

Below is the script used to find top5 destinations-

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 (int)\$1 as year, (int)\$10 as flight\_num, (chararray)\$17 as origin,(chararray) \$18 as dest;

C = filter B by dest is not null;

D = group C by dest;

E = foreach D generate group, COUNT(C.dest);

F = order E by \$1 DESC;

Result = LIMIT F 5;

A1 = load '/home/acadgild/airline\_usecase/airports.csv' USING  
org.apache.pig.piggybank.storage.CSVExcelStorage('','NO\_MULTILINE','UNIX','SKIP\_INPUT\_HEADER');

A2 = foreach A1 generate (chararray)\$0 as dest, (chararray)\$2 as city, (chararray)\$4 as country;

joined\_table = join Result by \$0, A2 by dest;

dump joined\_table;

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

Registering *piggybank* jar in order to use the CSVExcelStorage class-

```
grunt> REGISTER '/home/acadgild/airline_usecase/piggybank-0.15.0.jar'
2017-10-25 15:01:25,749 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 15:01:25,750 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 15:01:25,751 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> A = load '/home/acadgild/airline_usecase/DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage('','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
2017-10-25 15:19:15,250 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 15:19:15,251 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 15:19:15,251 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> limit1 = LIMIT A 10;
grunt> dump limit1;█
```

Contents of file-

```
(0,2008,1,3,4,2003.0,1955,2211.0,2225,WN,335,N712SW,128.0,150.0,116.0,-14.0,8.0,IAD,TPA,810,4.0,8.0,0,N,0,,,,,)
(1,2008,1,3,4,754.0,735,1002.0,1000,WN,3231,N772SW,128.0,145.0,113.0,2.0,19.0,IAD,TPA,810,5.0,10.0,0,N,0,,,,,)
(2,2008,1,3,4,628.0,620,804.0,750,WN,448,N428WN,96.0,90.0,76.0,14.0,8.0,IND,BWI,515,3.0,17.0,0,N,0,,,,,)
(4,2008,1,3,4,1829.0,1755,1959.0,1925,WN,3920,N464WN,90.0,90.0,77.0,34.0,34.0,IND,BWI,515,3.0,10.0,0,N,0,2.0,0.0,0.0,0.0,32.0)
(5,2008,1,3,4,1940.0,1915,2121.0,2110,WN,378,N726SW,101.0,115.0,87.0,11.0,25.0,IND,JAX,688,4.0,10.0,0,N,0,,,,,)
(6,2008,1,3,4,1937.0,1830,2037.0,1940,WN,509,N763SW,240.0,250.0,230.0,57.0,67.0,IND,LAS,1591,3.0,7.0,0,N,0,10.0,0.0,0.0,0.0,47.0)
(10,2008,1,3,4,706.0,700,916.0,915,WN,100,N690SW,130.0,135.0,106.0,1.0,6.0,IND,MCO,828,5.0,19.0,0,N,0,,,,,)
(11,2008,1,3,4,1644.0,1510,1845.0,1725,WN,1333,N334SW,121.0,135.0,107.0,80.0,94.0,IND,MCO,828,6.0,8.0,0,N,0,8.0,0.0,0.0,0.0,72.0)
(15,2008,1,3,4,1029.0,1020,1021.0,1010,WN,2272,N263WN,52.0,50.0,37.0,11.0,9.0,IND,MDW,162,6.0,9.0,0,N,0,,,,,)
(16,2008,1,3,4,1452.0,1425,1640.0,1625,WN,675,N286WN,228.0,240.0,213.0,15.0,27.0,IND,PHX,1489,7.0,8.0,0,N,0,3.0,0.0,0.0,0.0,12.0)
grunt> █
```

Extracting columns for year, flight number, origin and destination-

```
grunt> B = foreach A generate (int)$1 as year, (int)$10 as flight_num, (chararray)$17 as origin,(chararray) $18 as dest;
grunt> limit2 = LIMIT B 10;
grunt> dump limit2;
2017-10-25 15:27:06,420 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: LIMIT
2017-10-25 15:27:06,826 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 15:27:06,827 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-25 15:27:06,828 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 15:27:06,829 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2017-10-25 15:27:06,985 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 15:27:07,154 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 15:27:07,155 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-10-25 15:27:07,166 [main] INFO org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - Saved output of task 'attempt_0001_m_000001_1' to file:/tmp/temp1225608307/tmp2074148719/_temporary/0/task_0001_m_000001
2017-10-25 15:27:07,306 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 15:27:07,461 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 15:27:07,462 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(2008,335,IAD,TPA)
(2008,3231,IAD,TPA)
(2008,448,IND,BWI)
(2008,3920,IND,BWI)
(2008,378,IND,JAX)
(2008,509,IND,LAS)
(2008,100,IND,MCO)
(2008,1333,IND,MCO)
(2008,2272,IND,MDW)
(2008,675,IND,PHX)
```

Filter records which have destination as NULL-

```

grunt> C = filter B by dest is not null;
grunt> limit3 = LIMIT C 10;
grunt> dump limit3;
2017-10-25 15:28:37,447 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER,LIMIT
2017-10-25 15:28:37,780 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 15:28:37,781 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-25 15:28:37,782 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 15:28:37,785 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2017-10-25 15:28:37,903 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 15:28:38,070 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 15:28:38,071 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-10-25 15:28:38,084 [main] INFO org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - Saved output of task 'attempt_0001_m_000001_1' to file:/tmp/temp12
25608307/tmp439815025/ temporary/0/task_0001_m_000001
2017-10-25 15:28:38,143 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 15:28:38,326 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 15:28:38,326 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(2008,335,IAD,TPA)
(2008,3231,IAD,TPA)
(2008,448,IND,BWI)
(2008,3920,IND,BWI)
(2008,378,IND,JAX)
(2008,509,IND,LAS)
(2008,100,IND,MCO)
(2008,1333,IND,MCO)
(2008,2272,IND,MDW)
(2008,675,IND,PHX)
grunt>


```

## Group records by destination-

```
Quick connect... 2. VMWare(AG) 3. VMWare(AG)
>> grunt> D = group C by dest;
>> grunt> limit3 = LIMIT D 1;
>> grunt> dump limit3;
```

