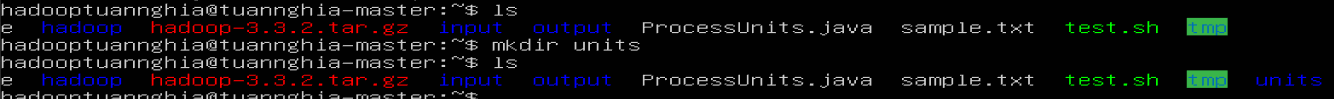
II. Tạo bộ dữ liệu

III. Tạo chương trình MapReduce



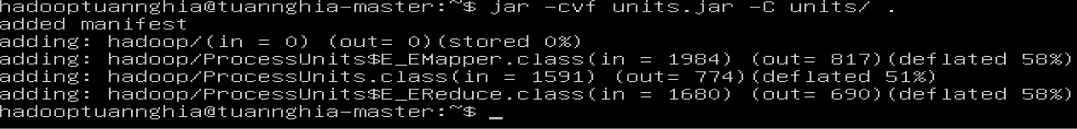


1. Download hadoop-core-1.2.1.jar;





1. Thông dịch file nguồn và tạo file jar:



1. Tạo thư mục input trong HDFS

$ hdfs dfs -mkdir input\_dir



1. Đưa dữ liệu sample.txt vào thư mục input\_dir trong HDFS

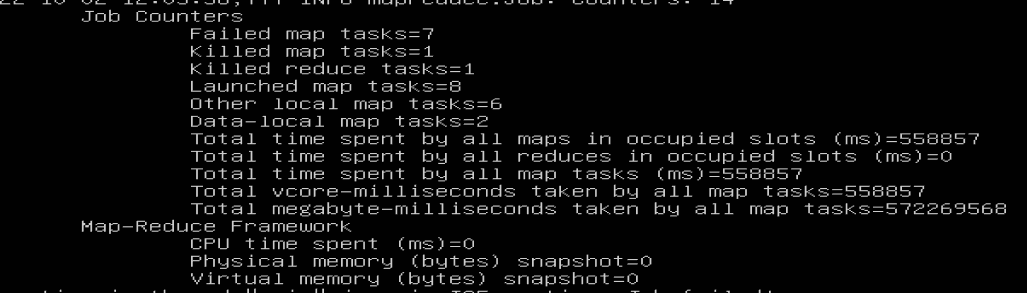
$ hdfs dfs -put sample.txt input\_dir



1. Thực thi chương trình Eleunit\_max trong Hadoop

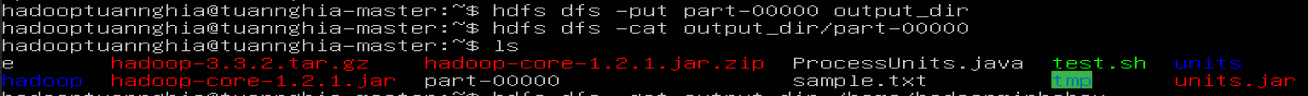
$ hadoop jar units.jar hadoop.ProcessUnits input\_dir output\_dir

Đợi cho chương trình thực thi xong sẽ cho ra kết quả số lượng input được chia tách, số lượng tác vụ Map và Reduce…



1. Xuất kết quả:

$ hdfs dfs -cat output\_dir/part-00000



1. Copy kết quả từ Hadoop ra thư mục bên

$ hdfs dfs -get output\_dir /home/hadoopminhchau

