**Problem Statement 1: Find out the top 5 most visited destinations.**

======================================================================

REGISTER '/home/cloudera/chhaya/aviation\_usecase/piggybank-0.15.0.jar';

A = LOAD '/home/cloudera/chhaya/aviation\_usecase/DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

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;

A1 = LOAD '/home/cloudera/chhaya/aviation\_usecase/airports.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

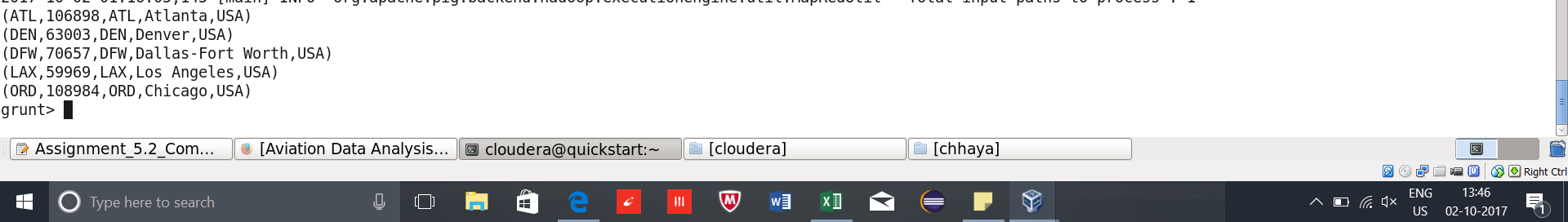
A2 = FOREACH A1 GENERATE (chararray)$0 AS dest, (chararray)$2 AS city, (chararray)$4 AS country;

joined\_table = JOIN Result BY $0, A2 BY dest;

DUMP joined\_table;

======================================================================

**Output Problem Statement 1:**



======================================================================

**Problem Statement 2: Which month has seen the most number of cancellations due to bad weather?**

REGISTER '/home/cloudera/chhaya/aviation\_usecase/piggybank-0.15.0.jar';

A = LOAD '/home/cloudera/chhaya/aviation\_usecase/DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

B = FOREACH A GENERATE (int)$2 AS month,(int)$10 AS flight\_num,(int)$22 AS cancelled,(chararray)$23 AS cancel\_code;

C = FILTER B BY cancelled == 1 AND cancel\_code =='B';

D = GROUP C BY month;

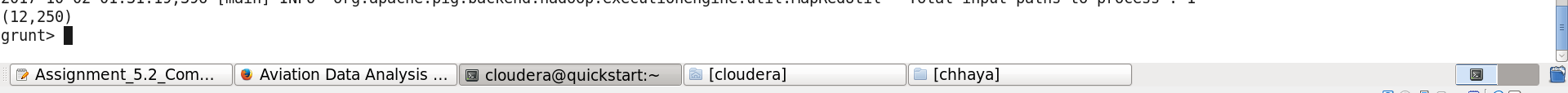
E = FOREACH D GENERATE group, COUNT(C.cancelled);

F = ORDER E BY $1 DESC;

Result = LIMIT F 1;

======================================================================

**Output Problem Statement 2:**



======================================================================

**Problem Statement 3: Top ten origins with the highest AVG departure delay**

REGISTER '/home/cloudera/chhaya/aviation\_usecase/piggybank-0.15.0.jar';

A1 = LOAD '/home/cloudera/chhaya/aviation\_usecase/DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX,'SKIP\_INPUT\_HEADER');

B1 = FOREACH A1 GENERATE (int)$16 as dep\_delay, (chararray)$17 as origin;

C1 = FILTER B1 BY (dep\_delay is not null) AND (origin is not null);

D1 = GROUP C1 BY origin;

E1 = FOREACH D1 GENERATE group, AVG(C1.dep\_delay);

Result = order E1 BY $1 DESC;

DUMP Result;

Top\_ten = LIMIT Result 10;

Lookup = LOAD '/home/cloudera/chhaya/aviation\_usecase/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;

======================================================================

**Output Problem Statement 3 Final\_Result**



======================================================================

**Problem Statement 4: Which route (origin & destination) has seen the maximum diversion?**

REGISTER '/home/cloudera/chhaya/aviation\_usecase/piggybank-0.15.0.jar';

A = LOAD '/home/cloudera/chhaya/aviation\_usecase/DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

B = FOREACH A GENERATE (chararray)$17 as origin, (chararray)$18 as dest, (int)$24 as diversion;

C = FILTER B BY (origin is not null) AND (dest is not null) AND (diversion == 1);

D = GROUP C BY (origin,dest);

E = FOREACH D GENERATE group, COUNT(C.diversion);

F = ORDER E BY $1 DESC;

Result = LIMIT F 10;

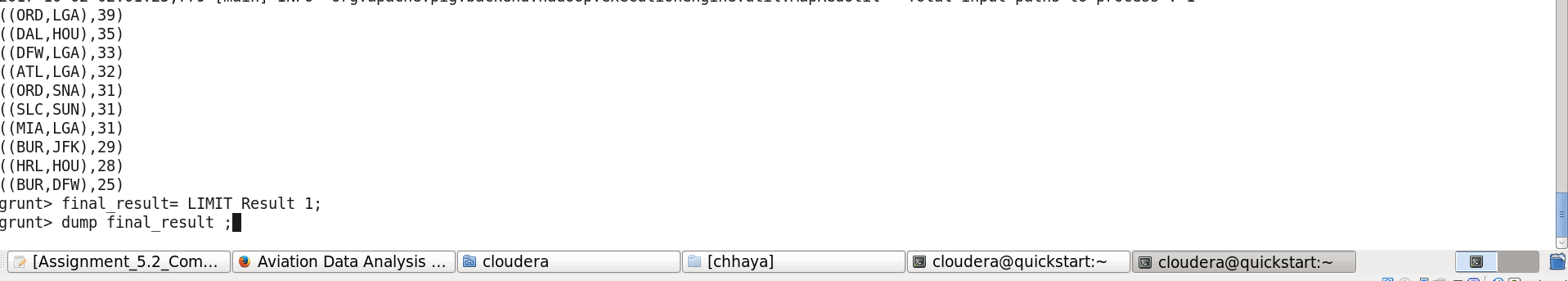
DUMP Result;

final\_result = LIMIT Result 1;

DUMP final\_result ;

======================================================================

**Output Problem Statement 4 Result:**



**Output Problem Statement 4 final\_result:**

