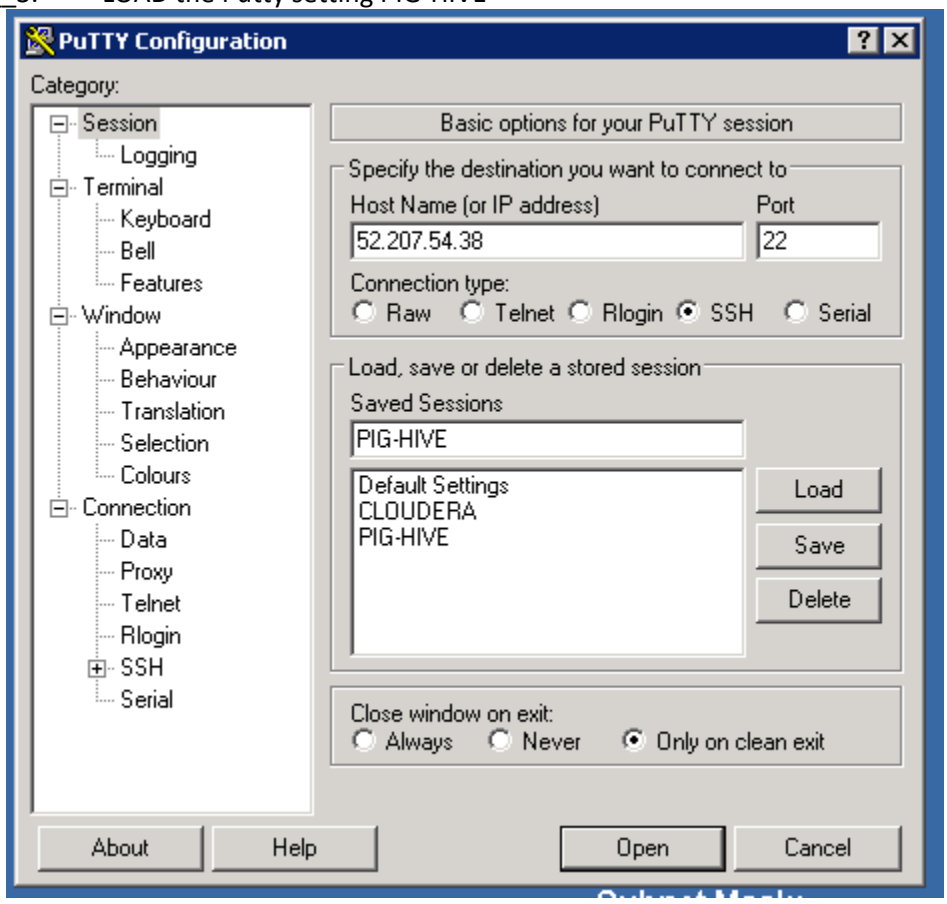
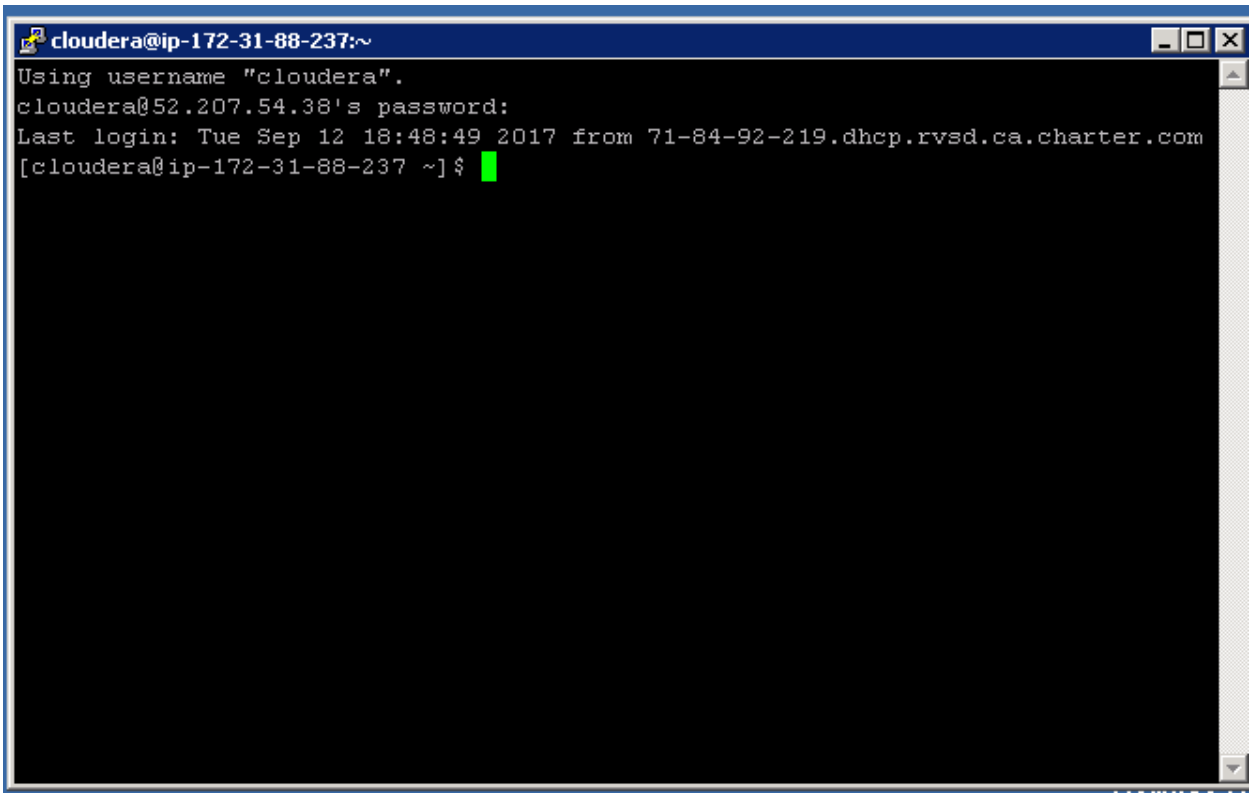
# Lab 2 – Introduction to Hive

