SESSION 5 ASSIGNMENT 2

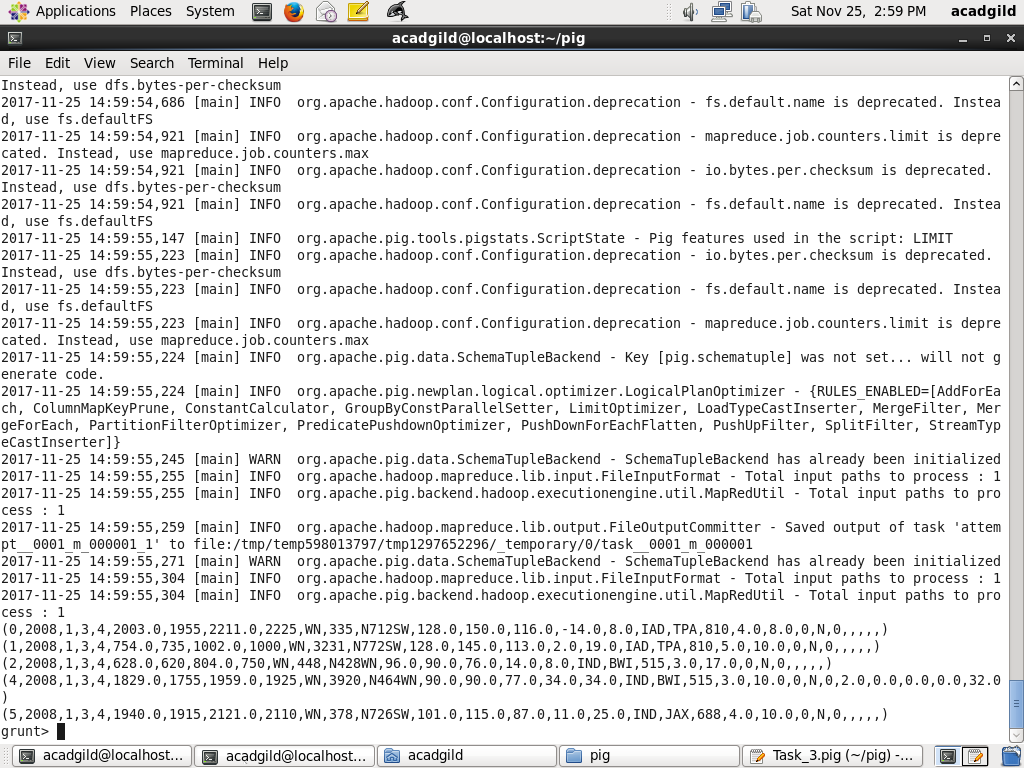
# PROBLEM 3

Q3. Top ten origins with the highest AVG departure delay.

REGISTER '/usr/local/pig/lib/piggybank.jar';

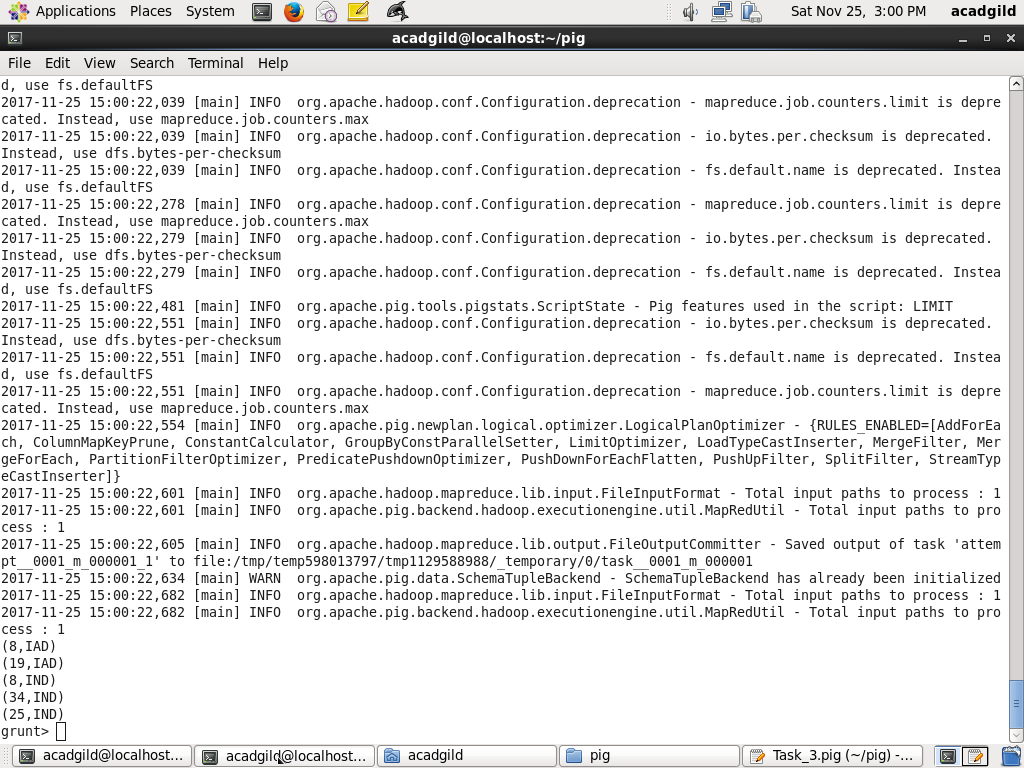
A = load 'DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

A\_LIMIT = LIMIT A 5;



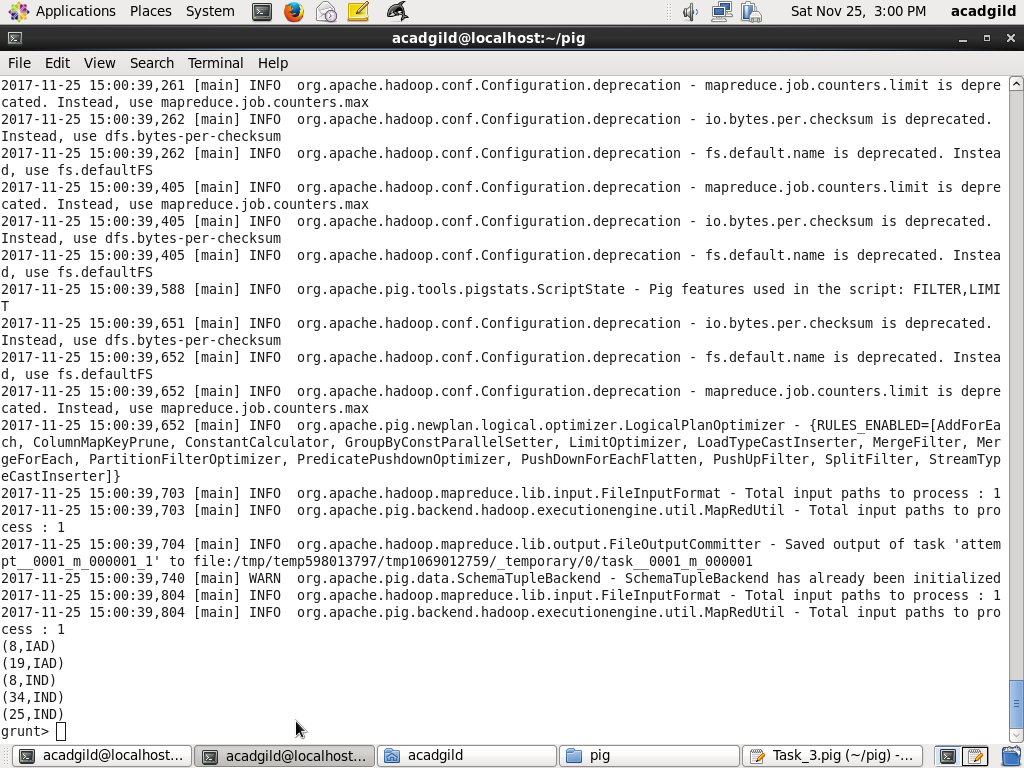
B1 = foreach A generate (int)$16 as dep\_delay, (chararray)$17 as origin;

B1\_LIMIT = LIMIT B1 5;



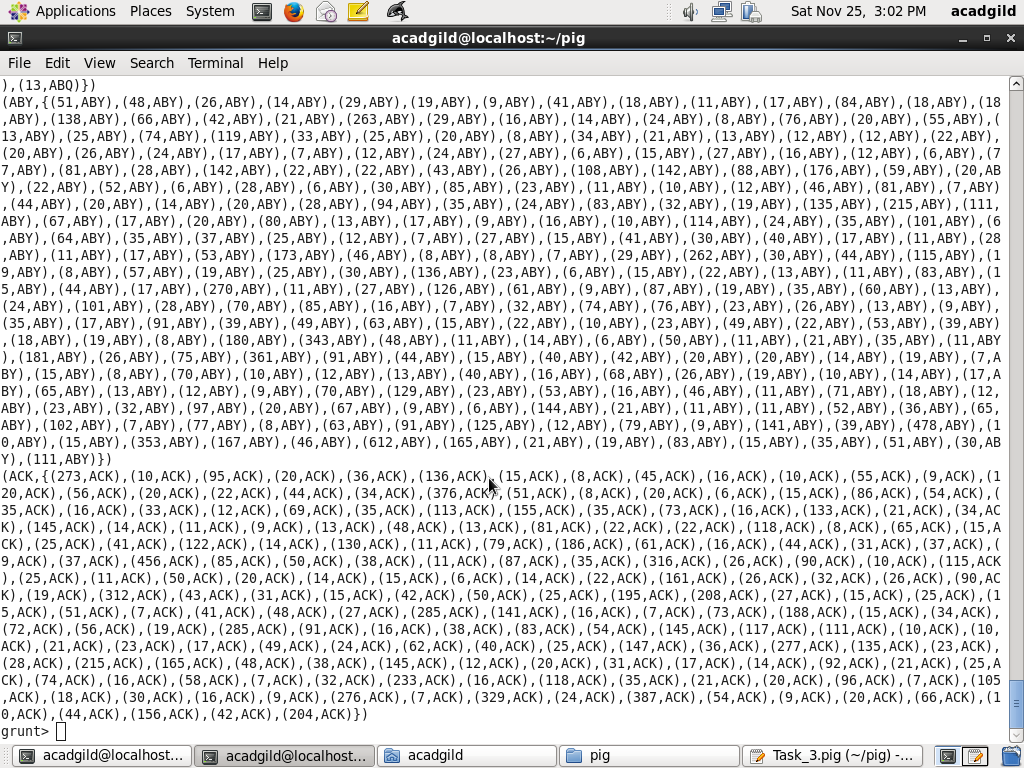
C1 = filter B1 by (dep\_delay is not null) AND (origin is not null);

C1\_LIMIT = LIMIT C1 5;



D1 = group C1 by origin;

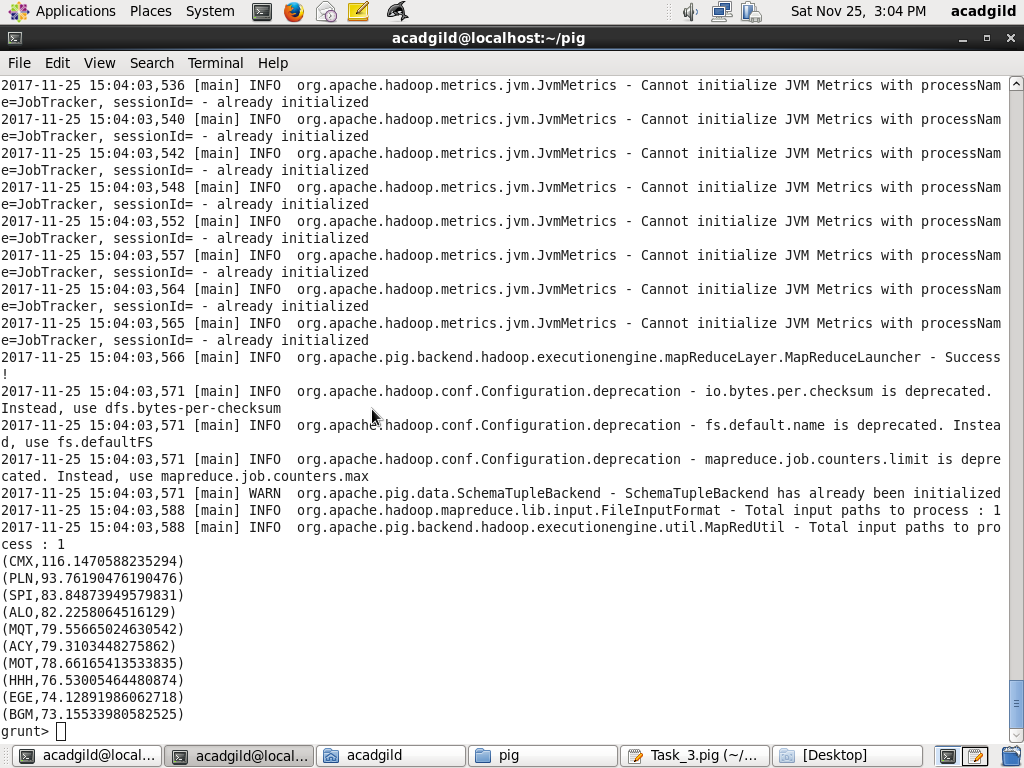
D1\_LIMIT = LIMIT D1 5;



E1 = foreach D1 generate group, AVG(C1.dep\_delay);

Result = order E1 by $1 DESC;

Top\_ten = limit Result 10;



Lookup = load 'airports.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

Lookup1 = foreach Lookup generate (chararray)$0 as origin, (chararray)$2 as city, (chararray)$4 as country;

Joined = join Lookup1 by origin, Top\_ten by $0;

Final = foreach Joined generate $0,$1,$2,$4;

Final\_Result = ORDER Final by $3 DESC;

dump Final\_Result;

