Lab Setup Instructions

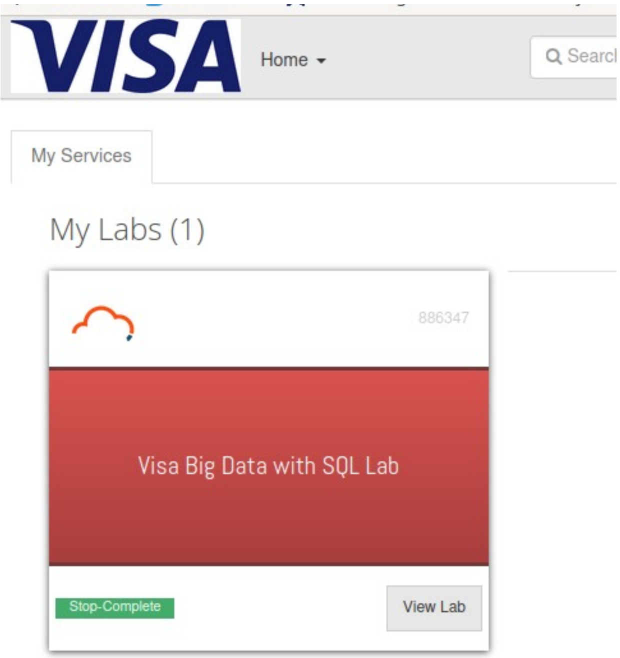
The Lab Environment

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

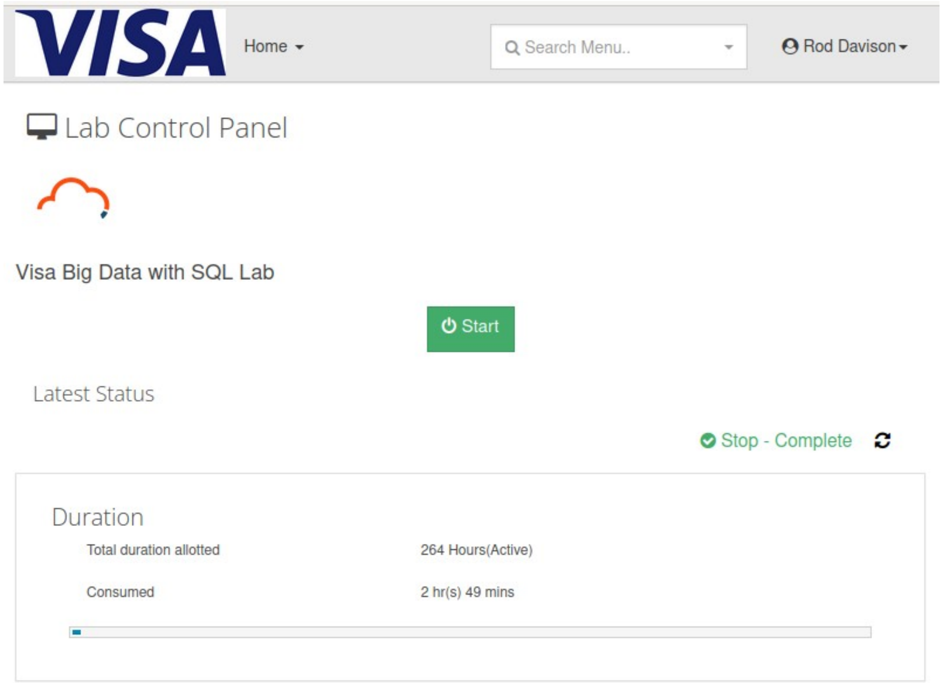
1. Host Machine. This is the machine you are accessing the online Nuvelab environment from, usually your normal work computer or laptop.
2. Nuvelab VM. This is the Nuvelab virtual machine you initially log into.
3. HPD VM. This is the Virtual Box VM that contains the HDP

Starting the Lab Environment

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



2. Click on the Lab Machine to get to the start button



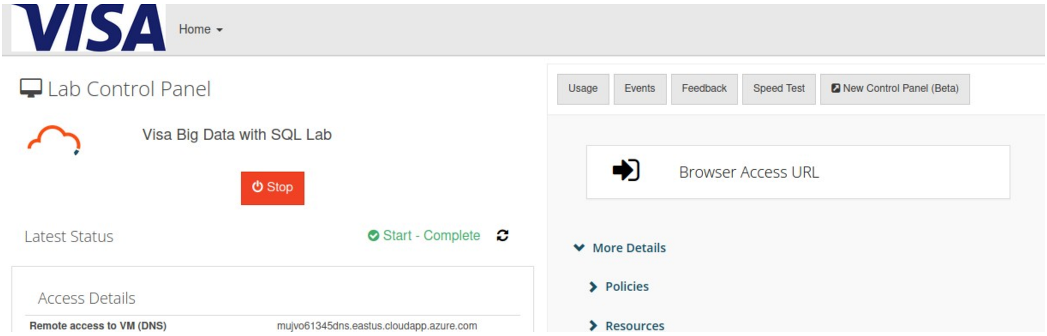
Latest Status

Stop - Complete

Duration

Total duration allotted	264 Hours(Active)
Consumed	2 hr(s) 49 mins

3. 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.



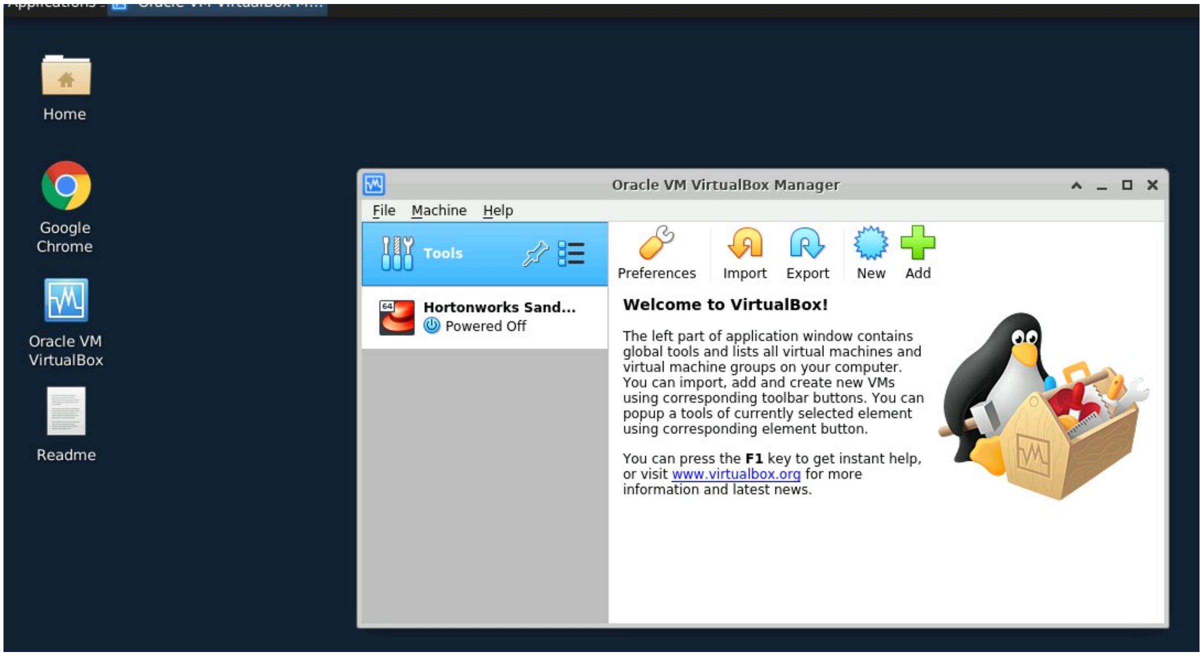
Latest Status

Start - Complete

Access Details

Remote access to VM (DNS)	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.*
5. Double click on the “Oracle VM Virtual Box” icon to start Virtual Box. You should see the start screen below.



