Cloudera Administrator Training for Apache Hadoop: EC2 and VMware Setup for Exercises

Setup Preparation

Training Environment

For these Hands-On Exercises, you will use a cloud training environment: five Amazon EC2 instances plus a VMware virtual machine.

Prerequisites for the Training Environment

For the cloud training environment, you need the following:

- An Amazon AWS account.
- A EC2 security group in the VPC that has all TCP, UDP, and ICMP ports open to other instances in the same security group, and has ports 22, 80 and 443 open to external access:
    - To create EC2 instances in a VPC, the ports must be open to inbound *and outbound* traffic. Be sure that the Inbound tab for your EC2 security group (or Inbound and Outbound tabs, if you are running VPC instances), appears as follows:

Type	Protocol	Port Range	Source
HTTP	TCP	80	0.0.0.0/0
All traffic	All	All	sg-07bcb162 (Admin VPC2)
SSH	TCP	22	0.0.0.0/0
HTTPS	TCP	443	0.0.0.0/0

In the preceding screen shot, sg-07bcb162 is the security group ID.

- A computer with:
    - At least 4GB of RAM
    - At least 10GB free disk space

- Support for the VT-x virtualization feature. This support must be enabled in the BIOS.
- VMware Player 6.x or above (Windows)/VMware Fusion 6.x or above (Mac)

Training Environment Setup

If you have never used Amazon EC2 before, we suggest you read one of the many tutorials available on the Web first. We regret that we cannot provide support if you choose to recreate the class environment yourself using EC2.

Launching EC2 instances will result in you being charged by Amazon for every hour each EC2 instance is running. This will be billed to the credit card associated with your EC2 account. ***It is your responsibility to ensure that you terminate the instances when you have finished using them. Cloudera is not responsible for any charges you may incur while using Amazon EC2***.

Perform the following steps:

1. Unzip the `Cloudera-Training-Get2EC2-VM-1.2-vmware.zip` file containing the Get2EC2 VM. You will use this VM to connect to EC2 instances.

   The unzip operation creates a directory containing the files for a virtual machine. The directory has a name similar to `Cloudera-Training-Get2EC2-VM-`*`n.n`*`-vmware`.

2. Move the virtual machine directory to any location you want.

3. Start the virtual machine.

4. Next, you will create five t2.large EC2 instances for the hosts `elephant`, `tiger`, `horse`, `monkey`, and `lion`.

   Start by logging in to the AWS Management Console.

   Note: If you are familiar with the AWS command-line interface, you can create your instances using command-line tools. Refer to "Using AWS Command-Line Tools" at the end of this section for examples.

5. Select **Services** > **EC2** to navigate to the EC2 Dashboard.

6. From the top-right menu, select the region closest to you (note that we do not have AMIs available in all regions – see the list below).

7. Click **Launch Instance**.

   The Step 1: Choose an Amazon Machine Image (AMI) page appears.

8. Choose the **Community AMIs** tab, and search for the AMI for the region you have selected:
   - Northern California – `ami-88fbaae8`
   - Virginia – `ami-c455b0d2`
   - Ireland – `ami-d0664fa3`
   - Tokyo – `ami-d8dcacbf`
   - Singapore – `ami-ccbb12af`

9. When the AMI information appears (it may take a minute or two to appear), choose **Select**.

10. On the next screen, choose the instance type **`t2.large`**, which is the desired type for these five instances.

11. Click **Next: Configure Instance Details** to accept the defaults.

    The Configure Instance Details page appears.

12. Specify **`5`** in the Number of Instances field.

13. For Network, choose an existing VPC network or click the link to create a new VPC if you wish. The default VPC is usually fine.

14. Choose an existing subnet in the VPC or click the link to create a new subnet.

15. For Auto-assign Public IP choose **Enable**.

16. Once you have completed the settings above, click **Next: Add Storage**

    The Step 4: Add Storage page appears.

17. Don't change anything on the 'Add Storage Configuration' page, just click **Next: Tag Instance**.

    The Step 5: Tag Instance page appears.

    No instance tags are required, but you can apply any tags you want to the instances.

18. Click **Next: Configure Security Group**.

    The Step 6: Configure Security Group page appears.

19. Click **Select an Existing Security Group**.

20. You should have created a security group according to the instructions in the section, "Prerequisites for the Training Environment." Select the check box next to this security group.

21. Click **Review and Launch**.

    The Step 7: Review Instance Launch page appears.

    Review the details on the page and make sure that the instance details are correct.

22. Click **Launch**.

    The "Select an existing key pair or create a new key pair" dialog box appears.

23. Select **Proceed without a key pair**.

24. Select the check box to acknowledge that you will not be able to connect to the instance without knowing the password built in to this AMI.

25. Click **Launch Instances**.

26. Click **View Instances**.

    The EC2 Dashboard appears.

27. Wait for all five instances to finish launching. The Instance State should have the value "running," and the Status Checks field should have the value "2/2 checks passed."

28. Note the public and private IP addresses for your five EC2 instances.

    To get the IP addresses for an EC2 instance, select the instance in the EC2 Dashboard. The instance information, including the public and private IP address, appears at the bottom of the page.

29. Proceed to the "Setup Activity: Configuring Networking" section in the Hands-On Exercises Manual.

Using Spot Instances

You can use spot instances for these exercises if you like, but be aware that because spot instances can be terminated at any time, you risk losing any work you have done in EC2.

Using AWS Command Line Tools to Create Instances

If you are an experienced AWS user and prefer to create the five EC2 instances using the AWS command line, you can do so. The following are example commands similar to commands you might use to create EC2 instances.

Create five t2.large EC2 instances (for `elephant`, `tiger`, `horse`, `monkey`, and `lion`) in the Northern California region:

```
$ ec2-run-instances ami-88fbaae8 -n 5 \
-g <MySecurityGroup> --region us-west-1 -t t2.large
```

Tailor the preceding example commands by specifying:

- Your desired AWS region.
- Your security group name.

- The appropriate AMI. Refer to the “Cloud Training Environment Setup” section for the AMI IDs.