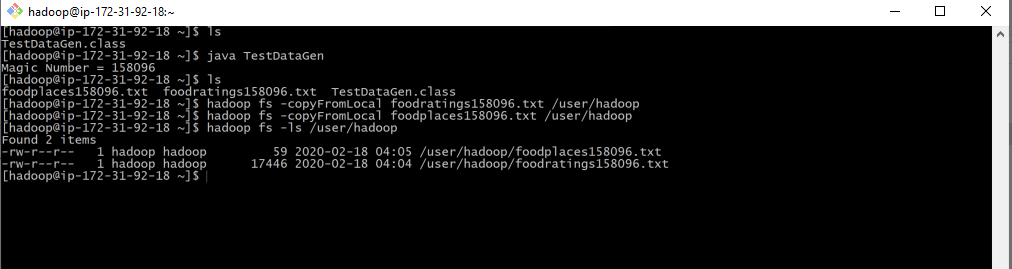
Rakshith Churchagundi Amarnath

A20424771

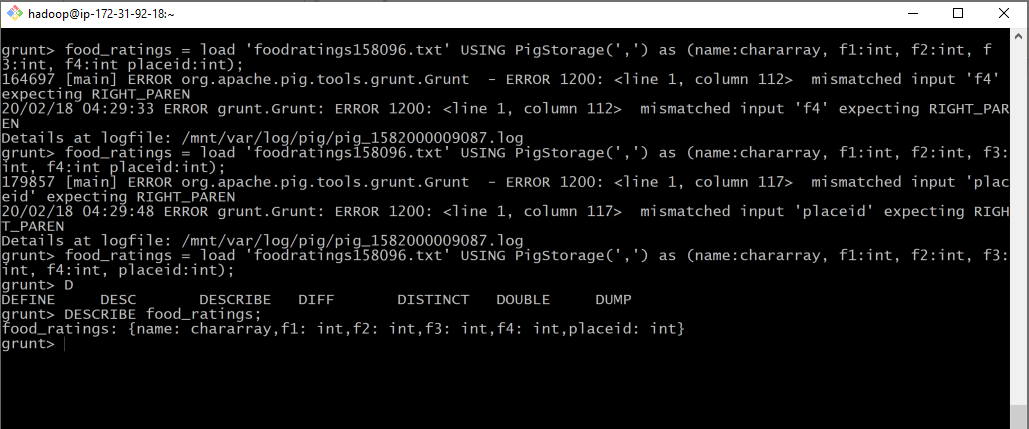
Assignment 5

1. Magic Number = 158096



food\_ratings = load 'foodratings158096.txt' USING PigStorage(',') as (name:chararray, f1:int, f2:int, f3:int, f4:int, placeid:int);

DESCRIBE food\_ratings;



2)

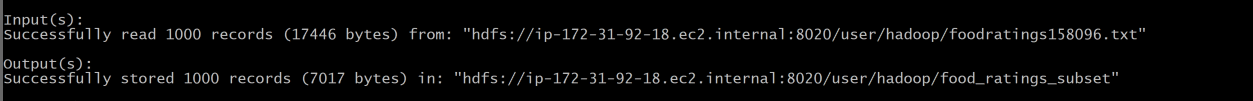
grunt> food\_ratings\_subset = FOREACH food\_ratings GENERATE name, f4;

grunt> store food\_ratings\_subset into 'food\_ratings\_subset';

