R·I·T

Name: _____

Date: _____

# Lab: How to process data with Apache HIVE

**Introduction**

Hive is a component of Hortonworks Data Platform (HDP). Hive provides a SQL like interface to data stored in HDP. Hive provides a database query interface to Apache Hadoop.

## Lab Goals

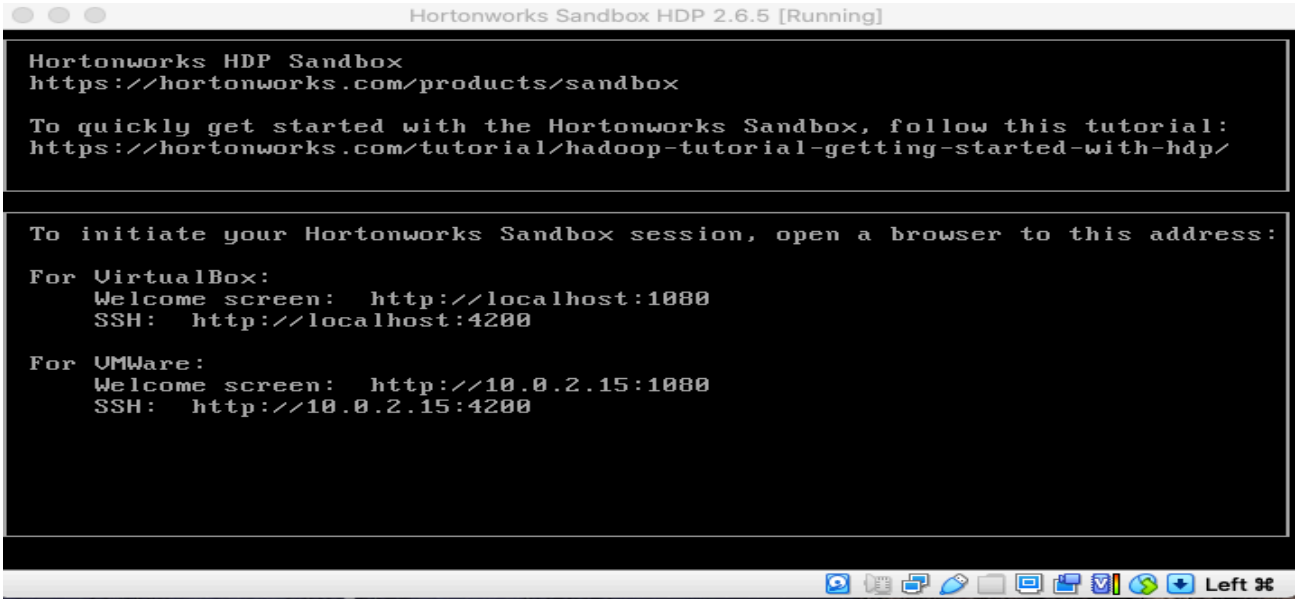
1. To be familiar with Hortonworks
2. Using the Hive SQL DB.
3. **Hortonworks version 2.6.5 is mandatory for this Lab**

## Activity 1 – Downloading the data

You can download the data files from MyCourses. The two files are Master.csv and Batting.csv.

## Activity 2 – Navigate to Hortonworks.

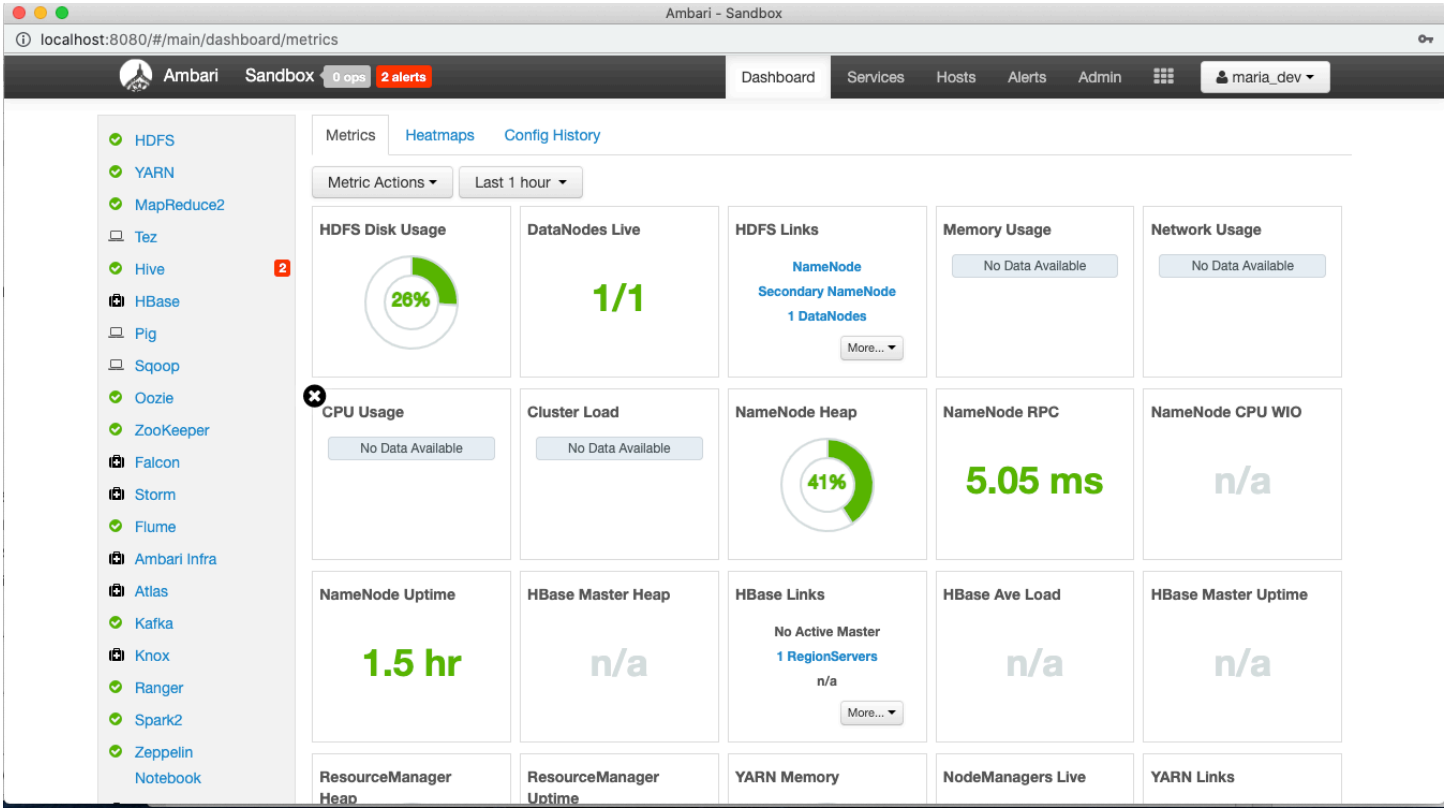
Start Hortonworks from Virtual Machine. The welcome screen is appeared as follows:




Open a browser with http://localhost:4200 and insert username and password **maria_dev.**

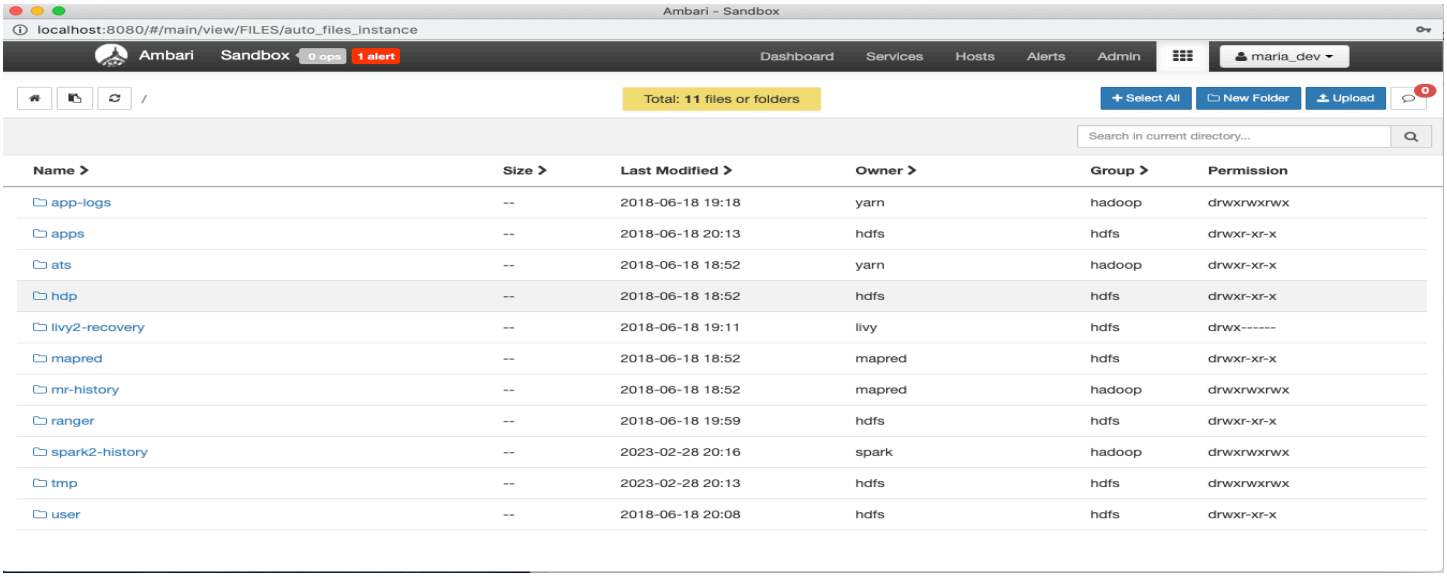
## Activity 3 – Starts HDP

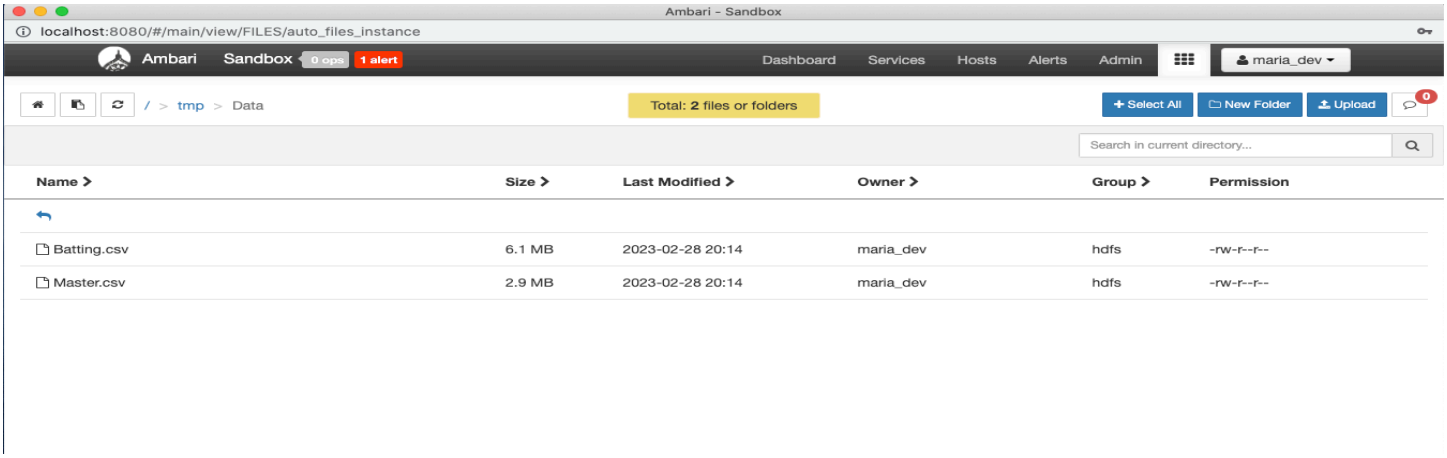
Open a web browser with localhost:1080 and select the HDP sandbox. After selecting HDP sandbox open a new page AMBARI portal asking user name and pwd : we insert both **maria_dev**




