Below is the script used to find top ten origins with the highest AVG departure delay-

**REGISTER '/home/acadgild/airline\_usecase/piggybank.jar';**

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

**B1 = foreach A 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;**

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

**Lookup = load '/home/acadgild/airline\_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;**

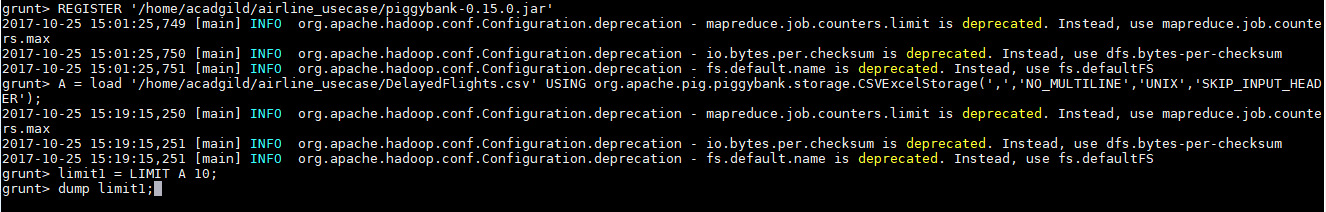
**Final = foreach Joined generate $0,$1,$2,$4;**

**Final\_Result = ORDER Final by $3 DESC;**

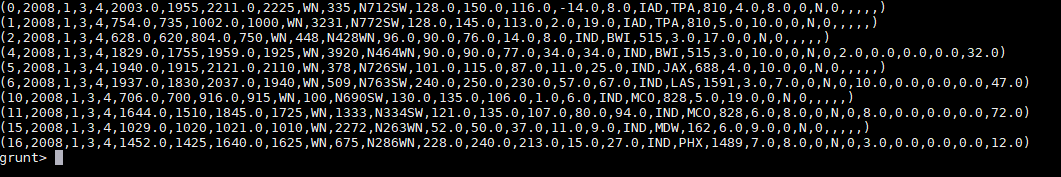
**dump Final\_Result;**

Each and every script has been explained below using intermediate outputs-

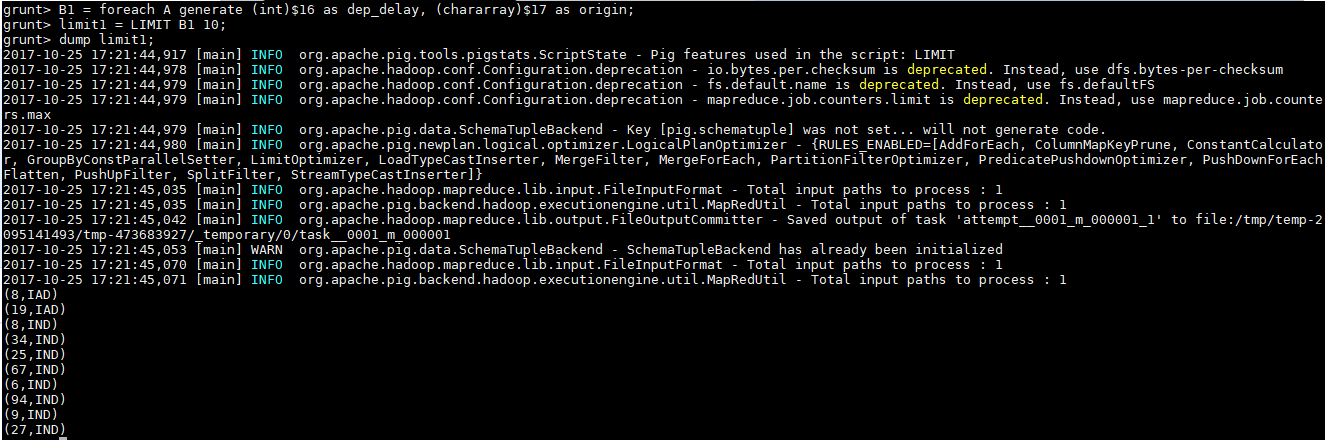
Registering piggybank jar in order to use the CSVExcelStorage class-



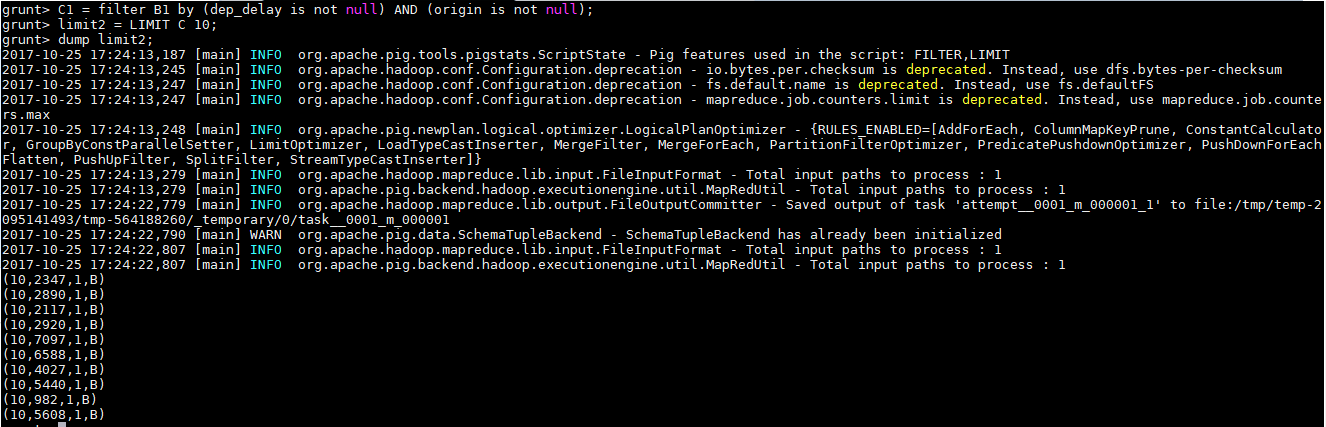
Contents of file-



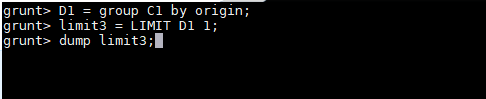
Extract records for departure delay and origin-

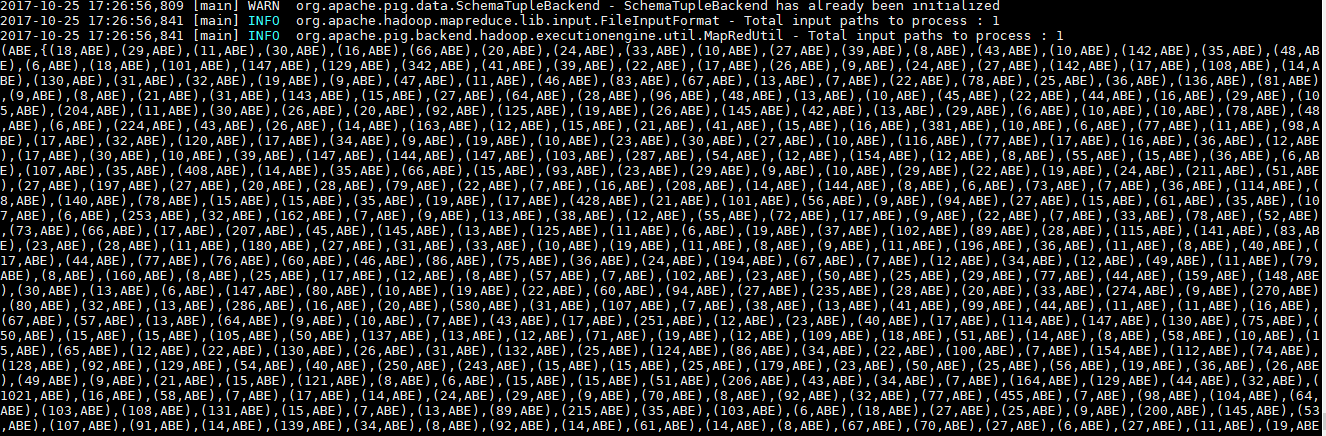


Filter those records which are NULL for departure delay and origin-

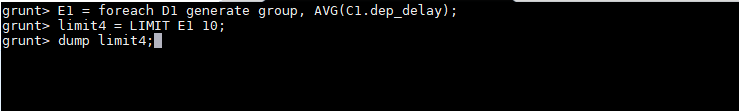


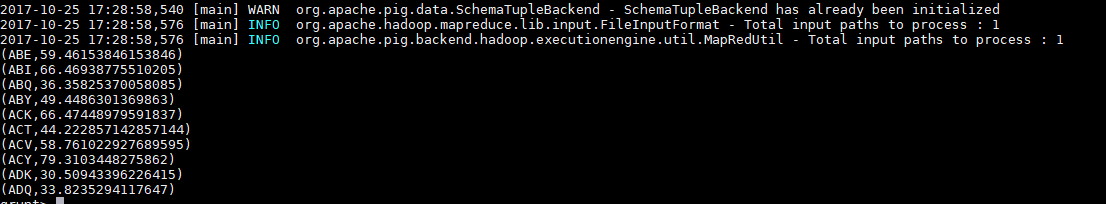
Group above result by origin-



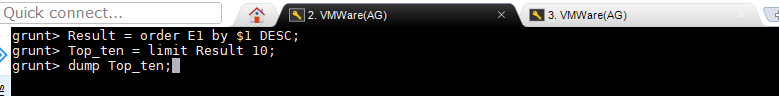


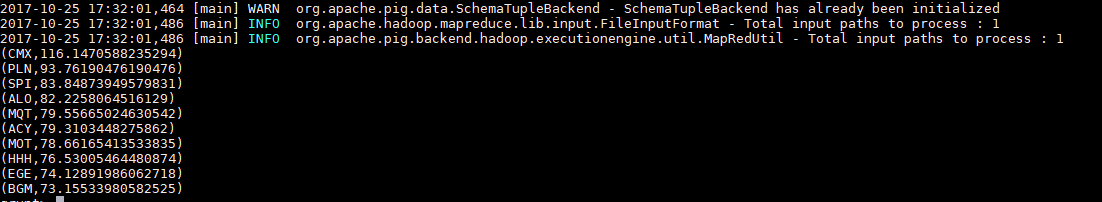
As required take average of departure delay for each element grouped by ORIGIN-



