

Find the proportion of flights delayed. A flight is delayed if the delay is greater than 15 minutes. Compute the fraction of delayed flights per month in the year 2008. The delay could be any kind of delay (carrier delay, weather delay, etc)

Hint:

```
-- A flight is delayed if the delay is greater than 15 minutes.  
-- delay = arrival time - scheduled arrival time  
-- Compute the fraction of delayed flights per different time  
-- granularities (hour, day, week, month, year)  
(let's focus on a month)
```

Foreach month:

```
-- compute the total number of flights  
-- compute delay relation: only those flights with delay > 15 min appear here  
-- compute the total number of delayed flights  
-- output relation: month, ratio delayed/total
```

You need to calculate:

- Fraction delay from the departure
- Fraction delay due to carrier
- Fraction delay due to weather

Since there is no weather delay, only carrier delay related information is calculated

1. Fraction delay due to carrier

```
air = FOREACH airlines GENERATE month, carrier, (int)(arrrtime-satime) as  
delay;
```

```
grouped= GROUP A BY carrier;
```

```
TOTAL = FOREACH grouped {
```

```
    A = FILTER air BY (delay >= 15);
```

```
    GENERATE group, COUNT(air) as air, COUNT(A) as A, (float)  
COUNT(A)/COUNT(air) as fraction;
```

```
}
```

```
dump TOTAL
```

```
hduser@vinod-virtual-machine: ~
File Edit Tabs Help
hduser@vinod-virtual-machine:~$ pig -x local
17/11/28 23:08:21 INFO pig.ExecTypeProvider: Trying ExecType : LOCAL
17/11/28 23:08:21 INFO pig.ExecTypeProvider: Picked LOCAL as the ExecType
2017-11-28 23:08:21,564 [main] INFO org.apache.pig.Main - Apache Pig version 0.15.0 (r1682971) compiled Jun 01 2015, 11:43:55
2017-11-28 23:08:21,564 [main] INFO org.apache.pig.Main - Logging error messages to: /home/hduser/pig_1511890701563.log
2017-11-28 23:08:21,579 [main] INFO org.apache.pig.impl.util.Utils - Default bootup file /home/hduser/.pigbootup not found
2017-11-28 23:08:21,663 [main] INFO org.apache.pig.backend.hadoop.executionengine.HExecutionEngine - Connecting to hadoop file system at: file:///
grunt> airlines = load 'airlines.csv' using PigStorage(',') AS (year:int,month:int,day:int,dow:int,dtime:int,sdtime:int,arrrtime:int,satime:int,carrier:chararray,fn:int,tn:chararray,etime:int,setttime:int,airtime:int,adelay:int,ddelay:int,scode:chararray,dcode:chararray,dist:int,tintime:int,touttime:int,cancel:chararray,cancelcode:chararray,diverted:int,cdelay:int,wdelay:int,ndelay:int,slatdelay:int,latdelay:int);
grunt> air = FOREACH airlines GENERATE month, carrier, (int)(arrrtime-satime) AS delay;
grunt> grouped= GROUP air BY carrier;
grunt> TOTAL = FOREACH grouped {
>> A = FILTER air BY (delay >= 15);
>> GENERATE group, COUNT(air) as air, COUNT(A) as A, (float) COUNT(A)/COUNT(air) as fraction;
>> }
2017-11-28 23:09:19,675 [main] WARN org.apache.pig.newplan.BaseOperatorPlan - Encountered Warning IMPLICIT_CAST_TO_FLOAT 1 time(s).
grunt> dump TOTAL;
```

```
hduser@vinod-virtual-machine: ~
File Edit Tabs Help

Input(s):
Successfully read 30 records from: "file:///home/hduser/airlines.csv"

Output(s):
Successfully stored 1 records in: "file:/tmp/temp2022113812/tmp-837476252"

Counters:
Total records written : 1
Total bytes written : 0
Spillable Memory Manager spill count : 0
Total bags proactively spilled: 0
Total records proactively spilled: 0

Job DAG:
job_local1187423219_0001

2017-11-28 23:09:28,354 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Encountered Warning FIELD DISCARDED TYPE CONVERSION FAILED 3 time(s).
2017-11-28 23:09:28,359 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer.MapReduceLauncher - Success!
2017-11-28 23:09:28,365 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized
2017-11-28 23:09:28,368 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1
2017-11-28 23:09:28,368 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1
(UA,30,8,0.26666668)
grunt>
```

2. Fraction delay from the departure

```
air= FOREACH airlines GENERATE month, carrier, (int)(dtime-sdtime) as  
delay;
```

```
grouped= GROUP air BY carrier;
```

