

Lab 1: EC2

DSCI 551 – Fall 2024

Due: 11:59pm, September 13, 2024, Friday

Points: 10

1. [8 points] Log into your EC2 using ssh (note that the connection string will change when you restart your instance. So make sure that you go to the AWS console to copy the new connection string: under Example of ssh client tab).
 - a. [2 points] There is a hidden file called `.bashrc` under your EC2 home directory. Explain what is the purpose of this file.
 - b. [2 points] Using nano text editor on EC2, add the following lines to the end of the file:
`export JAVA_HOME=/usr/lib/jvm/default-java`
`export PATH=$PATH:~/spark-3.5.2-bin-hadoop3/bin`
`export PATH=$PATH:~/hadoop-3.4.0/bin:~/hadoop-3.4.0/sbin:.`
 - c. [2 points] Explain what is the purpose of each line you added to `.bashrc` in the previous step.
 - d. [2 points] Execute `"source .bashrc"`. Explain what this command does.
2. [2 points] Install Spark by following the instructions in the handout and submit a screenshot showing the successful starting up of pyspark.

Note:

- **Ubuntu Software Installation Guide** is published in the Week 1 contents folder in D2L.
(<https://courses.uscdcn.net/d2l/le/content/28756/viewContent/499037/View?ou=28756>)
- Please follow the instructions in order and read each and every instruction. Do not skip any lines unless mentioned explicitly in the guide.

SUBMISSION INSTRUCTIONS:

1. Submit a PDF file with answers/screenshot for the above questions.
2. LATE submissions will not be accepted.
3. 10% PENALTY if submission instructions are violated.