**ASSIGNMENT-5.2**

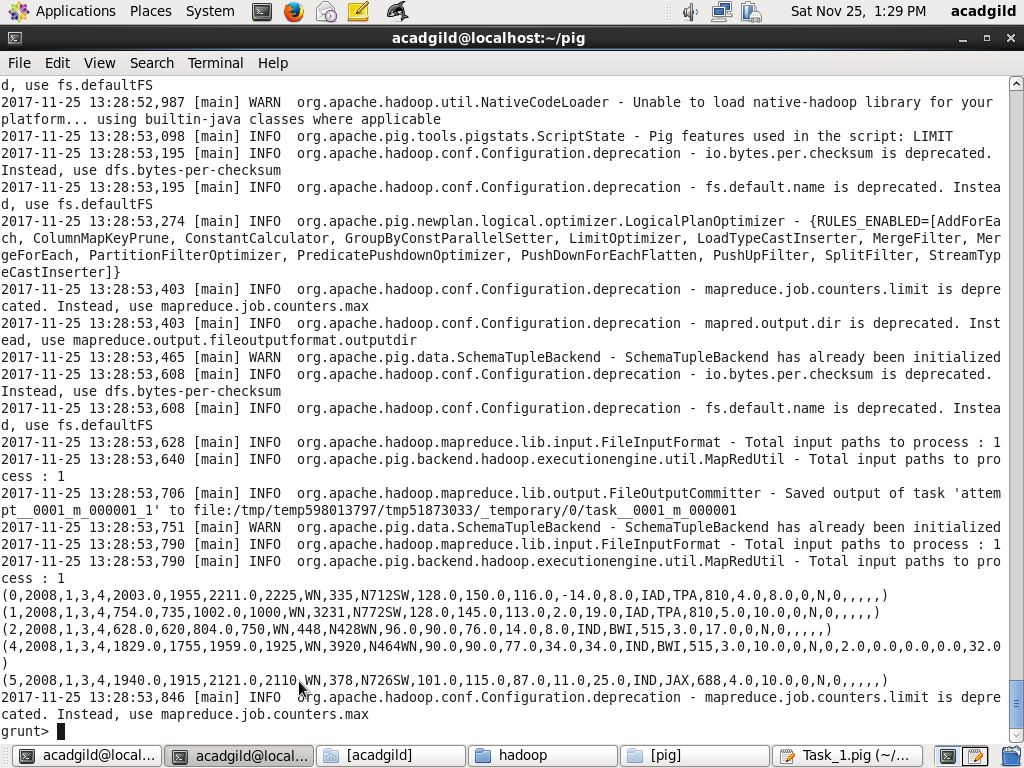
1. Find out the top 5 most visited destinations.

**QUERY 1:**

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

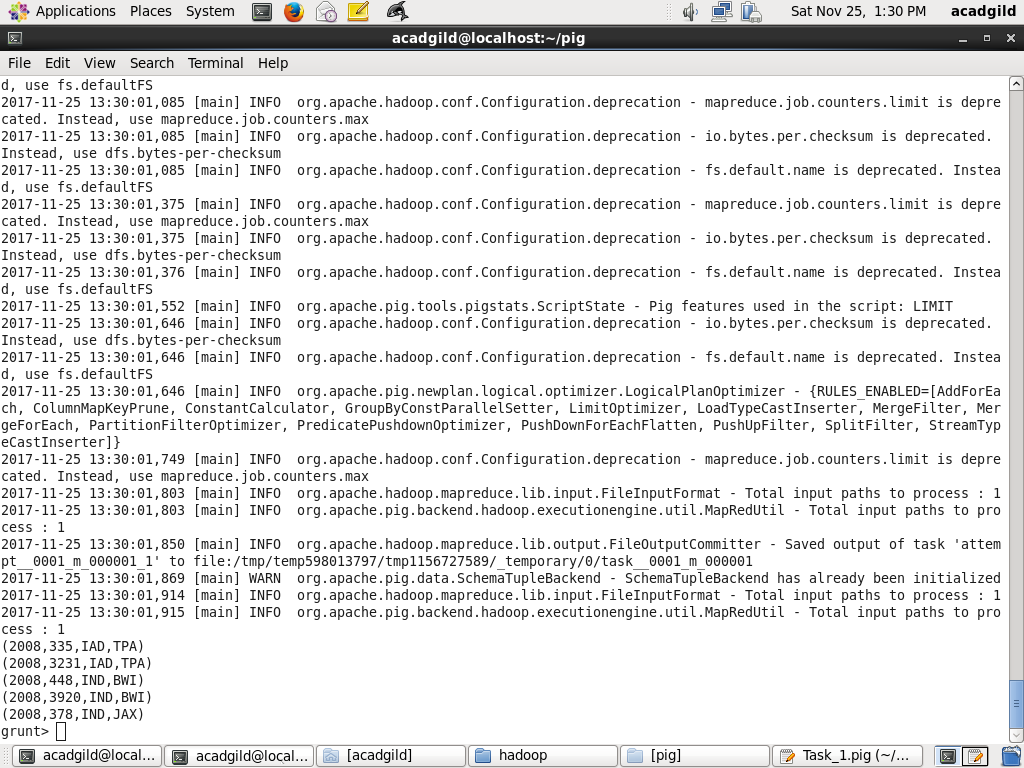
*A = load 'DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');*

*A\_LIMIT = LIMIT A 5;*

**

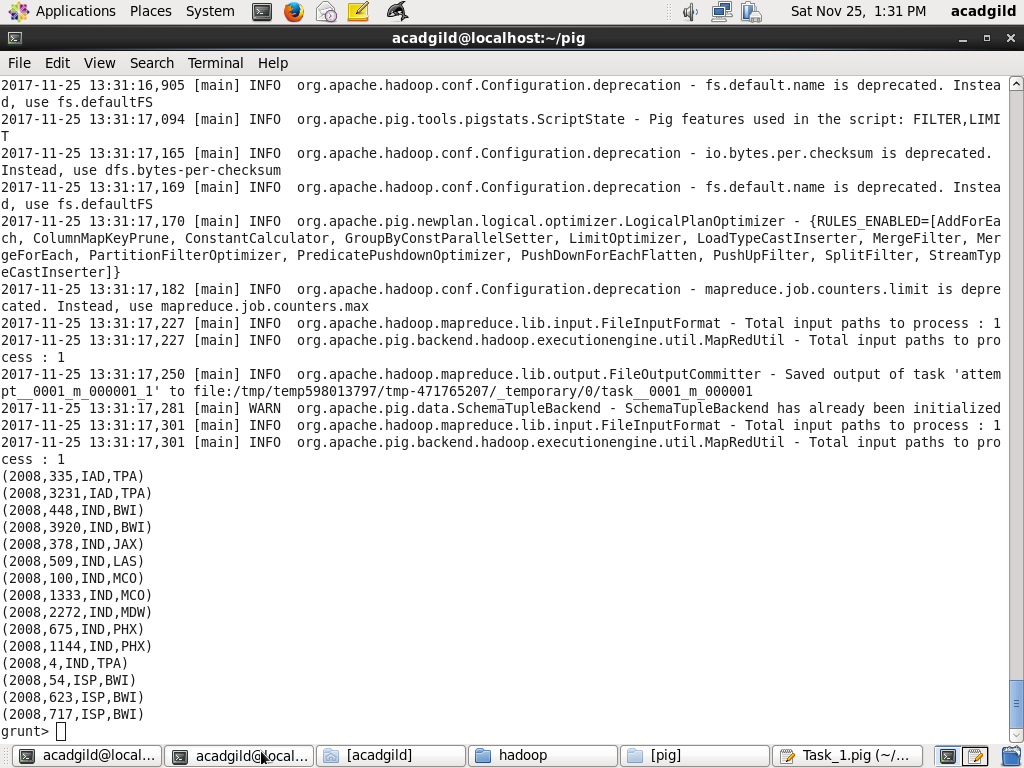
*B = foreach A generate (int)$1 as year, (int)$10 as flight\_num, (chararray)$17 as origin,(chararray) $18 as dest;*

