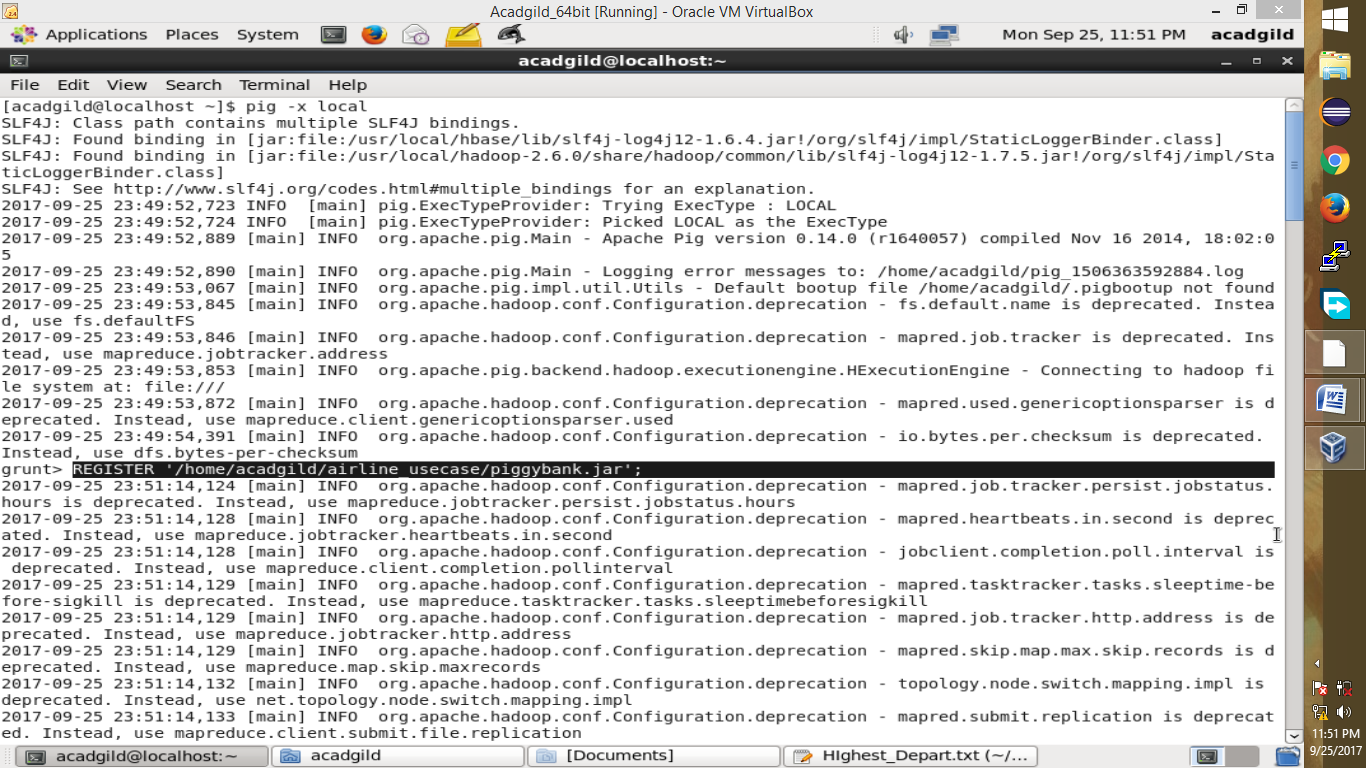
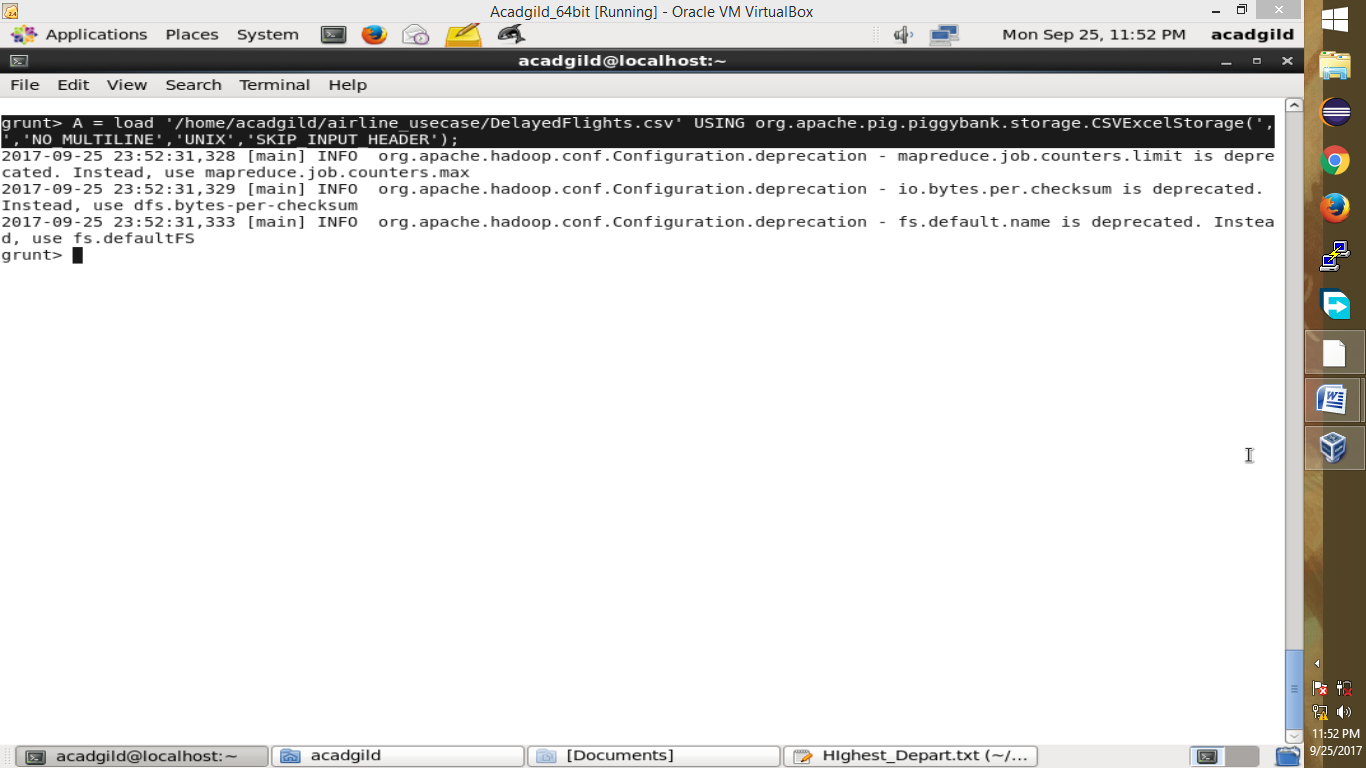
**Problem Statement 3**

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

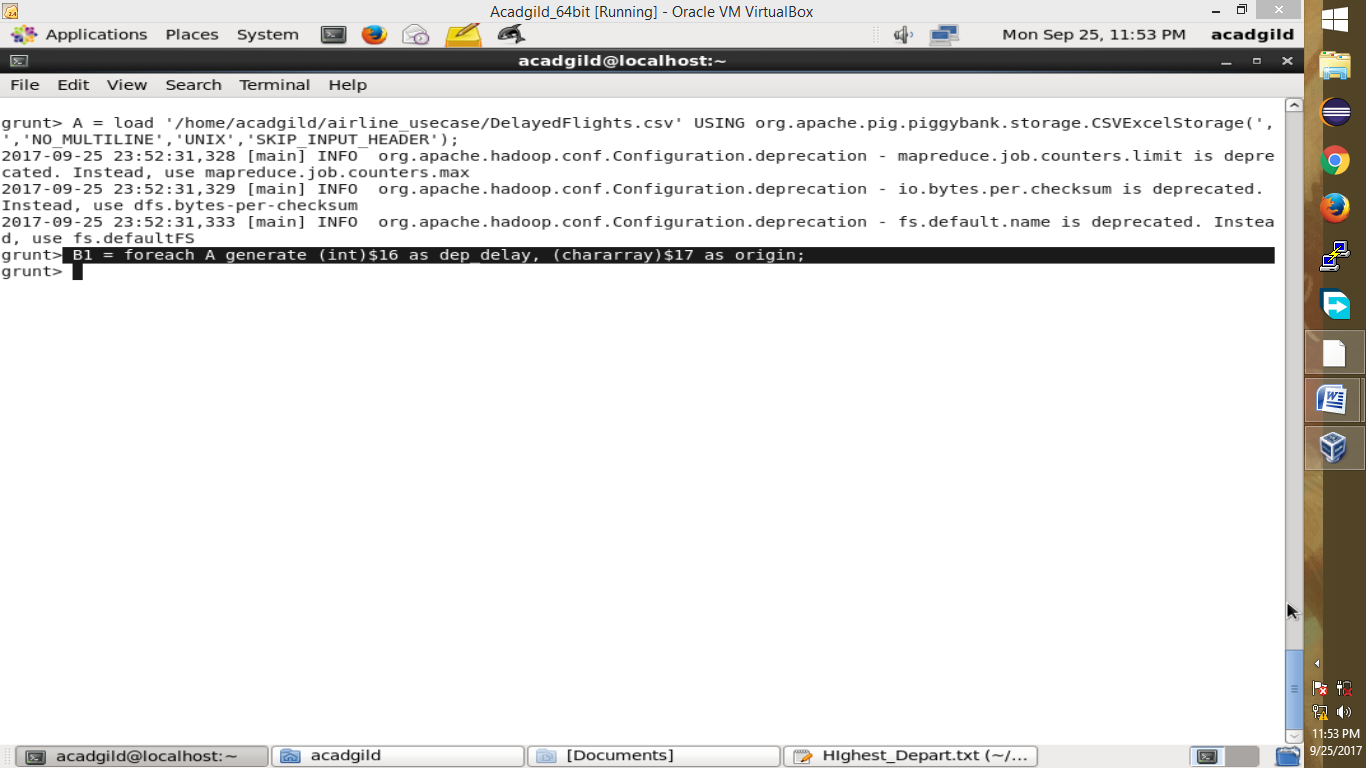
**Step 1: In Line 1**: We are registering *piggybank* jar in order to use the CSVExcelStorage class.



