



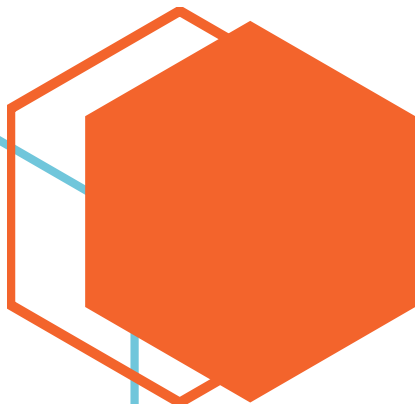
# PIG COMMANDS REVIEW

19CSE357  
BIG DATA ANALYTICS

CSE - E

DIVESH KOSURI  
CB.EN.U4CSE19422

09-04-2022



## ANALYSIS COMMANDS ON THE DATASETS

### DESCRIBE A;

```
grunt> describe A;
A: {route_lata: chararray,route_arid: chararray,route_source_lata: chararray,route_source_arid: chararray,route_codeshare: chararray,route_stops: chararray,route equip: chararray}
```

### EXPLAIN A;

```
#-----
# Map Reduce Plan
#-----
MapReduce node scope-203
Map Plan
routes: Store(fakefile:org.apache.pig.builtin.PigStorage) - scope-202
|
|---routes: New For Each(false,false,false,false,false,false,false)[bag] - scope-201
|
|   Cast[chararray] - scope-181
|   |---Project[bytearray][0] - scope-180
|   Cast[chararray] - scope-184
|   |---Project[bytearray][1] - scope-183
|   Cast[chararray] - scope-187
|   |---Project[bytearray][2] - scope-186
|   Cast[chararray] - scope-190
|   |---Project[bytearray][3] - scope-189
|   Cast[chararray] - scope-193
|   |---Project[bytearray][4] - scope-192
|   Cast[chararray] - scope-196
|   |---Project[bytearray][5] - scope-195
|   Cast[chararray] - scope-199
|   |---Project[bytearray][6] - scope-198
|
|---routes: Load(/CaseStudy/routes.dat:PigStorage(',')) - scope-179-----
Global sort: false
-----
```

### Illustrate A;

```
2022-04-04 22:26:43,748 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-04 22:26:43,748 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnlyMap - Allases being processed per job phase (AliasName[line,offset]): M: A[1,4] C: R:
2022-04-04 22:26:43,748 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MRCompiler - File concatenation threshold: 100 optimistic: false
2022-04-04 22:26:43,749 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size before optimization: 1
2022-04-04 22:26:43,750 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MultiQueryOptimizer - MR plan size after optimization: 1
2022-04-04 22:26:43,751 [main] INFO org.apache.pig.tools.pigstats.mapreduce.MRScriptState - Pig script settings are added to the job
2022-04-04 22:26:43,769 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.JobControlCompiler - mapred.job.reduce.markreset.buffer.percent is not set, set to default 0.3
2022-04-04 22:26:43,769 [main] INFO org.apache.pig.impl.util.SpillableMemoryManager - Selected heap (PS Old Gen) of size 699400192 to monitor. collectionUsageThreshold = 489580128, usageThreshold = 489580128
2022-04-04 22:26:43,769 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-04 22:26:43,769 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.PigMapOnlyMap - Allases being processed per job phase (AliasName[line,offset]): M: A[1,4] C: R:
-----
| A      | route_lata:chararray | route_arid:chararray | route_source_lata:chararray | route_source_arid:chararray | route_codeshare:chararray | route_stops:chararray | route equip:cha
|-----|
| AC     |                      |                      |                      |                      | YWL                      | 162                   | Y
|-----|
|-----|
```

## QUERY 1 ->

Query 1:

Query1 = FILTER finalairs BY territory == 'United States';

