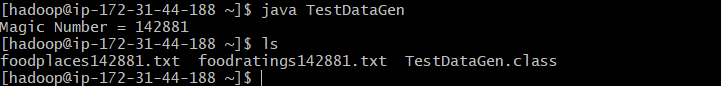
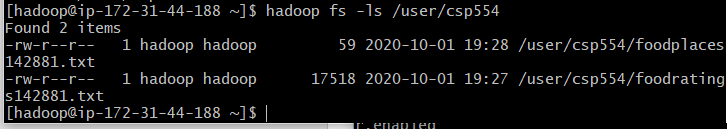
***1)Running TestDataGen***

***Magic Number:*** 142881

***Files Generated:*** foodplaces142881.txt and foodratings142881.txt

***Command Used:*** java TestDataGen

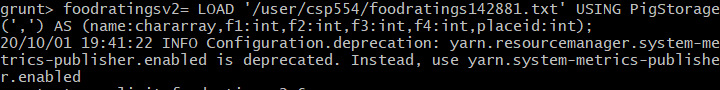




***Creating food\_ratings Relation:***

***Command Used to Create Relation:*** food\_ratings= LOAD 'user/hadoop/foodratings124881.txt' USING PigStorage(',') AS (name:CharArray,f1:int,f2:int,f3:int,f4:int,placeid:int);

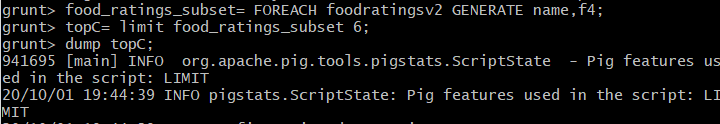
***Command Used to display Schema Relation:*** Describe food\_ratings;





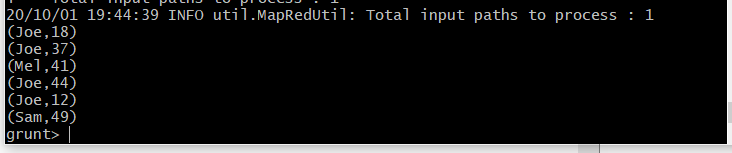
***2) Creating a new relation from existing***

***Command Used:*** food\_ratings\_subset= FOREACH foodratings GENERATE name,f4;



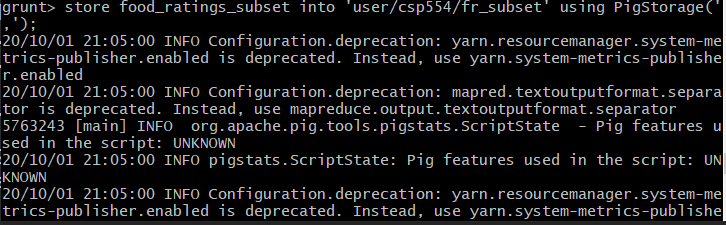
***Printing top 6 rows***

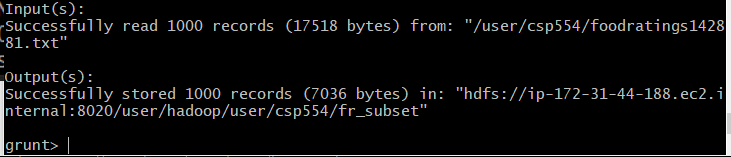
***Command Used:*** top6=limit food\_ratings\_subset 6;



