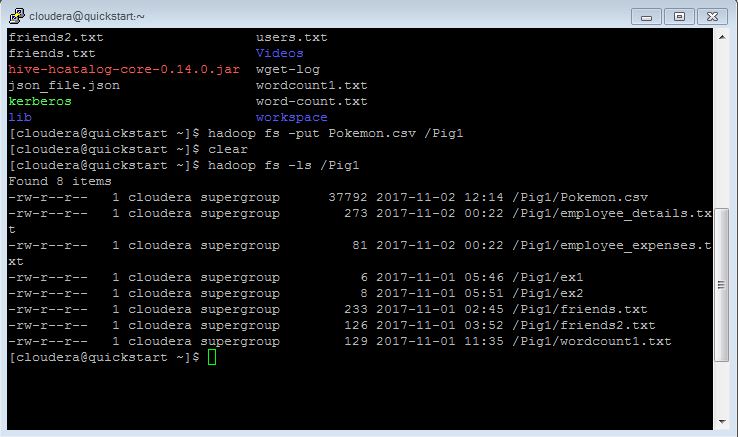
**PIG Use Case: Pokemon Data Analysis**

The Pokémon Fight League (PFL) management for the 2017 match has first of all decided a minimum criterion for the entry selection process that filters through the defense power for any Pokémon, which should ideally be greater than 55.

load the dataset inside PIG. We can either use the local mode or the MR mode. I am using the MR mode.

*Command*

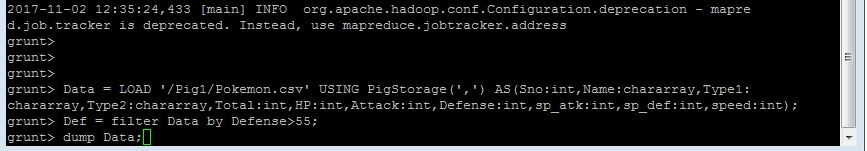
Data = LOAD ‘/Pig1/Pokemon.csv’ USING PigStorage(‘,’) AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int,sp\_atk:int,sp\_def:int,speed:int);



**Question-1: Find the list of players that have been selected in the qualifying round (DEFENCE>55).**

***Command***

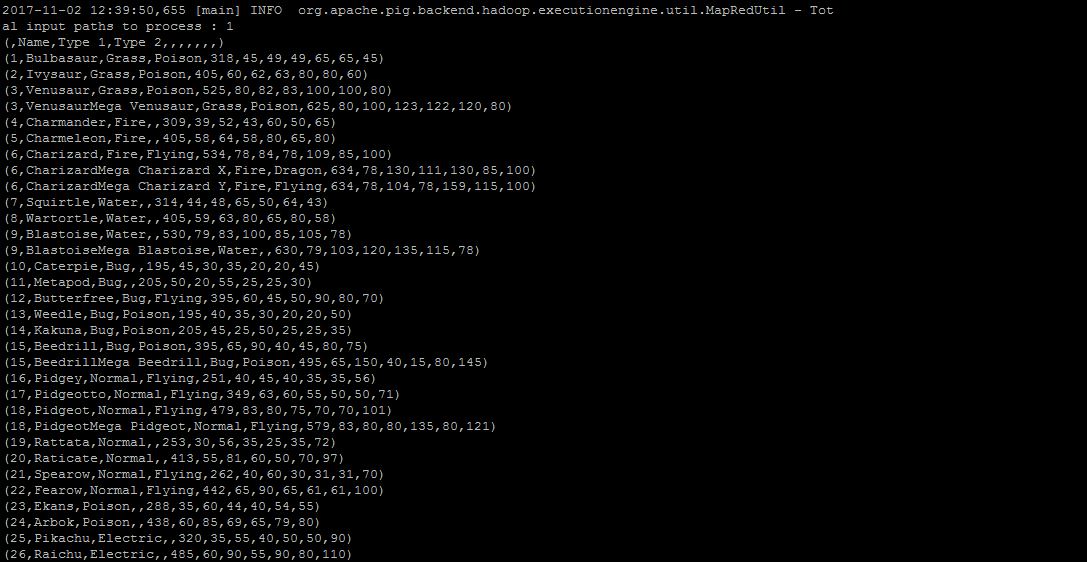
Def = FILTER Data BY Defense>55;



The dataset is filtered, and hence out of all the 800 Pokémons. Only pokemons with defense >55 will be filtered out. To get the count we have to use count built-in pig command COUNT.