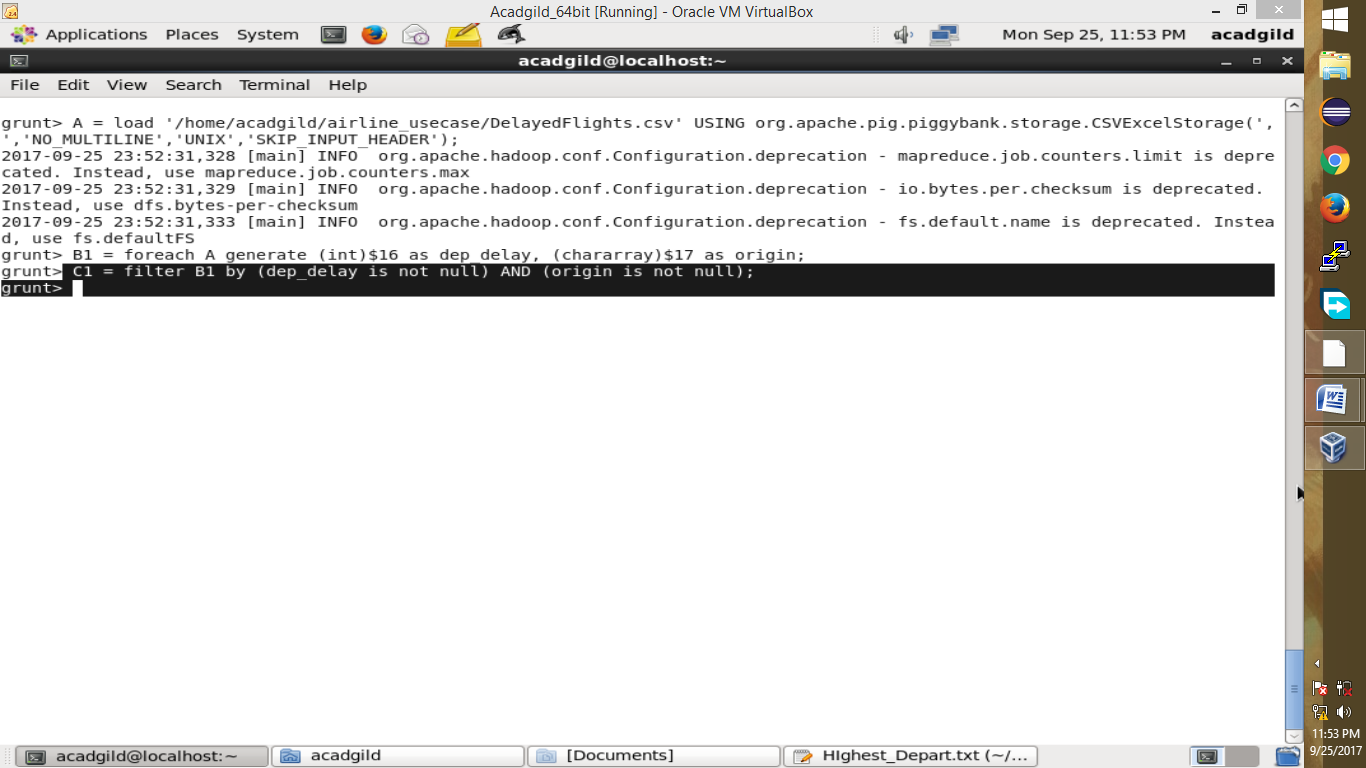
**Step 2:** In relation **A**, we are loading the dataset using CSVExcelStorage because of its effective technique to handle double quotes and header.