(ABE),(2008,2937,DTW,ABE),(2008,2937,DTW,ABE),(2008,2937,DTW,ABE),(2008,2939,DTW,ABE),(2008,5289,ATL,ABE),(2008,2939,DTW,ABE),(2008,2939,DTW,ABE),(2008,2939,DTW,ABE),  
(2008,2939,DTW,ABE),(2008,2939,DTW,ABE),(2008,2939,DTW,ABE),(2008,2939,DTW,ABE),(2008,2939,DTW,ABE),(2008,5289,ATL,ABE),(2008,2939,DTW,ABE),(2008,2941,DTW,ABE),(200  
8,2941,DTW,ABE),(2008,2941,DTW,ABE),(2008,2941,DTW,ABE),(2008,2941,DTW,ABE),(2008,2941,DTW,ABE),(2008,5338,ATL,ABE),(2008,5338,ATL,ABE),(2008,5338,ATL,ABE),(2008,533  
8,ATL,ABE),(2008,5338,ATL,ABE),(2008,5338,ATL,ABE),(2008,5338,ATL,ABE),(2008,5338,ATL,ABE),(2008,5461,ATL,ABE),(2008,5461,ATL,ABE),(2008,5461,ATL,ABE),(2008,5289,ATL,ABE),(2008,5461,ATL  
,ABE),(2008,5461,ATL,ABE),(2008,5289,ATL,ABE),(2008,5461,ATL,ABE),(2008,5289,ATL,ABE),(2008,5461,ATL,ABE),(2008,5461,ATL,ABE),(2008,5461,ATL,ABE),(2008,5461,ATL,ABE)  
,(2008,5461,ATL,ABE),(2008,2937,DTW,ABE),(2008,2937,DTW,ABE),(2008,2937,DTW,ABE),(2008,5289,ATL,ABE),(2008,2937,DTW,ABE),(2008,2937,DTW,ABE),(2008,2937,DTW,ABE),(200  
8,2937,DTW,ABE),(2008,2937,DTW,ABE),(2008,2652,CLE,ABE),(2008,2626,CLE,ABE),(2008,5816,ORD,ABE),(2008,7263,ORD,ABE),(2008,2672,CLT,ABE),(2008,2937,DTW,ABE),(2008,309  
4,CLE,ABE),(2008,2652,CLE,ABE),(2008,1032,CLT,ABE),(2008,7263,ORD,ABE),(2008,5816,ORD,ABE),(2008,5383,ATL,ABE),(2008,6006,ORD,ABE),(2008,6006,ORD,ABE),(2008,5816,ORD  
,ABE),(2008,2610,CLT,ABE),(2008,2578,CLE,ABE),(2008,1032,CLT,ABE),(2008,5508,ATL,ABE),(2008,5508,ATL,ABE),(2008,7263,ORD,ABE),(2008,7171,ORD,ABE),(2008,7263,ORD,ABE)  
,(2008,5508,ATL,ABE),(2008,5508,ATL,ABE),(2008,2578,CLE,ABE),(2008,7171,ORD,ABE),(2008,7291,ORD,ABE),(2008,5508,ATL,ABE),(2008,5508,ATL,ABE),(2008,2578,CLE,ABE),(200  
8,7263,ORD,ABE),(2008,7171,ORD,ABE),(2008,7263,ORD,ABE),(2008,3094,CLE,ABE),(2008,5816,ORD,ABE),(2008,7263,ORD,ABE),(2008,5383,ATL,ABE),(2008,7263,ORD,ABE),(2008,293  
9,DTW,ABE),(2008,5461,ATL,ABE),(2008,5461,ATL,ABE),(2008,6006,ORD,ABE),(2008,6006,ORD,ABE),(2008,2652,CLE,ABE),(2008,5841,ORD,ABE),(2008,5929,ORD,ABE),(2008,5929,ORD  
,ABE),(2008,1032,CLT,ABE),(2008,1032,CLT,ABE),(2008,2578,CLE,ABE),(2008,5338,ATL,ABE),(2008,1032,CLT,ABE),(2008,5816,ORD,ABE),(2008,6006,ORD,ABE),(2008,2626,CLE,ABE)  
,(2008,5383,ATL,ABE),(2008,2939,DTW,ABE),(2008,2941,DTW,ABE),(2008,2626,CLE,ABE),(2008,2610,CLT,ABE),(2008,7291,ORD,ABE),(2008,2941,DTW,ABE),(2008,2941,DTW,ABE),(200  
8,5841,ORD,ABE),(2008,5383,ATL,ABE),(2008,5816,ORD,ABE),(2008,2941,DTW,ABE),(2008,5289,ATL,ABE),(2008,5841,ORD,ABE),(2008,1032,CLT,ABE),(2008,6002,ORD,ABE),(2008,581  
6,ORD,ABE),(2008,2672,CLT,ABE),(2008,5816,ORD,ABE),(2008,5338,ATL,ABE),(2008,5929,ORD,ABE),(2008,2652,CLE,ABE),(2008,5841,ORD,ABE),(2008,7291,ORD,ABE),(2008,5508,ATL  
,ABE),(2008,5508,ATL,ABE),(2008,2672,CLT,ABE),(2008,2672,CLT,ABE),(2008,2672,CLT,ABE),(2008,5338,ATL,ABE),(2008,2652,CLE,ABE),(2008,1032,CLT,ABE),(2008,5338,ATL,ABE)  
,(2008,5338,ATL,ABE),(2008,5338,ATL,ABE),(2008,5338,ATL,ABE),(2008,5816,ORD,ABE),(2008,6002,ORD,ABE),(2008,5816,ORD,ABE),(2008,5816,ORD,ABE),(2008,3094,CLE,ABE),(200  
8,7171,ORD,ABE),(2008,2672,CLT,ABE),(2008,6002,ORD,ABE),(2008,2939,DTW,ABE),(2008,2672,CLT,ABE),(2008,7263,ORD,ABE),(2008,6002,ORD,ABE),(2008,6002,ORD,ABE),(2008,267  
2,CLT,ABE),(2008,7263,ORD,ABE),(2008,5289,ATL,ABE),(2008,7263,ORD,ABE),(2008,6002,ORD,ABE),(2008,7171,ORD,ABE),(2008,2626,CLE,ABE),(2008,6006,ORD,ABE),(2008,6002,ORD  
,ABE),(2008,5816,ORD,ABE),(2008,5929,ORD,ABE),(2008,5338,ATL,ABE),(2008,5338,ATL,ABE),(2008,2937,DTW,ABE),(2008,1032,CLT,ABE),(2008,2626,CLE,ABE),(2008,5461,ATL,ABE)  
,(2008,5461,ATL,ABE),(2008,5508,ATL,ABE),(2008,5841,ORD,ABE),(2008,5841,ORD,ABE),(2008,5841,ORD,ABE),(2008,2578,CLE,ABE),(2008,5841,ORD,ABE),(2008,5841,ORD,ABE),(2008  
8,2652,CLE,ABE),(2008,5841,ORD,ABE),(2008,2626,CLE,ABE),(2008,2626,CLE,ABE),(2008,2626,CLE,ABE),(2008,3094,CLE,ABE),(2008,5816,ORD,ABE),(2008,5461,ATL,ABE),(2008,729  
1,ORD,ABE),(2008,7263,ORD,ABE),(2008,7291,ORD,ABE),(2008,3094,CLE,ABE),(2008,5461,ATL,ABE),(2008,2626,CLE,ABE),(2008,7291,ORD,ABE),(2008,5461,ATL,ABE),(2008,1032,CLT

