**Project 1**:

**Problem description**:

The wages of the person should be updated with the current wage on daily execution of the csv file of employees using Hive.

**Datasets**:

1. The dataset resides on the linux file system and the path is /home/dell/data/hive\_data/datasets/
2. Data 1: emp\_data\_20230901.txt
3. Data 2: emp\_data\_20230902.txt

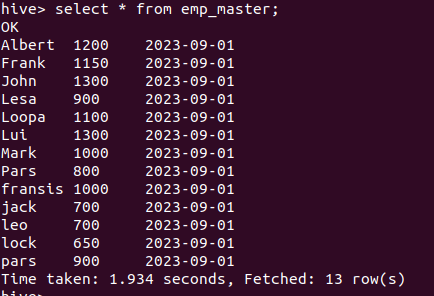
**Main files for execution**:

1. Hive query language file & shell script file resides on /home/dell/bin/
   1. Shell script file name (Main file): daily\_script.sh - This is the file responsible to execute .hql file.
   2. HQL file name: scd.hql

a) **do a select \* from emp\_master; [data should match with the code base project output]**

1. > sh daily\_script.sh 20230901

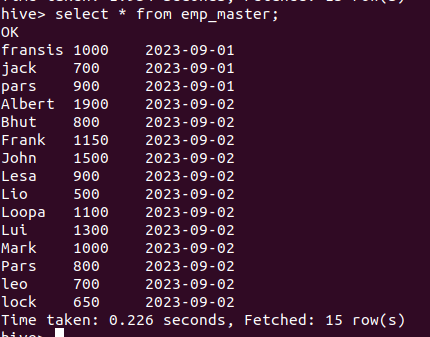
**Output**:



Can see all the records of employees on 2023-09-01, where there is no change in wages.

1. >sh daily\_script.sh 20230902

**Output**:



**Inference**: Here, after executing the file “emp\_data\_20230902.txt”, the existing employee named Albert & John has increase in their wages; lock, leo, Mark, Lui, Loopa, Lesa, Frank & Pars has maintained their wages; where as fransis, jack & pars have not worked on 2023-09-02.

b) **describe formatted emp\_master; [this should display your user id of table in the owner]**

1. hive >describe formatted emp\_master;

**Output**:

