS. To the right, there are sections for Developer Tools, Internet of Things, Game Development, Mobile Services, Application Services, and Security & Identity. A 'Resource Groups' sidebar provides information on creating resource groups and includes 'Create a Group' and 'Tag Editor' buttons. Additional resources like Getting Started, AWS Console Mobile App, AWS Marketplace, and AWS re:Invent Announcements are also listed.

2) Click -> Security Group -> Create New Security Group



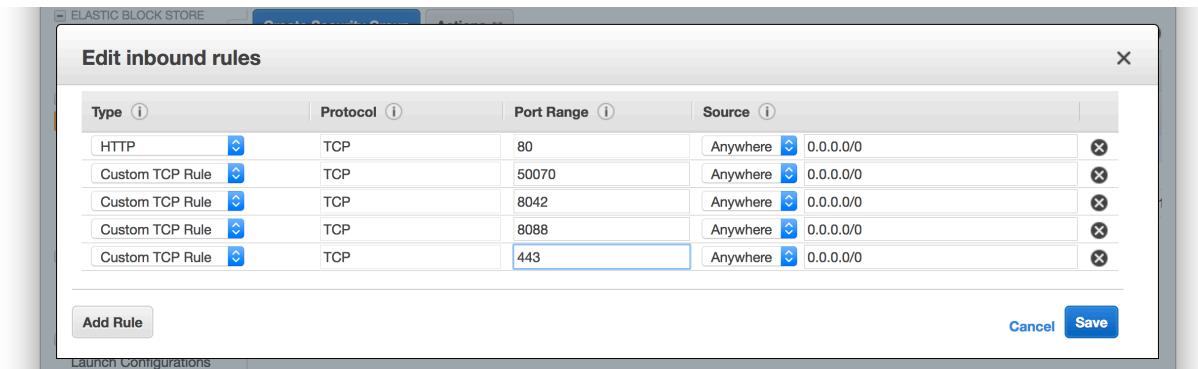
The screenshot shows the AWS EC2 Resources page. On the left, the 'NETWORK & SECURITY' section is expanded, showing 'Security Groups' which is highlighted in orange. Other options in this section include Elastic IPs, Placement Groups, and Key Pairs. The main content area displays 'Resources' for the US East (N. Virginia) region, showing 2 Running Instances, 0 Dedicated Hosts, 7 Volumes, and 2 Keypairs. It also shows 0 Elastic IPs, 1 Snapshots, 0 Load Balancers, and 3 Security Groups. On the right, 'Account Attributes' are listed, including Supported Platforms (VPC), Default VPC (vpc-3e054d5a), and a QR code.

2.1) Name Security Group - I use "EC2 Hadoop"



The screenshot shows the 'Create Security Group' dialog box. The 'Security group name' field is filled with 'EC2 Hadoop'. The 'Description' field is empty. The 'VPC' dropdown is set to 'vpc-3e054d5a (172.31.0.0/16) \*', where the asterisk denotes it as the default VPC. The dialog has a standard 'Create' and 'Cancel' button layout.

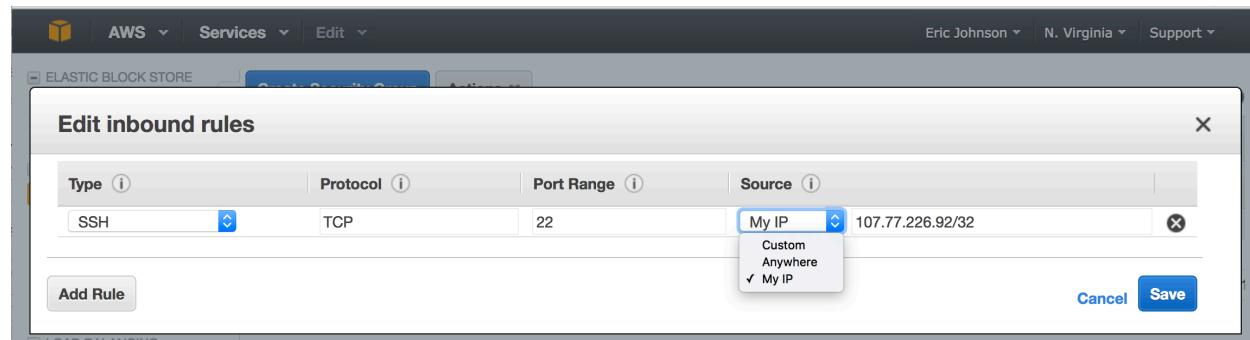
2.2) Click "Add Rule" under "Inbound" and do the following:



HTTP: 80

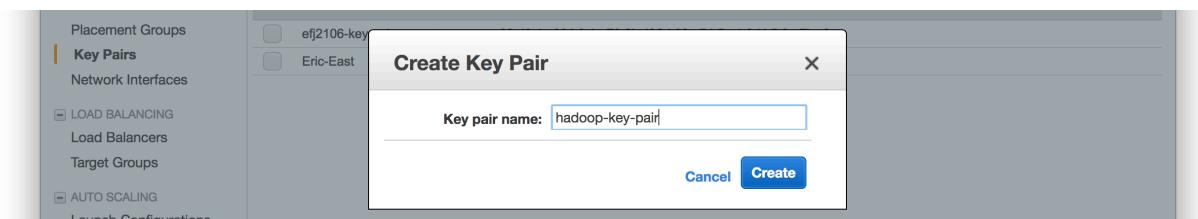
TCP: 50070, 8042, 8088, 443

As well as Port 22 for your IP to SSH



\*\* Note: Your IP will change each unique Wi-Fi network that you use to SSH into your machine. Because of this if you go to a new location and try ssh-ing in you will not be able to. In order to fix this, you need to do "My IP" for each new Wi-Fi address you use to access \*\*

3) Create a Key Pair -> Call "hadoop-key-pair"



File will download to your browser -> named: hadoop-key-pair.pem

\*\* Save this somewhere safe - you will need it whenever you SSH into your machine \*\*

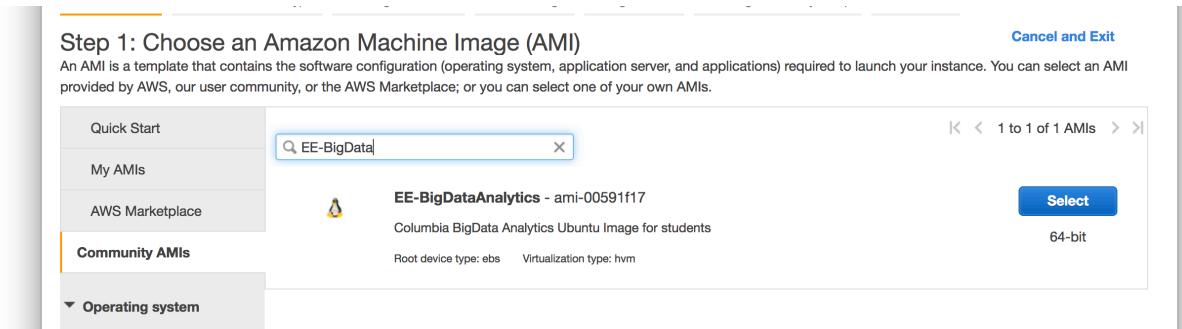
Change the permission of your ".pem" file by using the terminal window and executing the following line:

```
chmod 400 hadoop-key-pair.pem
```

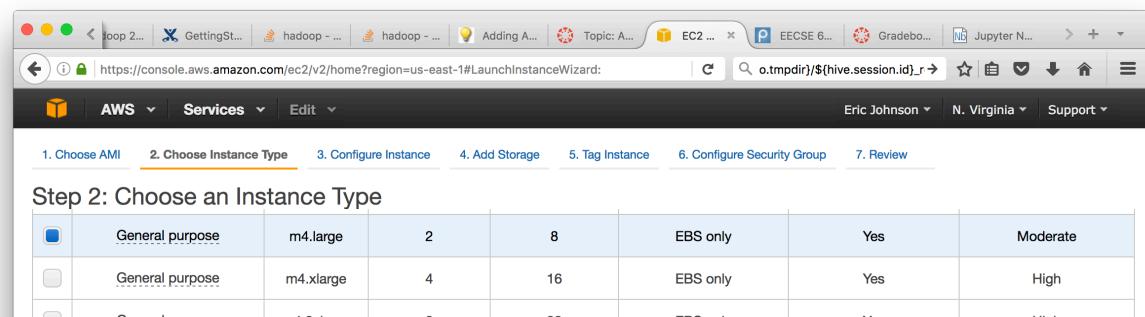
3) Click "Launch Instance"



4) Search AMI Community for "EE-BigData"



5) Recommend to use "m4.large" or "m4.xlarge"



- 6) Skip through and name your image for easy reference  
\*\*you may want to adjust storage - make 30+ GB

Step 5: Tag Instance

A tag consists of a case-sensitive key-value pair. For example, you could define a tag with key = Name and value = Webserver. [Learn more](#) about tagging your Amazon EC2 resources.