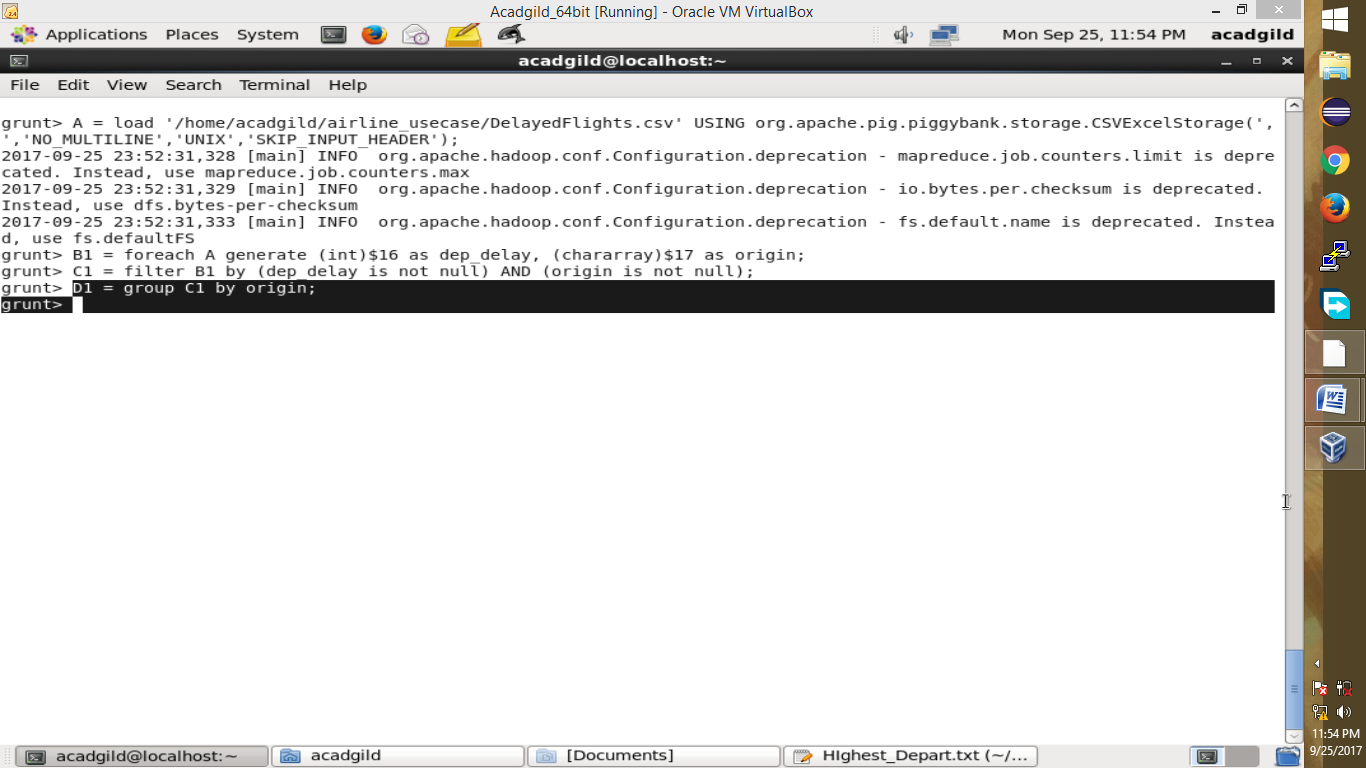
**Step 3:** In relation **B1**, we are generating the columns which are required for processing and explicitly typecasting each of them.



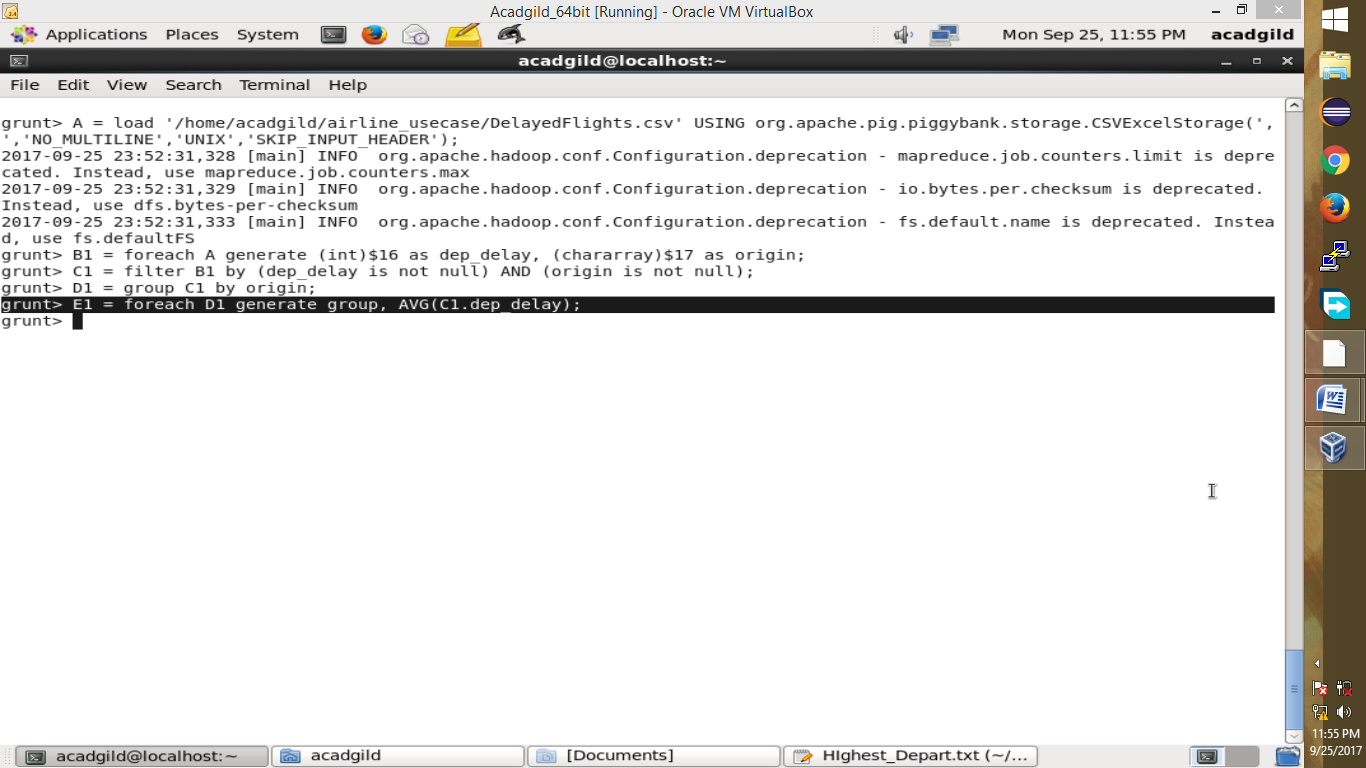
**Step 4:** In relation**C1**, we are removing the null values fields present if any.