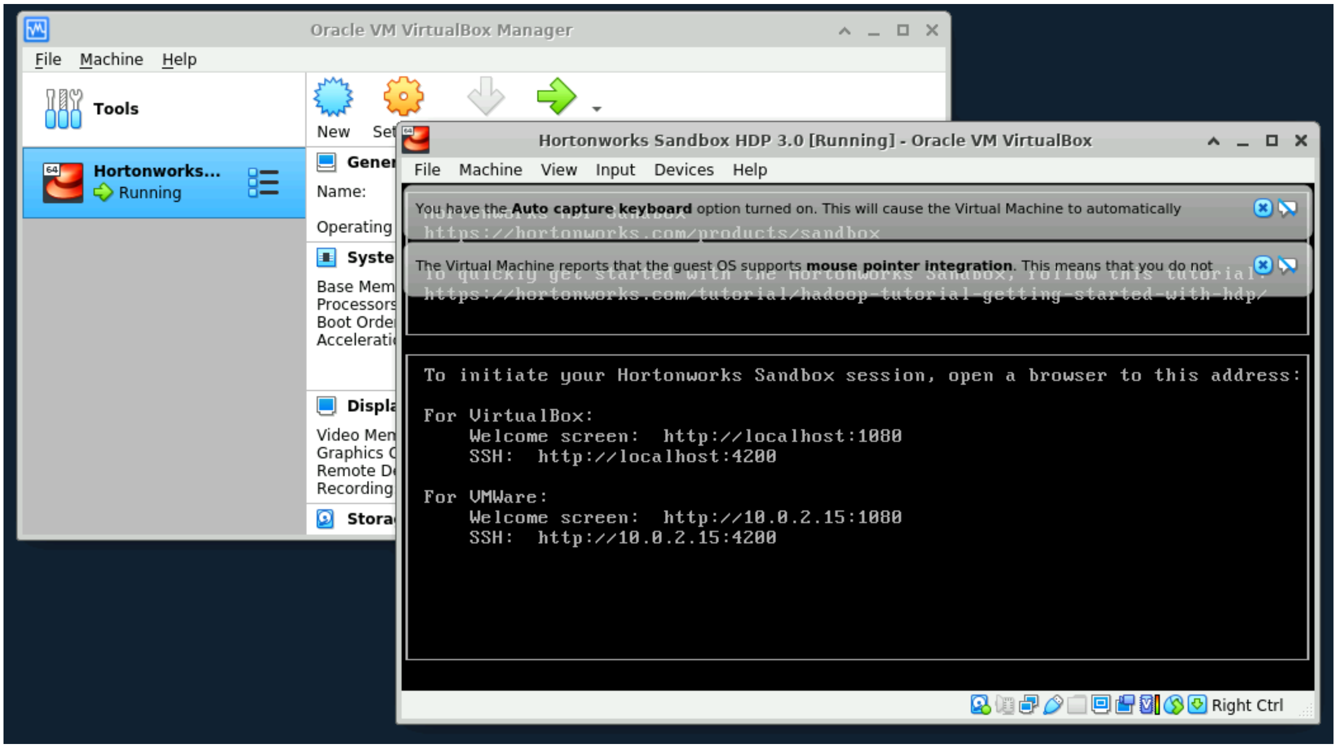
Welcome to VirtualBox!

The left part of application window contains global tools and lists all virtual machines and virtual machine groups on your computer. You can import, add and create new VMs using corresponding toolbar buttons. You can popup a tools of currently selected element using corresponding element button.

You can press the **F1** key to get instant help, or visit www.virtualbox.org for more information and latest news.

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

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



You have the **Auto capture keyboard** option turned on. This will cause the Virtual Machine to automatically

The Virtual Machine reports that the guest OS supports **mouse pointer integration**. This means that you do not

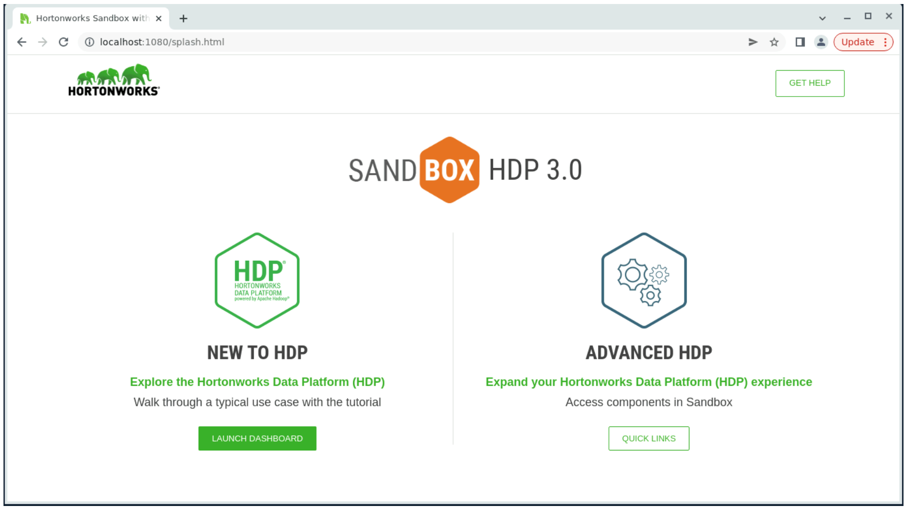
```
https://hortonworks.com/products/sandbox
To quickly get started with the Hortonworks Sandbox, follow this tutorial:
https://hortonworks.com/tutorial/hadoop-tutorial-getting-started-with-hdp/
```

```
To initiate your Hortonworks Sandbox session, open a browser to this address:

For VirtualBox:
    Welcome screen:  http://localhost:1080
    SSH:  http://localhost:4200

For VMWare:
    Welcome screen:  http://10.0.2.15:1080
    SSH:  http://10.0.2.15:4200
```

8. 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.



