Exercise 1:

**Commands -> foodratings = LOAD 'foodratings7936.txt'**

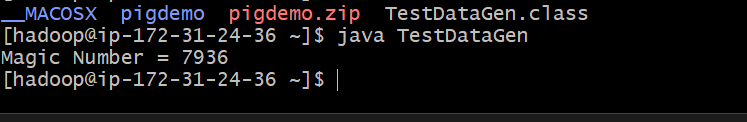
**USING PigStorage(',')**

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

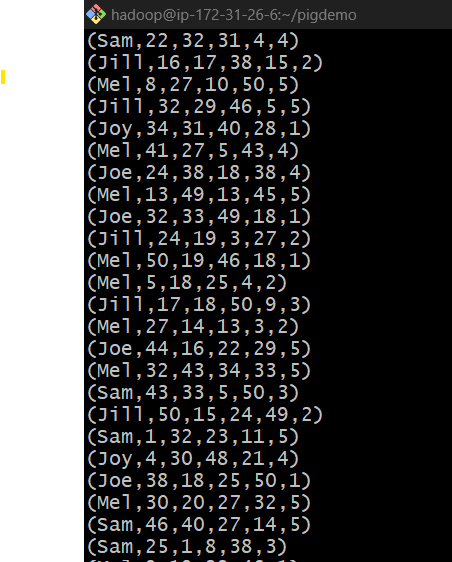
**-> food\_ratings = FOREACH foodratings GENERATE name as name, f1 as f1, f2 as f2, f3 as f3, f4 as f4, placeid as placeid;**

**->DUMP food\_ratings;**

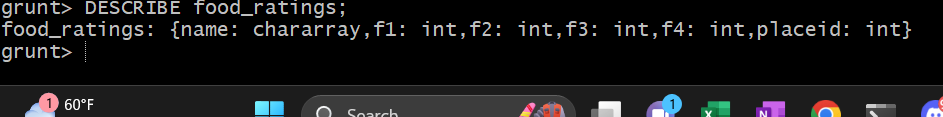
Magic Number:



Output:



Command -> DESCRIBE food\_ratings;



Exercise 2 :

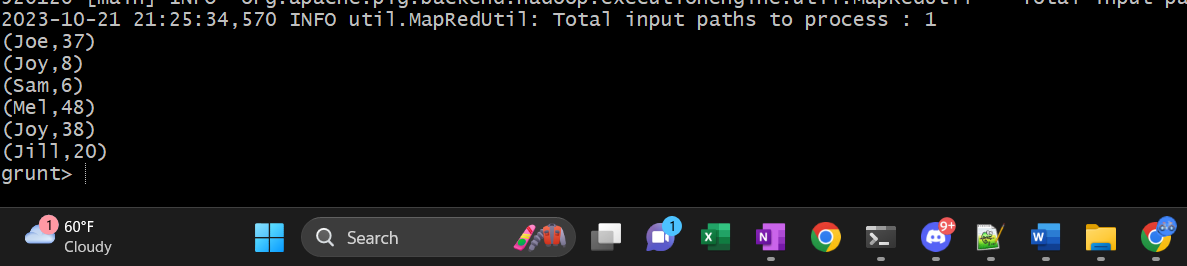
**Command -> food\_ratings\_subset = FOREACH food\_ratings GENERATE name, f4;**

**->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,

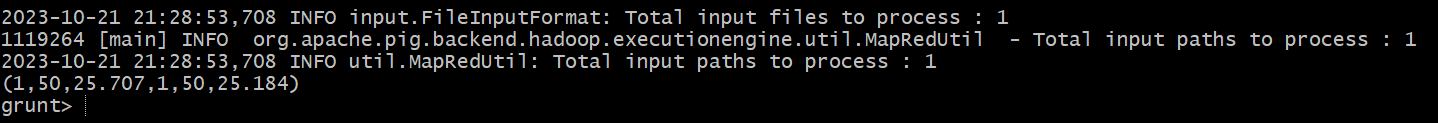
AVG(food\_ratings.f3) AS f3\_avg;

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;



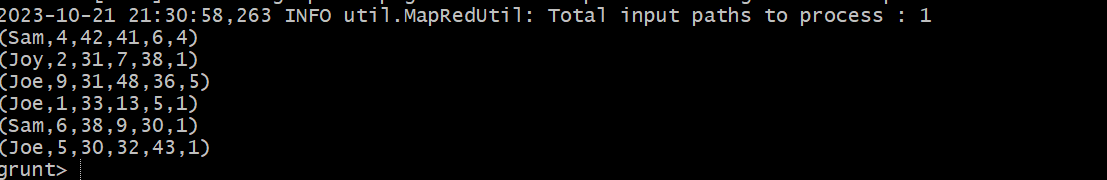
Exercise 4

Commands ->

food\_ratings\_filtered = FILTER food\_ratings BY (f1 < 20) AND (f3 > 5);

final\_output = LIMIT food\_ratings\_filtered 6;

DUMP final\_output;



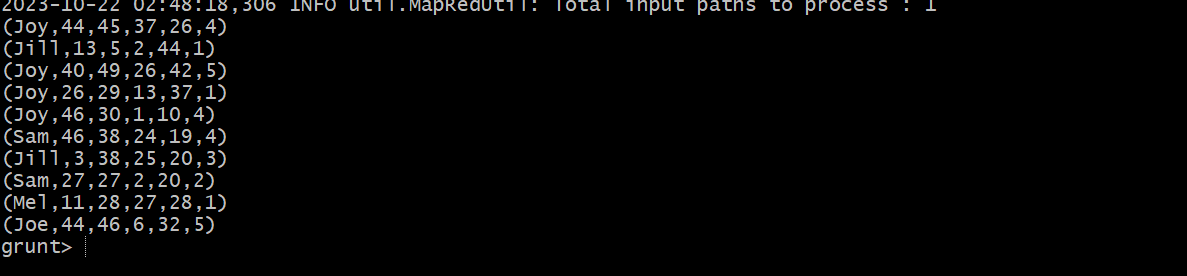
Exercise 5

Commands ->

food\_ratings\_2percent = SAMPLE food\_ratings 0.02;

final\_output = LIMIT food\_ratings\_2percent 10;

DUMP final\_output;



Exercise 6:

Commands->

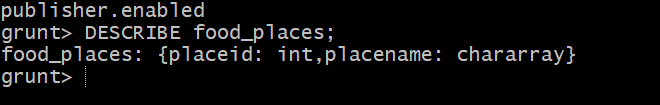
food\_places = LOAD 'foodplaces7936.txt' USING PigStorage(',') AS (placeid: int, placename: chararray);

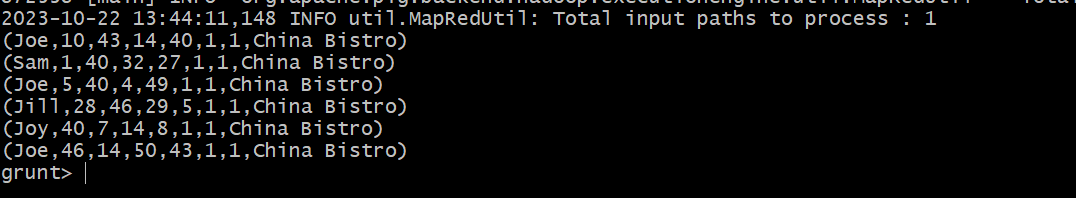
DESCRIBE food\_places;

