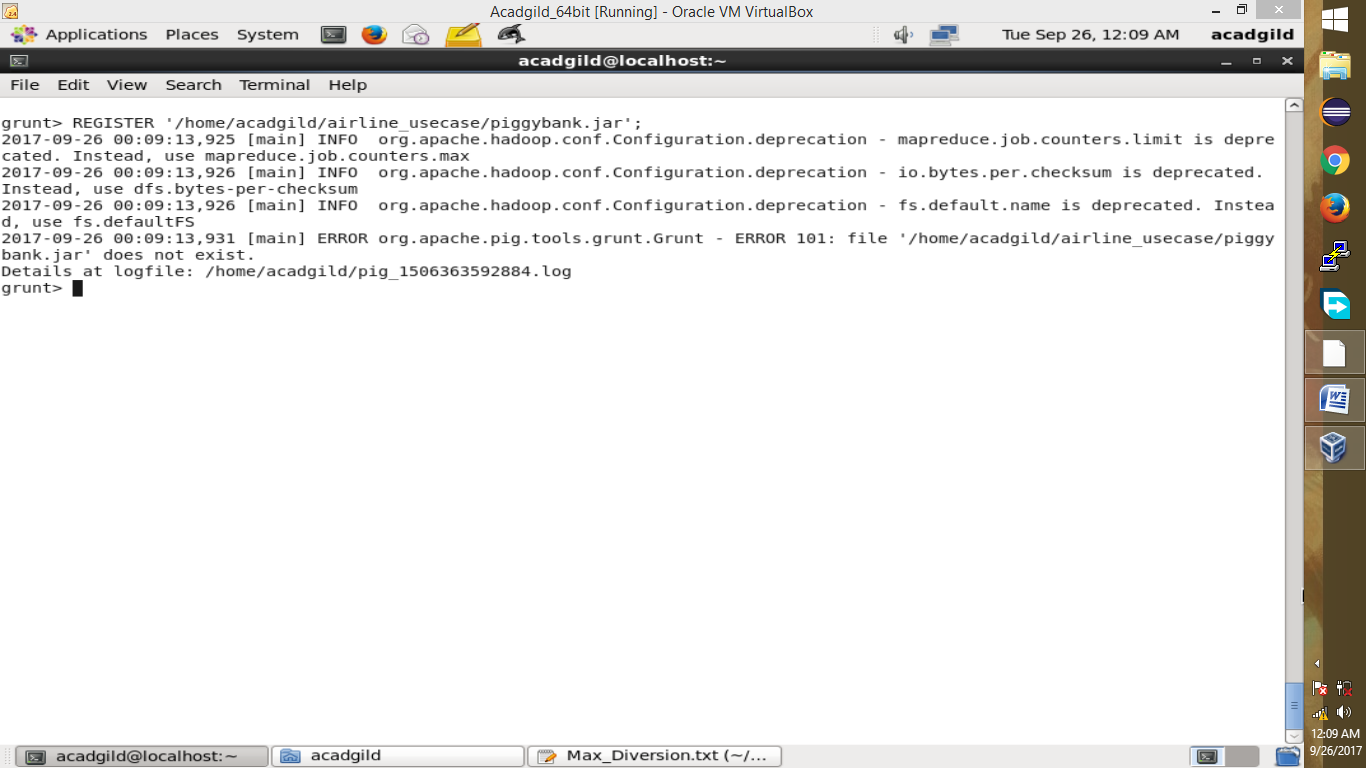
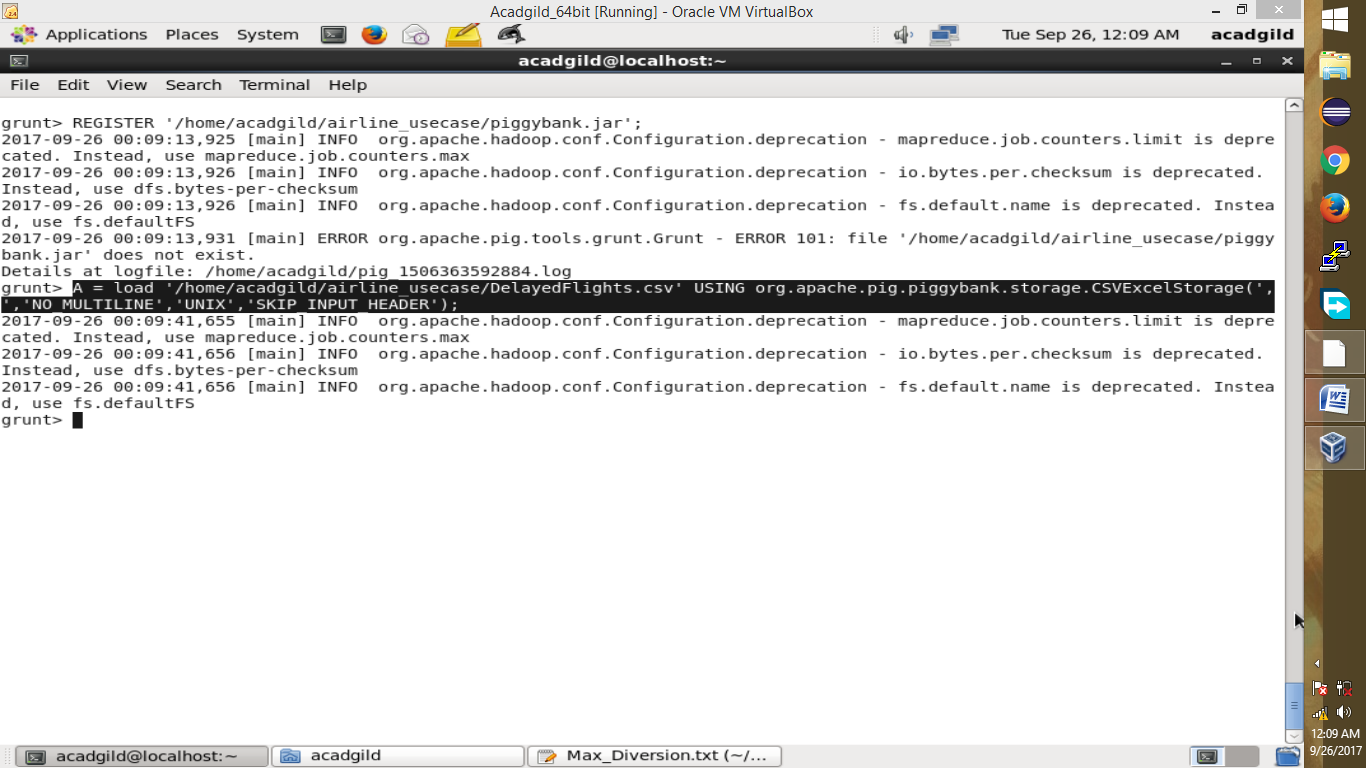
**Problem Statement 4**

Which route (origin & destination) has seen the maximum diversion?