*B\_LIMIT = LIMIT B 5;*

**

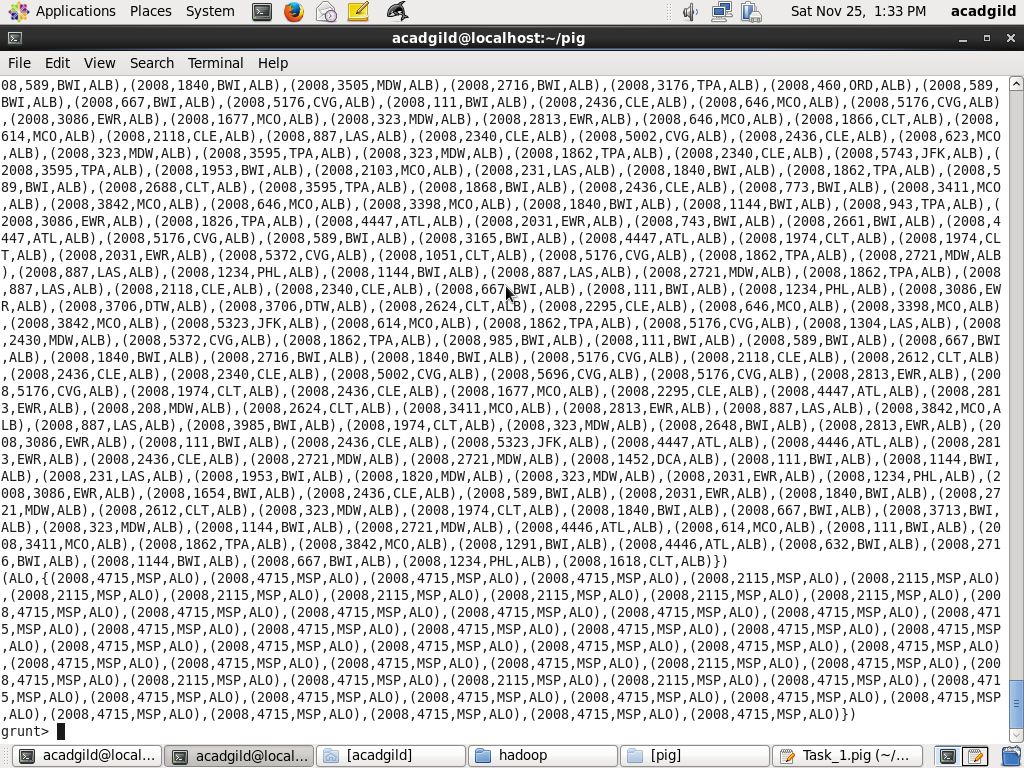
*C = filter B by dest is not null;*

*C\_LIMIT = LIMIT C 15;*

**

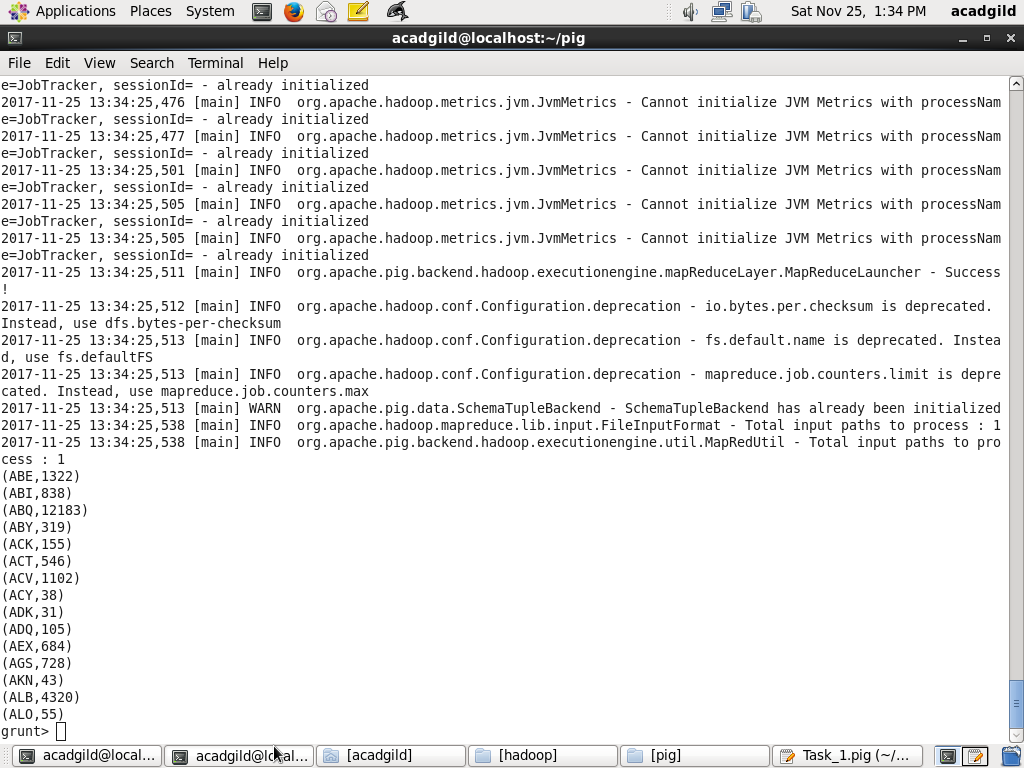
*D = group C by dest;*

*D\_LIMIT = LIMIT D 15;*

**

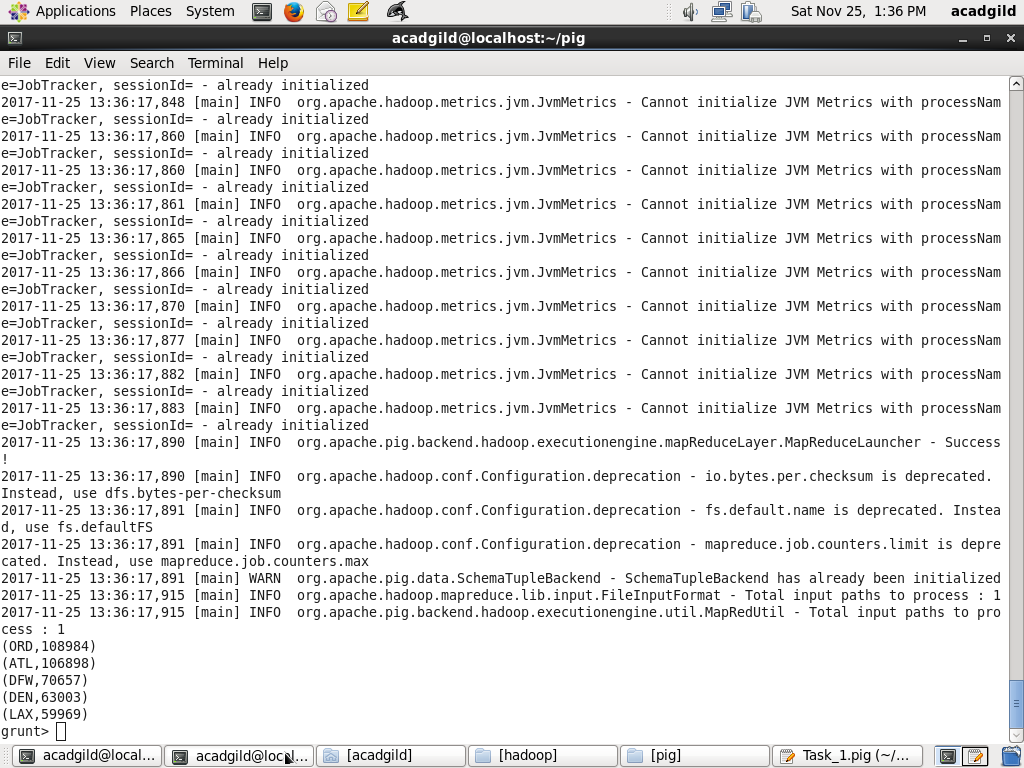
*E = foreach D generate group, COUNT(C.dest);*

*E\_LIMIT = LIMIT E 15;*

**

*F = order E by $1 DESC;*

*Result = LIMIT F 5;*

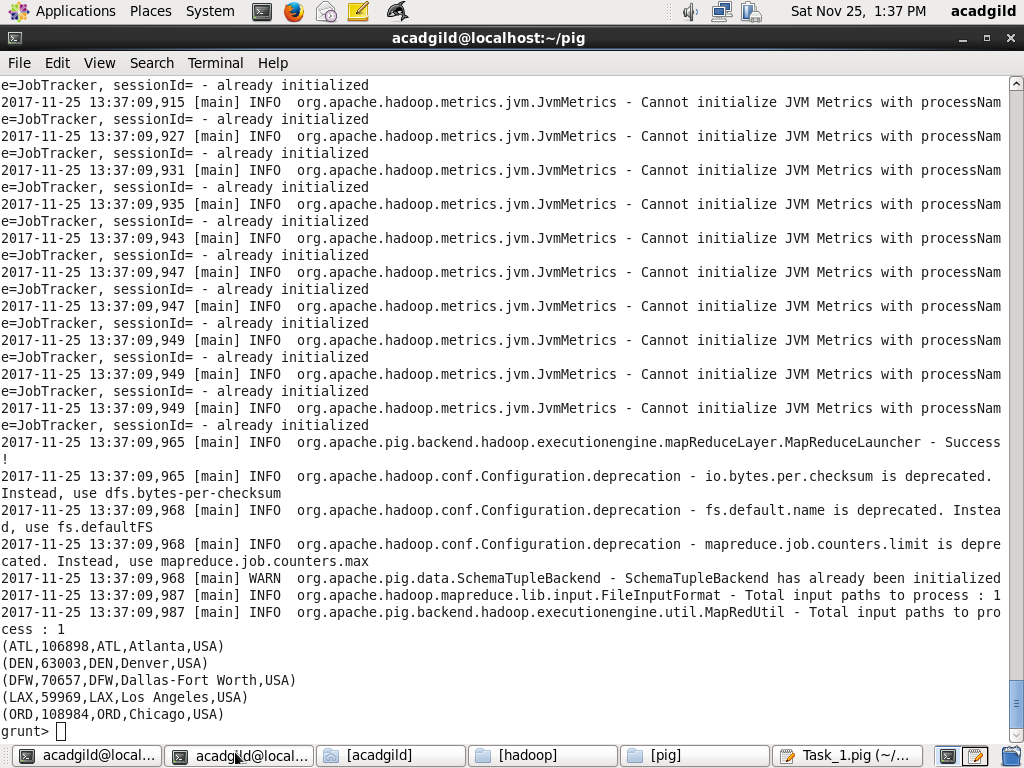
**

*A1 = load 'airports.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');*

*A2 = foreach A1 generate (chararray)$0 as dest, (chararray)$2 as city, (chararray)$4 as country;*