NEW TO HDP

Explore the Hortonworks Data Platform (HDP)

Walk through a typical use case with the tutorial

ADVANCED HDP

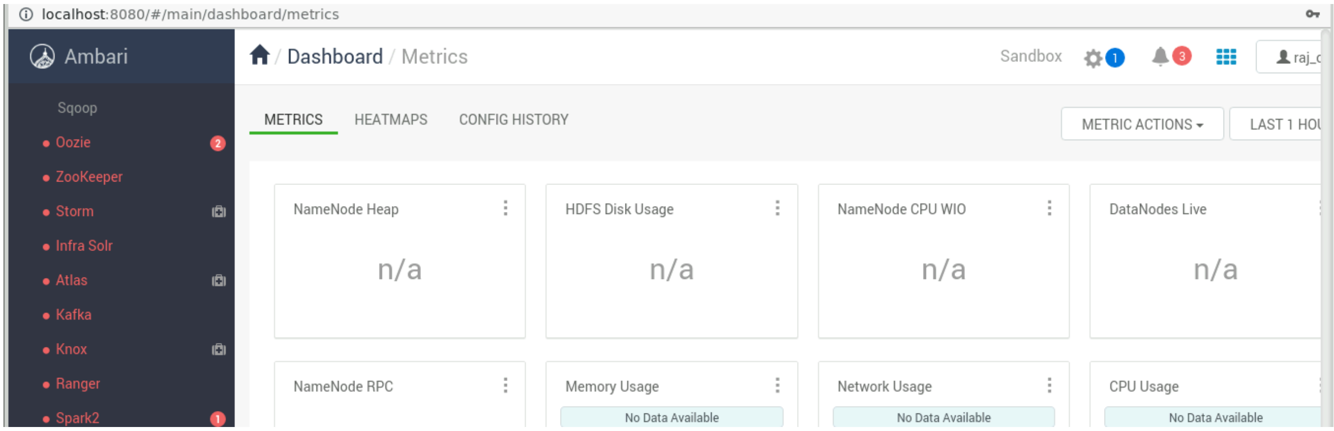
Expand your Hortonworks Data Platform (HDP) experience

Access components in Sandbox

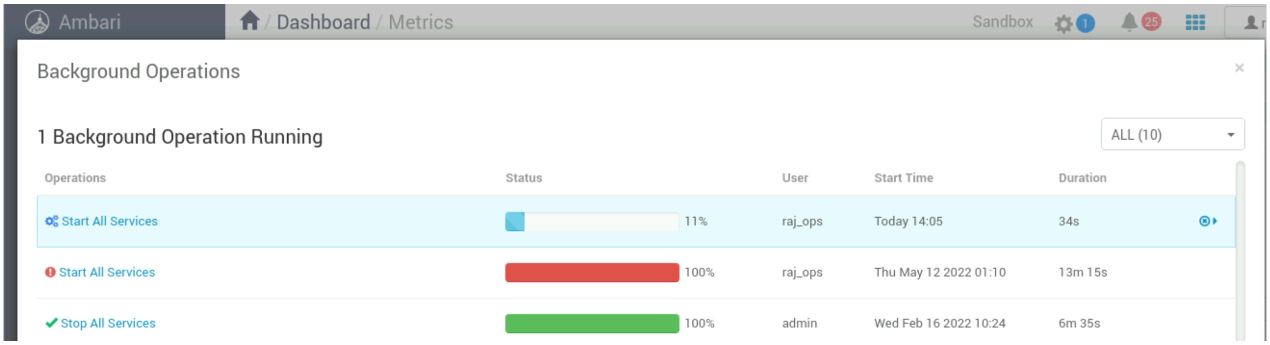
9. Click the option of the left. This will bring you to the Ambari management console. When prompted for a login use the credentials