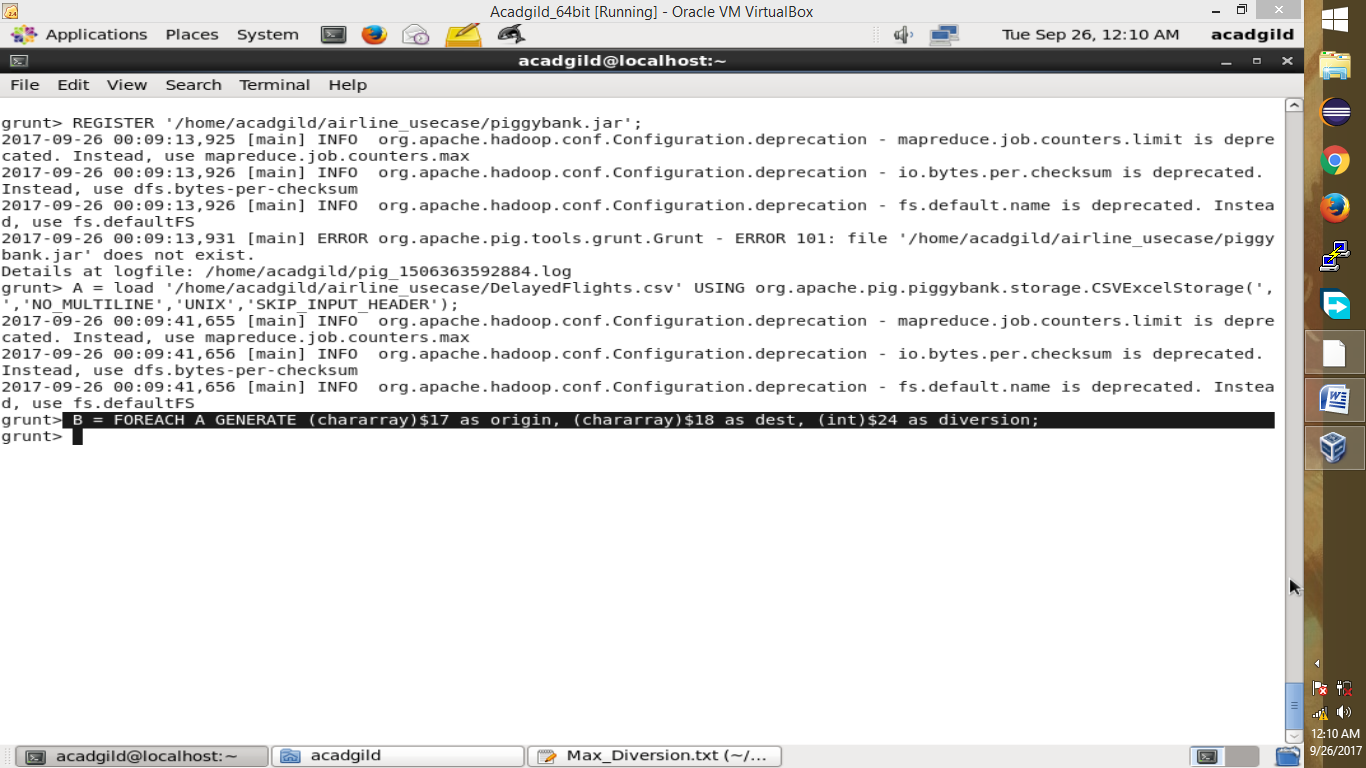
**Step 1: In Line 1**: We are registering *piggybank* jar in order to use CSVExcelStorage class.



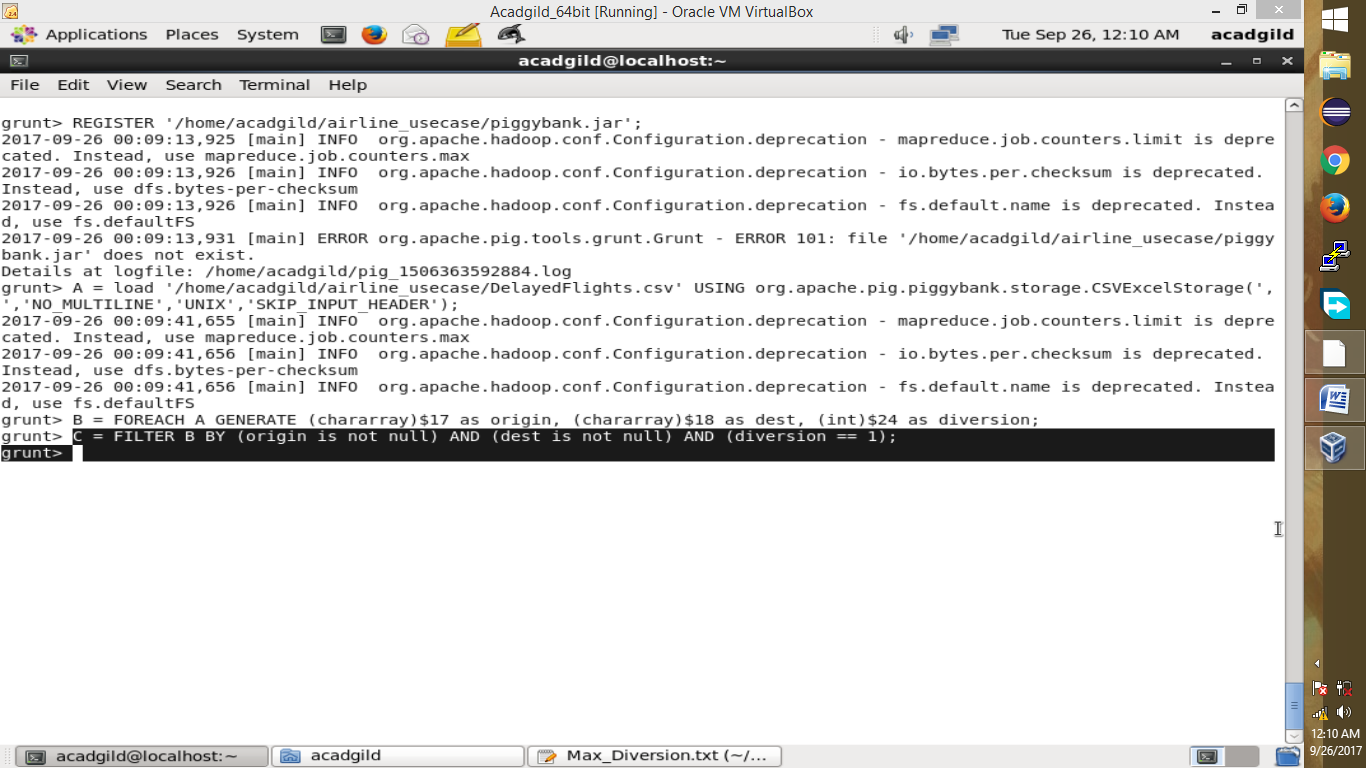