*joined\_table = join Result by $0, A2 by dest;*

*dump joined\_table;*

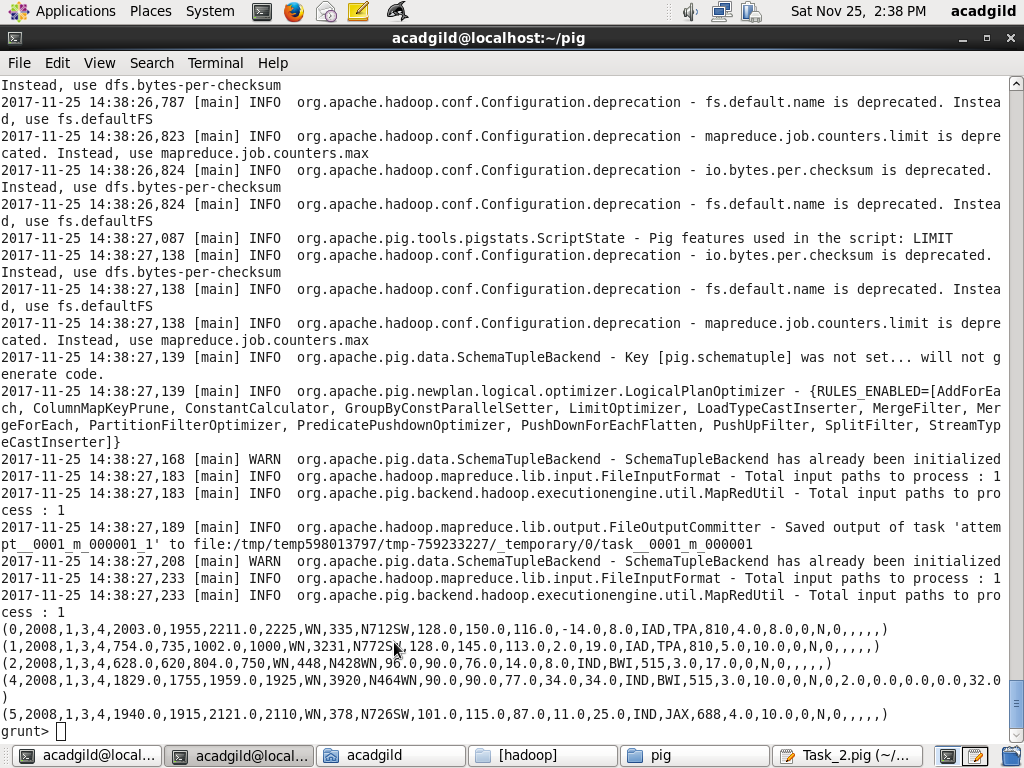
**

1. Which month has seen the most number of cancellations due to bad weather?

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

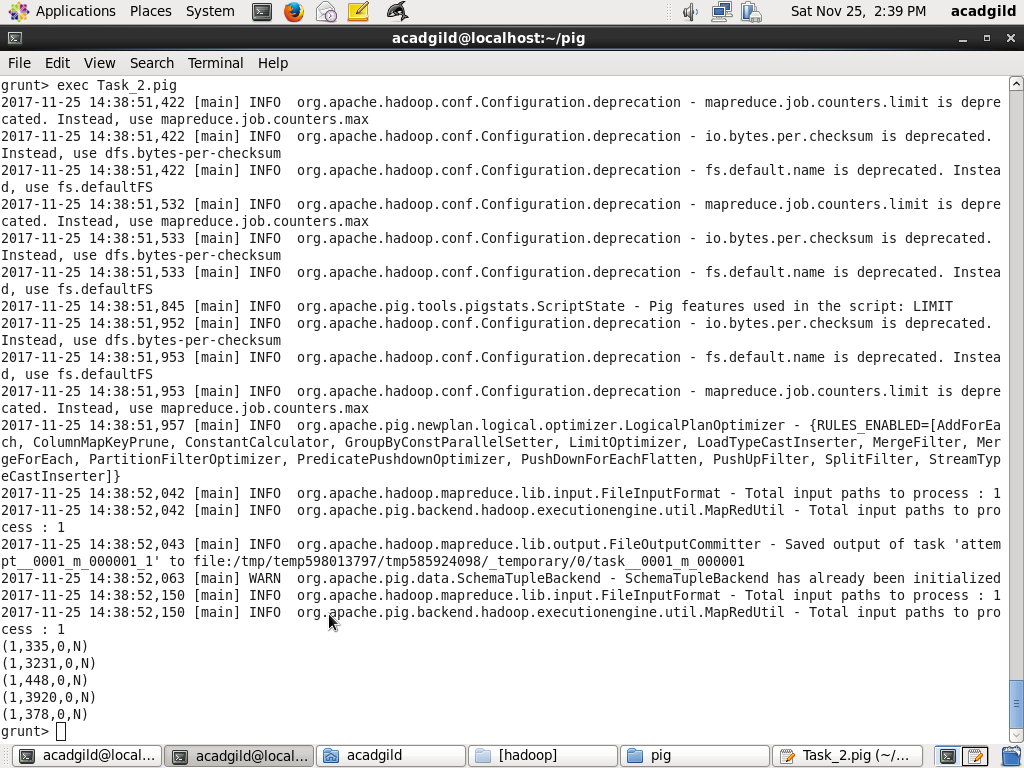
*A = load 'DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');*

*A\_LIMIT = LIMIT A 5;*

**

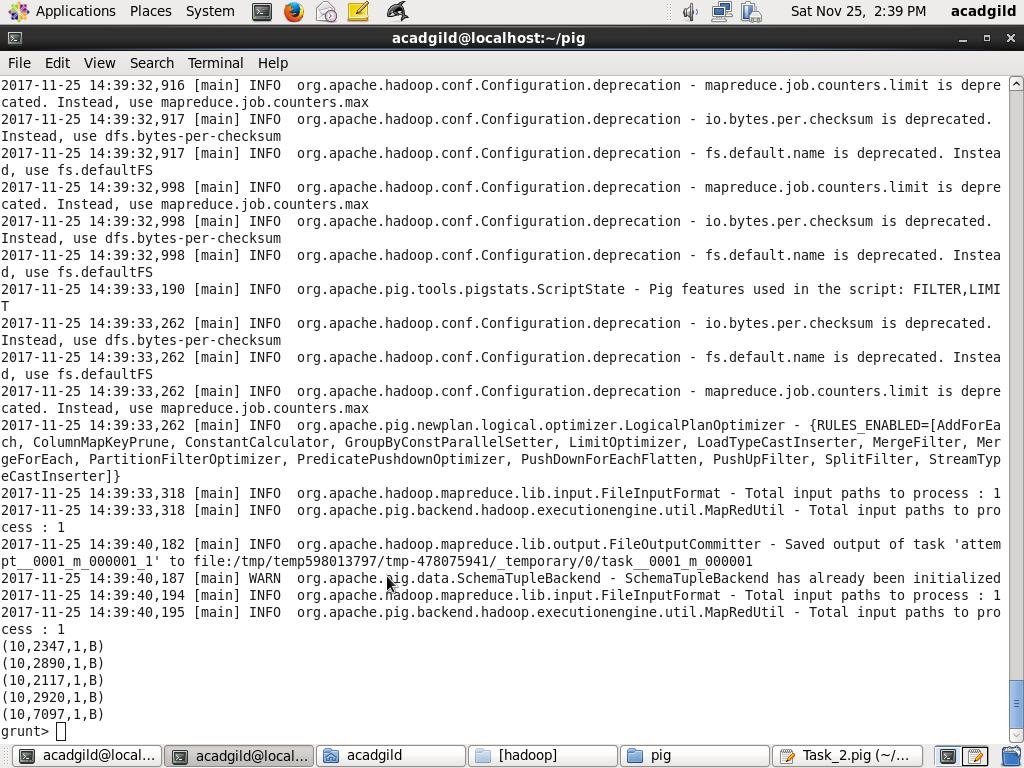
*B = foreach A generate (int)$2 as month,(int)$10 as flight\_num,(int)$22 as cancelled,(chararray)$23 as cancel\_code;*

