**PIG Use Case: Pokemon Data Analysis**

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

**First Load the pokemon data.**

load\_data = LOAD '/home/training/Desktop/pig/Pokemon.csv' USING PigStorage(',') as (Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int, SpAtk:int, SpDef:int, Speed:int);

1. **Find the list of players that have been selected in the qualifying round.**

To get into selected list, i.e to qualify condition is Defence > 55.

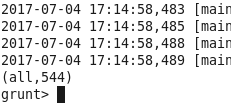
selected\_data = **FILTER** load\_data by Defense>55;

1. **State the number of players taking part in the competition after getting selected in the qualifying round.**

To get the no of players qualified, we need to use COUNT function. To use any aggregate function we need to first group the items.

group\_data = **GROUP** selected\_data all;

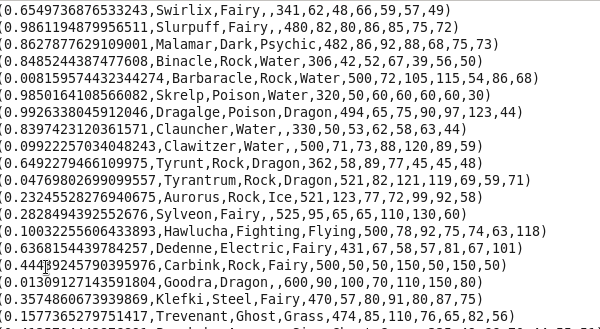
no\_of\_players\_qualified = FOREACH group\_data GENERATE group, **COUNT**(selected\_data);



1. **Using random() generate random numbers for each Pokémon on the selected list.**

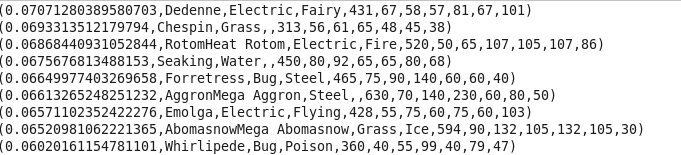


Sample random generated output:



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

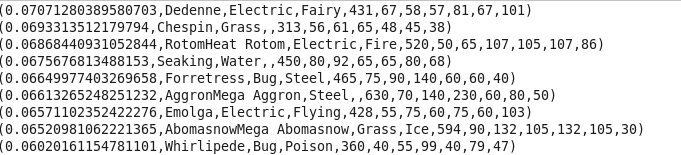




#### Ques 5: Now on a new relation again associate random numbers for each Pokémon and arrange in descending order according to column random.





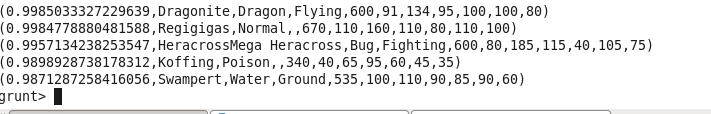


#### Ques: From the two different descending lists of random Pokémons, select the top 5 Pokémons for 2 different players.

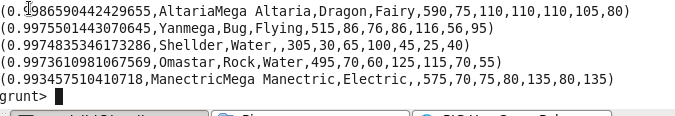
Selecting top 5 from random list 1 and list 2:



List 1:



List 2:



#### 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).

First fetch only name and HP.



Store the result in local:





Output has been successfully stored in local:

