

# Lab Setup Instructions

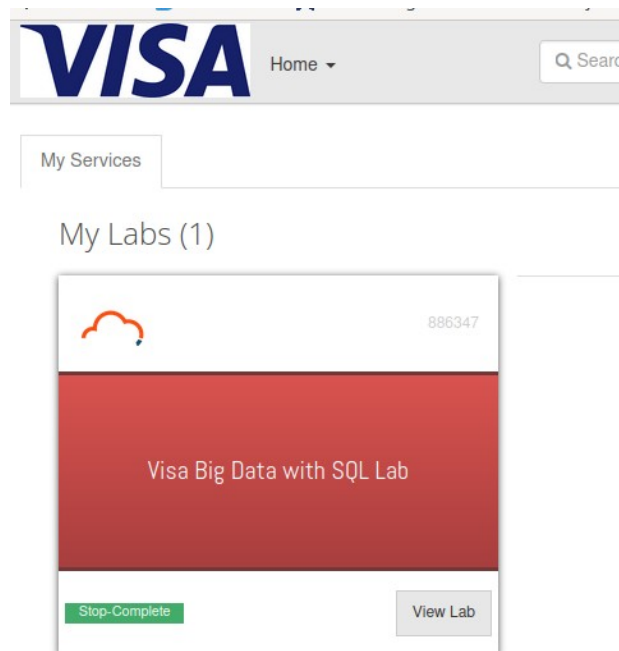
## The Lab Environment

The environment you will be using for the labs is a Horton Data Platform (HDP) virtual machine running inside Nuvelab virtual machine which you will access over a WEB URL. In the lab materials, the following terms are used consistently for clarity.

1. **Host Machine.** This is the machine you are accessing the online Nuvelab environment from, usually your normal work computer or laptop.
2. **Nuvelab VM.** This is the Nuvelab virtual machine you initially log into at the web url.
3. **HPD VM.** This is the Virtual Box VM that contains the HDP

## Starting the Lab Environment

1. Using the credentials provided, log into the Nuvelab site. You should see your Nuvelab VM in an off state.



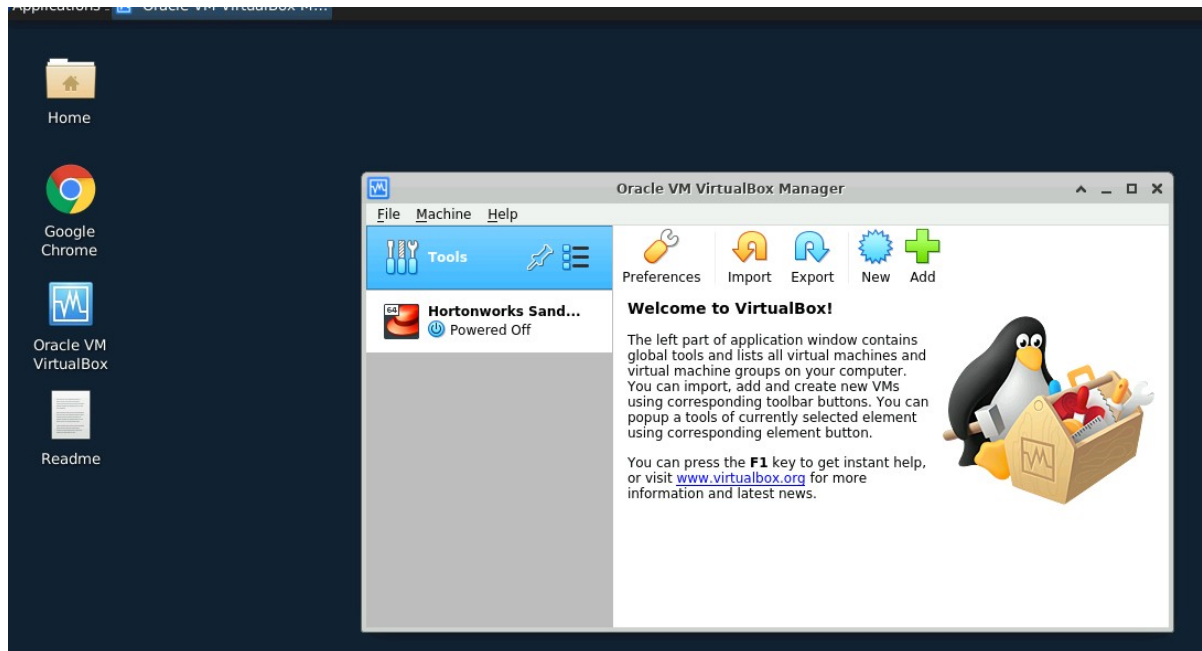
- Click on the Lab Machine to get to the start button

The screenshot shows the VISA Lab Control Panel interface. At the top, there is a header with the VISA logo, a 'Home' dropdown, a search bar, and a user profile for 'Rod Davison'. Below the header, the 'Lab Control Panel' title is followed by a cloud icon and the text 'Visa Big Data with SQL Lab'. A prominent green 'Start' button is visible. Underneath, the 'Latest Status' section shows a green checkmark and the text 'Stop - Complete' with a refresh icon. A 'Duration' box displays 'Total duration allotted' as 264 Hours(Active) and 'Consumed' as 2 hr(s) 49 mins, accompanied by a progress bar.

- Press the start button. It may take a while to boot up but eventually you will see a link to the URL you will use to access the Nuvelab VM.

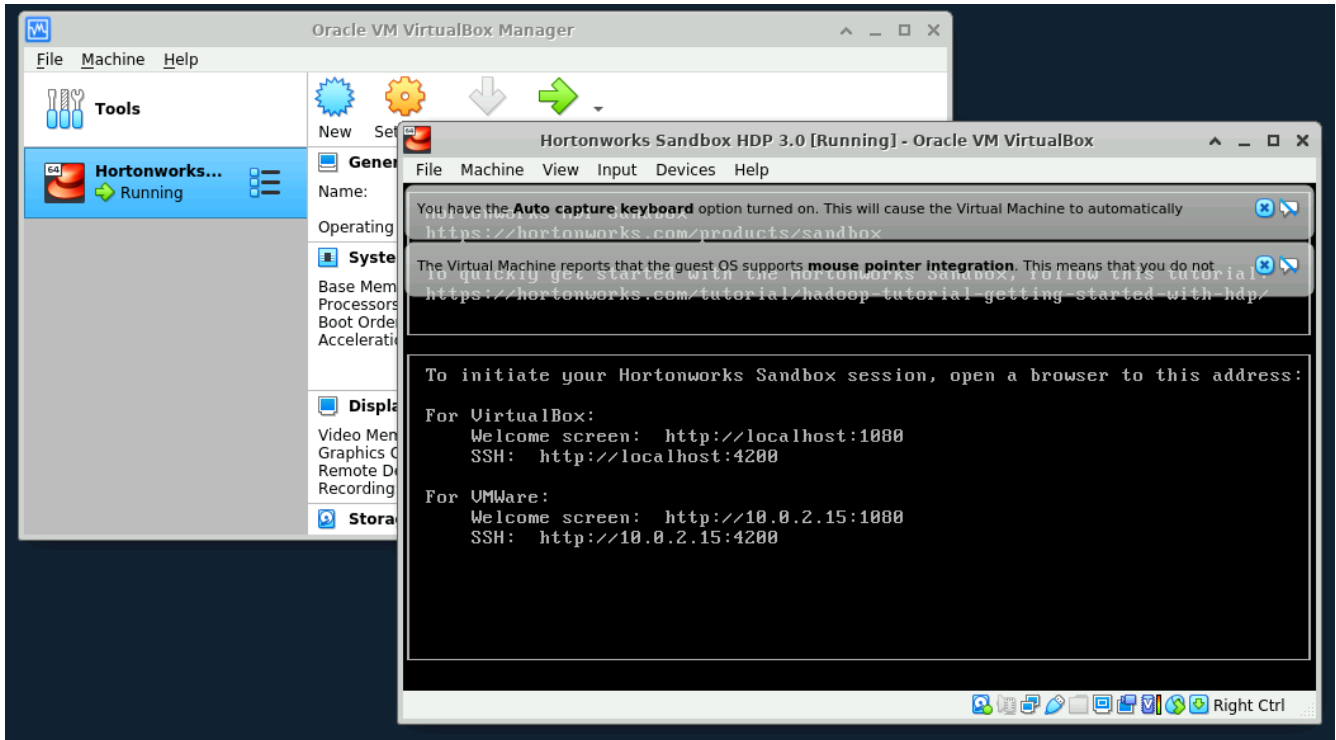
This screenshot shows the VISA Lab Control Panel after the lab has started. The 'Start' button has changed to a red 'Stop' button. The 'Latest Status' section now shows a green checkmark and the text 'Start - Complete' with a refresh icon. On the right side, there is a 'Browser Access URL' field with a copy icon. Below this, a 'More Details' section is expanded, showing links for 'Policies' and 'Resources'. At the bottom left, the 'Access Details' section shows 'Remote access to VM (DNS)' with the URL 'mujvo61345dns.eastus.cloudapp.azure.com'.

4. Going to the URL will bring up the Nuvelab VM desk top. *If you get a dialog box asking for credentials at this point, you are already logged in so just click on cancel. Or you can use the same password you used to login into the VM.*
5. Double click on the “Oracle VM Virtual Box” icon to start Virtual Box. You should see the start screen below.

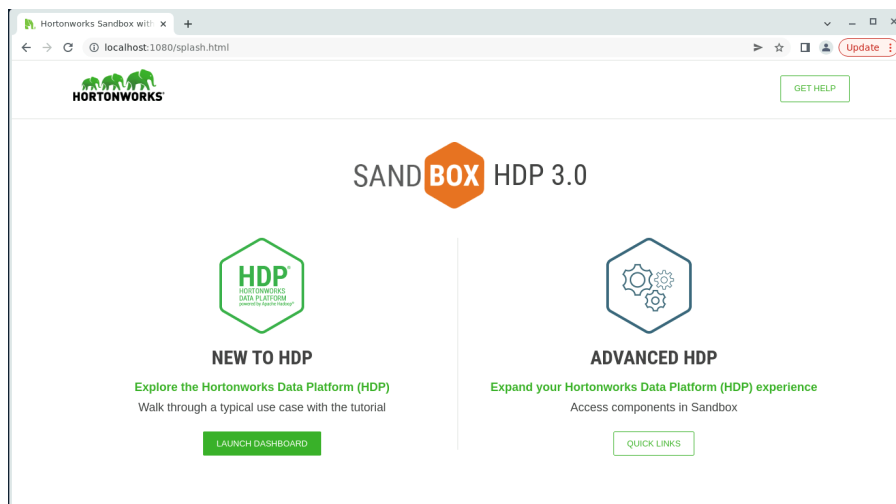


6. Select the Nortonworks Sandbox entry and a “Start” green arrow should appear which will start the HPD VM.

- The HPD VM may take a while to start up, but eventually you will see a display that looks like this.



- You are going to access the HPD VM via the browser *on the Nuvelab VM Desktop*. Open the Chrome browser and go to the “localhost:1080.” This will bring up the HDP splash screen. *Do not use a browser on your host machine desktop.*

