**Hive Project**

**Hive Incremental Data Load**

**Objective:** Write a shell script for the daily updating of a table that contains current wage data of employees.

**Given Data files:** emp\_data\_20230901.txt, emp\_data\_20230902.txt, emp\_data\_20230903.txt,

**Created execution files:**

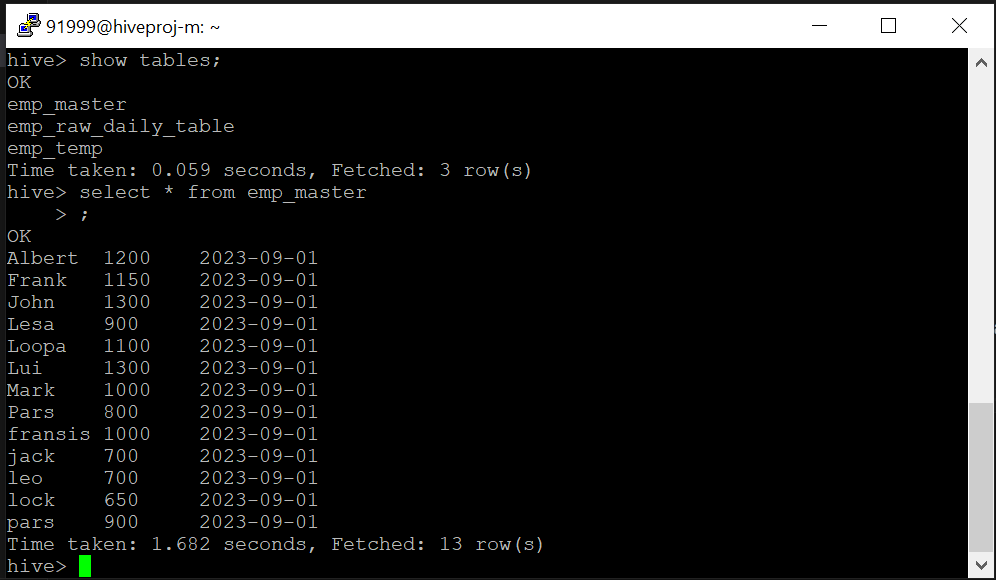
* **scd.hql** [consists of hive ql statements]
* **daily\_script.sh** [shell script required to execute the hql file]

**Steps:**

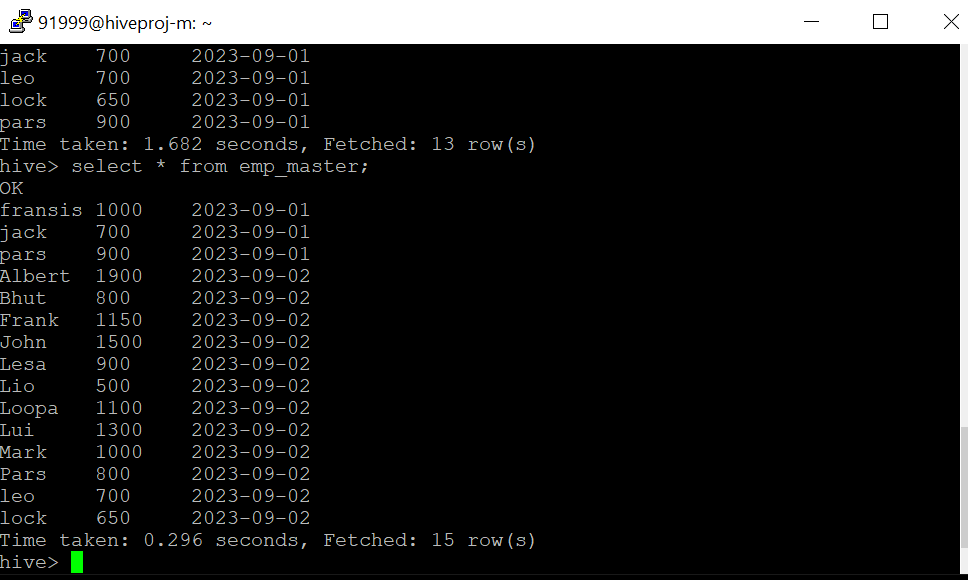
1. Loaded the 3 data files and scd.hql file from the local machine to the GCP Bucket.
2. Created bin and data folders inside the cluster VM which is a Linux machine.
3. Used Vim to write daily\_script.sh into the bin folder and copied scd.hql from GCP bucket to GCP cluster VM.
4. Copied the 3 data files from the GCP Bucket into the GCP Cluster VM.
5. Used the command **chmod -x daily\_script.sh** to give execution permissions.
6. Used the command **sh daily\_script.sh 20230901** to perform the data load.

**Results:**

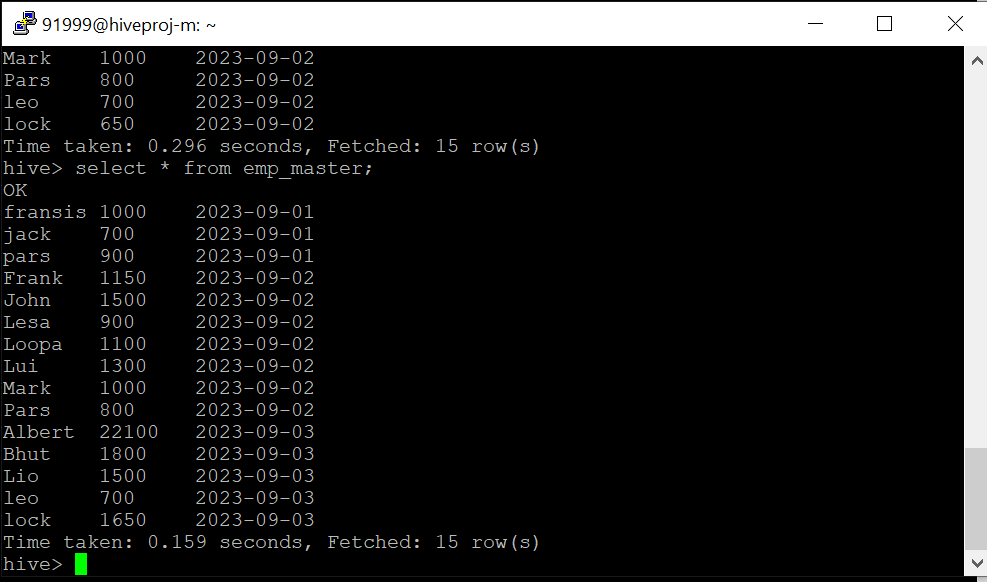
1. **sh daily\_script.sh 20230901**



1. **sh daily\_script.sh 20230902**



1. **sh daily\_script.sh 20230903**



1. **describe formatted emp\_master;**

