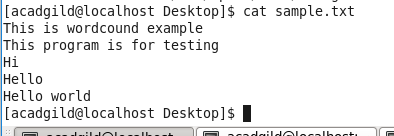
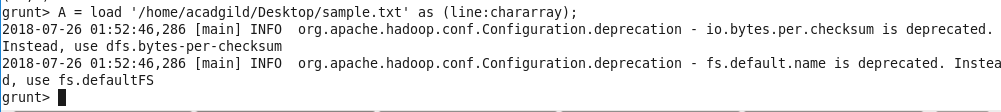
**Assignment 4(PIG and HIVE Basics)**

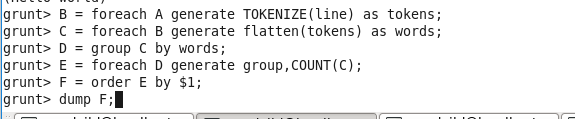
**Task 1.1**

Write a program to implement wordcount using Pig.

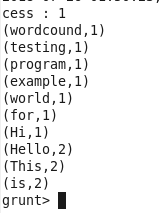
**Input File:-**







Result:-

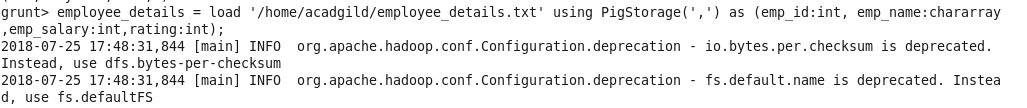


**Task 1.2**

(a) Top 5 employees (employee id and employee name) with highest rating. (In case two

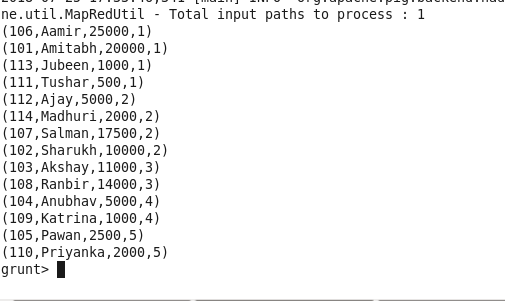
employees have same rating, employee with name coming first in dictionary should get

preference)





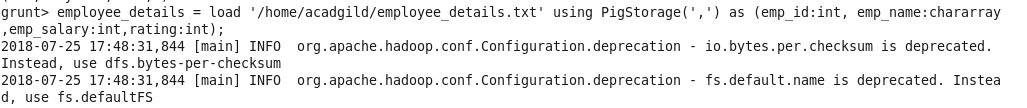
Result:-

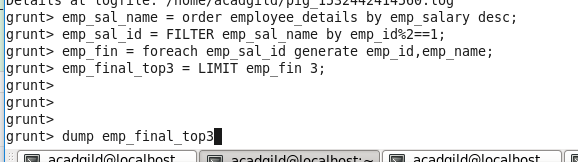


(b) Top 3 employees (employee id and employee name) with highest salary, whose employee id

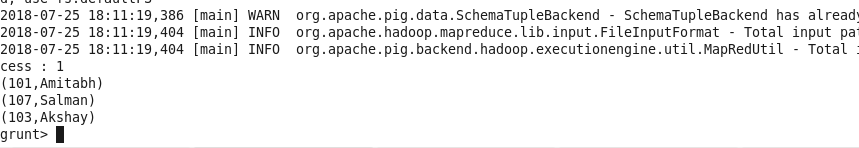
is an odd number. (In case two employees have same salary, employee with name coming first

in dictionary should get preference)



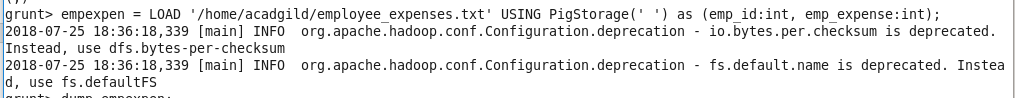


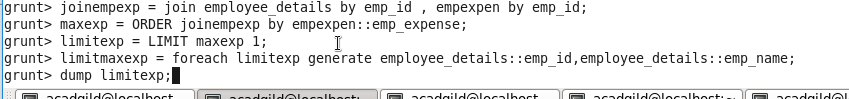
Result:-



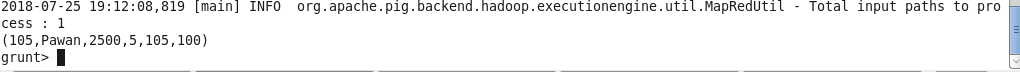
(c) Employee (employee id and employee name) with maximum expense (In case two employees have same expense, employee with name coming first in dictionary should get

preference)



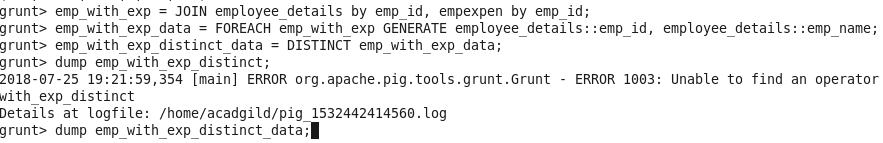


Result:-

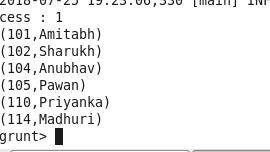


(d) List of employees (employee id and employee name) having entries in employee expenses

file.

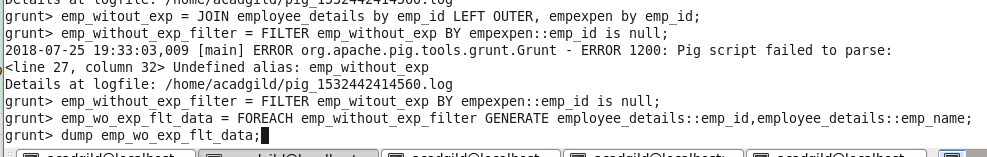


Result:-



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

file.



Result:-

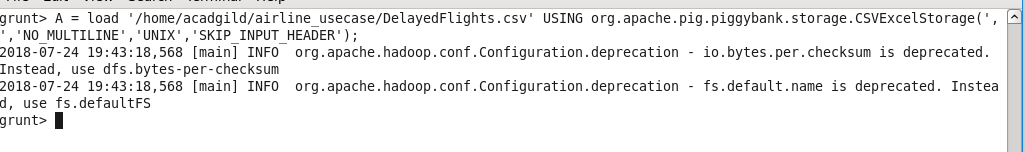


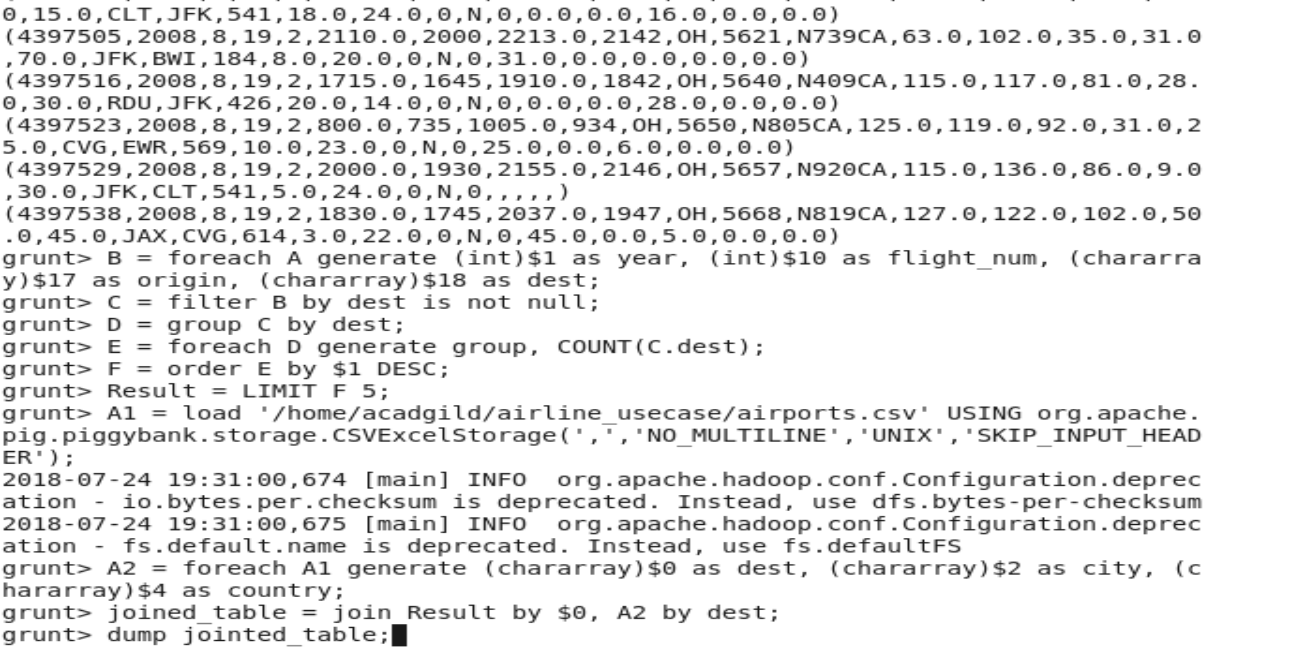
**Task 1.3**

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

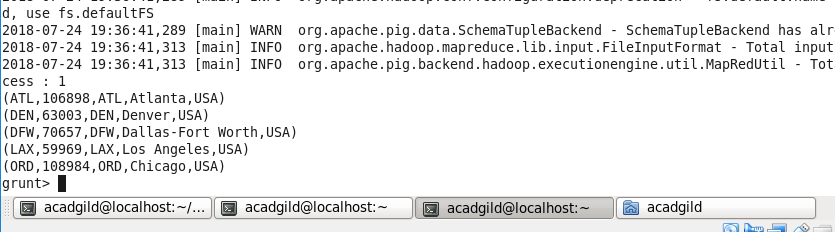
<https://acadgild.com/blog/aviation-data-analysis-using-apache-pig/>

**Problem Statement 1:**

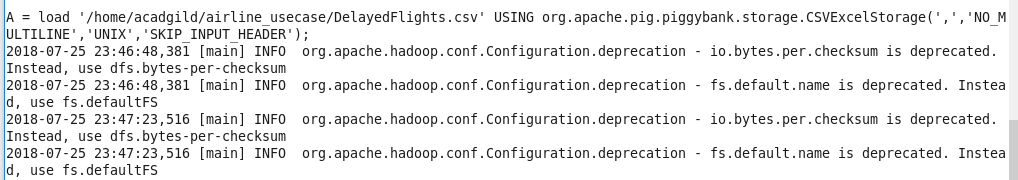


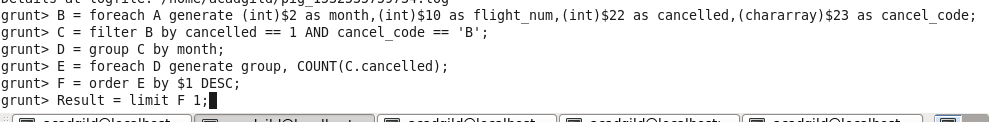


Result:-



**Problem Statement 2:**



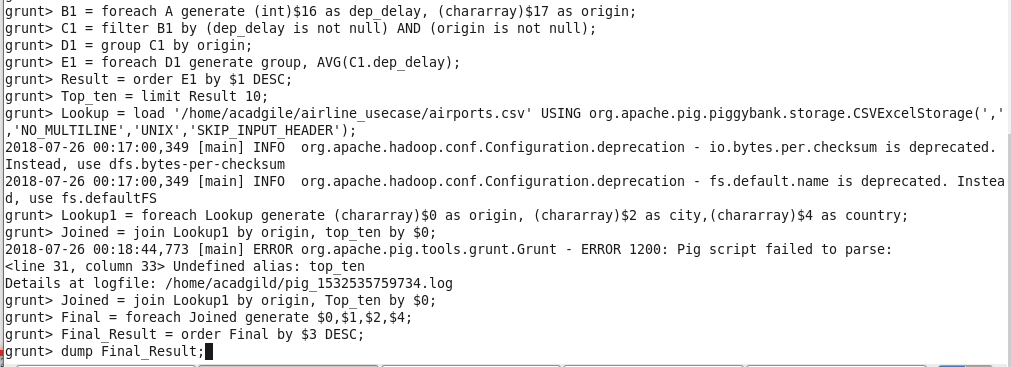


Grunt>dump Result;

