Below is the script used to find which route (origin & destination) has seen the maximum diversion?

**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');**

**B = FOREACH A GENERATE (chararray)$17 as origin, (chararray)$18 as dest, (int)$24 as diversion;**

**C = FILTER B BY (origin is not null) AND (dest is not null) AND (diversion == 1);**

**D = GROUP C by (origin,dest);**

**E = FOREACH D generate group, COUNT(C.diversion);**

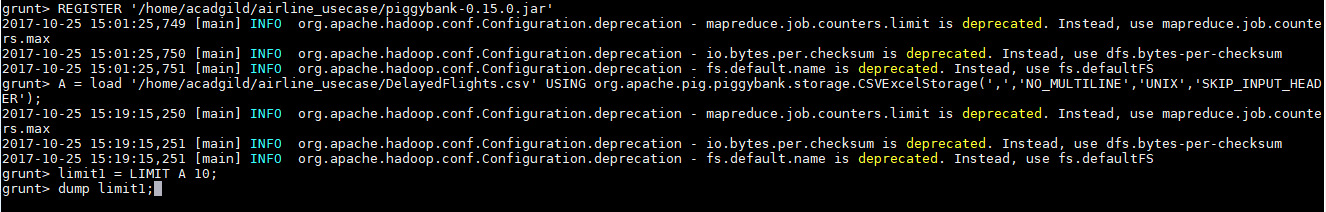
**F = ORDER E BY $1 DESC;**

**Result = limit F 10;**

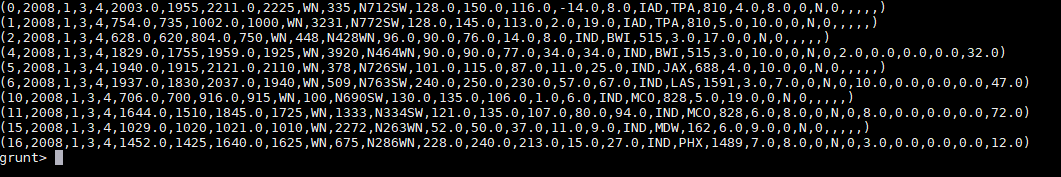
**dump Result;**

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

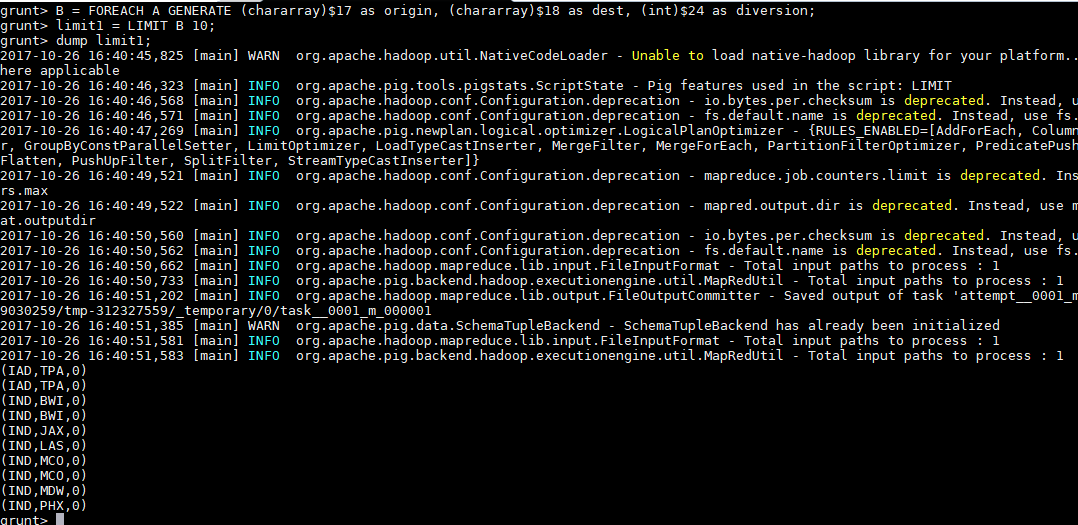
Registering piggybank jar in order to use the CSVExcelStorage class-



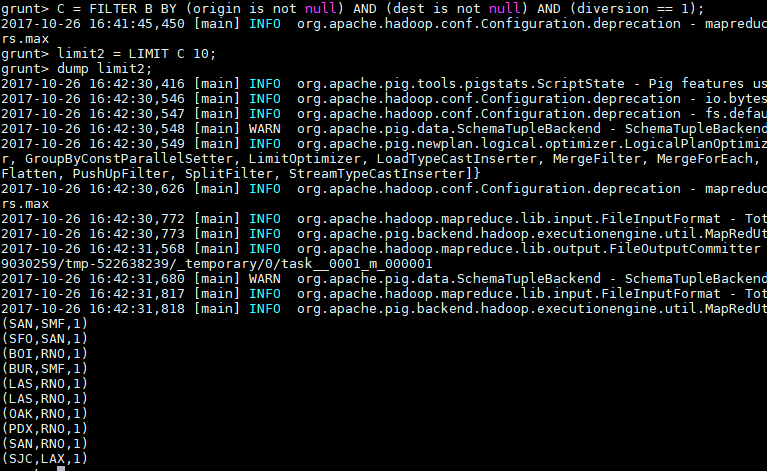
Contents of file-



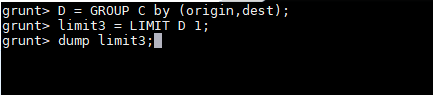
Extract origin, destination and diversion from the input file-

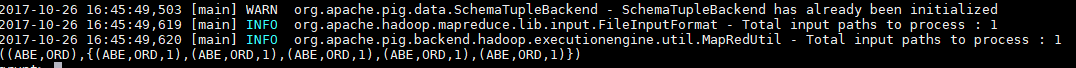


FILTER above result with origin as NOT NULL and DESTINATION as NOT NULL and DIVERSION value as 1-

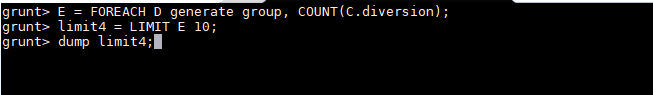


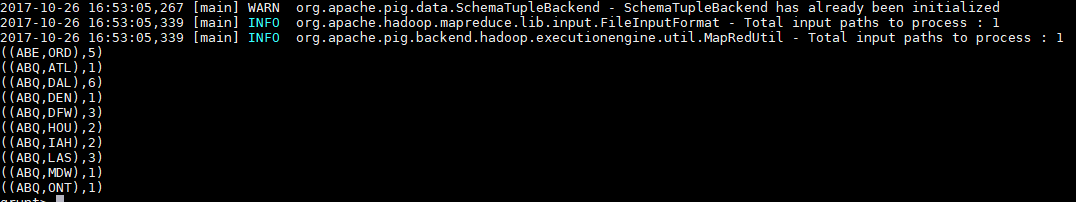
GROUP above result by origin and destination-





From the grouped record extract COUNT of diversions-





Arrange above result in descending order to sort it from maximum to minimum and extract first 10 rows-



**Final O/P-**