**Sample Output:**

dump Data;

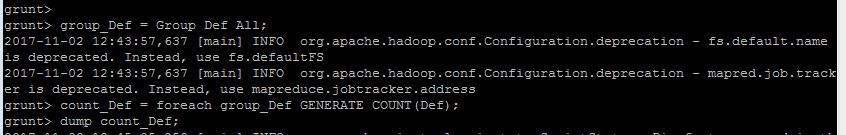


**Question-2: State the number of players taking part in the competition after getting selected in the qualifying round.**

***Command***

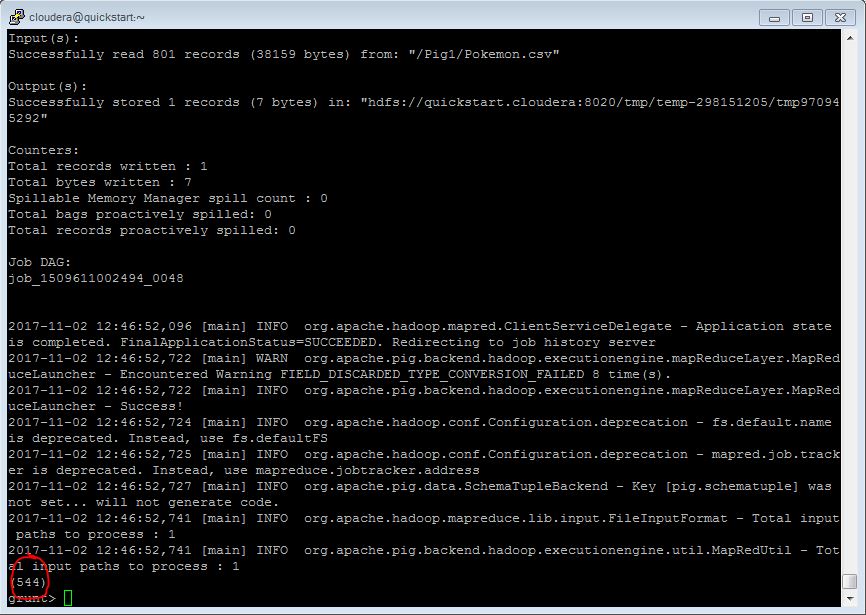
group\_Def = Group Def All;

count\_Def = foreach group\_Def GENERATE COUNT(Def);



**Sample Output:**

dump count\_Def;



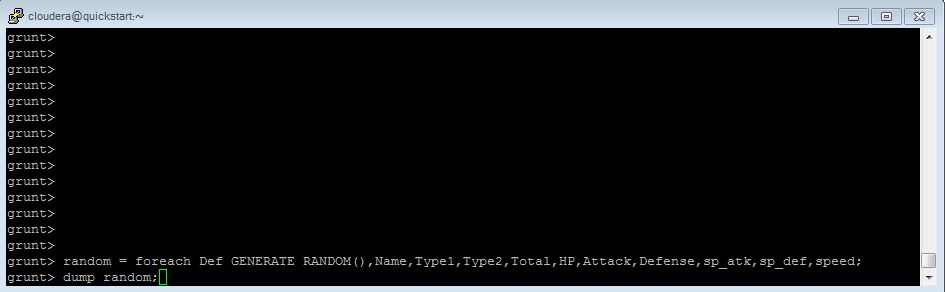
So,only 544 are eligible to take part in the tournament.

**Question-3: Using random() generate random numbers for each Pokémon on the selected list.**

***Command***

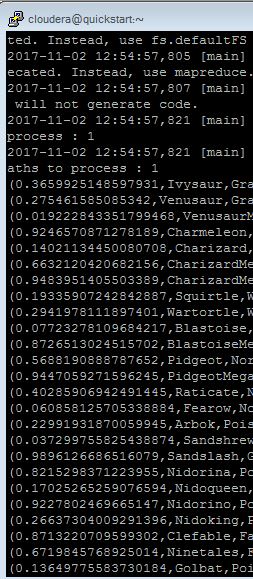
random = foreach Def GENERATE

RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,sp\_atk,sp\_def,speed;



**Sample Output:**

dump random;

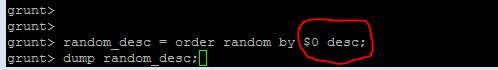


**Ques-4: Arrange the new list in a descending order according to a column randomly.**

This will give us the random list which 1st player will choose.

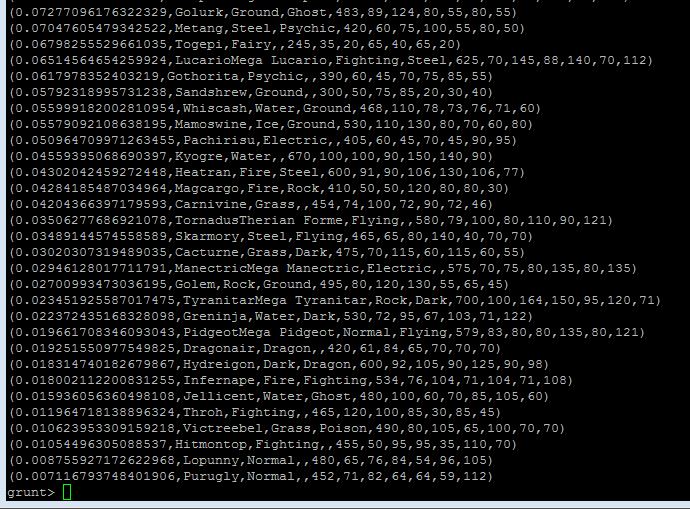
***Command***

random\_desc = ORDER random BY $0 DESC;



**Sample Output:**

Dump random\_desc;



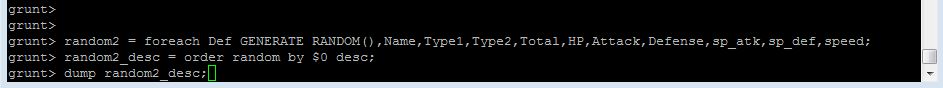
**Question 5: Now on a new relation again associate random numbers for each Pokémon and arrange in descending order according to column random.**

**Explanation**: We will be repeating above two steps again to form the 2nd list.

*Command*

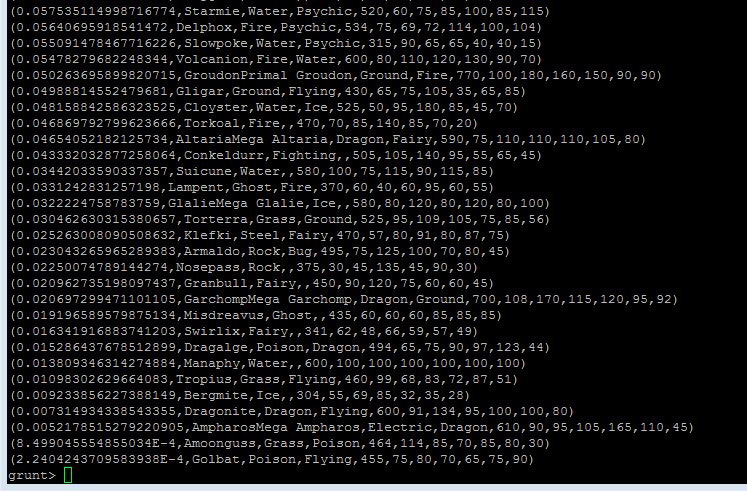
random2 = foreach Def GENERATE RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,sp\_atk,sp\_def,speed;

random2\_desc= ORDER random2 BY $0 DESC;



