**3. Count of page views by individual user**

**IP, count(\*)**

*Pig is used to solve the project 3-6.*

*Pig provides an engine for executing data flows in parallel on Hadoop. It includes a*

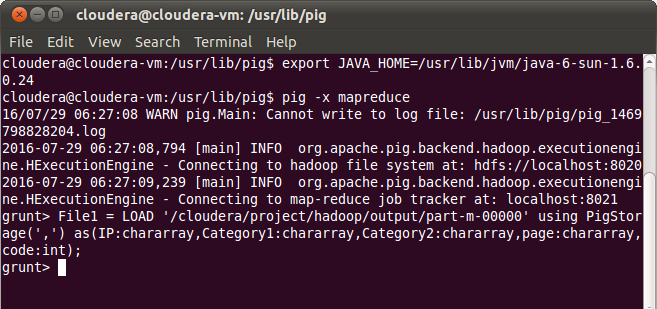
*language, Pig Latin, for expressing these data flows*

**Loading the output from project2 into pig:-**

**Command**:-

File1 = LOAD '/cloudera/project/hadoop/output/part-m-00000' using PigStorage(',') as(IP:chararray,Category1:chararray,Category2:chararray,page:chararray,code:int);

**Screenshot:-**



**Group the records as per the IP:-**

**Command:-**

GPRD = group File1 by IP;

**Count the grouped record in previous step:-**

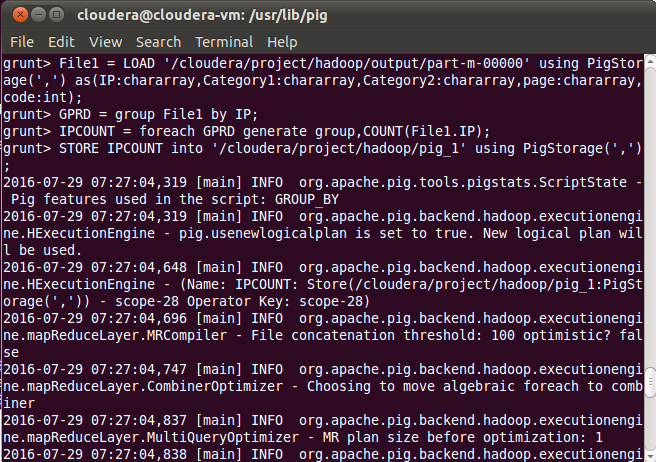
IPCOUNT = foreach GPRD generate group,COUNT(File1.IP)

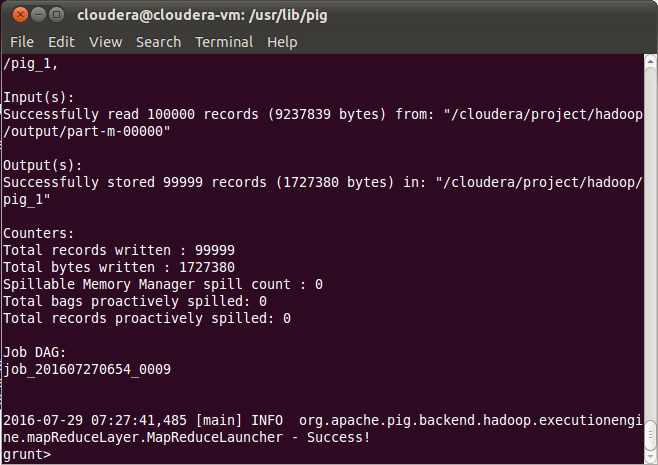
**Store the count results into HDFS:-**

**Command:-**

**STORE IPCOUNT into ‘/cloudera/project/hadoop/pig\_1’ using PigStorage(‘,’);**

**Screenshot:-**





**O\P:- IP and their count:-**