Taking number of count for each destination to find the frequency of destination visited-



Quick connect...

2. VMWare(AG) 3. VMWare(AG)

```
grunt> E = foreach D generate group, COUNT(C.dest);
>> grunt> limit4 = LIMIT E 10;
grunt> dump limit4;
```

Sessions

Tools

```
2017-10-25 15:50:59,649 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 15:50:59,714 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 15:50:59,714 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(ABE,1322)
(ABI,838)
(ABQ,12183)
(ABY,319)
(ACK,155)
(ACT,546)
(ACV,1102)
(ACY,38)
(ADK,31)
(ADQ,105)
```

### Sorting the count for destination

```
grunt> F = order E by $1 DESC;
grunt> Result = LIMIT F 5;
grunt> dump Result;
```



```
2017-10-25 15:53:04,121 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 15:53:04,173 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 15:53:04,173 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(ORD,108984)
(ATL,106898)
(DFW,70657)
(DEN,63003)
(LAX,59969)
grunt> █
```

Loading the airports.csv file to find other fields-

```
grunt> A1 = load '/home/acadgild/airline_usecase/airports.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
2017-10-25 15:54:21,327 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counte
rs.max
2017-10-25 15:54:21,328 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 15:54:21,328 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> A2 = foreach A1 generate (chararray)$0 as dest, (chararray)$2 as city, (chararray)$4 as country;
grunt> limit5 = LIMIT A2 10;
grunt> dump limit5;█
```

Contents of file-

```
2017-10-25 15:56:05,896 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 15:56:05,931 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 15:56:05,932 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(00M,Bay Springs,USA)
(00R,Livingston,USA)
(00V,Colorado Springs,USA)
(01G,Perry,USA)
(01J,Hilliard,USA)
(01M,Belmont,USA)
(02A,Clanton,USA)
(02C,Brookfield,USA)
(02G,East Liverpool,USA)
(03D,Memphis,USA)
grunt> █
```

Joining result from 1<sup>st</sup> file and 2<sup>nd</sup> file with destination id-

```
grunt> joined_table = join Result by $0, A2 by dest;
grunt> dump joined_table;█
```

Final Output-

```
2017-10-25 15:58:44,768 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 15:58:44,811 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 15:58:44,811 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(ATL,106898,ATL,Atlanta,USA)
(DEN,63003,DEN,Denver,USA)
(DFW,70657,DFW,Dallas-Fort Worth,USA)
(LAX,59969,LAX,Los Angeles,USA)
(ORD,108984,ORD,Chicago,USA)
grunt> █
```

Below is the script used to find which month has seen the most number of cancellations due to bad weather?

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 (int)\$2 as month,(int)\$10 as flight\_num,(int)\$22 as cancelled,(chararray)\$23 as cancel\_code;

C = filter B by cancelled == 1 AND cancel\_code == 'B';

D = group C by month;

E = foreach D generate group, COUNT(C.cancelled);

F= order E by \$1 DESC;

Result = limit F 1;

dump Result;

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

Registering piggybank jar in order to use the CSVExcelStorage class-

```
grunt> REGISTER '/home/acadgild/airline_usecase/piggybank-0.15.0.jar'
2017-10-25 15:01:25,749 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 15:01:25,750 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 15:01:25,751 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> A = load '/home/acadgild/airline_usecase/DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
2017-10-25 15:19:15,250 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 15:19:15,251 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 15:19:15,251 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> limit1 = LIMIT A 10;
grunt> dump limit1;█
```

Contents of file-

```
(0,2008,1,3,4,2003.0,1955,2211.0,2225,WN,335,N712SW,128.0,150.0,116.0,-14.0,8.0,IAD,TPA,810,4.0,8.0,0,N,0,,,,,)
(1,2008,1,3,4,754.0,735,1002.0,1000,WN,3231,N772SW,128.0,145.0,113.0,2.0,19.0,IAD,TPA,810,5.0,10.0,0,N,0,,,,,)
(2,2008,1,3,4,628.0,620,804.0,750,WN,448,N428WN,96.0,90.0,76.0,14.0,8.0,IND,BWI,515,3.0,17.0,0,N,0,,,,,)
(4,2008,1,3,4,1829.0,1755,1959.0,1925,WN,3920,N464WN,90.0,90.0,77.0,34.0,34.0,IND,BWI,515,3.0,10.0,0,N,0,2.0,0.0,0.0,0.0,32.0)
(5,2008,1,3,4,1940.0,1915,2121.0,2110,WN,378,N726SW,101.0,115.0,87.0,11.0,25.0,IND,JAX,688,4.0,10.0,0,N,0,,,,,)
(6,2008,1,3,4,1937.0,1830,2037.0,1940,WN,509,N763SW,240.0,250.0,230.0,57.0,67.0,IND,LAS,1591,3.0,7.0,0,N,0,10.0,0.0,0.0,0.0,47.0)
(10,2008,1,3,4,706.0,700,916.0,915,WN,100,N690SW,130.0,135.0,106.0,1.0,6.0,IND,MCO,828,5.0,19.0,0,N,0,,,,,)
(11,2008,1,3,4,1644.0,1510,1845.0,1725,WN,1333,N334SW,121.0,135.0,107.0,80.0,94.0,IND,MCO,828,6.0,8.0,0,N,0,8.0,0.0,0.0,0.0,72.0)
(15,2008,1,3,4,1029.0,1020,1021.0,1010,WN,2272,N263WN,52.0,50.0,37.0,11.0,9.0,IND,MDW,162,6.0,9.0,0,N,0,,,,,)
(16,2008,1,3,4,1452.0,1425,1640.0,1625,WN,675,N286WN,228.0,240.0,213.0,15.0,27.0,IND,PHX,1489,7.0,8.0,0,N,0,3.0,0.0,0.0,0.0,12.0)
grunt> █
```