   User name: **raj_ops** password: **raj_ops**

10. What 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.



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

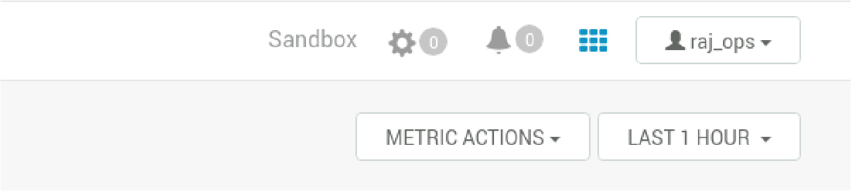


Background Operations

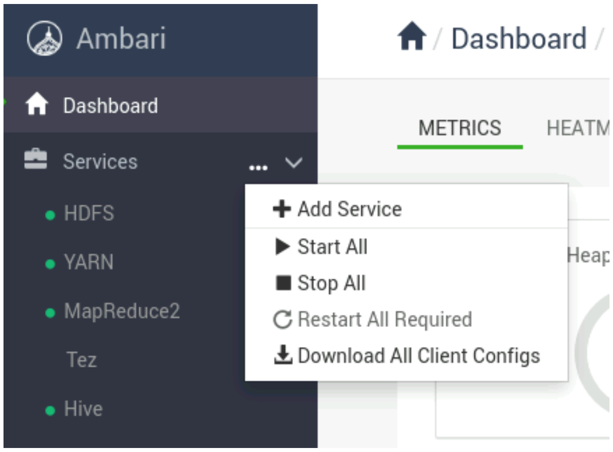
1 Background Operation Running

Operations	Status	User	Start Time	Duration
Start All Services	11%	raj_ops	Today 14:05	34s
Start All Services	100%	raj_ops	Thu May 12 2022 01:10	13m 15s
✓ Stop All Services	100%	admin	Wed Feb 16 2022 10:24	6m 35s

12. 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:

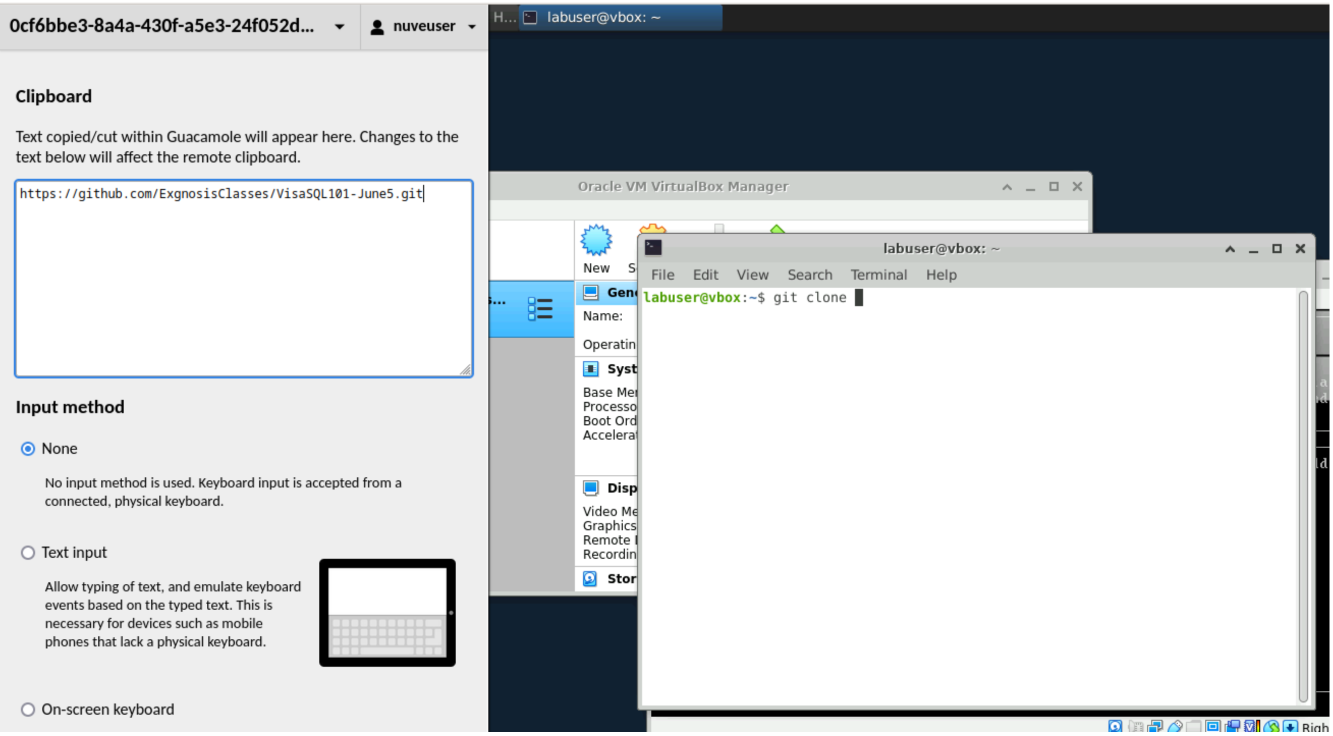
https://github.com/ExgnosisClasses/VisaSQL101-June5

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 it will be inaccessible from the Nuvelab VM and the HDP VM.
2. The Nuvelab VM. This the recommended place to clone the lab since the data in the repo can be easily uploaded into the HPD Hadoop file system through the graphical interace. This is the options 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 to your HPD VM.

Installing the repo on the Nuvelab VM

1. For convenience, you can copy data between your Host Machine and the Nuvelab VM.
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 inserted into the Nuvelab VM. Once entered, press ctl-alt-shift on the left side of the keyboard to remove the window.



Clipboard

Text copied/cut within Guacamole will appear here. Changes to the text below will affect the remote clipboard.

```
https://github.com/ExgnosisClasses/VisaSQL101-June5.git
```

Input method

● None

No input method is used. Keyboard input is accepted from a connected, physical keyboard.

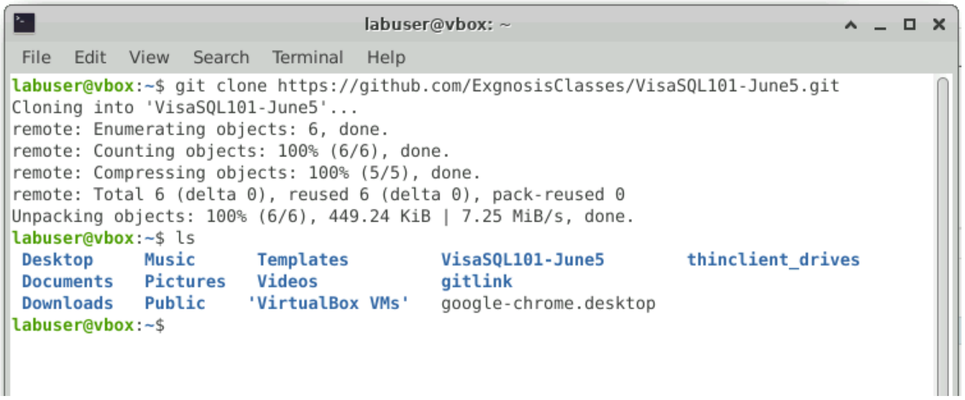
○ Text input

Allow typing of text, and emulate keyboard events based on the typed text. This is necessary for devices such as mobile phones that lack a physical keyboard.