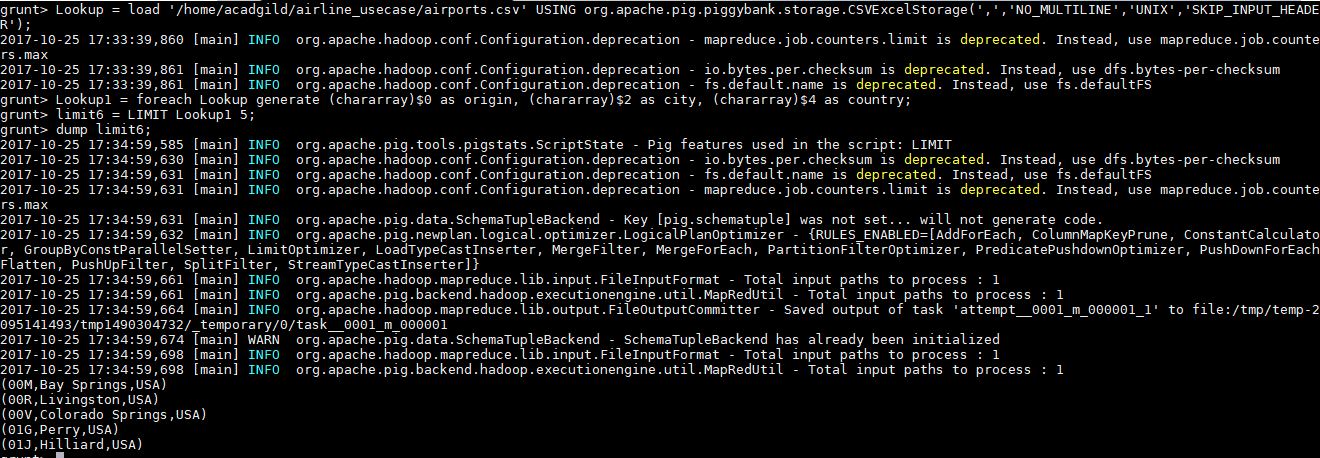


Arrange above result in Descending order and limit it to 10 to find the top 10 -

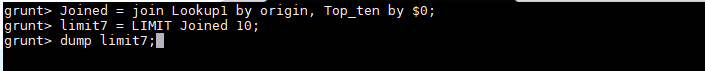


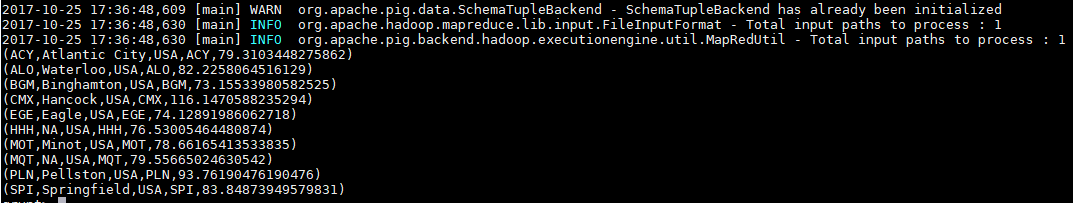


Load airports.csv file to find other records for the airports found from first file-

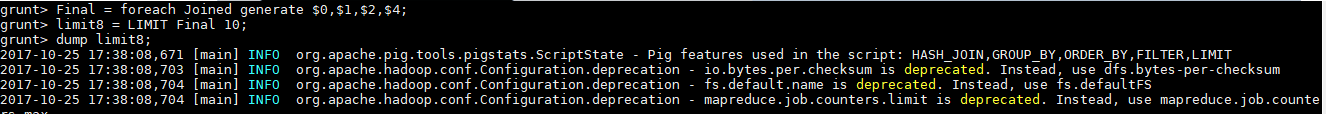


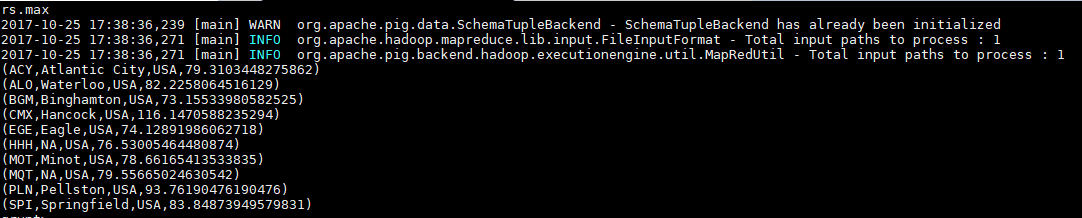
Join the results from first file with second file using origin-



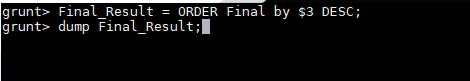


Extract first 4 rows from the result of JOIN- origin, city,country,average





ORDER the result by Average in descending manner-



**Final O/P**