Extract month, flight number, cancelled indicator and cancelled code from th above file-

```
grunt> B = foreach A generate (int)$2 as month,(int)$10 as flight_num,(int)$22 as cancelled,(chararray)$23 as cancel_code;
grunt> limit1 = LIMIT B 20;
grunt> dump limit1;
2017-10-25 16:43:32,931 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2017-10-25 16:43:33,035 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: LIMIT
2017-10-25 16:43:33,111 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 16:43:33,112 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-25 16:43:33,574 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInsertter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInsertter]}
2017-10-25 16:43:33,802 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 16:43:33,802 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.dir is deprecated. Instead, use mapreduce.output.fileoutputformat.outputdir
2017-10-25 16:43:33,975 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 16:43:33,976 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-25 16:43:33,988 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 16:43:33,998 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-10-25 16:43:34,054 [main] INFO org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - Saved output of task 'attempt__0001_m_000001_1' to file:/tmp/temp-2095141493/tmp144988200/_temporary/0/task__0001_m_000001
2017-10-25 16:43:34,100 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 16:43:34,145 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 16:43:34,151 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(1,335,0,N)
(1,3231,0,N)
(1,448,0,N)
(1,3920,0,N)
(1,378,0,N)
(1,509,0,N)
(1,100,0,N)
(1,1333,0,N)
(1,2272,0,N)
(1,675,0,N)
(1,1144,0,N)
(1,4,0,N)
(1,54,0,N)
(1,623,0,N)
```

Filter records with cancelled indicator as 1 and cancel code as B-

```
grunt> C = filter B by cancelled == 1 AND cancel_code == 'B';
2017-10-25 16:47:07,482 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
grunt> limit2 = LIMIT C 10;
grunt> dump limit2;
2017-10-25 16:47:22,751 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER,LIMIT
2017-10-25 16:47:22,809 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 16:47:22,810 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-25 16:47:22,810 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 16:47:22,811 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInsertter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInsertter]}
2017-10-25 16:47:22,826 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 16:47:22,892 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 16:47:22,892 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-10-25 16:47:41,589 [main] INFO org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - Saved output of task 'attempt__0001_m_000001_1' to file:/tmp/temp-2095141493/tmp1274654546/_temporary/0/task__0001_m_000001
2017-10-25 16:47:41,628 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 16:47:41,681 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 16:47:41,681 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(10,2347,1,B)
(10,2890,1,B)
(10,2117,1,B)
(10,2920,1,B)
(10,7097,1,B)
(10,6588,1,B)
(10,4027,1,B)
(10,5440,1,B)
(10,982,1,B)
(10,5608,1,B)
```

Group each record with month-

```
grunt> D = group C by month;
grunt> limit3 = LIMIT D 1;
grunt> dump limit3;█
```

```
2017-10-25 16:50:00,031 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 16:50:00,135 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 16:50:00,135 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(10,{(10,982,1,B),(10,5440,1,B),(10,4027,1,B),(10,6588,1,B),(10,7097,1,B),(10,2920,1,B),(10,2117,1,B),(10,2890,1,B),(10,2347,1,B),(10,5025,1,B),(10,5608,1,B),(10,1787,1,B),(10,43,1,B),(10,4715,1,B),(10,3179,1,B),(10,3031,1,B),(10,3010,1,B)})}
```

Find the number of months in each group-

```
grunt> E = foreach D generate group, COUNT(c.cancelled);
grunt> limit4 = LIMIT E 10;
grunt> dump limit4;
```

```
2017-10-25 16:53:25,050 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 16:53:25,094 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 16:53:25,094 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(10,17)
(11,40)
(12,250)
```

Arrange the above output in descending order and take the topmost to find the maximum count of cancellations-

```
grunt> F= order E by $1 DESC;
grunt> Result = limit F 1;
grunt> dump Result;
```

FINAL O/P-

```
2017-10-25 16:55:59,474 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 16:55:59,488 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 16:55:59,488 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(12,250)
```

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-

```
grunt> REGISTER '/home/acadgild/airline_usecase/piggybank-0.15.0.jar'
2017-10-25 15:01:25,749 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 15:01:25,750 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 15:01:25,751 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> A = load '/home/acadgild/airline_usecase/DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
2017-10-25 15:19:15,250 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 15:19:15,251 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 15:19:15,251 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> limit1 = LIMIT A 10;
grunt> dump limit1;
```

Contents of file-

```
(0,2008,1,3,4,2003.0,1955,2211.0,2225,WN,335,N712SW,128.0,150.0,116.0,-14.0,8.0,IAD,TPA,810,4.0,8.0,0,N,0,,,,)
(1,2008,1,3,4,754.0,735,1002.0,1000,WN,3231,N772SW,128.0,145.0,113.0,2.0,19.0,IAD,TPA,810,5.0,10.0,0,N,0,,,,)
(2,2008,1,3,4,628.0,620,804.0,750,WN,448,N428WN,96.0,90.0,76.0,14.0,8.0,IND,BWI,515,3.0,17.0,0,N,0,,,,)
(4,2008,1,3,4,1829.0,1755,1959.0,1925,WN,3920,N464WN,90.0,90.0,77.0,34.0,34.0,IND,BWI,515,3.0,10.0,0,N,0,2.0,0.0,0.0,0.0,32.0)
(5,2008,1,3,4,1940.0,1915,2121.0,2110,WN,378,N726SW,101.0,115.0,87.0,11.0,25.0,IND,JAX,688,4.0,10.0,0,N,0,,,,)
(6,2008,1,3,4,1937.0,1830,2037.0,1940,WN,509,N763SW,240.0,250.0,230.0,57.0,67.0,IND,LAS,1591,3.0,7.0,0,N,0,10.0,0.0,0.0,0.0,47.0)
(10,2008,1,3,4,706.0,700,916.0,915,WN,100,N690SW,130.0,135.0,106.0,1.0,6.0,IND,MCO,828,5.0,19.0,0,N,0,,,,)
(11,2008,1,3,4,1644.0,1510,1845.0,1725,WN,1333,N334SW,121.0,135.0,107.0,80.0,94.0,IND,MCO,828,6.0,8.0,0,N,0,8.0,0.0,0.0,0.0,72.0)
(15,2008,1,3,4,1029.0,1020,1021.0,1010,WN,2272,N263WN,52.0,50.0,37.0,11.0,9.0,IND,MDW,162,6.0,9.0,0,N,0,,,,)
(16,2008,1,3,4,1452.0,1425,1640.0,1625,WN,675,N286WN,228.0,240.0,213.0,15.0,27.0,IND,PHX,1489,7.0,8.0,0,N,0,3.0,0.0,0.0,0.0,12.0)
grunt>
```

