Task 3: Implement the use case present in below blog link and share the complete steps along with Screen shot(s) from your end.

https://acadgild.com/blog/aviation-data-analysis-using-apache-pig/

Problem Statement 1 : Find out the top 5 most visited destinations.

grunt> history

```
1  A= LOAD '/home/acadgild/Downloads/DelayedFlights.csv' Using org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');

2  B= foreach A generate (int)$1 as year, (int)$10 as flight_num, (chararray)$17 as origin, (chararray)$18 as dest;

3  C =filter B by dest is not null;

4  D = GROUP C by dest;

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

6  F =order E by $1 DESC;

7  Result = LIMIT F 5;

8  dump Result;

2018-05-13 22:15:46,787 [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)
```

Problem Statement 2 : Which month has been the most number of cancellations due to bad weather.

```
A= LOAD '/home/acadgild/Downloads/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;
```

```
2018-05-13 22:26:08,460 [main] INFO
org.apache.pig.backend.hadoop.executionengine.util.MapRedUtil - Total input paths to process : 1

(12,250)
```

PROBLEM STATEMENT 3 : Top ten origins with the highest AVG departure delay.

```
A= LOAD '/home/acadgild/Downloads/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/Downloads/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
```

2018-05-13 22:50:42,423 [main] WARN org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized

2018-05-13 22:50:42,491 [main] INFO org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-05-13 22:50:42,491 [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)

PROBLME STATEMENT 4 : Which route (origin&destination) has been the maximum diversion.

grunt> A= LOAD '/home/acadgild/Downloads/DelayedFlights.csv' Using org.apache.pig.piggybank.storage.CSVExcelStorage(',','NO_MULTILINE','UNIX','SKIP_INPUT_HEADER');

2018-05-13 22:54:32,442 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - io.bytes.per.checksum is deprecated. Instead, use dfs.bytes-per-checksum

2018-05-13 22:54:32,443 [main] INFO org.apache.hadoop.conf.Configuration.deprecation - fs.default.name is deprecated. Instead, use fs.defaultFS

grunt> B =FOREACH A GENERATE (chararray)$17 as origin, (chararray)$18 as dest, (int)$24 as diversion;

grunt> C = FILTER B BY (origin is not null) AND (dest is not null) AND (diversion ==1);

grunt> D = GROUP C by (origin,dest);

grunt> E =FOREACH D generate group,COUNT(C.diversion);

```
grunt> F = ORDER E by $1 DESC;

grunt> Res = limit F 1;

grunt> Result = limit F 10;

grunt> dump Result;


2018-05-13 23:04:24,145 [main] WARN  org.apache.pig.data.SchemaTupleBackend - SchemaTupleBackend has already been initialized

2018-05-13 23:04:24,182 [main] INFO  org.apache.hadoop.mapreduce.lib.input.FileInputFormat - Total input paths to process : 1

2018-05-13 23:04:24,182 [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>
```