****

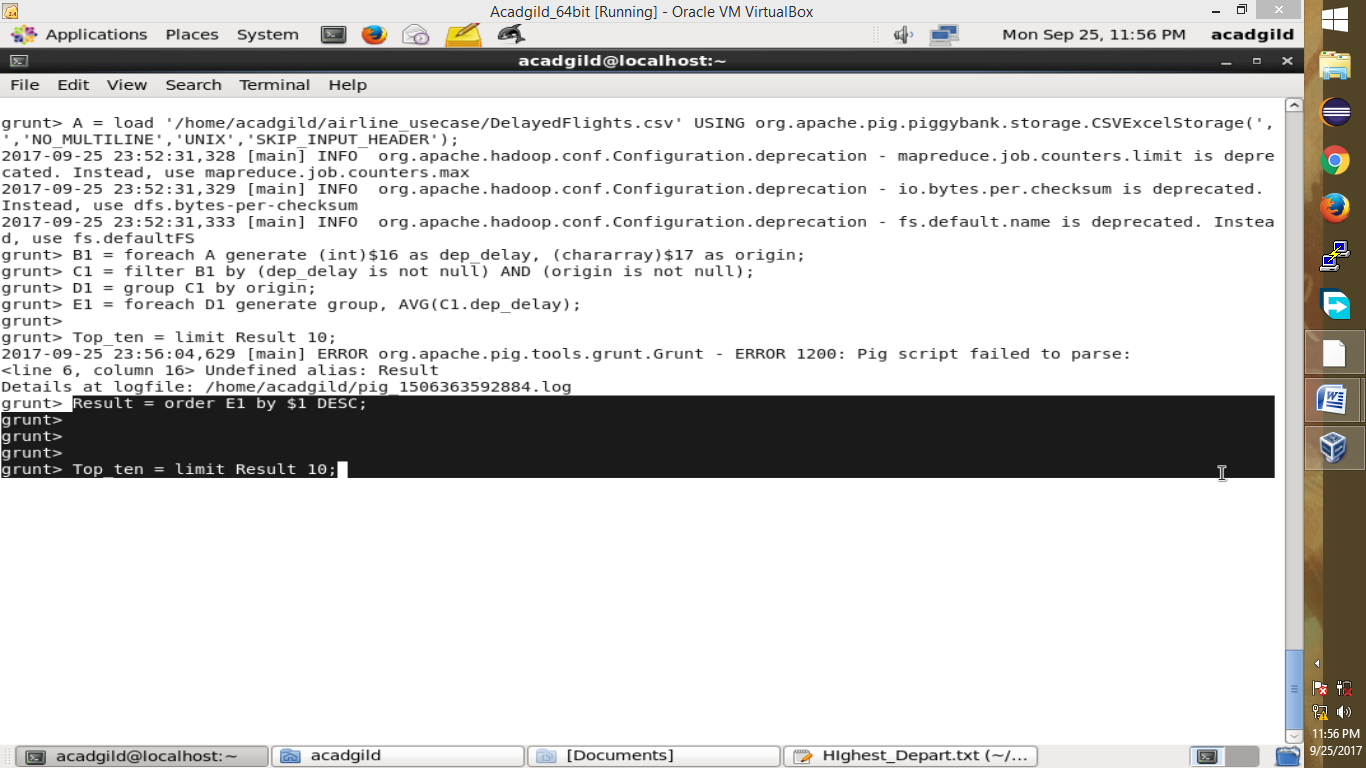
**Step 5:** In relation **D1**, we are grouping the data based on column “origin.”



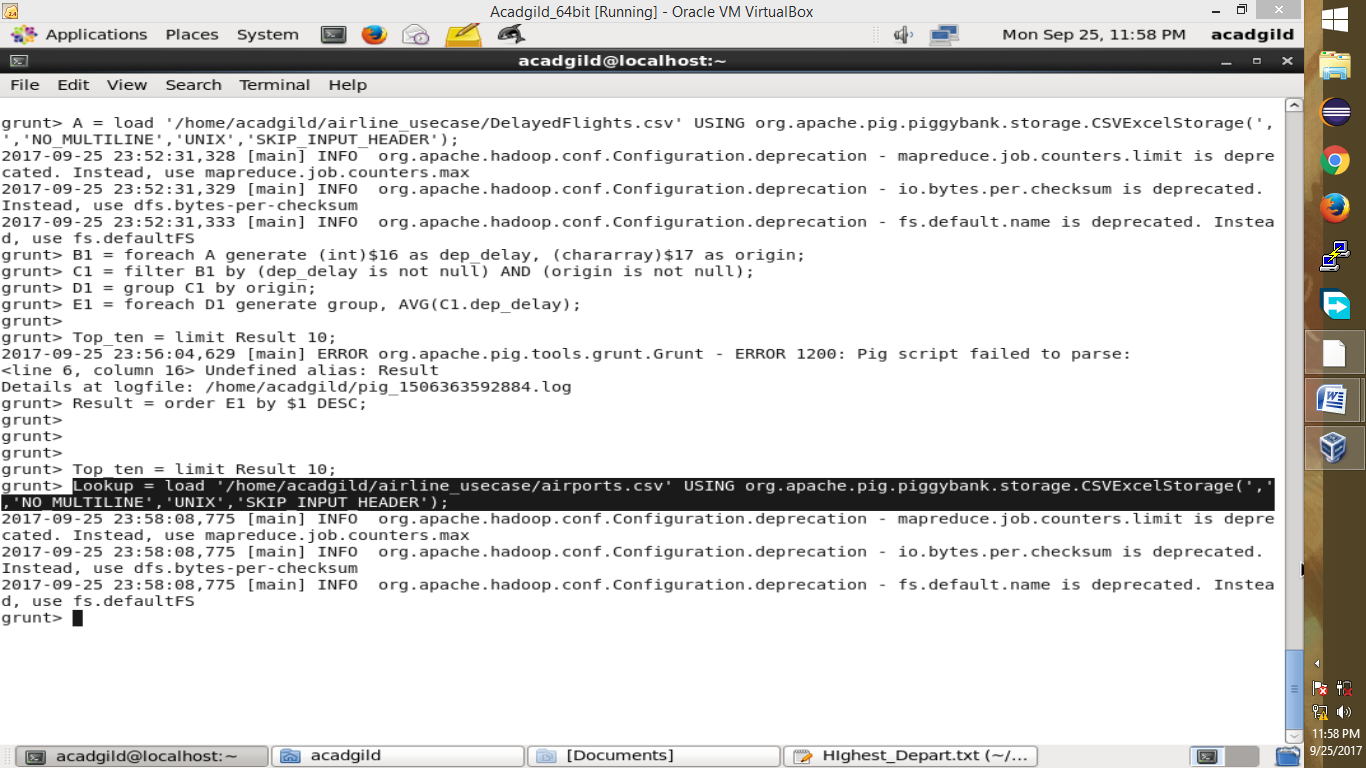
**Step 6:** In relation **E1**, we are finding average delay from each unique origin.