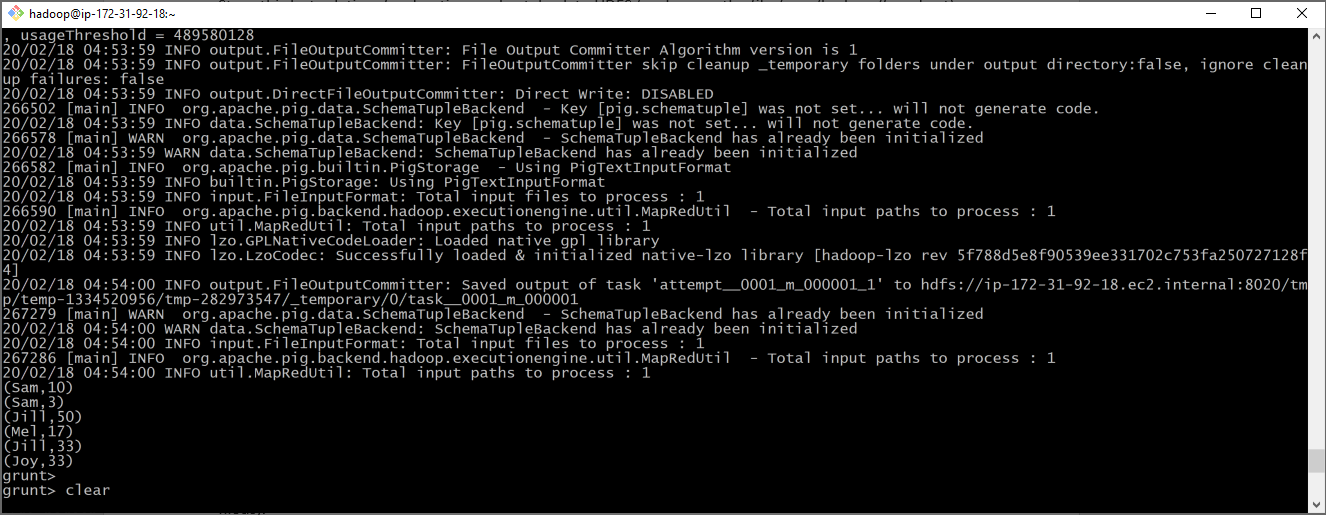
X = LIMIT food\_ratings\_subset 6;

DUMP x;