***Storing into food\_ratings\_subset into HDFS***

***Command Used:*** store food\_ratings\_subset into 'user/csp554/fr\_subset' using PigStorage(',');



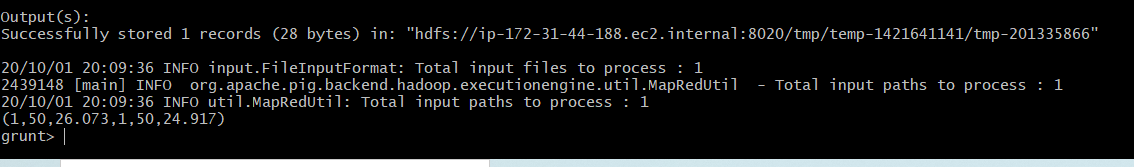


**3)** ***Calculating MIN,MAX,AVG of the f2,f3 in foodratings***

***COMMAND USED:*** food\_ratings\_profile= FOREACH foodratingsAll generate MIN(foodratingsv2.f2) as f2min,MAX(food\_ratingsv2.f2) as f2max,AVG(foodratingsv2.f2) as avgf2,MIN(foodratingsv2.f3) as f3min,MAX(foodratingsv2.f3) as f3max,AVG(foodratings.f3) as f3avg;

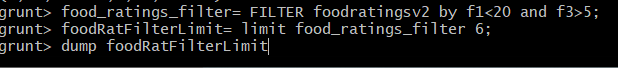






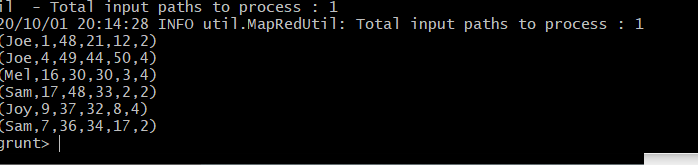
**4)*Filtering the records based on f1<20 and f3>5***

***Command Used:*** food\_ratings\_filter= FILTER foodratingsv2 by f1<20 and f3>5;



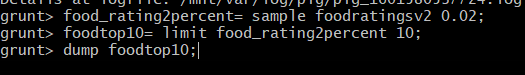
**Printing Top 6 rows**

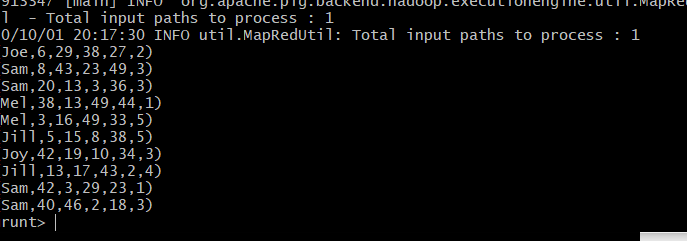
**Command Used:** foodRatFilterLimit= limit food\_ratings\_filter 6; dump foodRatFilterLimit



**5)** ***Sampling the 2 percent data randomly***

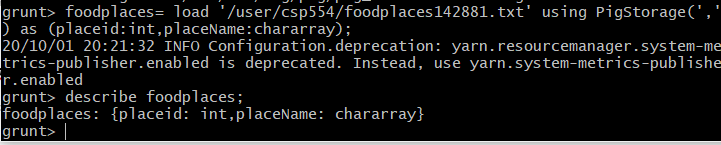
***Command Used :*** food\_rating2percent= sample foodratingsv2 0.02;





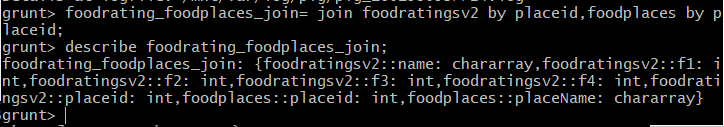
**6)** ***Loading foodplaces Data***

***Command Used:*** foodplaces= load '/user/csp554/foodplaces142881.txt' using PigStorage(',') as (placeid:int,placeName:chararray);

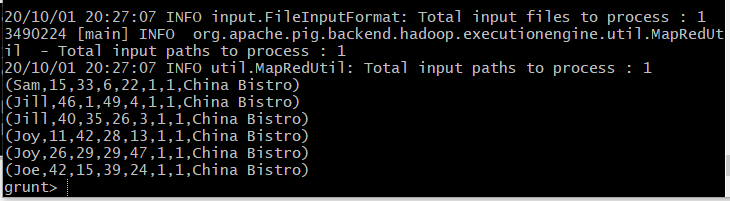


***Performing join between foodplaces and foodratings***

***Command Used:*** foodrating\_foodplaces\_join= join foodratingsv2 by placeid,foodplaces by placeid;







7)

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