Cloudera Administrator Training for Apache Hadoop

Using the Amazon EC2 Machine Images

Your training course used virtual machines created in Amazon's EC2 cloud. If you want to reproduce the exercise environment yourself, you can: we have made the AMIs (Amazon Machine Images) public, so if you have an Amazon Web Services account you can start up the VMs using your own account. Here's how to do that.

Disclaimer: 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.

NOTE: Launching EC2 Virtual Machines will result in you being charged by Amazon for every hour each EC2 instance is running. Even if you leave instances in a 'Stopped' state, you will still be charged by Amazon for storage. This will be billed to the credit card associated with your EC2 account. It is your responsibility to ensure that you shut down and delete the instances when you have finished using them. Cloudera is not responsible for any charges you may incur while using Amazon EC2.

1. Register for an Amazon AWS account if you don't already have one. The address is http://aws.amazon.com.
2. From the AWS Management Console, chose EC2.
3. 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).
4. Select 'Instances' from the menu on the left.
5. Click the 'Launch Instance' button.
6. Choose the 'Community APIs' tab, and search for the AMI for the region you have selected:
    - Northern California – ami-bbd610ff
    - Virginia – ami-672b7d02
    - Ireland – ami-2be8da5c
    - Tokyo – ami-1a3f5f1a
    - Sydney – ami-b3581289
7. When the AMI information appears (it may take a minute or two to appear), choose 'Select'.
8. On the next screen, you can specify the size of each instance. For the Administrator course exercises we suggest the 't2.medium' instance size.
9. Click 'Next: Configure Instance Details'.

10. In the 'Configure Instance Details' screen:
    a. Specify five instances.
    b. 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.
    c. Choose an existing subnet in the VPC or click the link to create a new subnet.
    d. Once you have completed the settings above, click 'Next: Add Storage'.
11. Don't change anything on the 'Add Storage Configuration' page, just hit 'Next: Tag Instance'.
12. On the next screen, you can optionally specify a name for your instances, such as 'Cloudera Administrator exercises'. Then hit 'Next: Configure Security Group'.
13. In the 'Configure Security Group' screen, either choose an existing VPC Security Group or define a new one. Regardless of which option you choose, make sure that ports 80 and 443 are accessible from outside EC2, and that all TCP, ICMP, and UDP ports are open within the VPC subnet. Also make sure the VPC assigns external IP addresses to the instances within it.
14. Click 'Review and Launch'.
15. After reviewing the details, click 'Launch'.
16. On the next screen, you select key pairs. Choose 'Proceed without a Key Pair', since you will be accessing the instances via the Get2EC2 virtual machine on your local computer. Hit 'Launch Instances', then click 'View Instances'
17. Launching the instances usually takes a minute or two. Once launched, you can use the Web UI to find the external IP addresses of the EC2 virtual machines, which you need in order to connect from the Get2EC2 VM on your local machine. **You will be charged for each EC2 instance per hour while it is running!**

Note: To access the VMs, you should download and install the Get2EC2 Virtual Machine on your local computer. You will need to install VMware Player (if you are using a PC) or VMware Fusion (if you are using a Mac) to run the local VM. You can then proceed just as you did in the classroom.

Do not forget to terminate the EC2 instances and the associated volumes from the AWS Control Panel when you have finished using them. You are charged by Amazon for every hour that each instance is running and will also be charged storage costs for stopped instances and any remaining EBS volumes.