

Problem Statement

Implement the use case present in below blog link

<https://acadgild.com/blog/pig-use-case-pokemon-data-analysis/>

Problem1:

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

Solution:

```
--Load Pokemon.csv Data
pokemon_data = LOAD '/home/acadgild/ankita/Assignment5_3/Pokemon' USING PigStorage(',')
AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int,splatkck:int
,spldefenct:int,speed:int);

-- filter qualifying data
qualifying_data = FILTER pokemon_data by Defense>55 ;

-- display results

dump qualifying_data;
```

Output:

```
Acadgild_64bit [Running] - Oracle VM VirtualBox
Help
(687,Malamar,Dark,Psychic,482,86,92,88,68,75,73)
(688,Binacle,Rock,Water,306,42,52,67,39,56,50)
(689,Barbaracle,Rock,Water,500,72,105,115,54,86,68)
(690,Skelp,Poison,Water,320,50,60,60,60,60,30)
(691,Dragalge,Poison,Dragon,494,65,75,90,97,123,44)
(692,Clauncher,Water,,330,50,53,62,58,63,44)
(693,Clawitzer,Water,,500,71,73,88,120,89,59)
(696,Tyrantrum,Rock,Dragon,362,58,89,77,45,45,48)
(697,Tyrantrum,Rock,Dragon,521,82,121,119,69,59,71)
(699,Aurorus,Rock,Ice,521,123,77,72,99,92,58)
(700,Sylveon,Fairy,,525,95,65,65,110,130,60)
(701,Hawlucha,Fighting,Flying,500,78,92,75,74,63,118)
(702,Dedenne,Electric,Fairy,431,67,58,57,81,67,101)
(703,Carbink,Rock,Fairy,500,50,50,150,50,150,50)
(706,Goodra,Dragon,,600,90,100,70,110,150,80)
(707,Klefki,Steel,Fairy,470,57,80,91,80,87,75)
(709,Trevenant,Ghost,Grass,474,85,110,76,65,82,56)
(710,PumpkabooAverage Size,Ghost,Grass,335,49,66,70,44,55,51)
(710,PumpkabooSmall Size,Ghost,Grass,335,44,66,70,44,55,56)
(710,PumpkabooLarge Size,Ghost,Grass,335,54,66,70,44,55,46)
(710,PumpkabooSuper Size,Ghost,Grass,335,59,66,70,44,55,41)
(711,GourgeistAverage Size,Ghost,Grass,494,65,90,122,58,75,84)
(711,GourgeistSmall Size,Ghost,Grass,494,55,85,122,58,75,99)
(711,GourgeistLarge Size,Ghost,Grass,494,75,95,122,58,75,69)
(711,GourgeistSuper Size,Ghost,Grass,494,85,100,122,58,75,54)
(712,Bergmite,Ice,,304,55,69,85,32,35,28)
(713,Avalugg,Ice,,514,95,117,184,44,46,28)
(715,Noivern,Flying,Dragon,535,85,70,80,97,80,123)
(716,Xerneas,Fairy,,680,126,131,95,131,98,99)
(717,Yveltal,Dark,Flying,680,126,131,95,131,98,99)
(718,Zygarde50% Forme,Dragon,Ground,600,108,100,121,81,95,95)
(719,Diancie,Rock,Fairy,600,50,100,150,100,150,50)
(719,DiancieMega Diancie,Rock,Fairy,700,50,100,110,160,110,110)
(720,HoopaHoopa Confined,Psychic,Ghost,600,80,110,60,150,130,70)
(720,HoopaHoopa Unbound,Psychic,Dark,680,80,160,60,170,130,80)
(721,Volcanion,Fire,Water,600,80,110,120,130,90,70)
grunt> █
```

Problem2:

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

Solution:

```
-- Load Data
pokemon_data = LOAD '/home/acadgild/ankita/Assignment5_3/Pokemon' USING PigStorage(',')
AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int,splatck:int
,spldefenct:int,speed:int);

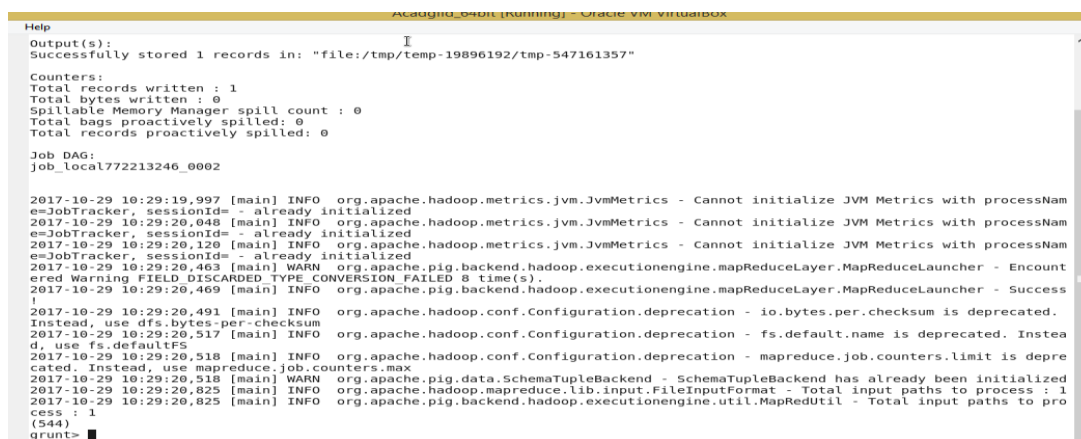
-- filter qualifying data
qualifying_data = FILTER pokemon_data by Defense>55 ;

-- group the qualifying data
group_qualifying_data = GROUP qualifying_data ALL ;

-- get count
count_qualifying_data = foreach group_qualifying_data GENERATE COUNT(qualifying_data) ;

-- display results
dump count_qualifying_data ;
```

Output:



```
Help
Output(s):
Successfully stored 1 records in: "file:/tmp/temp-19896192/tmp-547161357"

Counters:
Total records written : 1
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local772213246_0002

2017-10-29 10:29:19,997 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-10-29 10:29:20,048 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-10-29 10:29:20,120 [main] INFO org.apache.hadoop.metrics.jvm.JvmMetrics - Cannot initialize JVM Metrics with processNam
e=JobTracker, sessionId= - already initialized
2017-10-29 10:29:20,463 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encount
ered Warning FIELD DISCARDED TYPE CONVERSION FAILED 8 time(s).
2017-10-29 10:29:20,469 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success
!
2017-10-29 10:29:20,491 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated.
Instead, use dfs.bytes-per-checksum
2017-10-29 10:29:20,517 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instea
d, use fs.defaultFS
2017-10-29 10:29:20,518 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is depre
cated. Instead, use mapreduce.job.counters.max
2017-10-29 10:29:20,518 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-29 10:29:20,825 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-29 10:29:20,825 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to pro
cess : 1
(544)
grunt>
```

Problem3:

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

Solution:

```
-- Load Data
pokemon_data = LOAD '/home/acadgild/ankita/Assignment5_3/Pokemon' USING PigStorage(',')
AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int,splatck:int
,spldefence:int,speed:int);

-- filter qualifying data

qualifying_data = FILTER pokemon_data by Defense>55 ;

-- generate random for relation 1
random_include1 = foreach qualifying_data GENERATE
RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,splatck,spldefence,speed;

-- display result for random relation 1
dump random_include1;
```

Output:

```
Acadgild_64bit [Running] - Oracle VM VirtualBox
es Help
(0.4921839181055304,Heliolisk,Electric,Normal,481,62,55)
(0.7926474674145008,Tyrunt,Rock,Dragon,362,58,89)
(0.08082099418879596,Tyrantrum,Rock,Dragon,521,82,121)
(0.4087589485252534,Amaura,Rock,Ice,362,77,59)
(0.5124849550796733,Aurorus,Rock,Ice,521,123,77)
(0.8863224732512459,Sylveon,Fairy,,525,95,65)
(0.9038798587943884,Hawlucha,Fighting,Flying,500,78,92)
(0.45796493182085474,Dedenne,Electric,Fairy,431,67,58)
(0.7296286667423518,Carbink,Rock,Fairy,500,50,50)
(0.25423926194159197,Goomy,Dragon,,300,45,50)
(0.8816143865871554,Sliggoo,Dragon,,452,68,75)
(0.13584308211752605,Goodra,Dragon,,600,90,100)
(0.3401926418983452,Klefki,Steel,Fairy,470,57,80)
(0.7725912565675745,Phantump,Ghost,Grass,309,43,70)
(0.7767381475686043,Trevenant,Ghost,Grass,474,85,110)
(0.8002910429815886,PumpkabooAverage Size,Ghost,Grass,335,49,66)
(0.6189732718137687,PumpkabooSmall Size,Ghost,Grass,335,44,66)
(0.9012896802994927,PumpkabooLarge Size,Ghost,Grass,335,54,66)
(0.04385065088244722,PumpkabooSuper Size,Ghost,Grass,335,59,66)
(0.5036202187230231,GourgeistAverage Size,Ghost,Grass,494,65,90)
(0.3787710838374382,GourgeistSmall Size,Ghost,Grass,494,55,85)
(0.6892740933298985,GourgeistLarge Size,Ghost,Grass,494,75,95)
(0.16959369519127687,GourgeistSuper Size,Ghost,Grass,494,85,100)
(0.1967193498126456,Bergmite,Ice,,304,55,69)
(0.25683435937041743,Avalugg,Ice,,514,95,117)
(0.839523962003219,Noibat,Flying,Dragon,245,40,30)
(0.11588182244648282,Noivern,Flying,Dragon,535,85,70)
(0.6256247892402178,Xerneas,Fairy,,680,126,131)
(0.26143684137243517,Yveltal,Dark,Flying,680,126,131)
(0.6248172689201876,Zygarde50% Forme,Dragon,Ground,600,108,100)
(0.004176722232873908,Diancie,Rock,Fairy,600,50,100)
(0.6104692211592371,DiancieMega Diancie,Rock,Fairy,700,50,160)
(0.35796608507426664,HoopaHoopa Confined,Psychic,Ghost,600,80,110)
(0.5009464672683429,HoopaHoopa Unbound,Psychic,Dark,680,80,160)
(0.5235023253155175,Volcanion,Fire,Water,600,80,110)
grunt>
```

