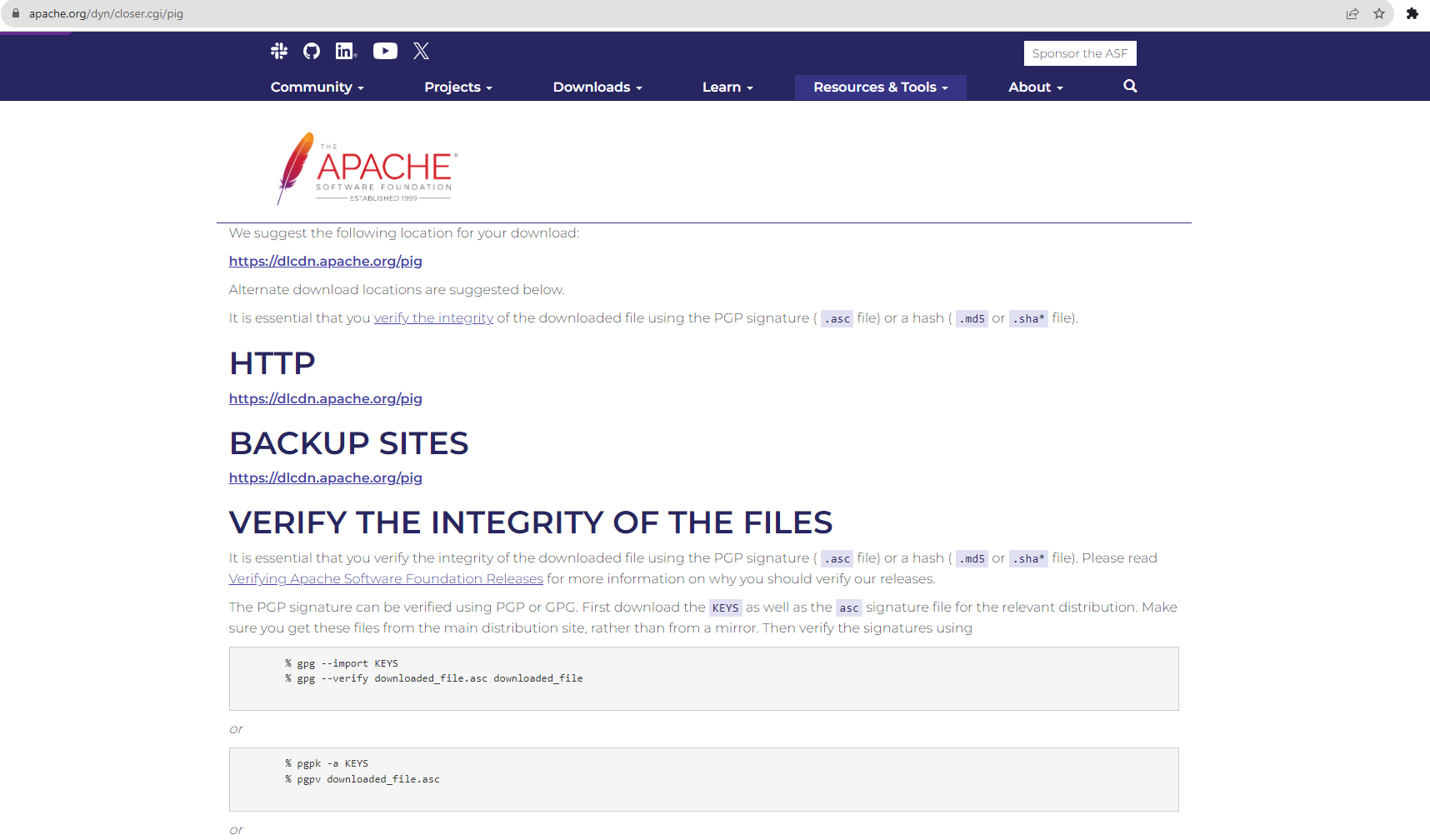
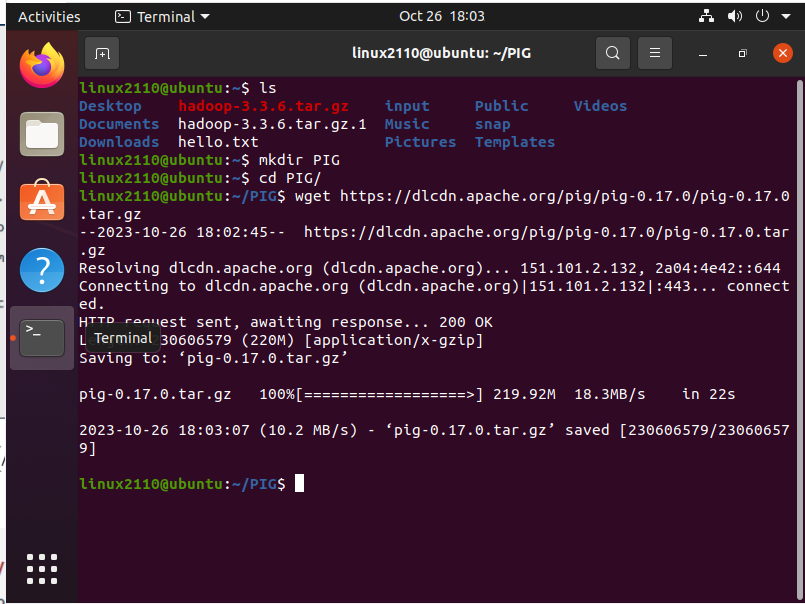
**Cài đặt Pig**  
Vào link :

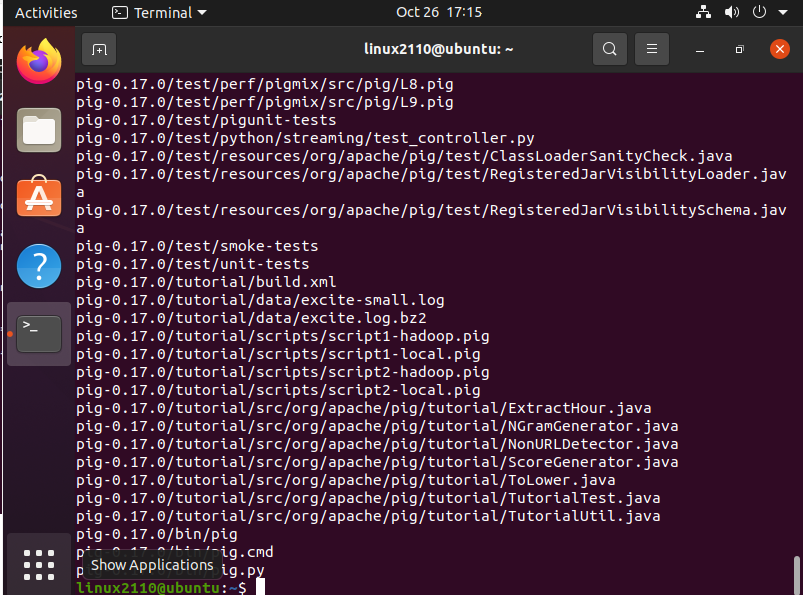
<https://pig.apache.org/releases.html>



Mở Ubuntu và dùng lệnh $ wget https://dlcdn.apache.org/pig/pig-0.17.0/pig-0.17.0.tar.gz



Dùng lệnh $ tar xvzf pig-0.17.0.tar.gz , để giải nén



Dùng lệnh để cấu hình

export JAVA\_HOME=/usr/local/java

export PATH=$PATH:/usr/local/java/bin

export HADOOP\_HOME=/usr/local/hadoop

export HADOOP\_CONF\_DIR=/usr/local/hadoop/etc/hadoop

export HADOOP\_MAPRED\_HOME=/usr/local/hadoop

export HADOOP\_COMMON\_HOME=/usr/local/hadoop

export HADOOP\_HDFS\_HOME=/usr/local/hadoop

export YARN\_HOME=/usr/local/hadoop

export PATH=$PATH:/usr/local/hadoop/bin

export PATH=$PATH:/usr/local/hadoop/sbin

export HADOOP\_COMMON\_LIB\_NATIVE\_DIR=$HADOOP\_HOME/lib/native

export HADOOP\_OPTS="-Djava.library.path=$HADOOP\_HOME/lib"

export PATH=$PATH:/home/hduser/idea/bin

export PIG\_HOME=/usr/local/pig

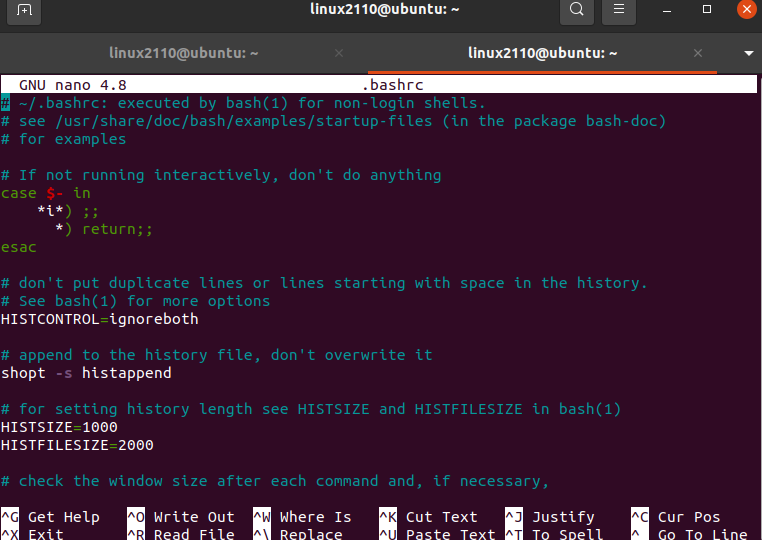
export PATH=$PATH:/usr/local/pig/bin

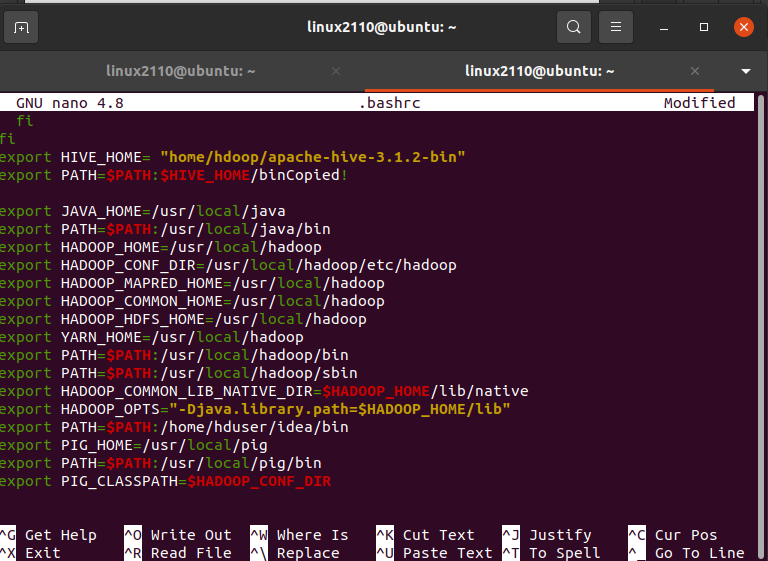
export PIG\_CLASSPATH=$HADOOP\_CONF\_DIR

sudo nano .bashrc

Cài đật môi trường

source ~/.bashrc





**Pig operations**

truck\_events = LOAD 'truck\_event\_text\_partition.csv' USING

PigStorage(',') AS (driverId:int, truckId:int, eventTime:chararray,

eventType:chararray, longitude:double, latitude:double,

eventKey:chararray, correlationId:long, driverName:chararray,

routeId:long,routeName:chararray,eventDate:chararray);`

drivers = LOAD 'drivers.csv' USING PigStorage(',') AS (driverId:int,

name:chararray, ssn:chararray, location:chararray,

certified:chararray, wage\_plan:chararray);

join\_data = JOIN truck\_events BY (driverId), drivers BY (driverId);

ordered\_data = ORDER join\_data BY truck\_events::driverName asc;

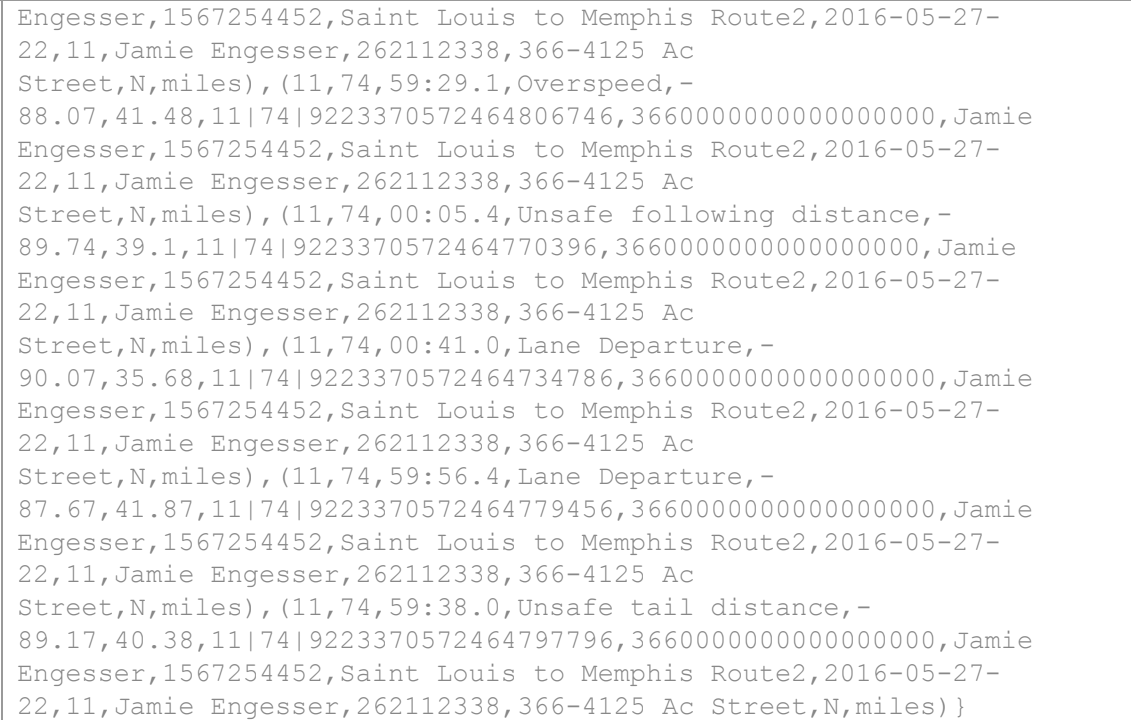
filtered = FILTER ordered\_data BY NOT (eventType MATCHES 'Normal');

grouped\_events = GROUP filtered BY truck\_events::driverId;

grouped\_events\_subset = LIMIT grouped\_events 5;

DUMP grouped\_events;

STORE grouped\_events INTO 'outut.txt' USING PigStorage(';');



**OUTPUT :**

10;{(10,85,59:46.9,Overspeed,-

95.5,36.37,10|85|9223370572464788896,3660000000000000000,George

Vetticaden,1390372503,Saint Louis to Tulsa,2016-05-27-22,10,George

Vetticaden,621011971,244-4532 Nulla

Rd.,N,miles),(10,85,00:13.1,Unsafe tail distance,-

91.18,38.22,10|85|9223370572464762694,3660000000000000000,George

Vetticaden,1390372503,Saint Louis to Tulsa,2016-05-27-22,10,George

Vetticaden,621011971,244-4532 Nulla

Rd.,N,miles),(10,85,00:39.7,Overspeed,-

94.23,37.09,10|85|9223370572464736126,3660000000000000000,George

Vetticaden,1390372503,Saint Louis to Tulsa,2016-05-27-22,10,George

Vetticaden,621011971,244-4532 Nulla Rd.,N,miles)}

2. 11;{(11,74,59:47.3,Unsafe tail distance,-

89.63,39.84,11|74|9223370572464788546,3660000000000000000,Jamie

Engesser,1567254452,Saint Louis to Memphis Route2,2016-05-27-

22,11,Jamie Engesser,262112338,366-4125 Ac

Street,N,miles),(11,74,00:49.6,Lane Departure,-

89.71,37.47,11|74|9223370572464726246,3660000000000000000,Jamie

Engesser,1567254452,Saint Louis to Memphis Route2,2016-05-27-

22,11,Jamie Engesser,262112338,366-4125 Ac

Street,N,miles),(11,74,00:14.1,Lane Departure,-

88.77,40.76,11|74|9223370572464761716,3660000000000000000,Jamie

Engesser,1567254452,Saint Louis to Memphis Route2,2016-05-27-

22,11,Jamie Engesser,262112338,366-4125 Ac

Street,N,miles),(11,74,00:23.1,Unsafe tail distance,-

88.42,41.11,11|74|9223370572464752715,3660000000000000000,Jamie

Engesser,1567254452,Saint Louis to Memphis Route2,2016-05-27-

22,11,Jamie Engesser,262112338,366-4125 Ac

Street,N,miles),(11,74,00:32.0,Unsafe tail distance,-

90.2,38.65,11|74|9223370572464743846,3660000000000000000,Jamie

Engesser,1567254452,Saint Louis to Memphis Route2,2016-05-27-

22,11,Jamie Engesser,262112338,366-4125 Ac

Street,N,miles),(11,74,59:29.1,Overspeed,-

88.07,41.48,11|74|9223370572464806746,3660000000000000000,Jamie

Engesser,1567254452,Saint Louis to Memphis Route2,2016-05-27-

22,11,Jamie Engesser,262112338,366-4125 Ac

Street,N,miles),(11,74,00:05.4,Unsafe following distance,-

89.74,39.1,11|74|9223370572464770396,3660000000000000000,Jamie

Engesser,1567254452,Saint Louis to Memphis Route2,2016-05-27-

22,11,Jamie Engesser,262112338,366-4125 Ac

Street,N,miles),(11,74,00:41.0,Lane Departure,-

90.07,35.68,11|74|9223370572464734786,3660000000000000000,Jamie

Engesser,1567254452,Saint Louis to Memphis Route2,2016-05-27-

22,11,Jamie Engesser,262112338,366-4125 Ac

Street,N,miles),(11,74,59:56.4,Lane Departure,-

87.67,41.87,11|74|9223370572464779456,3660000000000000000,Jamie

Engesser,1567254452,Saint Louis to Memphis Route2,2016-05-27-

22,11,Jamie Engesser,262112338,366-4125 Ac

Street,N,miles),(11,74,59:38.0,Unsafe tail distance,-

89.17,40.38,11|74|9223370572464797796,3660000000000000000,Jamie

Engesser,1567254452,Saint Louis to Memphis Route2,2016-05-27-

22,11,Jamie Engesser,262112338,366-4125 Ac Street,N,miles)}

29;{(29,66,00:47.8,Overspeed,-

94.57,35.37,29|66|9223370572464728016,3660000000000000000,Teddy

Choi,803014426,Wichita to Little Rock Route 2,2016-05-27-22,29,Teddy

Choi,185502192,P.O. Box 106- 7003 Amet Rd.,Y,hours)}

21. 30;{(30,58,00:49.3,Unsafe following distance,-

89.03,41.92,30|58|9223370572464726546,3660000000000000000,Dan

Rice,160779139,Des Moines to Chicago Route 2,2016-05-27-22,30,Dan

Rice,282307061,Ap #881-9267 Mollis Avenue,Y,hours)}

22. 31;{(31,18,00:47.6,Lane Departure,-

88.07,41.48,31|18|9223370572464728166,3660000000000000000,Rommel

Garcia,1594289134,Memphis to Little Rock Route 2,2016-05-27-

22,31,Rommel Garcia,858912101,P.O. Box 945- 6015 Sociis

St.,Y,hours)}

23. 32;{(32,42,00:48.7,Unsafe following distance,-

90.04,35.19,32|42|9223370572464727106,3660000000000000000,Ryan

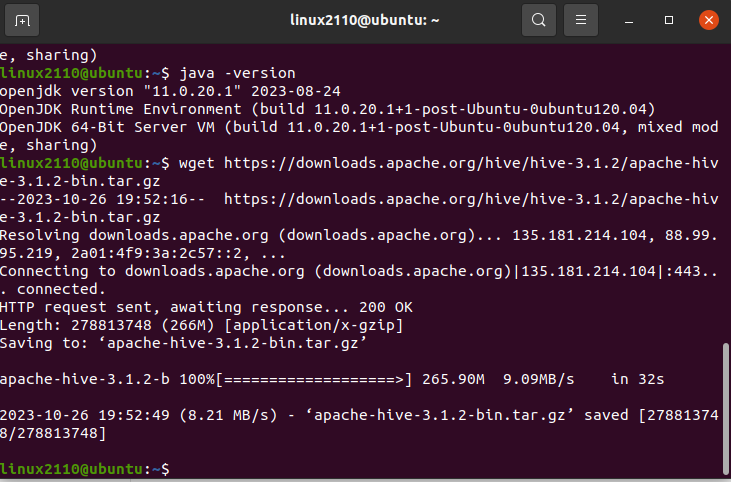
Templeton,1090292248,Peoria to Ceder Rapids Route 2,2016-05-27-

22,32,Ryan Templeton,290304287,765-6599 Egestas. Av.,Y,hours)}

**Cài đặt Hive**

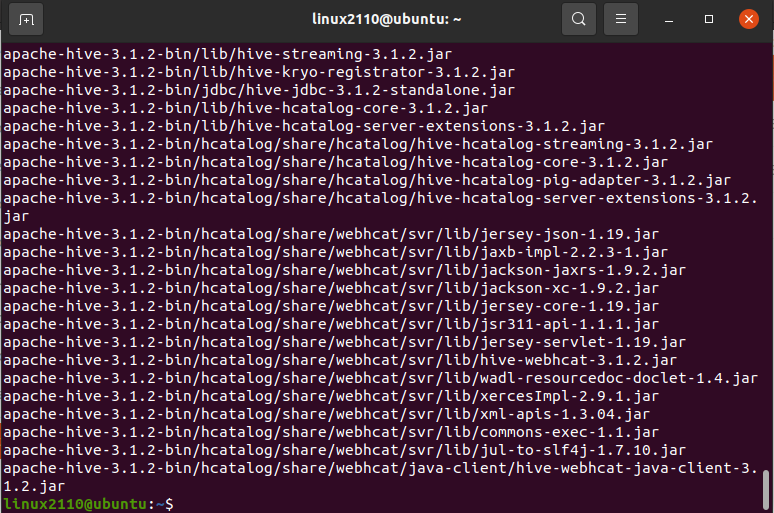
Dùng lệnh : wget <https://downloads.apache.org/hive/hive-3.1.2/apache-hive-3.1.2-bin.tar.gz>

Để download xuống

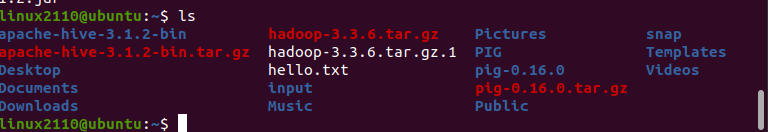


Dùng lệnh : tar xzf apache-hive-3.1.2-bin.tar.gz

Để giải nén ra



Dung $ ls để check



sudo nano .bashrc

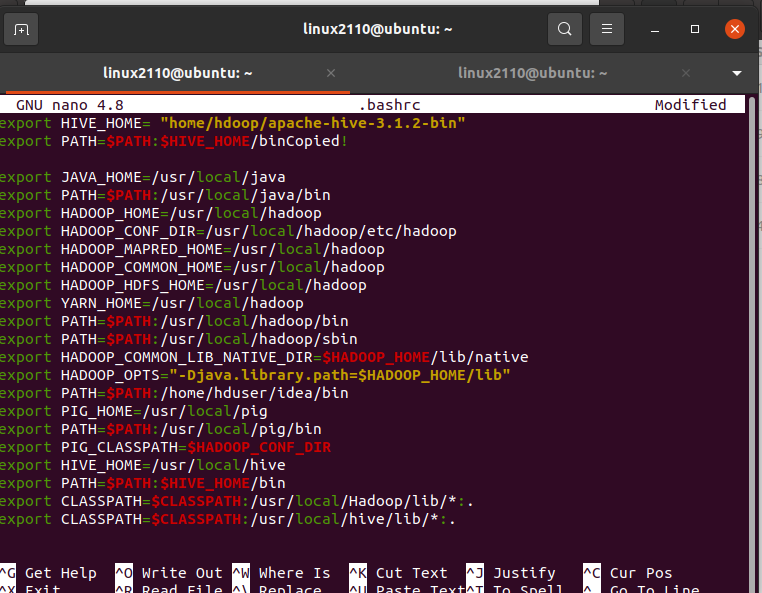
Thêm path :

export HIVE\_HOME=/usr/local/hive

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

export CLASSPATH=$CLASSPATH:/usr/local/Hadoop/lib/\*:.

export CLASSPATH=$CLASSPATH:/usr/local/hive/lib/\*:.



**Download và install Apache Derby**

$ cd ~

Để tải về :

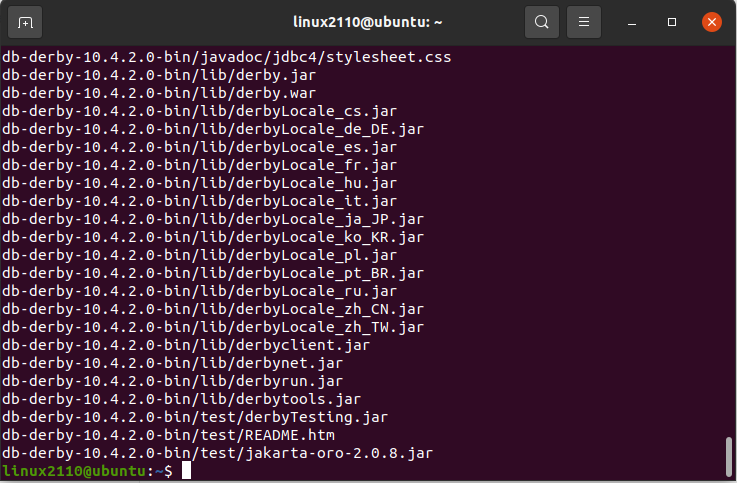
$ wget <http://archive.apache.org/dist/db/derby/db-derby-10.4.2.0/db-derby-10.4.2.0-bin.tar.gz>

Để giải nén ra :

$ tar zxvf db-derby-10.4.2.0-bin.tar.gz



Sau khi giải nén :



Dùng lệnh :

$ su -

Để chuyển về home / user :

# cd /home/user

# mv db-derby-10.4.2.0-bin /usr/local/derby

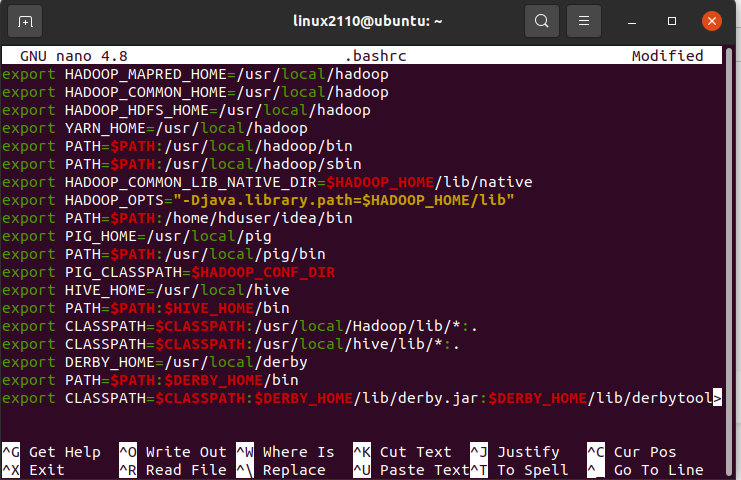
# exit

Sau đó là cài đặt môi trường cho Derby

export DERBY\_HOME=/usr/local/derby

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

export CLASSPATH=$CLASSPATH:$DERBY\_HOME/lib/derby.jar:$DERBY\_HOME/lib/derbytools.jar



**Hive Operations**

CREATE DATABASE IF NOT EXISTS userdb;

CREATE TABLE IF NOT EXISTS employee ( eid int, name String, salary

String, designation String) COMMENT 'Employee details' ROW FORMAT

DELIMITED FIELDS TERMINATED BY '\t' LINES TERMINATED BY '\n' STORED

AS TEXTFILE LOCATION '/user/input';

LOAD DATA LOCAL INPATH 'inputdata.txt' OVERWRITE INTO TABLE employee;

CREATE VIEW writer\_editor AS SELECT \* FROM employee WHERE

designation='Writer' or designation='Editor';

CREATE INDEX index\_salary ON TABLE employee(salary) AS

'org.apache.hadoop.hive.ql.index.compact.CompactIndexHandler' WITH

DEFERRED REBUILD;

SELECT \* from employee;

SELECT \* from writer\_editor;