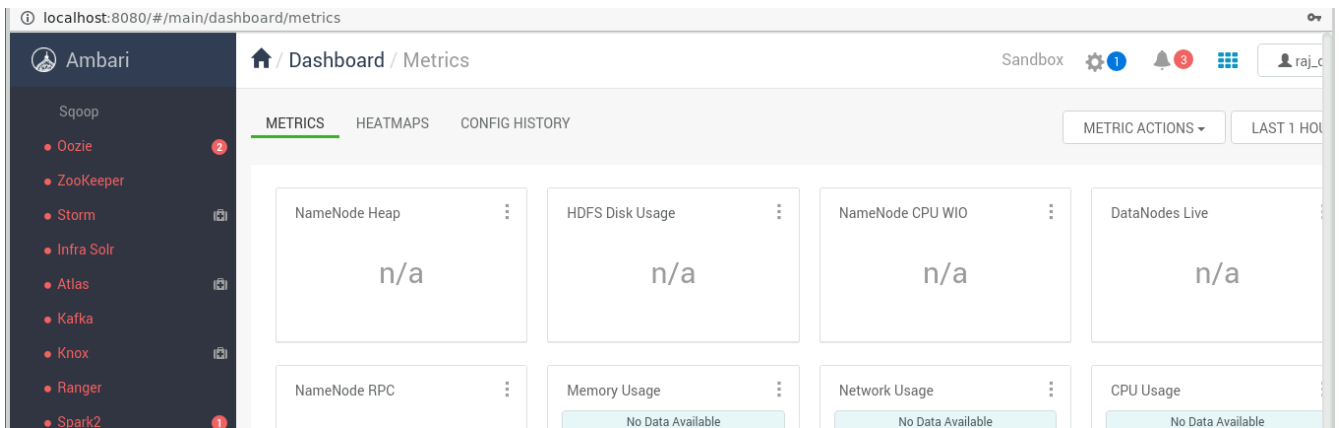


## Lab Setup Instructions

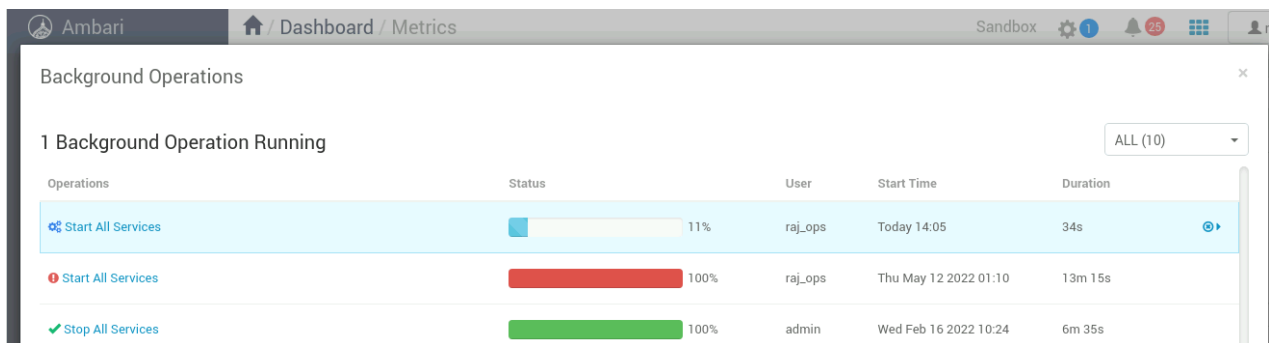
- Click the option of the left. This will bring up the Ambari management console. When prompted for a login use the credentials

User name: **raj\_ops**      password: **raj\_ops**

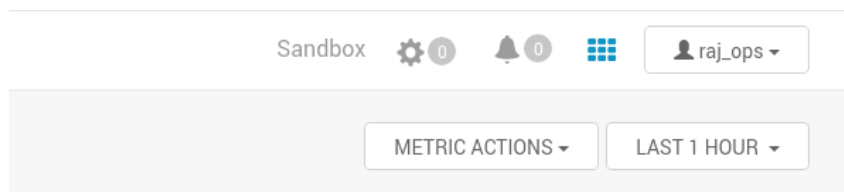
- You will see is the the display of the various services in the HPD starting up. It may take some time for them all to start.



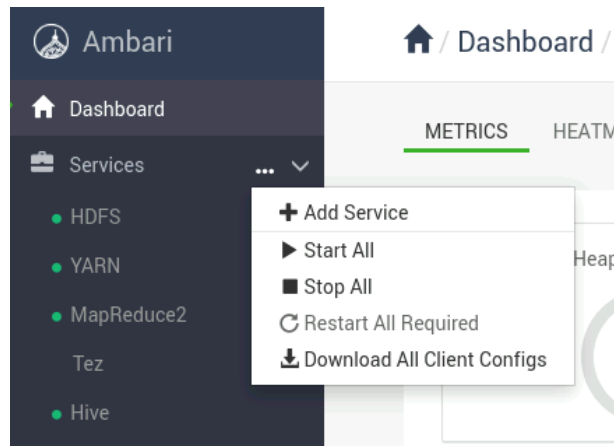
- What you are waiting for is the job “start all services” to finish



- Once finished, there should be no red “alarm” icon.



13. If there are any red alerts, you may need to use the drop down on the left hand side to restart all the services. If this doesn't work, then restart the HPD VM.



14. At this point, your HPD VM should be ready to use.

## Hive Users and Labs

The HDP VM has two predefined users: maria\_dev and raj\_ops. The password for each user is the same as their user ID. If you have to do any configuration of the HPD environment, like setting the security, then you should log in as raj\_ops since that ID has the access privileges needed. However, for the actual lab work, you should be logged in as maria\_dev.

You can create your own user account if you want, but it is not necessary for the labs.

