Exercise 1:

Commands -> foodratings = LOAD 'foodratings7936.txt'

USING PigStorage(',')

AS (name:chararray, f1:int, f2:int, f3:int, f4:int, placeid:int);

-> food_ratings = FOREACH foodratings GENERATE name as name, f1 as f1, f2 as f2, f3 as f3, f4 as f4, placeid as placeid;

->DUMP food_ratings;

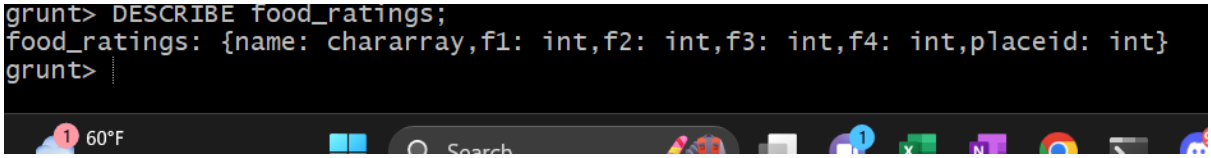
Magic Number:

```
__MACOSX  pigdemo  pigdemo.zip  TestDataGen.class
[hadoop@ip-172-31-24-36 ~]$ java TestDataGen
Magic Number = 7936
[hadoop@ip-172-31-24-36 ~]$
```

Output:

```
hadoop@ip-172-31-26-6:~/pigdemo
(Sam,22,32,31,4,4)
(Jill,16,17,38,15,2)
(Mel,8,27,10,50,5)
(Jill,32,29,46,5,5)
(Joy,34,31,40,28,1)
(Mel,41,27,5,43,4)
(Joe,24,38,18,38,4)
(Mel,13,49,13,45,5)
(Joe,32,33,49,18,1)
(Jill,24,19,3,27,2)
(Mel,50,19,46,18,1)
(Mel,5,18,25,4,2)
(Jill,17,18,50,9,3)
(Mel,27,14,13,3,2)
(Joe,44,16,22,29,5)
(Mel,32,43,34,33,5)
(Sam,43,33,5,50,3)
(Jill,50,15,24,49,2)
(Sam,1,32,23,11,5)
(Joy,4,30,48,21,4)
(Joe,38,18,25,50,1)
(Mel,30,20,27,32,5)
(Sam,46,40,27,14,5)
(Sam,25,1,8,38,3)
```

Command -> DESCRIBE food_ratings;



Exercise 2 :

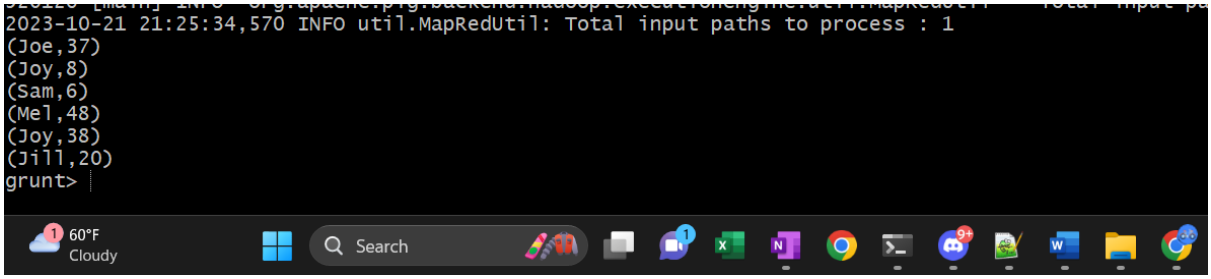
Command -> food_ratings_subset = FOREACH food_ratings GENERATE name, f4;

->STORE food_ratings_subset INTO '/user/hadoop/fr_subset' USING PigStorage();

->output = LIMIT food_ratings_subset 6;

->DUMP output;

Output:



Exercise 3

Commands ->

```
f2_stats = FOREACH (GROUP food_ratings ALL) GENERATE
   MIN(food_ratings.f2) AS f2_min,
   MAX(food_ratings.f2) AS f2_max,
   AVG(food_ratings.f2) AS f2_avg;
f3_stats = FOREACH (GROUP food_ratings ALL) GENERATE
   MIN(food_ratings.f3) AS f3_min,
   MAX(food_ratings.f3) AS f3_max,
   AVG(food_ratings.f3) AS f3_avg;
food_ratings_profile = JOIN f2_stats BY f2_min, f3_stats BY f3_min;
final_output = FOREACH food_ratings_profile GENERATE
   f2_min, f2_max, f2_avg, f3_min, f3_max, f3_avg;
DUMP final_output;
```

```
2023-10-21 21:28:53,708 INFO input.FileInputFormat: Total input files to process : 1
1119264 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths to process : 1
2023-10-21 21:28:53,708 INFO util.MapRedUtil: Total input paths to process : 1
(1,50,25.707,1,50,25.184)
grunt>
```

Exercise 4

Commands ->

```
food_ratings_filtered = FILTER food_ratings BY (f1 < 20) AND (f3 > 5);

final_output = LIMIT food_ratings_filtered 6;
DUMP final_output;
```

```
2023-10-21 21:30:58,263 INFO util.MapRedUtil: Total input paths to process : 1
(Sam,4,42,41,6,4)
(Joy,2,31,7,38,1)
(Joe,9,31,48,36,5)
(Joe,1,33,13,5,1)
(Sam,6,38,9,30,1)
(Joe,5,30,32,43,1)
grunt>
```

Exercise 5

Commands ->

```
food_ratings_2percent = SAMPLE food_ratings 0.02;


final_output = LIMIT food_ratings_2percent 10;

DUMP final_output;
```

```
2023-10-22 02:48:18,306 INFO util.MapRedUtil: Total input paths to process : 1
(Joy,44,45,37,26,4)
(Jill,13,5,2,44,1)
(Joy,40,49,26,42,5)
(Joy,26,29,13,37,1)
(Joy,46,30,1,10,4)
(Sam,46,38,24,19,4)
(Jill,3,38,25,20,3)
(Sam,27,27,2,20,2)
(Mel,11,28,27,28,1)
(Joe,44,46,6,32,5)
grunt>
```

Exercise 6:

Commands->

food_places = LOAD 'foodplaces7936.txt' USING PigStorage(',') AS (placeid: int, placename: chararray);

DESCRIBE food_places;

food_ratings_w_place_names = JOIN food_ratings BY placeid, food_places BY placeid;

final_output = LIMIT food_ratings_w_place_names 6;
DUMP final_output;

```
publisher.enabled
grunt> DESCRIBE food_places;
food_places: {placeid: int,placename: chararray}
grunt>
```

```
2023-10-22 13:44:11,148 INFO util.MapRedUtil: Total input paths to process : 1
(Joe,10,43,14,40,1,1,China Bistro)
(Sam,1,40,32,27,1,1,China Bistro)
(Joe,5,40,4,49,1,1,China Bistro)
(Jill,28,46,29,5,1,1,China Bistro)
(Joy,40,7,14,8,1,1,China Bistro)
(Joe,46,14,50,43,1,1,China Bistro)
grunt>
```

```
                    FileBytesWritten: 19441
                       HdfsBytesRead: 17537
                    HdfsBytesWritten: 210
     SpillableMemoryManager spill count: 0
               Bags proactively spilled: 0
            Records proactively spilled: 0

DAG Plan:
Tez vertex scope-462    ->      Tez vertex scope-464,
Tez vertex scope-463    ->      Tez vertex scope-464,
Tez vertex scope-464    ->      Tez vertex scope-466,
Tez vertex scope-466

Vertex Stats:
VertexId Parallelism TotalTasks   InputRecords   ReduceInputRecords  OutputRecords  FileBytesRead FileBytesWritten  HdfsBytesRead HdfsBytesWritten Alias     F
eature  Outputs
scope-462         1          1           1000                   0           1000           112            9770          17478               0 food_rating
s,food_ratings_w_place_names,foodratings
scope-463         1          1              5                   0              5           112             200             59               0 food_places
,food_ratings_w_place_names
scope-464         2          1              0                 176              6          9471            9471              0               0 food_rating
s_w_place_names,food_ratings_w_place_names_sample       HASH_JOIN
scope-466         1          1              6                   0              6             0               0              0             210 food_rating
s_w_place_names_sample  LIMIT   hdfs://ip-172-31-24-36.us-east-2.compute.internal:8020/tmp/temp1364657160/tmp2064759368,

Input(s):
Successfully read 5 records (59 bytes) from: "hdfs://ip-172-31-24-36.us-east-2.compute.internal:8020/user/hadoop/foodplaces7936.txt"
Successfully read 1000 records (17478 bytes) from: "hdfs://ip-172-31-24-36.us-east-2.compute.internal:8020/user/hadoop/foodratings7936.txt"

Output(s):
Successfully stored 6 records (210 bytes) in: "hdfs://ip-172-31-24-36.us-east-2.compute.internal:8020/tmp/temp1364657160/tmp2064759368"

2023-10-22 13:44:11,148 INFO input.FileInputFormat: Total input files to process : 1
872538 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  - Total input paths to process : 1
2023-10-22 13:44:11,148 INFO util.MapRedUtil: Total input paths to process : 1
(Joe,10,43,14,40,1,1,China Bistro)
(Sam,1,40,32,27,1,1,China Bistro)
(Joe,5,40,4,49,1,1,China Bistro)
(Jill,28,46,29,5,1,1,China Bistro)
(Joy,40,7,14,8,1,1,China Bistro)
(Joe,46,14,50,43,1,1,China Bistro)
grunt> [1]+  Killed                  pig
Killed
[hadoop@ip-172-31-24-36 pigdemo]$
```

Exercise 7:

1) A
2) C
3) B
4) B
5) B
6) A