PIG Use Case: Pokemon Data Analysis

Loading at local mode:

Pig –x local;

DataSet:

Sno:int,

Name:chararray,

Type1:chararray,

Type2:chararray,

Total:int,

HP:int,

Attact:int,

Defense:Int,

SpAtk:int,

SpDef:Int,

Speed:int);

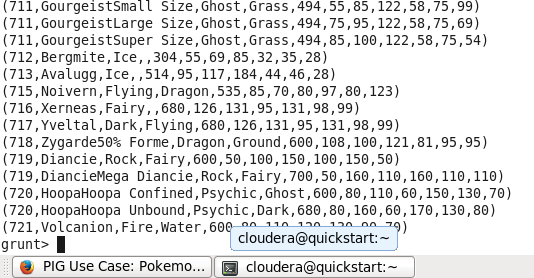
Load\_Data = load '/home/cloudera/Pokemon.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE')

as (Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attact:int,Defense:Int,SpAtk:int,SpDef:Int,Speed:int);

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

selected\_list = FILTER Load\_Data BY Defense>55;

dump selected\_list;



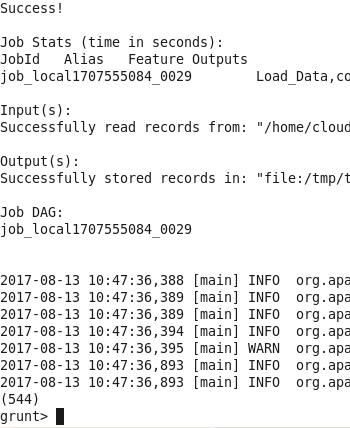
#### Ques 2: State the number of players taking part in the competition after getting selected in the qualifying round.

gourp\_selcted\_list = Group selected\_list All;

count\_selcted\_list = foreach gourp\_selcted\_list GENERATE COUNT(selected\_list);

dump count\_selcted\_list;

Result:

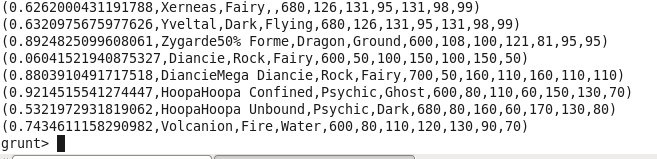


#### Ques 3: Using random() generate random numbers for each Pokémon on the selected list.

random\_include1 = foreach selected\_list GENERATE RANDOM(),Name,Type1,Type2,Total,HP,Attact,Defense,SpAtk,SpDef,Speed;

dump random\_include1;

Result:



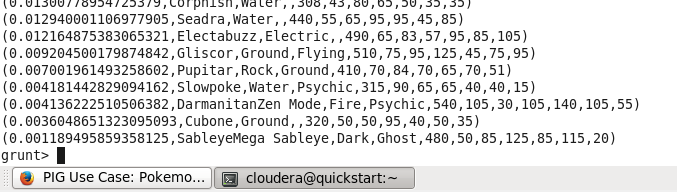
#### Ques 4: Arrange the new list in a descending order according to a column randomly.

**Explanation**: This will give us consequently a layer arranged to pick the random list which 1st player will

random1\_desending = ORDER random\_include1 BY $0 DESC;

dump random1\_desending;

Result:



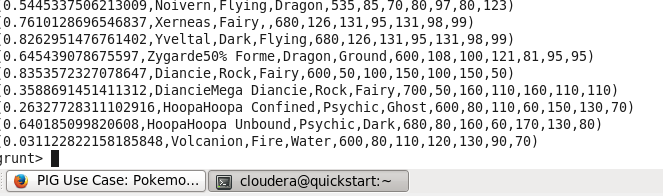
#### Ques 5: Now on a new relation again associate random numbers for each Pokémon and arrange in descending order according to column random.

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

random\_include2 = foreach selected\_list GENERATE RANDOM(),Name,Type1,Type2,Total,HP,Attact,Defense,SpAtk,SpDef,Speed;

dump random\_include2;

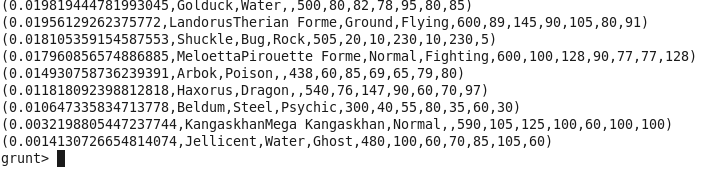
Result:



random2\_desending = ORDER random\_include2 BY $0 DESC;

dump random2\_desending;

Result:



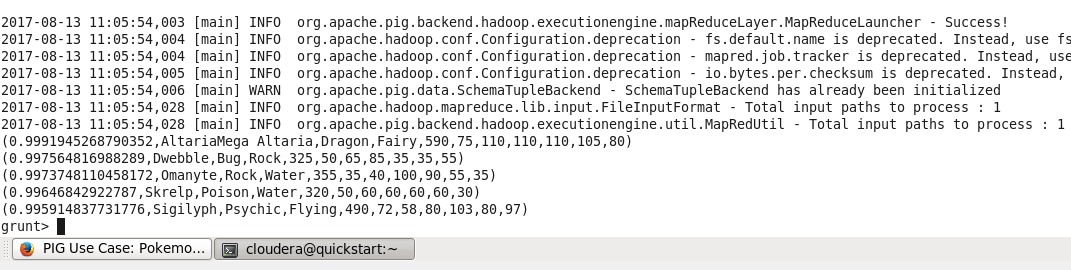
#### Ques: From the two different descending lists of random Pokémons, select the top 5 Pokémons for 2 different players.

