Index:

PIG

Ravella Abhinav
CB.EN.U4CSE19453

Load datasets:

Airlines:

Query:

airline = LOAD '/home/hduser/Downloads/airports_mod.dat' using PigStorage(',') as (airlineID:charArray,airline_name:charArray, airline_alias:charArray, airline_iata:charArray, airline_icao:charArray,callsign:charArray,territory:charArray, active:charArray);

Output:

```
grunt> airline = LOAD '/home/hduser/Downloads/airports_mod.dat' using PigStorage(',') as (airlineID:charArray,airline_name:charArray, airline_alias:char
Array, airline_iata:charArray, airline_icao:charArray,callsign:charArray,territory:charArray, active:charArray);
grunt> describe airline;
airline: {airlineID: chararray,airline_name: chararray,airline_alias: chararray,airline_iata: chararray,airline_icao: chararray,callsign: chararray,terr
itory: chararray,active: chararray}
grunt> describe airline;
airline: {airlineID: chararray,airline_name: chararray,airline_alias: chararray,airline_iata: chararray,airline_icao: chararray,callsign: chararray,terr
itory: chararray,active: chararray}
grunt> clear;
```

```
2022-04-05 06:36:51,148 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-05 06:36:51,179 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly$Map - Aliases being processed per job phase (AliasName[line,offset]): M: routes[1,9] C:  R:
2022-04-05 06:36:51,179 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-04-05 06:36:51,181 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-04-05 06:36:51,181 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-04-05 06:36:51,183 [main] INFO  org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2022-04-05 06:36:51,192 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-04-05 06:36:51,223 [main] INFO  org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen) of size 699400192 to monitor. collectionUsageThreshold = 489580128, usageThreshold = 4895
80128
2022-04-05 06:36:51,225 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-05 06:36:51,226 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly$Map - Aliases being processed per job phase (AliasName[line,offset]): M: routes[1,9] C:  R:
2022-04-05 06:36:51,226 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-04-05 06:36:51,231 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-04-05 06:36:51,231 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-04-05 06:36:51,239 [main] INFO  org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2022-04-05 06:36:51,267 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-04-05 06:36:51,324 [main] INFO  org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen) of size 699400192 to monitor. collectionUsageThreshold = 489580128, usageThreshold = 4895
80128
2022-04-05 06:36:51,326 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-05 06:36:51,332 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly$Map - Aliases being processed per job phase (AliasName[line,offset]): M: routes[1,9] C:  R:
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
---------------
| routes     | route_iata:chararray    | route_airid:chararray    | route_source_iata:chararray    | route_source_airid:chararray    | route_codeshare:chararray    | route_stops:chararray    | route_equi
p:chararray    |
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
---------------
|            | AF                      | 137                      | JFK                            | 3797                            | MSY                          | 3861                     | Y
               |
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
---------------

grunt>
```

Final Airlines:

Query:

final = LOAD '/home/hduser/Downloads/Final_airlines' using PigStorage(',') as (airlineID:charArray,airline_name:charArray, airline_alias:charArray, airline_iata:charArray, airline_icao:charArray,callsign:charArray,territory:charArray, active:charArray);

Output:

```
grunt> final = LOAD '/home/hduser/Downloads/Final_airlines' using PigStorage(',') as (airlineID:charArray,airline_name:charArray, airline_alias:charArra
y, airline_iata:charArray, airline_icao:charArray,callsign:charArray,territory:charArray, active:charArray);
grunt> describe final;
final: {airlineID: chararray,airline_name: chararray,airline_alias: chararray,airline_iata: chararray,airline_icao: chararray,callsign: chararray,territ
ory: chararray,active: chararray}
```

```
2022-04-05 06:37:48,929 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-05 06:37:48,930 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly$Map - Aliases being processed per job phase (AliasName[line,offset]): M: finalair[2,11] C:  R:

2022-04-05 06:37:48,930 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-04-05 06:37:48,937 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-04-05 06:37:48,955 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-04-05 06:37:48,958 [main] INFO  org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2022-04-05 06:37:48,958 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-04-05 06:37:49,025 [main] INFO  org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen) of size 699400192 to monitor. collectionUsageThreshold = 489580128, usageThreshold = 4895
80128
2022-04-05 06:37:49,026 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-05 06:37:49,029 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly$Map - Aliases being processed per job phase (AliasName[line,offset]): M: finalair[2,11] C:  R:

-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
---------------------
| finalair     | airlineID:chararray    | airline_name:chararray    | airline_alias:chararray    | airline_iata:chararray    | airline_icao:chararray    | callsign:chararray    | territory:chararray    |
 active:chararray    |
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
---------------------
|              | 3434                   | Maxjet Airways            | \N                        | MY                        | MXJ                       | MAX-JET               | United States          |
 N                   |
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
---------------------

grunt>
```