In this lab we will demonstrate the use of Impala within the Linux shell and also within the HUE UI of Cloudera

____1. Go to the desktop of your Windows VM
____2. Double click on the Putty shortcut
____3. LOAD the Putty setting PIG-HIVE

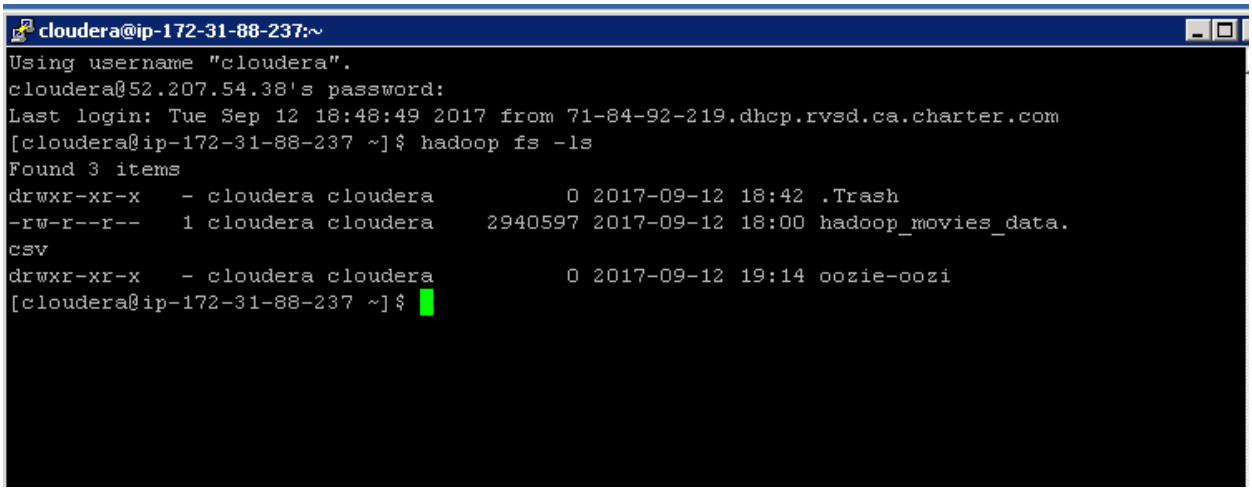



____4. The password to login to the Linux environment : cloudera

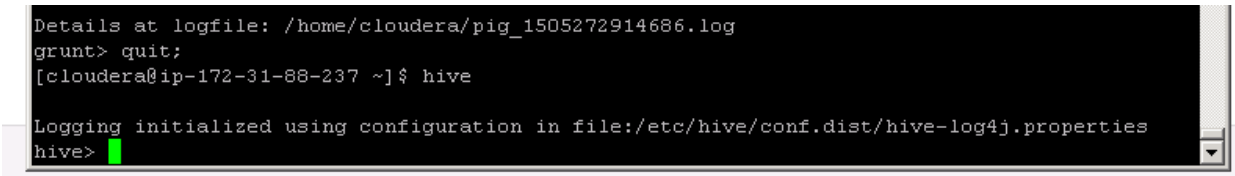



____5. The dataset we are using here is from Flickserv. Flickserv is a Netflix Search Engine. The dataset is a simple text(Hadoop_movies_data.csv) file that lists movie names and its details like release year, rating, and runtime. To view the dataset within the hdfs, please type :

$ hadoop fs -ls




____6. To start hive, type : $ hive

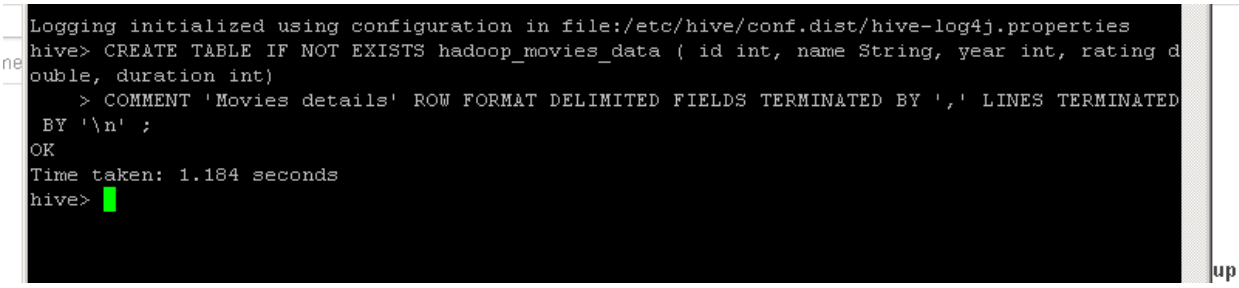



____7. Please type in:

CREATE TABLE IF NOT EXISTS hadoop_movies_data

( id int, name String, year int, rating double, duration int)

COMMENT 'Movies details' ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' STORED AS TEXTFILE;




____8. Generally, after creating a table in SQL, we can insert data using the Insert statement. But in Hive, we can insert data using the LOAD DATA statement.

While inserting data into Hive, it is better to use LOAD DATA to store bulk records. There are two ways to load data: one is from local file system and second is from Hadoop file system.

## Syntax

The syntax for load data is as follows:

```
LOAD DATA [LOCAL] INPATH 'filepath' [OVERWRITE] INTO TABLE tablename
[PARTITION (partcol1=val1, partcol2=val2 ...)]
```

- LOCAL is identifier to specify the local path. It is optional.
- OVERWRITE is optional to overwrite the data in the table.
- PARTITION is optional.

Check your File Browser in HUE to see if hadoop_movies_data.csv is located in /user/cloudera

If not you can upload the file, in the C:\movies folder on the Windows VM

The following query loads the given text into the table.

**hive> LOAD DATA INPATH 'hadoop_movies_data.csv' OVERWRITE INTO TABLE Hadoop_movies_data;**

Now the location of the file is located in the hdfs under the /user/hive/warehouse/

To verify exit out of hive with

hive> quit;

$ hadoop fs -ls /user/hive/warehouse

You should see the hadoop_movies_data folder

____9. Now we can re-enter hive

$hive

The metadata tables remain there in hive

____10. Let's run some queries

____11. The following query is executed to retrieve the employee details using the above table:

```
SELECT * FROM hadoop_movies_data WHERE rating > 4.0;
```

____12. ***List the movies that were released between 1950 and 1960***

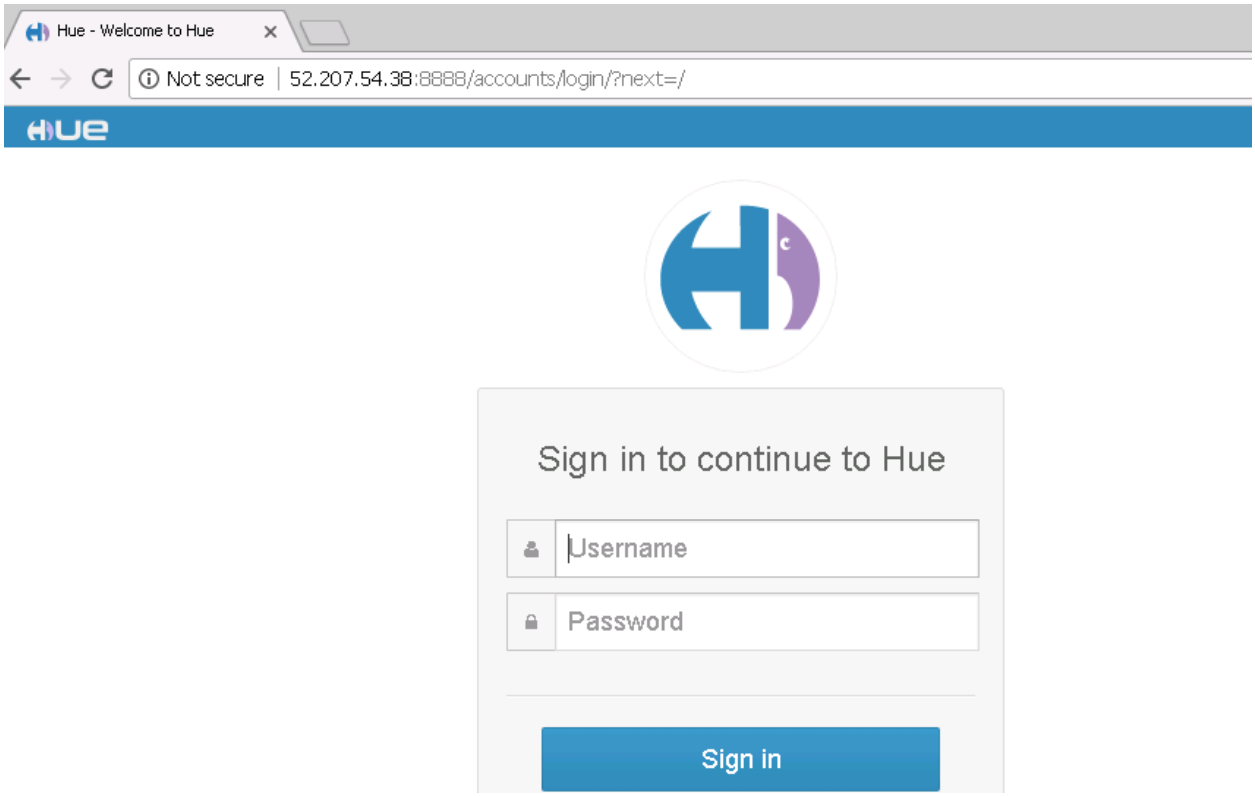
```
hive> SELECT * FROM hadoop_movies_data WHERE year > 1950 and year < 1960;
```

____13. Enter the following in a web browser:

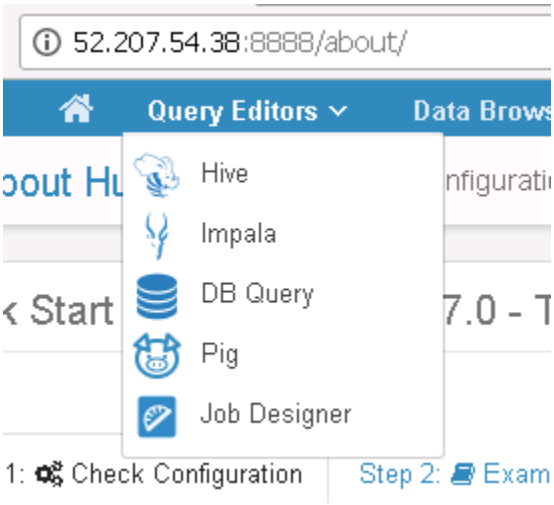
http://<ip-address supplied on desktop>>:8888