Problem4:

Arrange the new list in a descending order according to a column randomly.

Solution:

```
-- Load Data
pokemon_data = LOAD '/home/acadgild/ankita/Assignment5_3/Pokemon' USING PigStorage(',')
AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int,splatck:int
,spldefence:int,speed:int);

--Filter qualifying data

qualifying_data = FILTER pokemon_data by Defense>55 ;

-- generate random for relation 1
random_include1 = foreach qualifying_data GENERATE
RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,splatck,spldefence,speed;

-- sort by random function relation 1
random1_desending = ORDER random_include1 BY $0 DESC;

-- displaying result of random function relation 1
dump random1_desending;
```

Output:

```
(0.094299422053193811,spldefence,ice,water,290,70,40)
(0.04165872846874874,Amoonguss,Grass,Poison,464,114,85)
(0.04112818967833243,Liepard,Dark,,446,64,88)
(0.04036142442681656,Rhydon,Ground,Rock,485,105,130)
(0.040173611171296475,Magnemite,Electric,Steel,325,25,35)
(0.039372097531182204,Dusknoir,Ghost,,525,45,100)
(0.037787099482793685,Sunflora,Grass,,425,75,75)
(0.037366973223345146,Swablu,Normal,Flying,310,45,40)
(0.03729334521797689,Pyroar,Fire,Normal,507,86,68)
(0.036455393413030945,Numel,Fire,Ground,305,60,60)
(0.035944131069604035,ManectricMega Manectric,Electric,,575,70,75)
(0.03493765331300447,Roselia,Grass,Poison,400,50,60)
(0.03160041474391129,Simisage,Grass,,498,75,98)
(0.029942553769914615,Xatu,Psychic,Flying,470,65,75)
(0.027883790227097305,Typhlosion,Fire,,534,78,84)
(0.027694738598617152,Stunky,Poison,Dark,329,63,63)
(0.025001917548661323,MeloettaPirouette Forme,Normal,Fighting,600,100,128)
(0.024652717996032414,Bastiodon,Rock,Steel,495,60,52)
(0.022148741829219776,Fennekin,Fire,,307,40,45)
(0.02178287043957572,Frillish,Water,Ghost,335,55,40)
(0.02161278914864873,Makuhita,Fighting,,237,72,60)
(0.019561773562484674,Helioisk,Electric,Normal,481,62,55)
(0.018044909812275356,Sentret,Normal,,215,35,46)
(0.014442984279749105,Blissey,Normal,,540,255,10)
(0.013950454091826159,Scolipede,Bug,Poison,485,60,100)
(0.012985361702545717,Charmeleon,Fire,,405,58,64)
(0.009944563409446605,Archeops,Rock,Flying,567,75,140)
(0.009788104510038975,Lumineon,Water,,460,69,69)
(0.009692108790563925,Beedrill,Bug,Poison,395,65,90)
(0.007265466103542506,Dustox,Bug,Poison,385,60,50)
(0.004670869161197633,Deerling,Normal,Grass,335,60,60)
(0.0024079473065039547,MeowsticFemale,Psychic,,466,74,48)
grunt>
```

