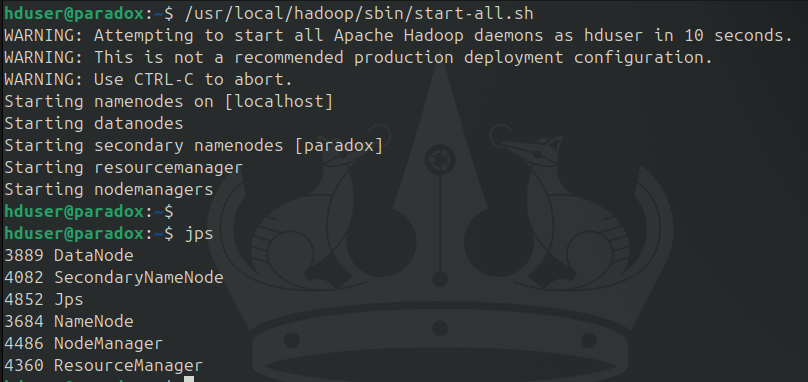
**Practical 6**

**Aim: Install, configure and run Pig. Execute Pig Latin scripts to sort, group, join, project and filter data.**

**Start the hadoop and verify all services are started**

\

/usr/local/hadoop/sbin/start-all.sh

**Download the Pig Package file:**

wget https://downloads.apache.org/pig/pig-0.17.0/pig-0.17.0.tar.gz

**Navigate to /usr/local/**

sudo tar xzvf /home/hduser/Downloads/pig-0.17.0.tar.gz

sudo mv pig-0.17.0-src/ pig

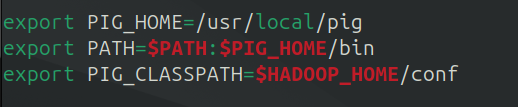
**Add the Pig environment variables in bashrc and check the pig version to verify the installation.**

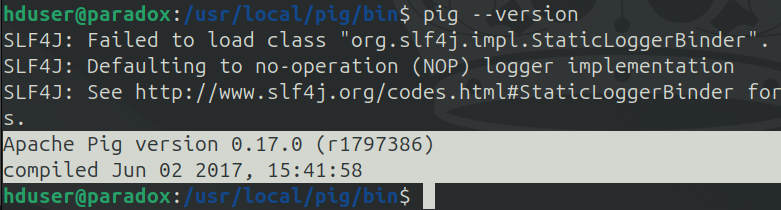
sudo nano ~/.bashrc

# Add the following lines

export PIG\_HOME=/usr/local/pig

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

export PIG\_CLASSPATH=$HADOOP\_HOME/conf

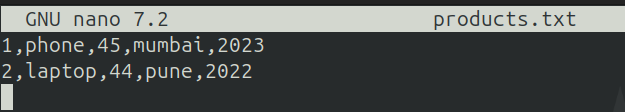
pig –-version

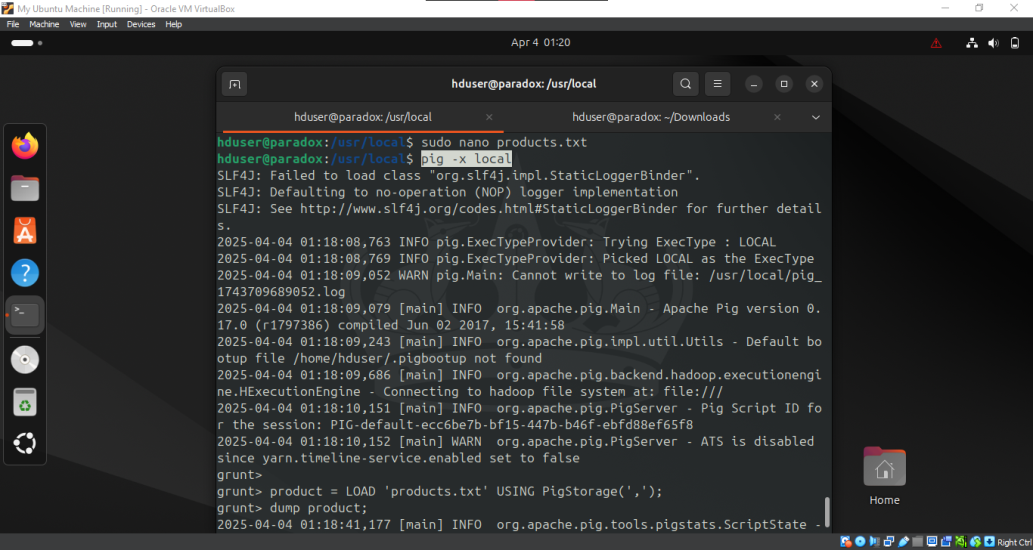
**Create a database file.**

sudo nano products.txt

Enter some text like (without spaces)

