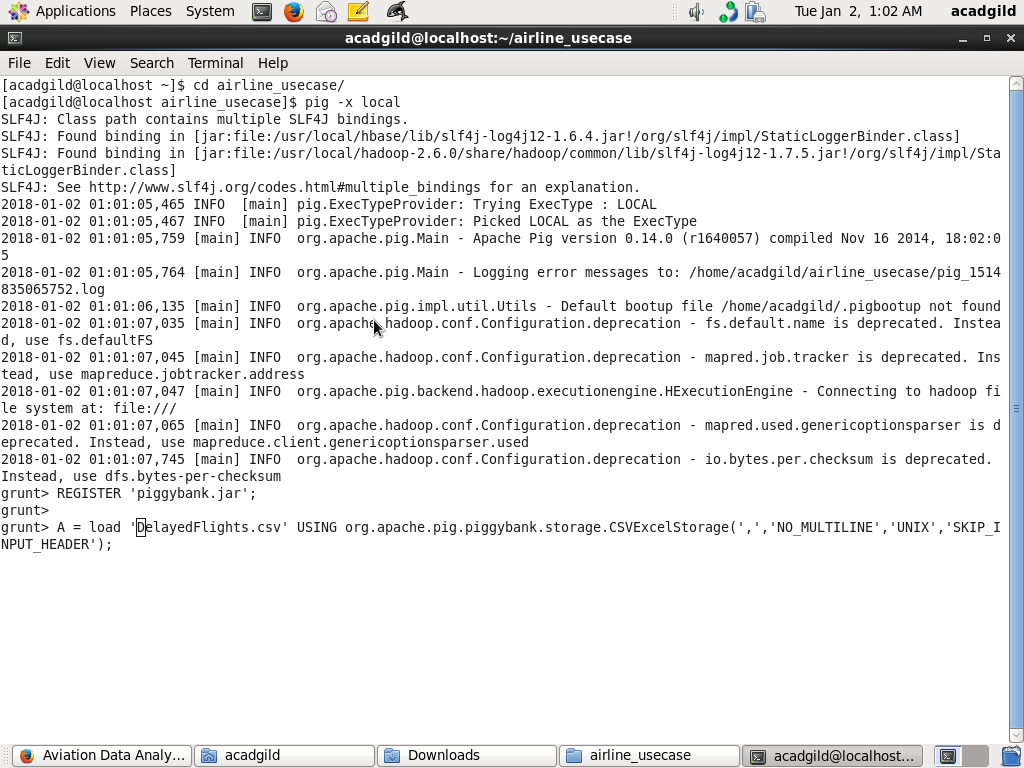
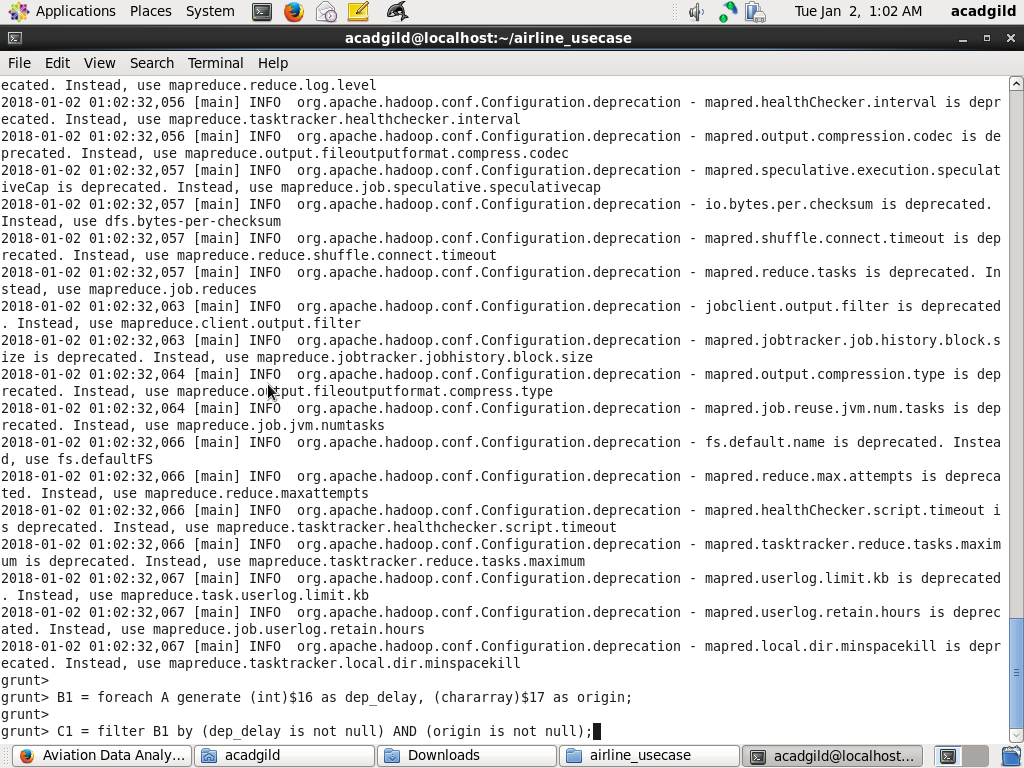
**Problem Statement 3**

