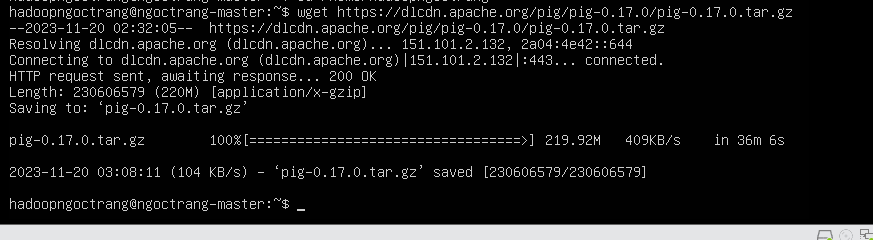
Trần Thị Ngọc Trang

21133109

CÀI ĐẶT APACHE PIG

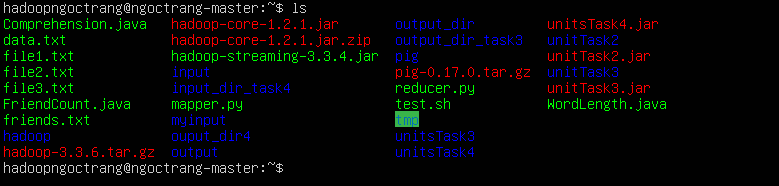
1**. Download Apache Pig**



**2. Giải nén và đổi tên thư mục** pig

$ tar -xzf pig-0.17.0.tar.gz

$ mv pig-0.17.0 pig

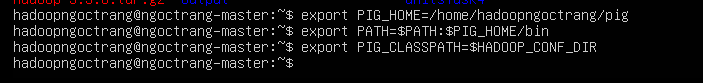


**3. Khai báo biến môi trường**

$ export PIG\_HOME=/home/hadoopngoctrang/pig

$ export PATH=$PATH:$PIG\_HOME/bin

$ export PIG\_CLASSPATH=$HADOOP\_CONF\_DIR



**4. Check Pig**

$ pig -help



**5. Thực thi Pig**

Local Mode $ pig -x local

Tez Local Mode $ pig -x spark local

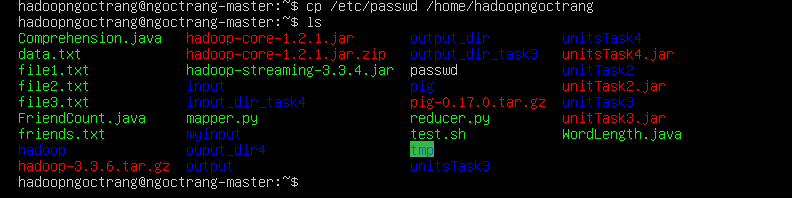
Map Reduce Mode $ pig Hoặc $ pig -x mapreduce

Tez Mode $ pig -x tez

Spark Mode $ pig -x spark

**6. Interactive Mode**

$ cp /etc/passwd /home/hadoopngoctrang

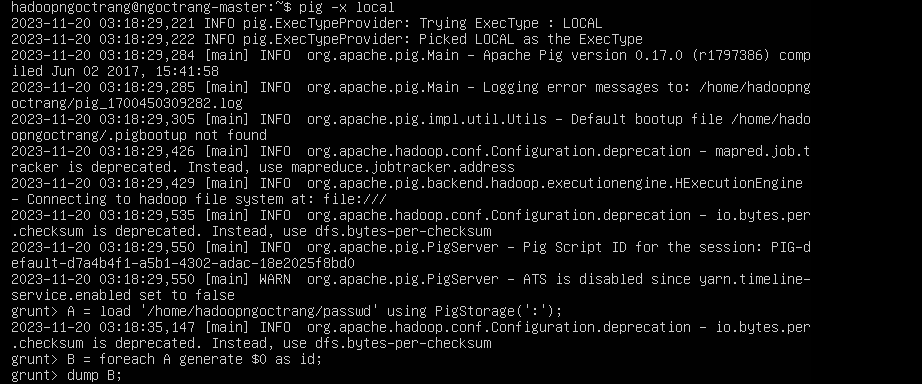


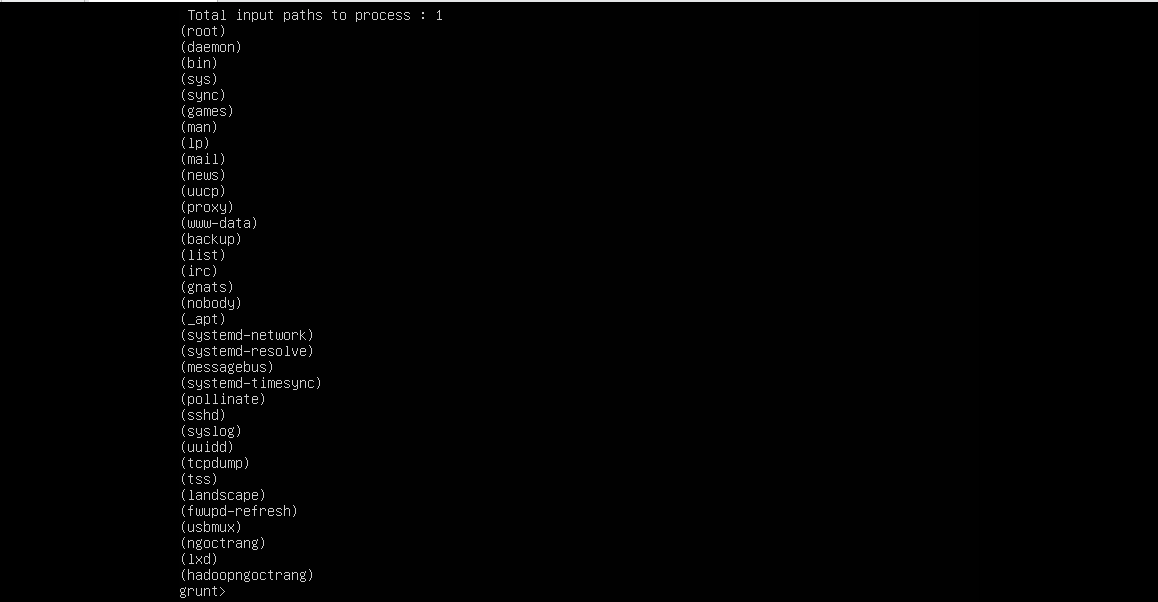
Khởi động Grunt Shell (trong cục bộ hoặc hadoop)và nhập các câu lệnh Pig trực tiếp