○ On-screen keyboard

```
labuser@vbox:~$ git clone
```

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:~$ 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:~$
```

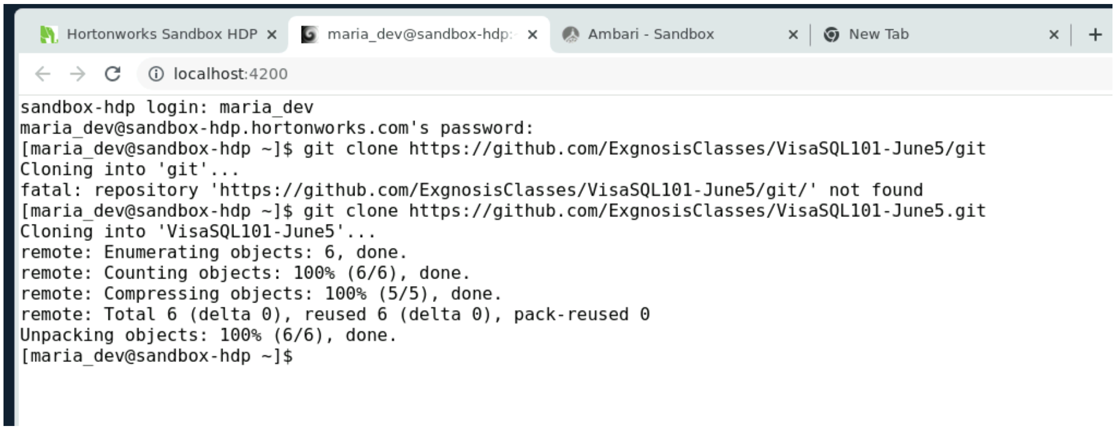
4.

Installing the 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.

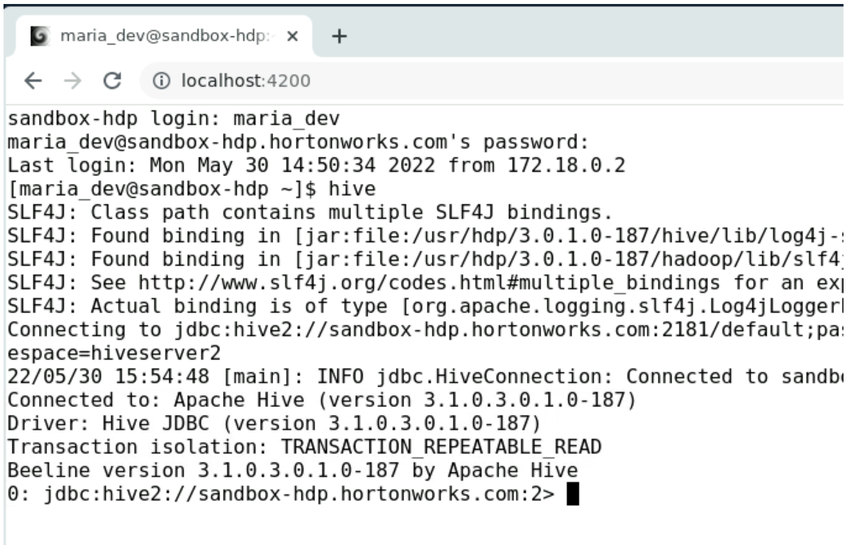
3. Then you can execute the git clone command to install the repo in the HDP VM. *However, it is not necessary to do this to complete the labs.*



```
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 ~]$
```

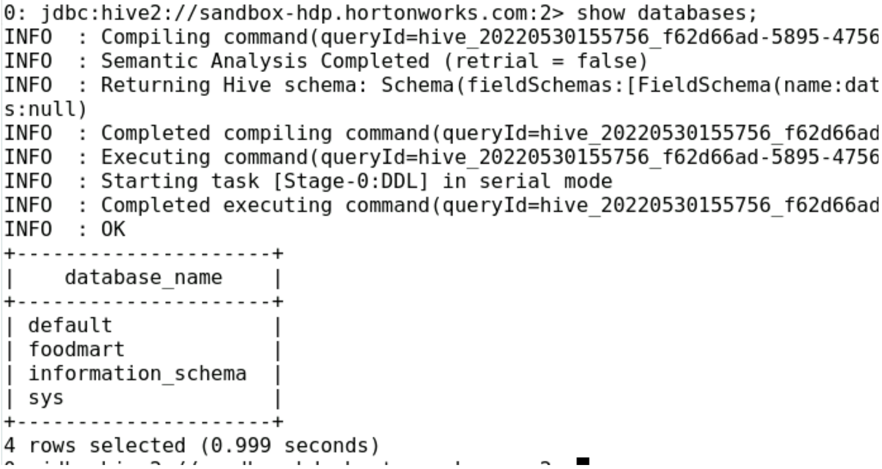
Starting Hive

1. 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. At the command prompt, enter hive to start the hive shell.



```
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 return the output below.



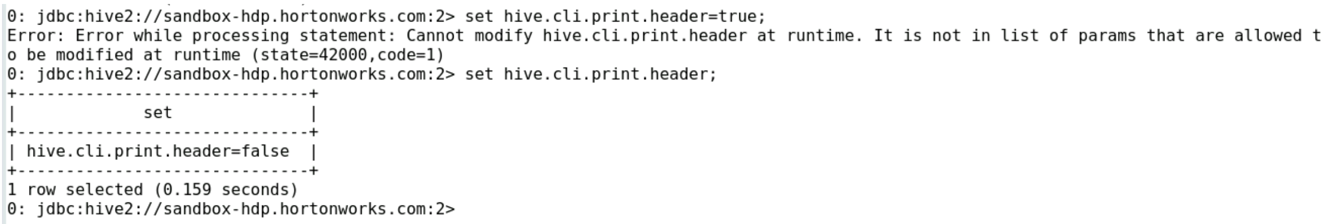
```
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2> show databases;
INFO  : Compiling command(queryId=hive_20220530155756_f62d66ad-5895-4756
INFO  : Semantic Analysis Completed (retrial = false)
INFO  : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:dat
s:null)
INFO  : Completed compiling 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)
0: jdbc:hive2://sandbox-hdp.hortonworks.com:2>
```

Final Details

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

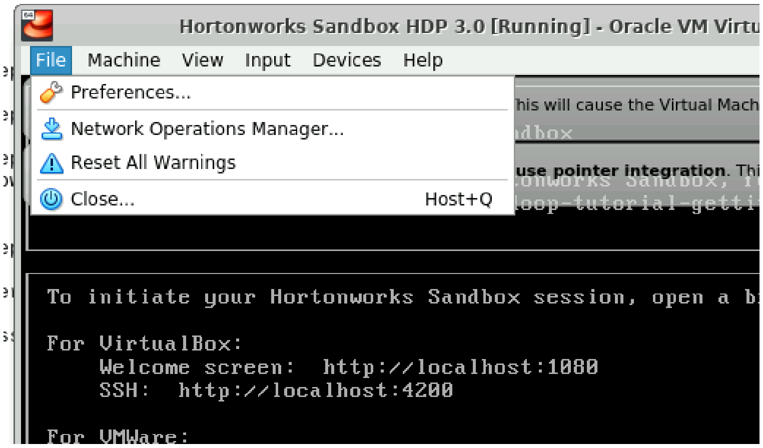
   set hive.cli.print.header=true;

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 t
o 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. 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.

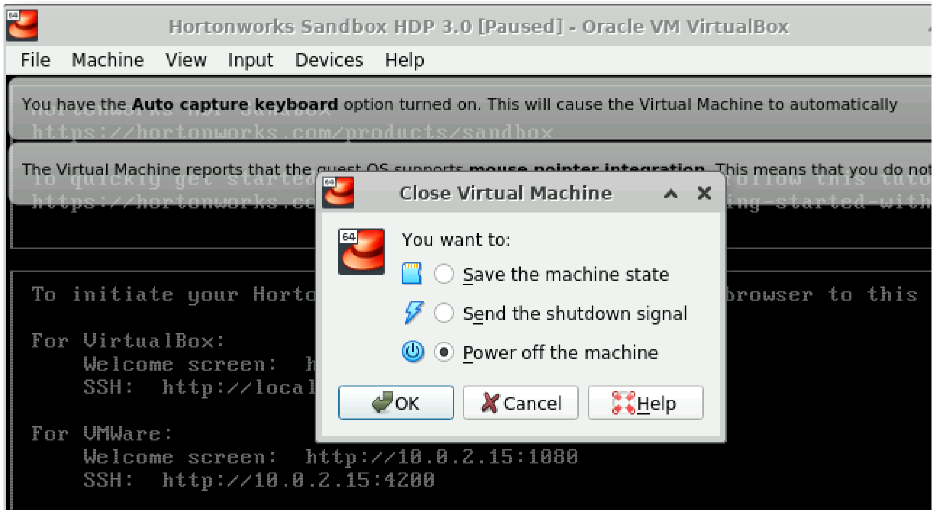


Hortonworks Sandbox HDP 3.0 [Running] - Oracle VM Virtu
File
Machine
View
Input
Devices
Help
Preferences...
Network Operations Manager...
Reset All Warnings
Close... Host+Q
his will cause the Virtual Mach
use pointer integration. Thi

```
To initiate your Hortonworks Sandbox session, open a b

For VirtualBox:
    Welcome screen:  http://localhost:1080
    SSH:  http://localhost:4200

For VMWare:
```



Hortonworks Sandbox HDP 3.0 [Paused] - Oracle VM VirtualBox

You have the **Auto capture keyboard** option turned on. This will cause the Virtual Machine to automatically

The Virtual Machine reports that the guest OS supports **mouse pointer integration**. This means that you do not

```
Hortonworks HDP Sandbox
https://hortonworks.com/products/sandbox

To quickly get started                                   follow this tuto
https://hortonworks.co                                  ing-started-with

To initiate your Horto                                  browser to this

For VirtualBox:
    Welcome screen:  h
    SSH:  http://local

For VMWare:
    Welcome screen:  http://10.0.2.15:1080
    SSH:  http://10.0.2.15:4200
```

Close Virtual Machine

You want to:

○ Save the machine state

○ Send the shutdown signal

● Power off the machine