Extract records for departure delay and origin-

```
grunt> B1 = foreach A generate (int)$16 as dep_delay, (chararray)$17 as origin;
grunt> limit1 = LIMIT B1 10;
grunt> dump limit1;
2017-10-25 17:21:44,917 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: LIMIT
2017-10-25 17:21:44,978 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 17:21:44,979 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-25 17:21:44,979 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 17:21:44,979 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-10-25 17:21:44,980 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2017-10-25 17:21:45,035 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 17:21:45,035 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-10-25 17:21:45,042 [main] INFO org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - Saved output of task 'attempt__0001_m_000001_1' to file:/tmp/temp-2095141493/tmp-473683927/ temporary/0/task__0001_m_000001
2017-10-25 17:21:45,053 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 17:21:45,070 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 17:21:45,071 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(8,IAD)
(19,IAD)
(8,IND)
(34,IND)
(25,IND)
(67,IND)
(6,IND)
(94,IND)
(9,IND)
(27,IND)
```

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

```
grunt> C1 = filter B1 by (dep_delay is not null) AND (origin is not null);
grunt> limit2 = LIMIT C 10;
grunt> dump limit2;
2017-10-25 17:24:13,187 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: FILTER,LIMIT
2017-10-25 17:24:13,245 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 17:24:13,247 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-25 17:24:13,247 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 17:24:13,248 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEachFlatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2017-10-25 17:24:13,279 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 17:24:13,279 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-10-25 17:24:22,779 [main] INFO org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - Saved output of task 'attempt__0001_m_000001_1' to file:/tmp/temp-2095141493/tmp-564188260/ temporary/0/task__0001_m_000001
2017-10-25 17:24:22,790 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 17:24:22,807 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 17:24:22,807 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(10,2347,1,B)
(10,2890,1,B)
(10,2117,1,B)
(10,2920,1,B)
(10,7097,1,B)
(10,6588,1,B)
(10,4027,1,B)
(10,5440,1,B)
(10,982,1,B)
(10,5608,1,B)
```