username: cloudera

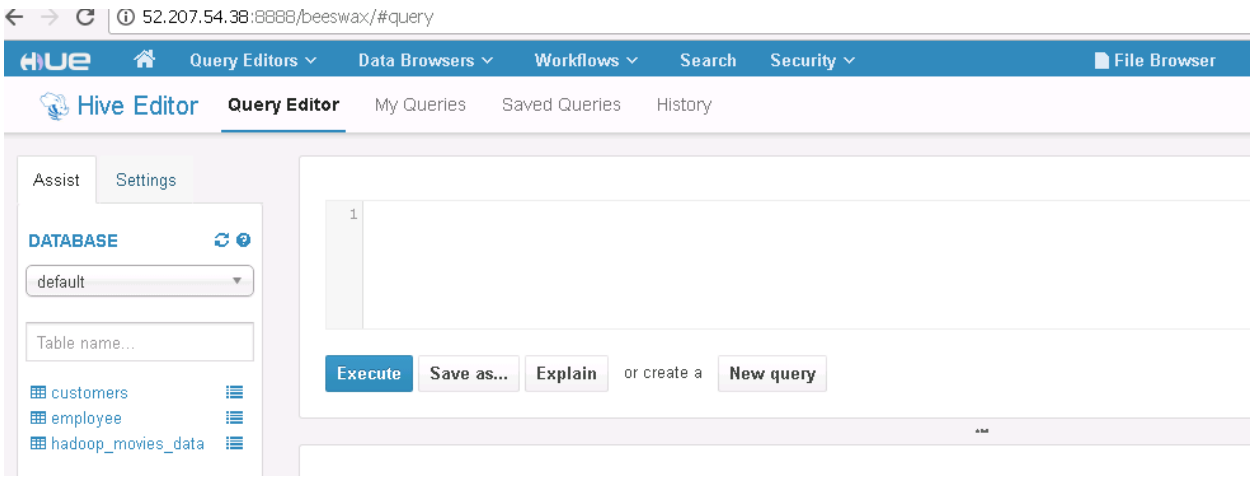
password: cloudera




____14. Go to Query Editors -> Hive



____15. You are now in the beeswax

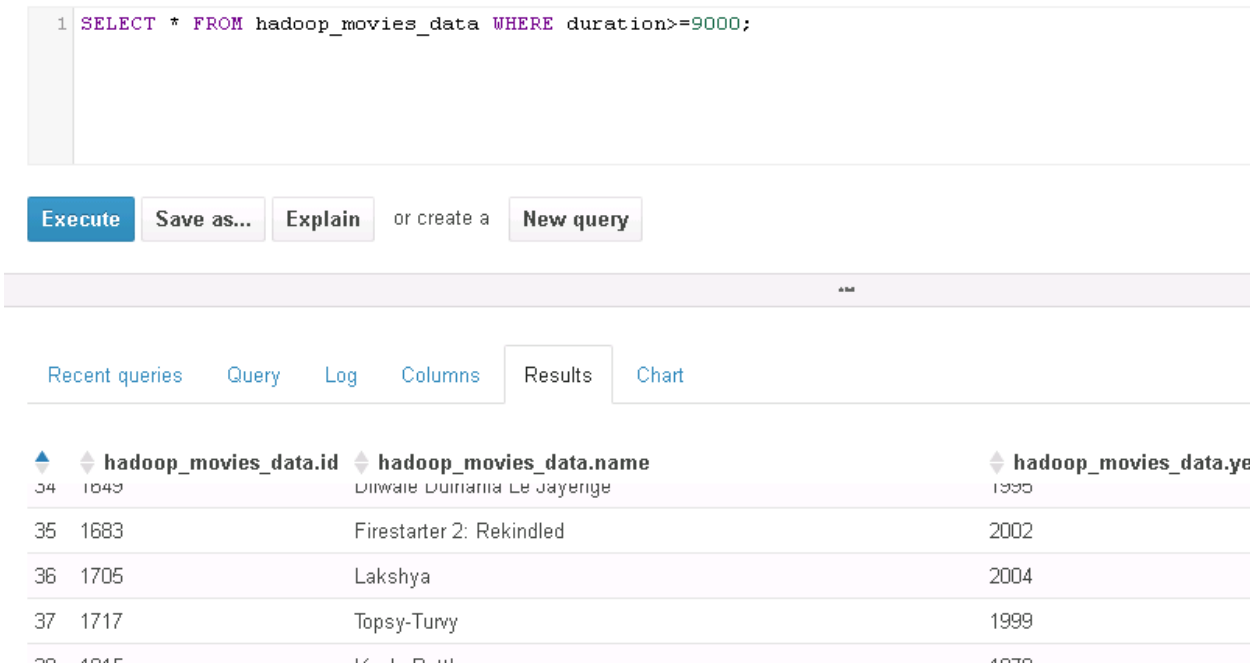



Take a look at the tables on the default database.

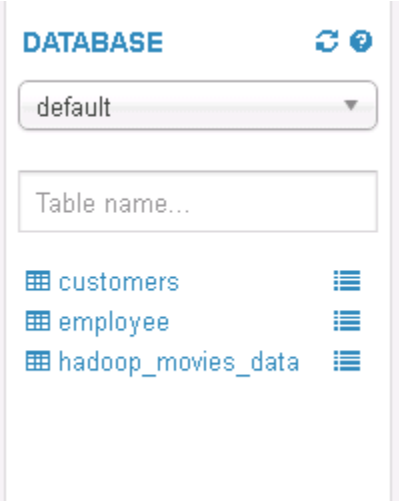
____16. Now enter a query within the Query Editor

```
SELECT * FROM hadoop_movies_data WHERE duration>=9000;
```

While processing, click on the Results Tab to see the results



____17. You can click on the tables on the left to see details




____18. Save your query as Query-movies, and then view your saved query by clicking on My Queries. You can also view the history there too.