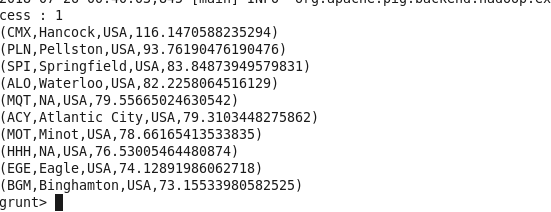
Result:-



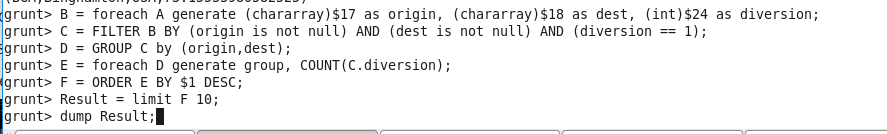
**Problem Statement 3:**



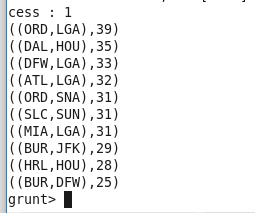
Result:-



**Problem Statement 4:**



Result:-



**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

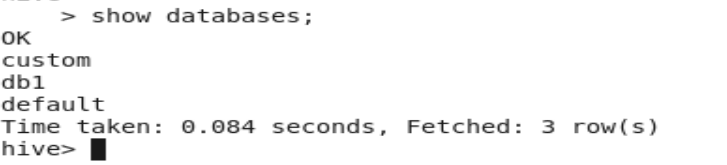
2. zip code

3. temperature

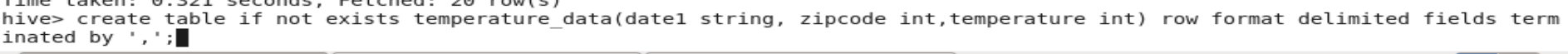
The table will be loaded from comma-delimited file.

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

* Custom data base created as shown below

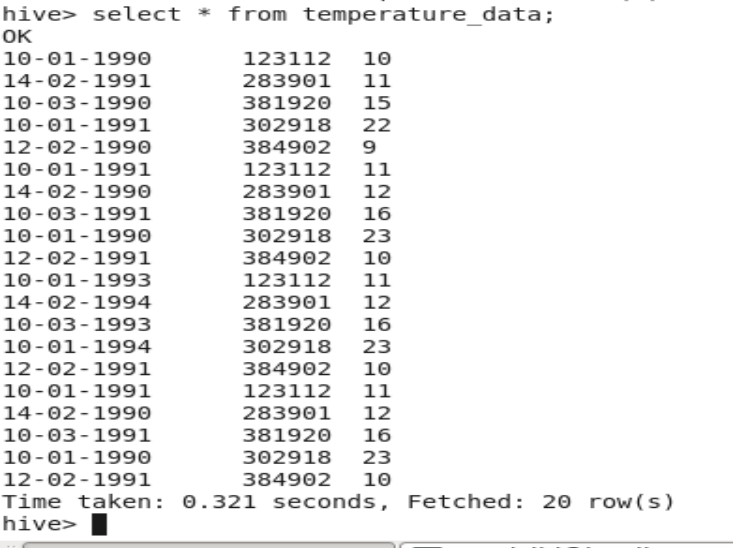


* Created the temperature\_data table



* Load the temperature\_data input file into the temperature\_data table

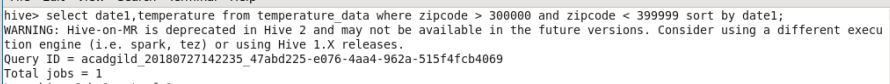




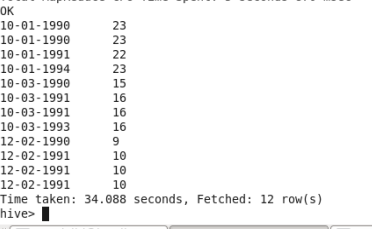
**Task 2.2**

* Fetch date and temperature from temperature data where zip code is greater than 300000 and less than 399999.

Hive Query:-

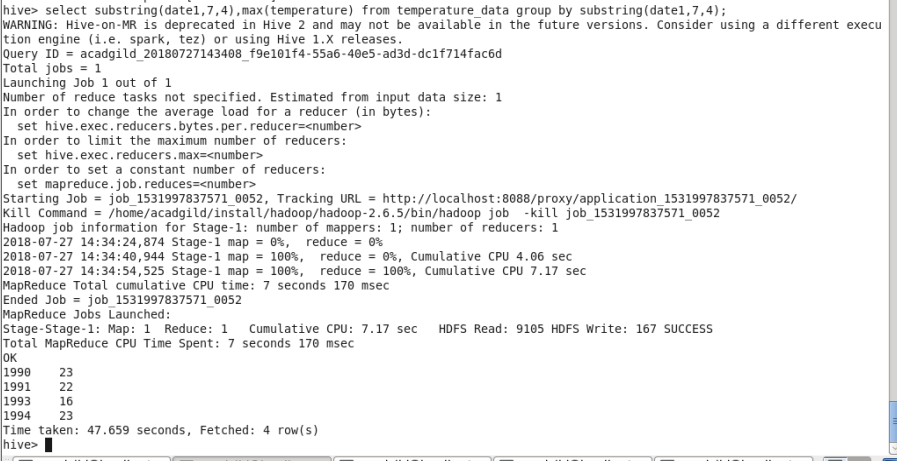


Result:-



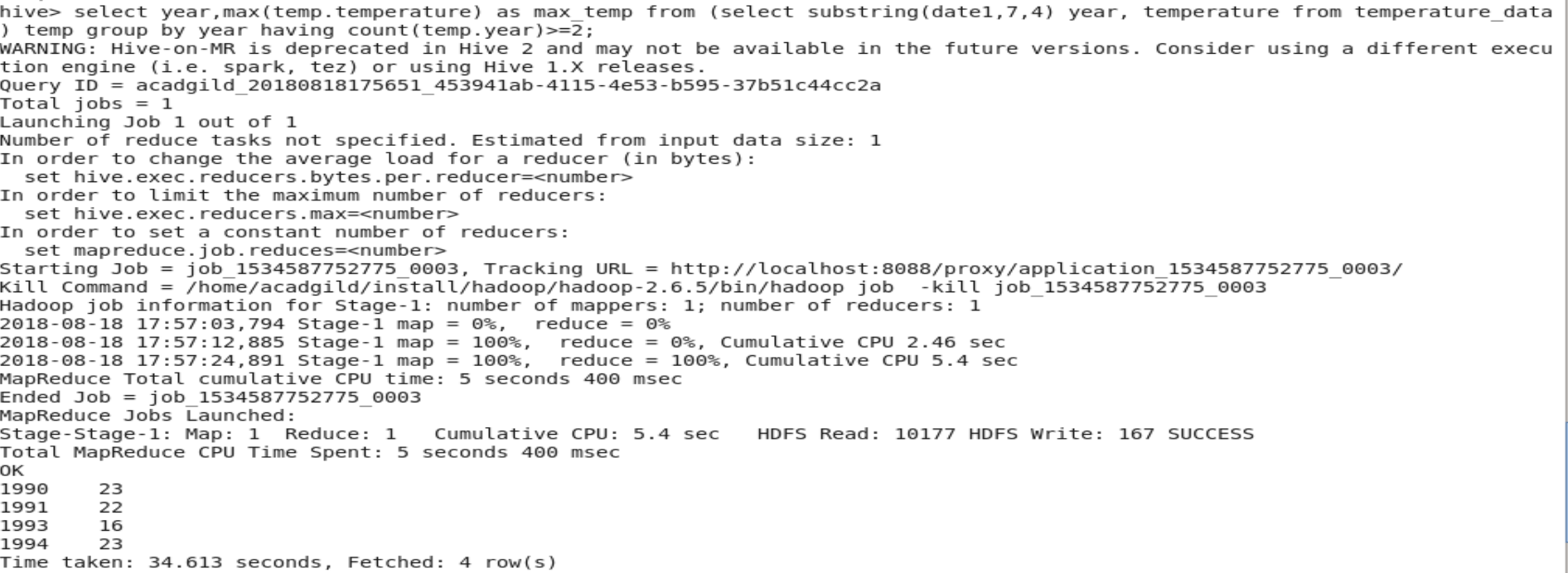
* Calculate maximum temperature corresponding to every year from temperature data table.

Hive Query and result:



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

Hive Query and result:



* Create a view on the top of last query, name it temperature\_data\_vw. Export contents from temperature\_data\_vw to a file in local file system, such that each file is '|' delimited.

View table result:-

