Lab 1: EC2

DSCI 551 – Spring 2024

Due: 11:59pm, January 19, 2024, Friday

Points: 10

1. [4 points] Explain what each of the following **commands doe**s, and what each of its **arguments means.**
    1. chmod 400 "dsci2024.pem"
    2. ssh -i "dsci2024.pem" ubuntu@ec2-54-183-13-46.us-west-1.compute.amazonaws.com
    3. wget https://dlcdn.apache.org/spark/spark-3.5.0/spark-3.5.0-bin-hadoop3.tgz
    4. tar xvf spark-3.5.0-bin-hadoop3.tgz

2. [4 points] Please add the following lines to the end of ~/.bashrc file (i.e., .bashrc file under the home directory). Show a screenshot of the file with modified content [last few lines of the file]. Explain what each of the following commands does.

    export JAVA_HOME=/usr/lib/jvm/default-java
    export PATH=$PATH:~/spark-3.5.0-bin-hadoop3/bin

3. [2 points] Install Spark by following the instructions in the handout (note the handout has been updated and make sure you use the latest version on d2l) and submit a screenshot showing the successful starting up of pyspark. [screenshot should include the prompt from "**pyspark**" command]

SUBMISSION INSTRUCTIONS:

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