Group above result by origin-

```
grunt> D1 = group C1 by origin;
grunt> limit3 = LIMIT D1 1;
grunt> dump limit3;
```

```
2017-10-25 17:26:56,809 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 17:26:56,841 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 17:26:56,841 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(ABE,{(18,ABE),(29,ABE),(11,ABE),(30,ABE),(16,ABE),(66,ABE),(20,ABE),(24,ABE),(33,ABE),(10,ABE),(27,ABE),(39,ABE),(8,ABE),(43,ABE),(10,ABE),(142,ABE),(35,ABE),(48,ABE),(6,ABE),(18,ABE),(101,ABE),(147,ABE),(129,ABE),(342,ABE),(41,ABE),(39,ABE),(22,ABE),(17,ABE),(26,ABE),(9,ABE),(24,ABE),(27,ABE),(142,ABE),(17,ABE),(108,ABE),(14,ABE),(130,ABE),(31,ABE),(32,ABE),(19,ABE),(9,ABE),(47,ABE),(11,ABE),(46,ABE),(83,ABE),(67,ABE),(13,ABE),(7,ABE),(22,ABE),(78,ABE),(25,ABE),(36,ABE),(136,ABE),(81,ABE),(9,ABE),(8,ABE),(21,ABE),(31,ABE),(143,ABE),(15,ABE),(27,ABE),(64,ABE),(28,ABE),(96,ABE),(48,ABE),(13,ABE),(10,ABE),(45,ABE),(22,ABE),(44,ABE),(16,ABE),(29,ABE),(105,ABE),(204,ABE),(11,ABE),(30,ABE),(26,ABE),(20,ABE),(92,ABE),(125,ABE),(19,ABE),(26,ABE),(145,ABE),(42,ABE),(13,ABE),(13,ABE),(29,ABE),(6,ABE),(10,ABE),(10,ABE),(78,ABE),(48,ABE),(6,ABE),(224,ABE),(43,ABE),(26,ABE),(14,ABE),(163,ABE),(12,ABE),(15,ABE),(21,ABE),(41,ABE),(15,ABE),(16,ABE),(381,ABE),(10,ABE),(6,ABE),(77,ABE),(11,ABE),(98,ABE),(17,ABE),(32,ABE),(120,ABE),(17,ABE),(34,ABE),(9,ABE),(19,ABE),(10,ABE),(23,ABE),(30,ABE),(27,ABE),(10,ABE),(116,ABE),(77,ABE),(17,ABE),(16,ABE),(36,ABE),(12,ABE),(17,ABE),(30,ABE),(10,ABE),(39,ABE),(147,ABE),(144,ABE),(147,ABE),(103,ABE),(287,ABE),(54,ABE),(12,ABE),(154,ABE),(12,ABE),(8,ABE),(55,ABE),(15,ABE),(36,ABE),(6,ABE),(107,ABE),(35,ABE),(408,ABE),(14,ABE),(35,ABE),(66,ABE),(15,ABE),(93,ABE),(23,ABE),(29,ABE),(9,ABE),(10,ABE),(29,ABE),(22,ABE),(19,ABE),(24,ABE),(211,ABE),(51,ABE),(27,ABE),(197,ABE),(27,ABE),(20,ABE),(28,ABE),(79,ABE),(22,ABE),(7,ABE),(16,ABE),(208,ABE),(14,ABE),(144,ABE),(8,ABE),(6,ABE),(73,ABE),(7,ABE),(36,ABE),(114,ABE),(8,ABE),(140,ABE),(78,ABE),(15,ABE),(15,ABE),(35,ABE),(19,ABE),(17,ABE),(428,ABE),(21,ABE),(101,ABE),(56,ABE),(9,ABE),(94,ABE),(27,ABE),(15,ABE),(61,ABE),(35,ABE),(107,ABE),(6,ABE),(253,ABE),(32,ABE),(162,ABE),(7,ABE),(9,ABE),(13,ABE),(38,ABE),(12,ABE),(55,ABE),(72,ABE),(17,ABE),(9,ABE),(22,ABE),(7,ABE),(33,ABE),(78,ABE),(52,ABE),(73,ABE),(66,ABE),(17,ABE),(207,ABE),(45,ABE),(145,ABE),(13,ABE),(125,ABE),(11,ABE),(6,ABE),(19,ABE),(37,ABE),(102,ABE),(89,ABE),(28,ABE),(115,ABE),(141,ABE),(83,ABE),(23,ABE),(28,ABE),(11,ABE),(180,ABE),(27,ABE),(31,ABE),(33,ABE),(10,ABE),(19,ABE),(11,ABE),(8,ABE),(9,ABE),(11,ABE),(196,ABE),(36,ABE),(11,ABE),(8,ABE),(40,ABE),(17,ABE),(44,ABE),(77,ABE),(76,ABE),(60,ABE),(46,ABE),(86,ABE),(75,ABE),(36,ABE),(24,ABE),(194,ABE),(67,ABE),(7,ABE),(12,ABE),(34,ABE),(12,ABE),(49,ABE),(11,ABE),(79,ABE),(8,ABE),(160,ABE),(8,ABE),(25,ABE),(17,ABE),(12,ABE),(8,ABE),(57,ABE),(7,ABE),(102,ABE),(23,ABE),(50,ABE),(25,ABE),(29,ABE),(77,ABE),(44,ABE),(159,ABE),(148,ABE),(30,ABE),(13,ABE),(6,ABE),(147,ABE),(80,ABE),(10,ABE),(19,ABE),(22,ABE),(60,ABE),(94,ABE),(27,ABE),(235,ABE),(28,ABE),(20,ABE),(33,ABE),(274,ABE),(9,ABE),(270,ABE),(80,ABE),(32,ABE),(13,ABE),(286,ABE),(16,ABE),(20,ABE),(580,ABE),(31,ABE),(107,ABE),(7,ABE),(38,ABE),(13,ABE),(41,ABE),(99,ABE),(44,ABE),(11,ABE),(11,ABE),(16,ABE),(67,ABE),(57,ABE),(13,ABE),(64,ABE),(9,ABE),(10,ABE),(7,ABE),(43,ABE),(17,ABE),(251,ABE),(12,ABE),(23,ABE),(40,ABE),(17,ABE),(114,ABE),(147,ABE),(130,ABE),(75,ABE),(50,ABE),(15,ABE),(15,ABE),(105,ABE),(50,ABE),(137,ABE),(13,ABE),(12,ABE),(71,ABE),(19,ABE),(12,ABE),(109,ABE),(18,ABE),(51,ABE),(14,ABE),(8,ABE),(58,ABE),(10,ABE),(15,ABE),(65,ABE),(12,ABE),(22,ABE),(130,ABE),(26,ABE),(31,ABE),(132,ABE),(25,ABE),(124,ABE),(86,ABE),(34,ABE),(22,ABE),(100,ABE),(7,ABE),(154,ABE),(112,ABE),(74,ABE),(128,ABE),(92,ABE),(129,ABE),(54,ABE),(40,ABE),(250,ABE),(243,ABE),(15,ABE),(15,ABE),(25,ABE),(179,ABE),(23,ABE),(50,ABE),(25,ABE),(56,ABE),(19,ABE),(36,ABE),(26,ABE),(49,ABE),(9,ABE),(21,ABE),(15,ABE),(121,ABE),(8,ABE),(6,ABE),(15,ABE),(15,ABE),(51,ABE),(206,ABE),(43,ABE),(34,ABE),(7,ABE),(164,ABE),(129,ABE),(44,ABE),(32,ABE),(1021,ABE),(16,ABE),(58,ABE),(7,ABE),(17,ABE),(14,ABE),(24,ABE),(29,ABE),(9,ABE),(9,ABE),(70,ABE),(8,ABE),(92,ABE),(32,ABE),(77,ABE),(455,ABE),(7,ABE),(98,ABE),(104,ABE),(64,ABE),(103,ABE),(108,ABE),(131,ABE),(15,ABE),(7,ABE),(13,ABE),(89,ABE),(215,ABE),(35,ABE),(103,ABE),(6,ABE),(18,ABE),(27,ABE),(25,ABE),(9,ABE),(200,ABE),(145,ABE),(53,ABE),(107,ABE),(91,ABE),(14,ABE),(139,ABE),(34,ABE),(8,ABE),(92,ABE),(14,ABE),(61,ABE),(14,ABE),(8,ABE),(67,ABE),(70,ABE),(27,ABE),(6,ABE),(27,ABE),(11,ABE),(19,ABE)
```

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

```
grunt> E1 = foreach D1 generate group, AVG(C1.dep_delay);
grunt> limit4 = LIMIT E1 10;
grunt> dump limit4;
```

```
2017-10-25 17:28:58,540 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 17:28:58,576 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 17:28:58,576 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(ABE,59.46153846153846)
(ABI,66.46938775510205)
(ABQ,36.35825370058085)
(ABY,49.4486301369863)
(ACK,66.47448979591837)
(ACT,44.222857142857144)
(ACV,58.761022927689595)
(ACY,79.3103448275862)
(ADK,30.50943396226415)
(ADQ,33.8235294117647)
grunt>
```

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

```
Quick connect...
2. VMWare(AG)
3. VMWare(AG)

grunt> Result = order E1 by $1 DESC;
> grunt> Top_ten = limit Result 10;
grunt> dump Top_ten;
```

```
2017-10-25 17:32:01,464 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 17:32:01,486 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 17:32:01,486 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(CMX,116.1470588235294)
(PLN,93.76190476190476)
(SPI,83.84873949579831)
(ALO,82.2258064516129)
(MQT,79.55665024630542)
(ACY,79.3103448275862)
(MOT,78.66165413533835)
(HHH,76.53005464480874)
(EGE,74.12891986062718)
(BGM,73.15533980582525)
grunt>
```

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

```
grunt> Lookup = load '/home/acadgild/airline_usecase/airports.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage('','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
2017-10-25 17:33:39,860 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 17:33:39,861 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 17:33:39,861 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> Lookup1 = foreach Lookup generate (chararray)$0 as origin, (chararray)$2 as city, (chararray)$4 as country;
grunt> limit6 = LIMIT Lookup1 5;
grunt> dump limit6;
2017-10-25 17:34:59,585 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: LIMIT
2017-10-25 17:34:59,630 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 17:34:59,631 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-25 17:34:59,631 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 17:34:59,631 [main] INFO org.apache.pig.data.SchemaTupleBackend - Key [pig.schematuple] was not set... will not generate code.
2017-10-25 17:34:59,632 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnMapKeyPrune, ConstantCalculator, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInsertter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdownOptimizer, PushDownForEach, Flatten, PushUpFilter, SplitFilter, StreamTypeCastInsertter]}
2017-10-25 17:34:59,661 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 17:34:59,661 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-10-25 17:34:59,664 [main] INFO org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - Saved output of task 'attempt_0001_m_000001_1' to file:/tmp/temp-2095141493/tmp1490304732/ temporary0/task_0001_m_000001
2017-10-25 17:34:59,674 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 17:34:59,698 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 17:34:59,698 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(00M,Bay Springs,USA)
(00R,Livingston,USA)
(00V,Colorado Springs,USA)
(01G,Perry,USA)
(01J,Hilliard,USA)
grunt>
```

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



```
grunt> Joined = join Lookup1 by origin, Top_ten by $0;
grunt> limit7 = LIMIT Joined 10;
grunt> dump limit7;
```

```
2017-10-25 17:36:48,609 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 17:36:48,630 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 17:36:48,630 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(ACY,Atlantic City,USA,ACY,79.3103448275862)
(ALO,Waterloo,USA,ALO,82.2258064516129)
(BGM,Binghamton,USA,BGM,73.15533980582525)
(CMX,Hancock,USA,CMX,116.1470588235294)
(EGE,Eagle,USA,EGE,74.12891986062718)
(HHH,NA,USA,HHH,76.53005464480874)
(MOT,Minot,USA,MOT,78.66165413533835)
(MQT,NA,USA,MQT,79.55665024630542)
(PLN,Pellston,USA,PLN,93.76190476190476)
(SPI,Springfield,USA,SPI,83.84873949579831)
```

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

```
grunt> Final = foreach Joined generate $0,$1,$2,$4;
grunt> limit8 = LIMIT Final 10;
grunt> dump limit8;
2017-10-25 17:38:08,671 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: HASH_JOIN,GROUP_BY,ORDER_BY,FILTER,LIMIT
2017-10-25 17:38:08,703 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 17:38:08,704 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-25 17:38:08,704 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.limit
```

```
rs.max
2017-10-25 17:38:36,239 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 17:38:36,271 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 17:38:36,271 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(ACY,Atlantic City,USA,79.3103448275862)
(ALO,Waterloo,USA,82.2258064516129)
(BGM,Binghamton,USA,73.15533980582525)
(CMX,Hancock,USA,116.1470588235294)
(EGE,Eagle,USA,74.12891986062718)
(HHH,NA,USA,76.53005464480874)
(MOT,Minot,USA,78.66165413533835)
(MQT,NA,USA,79.55665024630542)
(PLN,Pellston,USA,93.76190476190476)
(SPI,Springfield,USA,83.84873949579831)
```

ORDER the result by Average in descending manner-

```
grunt> Final_Result = ORDER Final by $3 DESC;
grunt> dump Final_Result;
```

Final O/P

```
2017-10-25 17:45:51,930 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-25 17:45:51,957 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-25 17:45:51,957 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(CMX,Hancock,USA,116.1470588235294)
(PLN,Pellston,USA,93.76190476190476)
(SPI,Springfield,USA,83.84873949579831)
(ALO,Waterloo,USA,82.2258064516129)
(MQT,NA,USA,79.55665024630542)
(ACY,Atlantic City,USA,79.3103448275862)
(MOT,Minot,USA,78.66165413533835)
(HHH,NA,USA,76.53005464480874)
(EGE,Eagle,USA,74.12891986062718)
(BGM,Binghamton,USA,73.15533980582525)
grunt>
```

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-

```
grunt> REGISTER '/home/acadgild/airline_usecase/piggybank-0.15.0.jar'
2017-10-25 15:01:25,749 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 15:01:25,750 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 15:01:25,751 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> A = load '/home/acadgild/airline_usecase/DelayedFlights.csv' USING org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');
2017-10-25 15:19:15,250 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-25 15:19:15,251 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-25 15:19:15,251 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
grunt> limit1 = LIMIT A 10;
grunt> dump limit1;
```

Contents of file-

```
(0,2008,1,3,4,2003.0,1955,2211.0,2225,WN,335,N712SW,128.0,150.0,116.0,-14.0,8.0,IAD,TPA,810,4.0,8.0,0,N,0,,,,)
(1,2008,1,3,4,754.0,735,1002.0,1000,WN,3231,N772SW,128.0,145.0,113.0,2.0,19.0,IAD,TPA,810,5.0,10.0,0,N,0,,,,)
(2,2008,1,3,4,628.0,620,804.0,750,WN,448,N428WN,96.0,90.0,76.0,14.0,8.0,IND,BWI,515,3.0,17.0,0,N,0,,,,)
(4,2008,1,3,4,1829.0,1755,1959.0,1925,WN,3920,N464WN,90.0,90.0,77.0,34.0,34.0,IND,BWI,515,3.0,10.0,0,N,0,2.0,0.0,0.0,0.0,32.0)
(5,2008,1,3,4,1940.0,1915,2121.0,2110,WN,378,N726SW,101.0,115.0,87.0,11.0,25.0,IND,JAX,688,4.0,10.0,0,N,0,,,,)
(6,2008,1,3,4,1937.0,1830,2037.0,1940,WN,509,N763SW,240.0,250.0,230.0,57.0,67.0,IND,LAS,1591,3.0,7.0,0,N,0,10.0,0.0,0.0,0.0,47.0)
(10,2008,1,3,4,706.0,700,916.0,915,WN,100,N690SW,130.0,135.0,106.0,1.0,6.0,IND,MCO,828,5.0,19.0,0,N,0,,,,)
(11,2008,1,3,4,1644.0,1510,1845.0,1725,WN,1333,N334SW,121.0,135.0,107.0,80.0,94.0,IND,MCO,828,6.0,8.0,0,N,0,8.0,0.0,0.0,0.0,72.0)
(15,2008,1,3,4,1029.0,1020,1021.0,1010,WN,2272,N263WN,52.0,50.0,37.0,11.0,9.0,IND,MDW,162,6.0,9.0,0,N,0,,,,)
(16,2008,1,3,4,1452.0,1425,1640.0,1625,WN,675,N286WN,228.0,240.0,213.0,15.0,27.0,IND,PHX,1489,7.0,8.0,0,N,0,3.0,0.0,0.0,0.0,12.0)
grunt>
```

Extract origin, destination and diversion from the input file-

```
grunt> B = FOREACH A GENERATE (chararray)$17 as origin, (chararray)$18 as dest, (int)$24 as diversion;
grunt> limit1 = LIMIT B 10;
grunt> dump limit1;
2017-10-26 16:40:45,825 [main] WARN org.apache.hadoop.util.NativeCodeLoader - Unable to load native-hadoop library for your platform...
here applicable
2017-10-26 16:40:46,323 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features used in the script: LIMIT
2017-10-26 16:40:46,568 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-26 16:40:46,571 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-26 16:40:47,269 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimizer - {RULES_ENABLED=[AddForEach, ColumnProject, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach, PartitionFilterOptimizer, PredicatePushdown, Flatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]}
2017-10-26 16:40:49,521 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce.job.counters.limit is deprecated. Instead, use mapreduce.job.counters.max
2017-10-26 16:40:49,522 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapred.output.dir is deprecated. Instead, use mapreduce.output.local.outputdir
2017-10-26 16:40:50,560 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum
2017-10-26 16:40:50,562 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS
2017-10-26 16:40:50,662 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-26 16:40:50,733 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
2017-10-26 16:40:51,202 [main] INFO org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter - Saved output of task 'attempt__0001_m_0000001_0030259/tmp-312327559/_temporary/0/task__0001_m_0000001'
2017-10-26 16:40:51,385 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-26 16:40:51,581 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-26 16:40:51,583 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(IAD,TPA,0)
(IAD,TPA,0)
(IND,BWI,0)
(IND,BWI,0)
(IND,JAX,0)
(IND,LAS,0)
(IND,MCO,0)
(IND,MCO,0)
(IND,MDW,0)
(IND,PHX,0)
grunt>
```

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

```
grunt> C = FILTER B BY (origin is not null) AND (dest is not null) AND (diversion == 1);
2017-10-26 16:41:45,450 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce
rs.max
grunt> limit2 = LIMIT C 10;
grunt> dump limit2;
2017-10-26 16:42:30,416 [main] INFO org.apache.pig.tools.pigstats.ScriptState - Pig features us
2017-10-26 16:42:30,546 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes
2017-10-26 16:42:30,547 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.defau
2017-10-26 16:42:30,548 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend
2017-10-26 16:42:30,549 [main] INFO org.apache.pig.newplan.logical.optimizer.LogicalPlanOptimiz
r, GroupByConstParallelSetter, LimitOptimizer, LoadTypeCastInserter, MergeFilter, MergeForEach,
Flatten, PushUpFilter, SplitFilter, StreamTypeCastInserter]]
2017-10-26 16:42:30,626 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - mapreduce
rs.max
2017-10-26 16:42:30,772 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Tot
2017-10-26 16:42:30,773 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUt
2017-10-26 16:42:31,568 [main] INFO org.apache.hadoop.mapreduce.lib.output.FileOutputCommitte
9030259/tmp-522638239/_temporary/0/task__0001_m_000001
2017-10-26 16:42:31,680 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend
2017-10-26 16:42:31,817 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Tot
2017-10-26 16:42:31,818 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUt
(SAN,SMF,1)
(SFO,SAN,1)
(BOI,RNO,1)
(BUR,SMF,1)
(LAS,RNO,1)
(LAS,RNO,1)
(OAK,RNO,1)
(PDX,RNO,1)
(SAN,RNO,1)
(SJC,LAX,1)
```

GROUP above result by origin and destination-

```
grunt> D = GROUP C by (origin,dest);
grunt> limit3 = LIMIT D 1;
grunt> dump limit3;

2017-10-26 16:45:49,503 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-26 16:45:49,619 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-26 16:45:49,620 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
((ABE,ORD),{(ABE,ORD,1),(ABE,ORD,1),(ABE,ORD,1),(ABE,ORD,1),(ABE,ORD,1)})
```

From the grouped record extract COUNT of diversions-

```
grunt> E = FOREACH D generate group, COUNT(C.diversion);
grunt> limit4 = LIMIT E 10;
grunt> dump limit4;

2017-10-26 16:53:05,267 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-26 16:53:05,339 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-26 16:53:05,339 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
((ABE,ORD),5)
((ABQ,ATL),1)
((ABQ,DAL),6)
((ABQ,DEN),1)
((ABQ,DFW),3)
((ABQ,HOU),2)
((ABQ,IAH),2)
((ABQ,LAS),3)
((ABQ,MDW),1)
((ABQ,ONT),1)
```

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



```
grunt> F = ORDER E BY $1 DESC;
grunt> Result = limit F 10;
grunt> dump Result;
```

Final O/P-

```
2017-10-26 16:58:22,022 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-10-26 16:58:22,072 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-10-26 16:58:22,072 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
((ORD,LGA),39)
((DAL,HOU),35)
((DFW,LGA),33)
((ATL,LGA),32)
((ORD,SNA),31)
((SLC,SUN),31)
((MIA,LGA),31)
((BUR,JFK),29)
((HRL,HOU),28)
((BUR,DFW),25)
grunt>
```