*B\_LIMIT = LIMIT B 5;*

**

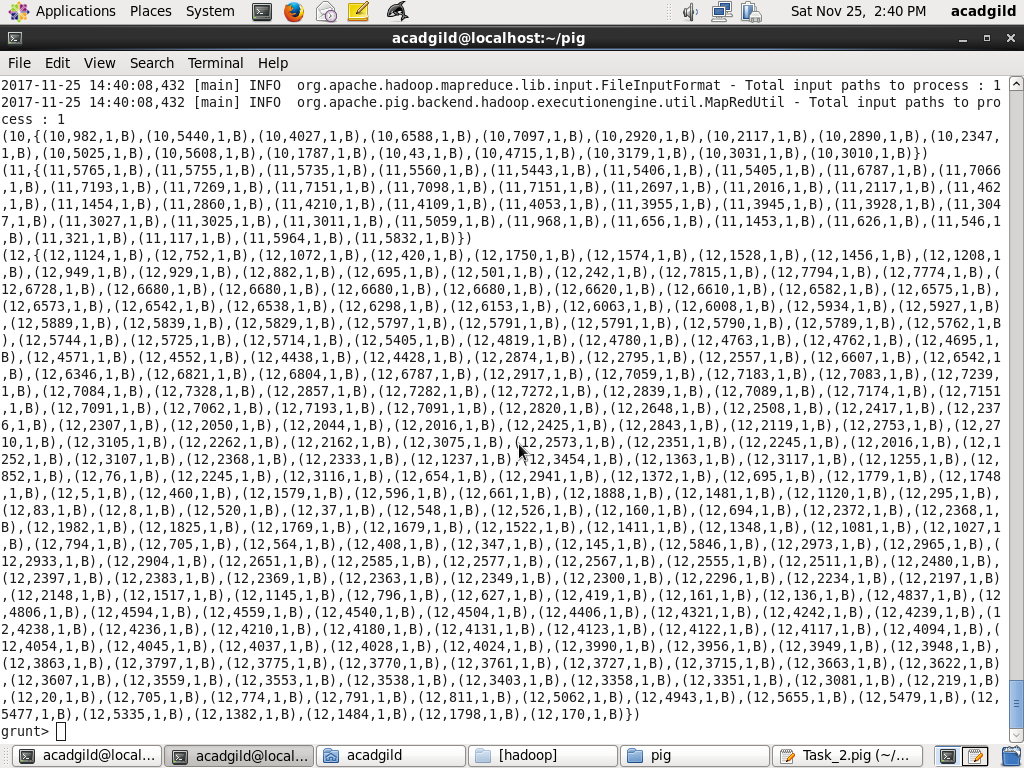
*C = filter B by cancelled == 1 AND cancel\_code =='B';*