The lab materials are located in a git hub repository here:

<https://github.com/ExgnosisClasses/VisaSQL101-June5>

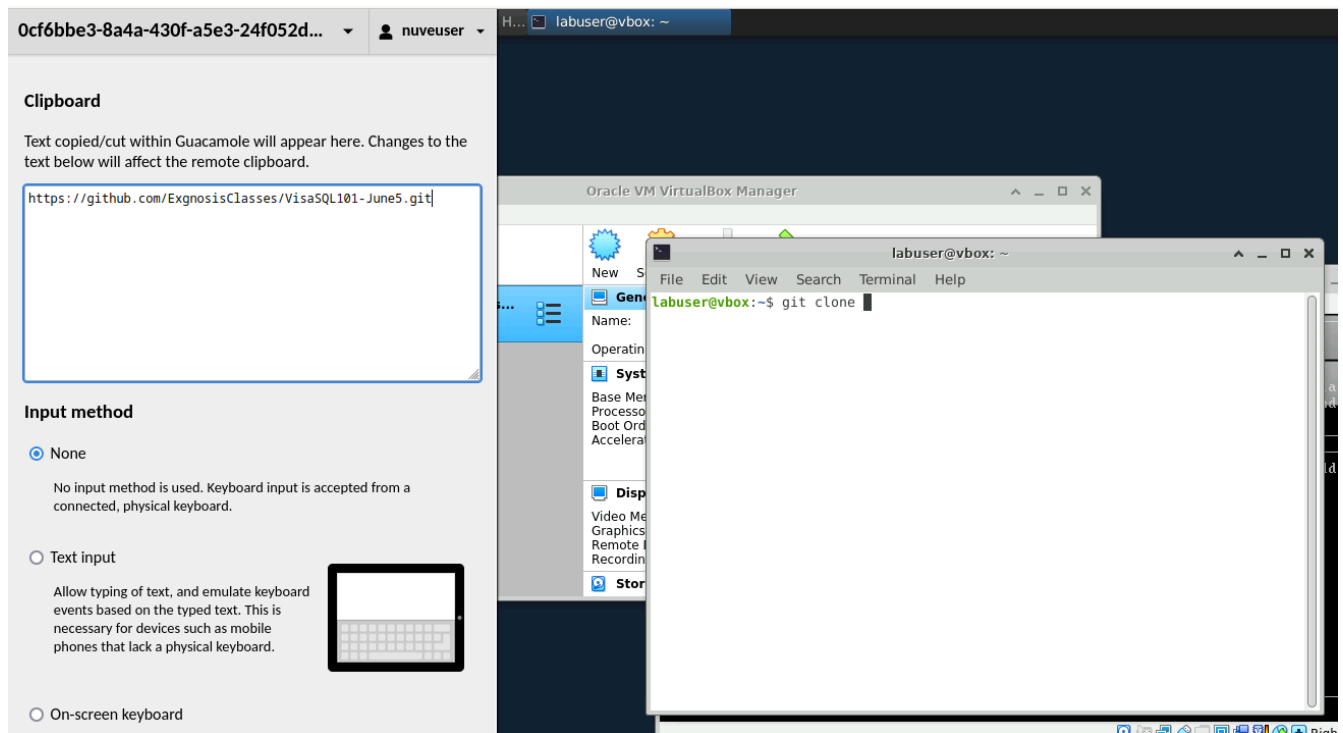
You can clone this repo to three different places.

1. *Your Host Machine.* This will enable you to keep a permanent copy of the repo but this copy will not be accessible from the Nuvelab VM and the HDP VM.
2. *The Nuvelab VM.* This the recommended place to clone the lab repository since the data in the repo can be easily uploaded into the HPD Hadoop file system through the graphical interlace. This is the option that will be used in all the lab instructions that follow.