**Explanation**:

limit\_data\_random1\_desending = LIMIT random1\_desending 5 ;

dump limit\_data\_random1\_desending;

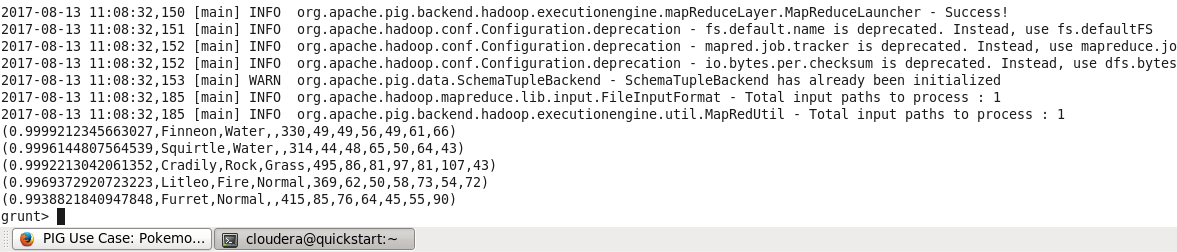
Result:



limit\_data\_random2\_desending = LIMIT random2\_desending 5 ;

dump limit\_data\_random2\_desending;

Result:



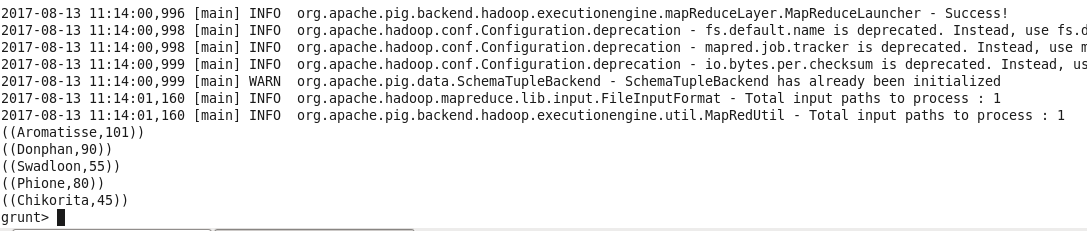
#### Ques: Store the data on a local drive to announce for the final match. By the name player1 and player2 (only show the NAME and HP).

**Explanation**:

filter\_only\_name1 = foreach limit\_data\_random1\_desending Generate ($1,HP);

dump filter\_only\_name1;

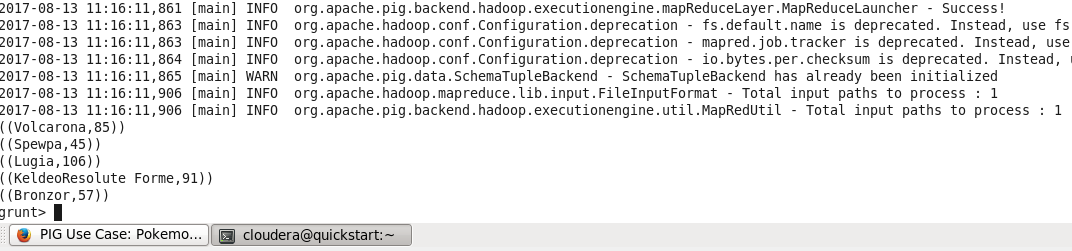
Result:



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

dump filter\_only\_name2;

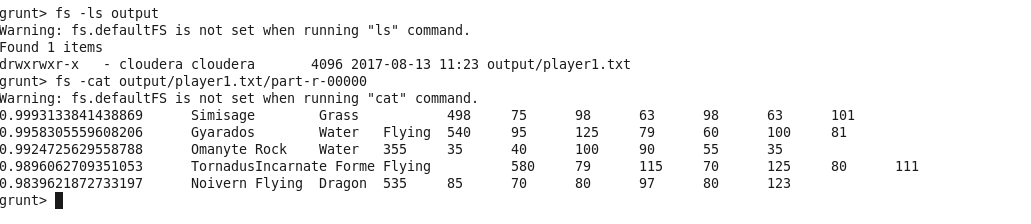
Result:



Store Result localally

STORE limit\_data\_random1\_desending INTO 'output/player1.txt'

Result:



STORE limit\_data\_random2\_desending INTO 'output/player2.txt'

Result:

