**QUESTION:-**

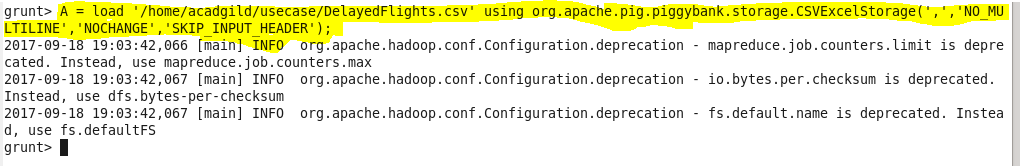


**Solution:-**

First we need to register piggybank.jar



We have created reference named A below and we have loaded the file DelayedFilights.csv and we have loaded it using **CSVExcelStorage()** function. Now relation is referring the file named **DelayedFlights.csv**.



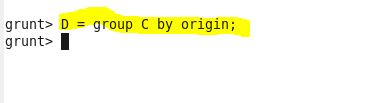
Totally there are 29 columns in DelayedFlights.csv. We can call them as $0 to $28 depending on the column position. Here we took 2 columns and named $16 as Dep\_delay, $17 as origin. Now we have **typecasted.** For each and every row of A, these 2 columns will be generated(retrieved) and it will be referred by **B**.



We have filtered relation B by **Dep\_delay as not null** and **origin as not null** to get the rows that contain valid Dep\_delays and origins and is referred by **relation C**.



We have grouped the origin in relation C and is referred by relation D.



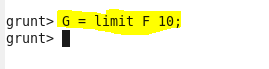
Now for each group in D, we are averaging the Dep\_delay. So that we can get the average departure delay for each and every origin and is **referred by relation E.**



Now we have ordered relation E by $1 (averaged Dep\_delay) in **descending order** and is referred by relation F.



Now we have retrieved first 10 rows of F to get the **top 10 origins with highest average departure delay**.



We have loaded the file that is referred by Relation G using **dump command.**



**OUTPUT:-**

