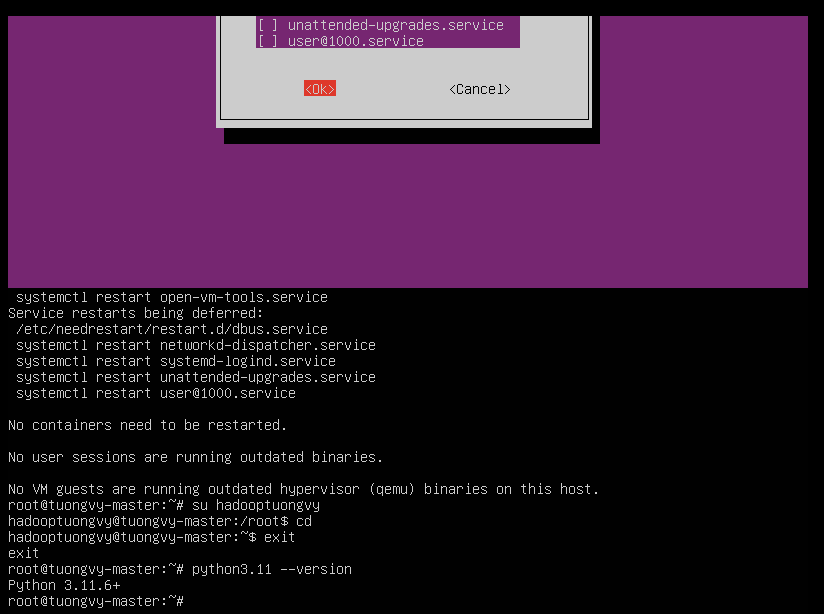
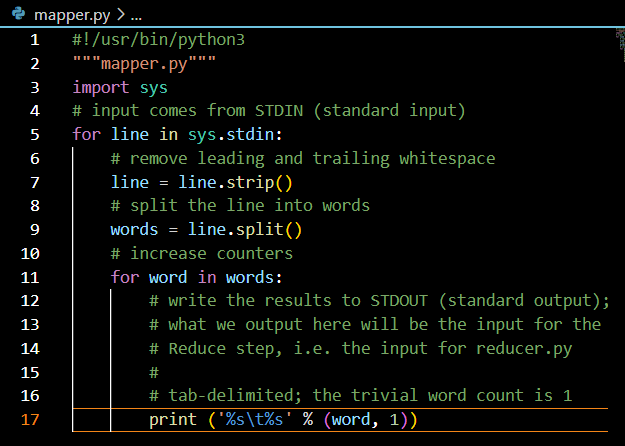
**HADOOP STREAMING**

1. **Install Python**

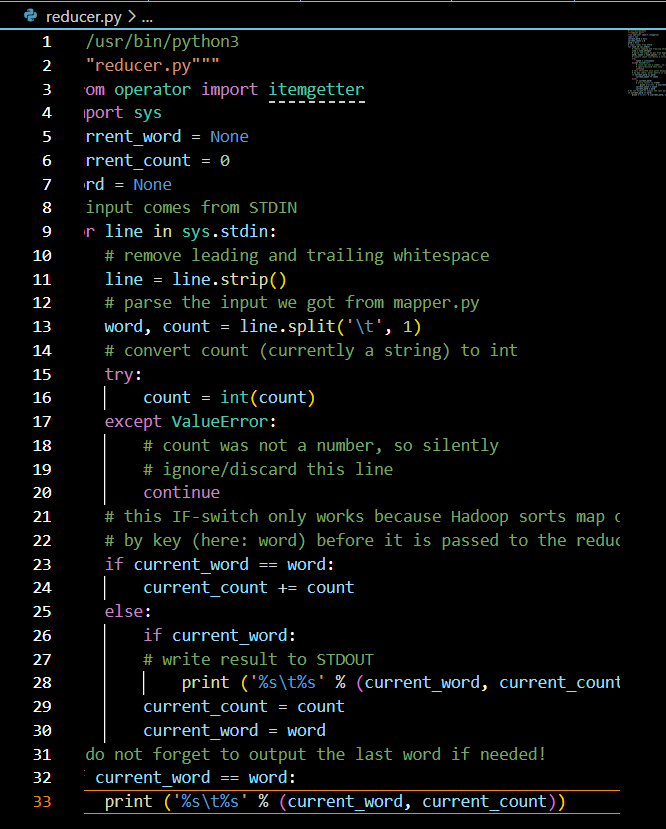


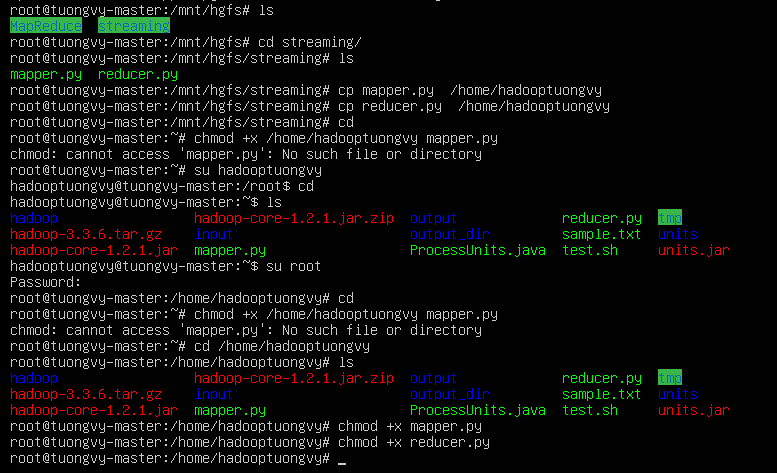
1. **Example Using Python WordCount**

**Mapper Phase Code**



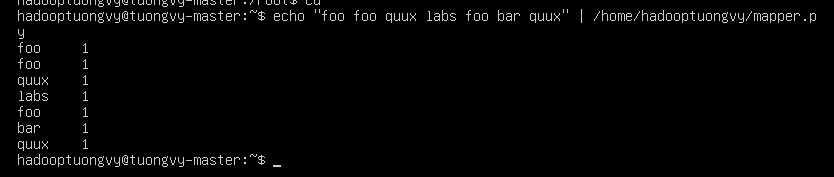
**Reducer Phase Code**



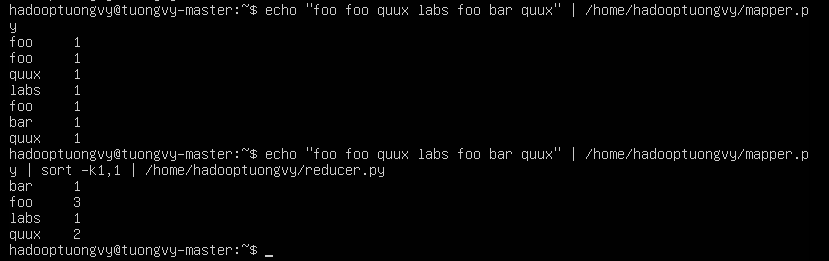


1. **Thực thi chương trình WordCount trên thư mục cục bộ**

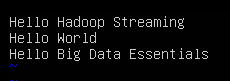
$ echo "foo foo quux labs foo bar quux" | /home/hadoopminhchau/mapper.py



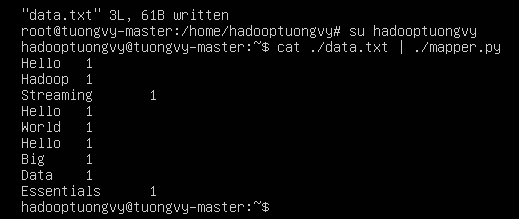
$ echo "foo foo quux labs foo bar quux" | /home/hadoopminhchau/mapper.py | sort -k1,1 | /home/hadoopminhchau/reducer.py



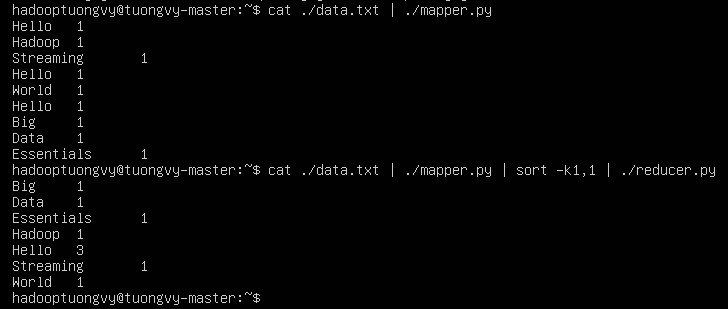
**Tao file data.txt**



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



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



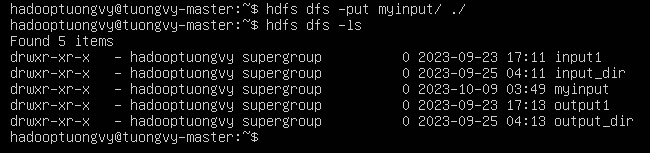
1. **Thực thi chương trình WordCount trên HDFS**

**Tạo thư mục myinput chứa dữ liệu**





**Copy thư mục myinput vào HDFS**



**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

