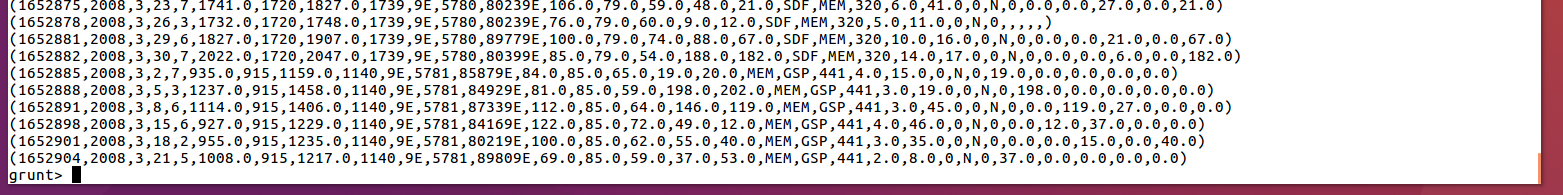
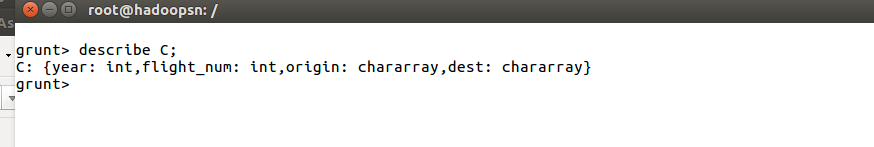
## Problem Statement 1

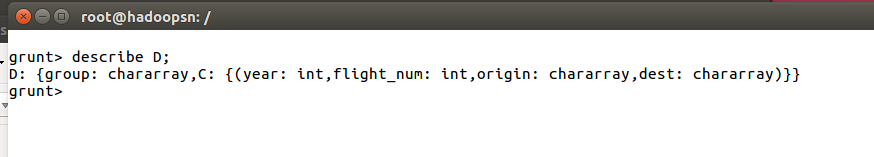
* REGISTER '/usr/local/pig/pig-0.17.0/lib/piggybank.jar';
* A = load '/hadooplocal/pig/DelayedFlights.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');



* dump A;
* B = foreach A generate (int)$1 as year, (int)$10 as flight\_num, (chararray)$17 as origin, (chararray)$18 as dest;
* C = filter B by dest is not null;