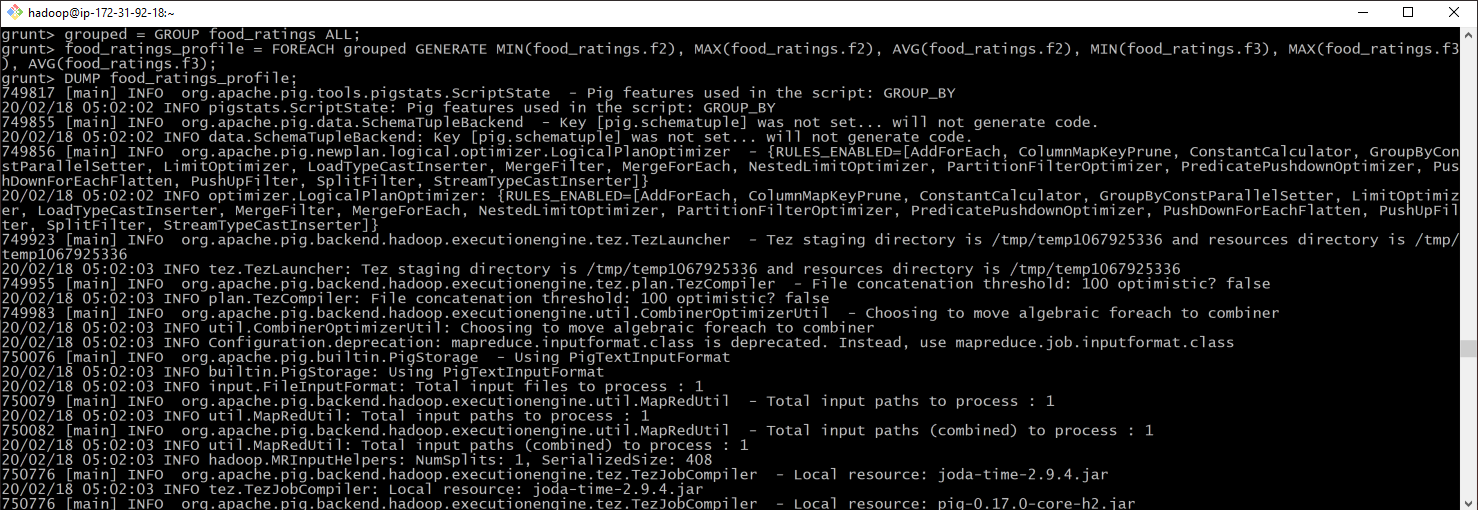


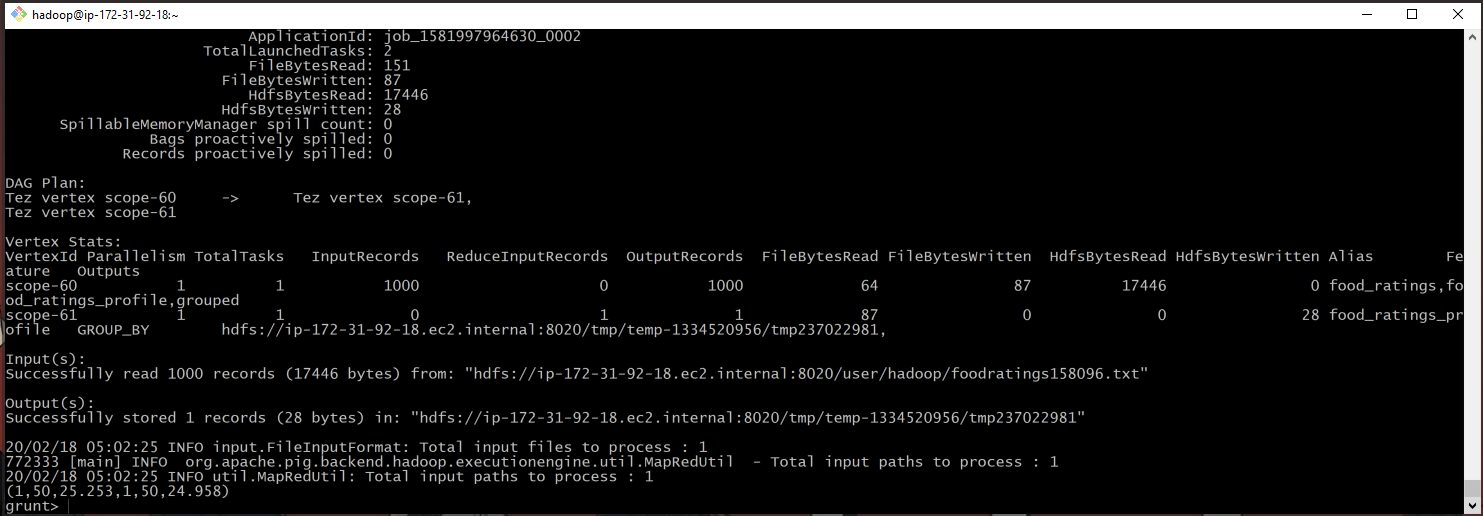
3)

grouped = GROUP food\_ratings ALL;

food\_ratings\_profile = FOREACH grouped GENERATE MIN(food\_ratings.f2), MAX(food\_ratings.f2), AVG(food\_ratings.f2), MIN(food\_ratings.f3), MAX(food\_ratings.f3), AVG(food\_ratings.f3);

DUMP food\_ratings\_profile



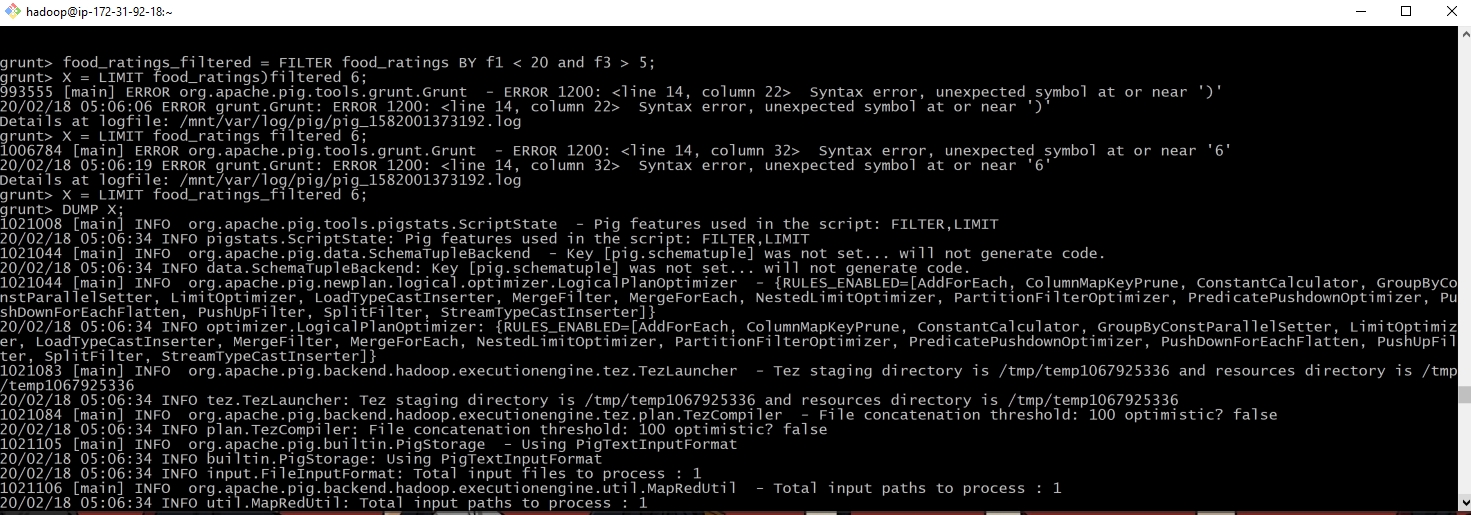


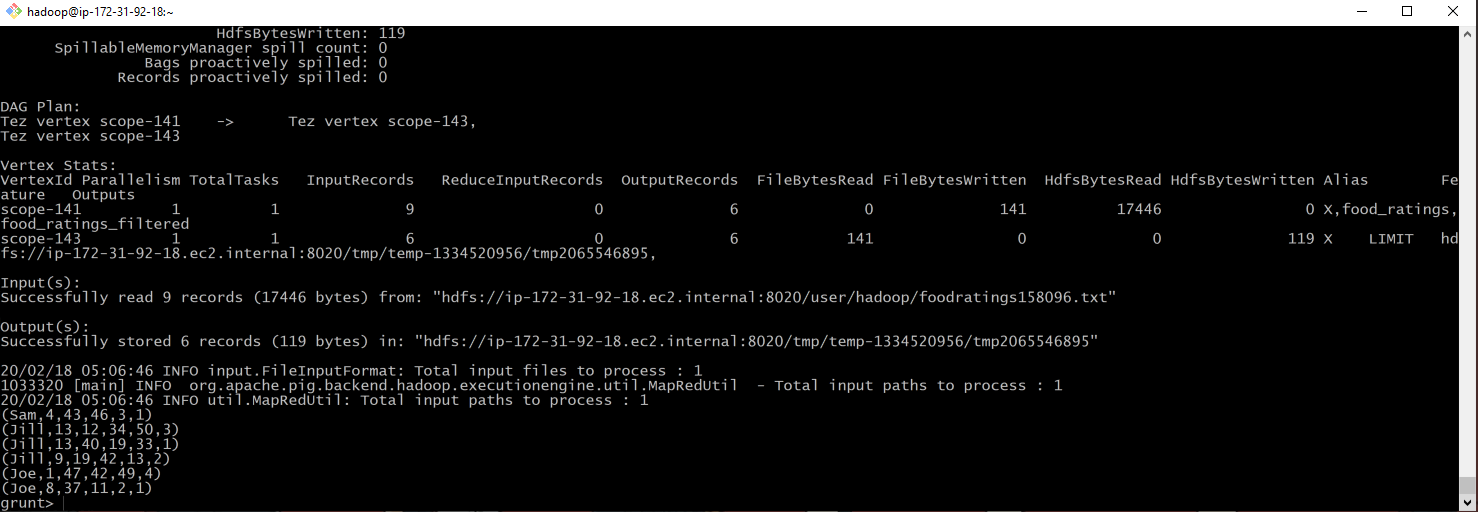
4)

food\_ratings\_filtered = FILTER food\_ratings BY f1 < 20 and f3 > 5;

X = LIMIT food\_ratings\_filtered 6;

DUMP X;



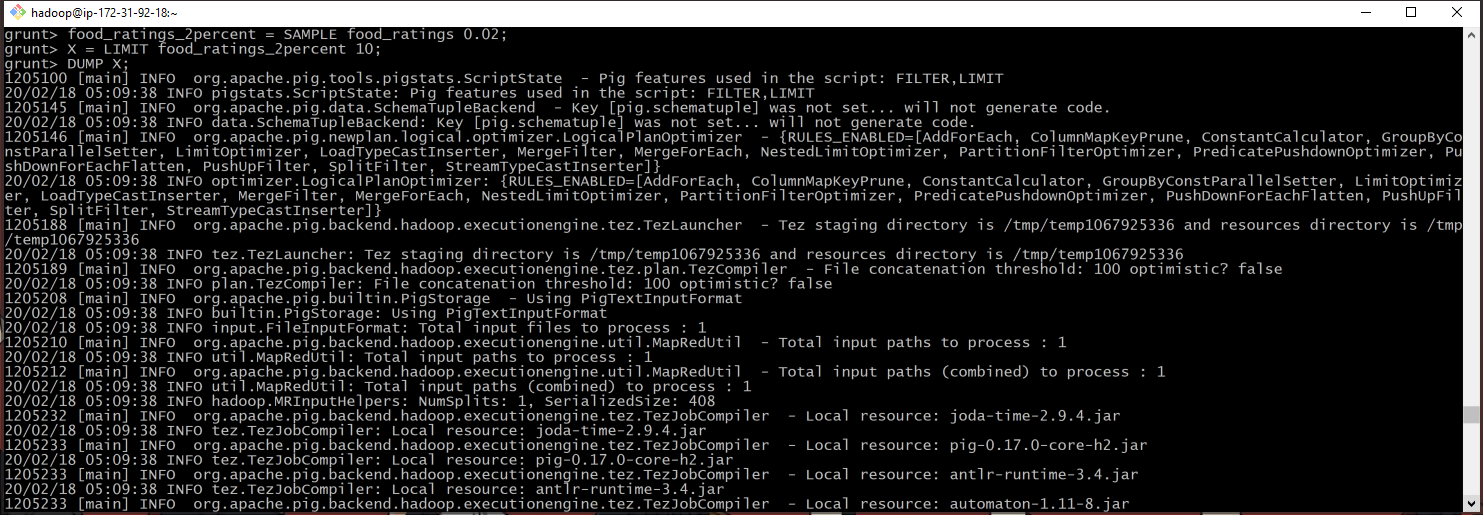


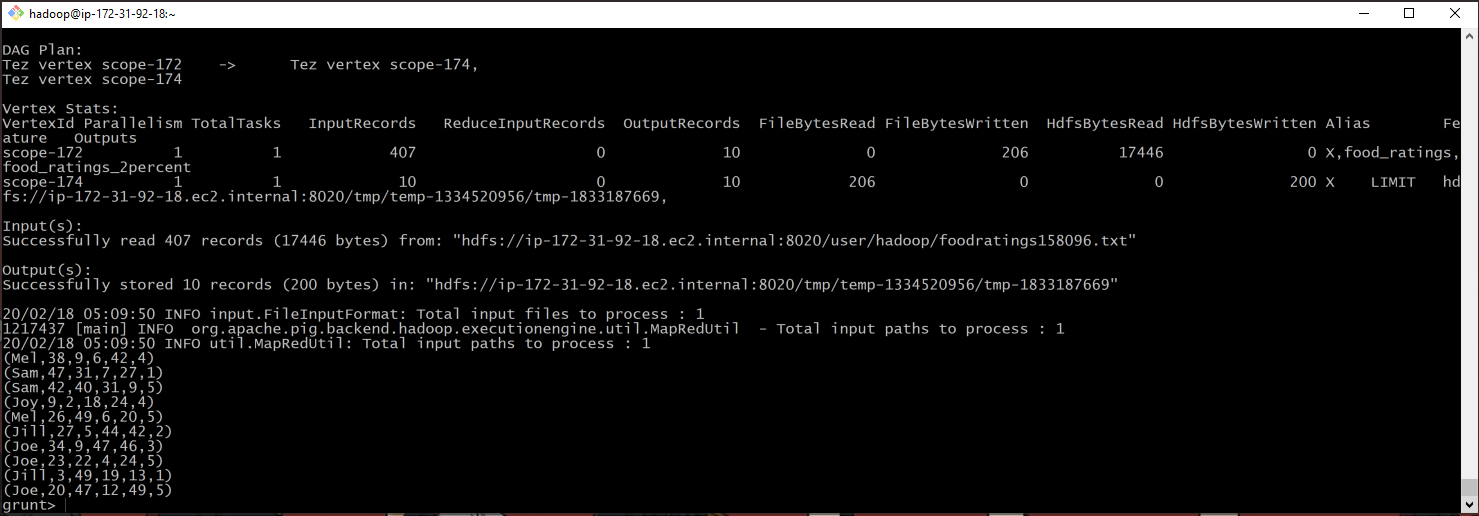
5.

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

X = LIMIT food\_ratings\_2percent 10;

DUMP X;





6)

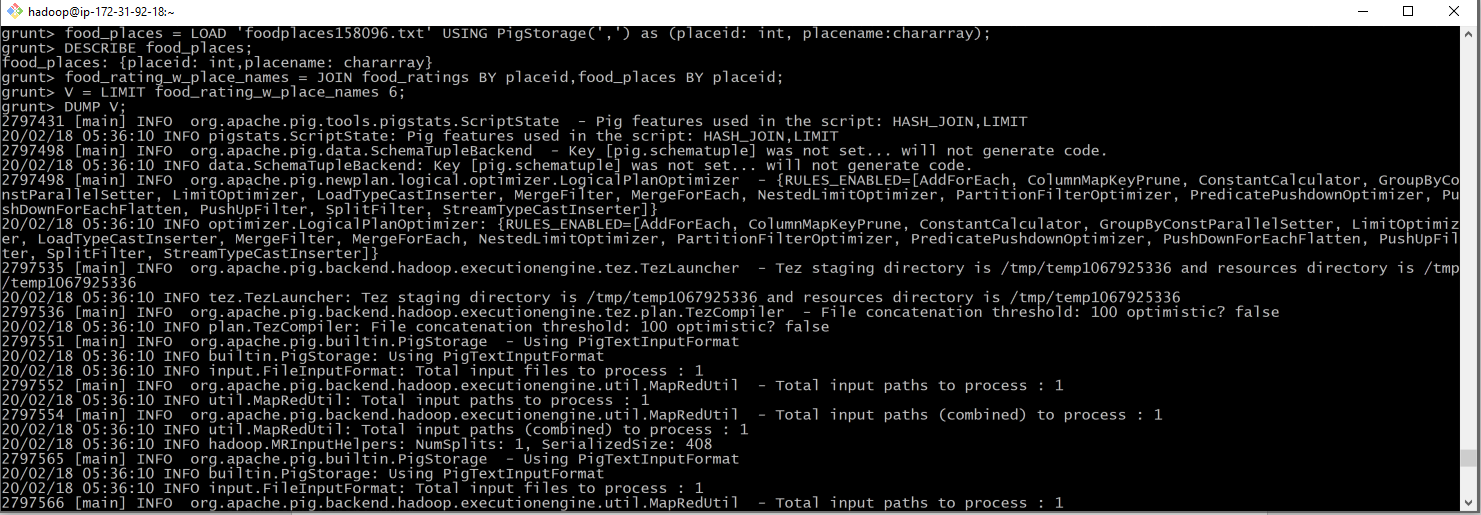
food\_places = load 'foodplaces158096.txt' USING PigStorage('.') as (placeid: int, placename:chararray);

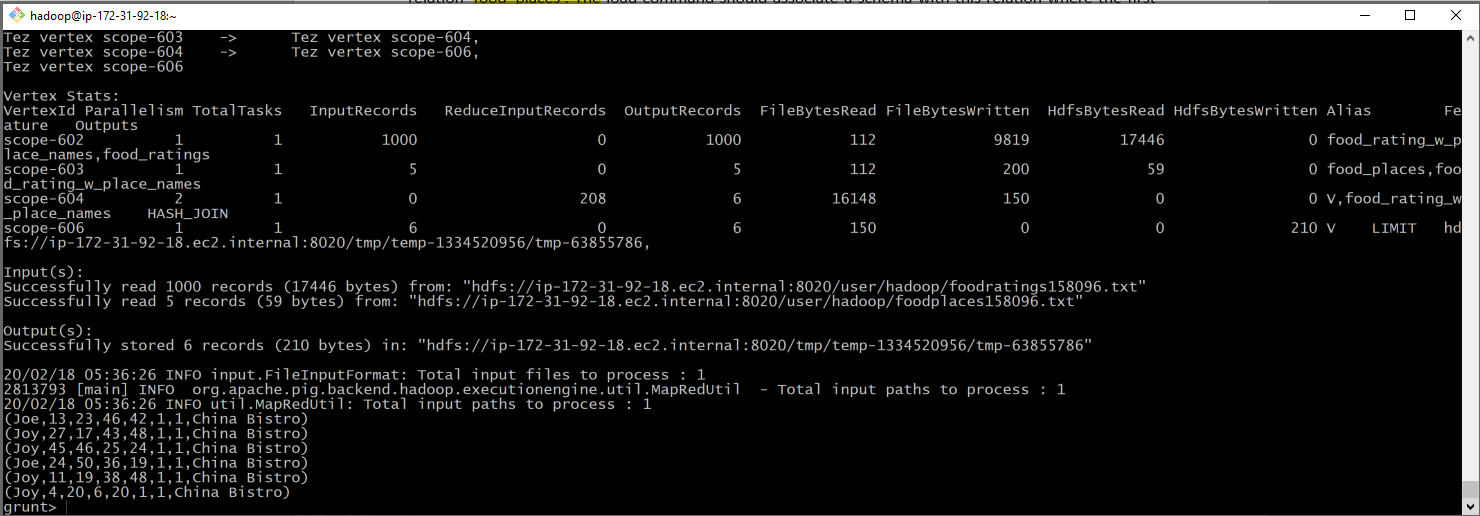
DESCRIBE food\_places;

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

V = LIMIT food\_ratings\_w\_place\_names 6;

DUMP V;





7)

I. (A) LIMIT

II. (C) DISTINCT

III. (B) (f1: STRING, f2: INT, f3: INT, f4: INT)

IV. (B) relB = FOREACH relA GENERATE $0, f3;

V. (B) data flow

VI. : (A) relB = FILTER relA by $0 < 20