Key (127 characters maximum)	Value (255 characters maximum)
Name	AWS-Hadoop-Image
<a href="#">Create Tag</a> (Up to 50 tags maximum)	

- 7) Assign your security group using "existing security group" and "EC2 Hadoop"

The screenshot shows the AWS Launch Instance Wizard at Step 6: Configure Security Group. The user has selected the 'Existing security group' option and chosen 'sg-3760314d' named 'EC2 Hadoop'. Below this, a table shows the security group's rules:

Type	Protocol	Port Range	Source
HTTP	TCP	80	0.0.0.0/0
Custom TCP Rule	TCP	50070	0.0.0.0/0

At the bottom, there are 'Cancel', 'Previous', and 'Review and Launch' buttons. The status bar at the bottom includes 'Feedback', 'English', 'Find in page', 'Highlight All', and 'Match Case'.

# Big Data Analytics

## AWS Image Setup Guide

Ching-Yung Lin  
Eric Johnson

### Launch Instance

The screenshot shows the AWS Launch Instance Wizard at Step 7: Review Instance Launch. It displays the following information:

- Step 7: Review Instance Launch**
- AMI Details:** EE-BigDataAnalytics - ami-00591f17 (Columbia BigData Analytics Ubuntu Image for students)
- Instance Type:** m4.large (6.5 ECUs, 2 vCPUs, 8 GiB Memory, EBS only storage, EBS-Optimized Available, Network Performance Moderate)
- Security Group:** EC2 Hadoop (open to the world)
- Launch Options:** Buttons for Cancel, Previous, and Launch.

Two warning messages are present:

- Your instance configuration is not eligible for the free usage tier.** To launch an instance that's eligible for the free usage tier, check your AMI selection, instance type, configuration options, or storage devices. Learn more about [free usage tier](#) eligibility and usage restrictions.
- Improve your instances' security.** Your security group, EC2 Hadoop, is open to the world. Your instances may be accessible from any IP address. We recommend that you update your security group rules to allow access from known IP addresses only. You can also open additional ports in your security group to facilitate access to the application or service you're running, e.g., HTTP (80) for web servers. [Edit security groups](#)

8) In the EC2 Dashboard select your instance and then click "Connect"

The screenshot shows the EC2 Dashboard with the following interface elements:

- Left sidebar:** EC2 Dashboard, Events, Tags, Reports, Limits, Instances (selected), Spot Requests, Reserved Instances, Scheduled Instances, Dedicated Hosts, Images (AMIs), Bundle Tasks, Elastic Block Store (Volumes, Snapshots), Network & Security (Security Groups).
- Center panel:** A modal window titled "Connect To Your Instance".
  - I would like to connect with:** A radio button is selected for "A standalone SSH client".
  - To access your instance:**
    - Open an SSH client. (find out how to [connect using PuTTY](#))
    - Locate your private key file (efj2106-key-pair-us-east.pem). The wizard automatically detects the key you used to launch the instance.
    - Your key must not be publicly viewable for SSH to work. Use this command if needed:  
`chmod 400 efj2106-key-pair-us-east.pem`
  - Example:**  
`ssh -i "efj2106-key-pair-us-east.pem" root@ec2-54-210-3-213.compute-1.amazonaws.com`
- Right panel:** A list of instances with their status (running, stopped, etc.) and counts.

In a blank terminal window copy and paste the last line (must be in the same directory of where you saved your "hadoop-key-pair.pem" file

\*\* before you hit enter change the name "root" to "hadoop"



If asked for a password enter: 'none'

9) Once in the machine execute the following commands in order:

```
ssh-keygen -f "/home/hadoop/.ssh/known_hosts" -R 0.0.0.0
ssh-keygen -f "/home/hadoop/.ssh/known_hosts" -R localhost

su hduser (if asked for password enter 'none')

ssh-keygen -t rsa (if prompted just keep hitting enter and overwrite Y)
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
su hadoop (if asked for password enter 'none')

ssh-keygen -t rsa (if prompted just keep hitting enter and overwrite Y)
cat ~/.ssh/id_rsa.pub >> ~/.ssh/authorized_keys
```

Finally, whenever you log into your machine you will need to execute the following two commands to set up your local environment variables:

```
source ~/.profile
source ~/.bashrc
```

**CONGRATULATIONS!! Your system is now setup and configured for your use**

You can use the following predefined commands that have been set up for your convenience

-- Starting Hadoop

namenode\_start

yarn\_start

-- Format HDFS

format\_namenode

-- Stopping Hadoop

yarn\_stop

namenode\_stop

-- Launch PIG

pig

-- *Launch HIVE*

**hive**

-- *Launch/Stop HBase*

**hbase\_start**

**hbase\_stop**