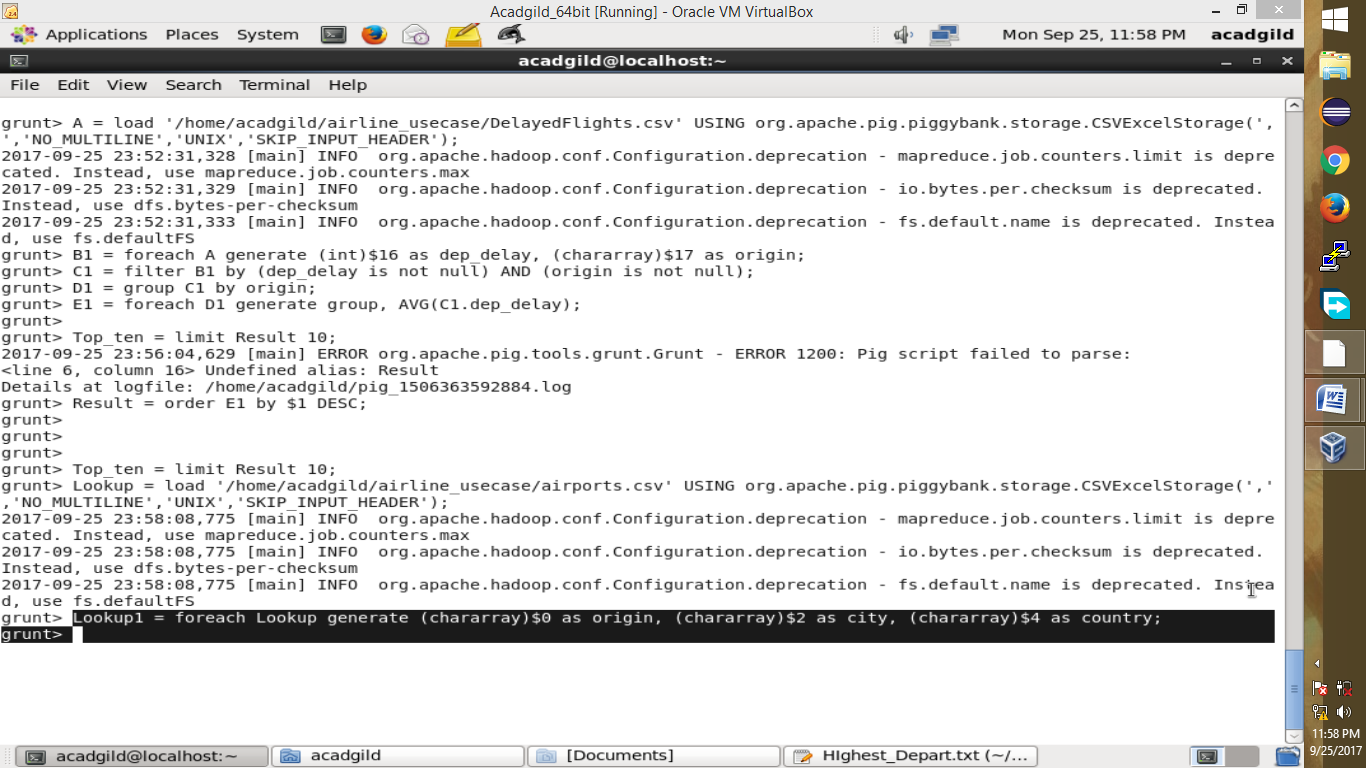
**Step 7:** Relations named **Result** and **Top\_ten** are ordering the results in descending order and printing the top ten values.



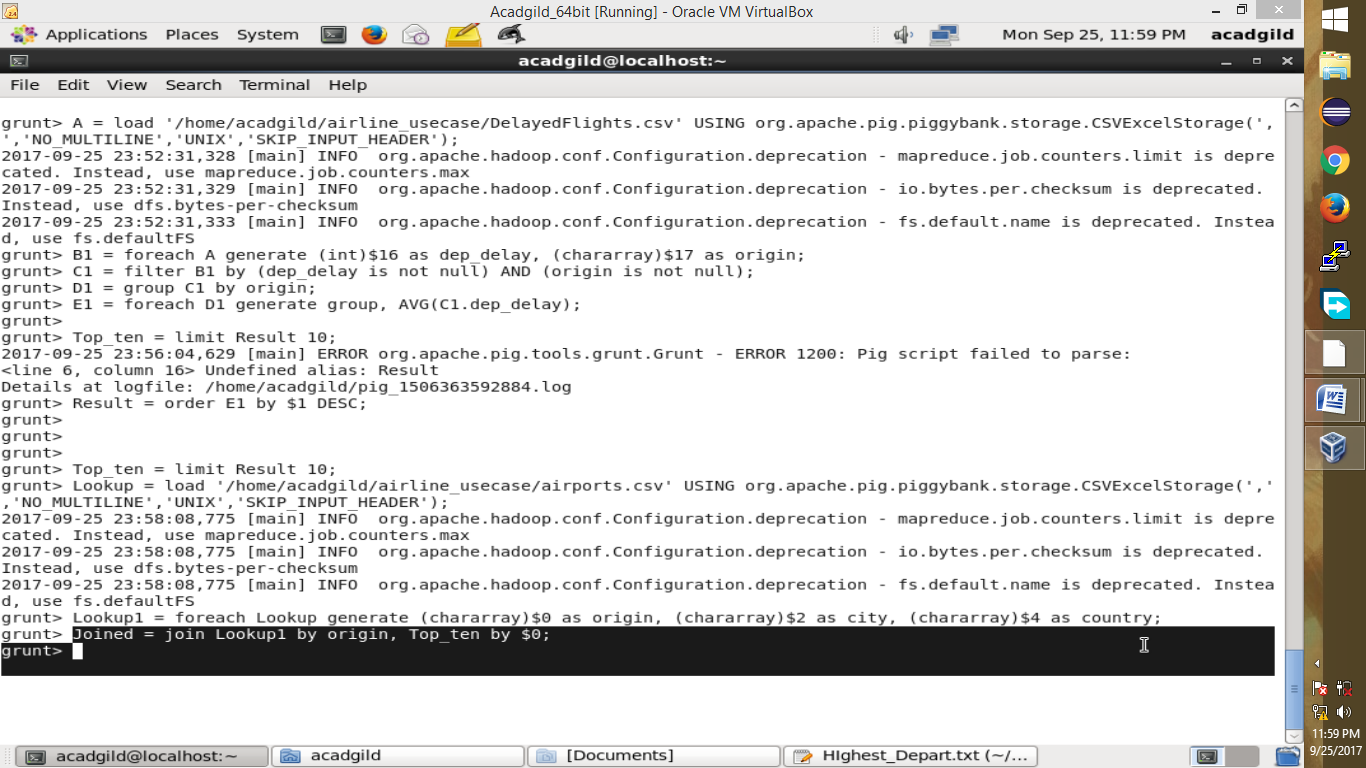
**Step 8:** In the relation **Lookup**, we are loading another table to which we will look up and find the city as well as the country.