*C\_LIMIT = LIMIT C 5;*

**

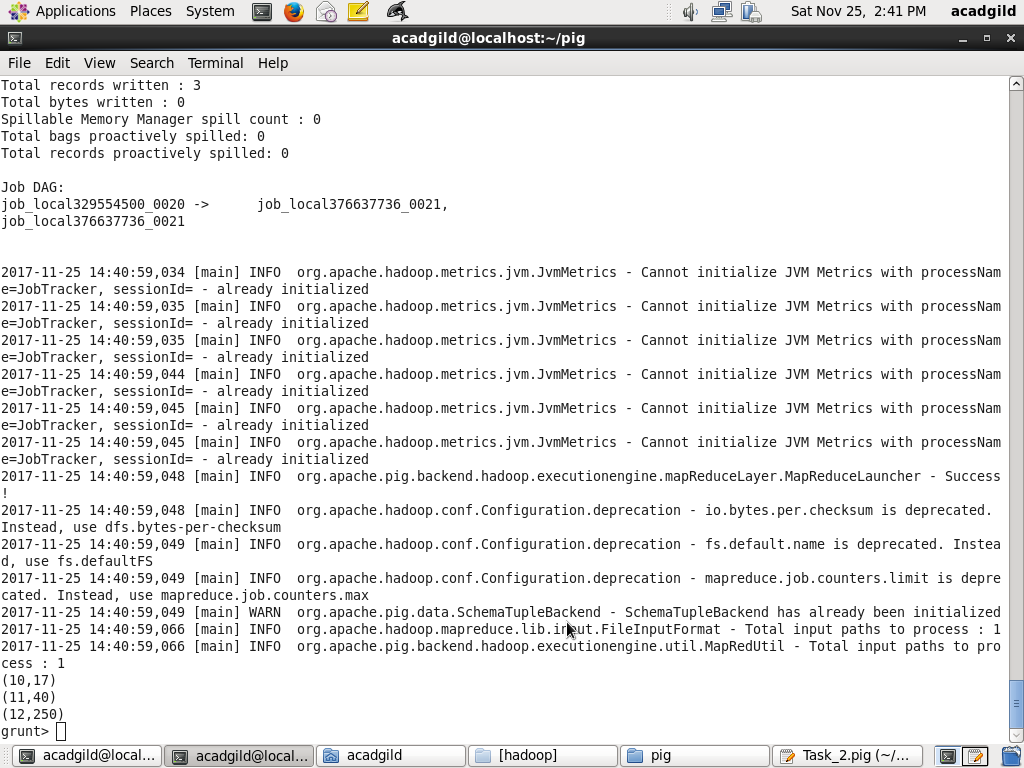
*D = group C by month;*

*D\_LIMIT = LIMIT D 5;*

**

*E = foreach D generate group, COUNT(C.cancelled);*

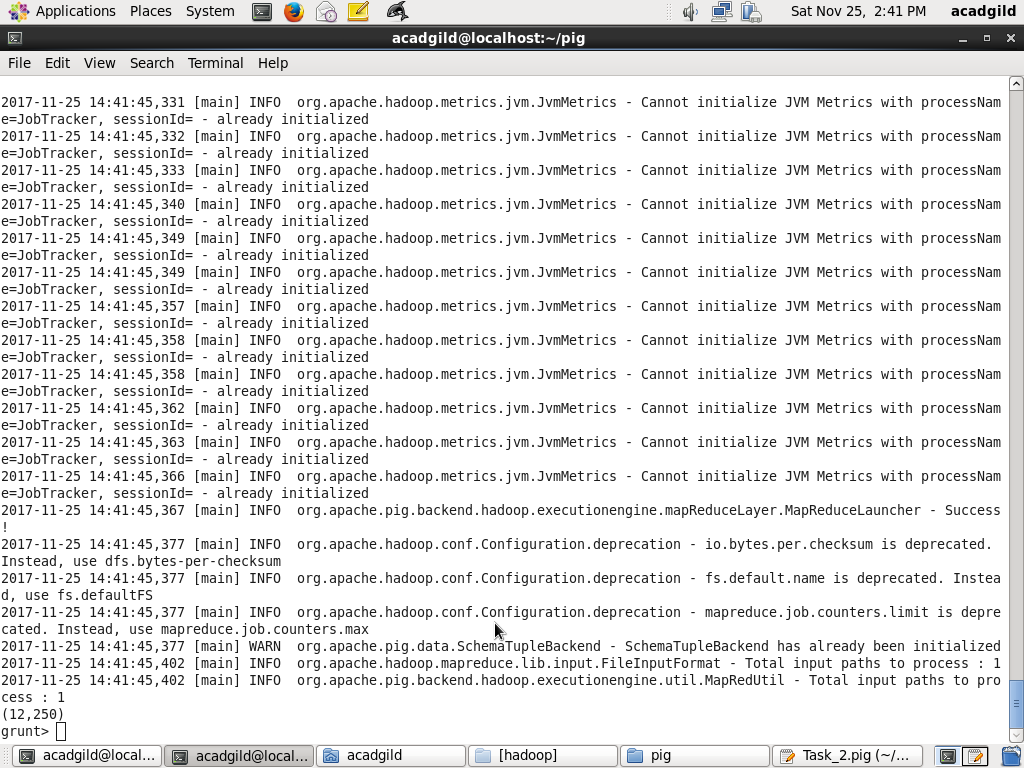
*E\_LIMIT = LIMIT E 5;*

**

*F= order E by $1 DESC;*

*Result = limit F 1;*

*dump Result;*

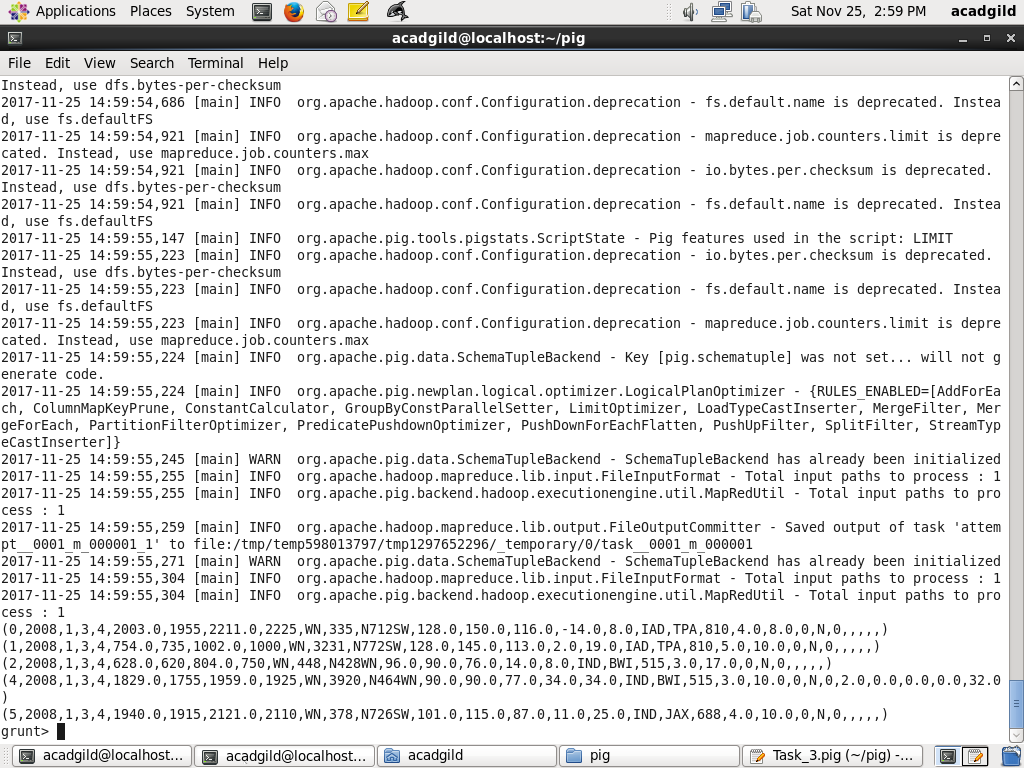