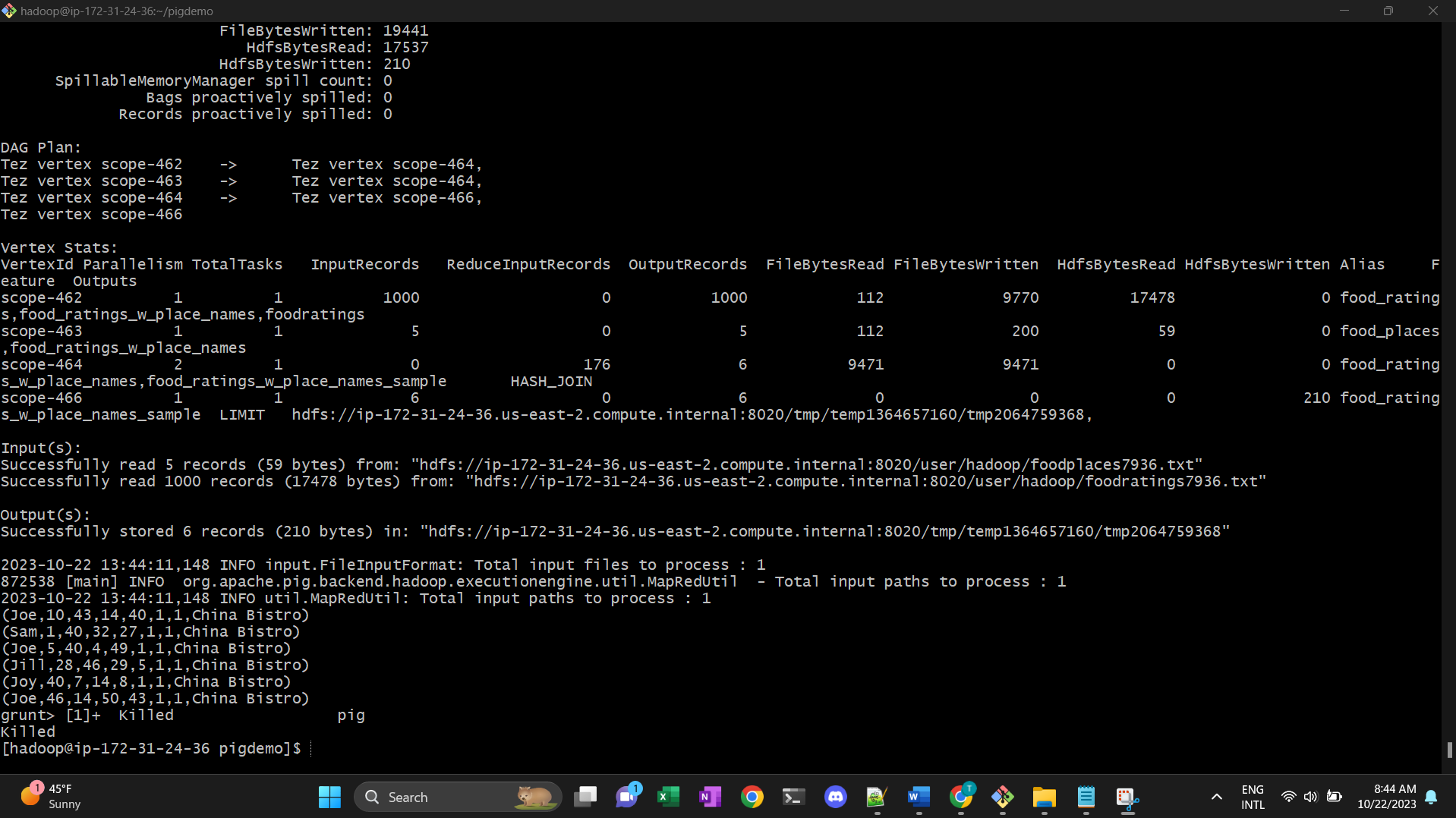
food\_ratings\_w\_place\_names = JOIN food\_ratings BY placeid, food\_places BY placeid;

final\_output = LIMIT food\_ratings\_w\_place\_names 6;

DUMP final\_output;







Exercise 7:

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