Dump Query1;

```
HadoopVersion  PigVersion  UserId  StartedAt  FinishedAt  Features
3-3.1  0-17.0  hduser  2022-04-05 00:08:39  2022-04-05 00:08:47  FILTER

Success!

Job Stats (time in seconds):
JobId  Maps  Reduces  MaxMapTime  MinMapTime  AvgMapTime  MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReduceTime  Alias  Feature  Outputs
job_local968342233_0002  1  0  n/a  n/a  n/a  n/a  0  0  MAP_ONLY  hdfs://localhost:54310/tmp/temp-1344964502/tmp522233176,

Input(s):
Successfully read 6048 records (14562740 bytes) from: "/CaseStudy/Final_airs"

Output(s):
Successfully stored 1080 records (12018703 bytes) in: "hdfs://localhost:54310/tmp/temp-1344964502/tmp522233176"

Counters:
Total records written : 1080
Total bytes written : 12018703
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local968342233_0002
```

```
2022-04-05 00:08:47,482 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-04-05 00:08:47,483 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-04-05 00:08:47,486 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-04-05 00:08:47,489 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2022-04-05 00:08:47,490 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2022-04-05 00:08:47,496 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input files to process : 1
2022-04-05 00:08:47,496 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(2,135 Airways,\N,,GNL,GENERAL,United States,N)
(9,3D Aviation,\N,,SEC,SECUREX,United States,N)
(10,40-Mile Air,\N,Q5,MLA,MILE-AIR,United States,Y)
(22,Aloha Airlines,\N,AQ,AAH,ALOHA,United States,Y)
(23,Alaska Island Air,\N,,AAK,ALASKA ISLAND,United States,N)
(24,American Airlines,\N,AA,AAL,AMERICAN,United States,Y)
(25,Aviation Management Corporation,\N,,AAM,AM CORP,United States,N)
(26,Atlantis Airlines (USA),\N,,AAO,ATLANTIS AIR,United States,N)
(31,Astro Air International,\N,,AAV,ASTRO-PHIL,United States,N)
(35,Allegiant Air,\N,G4,AAV,ALLEGIAN,United States,Y)
(49,Airborne Express,\N,GB,ABX,ABEX,United States,N)
(50,ABX Air,\N,GB,ABX,ABEX,United States,N)
(53,Academy Airlines,\N,,ACD,ACADEMY,United States,N)
(57,Air Aurora,\N,,AAI,BOREALIS,United States,N)
(66,Advance Leasing Company,\N,4G,,United States,N)
(67,Aztec Worldwide Airlines,\N,7A,,United States,N)
(81,Aerodyne,\N,,ADY,AERODYNE,United States,N)
(89,Anderson Aviation,\N,,ADX,ANDAX,United States,N)
(109,Alaska Central Express,\N,KO,AER,ACE AIR,United States,Y)
(115,Airway Express,\N,,AEX,AVCO,United States,N)
(117,Aerial Transit Company,\N,,AEZ,AERIAL TRANZ,United States,N)
(141,Air Guam,\N,,AGM,AIR GUAM,United States,N)
(145,Air Cargo America,\N,,MVM,PEGASUS,United States,N)
(149,Air Cargo Carriers,\N,2Q,SNC,NIGHT CARGO,United States,Y)
```

## QUERY 2 ->

Query2 :

Query2 = DISTINCT routes;

DUMP Query2;

```
(ZL,4178,SYD,3361,ARM,4062,)
(ZL,4178,SYD,3361,BHQ,6241,)
(ZL,4178,SYD,3361,BHS,6249,)
(ZL,4178,SYD,3361,BNK,4291,)
(ZL,4178,SYD,3361,DBO,3358,)
(ZL,4178,SYD,3361,GFF,6271,)
(ZL,4178,SYD,3361,LSY,6287,)
(ZL,4178,SYD,3361,MIM,6294,)
(ZL,4178,SYD,3361,MQL,6298,)
(ZL,4178,SYD,3361,MYA,6302,)
(ZL,4178,SYD,3361,NRA,6307,)
(ZL,4178,SYD,3361,NTL,4320,)
(ZL,4178,SYD,3361,OAG,6793,)
(ZL,4178,SYD,3361,PKE,6317,)
(ZL,4178,SYD,3361,TRO,6794,)
(ZL,4178,SYD,3361,WGA,3363,)
(ZL,4178,TRO,6794,GFN,6792,)
(ZL,4178,TRO,6794,SYD,3361,)
(ZL,4178,TSV,3330,ISA,3324,)
(ZL,4178,TSV,3330,WIN,6337,)
(ZL,4178,WGA,3363,MEL,3339,)
(ZL,4178,WGA,3363,SYD,3361,)
(ZL,4178,WIN,6337,LRE,6289,)
(ZL,4178,WIN,6337,TSV,3330,)
(ZL,4178,WYA,6334,ADL,3341,)
(ZM,19016,DME,4029,FRU,2912,)
(ZM,19016,FRU,2912,DME,4029,)
(ZM,19016,FRU,2912,OSS,2913,)
```

```
2022-04-05 00:16:38,479 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.LocalJobRunner - map
2022-04-05 00:16:38,479 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.Task - Task 'attempt_local344417590_0005_m_000000_0' done.
2022-04-05 00:16:38,481 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.Task - Final Counters for attempt_local344417590_0005_m_000000_0: Counters: 21
File System Counters
  FILE: Number of bytes read=2158
  FILE: Number of bytes written=32777528
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=44860897
  HDFS: Number of bytes written=38020006
  HDFS: Number of read operations=493
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=05
  HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=67663
  Map output records=67663
  Input split bytes=370
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=60
  Total committed heap usage (bytes)=396886016
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0
```

## QUERY 3 ->

Query 3:

Query3 = JOIN finalairs BY airline\_iata, routes BY route\_iata;

DUMP Query3;

```
2022-04-05 00:28:13,840 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.Task - Final Counters for attempt_local2059191319_0007_m_000000_0: Counters: 24
File System Counters
  FILE: Number of bytes read=6834859
  FILE: Number of bytes written=55729492
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=72931082
  HDFS: Number of bytes written=58111619
  HDFS: Number of read operations=641
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=89
  HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=67663
  Map output records=67663
  Map output bytes=2843656
  Map output materialized bytes=2978988
  Input split bytes=370
  Combine input records=0
  Spilled Records=67663
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=20
  Total committed heap usage (bytes)=652214272
MultiInputCounters
  Input records from _1_routes.dat=67663
File Input Format Counters
  Bytes Read=0
```

```
(4178,Regional Express,\N,ZL,RXA,REX,Australia,Y,ZL,4178,NTL,4320,BNK,4291,)
(4178,Regional Express,\N,ZL,RXA,REX,Australia,Y,ZL,4178,NTL,4320,SYD,3361,)
(4178,Regional Express,\N,ZL,RXA,REX,Australia,Y,ZL,4178,OAG,6793,SYD,3361,)
(4178,Regional Express,\N,ZL,RXA,REX,Australia,Y,ZL,4178,PKE,6317,BHS,6249,)
(4178,Regional Express,\N,ZL,RXA,REX,Australia,Y,ZL,4178,PKE,6317,SYD,3361,)
(4178,Regional Express,\N,ZL,RXA,REX,Australia,Y,ZL,4178,PLO,6318,ADL,3341,)
(4178,Regional Express,\N,ZL,RXA,REX,Australia,Y,ZL,4178,RCM,3360,JCK,6276,)
(19016,Apache Air,Apache,ZM,IWA,APACHE,United States,Y,ZM,19016,DME,4029,FRU,2912,)
(19016,Apache Air,Apache,ZM,IWA,APACHE,United States,Y,ZM,19016,OSS,2913,FRU,2912,)
(19016,Apache Air,Apache,ZM,IWA,APACHE,United States,Y,ZM,19016,FRU,2912,OSS,2913,)
```

## QUERY 4 ->

Query 4:

Query4 = FOREACH routes GENERATE route\_airid, route\_source\_iata, route\_stops;