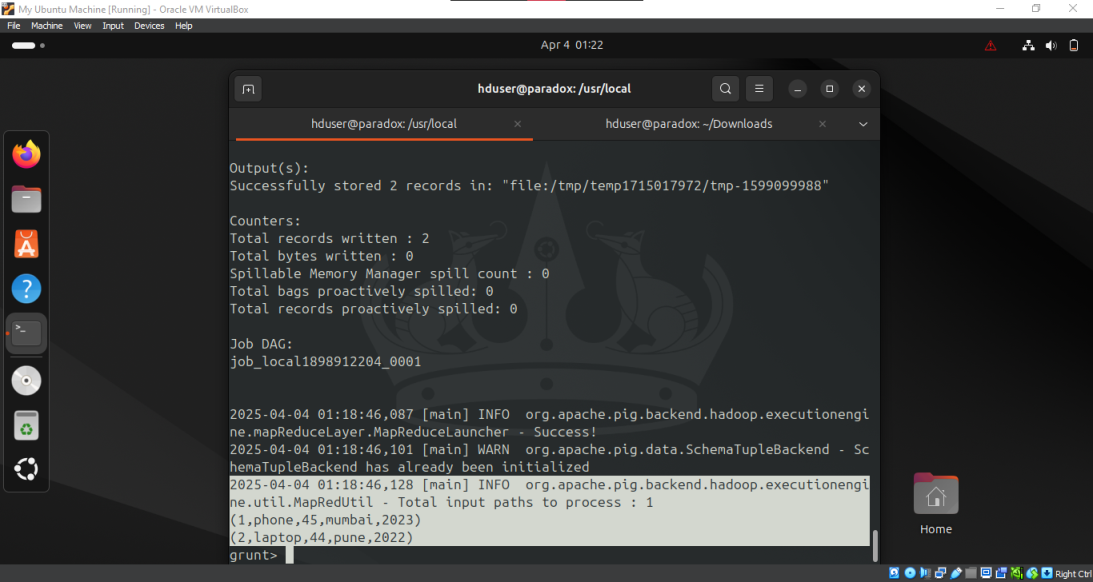
1,phone,45,mumbai,2023

2,laptop,44,pune,2022

**Run the pig in Local mode and load the products file**pig -x local

product = LOAD 'products.txt' USING PigStorage(',');

dump product;



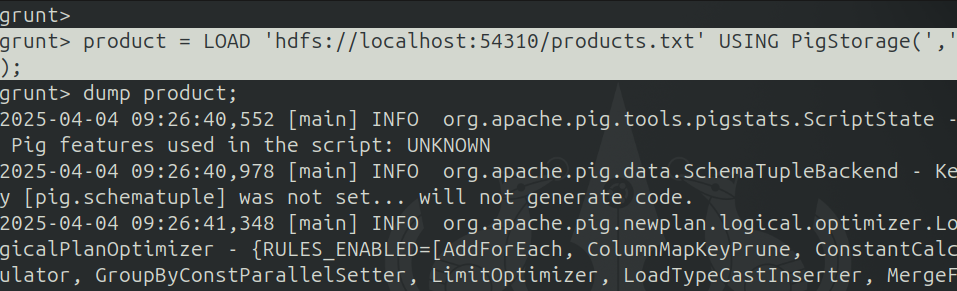
**Running in HDFS mode**

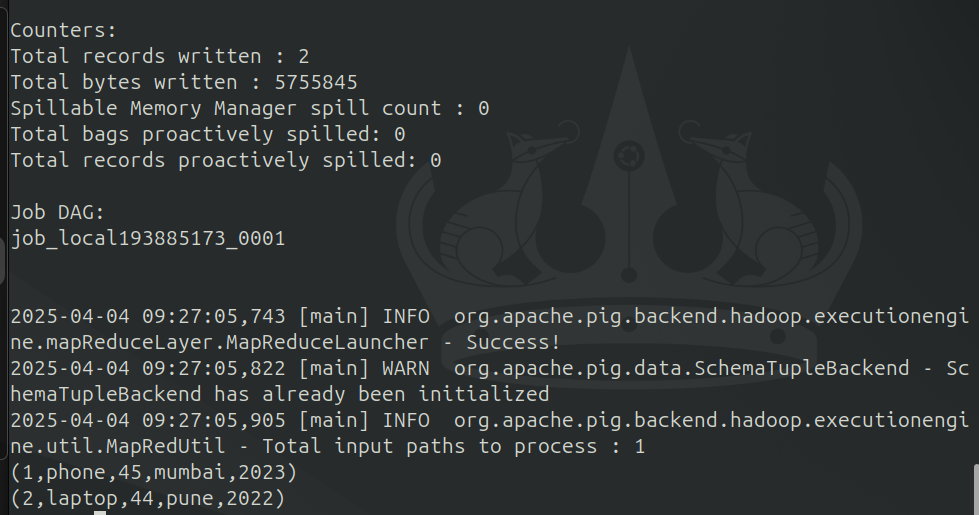
First we need to move products.txt to HDFS

hdfs dfs -put /usr/local/products.txt /

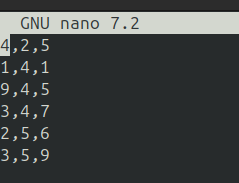
pig

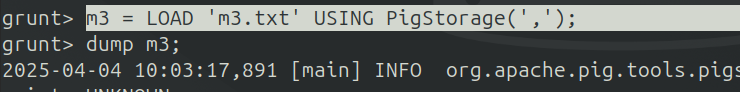
product = LOAD 'hdfs://localhost:54310/products.txt' USING PigStorage(',');

dump product;



**Use of DISTINCT operator in PIG. Assume appropriate data in text files.**

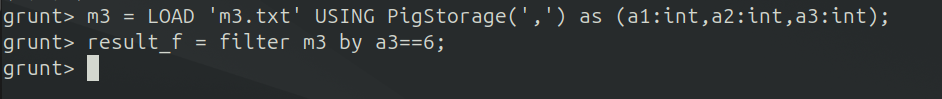
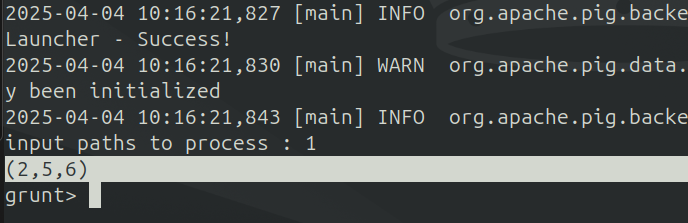
sudo nano m3.txt



**Use of FILTER operator in PIG**

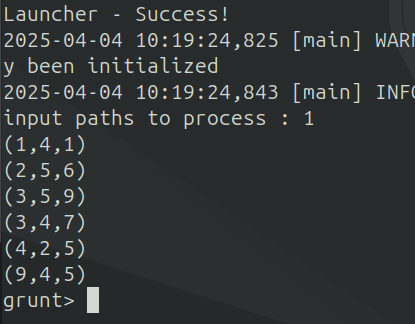
m3 = LOAD 'm3.txt' USING PigStorage(',') as (a1:int,a2:int,a3:int);

result\_f = filter m3 by a3==6;

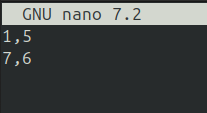
dump result\_f

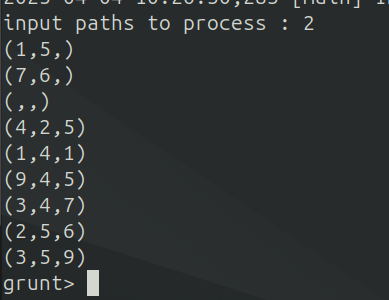
**Use of ORDERBY operator in PIG**  
result\_ob = ORDER m3 BY a1 ASC;

dump result\_ob;



**Use of UNION operator in PIG**

Sudo nano m1.txt



**Conclusion:** Practical to Install, configure and run Pig. Execute Pig Latin scripts to sort, group, join, project and filter data, successfully executed.