**Sample Output:**

dump random2\_desc;



**Question-6: From the two different descending lists of random Pokémons, select the top 5 Pokémons for 2 different players.**

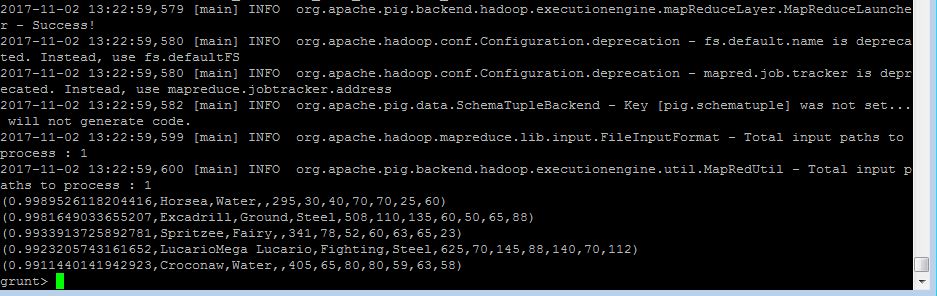
**Explanation**:

***Commands***

limit\_random\_desc = LIMIT random\_desc 5 ;



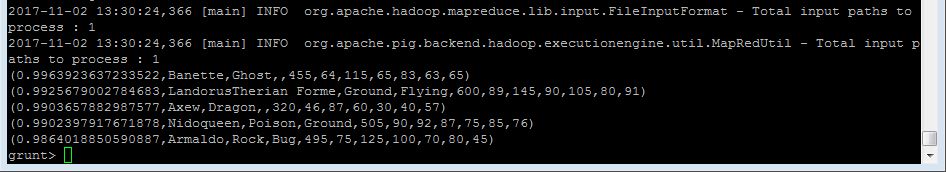
**Sample Output:**



limit\_random2\_desc = LIMIT random2\_desc 5 ;



**Sample Output:**



**Question-7: Store the data on a local drive to announce for the final match. By the name player1 and player2 (only show the NAME and HP).**

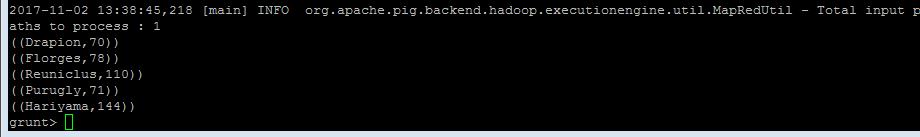
**Explanation**:

***Command***

filter\_only\_name1 = foreach limit\_random\_desc GENERATE ($1,HP);



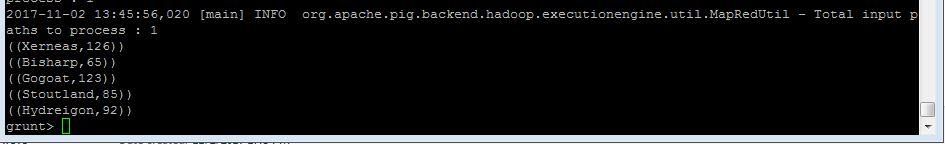
**For Player1 Output :**



filter\_only\_name2 = foreach limit \_random2\_desc Generate ($1,HP);



**For Player2 Output :**



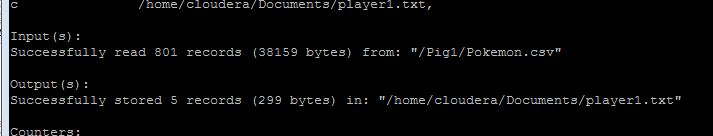
Therefore querying to get 2 sets of 5 Pokémons, which get select randomly is completed. So now let’s store this result in our local system

***command***

STORE limit\_ random\_desc INTO ‘/home/cloudera/Documents/player1.txt’;



**Output:**



***command***

STORE limit \_random2\_desc INTO ‘/home/cloudera/Documents/player2.txt’;



**Output:**