**

1. Top 10 origins with the highest average departure delay

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

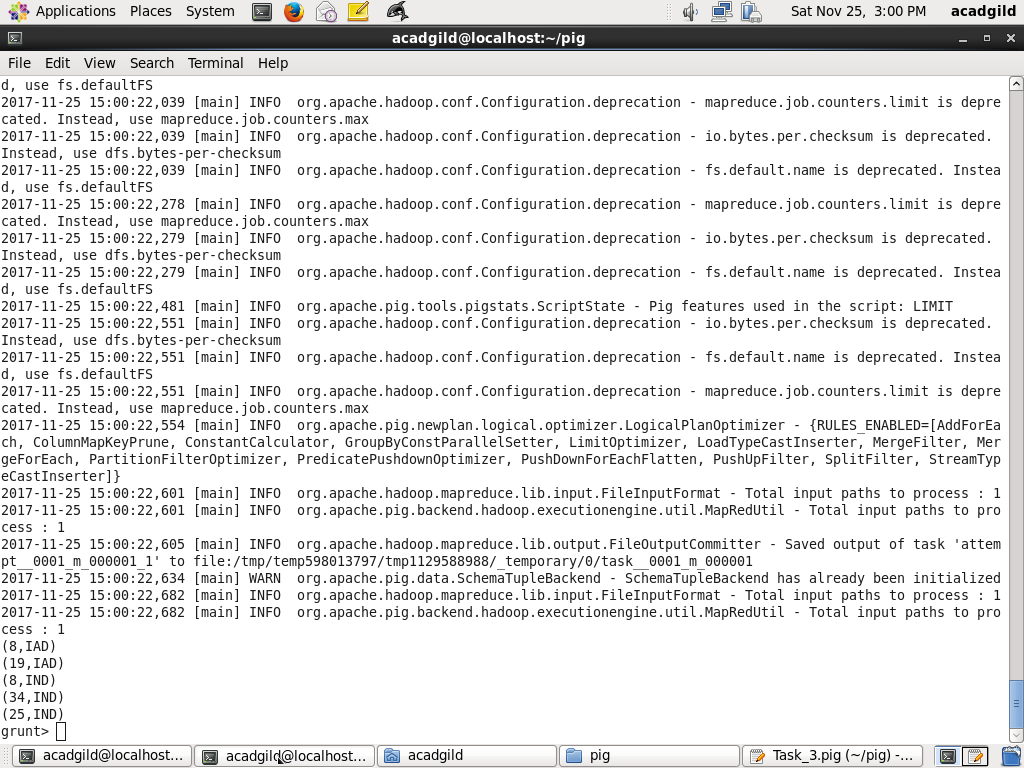
*A = load 'DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');*

*A\_LIMIT = LIMIT A 5;*

**

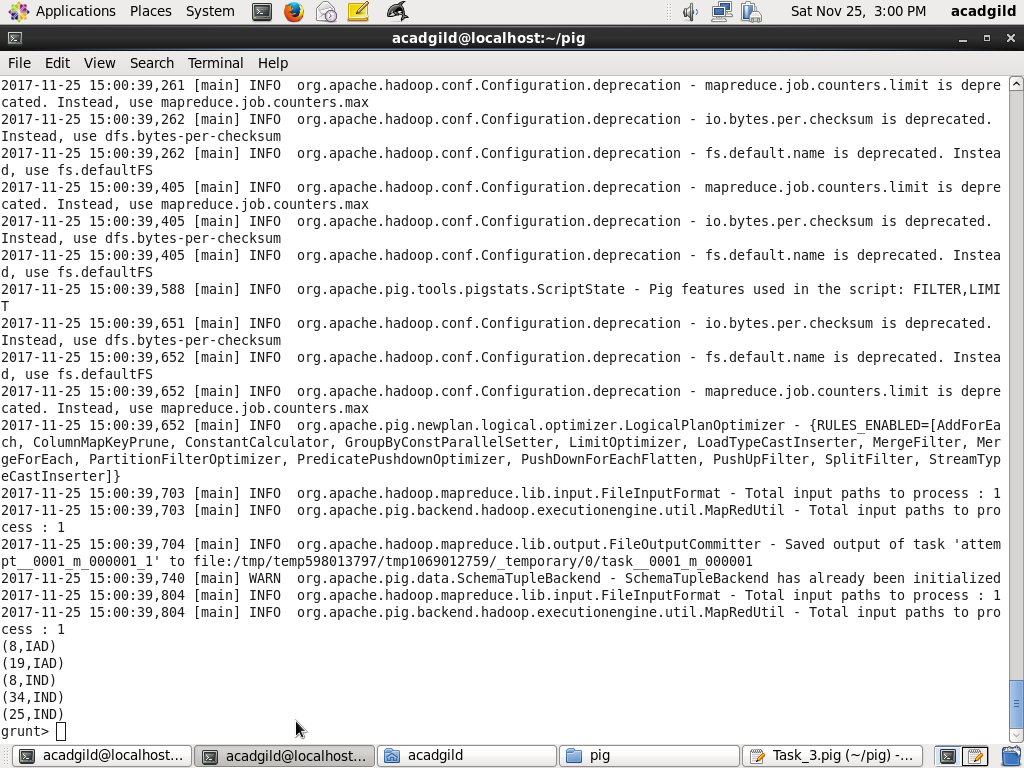
*B1 = foreach A generate (int)$16 as dep\_delay, (chararray)$17 as origin;*