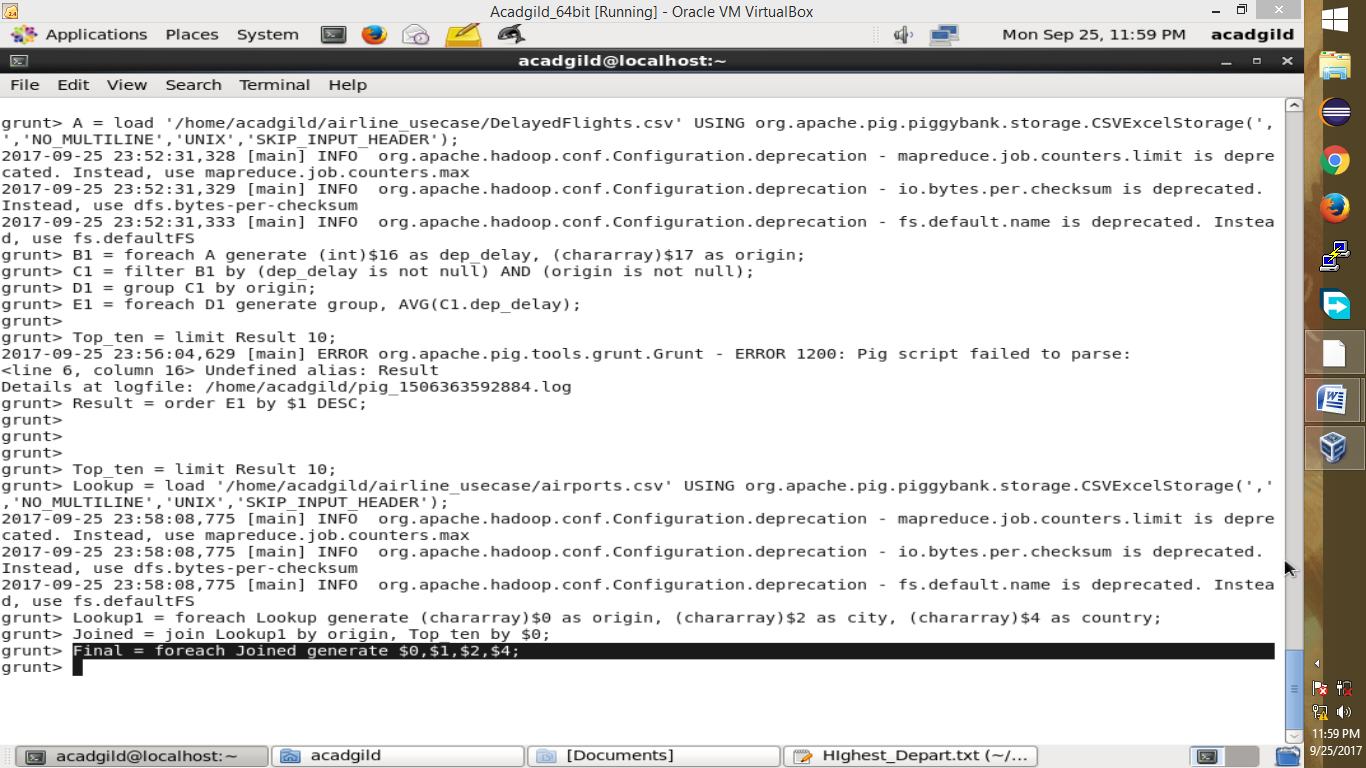
**Step 9:** In the relation L**ookup1,** we are generating the destination, city, and country from the previous relation.



**Step 10:** In the relation **Joined**, we are joining relation Top\_ten and Lookup1 based on common a column, i.e., “origin.”

****

**Step 11:** In the relation**Final,**we are generating required columns from the Joined table.

****

**Step 12:** Finally, we are ordering and printing the results.

