Assignment 2-1: Run a Hive Job

New Attempt

Due   Apr 25 by 11:59pm   **Points**   10   **Submitting**   a text entry box or a file upload

Available   Apr 16 at 12am - May 9 at 11:59pm 24 days

Description:

Complete the following tutorial: **https://www.cloudera.com/tutorials/how-to-process-data-with-apache-hive.html**   (edit: the tutorials are now walled off - when I was prompted for registration information for tutorial access I put in dummy values to access the tutorial)

NOTE: the version of the hortonworks vm is slightly older than the current - as such instead of "DAS UI" for "**Start DAS**   " in the tutorial, you can use the regular HIVE and HDFS ui and show in this weeks lecture to accomplish the same thing.

You can complete the tutorial using your "virtual environment"

Locally:

1) If you don't already have a virtual machine program you can use virtual box, **https://www.virtualbox.org/wiki/Downloads**
2) Download the hortonwork sandbox for virtual box, **http://hortonassets.s3.amazonaws.com/2.4/HDP_2.4_virtualbox_v3.ova**
3) Once the virtual machine is running, in a browser visit **http://IPAddress_of_VirtualMachine:8080**
4) Login with "maria_dev" as both the username and password
5) Follow the tutorial starting from "Step 1: Download The Data"

Virtual Environment:

1) Login to the virtual environment using your assigned virtual environment ID, this can be found in "class resources"
2) Once in the virtual environment, open your student VM, login as administrator, credentials in "class resources"
3) Once in the student virtual machine, in a browser visit **http://ip_of_your_hortonworksvm:8080**
4) Login with you assigned virtual environment ID, e.g: "student99", as both the username and password
5) Follow the tutorial starting from **"Step 1: Download The Data"**

-- Please Use Custom Names when following tutorial! --

In the Virtual Environment you are using a shared environment with other students, to be sure you are not overwriting another persons data be sure to use unique names and variables.

1) Please create your own personal folder when you upload data
[you can skip this specific step if your using your dedicated/personal hortonworks VM]

e,g: "/user/maria_dev" create a new folder matching your assigned virtual environment ID, e.g:/user/student99

2) Please append you virtual environment ID, (e.g: _student99), to the word "drivers" or "timesheet" in any code/sql the tutorial asks you to execute.

e.g:

```
create table temp_drivers (col_value STRING);
```

Please rename such as:

```
create table temp_drivers_student99 (col_value STRING);
```

What to turn in:

Screenshots of you completing the following steps in the tutorial:

3.3 Create Query to Populate Hive Table temp_drivers or temp_timesheet with drivers.csv or timesheet.csv (respectively) data

3.5 Create Query to Extract Data from temp_drivers or temp_timesheet and store it drivers or timesheet (respectively)

Any reasonable text format is acceptable. (e.g: doc, docx, rtf, txt, etc).

To submit your assignment:

- Use the +Submit Assignment link located in the top right corner.
- Type your submission or find and upload your saved file.
- Click the Submit Assignment button to turn in your assignment.