## Activity 4 – Uploading the data files

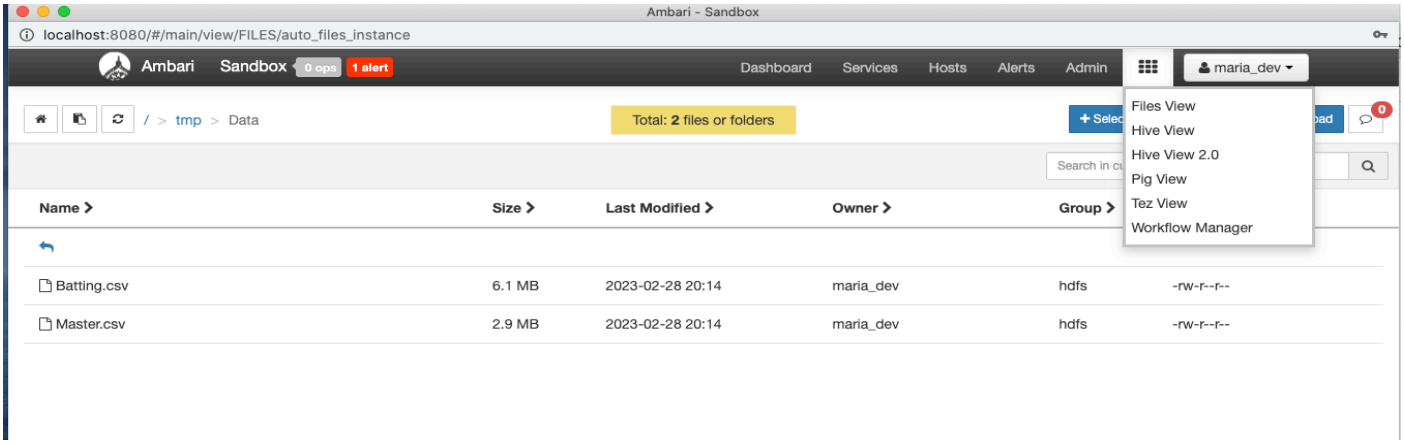
We start by selecting the ***HDFS Files view*** from the menu at the top. The store resides in the Hortonworks Sandbox VM. The HDFS Files view shows you the files in the HDP file store. Create a new folder in /tmp/Data to upload files. Navigate to where you stored the Batting.csv file on your local disk and select Batting.csv. Do the same thing for Master.csv . When you are done you will see there are two files in your directory.