DUMP Query4;



```
2022-04-05 00:35:44,620 [LocalJobRunner Map Task Executor #0] INFO org.apache.hadoop.mapred.Task - Final Counters for attempt_local727796562_0008_m_000000_0: Counters: 21
File System Counters
  FILE: Number of bytes read=13666332
  FILE: Number of bytes written=66133762
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=89958661
  HDFS: Number of bytes written=74041607
  HDFS: Number of read operations=733
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=104
  HDFS: Number of bytes read erasure-coded=0
Map-Reduce Framework
  Map input records=67663
  Map output records=67663
  Input split bytes=370
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=5
  Total committed heap usage (bytes)=513277952
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=0
```

```
(4178,TRO,3361)
(4178,TSV,3324)
(4178,TSV,6337)
(4178,WGA,3339)
(4178,WGA,3361)
(4178,WIN,6289)
(4178,WIN,3330)
(4178,WYA,3341)
(19016,DME,2912)
(19016,FRU,4029)
(19016,FRU,2913)
```

## QUERY 6 ->

Query 6:

Ter 1->

`hadoop fs -mkdir /home/hduser/OutPut`

`STORE Query1 INTO '/home/hduser/OutPut' USING PigStorage(',');`

Ter 2->

`hadoop fs -ls /home/hduser/OutPut`

`hadoop fs -cat /home/hduser/OutPut/part-m-00000`

DUMP Query6;

```

hduser@DiveshKosuri:~$ hadoop fs -mkdir /home/hduser/OutPut
2022-04-08 22:12:22,188 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
mkdir: '/home/hduser/OutPut': File exists
hduser@DiveshKosuri:~$ hadoop fs -cat /home/hduser/OutPut/part-m-00000
2022-04-08 22:13:35,862 WARN util.NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
2,135 Airways,\N,,GNL,GENERAL,United States,N
9,3D Aviation,\N,,SEC,SECUREX,United States,N
10,40-Mile Air,\N,Q5,MLA,MILE-AIR,United States,Y
22,Aloha Airlines,\N,AQ,AAH,ALOHA,United States,Y
23,Alaska Island Air,\N,,AAK,ALASKA ISLAND,United States,N
24,American Airlines,\N,AA,AAL,AMERICAN,United States,Y
25,Aviation Management Corporation,\N,,AAM,AM CORP,United States,N
26,Atlantis Airlines (USA),\N,,AAO,ATLANTIS AIR,United States,N
31,Astro Air International,\N,,AAV,ASTRO-PHIL,United States,N
35,Allegiant Air,\N,G4,AAV,ALLEGIAN,United States,Y
49,Airborne Express,\N,GB,ABX,ABEX,United States,N
50,ABX Air,\N,GB,ABX,ABEX,United States,N
53,Academy Airlines,\N,,ACD,ACADEMY,United States,N
57,Air Aurora,\N,,AAI,BOREALIS,United States,N
66,Advance Leasing Company,\N,4G,,United States,N
67,Aztec Worldwide Airlines,\N,7A,,United States,N
81,Aerodyne,\N,,ADV,AERODYNE,United States,N
89,Anderson Aviation,\N,,ADX,ANDAX,United States,N
109,Alaska Central Express,\N,KO,AER,ACE AIR,United States,Y
115,Airway Express,\N,,AEX,AVCO,United States,N
117,Aerial Translt Company,\N,,AEZ,AERIAL TRANZ,United States,N
141,Air Guam,\N,,AGM,AIR GUAM,United States,N
145,Air Cargo America,\N,,MVM,PEGASUS,United States,N
149,Air Cargo Carriers,\N,2Q,SNC,NIGHT CARGO,United States,Y
164,Atlantic Gulf Airlines,\N,,AGF,ATLANTIC GULF,United States,N
177,Atlantic Coast Airlines,\N,,BLR,BLUE RIDGE,United States,N
182,Aero Flight Service,\N,,ACV,FLIGHT GROUP,United States,N
188,Aspen Helicopters,\N,,AHF,ASPEN,United States,N
205,Alpine Air Express,\N,5A,AIP,ALPINE AIR,United States,N
210,Airlift International,\N,,AIR,AIRLIFT,United States,Y
217,Air Illinois,\N,,AIL,AIR ILLINOIS,United States,N
236,Ameristar Jet Charter,\N,,AJI,AMERISTAR,United States,N
237,A2 Jet Leasing,\N,,AJJ,ATLANTIC JET,United States,N

```

```

2022-04-08 22:31:41,493 [pool-24-thread-1] INFO org.apache.hadoop.mapred.LocalJobRunner - Finishing task: attempt_local1641374742_0007_r_000000_0
2022-04-08 22:31:41,495 [Thread-516] INFO org.apache.hadoop.mapred.LocalJobRunner - reduce task executor complete.
2022-04-08 22:31:41,799 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Running jobs are [job_local1641374742_0007]
2022-04-08 22:31:46,327 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-04-08 22:31:46,332 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-04-08 22:31:46,335 [main] WARN org.apache.hadoop.metrics2.impl.MetricsSystemImpl - JobTracker metrics system already initialized!
2022-04-08 22:31:46,359 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - 100% complete
2022-04-08 22:31:46,366 [main] INFO org.apache.pig.tools.pigstats.mapreduce.SimplePigStats - Script Statistics:

HadoopVersion  PigVersion  UserId  StartedAt  FinishedAt  Features
3.3.1  0.17.0  hduser  2022-04-08 22:31:33  2022-04-08 22:31:46  GROUP_BY,LIMIT

Success!

Job Stats (time in seconds):
JobId  Maps  Reduces  MaxMapTime  MinMapTime  AvgMapTime  MedianMapTime  MaxReduceTime  MinReduceTime  AvgReduceTime  MedianReductime  Alias  Feature Outputs
job_local1512289408_0006  1  1  n/a  n/a  n/a  n/a  n/a  n/a  n/a  n/a  Query7,Query8,finalairs  GROUP_BY
job_local1641374742_0007  1  1  n/a  n/a  n/a  n/a  n/a  n/a  n/a  n/a  hdfs://localhost:54310/tmp/temp1955396190/tmp1969260916,

Input(s):
Successfully read 6048 records (83000378 bytes) from: "/CaseStudy/Final_airs"

```

## QUERY 7 ->

Query 7:

Query7 = GROUP finalairs by territory;

DUMP Query7;

```

Total Length = 316243
Input split[0]:
  Length = 316243
  ClassName: org.apache.hadoop.mapreduce.lib.input.FileSplit
  Locations:

```



```
Democratic Republic of the Congo,((486,Air Transport Service,N,,ATS,,Democratic Republic of the Congo,N),(396,Air Plan International,N,,APV,AIR PLAN,Democratic Republic of the Congo,N),(764,Air Charter Services,N,,CHB,ZAIRE CHARTER,Democratic Republic of the Congo,N),(4338,Scibe Airlift,N,,SBZ,SCIBE AIRLIFT,Democratic Republic of the Congo,N),(1953,DAS Airlines,N,,RKC,DAS CONGO,Democratic Republic of the Congo,N),(888,African Company Airlines,N,,FPY,AFRICOMPANY,Democratic Republic of the Congo,N),(3146,Kinshasa Airways,N,,KNS,KINSHASA AIRWAYS,Democratic Republic of the Congo,N),(1479,Bravo Air Congo,N,,BRV,BRAVO,Democratic Republic of the Congo,N),(1502,Blue Airlines,N,,BUL,BLUE AIRLINES,Democratic Republic of the Congo,N),(3261,Lignes Aeriennes Congolaises,N,,LCC,CONGOLAISE,Democratic Republic of the Congo,N),(2756,Hewa Bora Airways,N,,CO,ALX,ALLCONGO,democratic Republic of the Congo,N),(38,African Business and Transportations,N,,ABB,AFRICAN BUSINESS,Democratic Republic of the Congo,N),(4412,Shabair,N,,SHB,SHABAIR,Democratic Republic of the Congo,N),(1718,Cetraca Aviation Service,N,,CER,CETRACA,Democratic Republic of the Congo,N),(2478,Funtshi Aviation Service,N,,FUN,FUNTSHI,Democratic Republic of the Congo,N),(2215,Espace Aviation Services,N,,EPC,ESPACE,Democratic Republic of the Congo,N),(2204,Enterprise World Airways,N,,EWS,WORLD ENTERPRISE,Democratic Republic of the Congo,N),(3382,Mallia Airlift,N,,MLC,MALLIA,Democratic Republic of the Congo,N),(758,Africa One,N,,CFR,Democratic Republic of the Congo,N))
(Saint Vincent and the Grenadines,((4534,St. Vincent Grenadines Air (1999),N,,SVO,GRENADINES,Saint Vincent and the Grenadines,N),(11724,SVG Air,,SVG,Grenadines,Saint Vincent and the Grenadines,Y)))
(Democratic People's Republic of Korea,((998,Air Koryo,N,,J5,KOR,AIR KORYO,Democratic People's Republic of Korea,Y)))
((415,Aerojet de Costa Rica,N,,ARJ,S.A.,N),(1,Private flight,N,,N/A,,Y),(4430,Scorpio Aviation,N,,8S,,N),(4854,Sirin,N,,R1,,N),(4051,Princess Air,N,,8Q,,N),(3562,Multiflight,N,,HFT,YORKAIR,N),(5199,Transports et Travaux A,N,,OF,,N),(2884,ICAO,N,,YYY,,N),(2649,Guyana Airways 2800,N,,GY,,N),(2636,Gulf African Airlines - Gambia,N,,GUF,GULF AFRICAN,N),(2080,Electronic Data Systems,N,,N),(5199,Transports et Travaux A,N,,OF,,N),(2884,ICAO,N,,YYY,,N),(2649,Guyana Airways 2800,N,,GY,,N),(2636,Gulf African Airlines - Gambia,N,,GUF,GULF AFRICAN,N),(2080,Electronic Data Systems,N,,N)
```

## QUERY 8 ->

Query 8:

Query8 = LIMIT Query7 15;

### File System Counters

```
FILE: Number of bytes read=8091982
FILE: Number of bytes written=34351155
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=29671070
HDFS: Number of bytes written=26409654
HDFS: Number of read operations=570
HDFS: Number of large read operations=0
HDFS: Number of write operations=52
HDFS: Number of bytes read erasure-coded=0
```

### Map-Reduce Framework

```
Combine input records=0
Combine output records=0
Reduce input groups=277
Reduce shuffle bytes=431550
Reduce input records=6048
Reduce output records=4
Spilled Records=6048
Shuffled Maps =1
Failed Shuffles=0
Merged Map outputs=1
GC time elapsed (ms)=0
Total committed heap usage (bytes)=697303040
```

### Shuffle Errors

```
BAD_ID=0
CONNECTION=0
IO_ERROR=0
WRONG_LENGTH=0
WRONG_MAP=0
WRONG_REDUCE=0
```

### File Output Format Counters

```
Bytes Written=0
```

```
2022-04-08 22:31:46,403 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(S,((1390,British Gulf International,N,,BGI,BRITISH GULF,S,N),(2290,Express International Cargo,N,,EIC,EXCARGO,S,N),(3591,Natalco Air Lines,N,,NCO,NATALCO,S,N),(3286,Linhas Aereas Santonenses,N,,SMS,SANTONENSES,S,N),(3288,LASTP,N,,OTH,LASTP,S,N)))
(N,((5533,Tyrolean Airways,N,,TYR,TYRDLIN,N,Y),(5556,buzz,N,,UK,BUZ,N,N,N),(-1,Unknown,N,,N/A,N,N,Y)))
(LAP,((3207,LAP Colombia - Lineas Aereas Petroleras,N,,APT,S.A.,LAP,N)))
(ACON,((4456,Southern Jersey Airways,N,,ALC,Inc.,ACON,N)))
(ASUR,((1851,Compania Boliviana de Transporte Aereo Privado Aerosur,N,,ASU,S.A.,ASUR,N)))
(ATCO,((480,Aerotaas,N,,AQA,S.A.,ATCO,N)))
(AUSA,((513,Aerolineas Uruguayas,N,,AUY,S.A.,AUSA,N)))
(Chad,((3577,Nada Air Service,N,,NHZ,NADA AIR,Chad,N),(894,Air Affaires Tchad,N,,FTC,AFFAIRES TCHAD,Chad,N),(1733,Chart Aviation Services,N,,SB,CSU,CHART SERVICE,Chad,N),(3262,Lignes Aeriennes Du Tchad,N,,LAD,LATCHAD,Chad,N),(4807,Tchad Airlines,N,,TCD,TCHADLINES,Chad,N),(4942,Toumai Air Tchad,N,,THE,TOUMAI AIR,Chad,N),(1312,Air Horizon,N,,TPK,TCHAD-HORIZON,Chad,N),(433,Air Tchad,N,,HTT,HOTEL TAN GO,Chad,N)))
```



## QUERY 9 ->

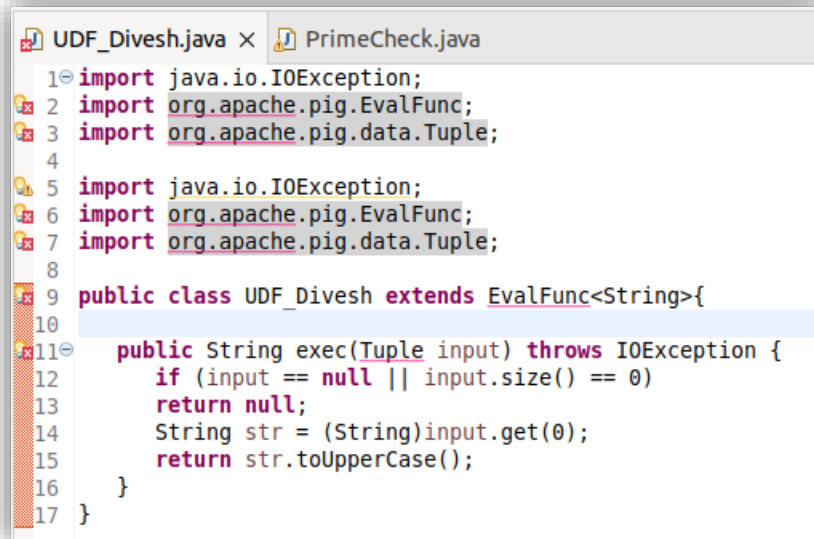
QUERY 9:

Loading JAR file in PIG

REGISTER '/home/hduser/BigData/JARs/UDF\_Divesh.jar';

PIG\_Input = LOAD '/CaseStudy/PigInput.txt' using PigStorage() AS (line:chararray);

PIG\_UDF = FOREACH PIG\_Input GENERATE UDF\_Divesh(line);



```
UDF_Divesh.java x PrimeCheck.java
1 import java.io.IOException;
2 import org.apache.pig.EvalFunc;
3 import org.apache.pig.data.Tuple;
4
5 import java.io.IOException;
6 import org.apache.pig.EvalFunc;
7 import org.apache.pig.data.Tuple;
8
9 public class UDF_Divesh extends EvalFunc<String>{
10
11     public String exec(Tuple input) throws IOException {
12         if (input == null || input.size() == 0)
13             return null;
14         String str = (String)input.get(0);
15         return str.toUpperCase();
16     }
17 }
```

```
grunt> REGISTER '/home/hduser/BigData/JARs/UDF_Divesh.jar';
grunt> PIG_Input = LOAD '/CaseStudy/PigInput.txt' using PigStorage() AS (line:chararray);
grunt> PIG_UDF = FOREACH PIG_Input GENERATE UDF_Divesh(line);
```