SESSION 5 ASSIGNMENT 2

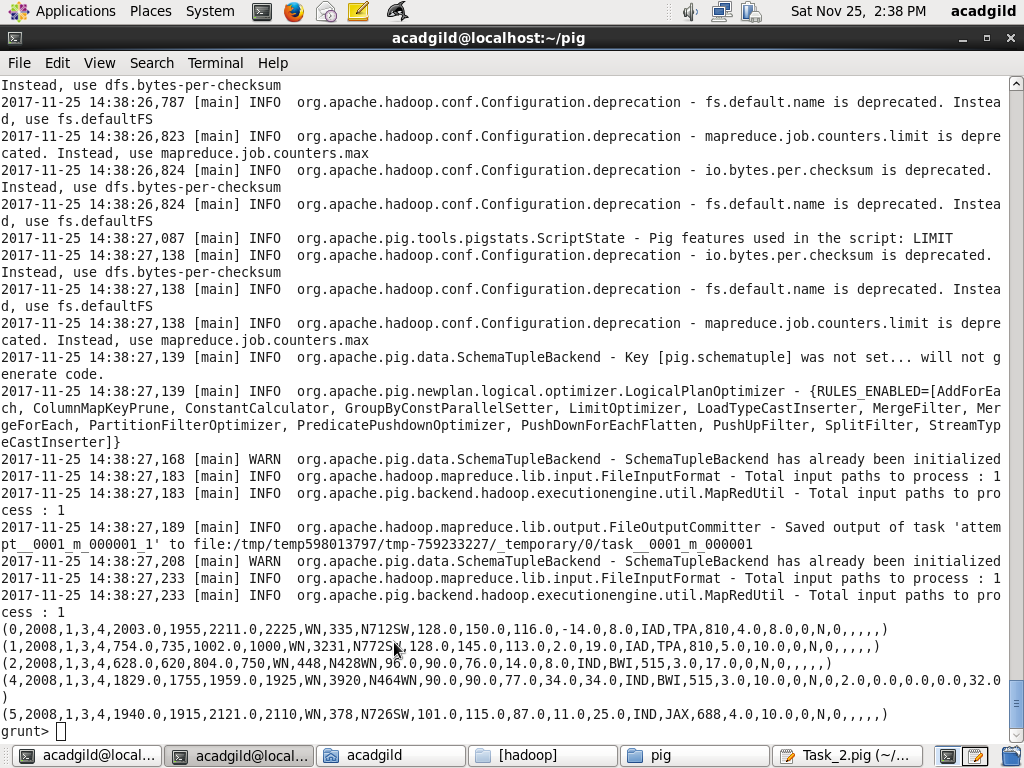
## PROBLEM 2

### Q2. Which month has seen the most number of cancellations due to bad weather?

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

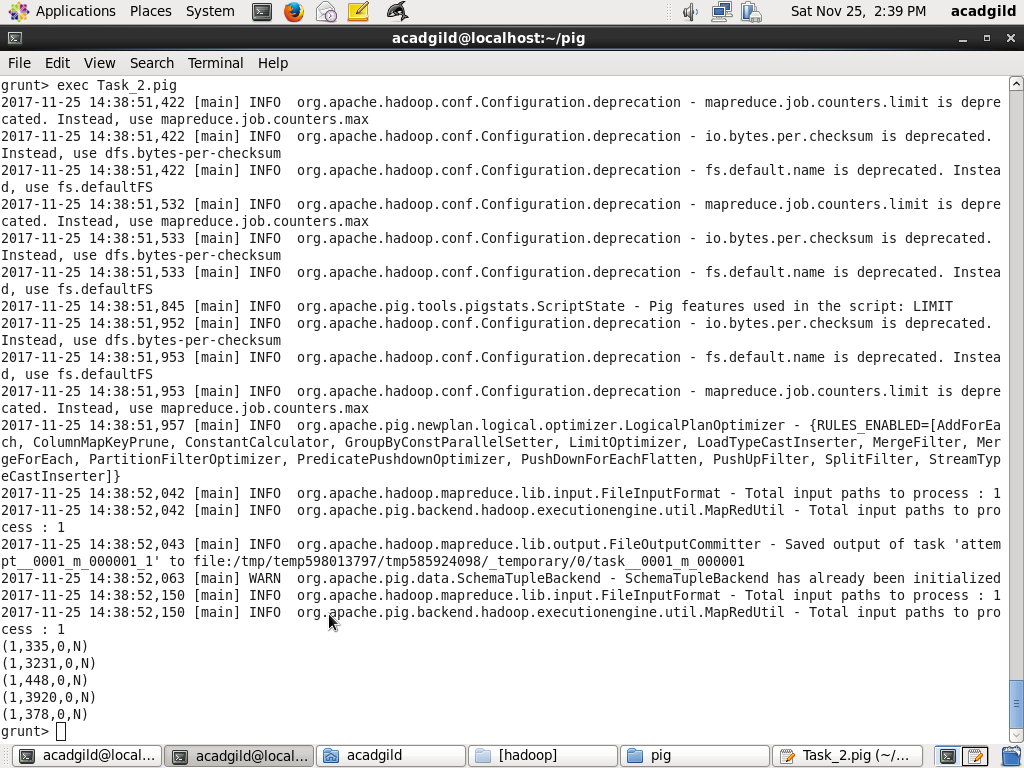
A = load 'DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

A\_LIMIT = LIMIT A 5;



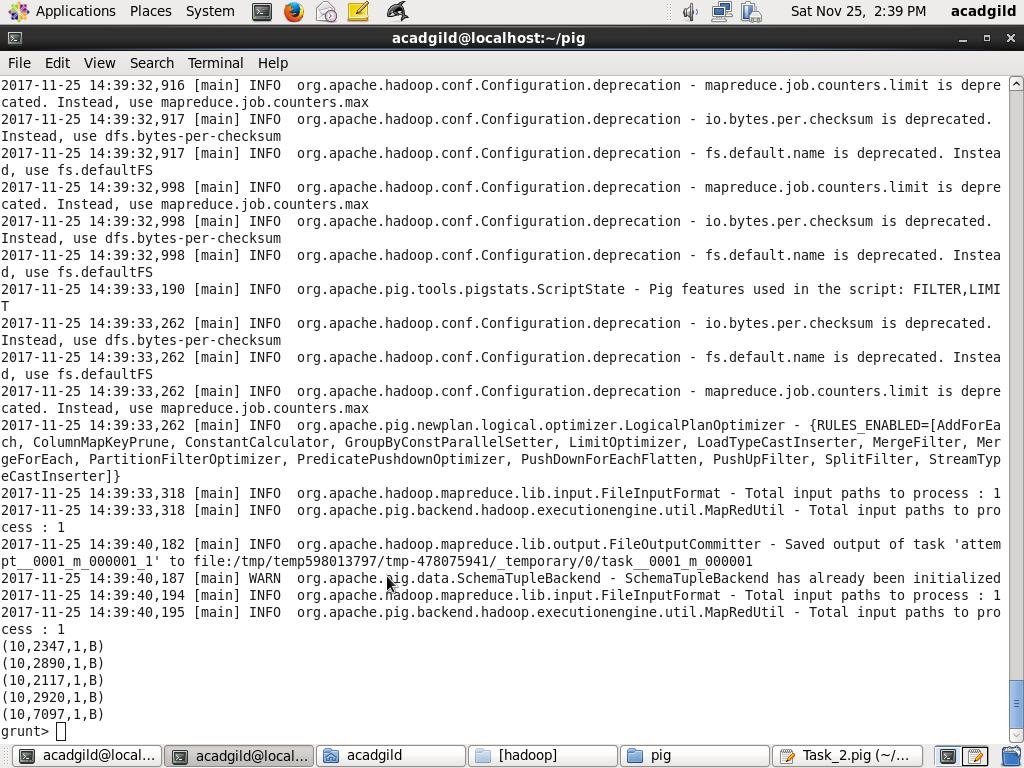
B = foreach A generate (int)$2 as month,(int)$10 as flight\_num,(int)$22 as cancelled,(chararray)$23 as cancel\_code;

B\_LIMIT = LIMIT B 5;



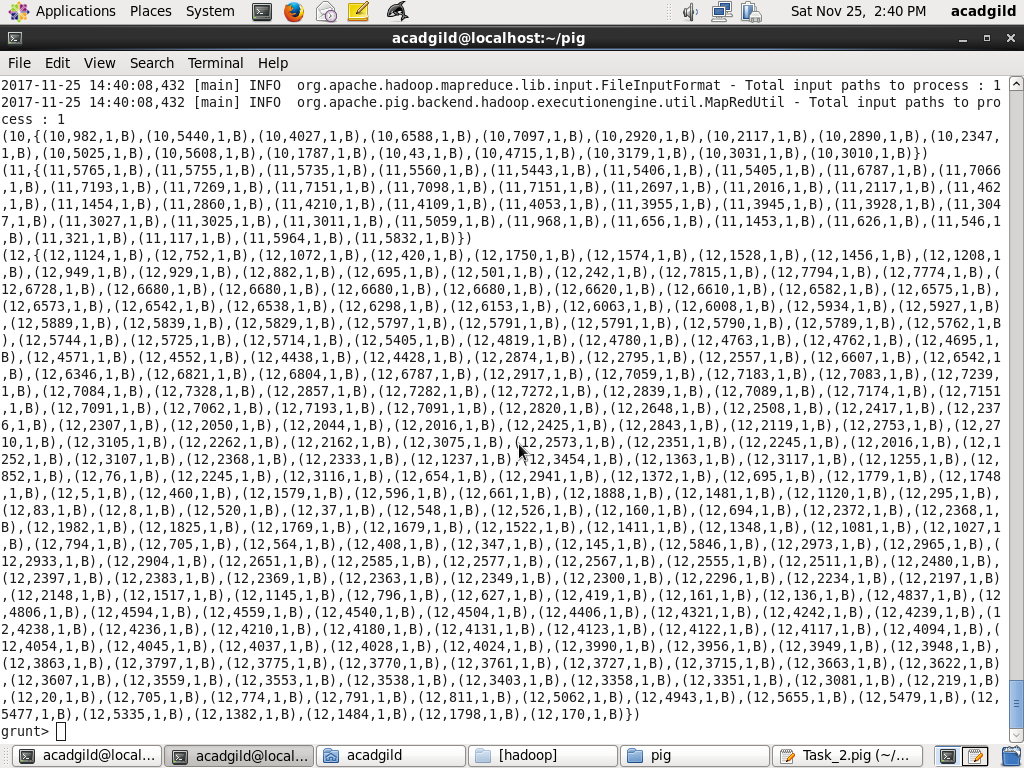
C = filter B by cancelled == 1 AND cancel\_code =='B';

C\_LIMIT = LIMIT C 5;



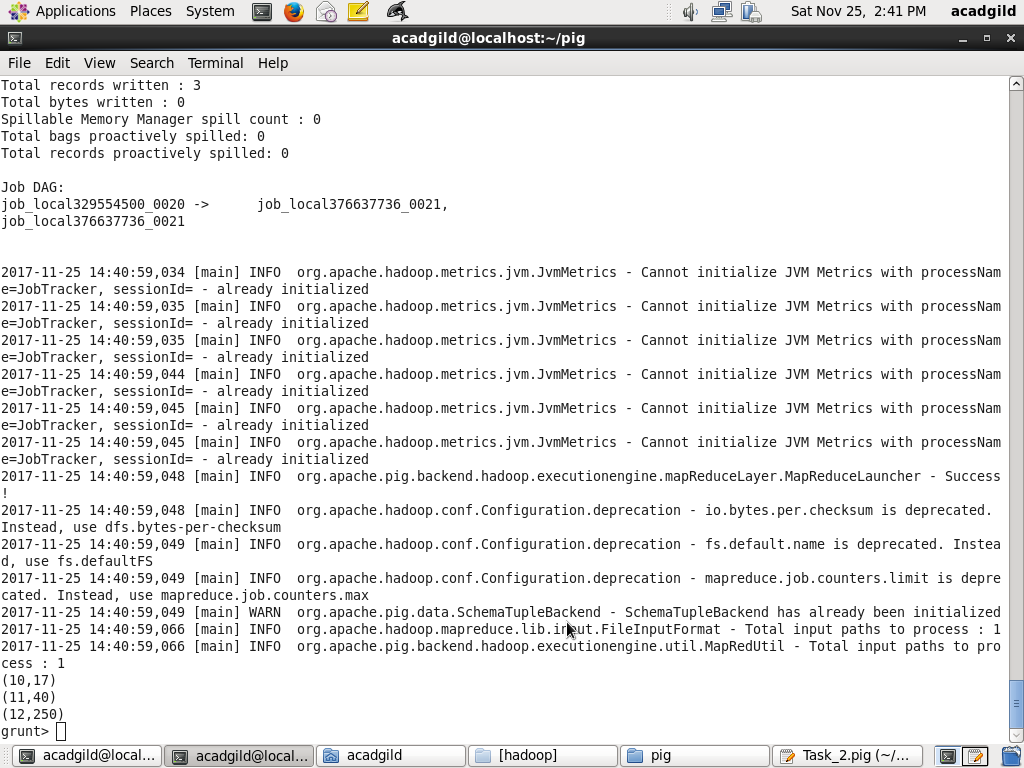
D = group C by month;

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;