Problem5:

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

Solution:

```
-- Load Data
pokemon_data = LOAD '/home/acadgild/ankita/Assignment5_3/Pokemon' USING PigStorage(',')
AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int,splatck:int
,spldefence:int,speed:int);

qualifying_data = FILTER pokemon_data by Defense>55 ;

-- generate random for relation 2
random_include2 = foreach qualifying_data GENERATE
RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,splatck,spldefence,speed;

-- sort by random function relation 2
random2_desending = ORDER random_include2 BY $0 DESC;

-- display results of random function for relation 2
dump random2 desending;
```

Output:

```
Acadgild_64bit [Running] - Oracle VM VirtualBox
Help
(0.061221507419750854,Slaking,Normal,,670,150,160,100,95,65,100)
(0.05973322167998496,Shieldon,Rock,Steel,350,30,42,118,42,88,30)
(0.05877625974413436,ShayminSky Forme,Grass,Flying,600,100,103,75,120,75,127)
(0.057843802006208644,GourgeistAverage Size,Ghost,Grass,494,65,90,122,58,75,84)
(0.05146102507206263,Vullaby,Dark,Flying,370,70,55,75,45,65,60)
(0.051023371370402212,Blastoise,Water,,530,79,83,100,85,105,78)
(0.0501736619242554,Ludicolo,Water,Grass,480,80,70,70,90,10,70)
(0.04621713179729496,Roggenrola,Rock,,280,55,75,85,25,25,15)
(0.045880188729285454,LameneoTherianForme,Ground,Flying,600,89,145,90,105,80,91)
(0.04527653315202207,Milotic,Water,,540,95,60,79,100,125,81)
(0.04356814012718124,Staraptor,Normal,Flying,485,85,120,70,50,60,100)
(0.04216791676547238,Aron,Steel,Rock,330,50,70,100,40,40,30)
(0.040810406128500176,Steelix,Steel,Ground,510,75,85,200,55,65,30)
(0.037225250533054144,Stantler,Normal,,465,73,95,62,85,65,85)
(0.035704358086331434,Uxie,Psychic,,580,75,75,130,75,130,95)
(0.03303774562784334,Rhydon,Ground,Rock,485,105,130,120,45,45,40)
(0.03060338073657687,Dragonair,Dragon,,420,61,84,65,70,70,70)
(0.02926342997471676,SlowbroMega Slowbro,Water,Psychic,590,95,75,180,130,80,30)
(0.027117063636353534,Rhyperior,Ground,Rock,535,115,140,130,55,55,40)
(0.024692626829211695,Clefable,Fairy,,483,95,70,73,95,90,60)
(0.022215360835222242,AggronMega Aggron,Steel,,630,70,140,230,60,80,50)
(0.02205313006842413,Watchog,Normal,,420,60,85,69,60,69,77)
(0.01833620329872876,Hypno,Psychic,,483,85,73,70,73,115,67)
(0.018012100736982672,MedichamMega Medicham,Fighting,Psychic,510,60,100,85,80,85,100)
(0.017565557376526875,Leavanny,Bug,Grass,500,75,103,80,70,80,92)
(0.01637905407842899,Escavalier,Bug,Steel,495,70,135,105,60,105,20)
(0.015511508634055926,Axew,Dragon,,320,46,87,60,30,40,57)
(0.015143532281319927,Swanna,Water,Flying,473,75,87,63,87,63,98)
(0.014489447464343774,Magcargo,Fire,Rock,410,50,50,120,80,80,30)
(0.01322770433784004,RotomFan Rotom,Electric,Flying,520,50,65,107,105,107,86)
(0.009516746126589926,PinsirMega Pinsir,Bug,Flying,600,65,155,120,65,90,105)
(0.004090791779631231,Skuntank,Poison,Dark,479,103,93,67,71,61,84)
(0.0031295903100670985,Regirock,Rock,,580,80,100,200,50,100,50)
(0.002035366013455331,Weavile,Dark,Ice,510,70,120,65,45,85,125)
(0.786738501489697E-4,Skrelp,Poison,Water,320,50,60,60,60,60,30)
grunt>
```

Problem6:

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

Solution:

```
--Limit random function relation1

limit_random1 = LIMIT random1_descending 5;

--display result for random function for relation1

dump limit_random1;

--Limit random function relation2

limit_random2 = LIMIT random1_descending 5;

--display result for random function for relation1

dump limit_random2;
```

Output:

Result for random relation1---

```
2017-10-29 12:17:34,345 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total
2017-10-29 12:17:34,346 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil
cess : 1
(0.9996741787224201,Milotic,Water,,540,95,60,79,100,125,81)
(0.9988937827689965,LucarioMega Lucario,Fighting,Steel,625,70,145,88,140,70,112)
(0.9909755224168616,Gligar,Ground,Flying,430,65,75,105,35,65,85)
(0.9902450737069991,Empoleon,Water,Steel,530,84,86,88,111,101,60)
(0.9854132973103437,Gorebyss,Water,,485,55,84,105,114,75,52)
grunt> █
```

Result for random relation2---

```
2017-10-29 12:24:31,898 [main] INFO  org.apache.pig.backend.hadoop.executionengine.
cess : 1
(0.9994410696471268,Genesect,Bug,Steel,600,71,120,95,120,95,99)
(0.9974946918209349,Kyurem,Dragon,Ice,660,125,130,90,130,90,95)
(0.9974218714984913,Dedenne,Electric,Fairy,431,67,58,57,81,67,101)
(0.9967180869498755,Mesprit,Psychic,,580,80,105,105,105,105,80)
(0.9901542597845593,Torterra,Grass,Ground,525,95,109,105,75,85,56)
grunt> █
```

Problem7:

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).

Solution:

```
-- Load Data
pokemon_data = LOAD '/home/acadgild/ankita/Assignment5_3/Pokemon' USING PigStorage(',')
AS(Sno:int,Name:chararray,Type1:chararray,Type2:chararray,Total:int,HP:int,Attack:int,Defense:int,splatck:int
,spldefence:int,speed:int);

qualifying_data = FILTER pokemon_data by Defense>55 ;

-- generate random for relation 1
random_include1 = foreach qualifying_data GENERATE
RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,splatck,spldefence,speed;

-- sort by random function relation 1
random1_desending = ORDER random_include1 BY $0 DESC;

-- generate random for relation 2
random_include2 = foreach qualifying_data GENERATE
RANDOM(),Name,Type1,Type2,Total,HP,Attack,Defense,splatck,spldefence,speed;

-- sort by random function relation 2
random2_desending = ORDER random_include2 BY $0 DESC;

-- top 5 from relation 1
limit_random1 = LIMIT random1_desending 5 ;

-- top 5 from relation 2
limit_random2 = LIMIT random2_desending 5 ;

-- filter name from relation 1
filter_only_name1 = foreach limit_random1 Generate ($1,HP);

-- filter name from relation 2
filter_only_name2 = foreach limit_random2 Generate ($1,HP);

-- store relation 1 data to local file
STORE limit_random1 INTO '/home/acadgild/ankita/Assignment5_3/player1.txt';

-- store relation 2 data to local file
STORE limit_random2 INTO '/home/acadgild/ankita/Assignment5_3/player2.txt';
```

Output:

Result of Player1.txt—

```
2017-10-29 12:35:11,223 [main] INFO c
!
grunt> cat Player1.txt;
0.9992667743956373 Malamar Dark Psychic 482 86 92 88 68 75 73
0.9955287665740217 GlalieMega Glalie Ice 580 80 120 80 120 80 100
0.9942589930765201 Charmeleon Fire 405 58 64 58 80 65 80
0.9937828016672765 Gardevoir Psychic Fairy 518 68 65 65 125 115 80
0.9933297291887441 Gigalith Rock 515 85 135 130 60 80 25
grunt> █
```

```
2017-10-29 12:35:11,223 [main] INFO c
cess : 1
((Swanna,75))
((Gabite,68))
((Exeggutor,95))
((Dragonair,61))
((Shellder,30))
grunt> █
```

Result of Player2.txt—

```
grunt> cat player2.txt;
0.9990708326219241 Bronzong Steel Psychic 500 67 89 116 79 116 33
0.9990272659534193 Crawdaunt Water Dark 468 63 120 85 90 55 55
0.9987790959169232 Exploud Normal 490 104 91 63 91 73 68
0.9928881560154161 Bronzor Steel Psychic 300 57 24 86 24 86 23
0.9927841521844853 Reuniclus Psychic 490 110 65 75 125 85 30
grunt> █
```

```
2017-10-29 12:37:54,966 [main] INFO o
cess : 1
((Gothitelle,70))
((Slowking,95))
((Scizor,70))
((Gliscor,75))
((Nidoqueen,90))
grunt> █
```