*B1\_LIMIT = LIMIT B1 5;*

**

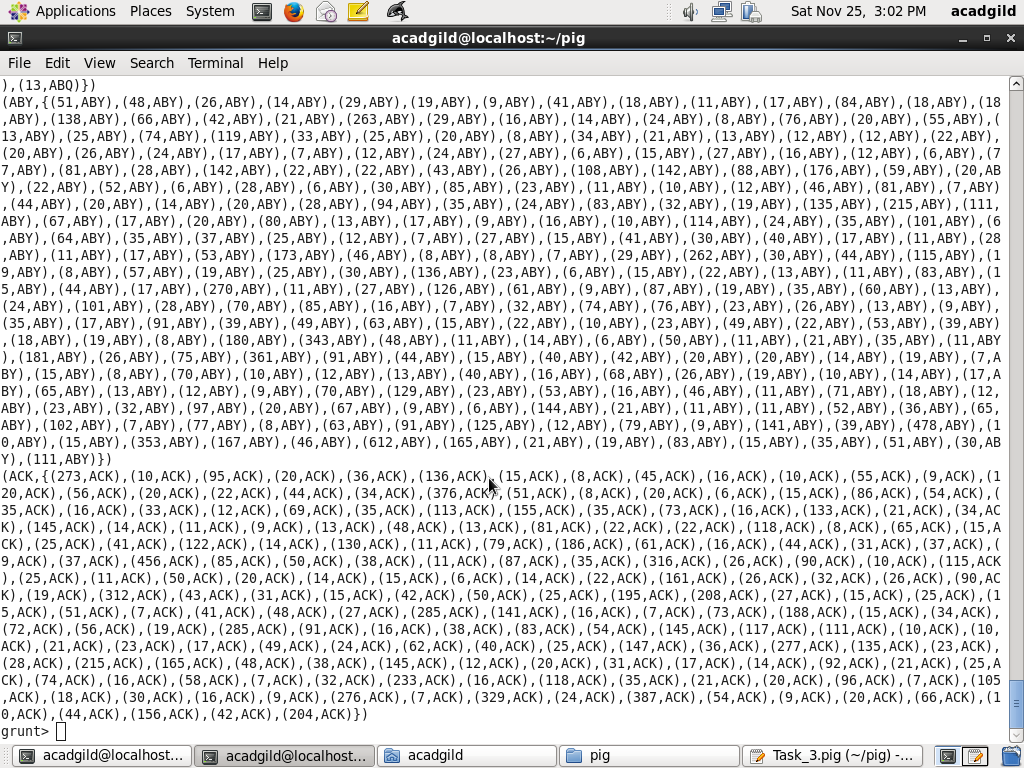
*C1 = filter B1 by (dep\_delay is not null) AND (origin is not null);*