**Top ten origins with the highest AVG departure delay**

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

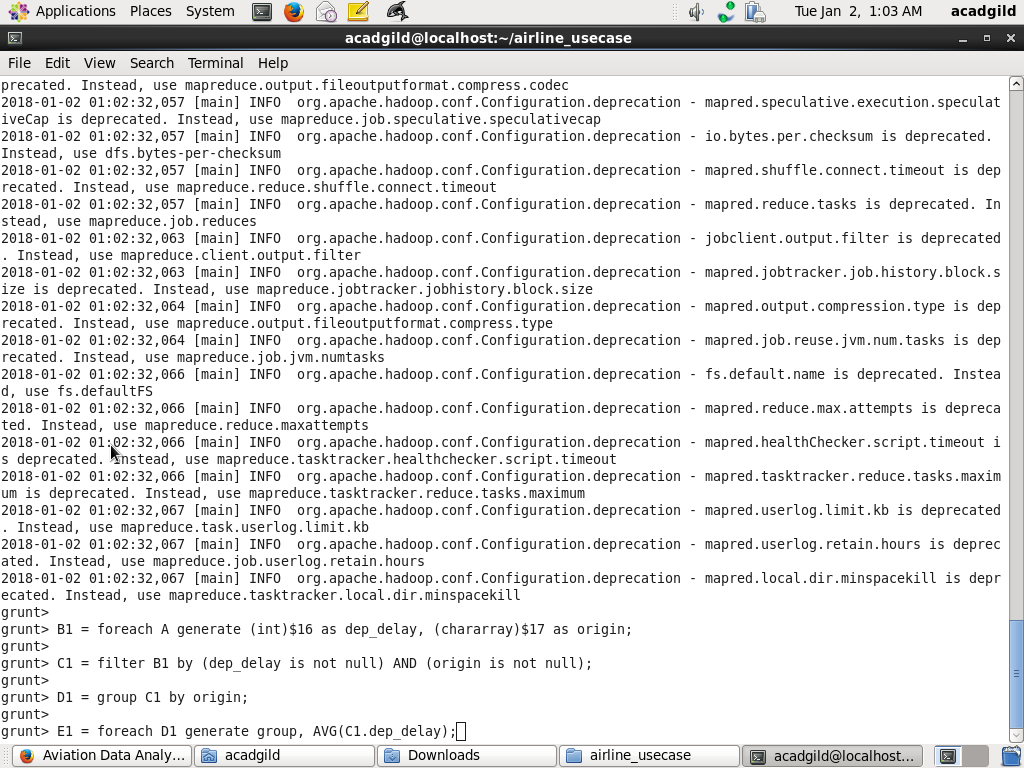


1. **B1 = foreach A generate (int)$16 as dep\_delay, (chararray)$17 as origin;**
2. **C1 = filter B1 by (dep\_delay is not null) AND (origin is not null);**



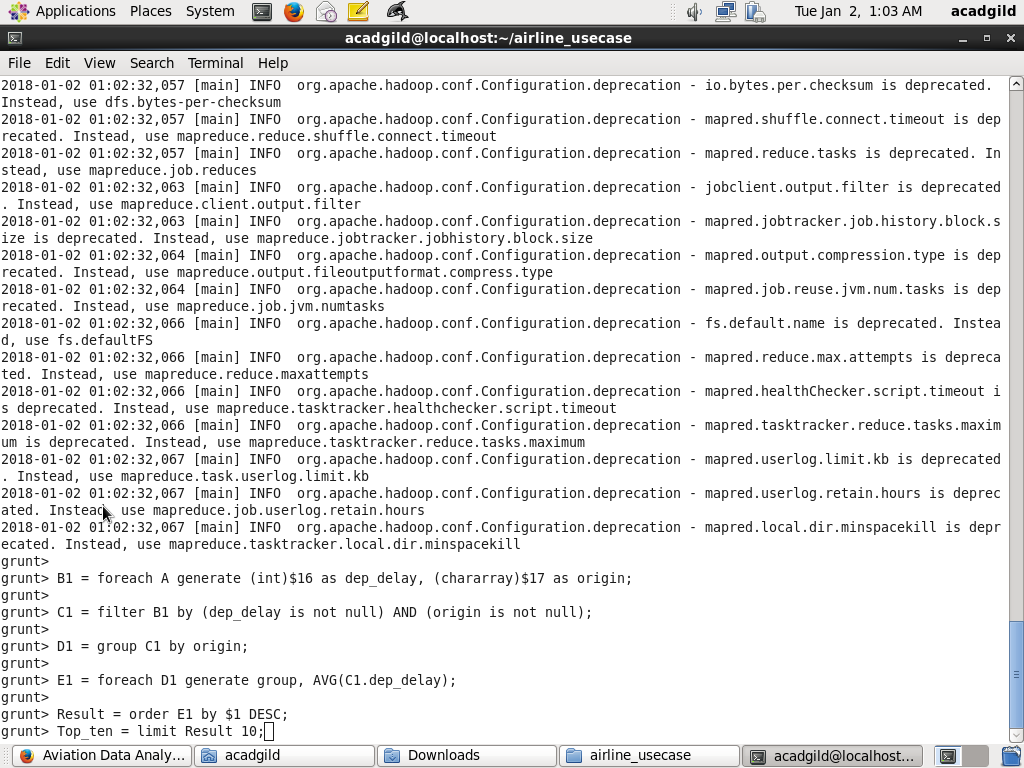
1. **D1 = group C1 by origin;**

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

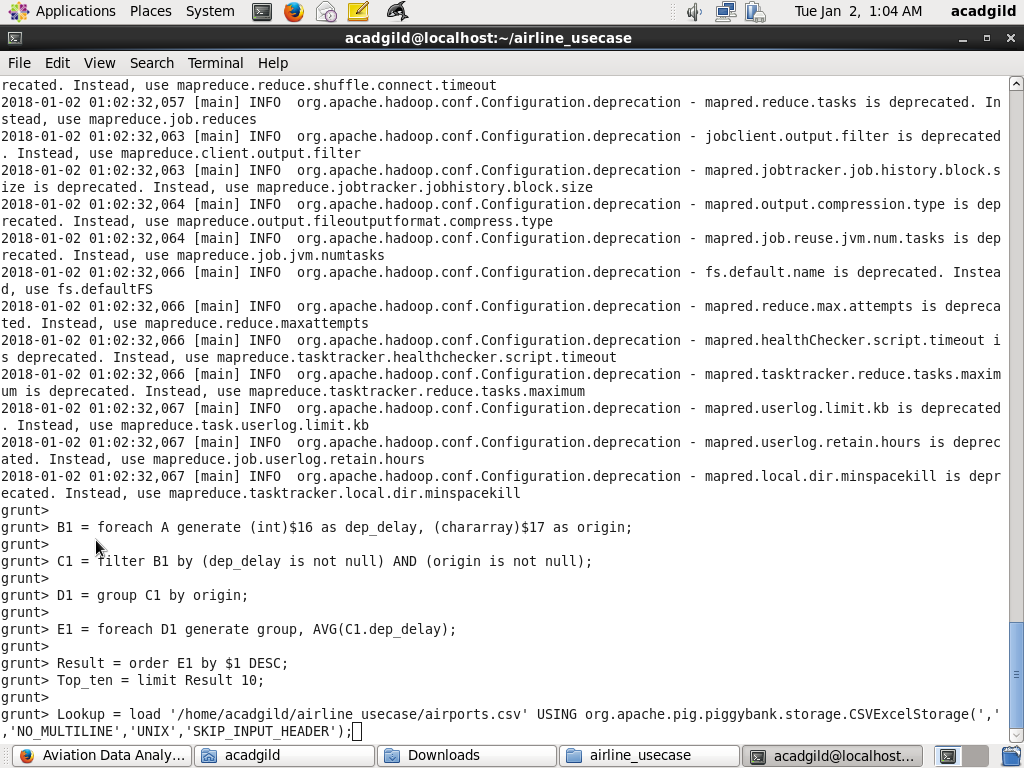


1. **Result = order E1 by $1 DESC;**

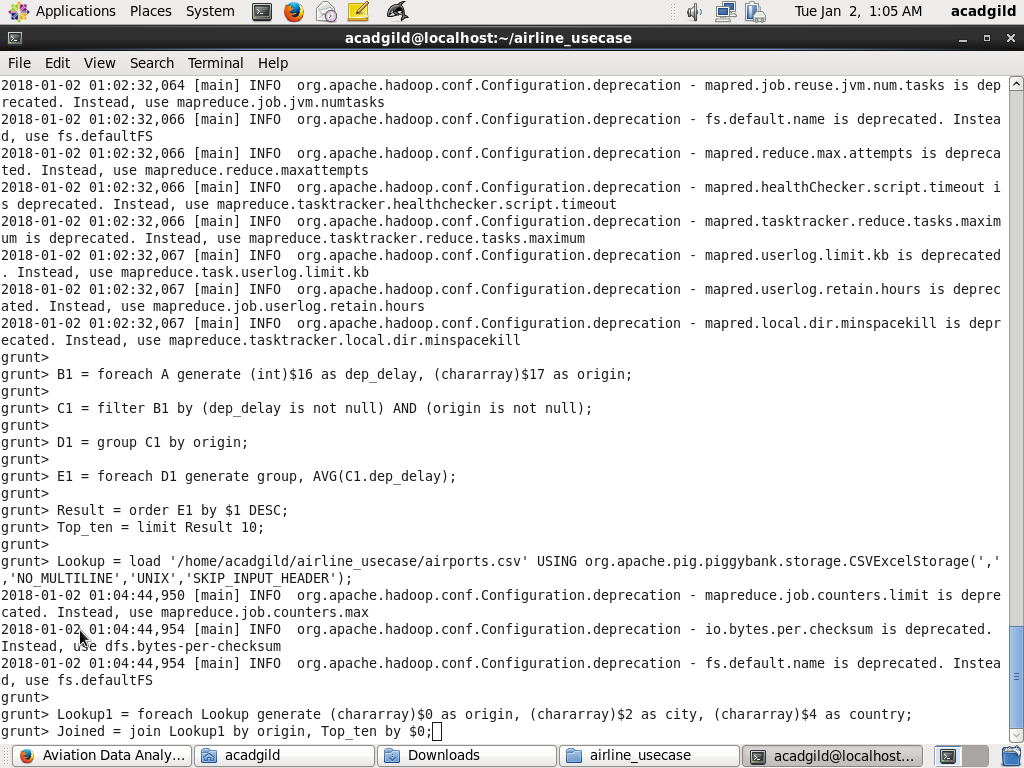
1. **Top\_ten = limit Result 10;**



1. **Lookup = load '/home/acadgild/airline\_usecase/airports.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');**



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



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

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

1. Final\_Result = ORDER Final by $3 DESC;

1. dump Final\_Result;

