**Assgn : 5.2**

**Problem Statement 1:**

**delFlights = LOAD '/home/acadgild/hadoop/DelayedFlights.csv' USING PigStorage(',');**

**selRel = FOREACH delFlights GENERATE (int)$1 as year, (int)$10 as flight\_num, (chararray)$17 as origin, (chararray)$18 as dest;**

**filRel = FILTER selRel by dest is not null;**

**grpRel = group filRel by dest;**

**countRel = FOREACH grpRel GENERATE group,COUNT(filRel.dest);**

**orderRel = order countRel by $1 desc;**

**resRel = LIMIT orderRel 5;**

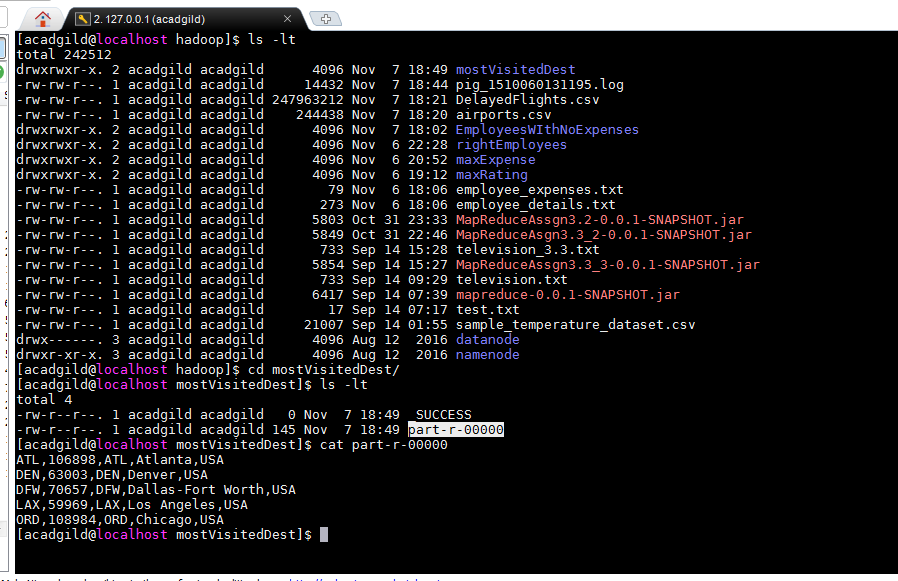
**airportRel = LOAD '/home/acadgild/hadoop/airports.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage**

**(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');**

**aselRel = FOREACH airportRel GENERATE (chararray)$0 as dest, (chararray)$2 as city, (chararray)$4 as country;**

**joined\_table = join resRel by $0, aselRel by dest;**

**STORE joined\_table INTO '/home/acadgild/hadoop/mostVisitedDest' USING PigStorage(',');**



**Problem Statement 2:**

**delFlights = LOAD '/home/acadgild/hadoop/DelayedFlights.csv' USING PigStorage(',');**

**selRel = FOREACH delFlights GENERATE (int)$2 as month, (int)$10 as flight\_num, (int)$22 as cancelled, (chararray)$23 as cancel\_code;**