**Step 2:** In relation **A**, we are loading the dataset using CSVExcelStorage because of its effective technique to handle double quotes and headers.



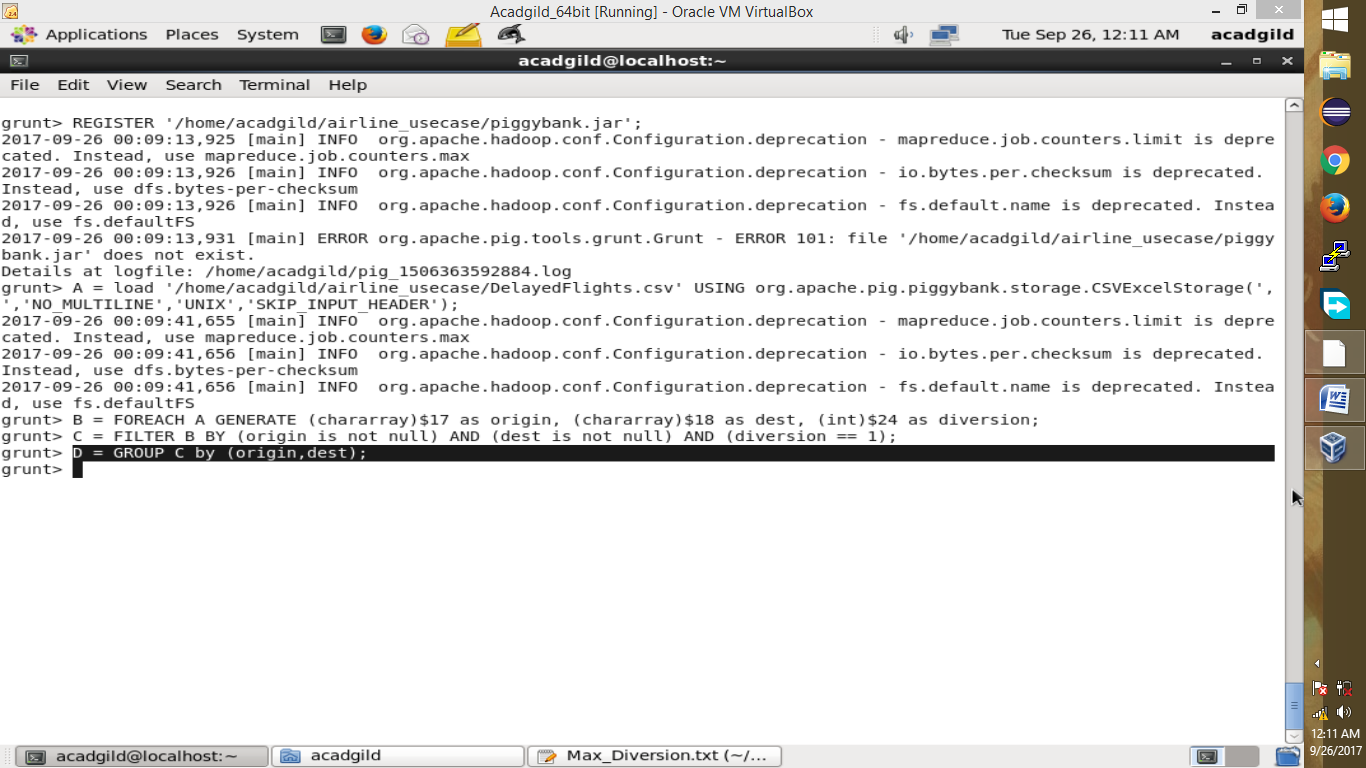
**Step 3:** In relation **B**, we are generating the columns which are required for processing and explicitly type-casting each of them.

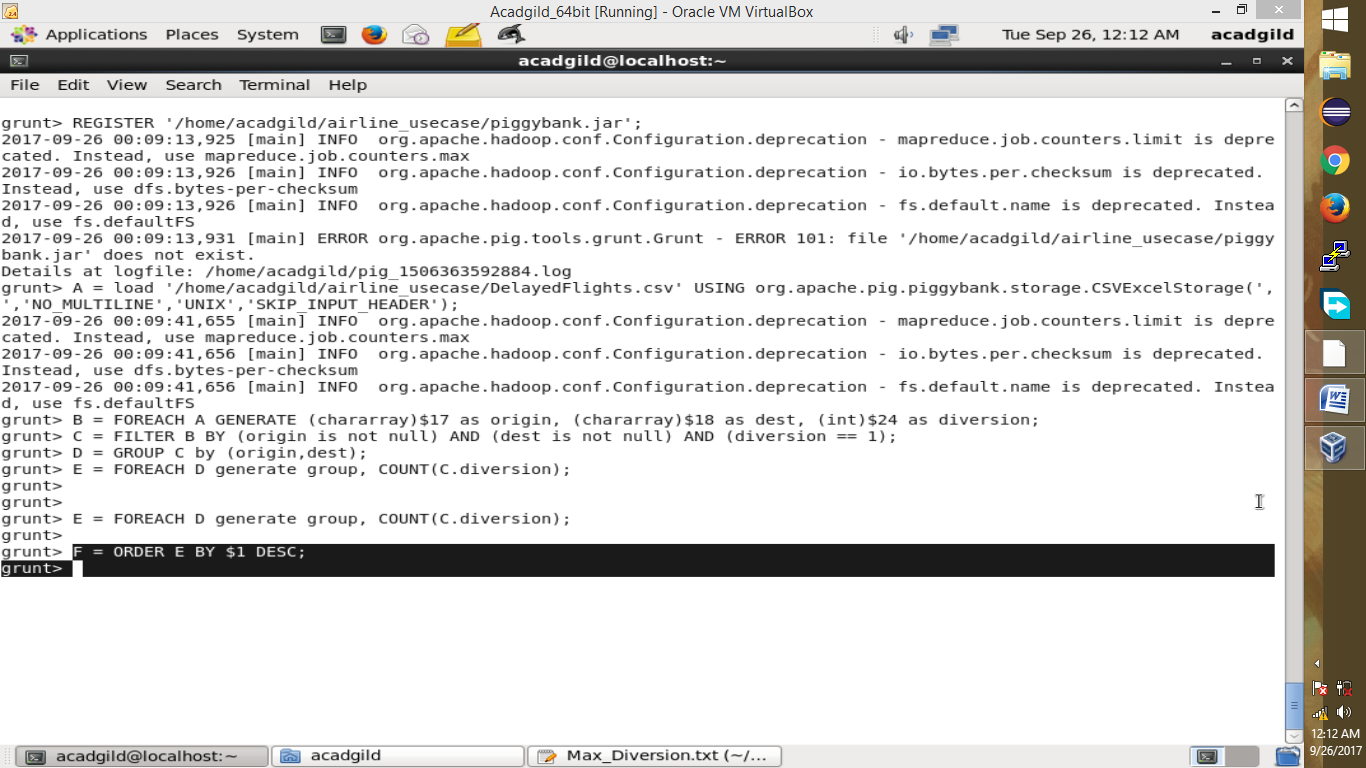
****

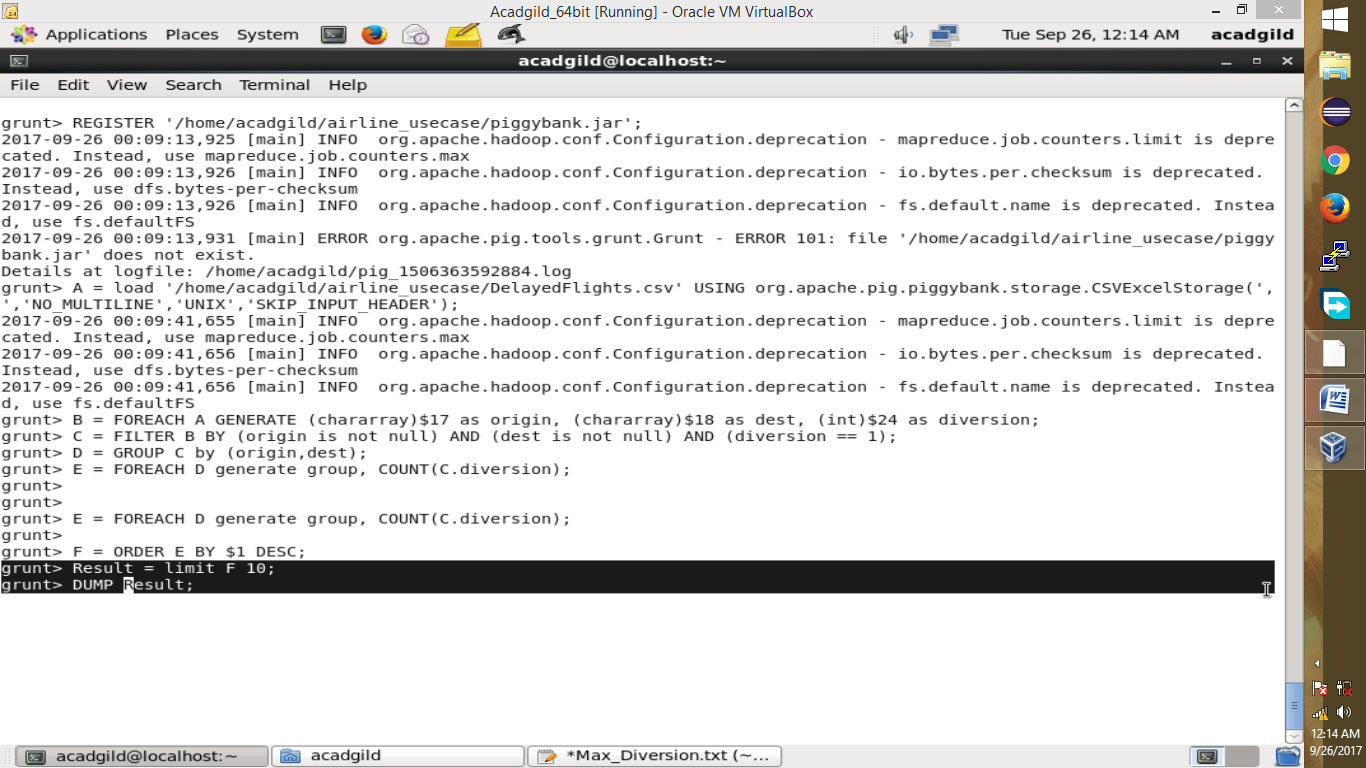
**Step 4:** In relation **C**, we are filtering the data based on “not null” and diversion =1. This will remove the null records, if any, and give the data corresponding to the diversion taken.



**Step 5:** In relation **D**, we are grouping the data based on origin and destination.

****

**Step 6:** Relations **F** and **Result** orders the result and produces top 10 results. ****



Maximum diversion are printed in the below screen:

