Tutorial:
Using HortonWorks Sandbox
on Amazon Web Services

1   Overview

Welcome

Before diving into Cloud Applications, we need to set up the environment for doing tutorials or programming assignments. This course uses an all-in-one virtual machine made by Hortonworks. This tutorial covers the critical skills needed to work with this VM on Amazon Web Service (AWS).

Objectives

Upon completing this tutorial, students will be able to:

- Set up an all-in-one Hadoop installation on AWS
- Start, Stop, and Terminate the Virtual Machine
- Harden the Virtual Machine to prevent unauthorized accesses
- Install "nano" Text Editor
- Connect to the VM through SSH

2   Requirements

AWS Account

You will need to sign up for an AWS Educate account. You'll get access upon a successful MP1 submission. For more information, visit the following address:

http://aws.amazon.com/education/awseducate/

SSH Client

For command line steps, you need a SSH client:

- For Linux and OS X, you should already have it installed.
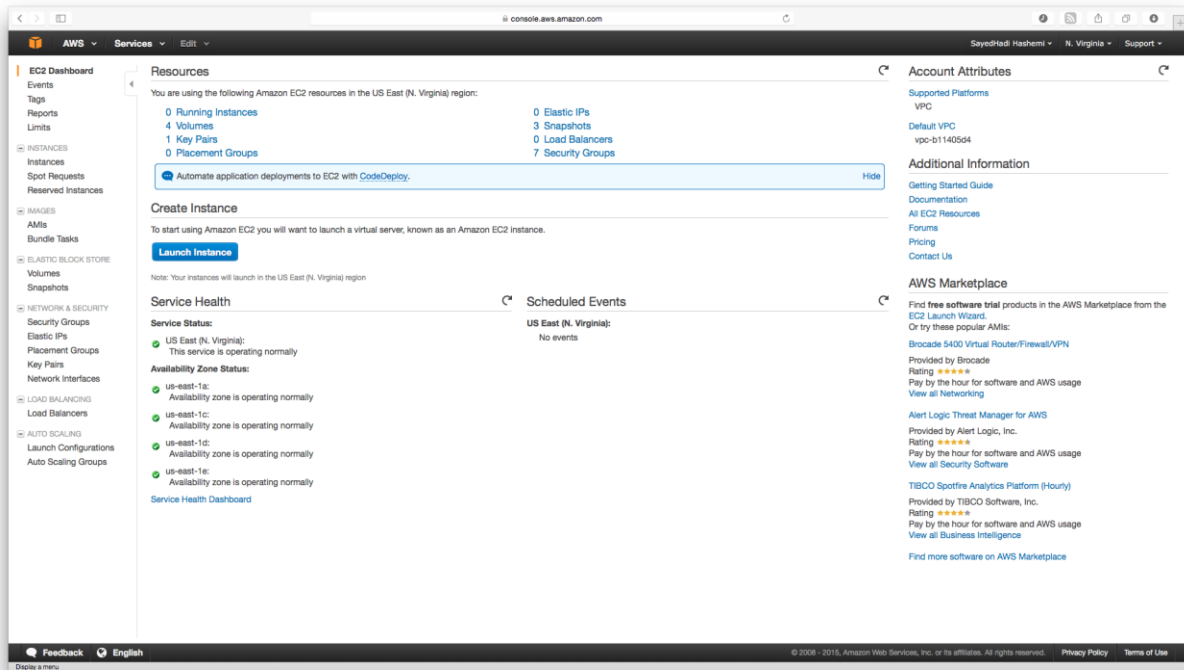- For Windows, you can download a free copy of the **Putty** from the following URL.

http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html

3 Setup Hadoop Virtual Machine

Step 1: Log in to the AWS EC2 Management Dashboard using the following link:

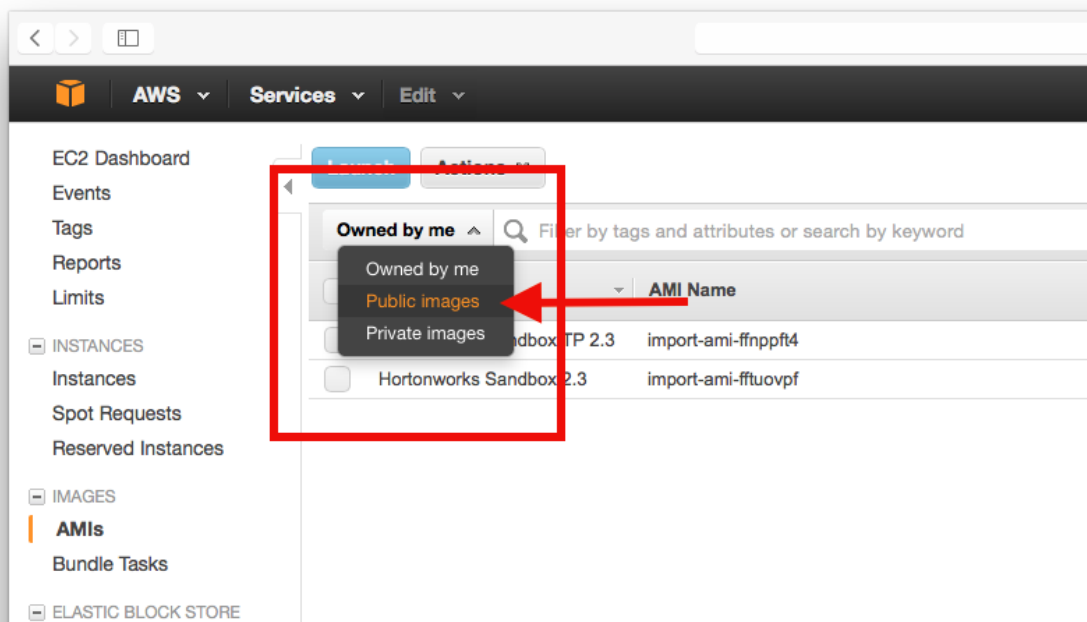🌐   https://console.aws.amazon.com/ec2/v2/home?region=us-east-1

💡   For more information about this AWS, please refer to:

http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/concepts.html

Step 2: Make sure your current region is set to **US East**. This information can be checked in the top navigation menu.
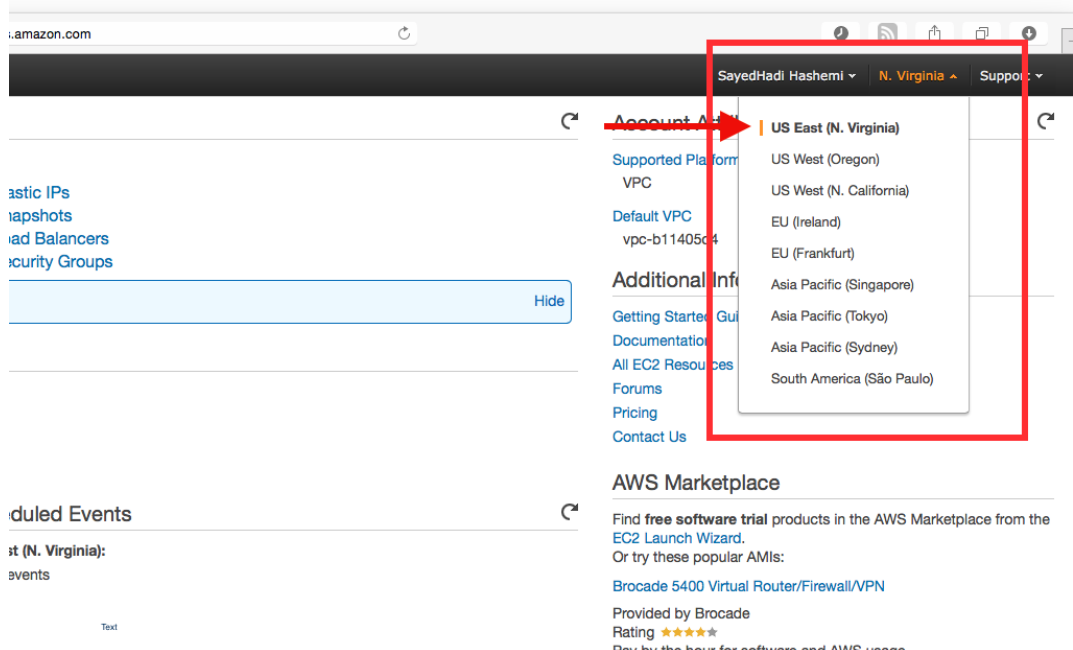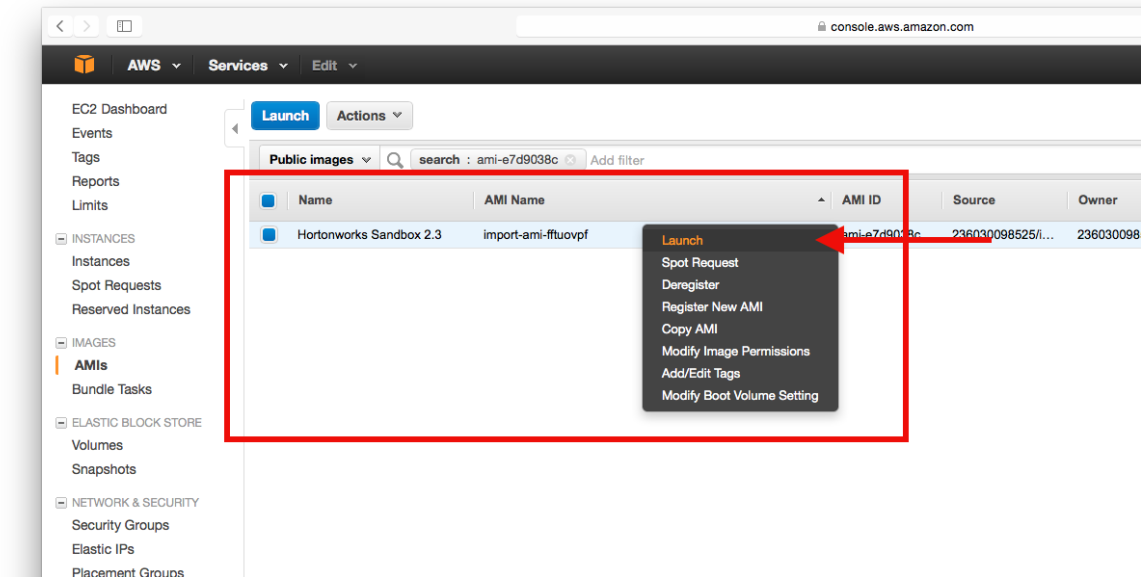
Step 3: Open **AMIs** from the left panel; then select **Public images**.

Step 4: Search for the following image. Then right click on the image and select **Launch.**

```
ami-e7d9038c
```

Step 5: Select **c3.xlarge** for Instance Type, and then click on **Review and Launch**.

Step 6: Select **Edit security groups** and add the following three additional rules. Then click on **Review and Launch**:

Type	Protocol	Port Range	Source
Custom TCP Rule	TCP	8080	Anywhere
HTTP	TCP	80	Anywhere
HTTPS	TCP	443	Anywhere
SSH	TCP	22	Anywhere

Step 5: Click on **Launch.** When prompted, select **"Proceed without a key pair"** and continue until you see the confirmation.

4   Start HDP Virtual Machine

Step 1: From the main windows of **AWS EC2,** open **Instances** from the left panel. Alternatively, use the following link:

🌍   https://console.aws.amazon.com/ec2/v2/home?region=us-east-1#Instances
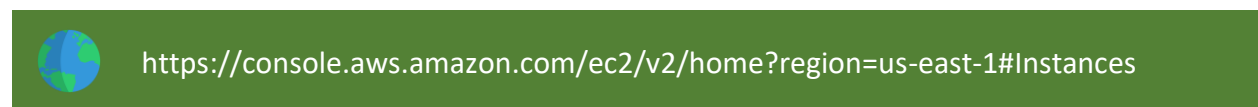
Step 2: Make sure your current region is set to **US East**. This information can be checked in the top navigation menu.

Step 3: Right click on the instance and select **Instance State > Start.**

Step 4: Wait for a few moments until you see that you have passed all the **Status Checks**. Note the value under **Public IP**.

5   Connect to the Virtual Machine Using SSH

Step 1 (OS X, Linux): Open the terminal on your local machine, and log in to the virtual machine via SSH protocol. Use the **Public IP** Address you noted earlier from AWS EC2 Dashboard. The default password is **hadoop**.

```
# ssh root@<Public IP>
```

Step 1 (Windows): Open **Putty** on your machine, and log in to the virtual machine via SSH protocol using the following information:

Server	<Public IP>
Port	22
Username	root
Password	hadoop

Step 2: After successfully logging in, you should see a prompt similar to the following:

```
[root@sandbox ~]#
```

Ignore the leading # in the commands.
It is simply an indicator that the command has to run in a terminal.

6   Connect to the Virtual Machine Using HTTP

Step 1: Open a web browser on your local machine and browse to the following URL. Replace the **<Public IP>** with the **Public IP** Address you noted earlier from AWS EC2 Dashboard. The default username/password is **admin/admin**.

```
http://<Public IP>:8080/
```

Step 2: After successfully logging in, you should see a screen similar to the following:

7 Install "nano" Text Editor

If you are new to Linux, you might find it challenging to use the default text editor in the terminal. Therefore, it is recommended for you to use **nano.** Unfortunately, nano is not installed by default in the HDP Sandbox. Fortunately, it is relatively easy to install.

Step 1: Run this command and follow the installation instructions:

```
# yum install nano
```

Step 2: After the installation is finished, check the installation:

```
# nano
```

Step 3: Quit **nano**.

8 Stop HDP Virtual Machine

Step 1: Run the following command from the SSH Terminal:

```
# poweroff
```

Step 1 (Alternative): From the main windows of **AWS EC2** open **Instances** from the left panel. Make sure your current region is set to **US East**. This information can be checked in the top navigation menu. Right click on the instance, and select **Instance State > Stop.**

Step 2: Wait for a few moments until the VM has fully shut down.

> ⚠️ You are billed hourly on AWS. Always stop the instance as soon as you are finished. Otherwise, you might run out of credit to use the instances.

9   Stop HDP Virtual Machine

When you are completely finished, you can fully remove the instance from your AWS account. After that, you will not be able to access to instance data or recover it.

Step 1: From the main windows of **AWS EC2** open **Instances** from the left panel. Make sure your current region is set to **US East**. This information can be checked in the top navigation menu. Right click on the instance, and select **Instance State > Terminate.**

Step 2: Wait for a few moments until the VM has been fully removed.

10 Make Virtual Machine Safer and More Secure (optional)

Since the VM will be accessible to the Internet, additional **essential** steps are needed to ensure the security and safety of VM.

10.1 Change Default Passwords

Step 1: Start the virtual machine, and then connect to it through the SSH.

Step 2: Change the default **root** password using the following command:

```
# passwd root
```

> 💡 Don't forget this password. If you do, you might need to initiate a new instance.

Step 3: Change the default Hadoop password using the following command:

```
# passwd Hadoop
```

Step 4: Open a web browser on your local machine and browse to the following URL. Replace the **<Public IP>** with the **Public IP** Address you noted earlier from AWS EC2 Dashboard. The default username/password is **admin/admin**.

```
http://<Public IP>:8080/views/ADMIN_VIEW/2.1.0/INSTANCE/#/
```

Step 5: From the right panel select "**Users**". Then click on "**admin**" user in the main panel. After that, use "**Change Password"** to change the admin password.

Step 6: Stop the Virtual Machine.

10.2 Alarms

Step 1: From the main windows of **AWS EC2,** open **Instances** from the left panel. Alternatively use the following link:

🌍   https://console.aws.amazon.com/ec2/v2/home?region=us-east-1#Instances

Step 2: Make sure your current region is set to **US East**. This information can be checked in the top navigation menu.

Step 3: Right click on the instance, and select **CloudWatch Monitoring > Add/Edit Alarms.** Then, click on **Create Alarm.**

Step 3: Right click on the instance, and select **CloudWatch Monitoring > Add/Edit Alarms.** Then click on **Create Alarm.**

Step 4 (IDLE Protection): Enter the following values, and then click **Create Alarm.**

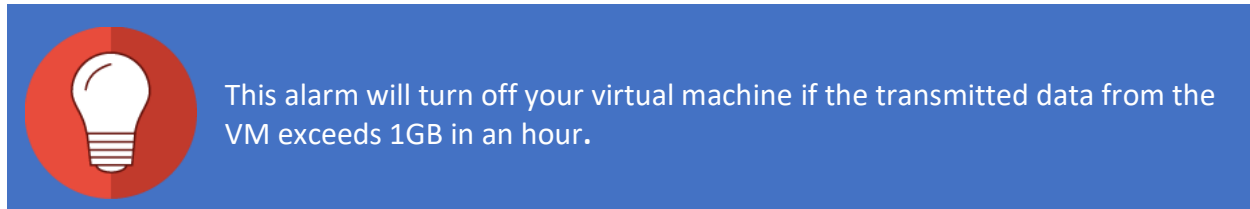|	
Send Notification to	Select your email address, or add yours by clicking on **create topic**.
Take Action	Stop this instance
Whenever	**Sum** of **CPU Utilization**
Is	>= 0 percent
For at least	**6** consecutive period(s) of **1 hour**
Name of alarm	IDLE Protection

💡 This alarm will turn off your virtual machine every **5** hours to protect your credits from being wasted on an idle virtual machine. If 5 hours is too little for you, feel free change it to a more convenient value.

Step 3: Right click on the instance, and select **CloudWatch Monitoring > Add/Edit Alarms.** Then click on **Create Alarm.**

Step 4 (Zombie Protection): Enter the following values, and then click **Create Alarm.**
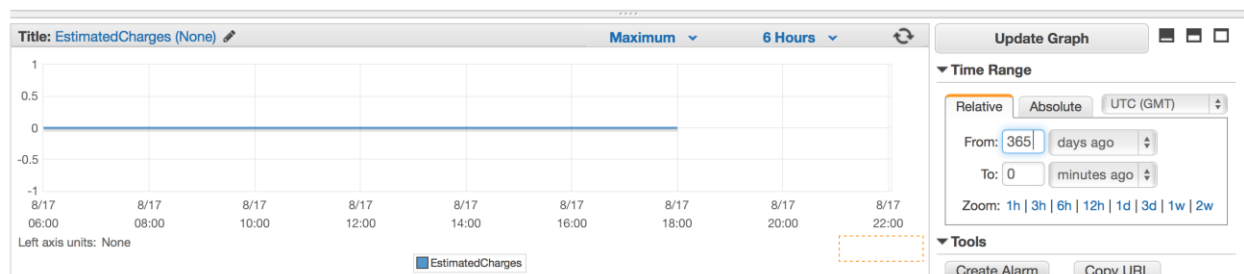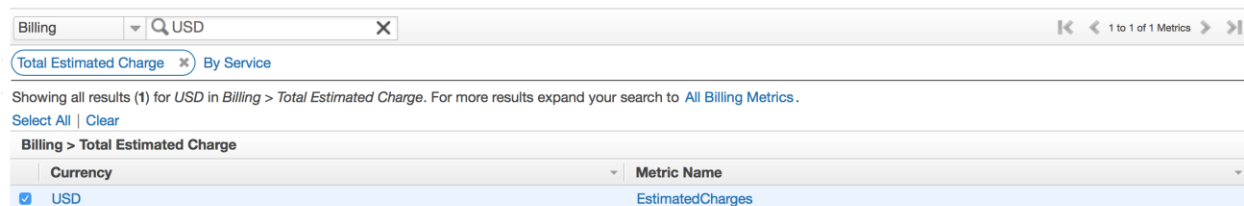
Send Notification to	Select your email address, or add yours by clicking on **create topic**.
Take Action	Stop this instance
Whenever	**Sum** of **Network Out**
Is	>= 1000000000 bytes
For at least	**1** consecutive period(s) of **1 hour**
Name of alarm	Zombie Protection

This alarm will turn off your virtual machine if the transmitted data from the VM exceeds 1GB in an hour.

Step 3: From the top panel select **Services > CloudWatch.** Then click on **Create Alarm**.

Step 4 (Billing Warning): Select **Total Estimated Charge**, then check the **USD.** In the bottom panel, enter the following values. Finally, click on **Create Alarm.**

Time Range	Relative
From	365 Days ago
To	0 minutes ago

Step 5: In the next screen, enter the following values**, and** then click on **Create Alarm**.

Exceed	40 USD

send a notification to	Select your email address, or add yours by clicking on **create topic.**

This alarm will only notify you when your bill is about to exceed your AWS credit. It is still your responsibility to release AWS resources before your credit completely runs out.

11 Best Practices

In addition to the essential steps described in the previous section, there are a few more recommended steps to increase the security level of the AWS and VM.

11.1 AWS Authentication

To reduce the risk of someone taking control of your AWS account, it is recommended that you enable Multi Factor Authentication on your account.

Refer to the following URL for more information:
https://aws.amazon.com/iam/details/mfa/

11.2 Public Key Authentication

There are many ways to improve the security of the SSH service on your VM. Disable password login and Public Key Authentication are some of the well-know mechanisms.

Refer to the following URL for more information:
http://wiki.centos.org/HowTos/Network/SecuringSSH

11.3 Patch Management

It is highly recommended that you keep the VM Operating System and applications updated.

Refer to the following URL for more information:
https://www.centos.org/docs/5/html/yum/sn-updating-your-system.html