**Write PIG scripts for following tasks.**

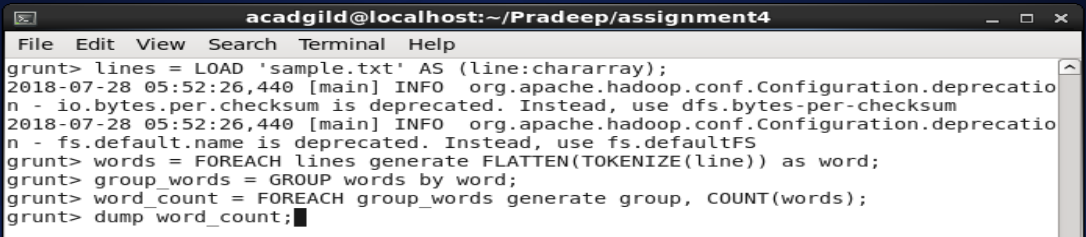
**Task 1.1**

Write a program to implement wordcount using Pig.

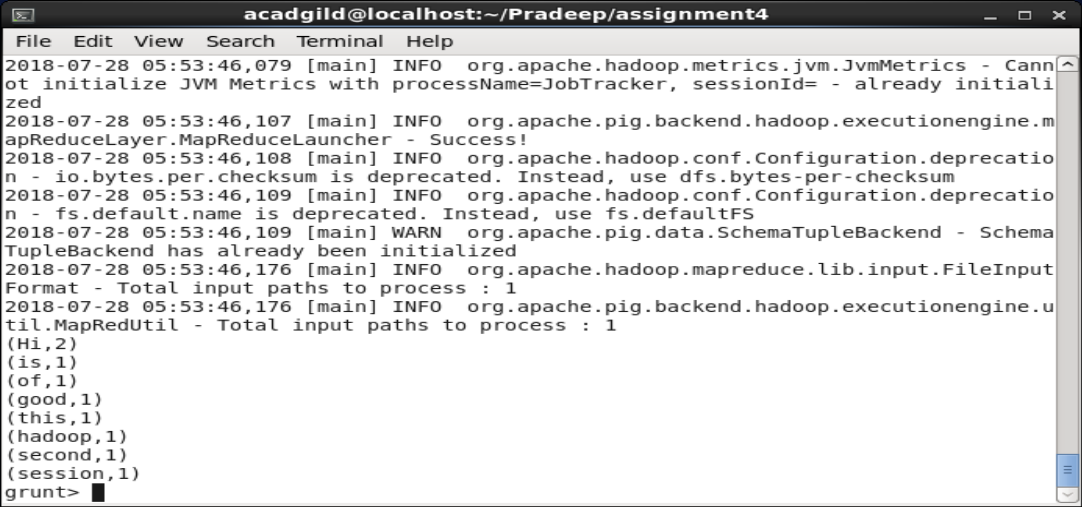
Input File – sample.txt



PIG SCRIPT:



**OUTPUT:**

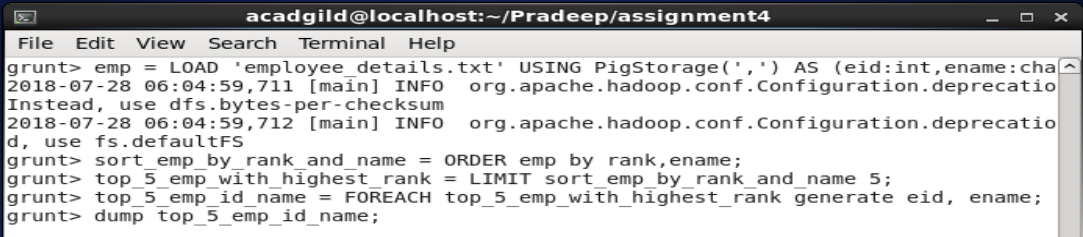


**Task 1.2**

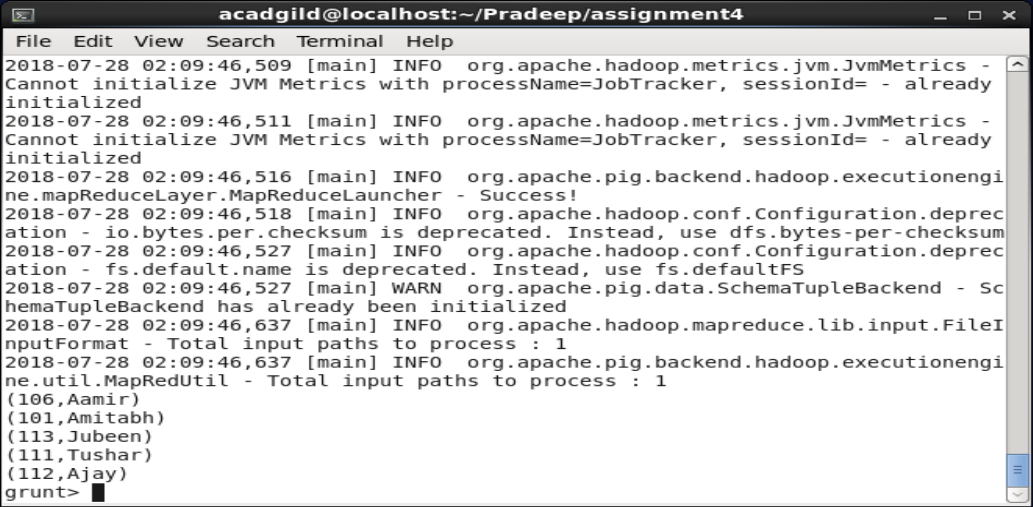
**We have employee\_details and employee\_expenses files. Use local mode while running Pig and write Pig Latin script to get below results:**

(a) Top 5 employees (employee id and employee name) with highest rating. (In case two employees have same rating, employee with name coming first in dictionary should get preference)

PIG Script:

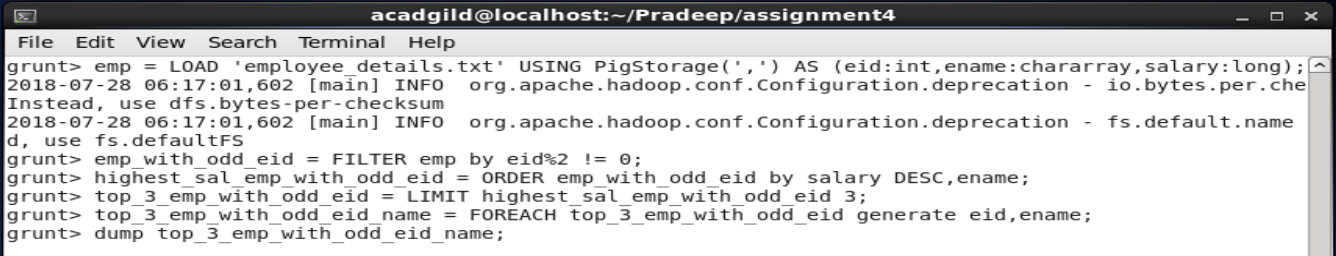


**OUTPUT:**

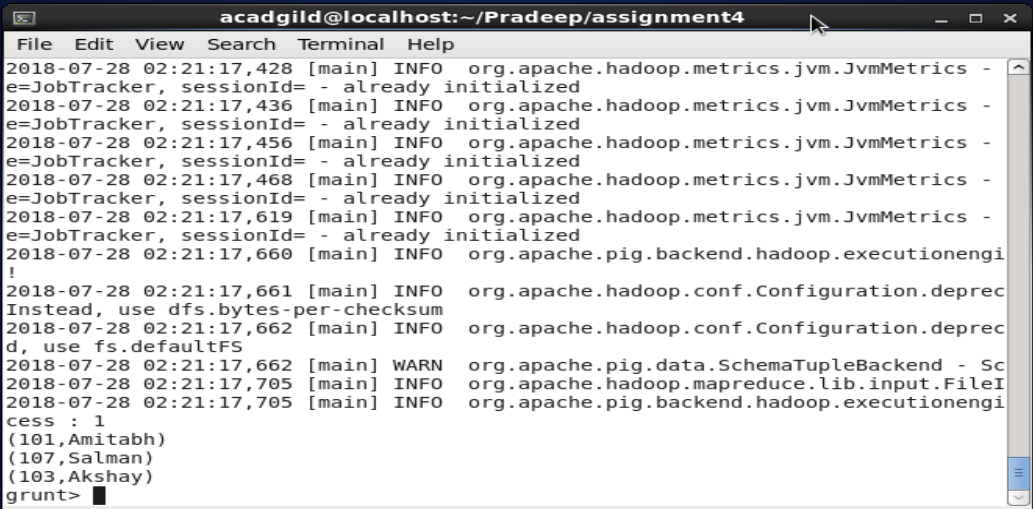


(b) Top 3 employees (employee id and employee name) with highest salary, whose employee id is an odd number. (In case two employees have same salary, employee with name coming first in dictionary should get preference)

PIG Script:

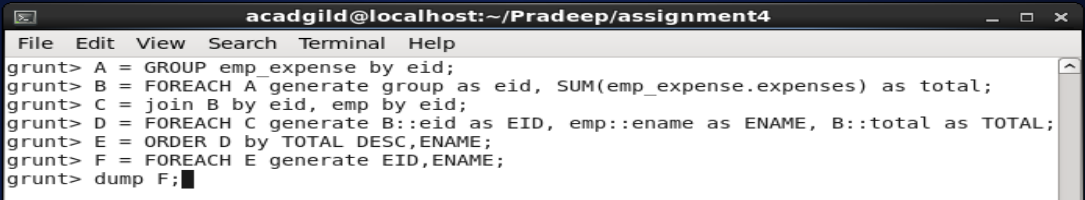


**OUPUT:**

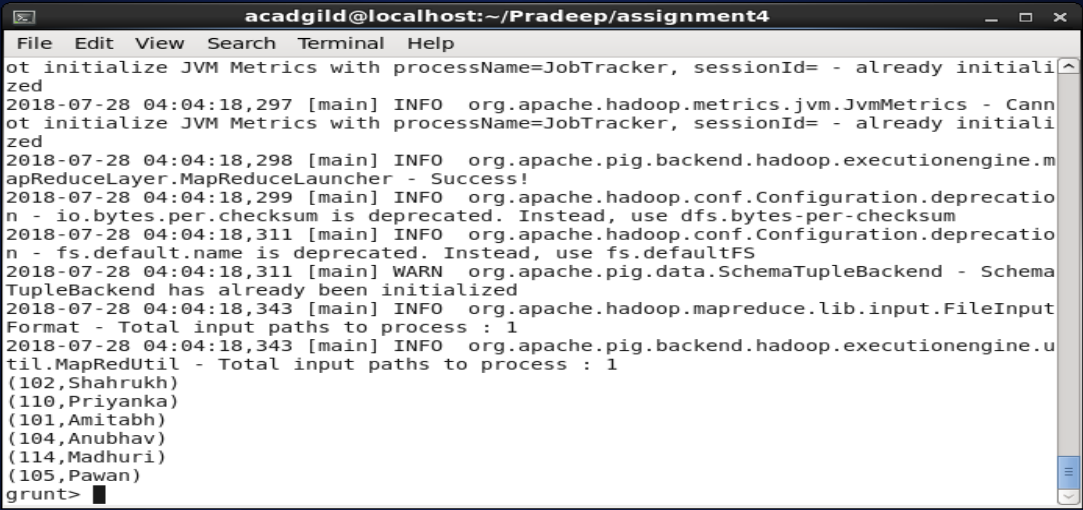


(c) Employee (employee id and employee name) with maximum expense (In case two employees have same expense, employee with name coming first in dictionary should get preference)

PIG Script:

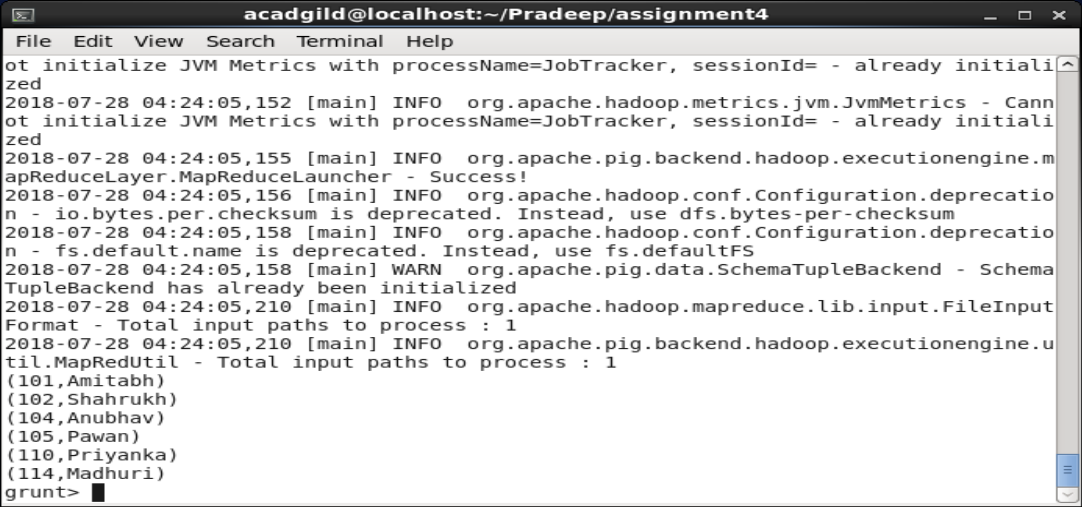


**OUPUT:**

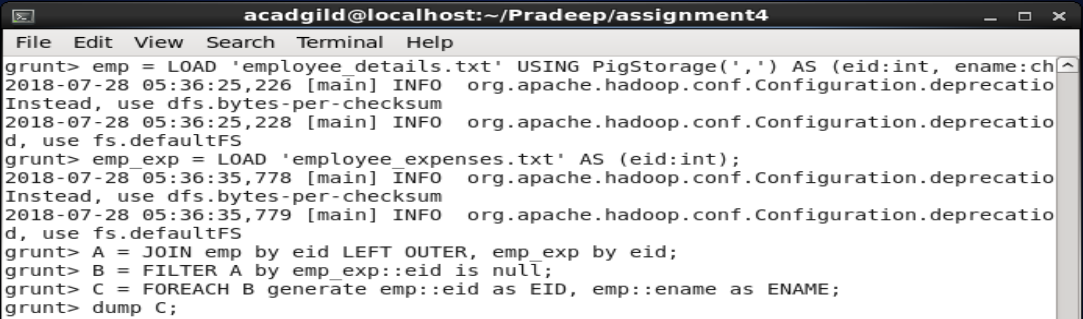


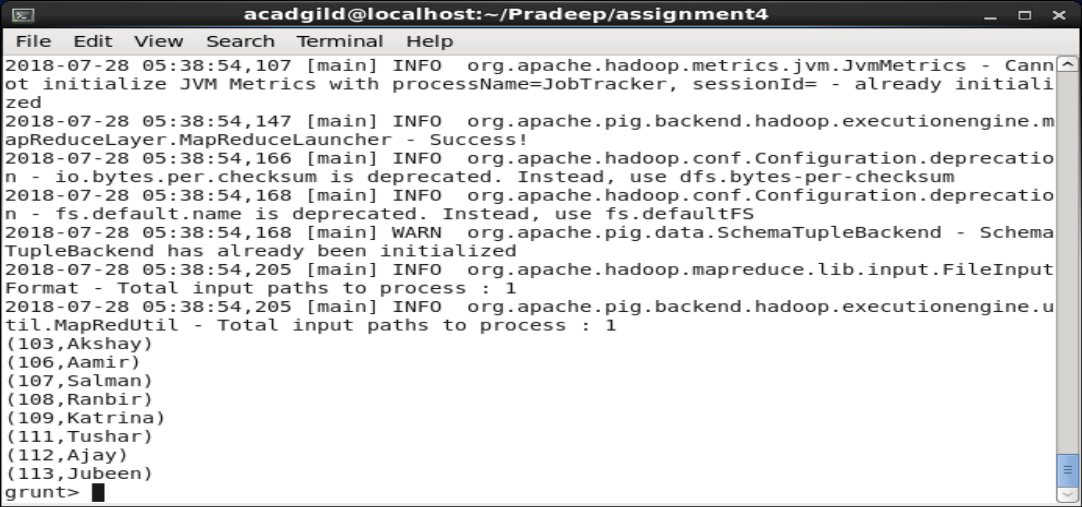
(d) List of employees (employee id and employee name) having entries in employee\_expenses file.





(e) List of employees (employee id and employee name) having no entry in employee\_expenses file.





Task 1.3

Implement the use case present in below blog link and share the complete steps along with screenshot(s) from your end.

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

**Problem Statement 1**

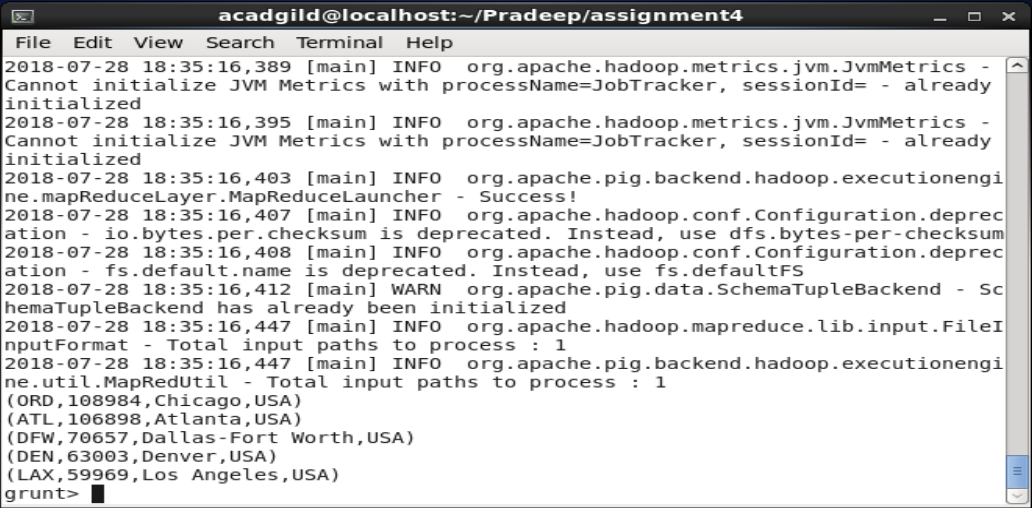
Find out the top 5 most visited destinations:

**NOTE:** In the blog with the given command we are not getting the most visited destinations in DESC order so I have added the additional commands to get that. And there were repetitions of destination code in the blog I have taken care of that also.

PIG Script:



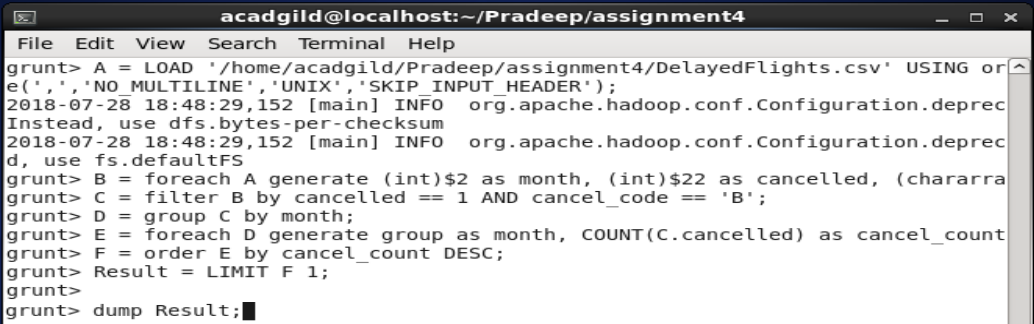
**OUPUT:**



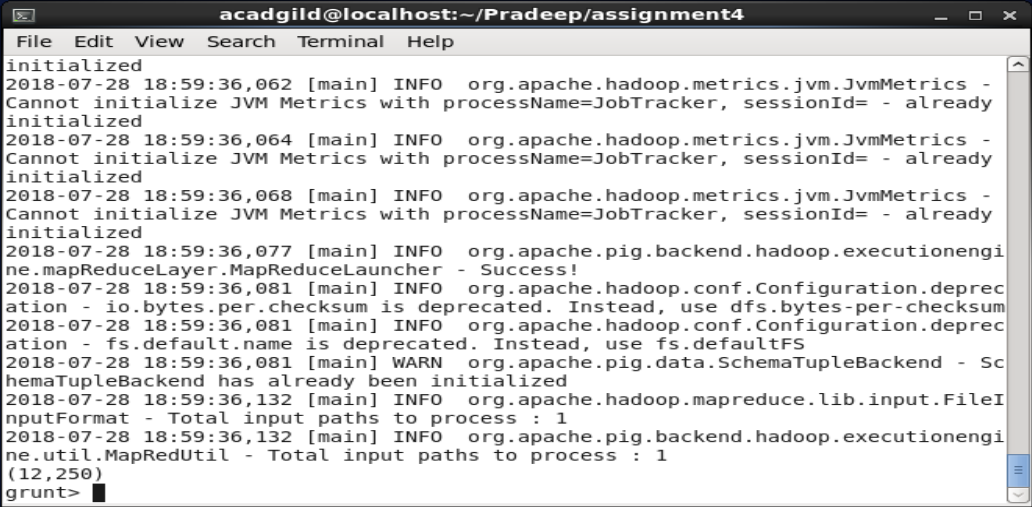
**Problem Statement 2**

Which month has seen the most number of cancellations due to bad weather?

PIG Script:

****

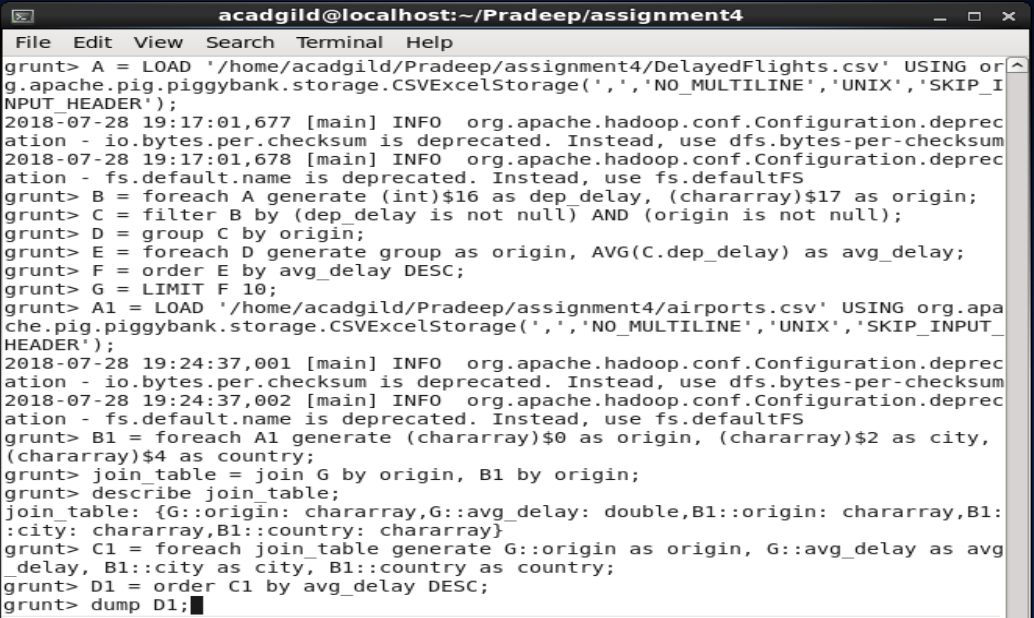
**OUPUT:**



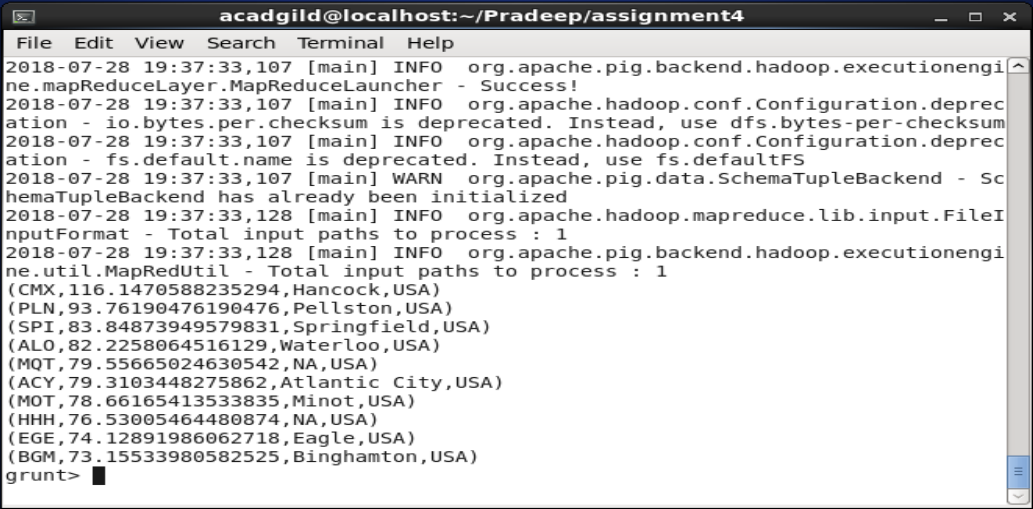
**Problem Statement 3**

Top ten origins with the highest AVG departure delay

PIG Script:



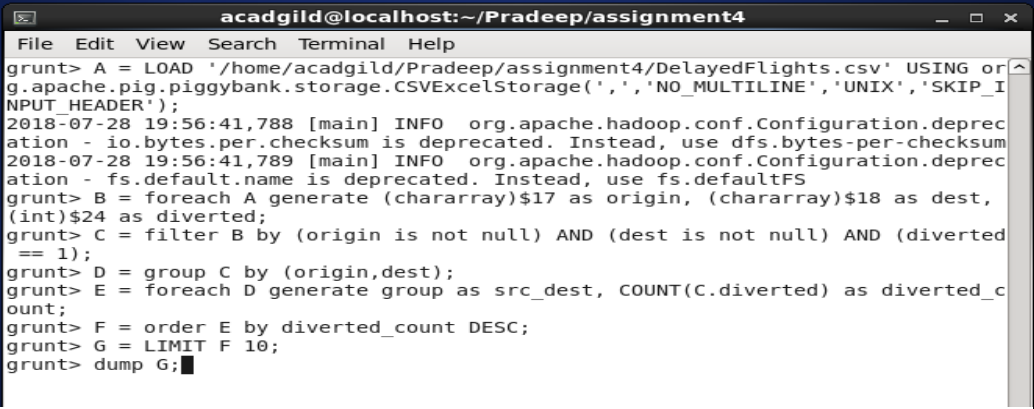
**OUPUT:**



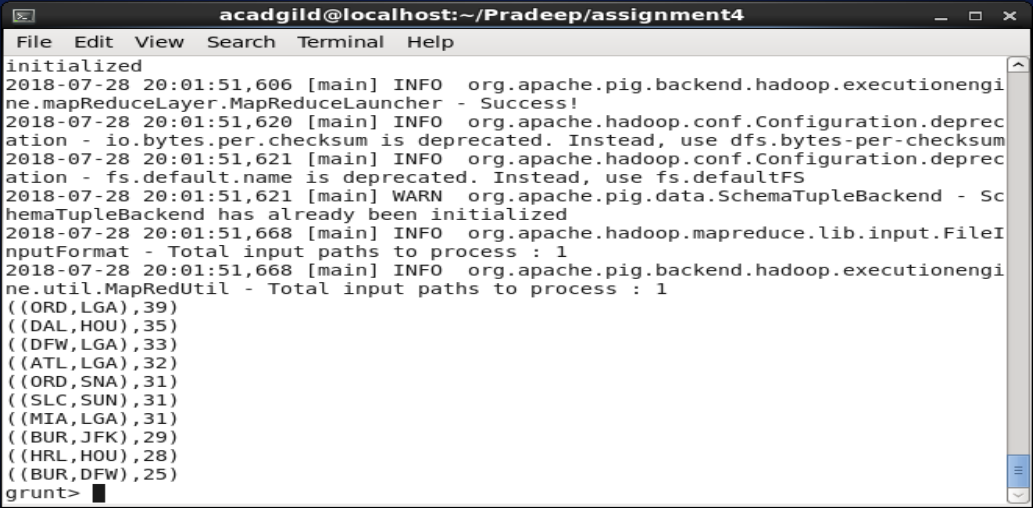
**Problem Statement 4**

Which route (origin & destination) has seen the maximum diversion?

PIG Script:



**OUPUT:**



**Task2. 1**

<https://drive.google.com/file/d/0Bxr27gVaXO5sa0JBamZXdkpYUFk/view?usp=sharing>

Create a database named 'custom'.

Create a table named temperature\_data inside custom having below fields:

1. date (mm-dd-yyyy) format

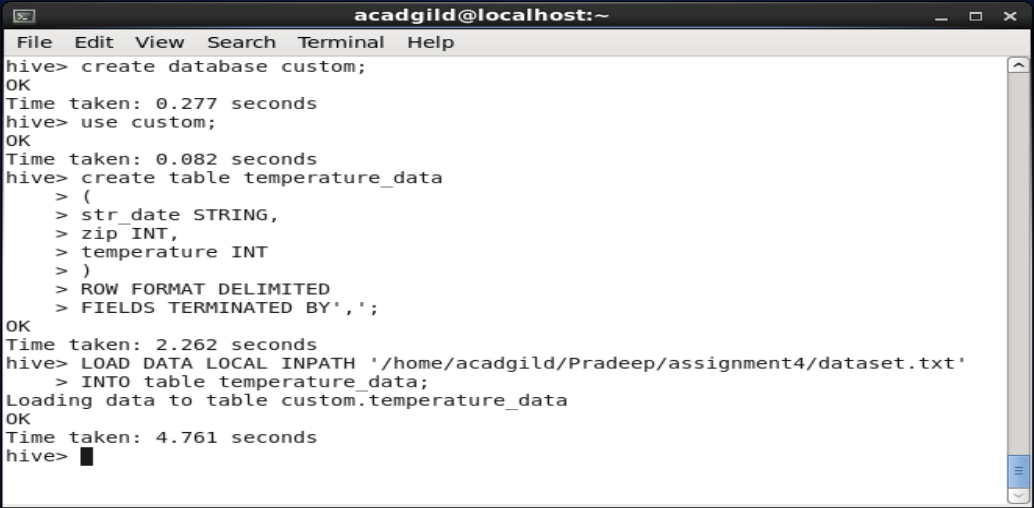
2. zip code

3. temperature

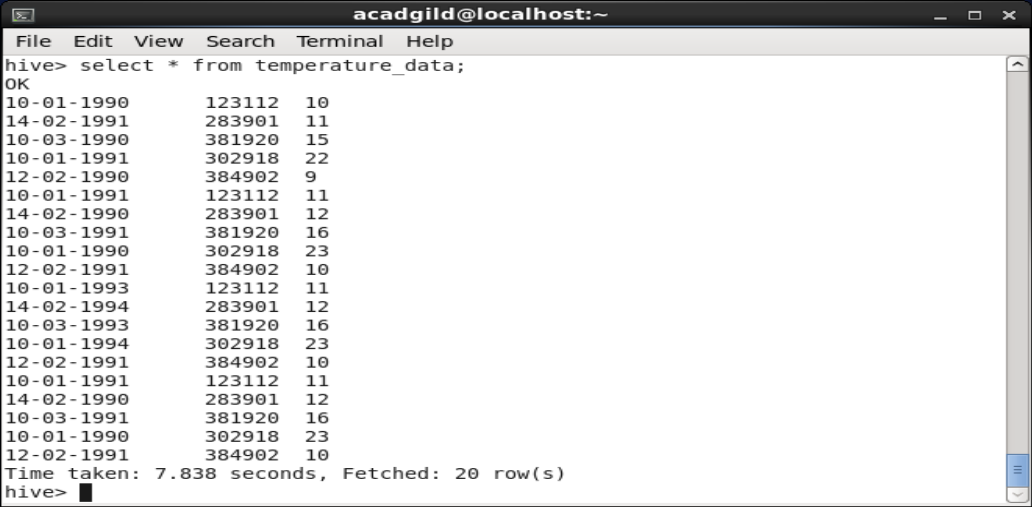
The table will be loaded from comma-delimited file.

Load the dataset.txt (which is ',' delimited) in the table.

HIVE Commands:



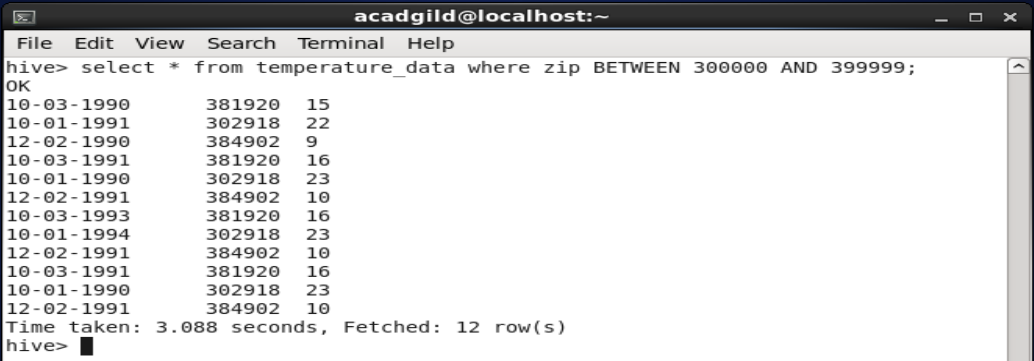
**OUPUT:**



**Task 2.2**

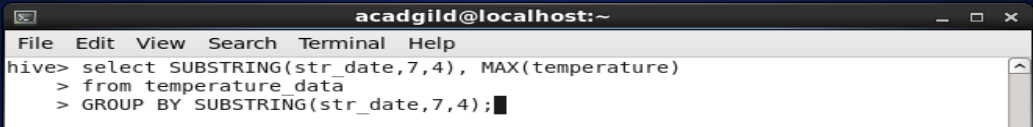
1. Fetch date and temperature from temperature\_data where zip code is greater than 300000 and less than 399999.

HIVE Command and **OUPUT:**

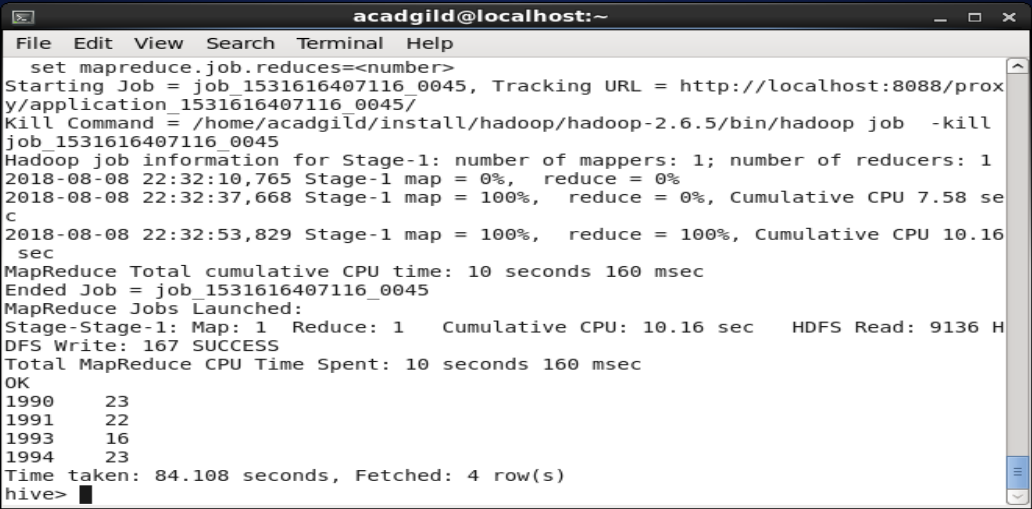


1. Calculate maximum temperature corresponding to every year from temperature\_data table.

HIVE Commands:

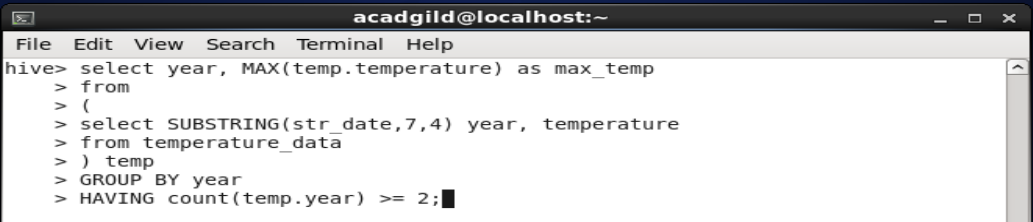


**OUPUT:**

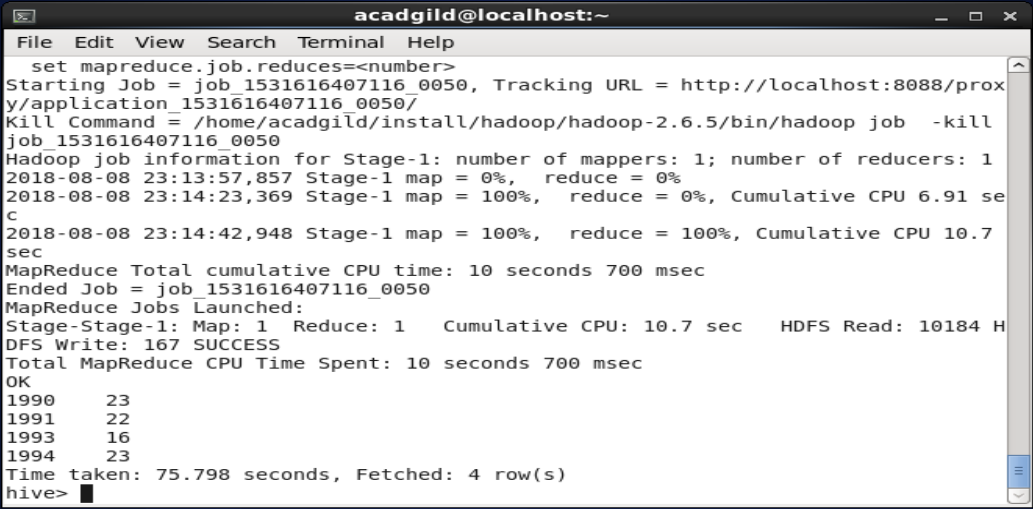
****

1. Calculate maximum temperature from temperature\_data table corresponding to those years which have at least 2 entries in the table.

HIVE Commands:

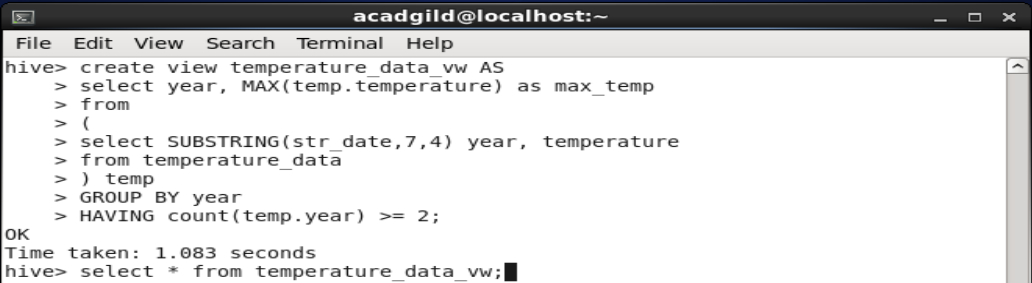


**OUPUT:**

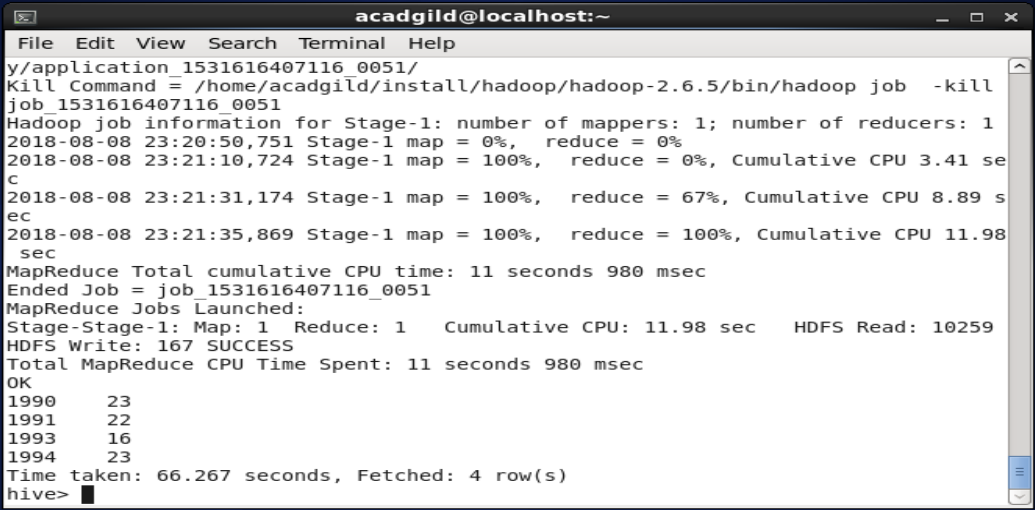
****

1. Create a view on the top of last query, name it temperature\_data\_vw.

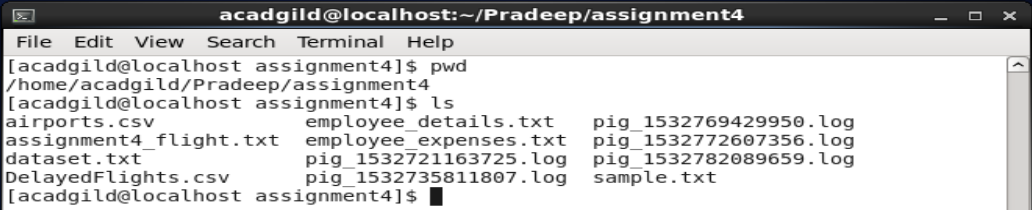
HIVE Commands:



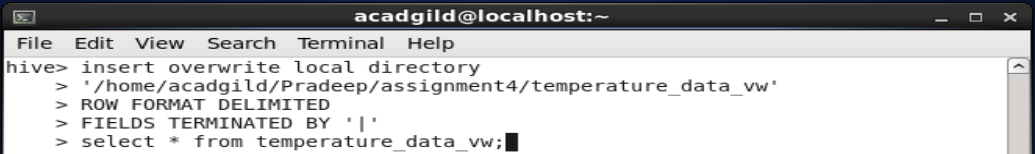
**OUPUT:**

****

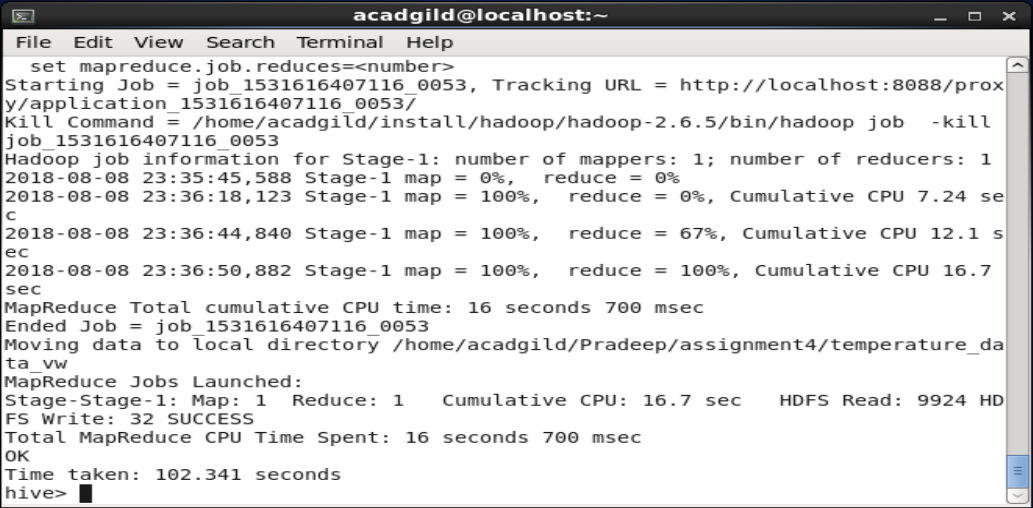
1. Export contents from temperature\_data\_vw to a file in local file system, such that each file is '|' delimited.

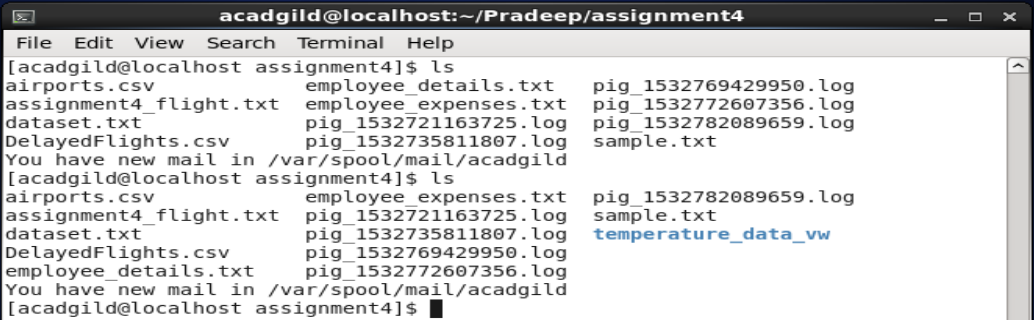


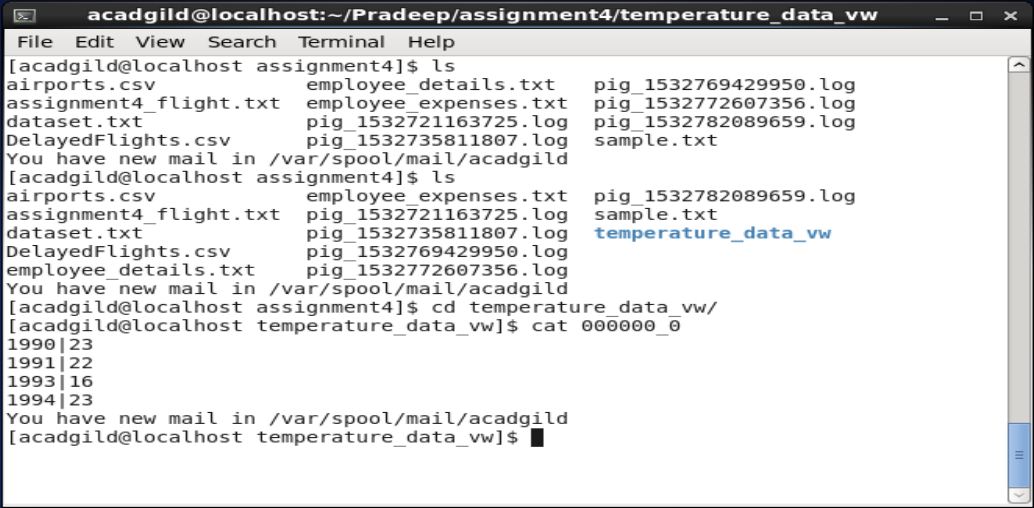
HIVE Commands:



**OUPUT:**

****

****

****