**filRel = FILTER selRel by cancelled == 1 AND cancel\_code == 'B';**

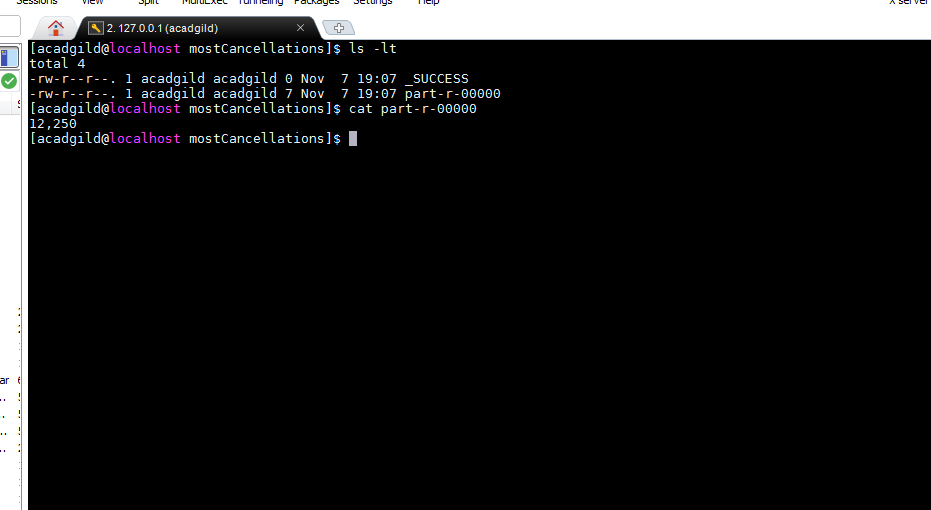
**grpRel = group filRel by month;**

**countRel = FOREACH grpRel GENERATE group,COUNT(filRel.cancelled);**

**orderRel = order countRel by $1 desc;**

**resRel = LIMIT orderRel 1;**

**STORE resRel INTO '/home/acadgild/hadoop/mostCancellations' USING PigStorage(',');**



**Problem Statement 3:**

**A = LOAD '/home/acadgild/hadoop/DelayedFlights.csv' USING PigStorage(',');**

**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/hadoop/airports.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');**

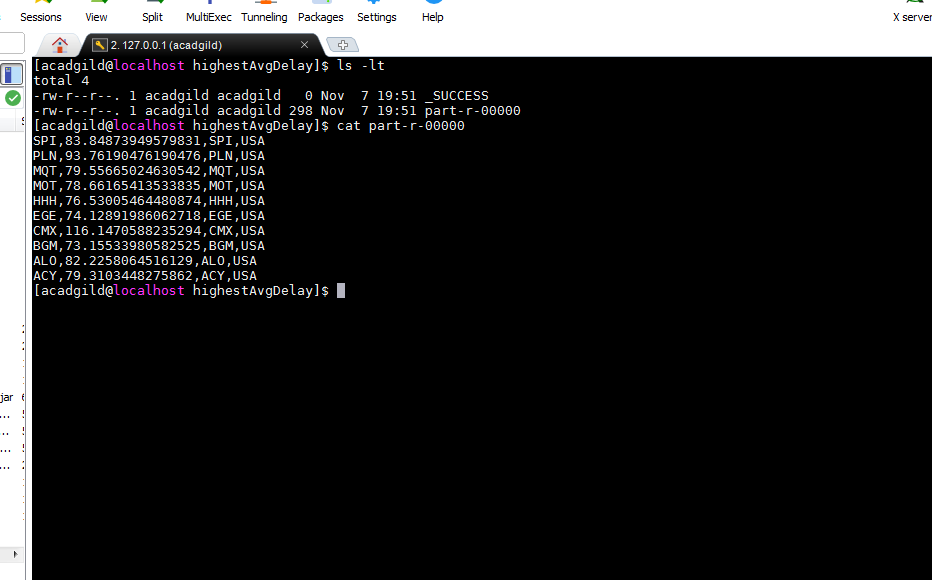
**Lookup1 = FOREACH Lookup GENERATE (chararray)$0 as origin, (chararray)$2 as city, (chararray)$4 as country;**

**joined = join top\_ten by $0, Lookup1 by origin;**

**final = FOREACH joined GENERATE $0,$1,$2,$4;**

**final\_result = ORDER final BY $3 desc;**

**STORE final\_result INTO '/home/acadgild/hadoop/highestAvgDelay' USING PigStorage(',');**



**Problem Statement 4:**

**A = LOAD '/home/acadgild/hadoop/DelayedFlights.csv' USING PigStorage(',');**

**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;**

**STORE result INTO '/home/acadgild/hadoop/maxDiversion' USING PigStorage(',');**