$ grunt> A = load '/home/hadoopngoctrang/passwd' using PigStorage(':');

$ grunt> B = foreach A generate $0 as id;

$ grunt> dump B;



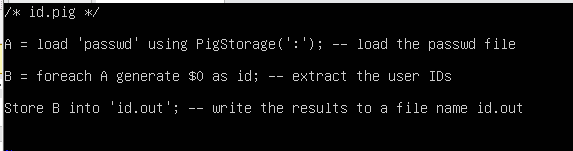


**7. Batch Mode**

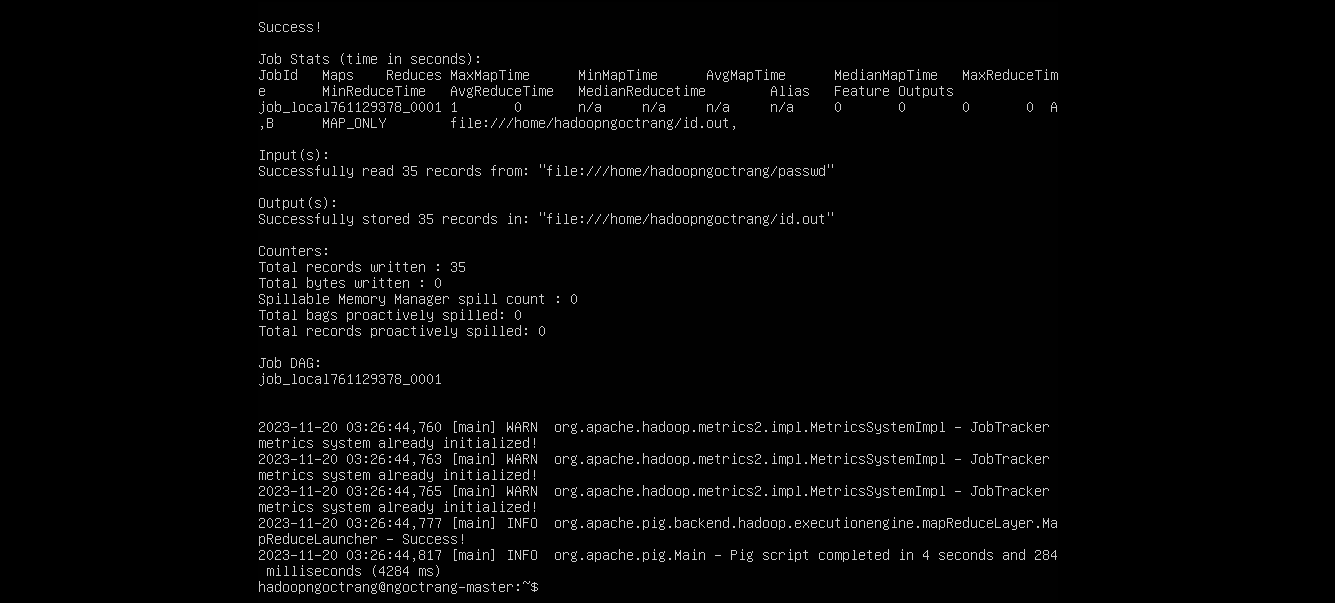
$ vim id.pig



Thực thi file script

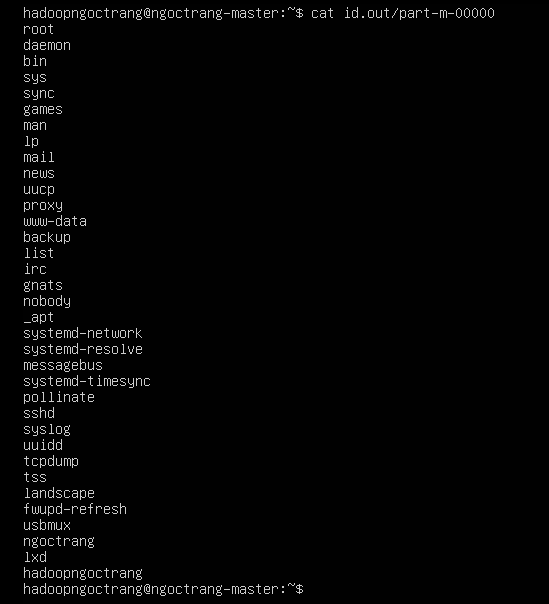


$ pig -x local id.pig



Xuất kết quả





**8. Chạy Pig trên mapreduce**

$ hadoop fs -mkdir pigdata

$ hadoop fs -put /home/hadoopngoctrang/passwd pigdata

$ pig

grunt> A = load ‘./pigdata/passwd’ using PigStorage (‘:’);

grunt> B = foreach A generate $0 as id;

grunt> dump B;