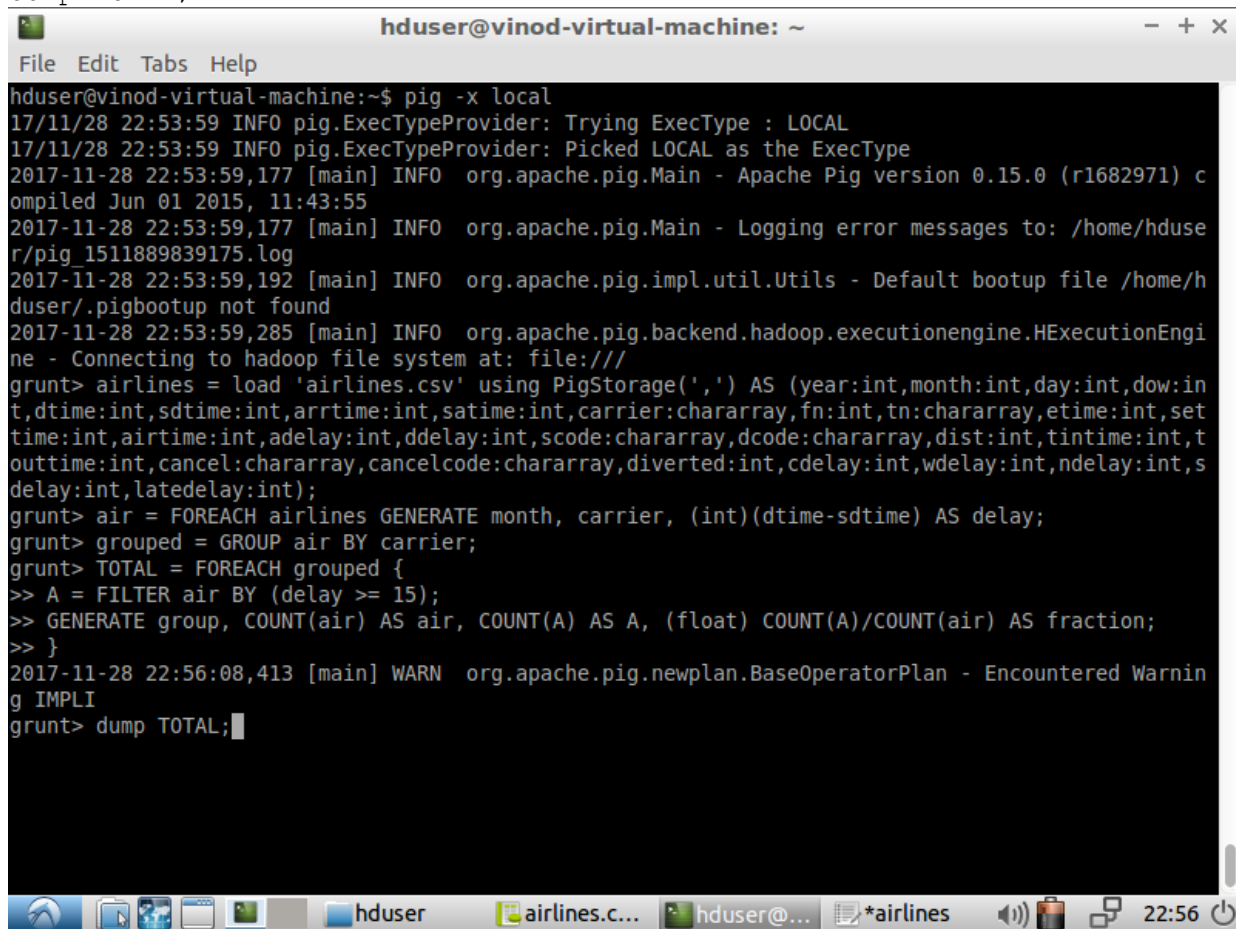
```
TOTAL = FOREACH grouped {
```

```
    A = FILTER air BY (delay >= 15);
```

```
    GENERATE group, COUNT(A) as air, COUNT(A) as A, (float)  
COUNT(A)/COUNT(air) as fraction;
```

```
}
```

```
dump TOTAL;
```

A screenshot of a terminal window titled "hduser@vinod-virtual-machine: ~". The terminal shows the execution of a Pig script. The script starts with loading the 'airlines.csv' file into a table named 'airlines'. It then generates a new table 'air' with columns for month, carrier, and delay (calculated as dtime-sdtime). The 'air' table is grouped by carrier into a table named 'grouped'. A new table 'TOTAL' is created by iterating over 'grouped'. For each group, it filters for rows where delay is greater than or equal to 15, counts the number of rows (A), and calculates the fraction of rows with delay >= 15 (COUNT(A)/COUNT(air)). The script ends with a 'dump TOTAL;' command. The terminal output shows various log messages from the Pig execution engine, including information about the local execution type, version, and file system connections. A warning message is also displayed: "Encountered Warning IMPLI". The terminal window has a standard Linux desktop environment with a taskbar at the bottom showing icons for the terminal, a file manager, and other applications. The system clock in the bottom right corner shows 22:56.

```
hduser@vinod-virtual-machine: ~  
File Edit Tabs Help  
Input(s):  
Successfully read 30 records from: "file:///home/hduser/airlines.csv"  
Output(s):  
Successfully stored 1 records in: "file:/tmp/temp400572078/tmp-979826576"  
Counters:  
Total records written : 1  
Total bytes written : 0  
Spillable Memory Manager spill count : 0  
Total bags proactively spilled: 0  
Total records proactively spilled: 0  
Job DAG:  
job_local273408587_0001  
2017-11-28 22:56:54,286 [main] WARN org.apache.pig.backend.hadoop.executionengine.mapReduceLayer  
.MapReduceLauncher - Encountered Warning FIELD_DISCARDED_TYPE_CONVERSION_FAILED 3 time(s).  
2017-11-28 22:56:54,286 [main] INFO org.apache.pig.backend.hadoop.executionengine.mapReduceLayer  
.MapReduceLauncher - Success!  
2017-11-28 22:56:54,288 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend  
has already been initialized  
2017-11-28 22:56:54,290 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total  
input paths to process : 1  
2017-11-28 22:56:54,290 [main] INFO org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil  
- Total input paths to process : 1  
(UA,30,4,0.13333334)  
grunt>
```