**CE066 | BDA LAB7**

**Aim :** "Crud operations and more using hive query language."

* In Hadoop Eco System, out of many, one more matured component is Hive. Hive is mainly used for performing SQL on data residing distributed.

Installed **apache-hive-3.1.2-bin** with **Hadoop-3.2.1** using WSL.

**Installation steps:**

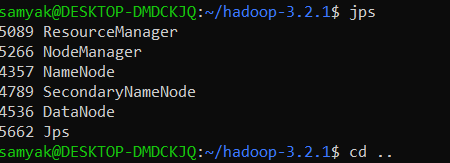
1. wget <http://archive.apache.org/dist/hive/hive-3.1.2/apache-hive-3.1.2-bin.tar.gz>
2. tar -xzf apache-hive-3.1.2-bin.tar.gz
3. Set environment variable
4. hdfs dfs -mkdir -p /user/hive/warehouse

hdfs dfs -mkdir /tmp

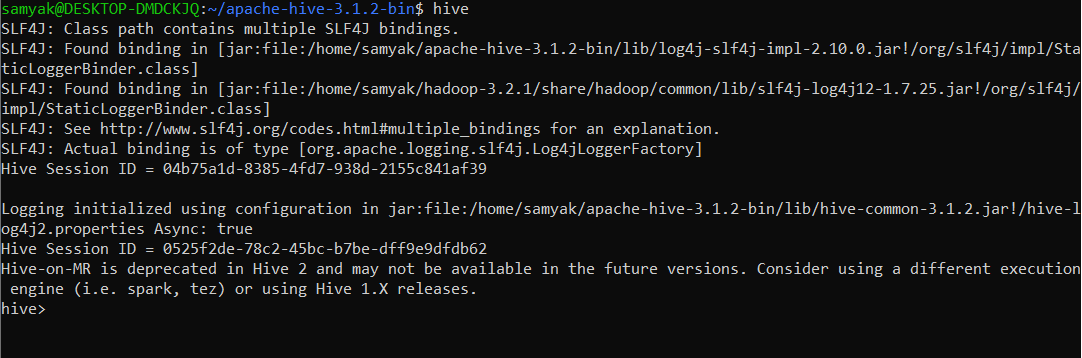
5)hdfs dfs -chmod g+w /user/hive/warehouse  
 hdfs dfs -chmod g+w /tmp

6) bin/schematool -initSchema -dbType derby

7) Launch **Hive.**



Started all Nodes!!



Launched Hive!

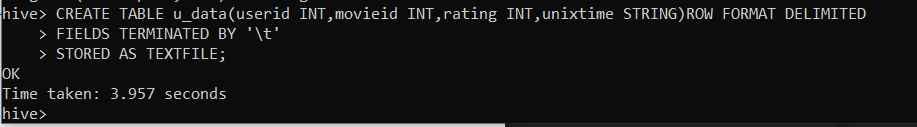
**Exercise:**

1. Retrieve the data set.

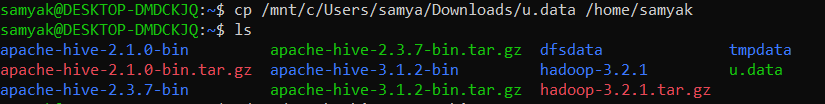
wget <http://files.grouplens.org/datasets/movielens/ml-100k.zip>

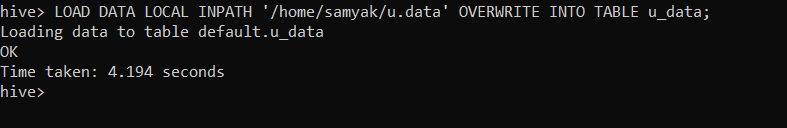
unzip ml-100k.zip

1. Run below HQL

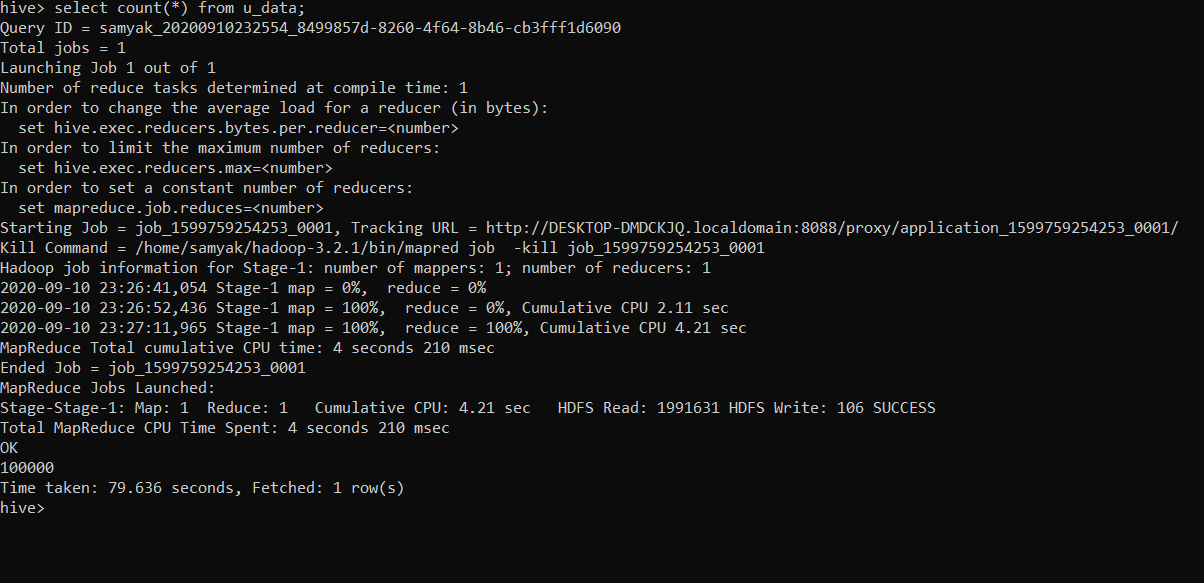


1. Load the data from local file u.data to table u\_data.





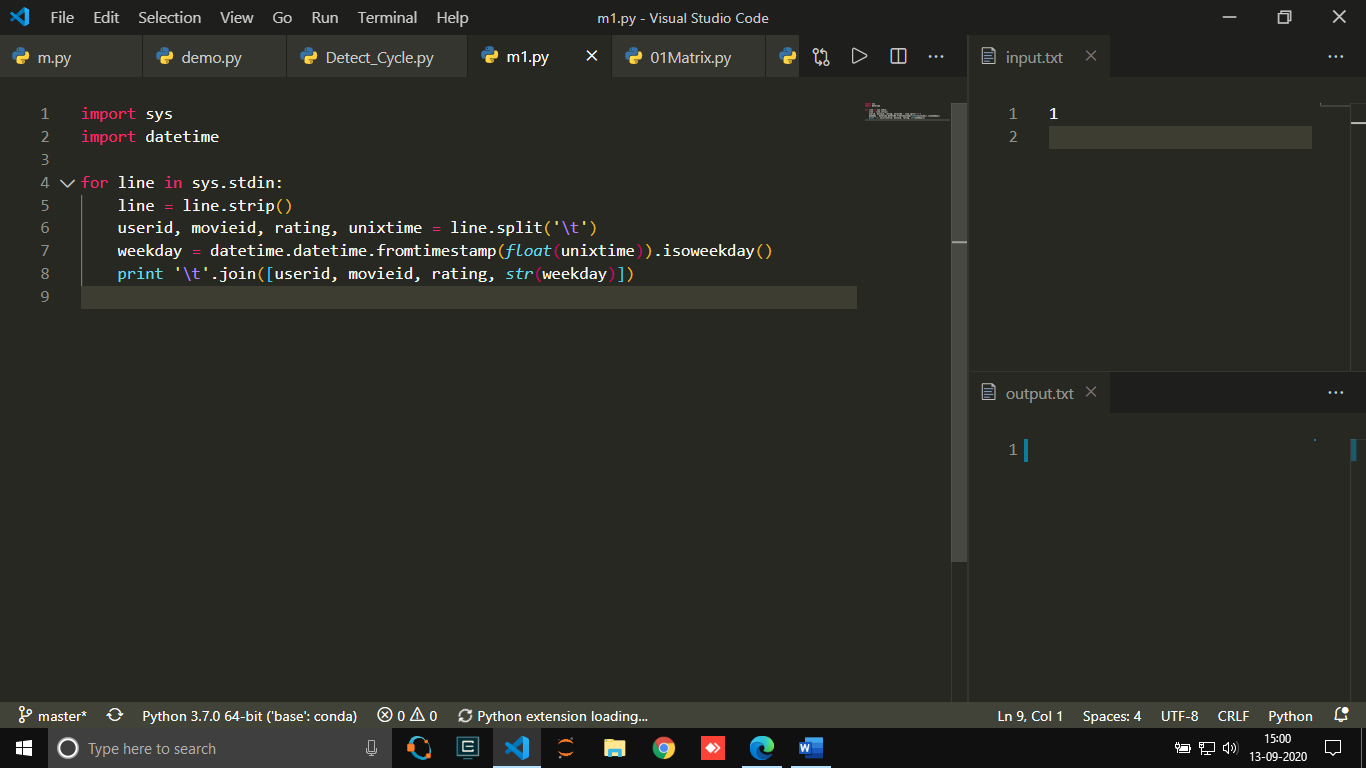
1. Find total number of records u\_data table.

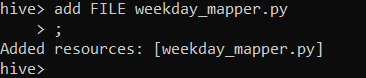


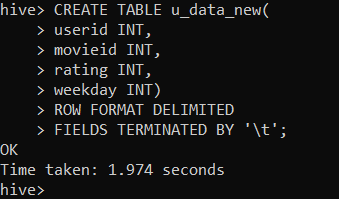
1. Find weekday and corresponding count group by weekday.

**Solution:**

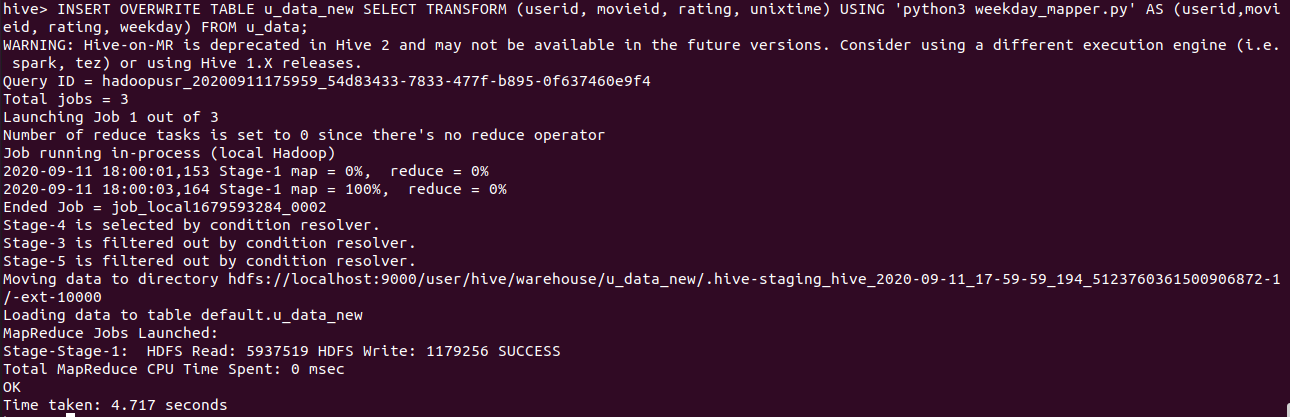
Create weekday\_mapper.py:



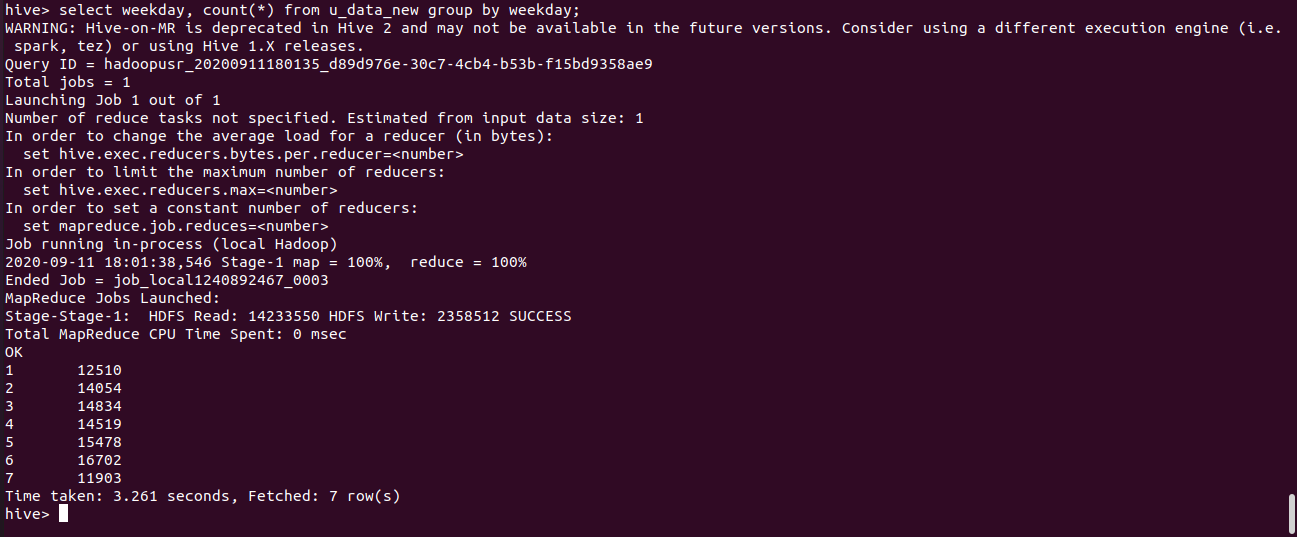




* Table u\_data\_new created!!!



* Table is overwritten from u\_data!!



* Weekday and Count is shown grouped by weekday!