Routes:

Query:

routes = LOAD '/home/hduser/Downloads/routes' using PigStorage(',') as (route_iata:charArray, route_airid:charArray, route_source_iata:charArray, route_source_airid:charArray, route_codeshare:charArray, route_stops:charArray, route_equip:charArray);

Output:

```
grunt> routes = LOAD '/home/hduser/Downloads/routes' using PigStorage(',') as (route_iata:charArray,route_airid:charArray, route_source_iata:charArray,
route_source_airid:charArray, route_codeshare:charArray,route_stops:charArray,route_equip:charArray);
grunt> describe routes;
routes: {route_iata: chararray,route_airid: chararray,route_source_iata: chararray,route_source_airid: chararray,route_codeshare: chararray,route_stops:
 chararray,route_equip: chararray}
grunt>
```

```
2022-04-05 06:34:49,526 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-05 06:34:49,527 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly$Map - Aliases being processed per job phase (AliasName[line,offset]): M: routes[1,9] C:  R:
2022-04-05 06:34:49,527 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-04-05 06:34:49,534 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-04-05 06:34:49,534 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-04-05 06:34:49,546 [main] INFO  org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2022-04-05 06:34:49,546 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-04-05 06:34:49,621 [main] INFO  org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen) of size 699400192 to monitor. collectionUsageThreshold = 489580128, usageThreshold = 4895
80128
2022-04-05 06:34:49,621 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-05 06:34:49,623 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly$Map - Aliases being processed per job phase (AliasName[line,offset]): M: routes[1,9] C:  R:
2022-04-05 06:34:49,623 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic? false
2022-04-05 06:34:49,636 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-04-05 06:34:49,636 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-04-05 06:34:49,647 [main] INFO  org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2022-04-05 06:34:49,647 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-04-05 06:34:49,698 [main] INFO  org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen) of size 699400192 to monitor. collectionUsageThreshold = 489580128, usageThreshold = 4895
80128
2022-04-05 06:34:49,701 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-05 06:34:49,702 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnly$Map - Aliases being processed per job phase (AliasName[line,offset]): M: routes[1,9] C:  R:
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
-------------
| routes     | route_iata:chararray   | route_airid:chararray   | route_source_iata:chararray    | route_source_airid:chararray    | route_codeshare:chararray    | route_stops:chararray    | route_equip:
chararray    |
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
-------------
|            | AF                     | 137                     | ATL                            | 3682                            | BDL                          | 3825                     | Y
             |
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------
-------------

grunt>
```

Queries:

1. **Find the airlines that are active:**
    - active = filter final by Active == 'Y';
    - limit = LIMIT active 50;
    - dump active;
    - dump limit;

Output:

```
HadoopVersion   PigVersion      UserId  StartedAt       FinishedAt      Features
3.3.1   0.17.0  hduser  2022-04-05 06:39:29     2022-04-05 06:39:40     FILTER

Success!

Job Stats (time in seconds):
JobId   Maps    Reduces MaxMapTime      MinMapTime      AvgMapTime      MedianMapTime   MaxReduceTime   MinReduceTime   AvgReduceTime   MedianReducetime        Alias   Feature Outputs
job_local242044014_0001 1       0       n/a     n/a     n/a     n/a     0       0       0       0       active,finalair MAP_ONLY        hdfs://localhost:54310/tmp/temp470170944/tmp1309034107,

Input(s):
Successfully read 6048 records (7432748 bytes) from: "/pig/Final_airlines"

Output(s):
Successfully stored 1161 records (5834878 bytes) in: "hdfs://localhost:54310/tmp/temp470170944/tmp1309034107"

Counters:
Total records written : 1161
Total bytes written : 5834878
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local242044014_0001
```

```
(19337,Transair,,,TTZ,,Canada,Y)
(19350,Comfort Express Virtual Charters Albany,,,EVC,Comfort Express,United States,Y)
(19351,Comfort Express Virtual Charters,,,CEO,,United States,Y)
(19359,FLYJET,,,FYJ,Fast Jet,Poland,Y)
(19361,Snowbird Airlines,,S8,SBD,,Finland,Y)
(19363,Russkie Krylya,,,KRY,,Russia,Y)
(19367,Kharkiv Airlines,,KH,KHK,,Ukraine,Y)
(19433,XAIR USA,,XA,XAU,XAIR,United States,Y)
(19451,Air Costa,,LB,\N,,India,Y)
(19459,Simrik Airlines,,,RMK,,Nepal,Y)
(19473,XPTO,XPTO  ,XP,XPT,XPTO,Portugal,Y)
(19474,Royal Flight,,,DME,,Russia,Y)
(19525,BBN-Airways,BlackBurn,,EGH,BBN,United Kingdom,Y)
(19531,Tomsk-Avia,,,TKS,,Russia,Y)
(19541,Malawian Airlines,,3W,\N,,Malawi,Y)
(19548,Yeti Airlines  ,,,NYT,,Nepal,Y)
(19567,Avilu,Avilu' SA,..,...,,Switzerland,Y)
(19599,Skyline Ulasim Ticaret A.S.,Skyline Ulasim Ticaret A.S.,,KCU,Kocoglu,Turkey,Y)
(19619,Envoy Air,,,ENY,Envoy,United States,Y)
(19651,CARICOM AIRWAYS (BARBADOS) INC.,CARICOM AIRWAYS,,CCB,,Barbados,Y)
(19674,Rainbow Air (RAI),Rainbow Air (RAI),RN,RAB,Rainbow,United States,Y)
(19675,Rainbow Air Canada,Rainbow Air CAN,RY,RAY,Rainbow CAN,Canada,Y)
(19676,Rainbow Air Polynesia,Rainbow Air POL,RX,RPO,Rainbow Air,United States,Y)
(19677,Rainbow Air Euro,Rainbow Air EU,RU,RUE,Rainbow Air,United Kingdom,Y)
(19678,Rainbow Air US,Rainbow Air US,RM,RNY,Rainbow Air,United States,Y)
(19745,Transilvania,,,TNS,,Romania,Y)
(19751,Dobrolet,Добролёт,QD,DOB,DOBROLET,Russia,Y)
(19774,Spike Airlines,Aero Spike,S0,SAL,Spike Air,United States,Y)
(19776,Grand Cru Airlines,,,GCA,,Lithuania,Y)
(19785,Go2Sky,,,RLX,RELAX,Slovakia,Y)
(19803,All Argentina,All Argentina,L1,AL1,,Argentina,Y)
(19804,All America,All America,A2,AL2,,United States,Y)
(19805,All Asia,All Asia,L9,AL3,,China,Y)
(19806,All Africa,All Africa,9A,99F,,South Africa,Y)
(19807,Regionalia México,Regionalia México,N4,J88,,Mexico,Y)
(19808,All Europe,All Europe,N9,N99,,United Kingdom,Y)
(19809,All Spain,All Spain,N7,N77,,Spain,Y)
(19810,Regional Air Iceland,Regional Air Iceland,9N,N78,,Iceland,Y)
(19812,Voestar,Voestar Brasil,8K,K88,,Brazil,Y)
(19813,All Colombia,All Colombia,7O,7KK,,Colombia,Y)
(19814,Regionalia Uruguay,Regionalia Uruguay,2X,2K2,,Uruguay,Y)
(19815,Regionalia Venezuela,Regionalia Venezuela,9X,9XX,,Venezuela,Y)
(19827,Regionalia Chile,Regionalia Chile,9J,CR1,,Chile,Y)
(19828,Vuela Cuba,Vuela Cuba,6C,6CC,,Cuba,Y)
(19830,All Australia,All Australia,88,8K8,,Australia,Y)
(19831,Fly Europa,,ER,RWW,,Spain,Y)
(19834,FlyPortugal,,PO,FPT,FlyPortugal,Portugal,Y)
grunt>
```

2. **Find the names of active airlines in territory ‘Unites States’?**

Query:

- active_airlines = filter finalair by active == 'Y';
- active_airlines_usa = filter active_airlines by territory == 'United States';
- find_active_airline_names_in_usa = foreach active_airlines_usa generate airline_name;
- dump find_active_airline_names_in_usa;

```
2022-04-05 06:43:18,681 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-04-05 06:43:18,689 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-05 06:43:18,725 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-04-05 06:43:18,725 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(135 Airways)
(3D Aviation)
(40-Mile Air)
(Aloha Airlines)
(Alaska Island Air)
(American Airlines)
(Aviation Management Corporation)
(Atlantis Airlines (USA))
(Astro Air International)
(Allegiant Air)
(Airborne Express)
(ABX Air)
(Academy Airlines)
(Air Aurora)
(Advance Leasing Company)
(Aztec Worldwide Airlines)
(Aerodyne)
(Anderson Aviation)
(Alaska Central Express)
(Airway Express)
(Aerial Transit Company)
(Air Guam)
(Air Cargo America)
(Air Cargo Carriers)
(Atlantic Gulf Airlines)
(Atlantic Coast Airlines)
(Aero Flight Service)
(Aspen Helicopters)
(Alpine Air Express)
(Airlift International)
(Air Illinois)
(Ameristar Jet Charter)
(A2 Jet Leasing)
(Amerijet International)
(Alpha Jet International)
(American Flyers)
(Aero Coach Aviation)
(Atlantic Southeast Airlines)
(Air Logistics)
(Air Lincoln)
(America West Airlines)
```

Explanation:

Job is to find all the airlines that are active in the United States which are currently active. So first filtering all the airlines that are active using filter and by keywords and saving them in a table and then querying that table to check for the United States territory using the same keywords and printing the airline names only using foreach and generate keywords.

3. **Get all the alias names of all airlines with airlineID**

Query:

- foreach_data = FOREACH finalair GENERATE airlineID,airline_name,airline_alias;

- dump foreach_data;

Output:

```
(5230,University of North Dakota,\N)
(5231,Universal Airlines,\N)
(5232,Uganda Royal Airways,\N)
(5233,United Aviation Services,\N)
(5234,Ural Airlines,\N)
(5235,Ukraine Transavia,\N)
(5236,United Arab Emirates Air Force,\N)
(5237,Union Africaine des Transports,\N)
(5238,Uganda Air Cargo,\N)
(5239,US Airports Air Charter,\N)
(5240,Ucoaviacion,\N)
(5241,Ues-Avia Aircompany,\N)
(5242,UFS,\N)
(5243,Uganda Airlines,\N)
(5244,Urgemer Canarias,\N)
(5245,Ukrainian Helicopters,\N)
(5246,Ulyanovsk Higher Civil Aviation School,\N)
(5247,Universal Jet Rental de Mexico,\N)
(5248,Universal Jet Aviation,\N)
(5249,UK International Airlines,\N)
(5250,Ukraine Air Alliance,\N)
(5251,UM Airlines,\N)
(5252,Ukraine Air Enterprise,\N)
(5253,United Kingdom Home Office,\N)
(5254,Ukrainian Cargo Airways,\N)
(5255,Ultrair,\N)
(5256,Unitemp-M,\N)
(5257,Uni-Fly,\N)
(5258,Union Flights,\N)
(5259,Universal Jet,\N)
(5260,Unsped Paket Servisi,\N)
(5261,Unifly Servizi Aerei,\N)
(5262,Ukrainian Pilot School,\N)
(5263,United Parcel Service,\N)
(5264,Uraiavia,\N)
(5265,US Airways,\N)
(5266,US Check Airlines,\N)
(5267,USAfrica Airways,\N)
(5268,US Helicopter,\N)
(5269,US Jet,\N)
(5270,US Express,\N)
(5271,UTair Aviation,\N)
(5272,UTAGE,\N)
(5273,Utair South Africa,\N)
(5274,Ukrainian State Air Traffic Service Enterprise,\N)
(5275,Urartu-Air,\N)
(5276,Universal Airways,\N)
(5277,Universal Airlines,\N)
```

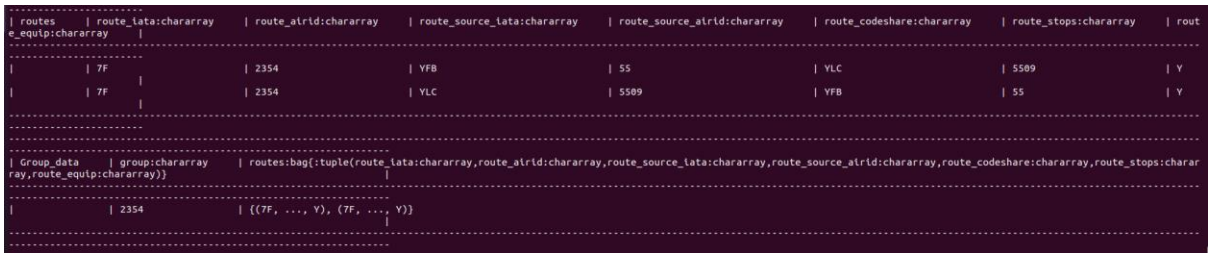
4. Exploring group command:

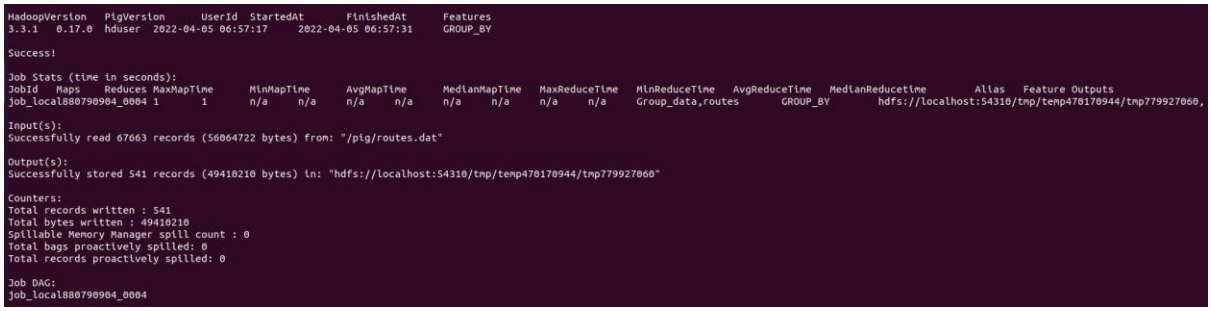
Query:

- Group_data = GROUP routes BY (route_iata,route_airid);
- Limit_group = LIMIT Group_data 10;
- Dump Limit_group;

- group_all = GROUP routes All;

Output:



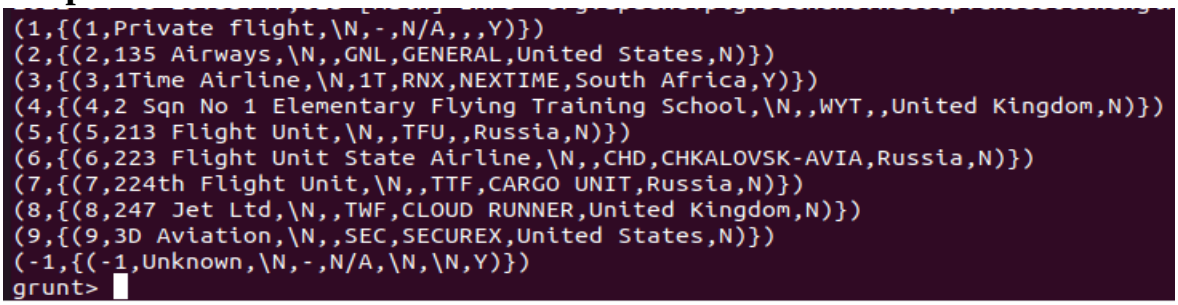


5. **Group airline id by active status;**

Query:

- ae = group finalair by airlineID;
- ans = limit ae 10;
- illustrate as;
- dump as;

Output:

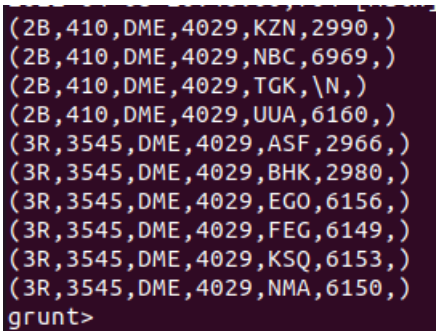


6. **Split relation based on src airport id i.e, equal to DME to get exclusive info.**

Query:

- split routes into ru_dme if route_source_iata == 'DME' , ru_not_dme if route_source_iata != 'DME';
- dump ru_dme;
- dump ru_not_dme;

Output:



Explaination:

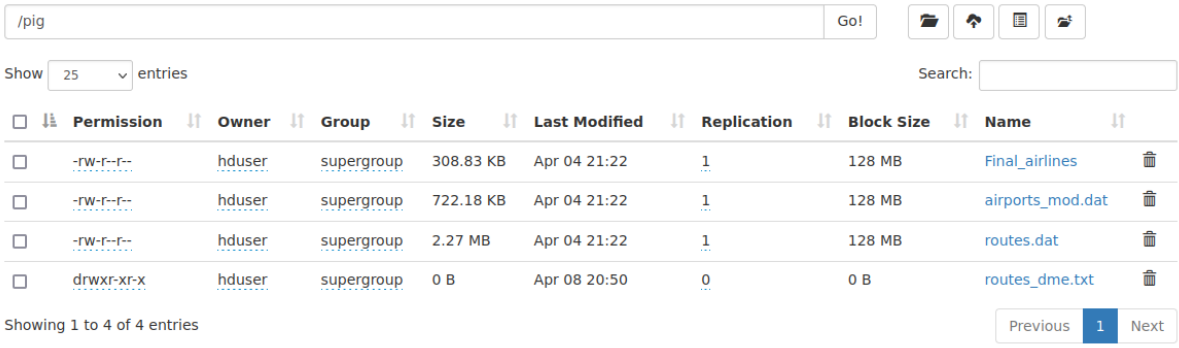
- We are trying to split our relation routes in source airport as DME and not equal to DME using split operator.

7. **Store ru_dme relation into text file separated by delimiter “,”.**

Query:

- store ru_dme into '/pig/routes_dme.txt' using PigStorage(',');

Output:



8. **Find the flights that has most number of stops?**

Query:

- stops = ORDER routes by route_stops DESC;
- dump stops;

Output:

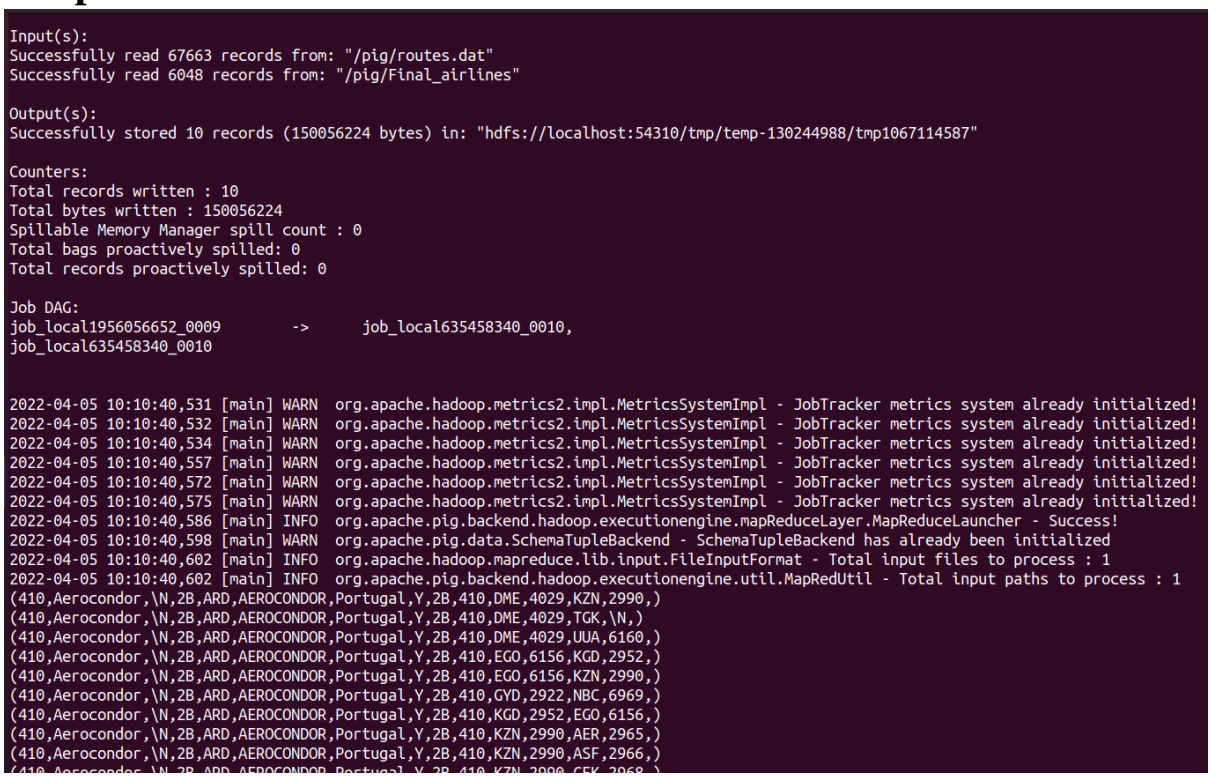
```
(AC,330,YQB,111,YOW,100,Y)
(AC,330,ORD,3830,YOW,100,Y)
(PD,4021,YQM,117,YOW,100,)
(UA,5209,ORD,3830,YOW,100,Y)
(AC,330,IAD,3714,YOW,100,Y)
(PD,4021,YTZ,144,YOW,100,)
(WS,5416,YYT,189,YOW,100,)
(AC,330,YQM,117,YOW,100,Y)
(WS,5416,YEG,49,YOW,100,)
(AC,330,BOS,3448,YOW,100,Y)
(AC,330,FRA,340,YOW,100,)
(US,5265,CLT,3876,YOW,100,Y)
(AA,24,PHL,3752,YOW,100,Y)
(AC,330,YYZ,193,YOW,100,)
(AC,330,YWG,160,YOW,100,Y)
(WS,5416,YHZ,73,YOW,100,)
(MO,502,YZF,196,YOW,100,Y)
(AC,330,YUL,146,YOW,100,)
(WS,5416,YYC,178,YOW,100,)
(AC,330,DCA,3520,YOW,100,Y)
(5T,1623,YFB,55,YOW,100,)
(AC,330,YFC,56,YOW,100,Y)
(UA,5209,EWR,3494,YOW,100,Y)
(AC,330,YXU,174,YOW,100,Y)
(7F,2354,YFB,55,YOW,100,)
(US,5265,PHL,3752,YOW,100,Y)
(4N,341,YZF,196,YOW,100,)
(WS,5416,YYZ,193,YOW,100,)
(WS,5416,YWG,160,YOW,100,)
(AC,330,YHZ,73,YOW,100,)
(DL,2009,DTW,3645,YOW,100,Y)
(AC,330,YYC,178,YOW,100,)
(AA,24,CLT,3876,YOW,100,Y)
(UA,5209,BOS,3448,YOW,100,Y)
(UA,5209,IAD,3714,YOW,100,Y)
(LH,3320,LHR,507,YOW,100,Y)
(UA,5209,LGA,3697,YOW,100,Y)
(AC,330,YYT,189,YOW,100,Y)
(PD,4021,YHZ,73,YOW,100,)
(AC,330,YEG,49,YOW,100,)
(GL,921,NAQ,5446,THU,10,)
(GL,921,SVR,\N,THU,10,)
(CG,1308,LAE,4,GKA,1,)
(PX,328,POM,5,GKA,1,)
(CG,1308,POM,5,GKA,1,)
(CG,1308,MAG,2,GKA,1,)
(CG,1308,HGU,3,GKA,1,)
grunt>
```

9. JOIN Command:

Query:

- right_join = JOIN finalair BY airline_iata RIGHT OUTER, routes BY route_iata;
- LIM = LIMIT right_join 10;
- dump LIM;

Output:



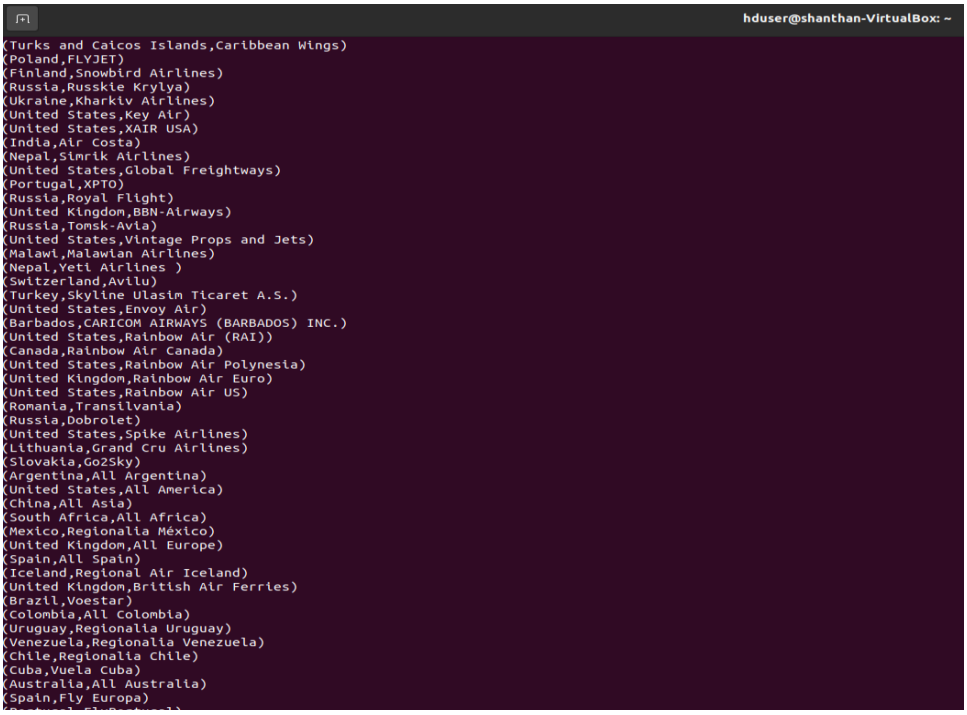
```
Input(s):
Successfully read 67663 records from: "/pig/routes.dat"
Successfully read 6048 records from: "/pig/Final_airlines"

Output(s):
Successfully stored 10 records (150056224 bytes) in: "hdfs://localhost:54310/tmp/temp-130244988/tmp1067114587"

Counters:
Total records written : 10
Total bytes written : 150056224
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1956056652_0009        ->      job_local635458340_0010,
job_local635458340_0010


2022-04-05 10:10:40,531 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-04-05 10:10:40,532 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-04-05 10:10:40,534 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-04-05 10:10:40,557 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-04-05 10:10:40,572 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-04-05 10:10:40,575 [main] WARN  org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-04-05 10:10:40,586 [main] INFO  org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-04-05 10:10:40,598 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-05 10:10:40,602 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-04-05 10:10:40,602 [main] INFO  org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(410,Aerocondor,\N,2B,ARD,AEROCONDOR,Portugal,Y,2B,410,DME,4029,KZN,2990,)
(410,Aerocondor,\N,2B,ARD,AEROCONDOR,Portugal,Y,2B,410,DME,4029,TGK,\N,)
(410,Aerocondor,\N,2B,ARD,AEROCONDOR,Portugal,Y,2B,410,DME,4029,UUA,6160,)
(410,Aerocondor,\N,2B,ARD,AEROCONDOR,Portugal,Y,2B,410,EGO,6156,KGD,2952,)
(410,Aerocondor,\N,2B,ARD,AEROCONDOR,Portugal,Y,2B,410,EGO,6156,KZN,2990,)
(410,Aerocondor,\N,2B,ARD,AEROCONDOR,Portugal,Y,2B,410,GYD,2922,NBC,6969,)
(410,Aerocondor,\N,2B,ARD,AEROCONDOR,Portugal,Y,2B,410,KGD,2952,EGO,6156,)
(410,Aerocondor,\N,2B,ARD,AEROCONDOR,Portugal,Y,2B,410,KZN,2990,AER,2965,)
(410,Aerocondor,\N,2B,ARD,AEROCONDOR,Portugal,Y,2B,410,KZN,2990,ASF,2966,)
(410,Aerocondor,\N,2B,ARD,AEROCONDOR,Portugal,Y,2B,410,KZN,2990,CEK,2968,)
```


hduser@shanthan-VirtualBox: ~
(Turks and Caicos Islands,Caribbean Wings)
(Poland,FLYJET)
(Finland,Snowbird Airlines)
(Russia,Russkie Krylya)
(Ukraine,Kharkiv Airlines)
(United States,Key Air)
(United States,XAIR USA)
(India,Air Costa)
(Nepal,Simrik Airlines)
(United States,Global Freightways)
(Portugal,XPTO)
(Russia,Royal Flight)
(United Kingdom,BBN-Airways)
(Russia,Tomsk-Avia)
(United States,Vintage Props and Jets)
(Malawi,Malawian Airlines)
(Nepal,Yeti Airlines)
(Switzerland,Avilu)
(Turkey,Skyline Ulasim Ticaret A.S.)
(United States,Envoy Air)
(Barbados,CARICOM AIRWAYS (BARBADOS) INC.)
(United States,Rainbow Air (RAI))
(Canada,Rainbow Air Canada)
(United States,Rainbow Air Polynesia)
(United Kingdom,Rainbow Air Euro)
(United States,Rainbow Air US)
(Romania,Transilvania)
(Russia,Dobrolet)
(United States,Spike Airlines)
(Lithuania,Grand Cru Airlines)
(Slovakia,Go2Sky)
(Argentina,All Argentina)
(United States,All America)
(China,All Asia)
(South Africa,All Africa)
(Mexico,Regionalia México)
(United Kingdom,All Europe)
(Spain,All Spain)
(Iceland,Regional Air Iceland)
(United Kingdom,British Air Ferries)
(Brazil,Voestar)
(Colombia,All Colombia)
(Uruguay,Regionalia Uruguay)
(Venezuela,Regionalia Venezuela)
(Chile,Regionalia Chile)
(Cuba,Vuela Cuba)
(Australia,All Australia)
(Spain,Fly Europa)