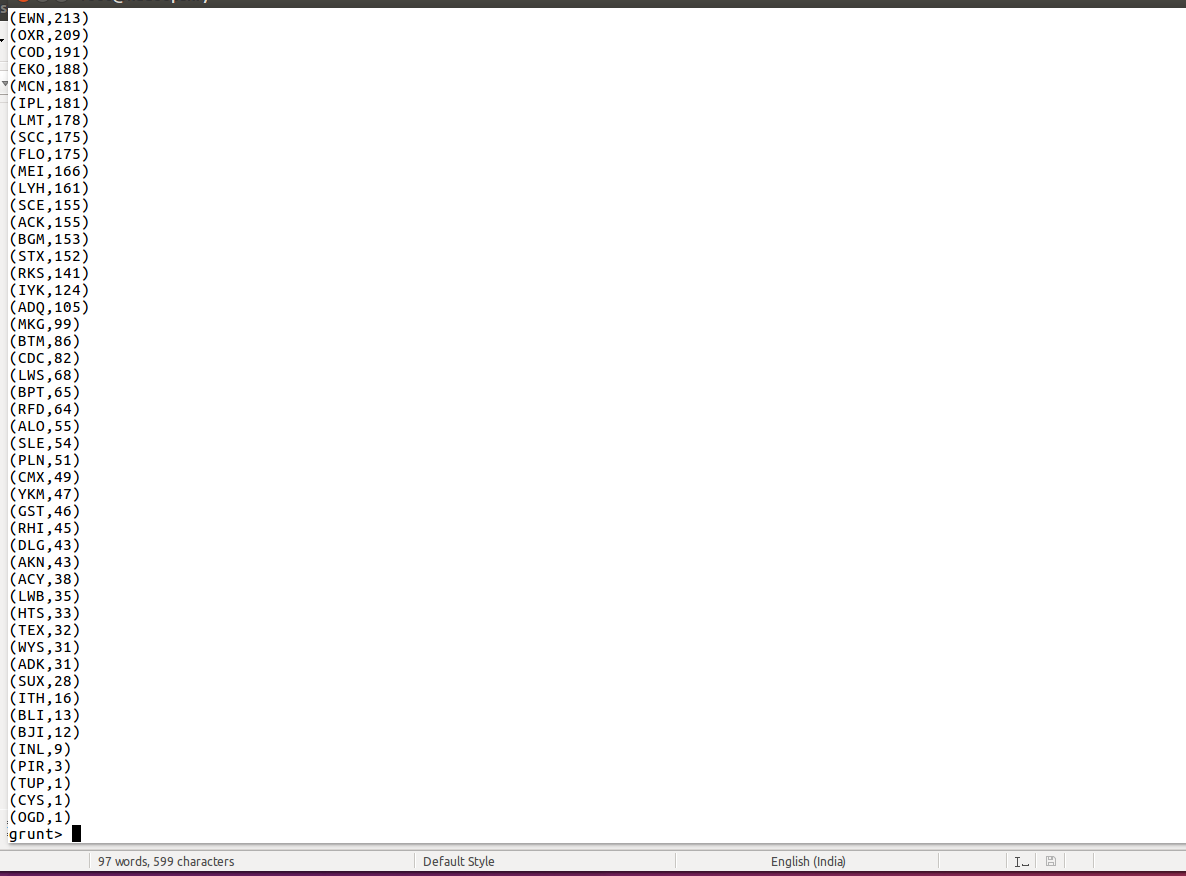
* D = group C by dest;



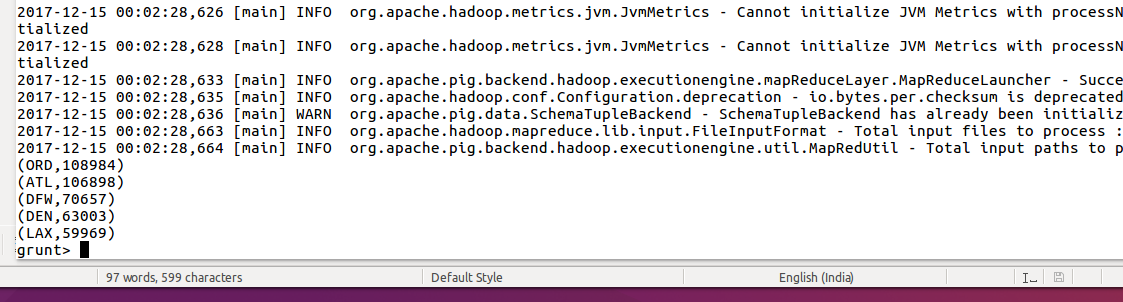
* E = foreach D generate group, COUNT(C.dest);



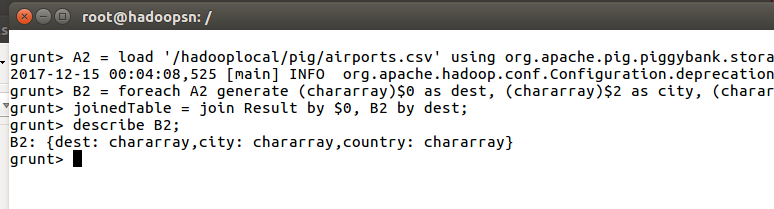
* F = order E by $1 desc;



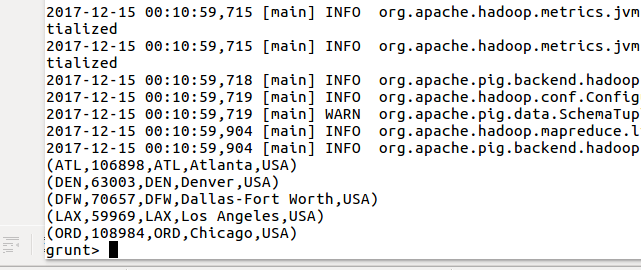
* Result = limit F 5;



* A2 = load '/hadooplocal/pig/airports.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');
* B2 = foreach A2 generate (chararray)$0 as dest, (chararray)$2 as city, (chararray)$4 as country;

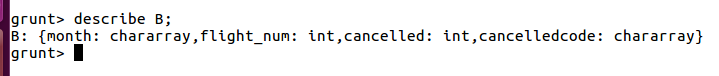


* joinedTable = join Result by $0, B2 by dest;

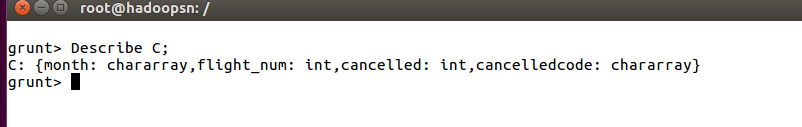


## Problem Statement 2

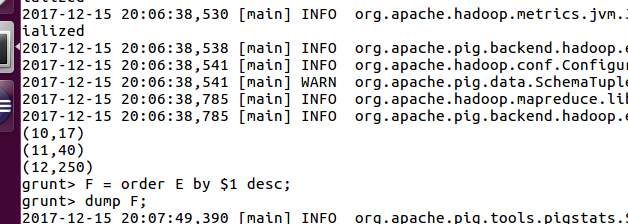
* REGISTER '/usr/local/pig/pig-0.17.0/lib/piggybank.jar';
* A = load '/hadooplocal/pig/DelayedFlights.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');
* B = foreach A generate (chararray)$2 as month, (int)10 as flight\_num, (int)$22 as cancelled, (chararray)$23 as cancelledcode;



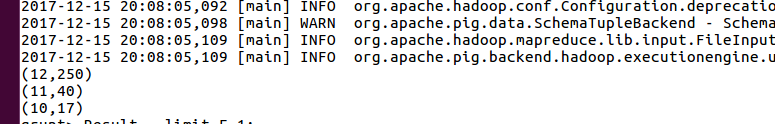
* C = filter B by cancelled == 1 and cancelledcode == 'B';



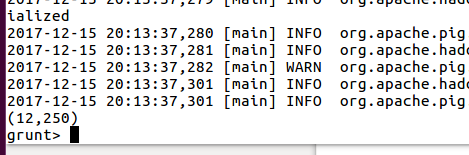
* D = group C by month;
* E = foreach D generate group, COUNT(C.cancelled);



* F = order E by $1 desc;

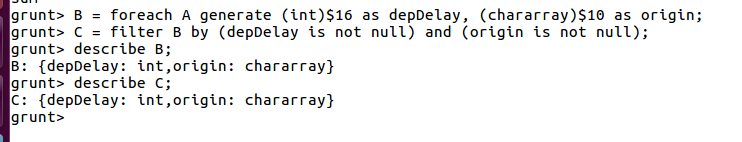


* Result = limit F 1;

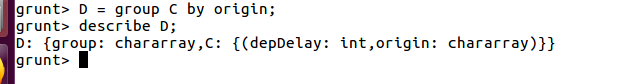


## Problem Statement 3

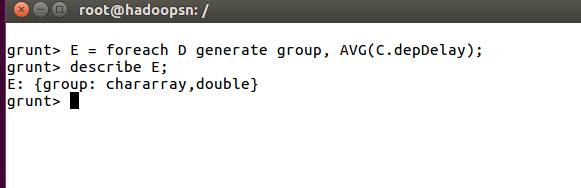
* REGISTER '/usr/local/pig/pig-0.17.0/lib/piggybank.jar';
* A = load '/hadooplocal/pig/DelayedFlights.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');
* B = foreach A generate (int)$16 as depDelay, (chararray)$17 as origin;
* C = filter B by (depDelay is not null) and (origin is not null);



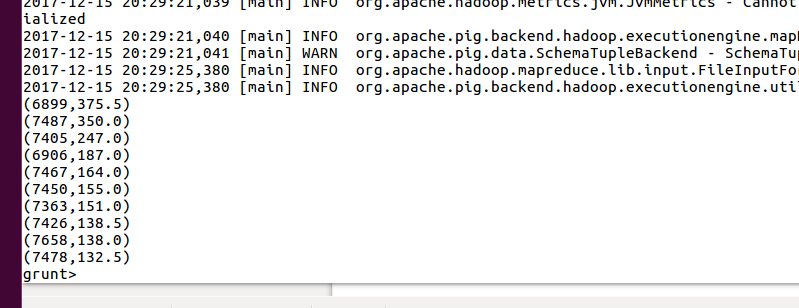
* D = group C by origin;



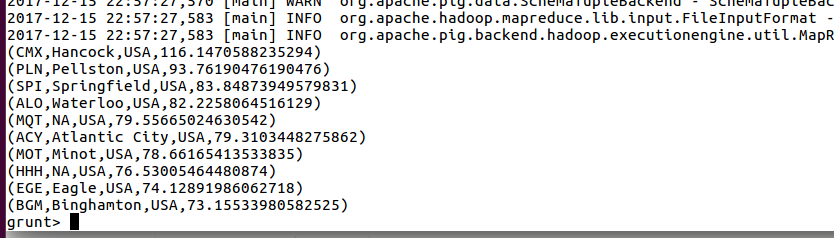
* E = foreach D generate group, AVG(C.depDelay);



* F = order E by $1 desc;
* ResultTopTen = limit F 10;

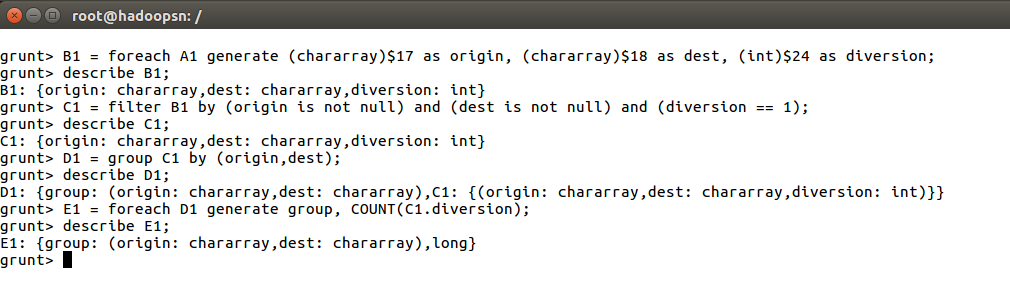


* A2 = load '/hadooplocal/pig/airports.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');
* B2 = foreach A2 generate (chararray)$0 as origin, (chararray)$2 as city, (chararray)$4 as country;
* joined = join B2 by origin, ResultTopTen by $0;
* Final = foreach joined generate $0,$1,$2,$4;
* FinalResult = order Final by $3 desc;

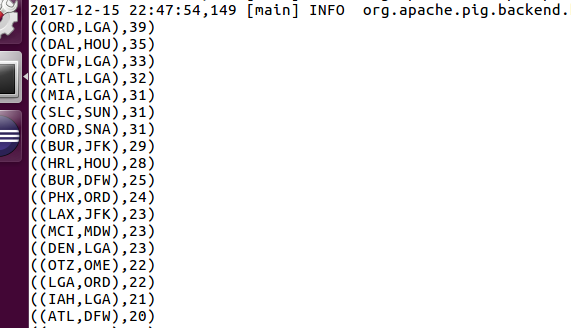


## Problem Statement 4

* REGISTER '/usr/local/pig/pig-0.17.0/lib/piggybank.jar';
* A1 = load '/hadooplocal/pig/DelayedFlights.csv' using org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');
* B1 = foreach A1 generate (chararray)$17 as origin, (chararray)$18 as dest, (int)$24 as diversion;
* C1 = filter B1 by (origin is not null) and (dest is not null) and (diversion == 1);
* D1 = group C1 by (origin,dest);
* E1 = foreach D1 generate group, COUNT(C1.diversion);



* f1 = ORDER E1 by $1 desc;



* Result = limit f1 1;

