

Lab 01: Setup Hadoop

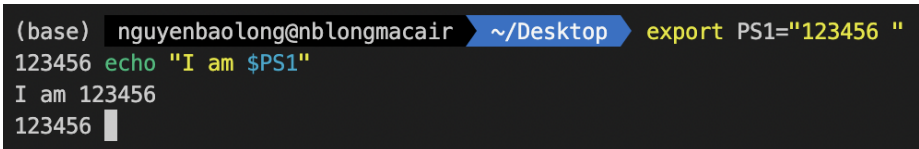
Nguyen Bao Long (@nblong)

February 21, 2023

0.1 General information

- In this assignment, you will learn:
 - How to setup an environment for running Hadoop.
 - Run an example using Hadoop.
 - Write a report in English.
- Report:
 - Write your report using the provided template. If you are not familiar with Markdown, take a look on this cheat sheet.
 - Build your *.pdf file by running `build.sh`. See this site for more information.
- Submission guideline:
 - Since Hadoop needs to be installed on laptop of each member in your team, there will be **3 reports submitted separately**. Save your report as `ID.pdf` (e.g. you id is 123456 then your submission is `123456.pdf`) and submit it on Moodle.
 - Note that each report has to be included teammates information.
- You can discuss with your teammate, but the **result must be yours**. That means you will write the report on your own (computer and words). **One copy means zero for all members**.
- Since your report will contain a lot of screenshots with terminal, your ID has to be included in the terminal. In order to do that, simply run the following command:

```
export PS1="<your id>"
```
- For example:



```
(base) nguyenbaolong@nblongmacair ~/Desktop export PS1="123456 "
123456 echo "I am $PS1"
I am 123456
123456 █
```

Figure 1: Add your ID into your terminal

1 Install Java & Hadoop

1.1 Install Java (2p)

- You should use Linux (install dual boot, don't use VMWare) for this assignment (Windows is good but I find it hard to set environment variables). Furthermore, linux has some package managers (**apt**, **sdkman**,...) that help installing things easier.
- In order to run Hadoop, your laptop has to be installed Java. I highly recommend **Java 8**. Remember to set **\$JAVA_HOME** variable.
- Summarize your installation process (attach screenshot as proof) in your report.

1.2 Install Hadoop (3p)

- You can download a Hadoop release. You should use **versions 3.3.x**. For Linux, you just have to double click on the file **hadoop-3.3.4.tar.gz** to decompress it. Then, simply put the output folder (**hadoop-3.3.4**) wherever you want your Hadoop to be installed.
- Setup Hadoop and environment for running it based on this tutorial posted on **hadoop.apache.org**.
- Summarize your installation process (attach screenshot as proof) in your report.

2 Run Hadoop (5p)

- Local mode and Pseudo-Distributed mode: Follow the instruction in the above tutorial. (4p)
- Fully-distributed mode (*Optional*): You can try this mode with your teammates. (1p)
- Summarize your running process (attach screenshot as proof) in your report.