*C1\_LIMIT = LIMIT C1 5;*

**

*D1 = group C1 by origin;*

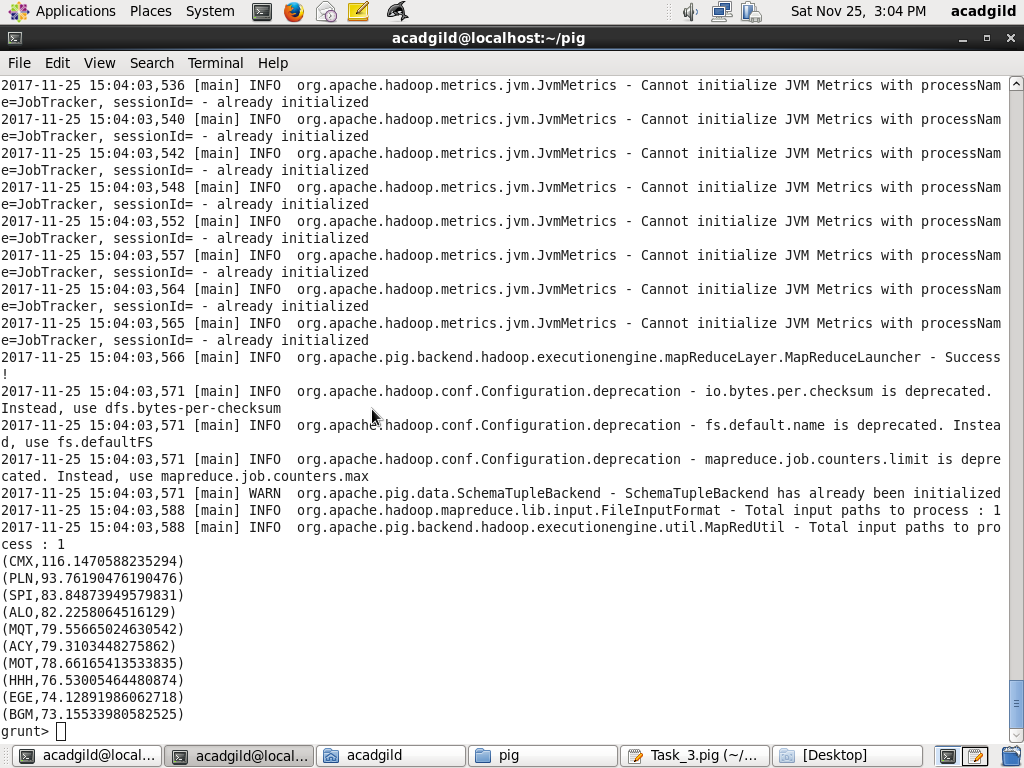
*D1\_LIMIT = LIMIT D1 5;*

**

*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');*

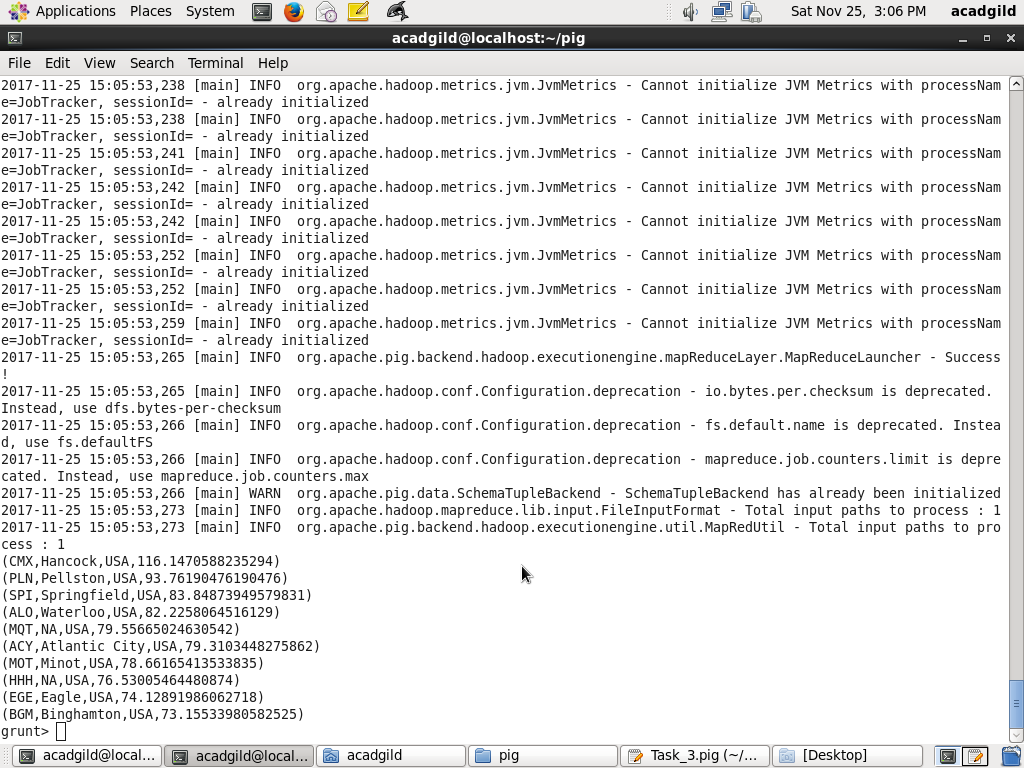
*Lookup1 = foreach Lookup generate (chararray)$0 as origin, (chararray)$2 as city, (chararray)$4 as country;*

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

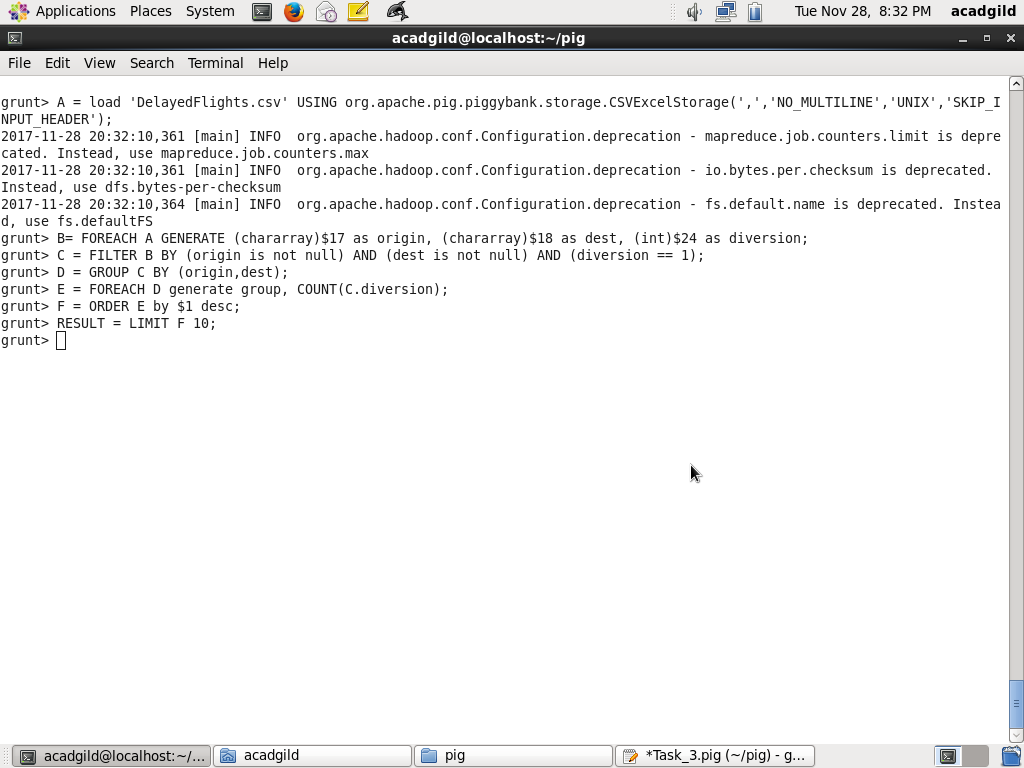
*Final = foreach Joined generate $0,$1,$2,$4;*

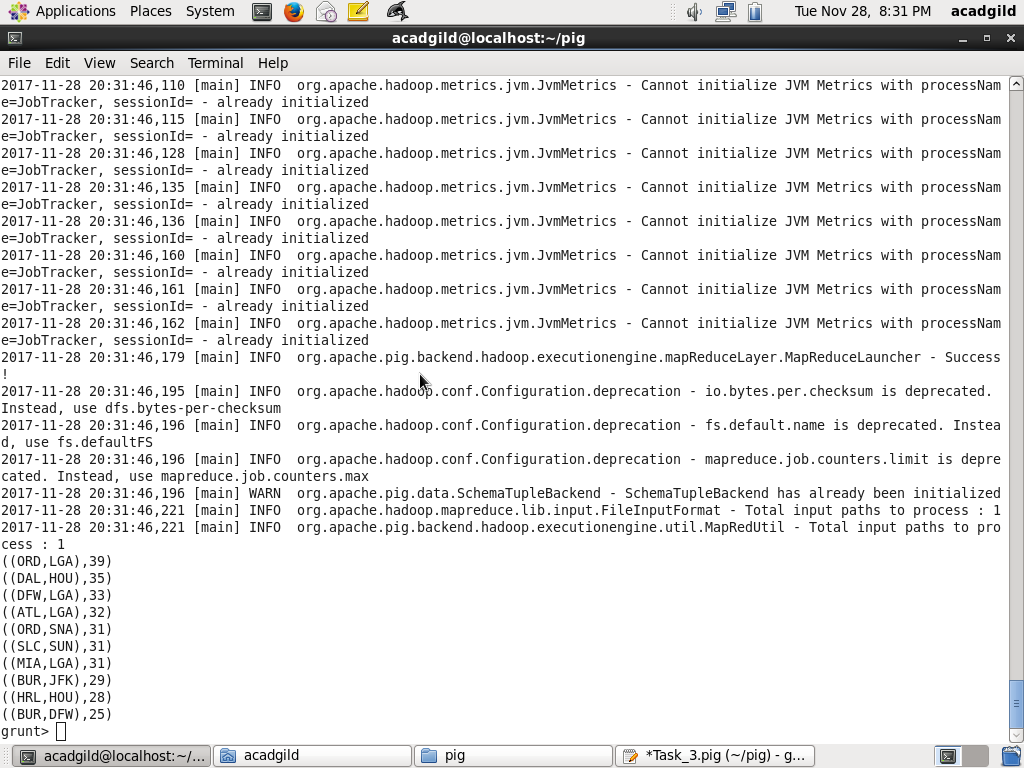
*Final\_Result = ORDER Final by $3 DESC;*

*dump Final\_Result;*

**

1. Which route (origin & destination) has seen the maximum diversion?

**

**