**HADOOP STREAMING**

**2. Example Using Python WordCount :**

**+** Tạo mapper.py



**+** Tạo reducer.py:

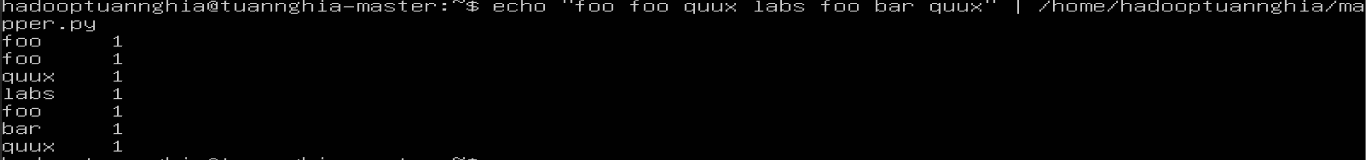


+ cấp quyền cho mapper.py và reducer.py:

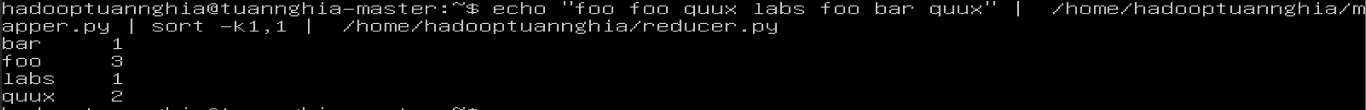


1. **Thực thi chương trình WordCount trên thư mục cục bộ:**

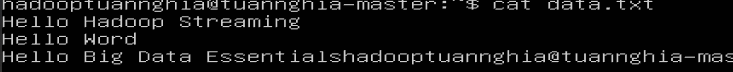
echo "foo foo quux labs foo bar quux" | /home/hadooptuannghia/mapper.py



- echo "foo foo quux labs foo bar quux" | /home/hadooptuannghia/mapper.py | sort -k1,1 | /home/hadooptuannghia/reducer.py



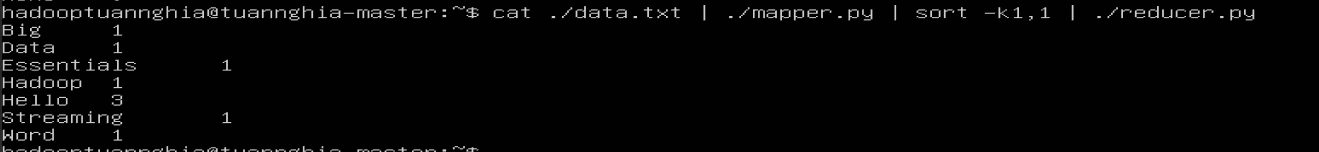
- Tạo file data.txt chứa dữ liệu:



- cat ./data.txt | ./mapper.py



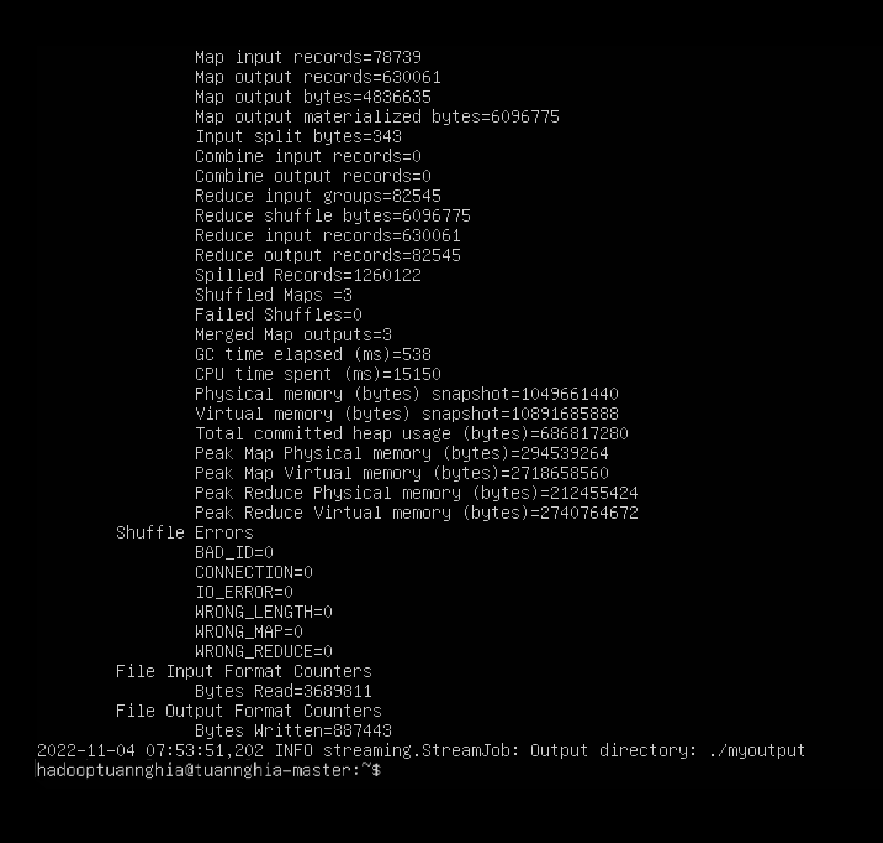
- cat ./data.txt | ./mapper.py | sort -k1,1 | ./reducer.py



1. **Thực thi chương trên WordCount tręn HDFS**- Tạo thư mục myinput chứa dữ liệu

- Chạy MapReduce job

$ hadoop jar hadoop-streaming-3.3.4.jar -file mapper.py -mapper mapper.py -file reducer.py -reducer reducer.py -input ./myinput -output ./myoutput



- Hiển thị kết quả

$ hdfs dfs -cat ./myoutput/part-00000

