CINF/CSCI 5931 Big Data Analytics Fall 2017

Assignment 1 MapReduce Programming Using AWS EMR-Part 1

Post Date: Sep 14th,2017 **Due Date**: Sep 28th, 2017 by 11:59pm

In this assignment, you will practice writing MapReduce programs on Amazon Web Services (AWS). You will work in **team of two** for this assignment.

Important Notes:

1. Each of you will get a $40 free credit for AWS when you sign up for Student Starter account. This amount should be sufficient to finish this assignment. It is your responsibility to check your balance frequently. You will be responsible for any overcharges occur.
2. Make sure you debug your program in local mode first before running it on AWS.
3. In order to cut unnecessary spending, you can use one student's account for exploring and testing. Then use the second one for final work.

Assignment Tutorial

1. Getting started with Amazon AWS

1.1 Signing up for AWS

1. Go to https://aws.amazon.com/, and click on Create A Free Account.
2. Register. Fill out all information fields. You must provide your credit card.
3. Verify your registration by phone. AWS will call you. When they do, enter the pin shown on the screen.
4. Once you registered an account, log into AWS Management Console (https://aws.amazon.com/console/) and then go to Credits (under Billing). Enter your credit code you received via email for AWS Educate. You should have $40 on your balance.

1.2 Apply for AWS Educate

In this step, you can receive $40 of AWS credit by applying for AWS Educate at https://aws.amazon.com/education/awseducate/apply/.

1. Click on 'Apply for AWS Educate for Students'.
2. Step 1: Choose Your Role. Select 'Student'.
3. Step 2: Fill in your information as in Figure 1. You account id can be found at 'My Account' under your user name after you log in AWS Management Console.
4. Step 3: An email address verification message will be sent to your mail box, and enter this verification code sent to your email address. Check 'I'm not Robot'.

Note: to make sure your verification emails (from AWS and piazza) are not blocked by UHCL mail server, you should do the following:

- Go to: https://www.uhcl.edu/computing/resources/documentation/spam

- Click on the link in ‘Log in to Spam Management’ (end of the first paragraph).
- You will be redirected to Pure Message Login Portal and fill out your credentials. Use your UHCL computing id and password.
- Once you are logged in, you will see a list of blocked messages. Select the emails from senders you do NOT wish to block, and then click on ‘Deliver & Approve Sender’.
- To make sure you will received emails from specific senders from now on, click the ‘Approved Senders’ tab on the left, and fill out the ‘Add Sender’ form.

5. Step 4: After one day, you will received an email titled ‘AWS Educate Application Approved’, with Credit Code for $40 of AWS credit.