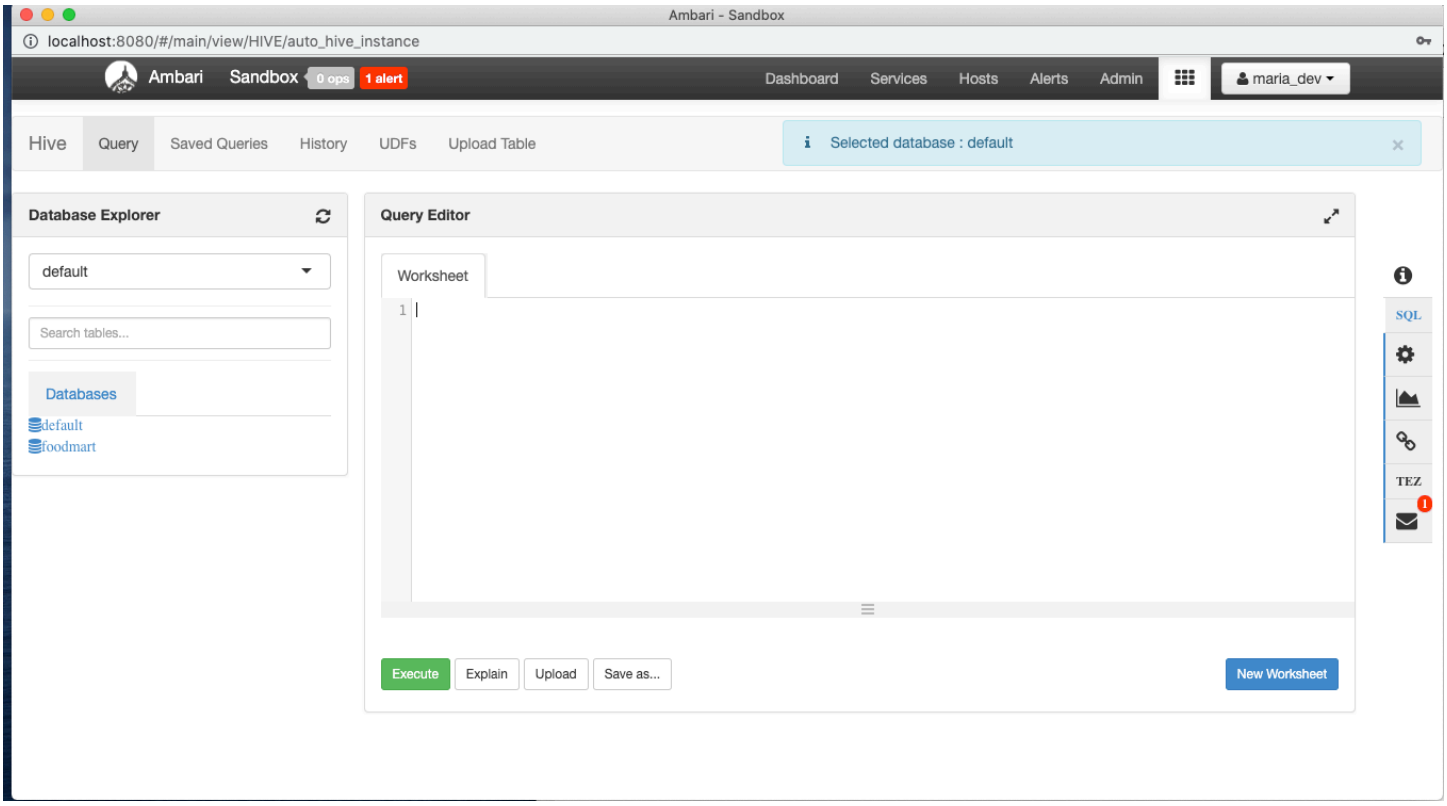

## Activity 5 – Open Hive Views

Let's open the by clicking on the Hive button in the top bar as previously when we selected the ***HDFS Files*** view. The Hive view provides a user interface to the Hive data warehouse system for Hadoop.



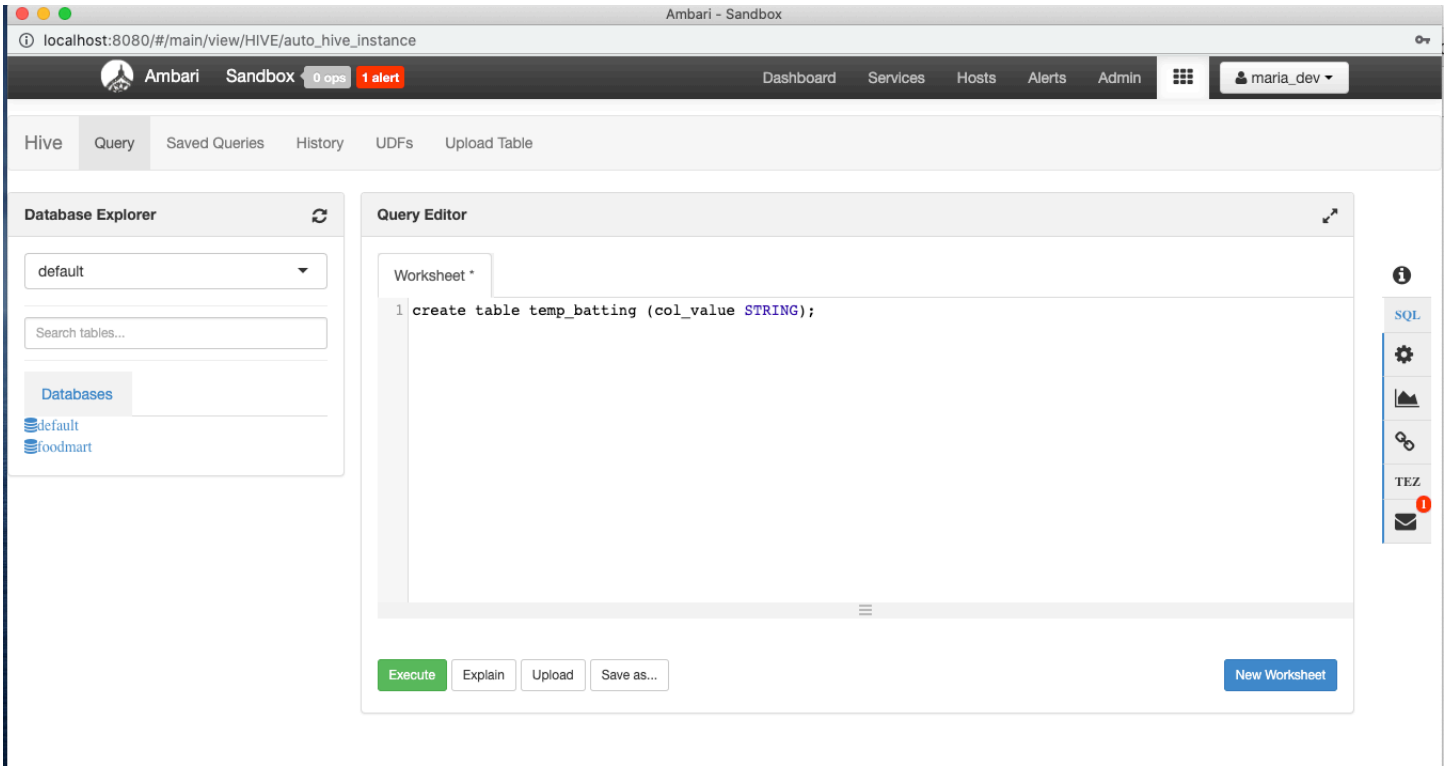

## Activity 6 – Open Hive Views

On right is a query editor. A query may span multiple lines. At the bottom there are buttons to Execute the query, Explain the query, Save the query with a name and to open a new Worksheet window for another query.



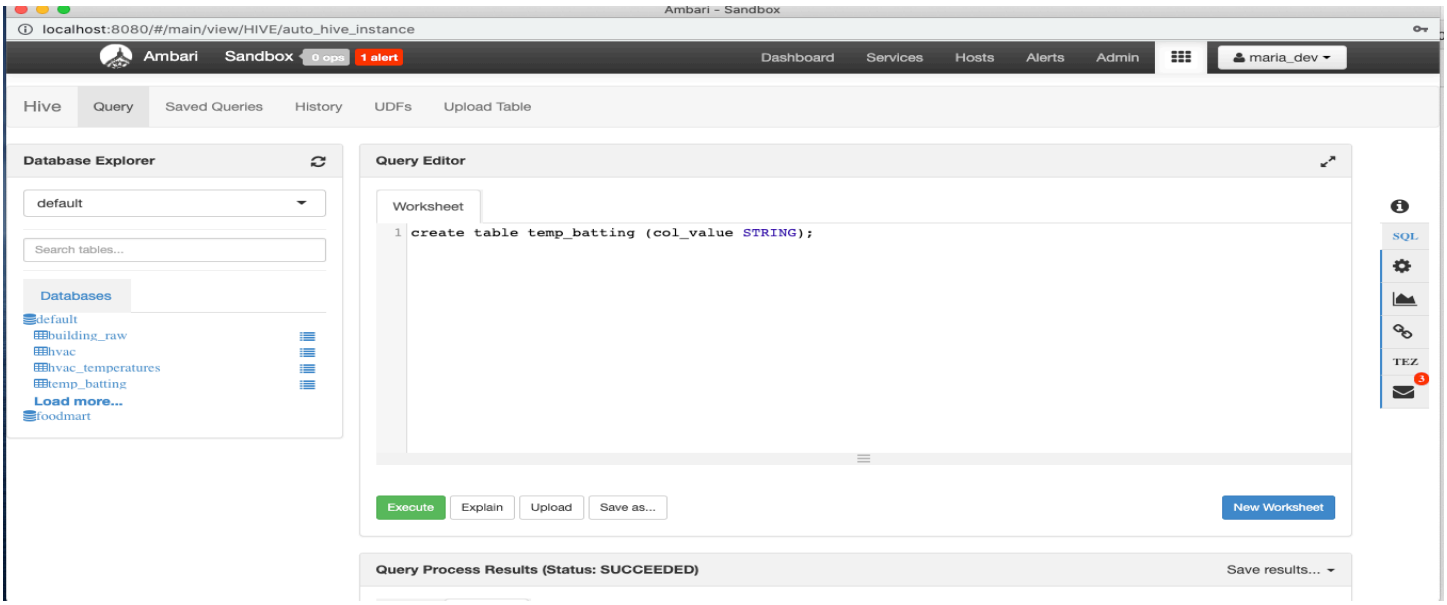
## Activity 7 – Create table

The first task we will do is create a table to hold the data. We will type the query into the composition area on the right hand-side. Once you have typed in the query hit the Execute button at the bottom



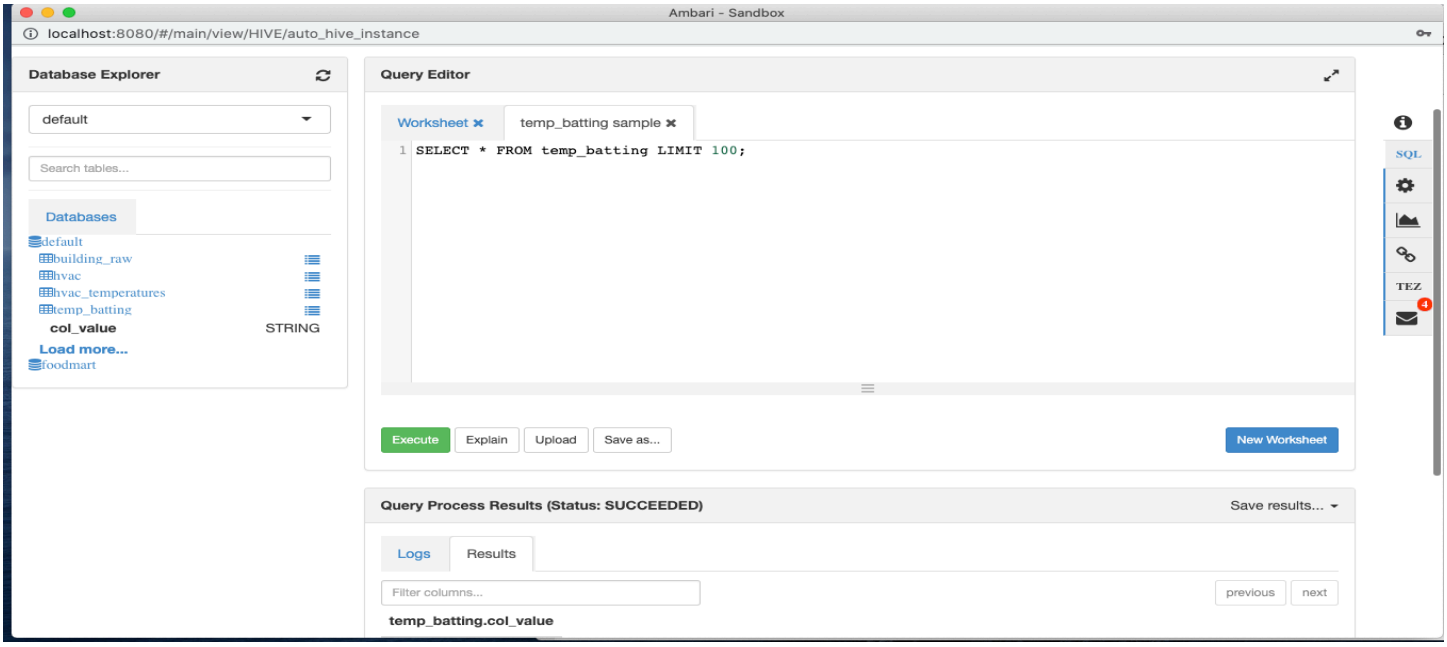

The query does not return any results because at this point we just created an empty table and we have not copied any data in it.

Once the query has executed we can refresh the Database Explorer at the left of the composition area and when folding down the default database we will see we have a new table called temp_batting .



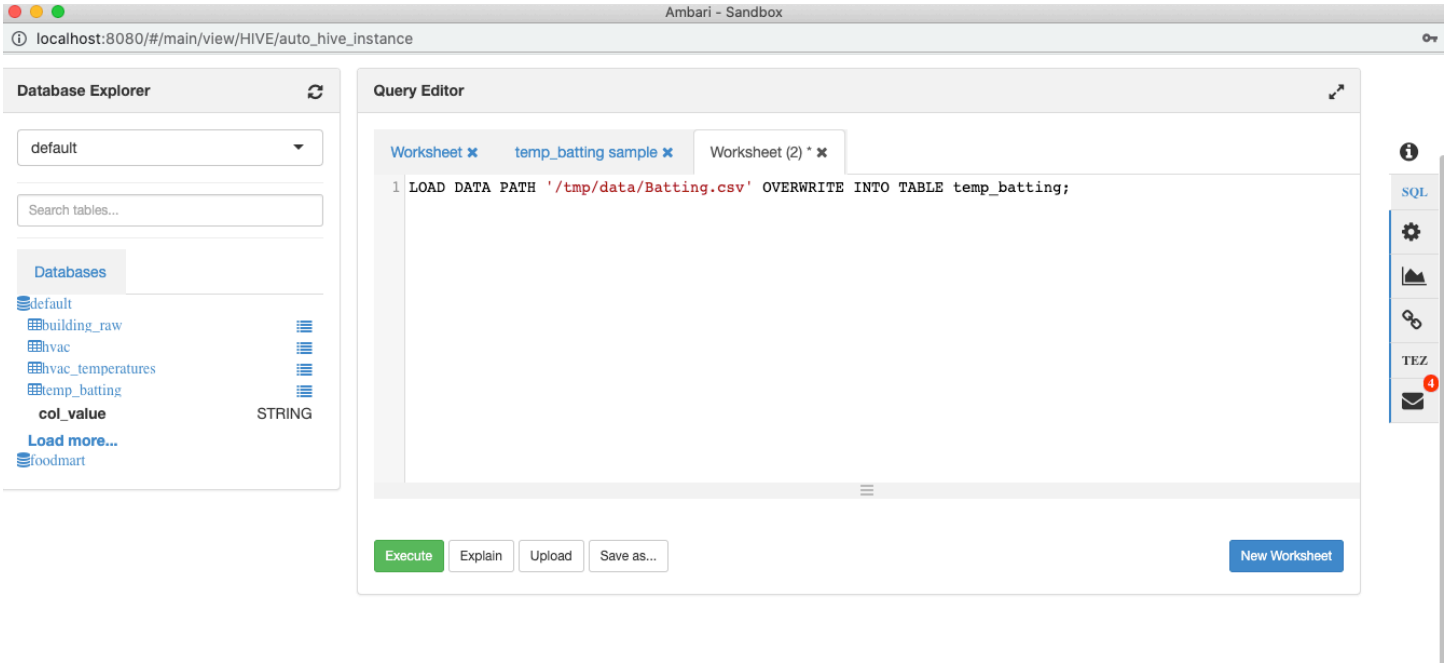

## Activity 8 – Check the sample data

Clicking on the icon next to the table name a new Worksheets opens, which loads sample data from this table. We see the table is empty right now. This is a good example of the interactive feel you get with using Hive.

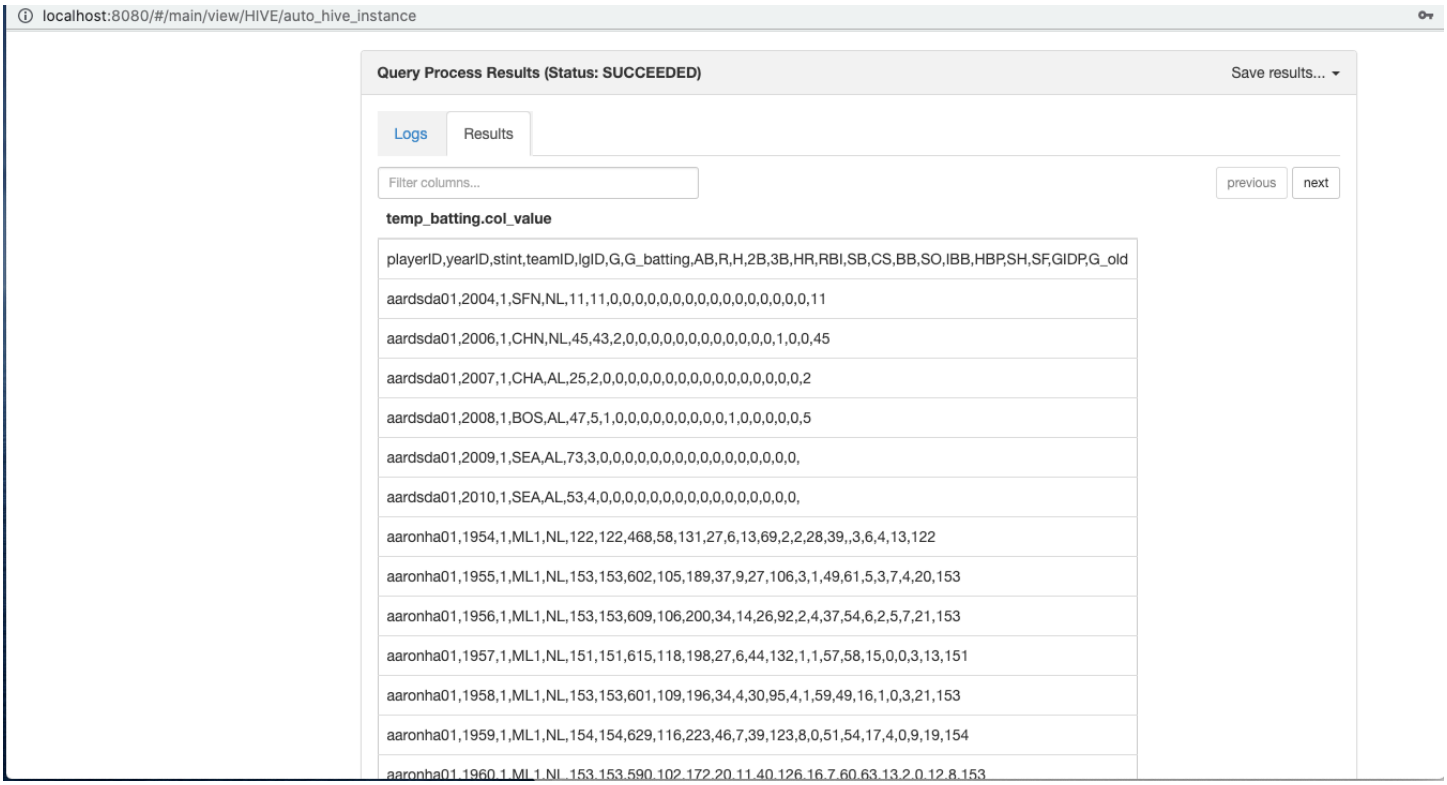


## Activity 9 – Load data

The next line of code will load the data file Batting.csv into the table temp_batting . After executing the query we can look at the Tables again and when we browse the data for temp_batting we see that the data has been read in.

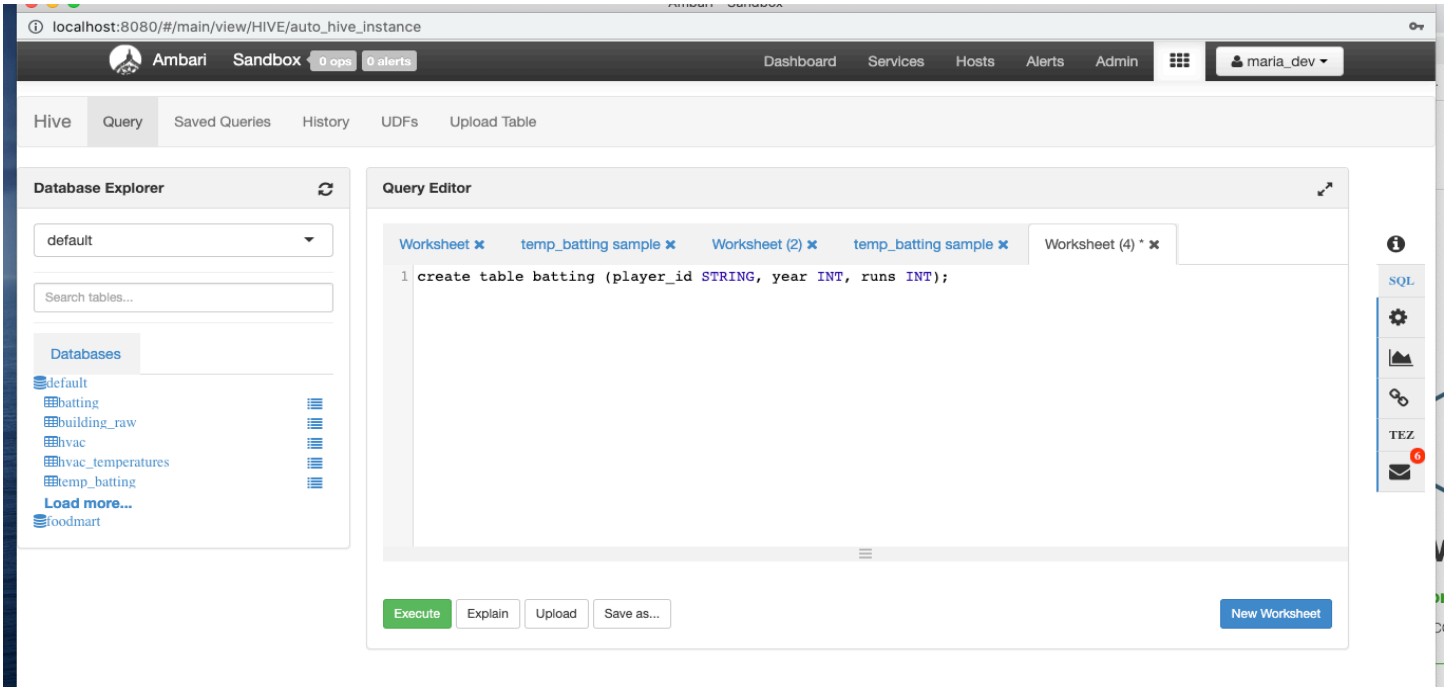


## Activity 10 – Sample data



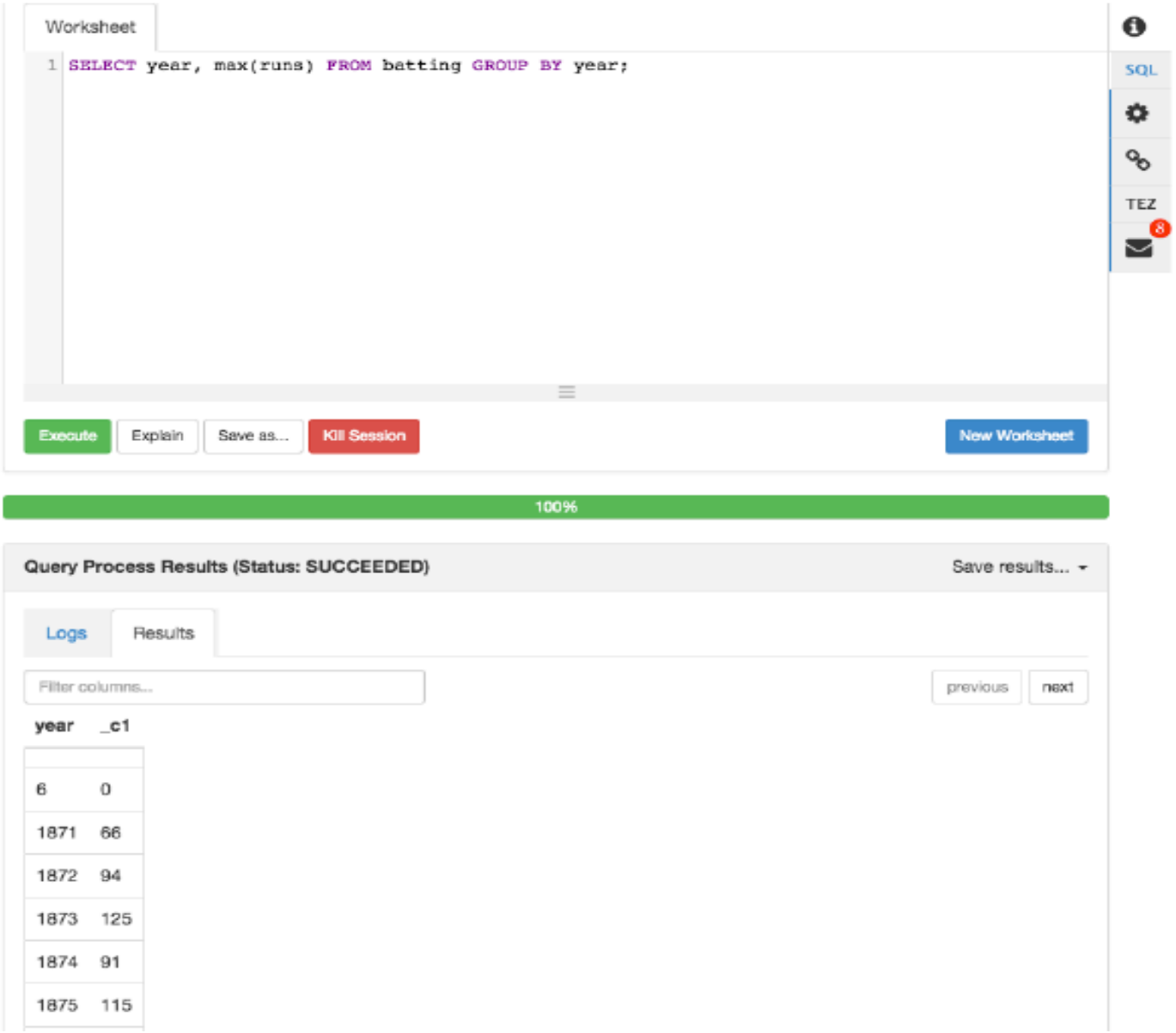

## Activity 11 – Sample data

Now that we have read the data in, we can start working with it. The next thing we want to do extract the data. So first we will type in a query to create a new table called batting to hold the data. That table will have three columns for player_id, year and the number of runs.


localhost:8080/#/main/view/HIVE/auto_hive_instance
Ambari
Sandbox
0 ops
0 alerts
Dashboard
Services
Hosts
Alerts
Admin
maria_dev
Hive
Query
Saved Queries
History
UDFs
Upload Table
Database Explorer
default
Search tables...
Databases
default
batting
building_raw
hvac
hvac_temperatures
temp_batting
Load more...
foodmart
Query Editor
Worksheet
temp_batting sample
Worksheet (2)
temp_batting sample
Worksheet (4) *
1 create table batting (player_id STRING, year INT, runs INT);
SQL
TEZ
Execute
Explain
Upload
Save as...
New Worksheet

## Activity 12 – Execute the query

Execute the query and look at the batting table. You should see data that looks like this.




==========END OF THE LAB==========