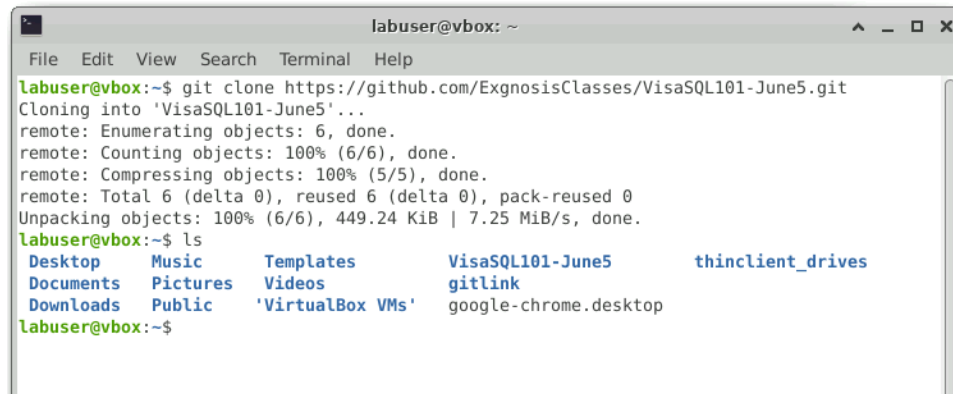
3. *The HDP VM.* If you want to use the command line HDFS commands to load and work with the data, you can also clone the repository directly into your HPD VM.

### ***Installing the repo on the Nuvelab VM***

1. For convenience, you can cut and paste between your Host Machine and the Nuvelab VM clipboards.
2. To copy to the Nuvelab VM, press *ctl-alt-shift* on *only* the left side of the keyboard. This will open up a window on the left side of the screen. Anything you put into that window will be copied to the Nuvelab VM clipboard. Once entered, press *ctl-alt-shift* on the left side of the keyboard to remove the window. Similarly, anything currently on the VM clipboard will appear on the same screen.



3. To add the repo to the Nuvelab VM, open up a terminal window from the menu at the bottom of the screen and execute the git clone command.



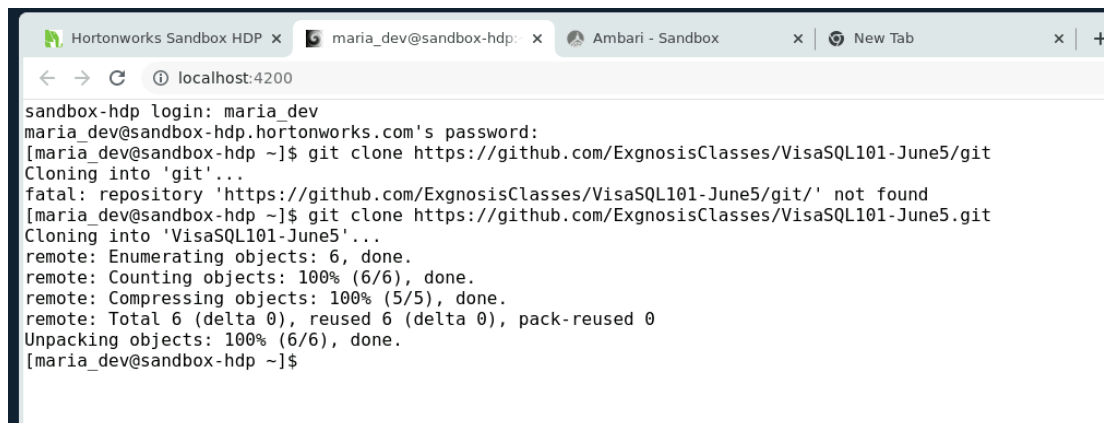
```

labuser@vbox: ~
File Edit View Search Terminal Help
labuser@vbox:~$ git clone https://github.com/ExgnosisClasses/VisaSQL101-June5.git
Cloning into 'VisaSQL101-June5'...
remote: Enumerating objects: 6, done.
remote: Counting objects: 100% (6/6), done.
remote: Compressing objects: 100% (5/5), done.
remote: Total 6 (delta 0), reused 6 (delta 0), pack-reused 0
Unpacking objects: 100% (6/6), 449.24 KiB | 7.25 MiB/s, done.
labuser@vbox:~$ ls
Desktop      Music        Templates    VisaSQL101-June5  thinclient_drives
Documents    Pictures     Videos      gitlink
Downloads    Public      'VirtualBox VMs' google-chrome.desktop
labuser@vbox:~$

```

### Installing the Lab Repo in the HDP VM

1. First, open a web shell for the maria\_dev user since that is the id that will be used in the labs in the HDP VM.
2. In a new browser window, go to localhost:4200. When prompted, enter maria\_dev for both the user id and password.
3. Then you can execute the git clone command to install the repo in the HDP VM. *However, it is not necessary to do this to complete the labs.*



```

Hortonworks Sandbox HDP x maria_dev@sandbox-hdp: x Ambari - Sandbox x New Tab x +
localhost:4200
sandbox-hdp login: maria_dev
maria_dev@sandbox-hdp.hortonworks.com's password:
[maria_dev@sandbox-hdp ~]$ git clone https://github.com/ExgnosisClasses/VisaSQL101-June5/git
Cloning into 'git'...
fatal: repository 'https://github.com/ExgnosisClasses/VisaSQL101-June5/git/' not found
[maria_dev@sandbox-hdp ~]$ git clone https://github.com/ExgnosisClasses/VisaSQL101-June5.git
Cloning into 'VisaSQL101-June5'...
remote: Enumerating objects: 6, done.
remote: Counting objects: 100% (6/6), done.
remote: Compressing objects: 100% (5/5), done.
remote: Total 6 (delta 0), reused 6 (delta 0), pack-reused 0
Unpacking objects: 100% (6/6), done.
[maria_dev@sandbox-hdp ~]$

```



## Starting Hive

1. For these labs, it is assumed that you will be running Hive as the user `maria_dev`.
2. In a new browser window in the Nuvelab VM, go to `localhost:4200`. When prompted, enter `maria_dev` for both the user id and password.
3. At the command prompt, enter `hive` to start the hive shell.

```

maria_dev@sandbox-hdp: x +
localhost:4200
sandbox-hdp login: maria_dev
maria_dev@sandbox-hdp.hortonworks.com's password:
Last login: Mon May 30 14:50:34 2022 from 172.18.0.2
[maria_dev@sandbox-hdp ~]$ hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hdp/3.0.1.0-187/hive/lib/log4j-:
SLF4J: Found binding in [jar:file:/usr/hdp/3.0.1.0-187/hadoop/lib/slf4:
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an ex
SLF4J: Actual binding is of type [org.apache.logging.slf4j.Log4jLogger
Connecting to jdbc:hive2://sandbox-hdp.hortonworks.com:2181/default;pa
espace=hiveserver2
22/05/30 15:54:48 [main]: INFO jdbc.HiveConnection: Connected to sandb
Connected to: Apache Hive (version 3.1.0.3.0.1.0-187)
Driver: Hive JDBC (version 3.1.0.3.0.1.0-187)
Transaction isolation: TRANSACTION_REPEATABLE_READ
Beeline version 3.1.0.3.0.1.0-187 by Apache Hive
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>

```

4. To ensure hive is running, execute the command “show databases;” which should re- turn the output below.

```

0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> show databases;
INFO : Compiling command(queryId=hive_20220530155756_f62d66ad-5895-4756
INFO : Semantic Analysis Completed (retrial = false)
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:dat
s:null)
INFO : Completed compiling command(queryId=hive_20220530155756_f62d66ad
INFO : Executing command(queryId=hive_20220530155756_f62d66ad-5895-4756
INFO : Starting task [Stage-0:DDL] in serial mode
INFO : Completed executing command(queryId=hive_20220530155756_f62d66ad
INFO : OK
+-----+
| database_name |
+-----+
| default      |
| foodmart     |
| information_schema |
| sys          |
+-----+
4 rows selected (0.999 seconds)

```

## Final Details

1. The README document on the Nuvelab VM desktop is incorrect. It gives the root password for the HDP VM as “labuser123\$”. This is wrong. The correct password is the same as for the Nuvelab VM “Nuvelabs123\$”.
2. At the hive shell, run the following command.

```
set hive.cli.print.header=true;
```

3. If you get an error like below, then you need to make the modifications described in the other pdf document “Security fix for labs.pdf”

```
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> set hive.cli.print.header=true;
Error: Error while processing statement: Cannot modify hive.cli.print.header at runtime. It is not in list of params that are allowed to be modified at runtime (state=42000,code=1)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> set hive.cli.print.header;
+-----+
|          set          |
+-----+
| hive.cli.print.header=false |
+-----+
1 row selected (0.159 seconds)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

When you are done, you should close the HDP VM window and select the option to power off the VM as shown on the next page. This will ensure the next time you use the machine, all of the services will be initialized correctly. If you just suspend the machine, not all of the services may resume correctly leaving your VM in an unstable state.

