**ASSIGNMENT 25.4**

**Problem Statement:**

Data Definition:

Column 1: City pair (Combination of from and to): String

Column 2: From location: String

Column 3: To Location: String

Column 4: Product type: Integer (1=Air, 2=Car, 3 =Air+Car, 4 =Hotel, 5=Air+Hotel, 6=Hotel +Car, 7 =Air+Hotel+Car)

Column 5: Adults traveling: Integer

Column 6: Seniors traveling: Integer

Column 7: Children traveling: Integer

Column 8: Youth traveling: Integer

Column 9: Infant traveling: Integer

Column 10: Date of travel: String

Column 11: Time of travel: String

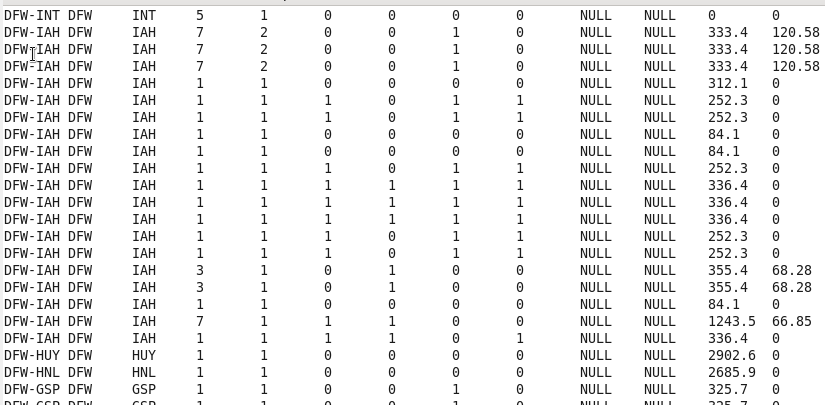
Column 12: Date of Return: String

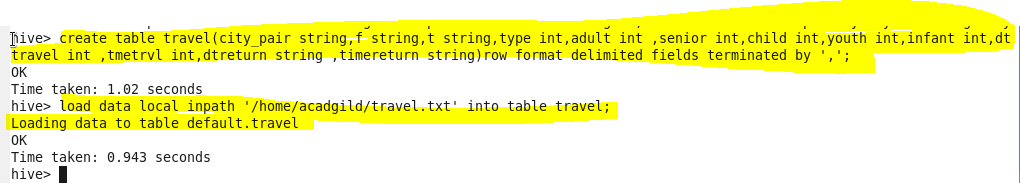
Column 13: Time of Return: String

Column 14: Price of booking: Float

Column 15: Hotel name: String

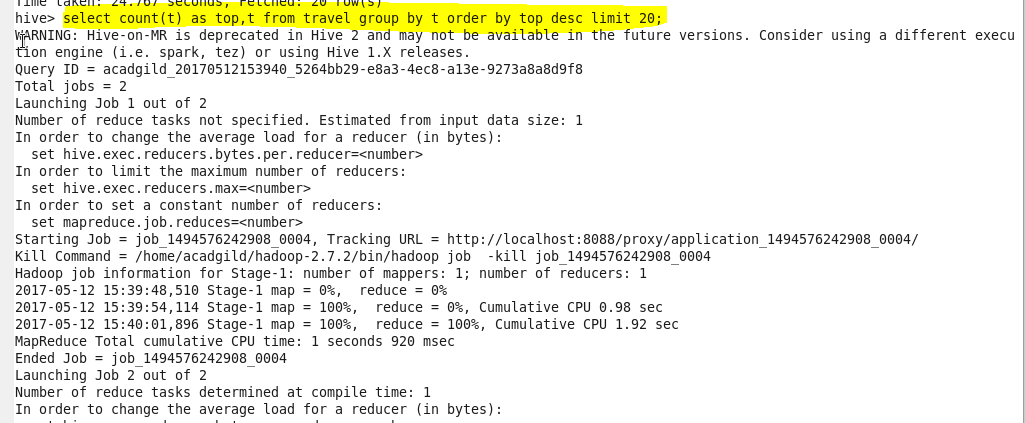
**Input File-**

****

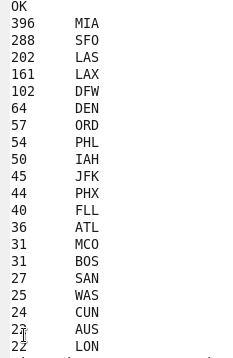
****

1. **Write a hive query to find the top 20 destination people travel the most: Based on the given data.**

* **Count is used to count the destinations.**
* **Group by is used to group the data on the basis of destinations.**
* **Order by is used to find the top most travelled destinations.**
* **Limit is used to display only 20 entries from top.**

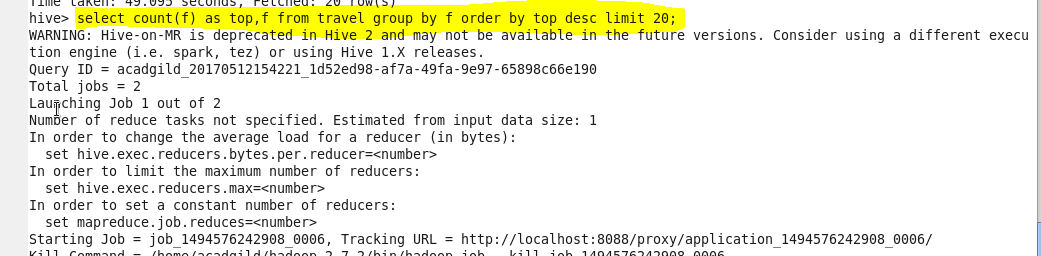
****

**Output-**

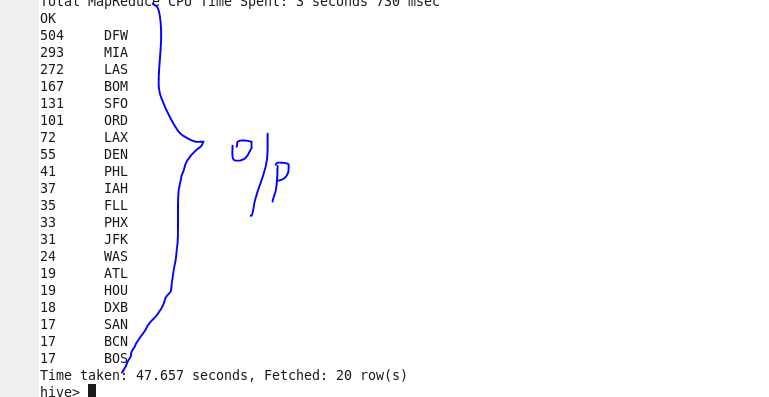
****

1. **Write a hive query to find the top 20 locations from where people travel the most.**

* **Count is used to count the source location.**
* **Group by is used to group the data on the basis of source location.**
* **Order by is used to find the top most source locations.**
* **Limit is used to display only 20 entries from top.**

****

**Output-**

****