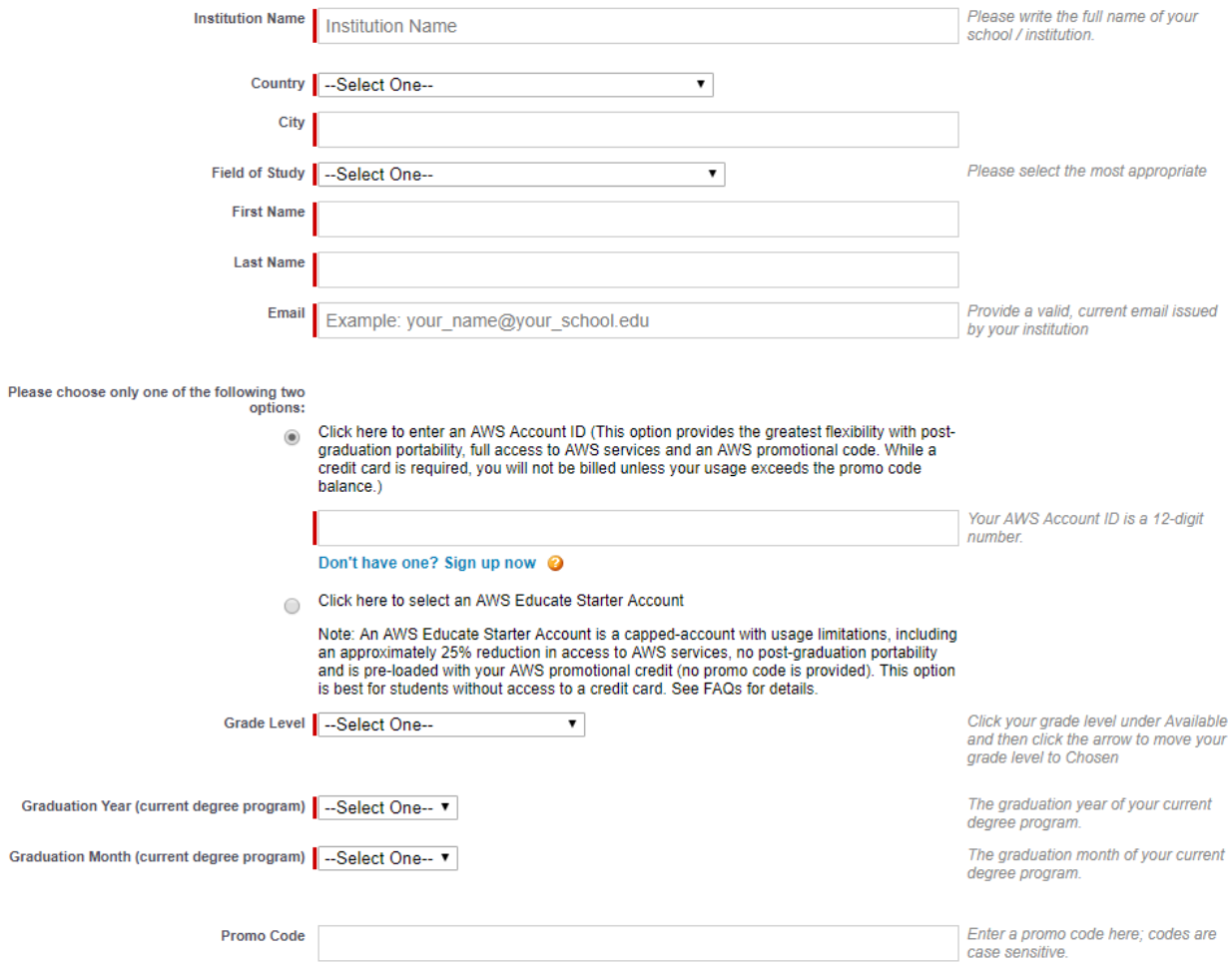

Figure 1. Apply for AWS Educate for Students Step 2.

1.3 Getting Ready for Your Assignment

1. Set availability zone: Inside the AWS Management Console, you set your availability zone to US East (Ohio). You should find this choice in the upper-right corner of the webpage in a pull-down menu (next to Support).
2. Getting access key id and secret access key: Go to AWS Security Credentials (https://console.aws.amazon.com/iam/home?#security_credential). Click on “Access Keys", and then click on “Create New Access Key". Record both your access key id and secrete access key; you will need them to configure your working environment (There is no way to retrieve an old secrete access key, but you can create a new key to use.)
3. Creating a key pair: Go the Amazon EC2 Dashboard (https://console.aws.amazon.com/ec2/) from AWS console. In the vertical menu bar on the left, choose “Key Pairs" under \NETWORK AND SECURITY". Click on “Create Key Pair", and name your new keypair“5931FA17Keypair". You will be prompted to download a file named “5931Fa17Keypair.pem". Save this file in a convenient place.

1.4 Launching an Instance

At Amazon EC2 Dashboard, you can start your first free instance by clicking on “Launch Instance'.

1. Step 1: Choose an Amazon Machine Image(AMI), you may select “Ubuntu Server 14.04 LTS(PV), SSD Volume Type".
2. Step 2: Choose an Instance Type, you may select “t2.micro" which is free. You can click“Review and Launch" or “Configure Instance Details" which follows by “Add Storage" (step 4), “Tag Instance" (step 5); “Configure Security Group" (step 6).
3. Step 7: Review and Launch.

When you return to EC2 Dashboard, you will have 1 Running Instances under “Resources". Click on “Instances", you will see the detailed information of the instance you just launch, as shown in Figure 2. You can click on “Connect " button to view how to connect to this instance using a standalone SSH, e.g. Figure 3, where your private key file is generated by following steps in Section 1.3. More details can be found at http://docs.aws.amazon.com/AWSEC2/latest/UserGuide/AccessingInstancesLinux.html including how to transfer a local file to an EC2 instance. (Note: record your Public DNS (IPv4) for future reference).

Figure 2: EC2 Dashboard

Note:

Running instance for a longer period of time will charge you. Before you move to another step, it is important to STOP the instance to avoid charges. To stop an instance, from the "Actions" tab, choose "Instance State" and click on "Stop"

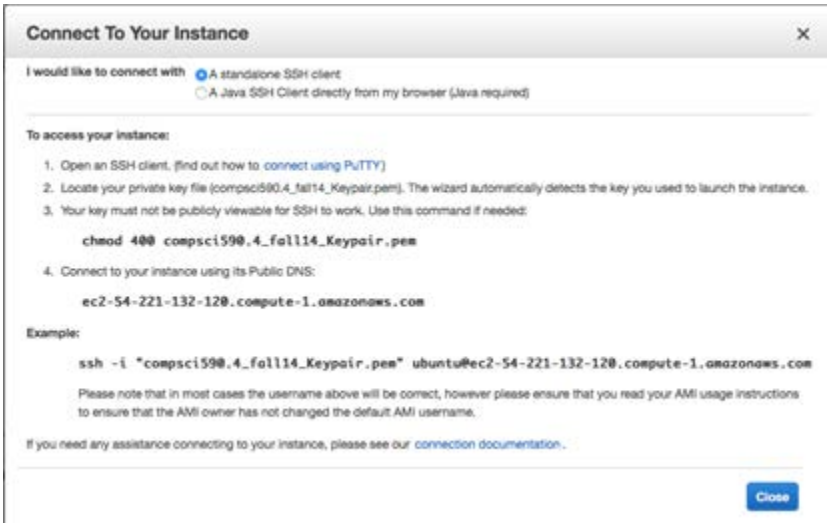

Figure 3. Connect Instance Using SSH