1. Setting up Hortonworks Sandbox on VMware

This tutorial helps you in installing Hortonworks Sandbox on VMware Workstation. This tutorial is designed for Windows users only. The tutorial is simple and straight forward with lots of screenshots and illustrations to help you in the installation.

The VMware Workstation is a hosted hypervisor for the Windows and Linux platforms. In simple terms, it provides you a virtual environment on your computer so that you can test operating systems and software on it without affecting the rest of the computer. We will be using it to run Hortonworks Sandbox so as to give you a hands-on experience of Big Data and give you ample practice.

Hortonworks Sandbox is a stable distribution of Apache Hadoop. It has all the tools for Big Data configured and bundled into a single distribution so that you don't have to install each of them separately and configure them one by one. Sandbox is the perfect tool for beginners to just dive in and start learning without the hassle of configuration.

1.1. Pre-requisites

Your device must have

- 12Gb or above of RAM (16Gb is recommended)
- 50Gb of free hard disk space
- VMware workstation 11.0 or above installed
- .ova file for Hortonworks Sandbox 2.4 or above
- Windows 7 64-bit or higher is recommended. Older Windows can have issues with some of VMware's features.

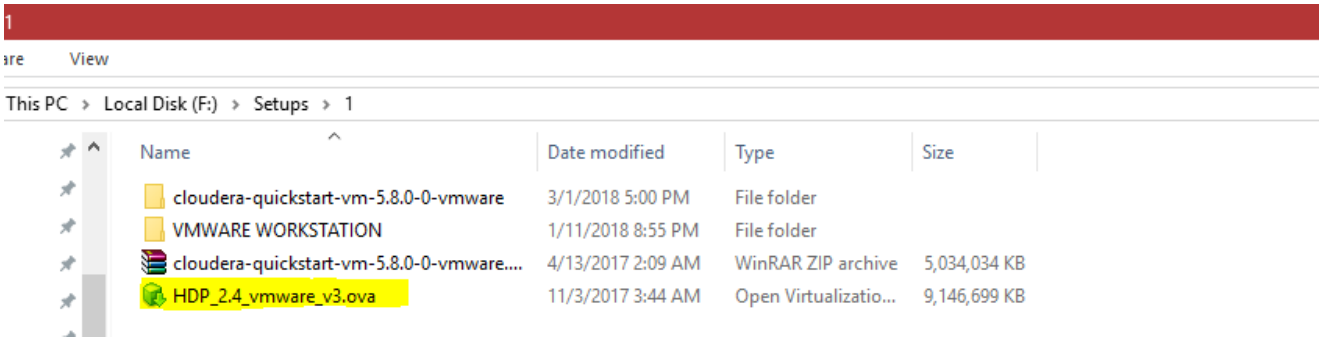
prior to the installation. It is also recommended to have an Intel 4th generation core i3 processor or higher for optimal performance. Lower specifications can work but they will be slow while running jobs.

If VMware Workstation is not installed on your device then install it using the installer provided. When prompted for the serial key input the key in the accompanying serial.txt file.

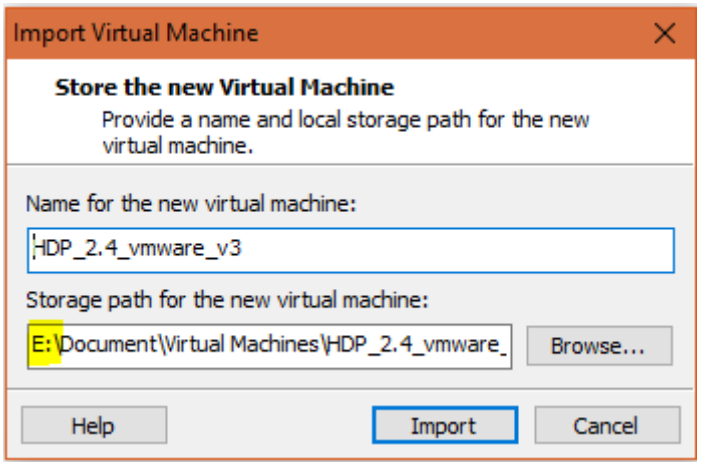
The .ova file for Hortonworks Sandbox on your device will also be provided along with the VMware installer. We will be using Hortonworks Sandbox 2.4 in the class. You can also download the latest version using the link below.
https://hortonworks.com/downloads/#sandbox

1.2. Setting up Hortonworks Sandbox

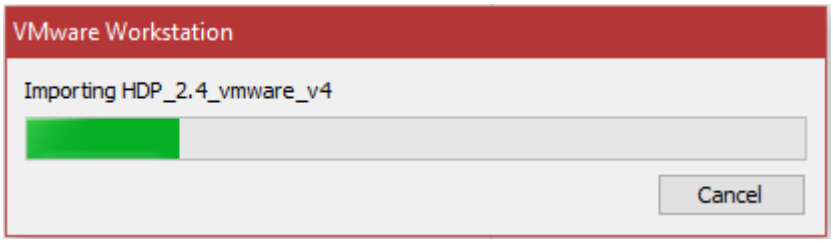
- Navigate to the Hortonworks Sandbox .ova file and open it.



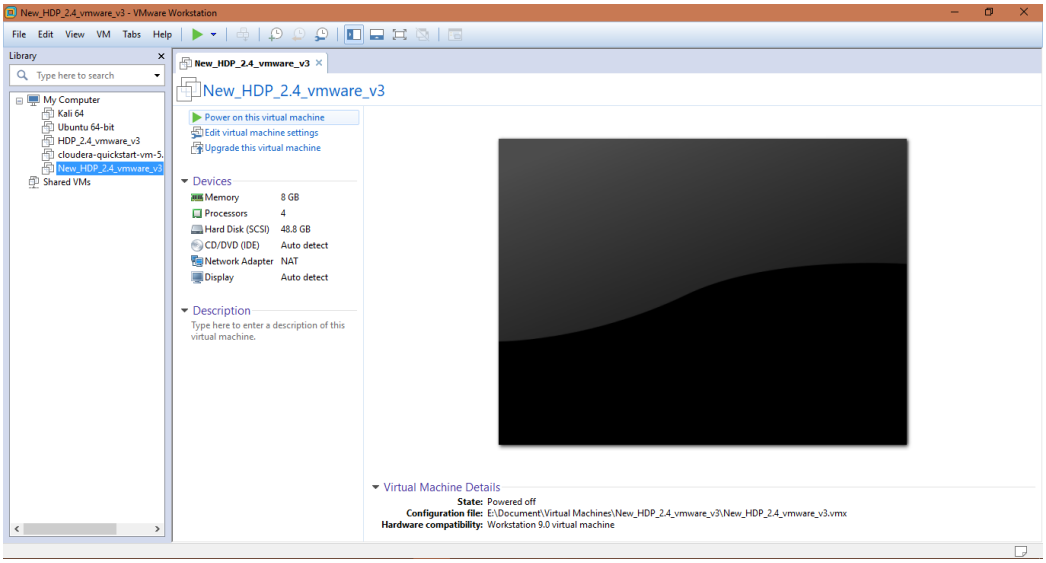
- It should open the VMware workstation and open the dialog below. Make sure the disk you are specifying in the storage path field has more than 50Gb of free space available.



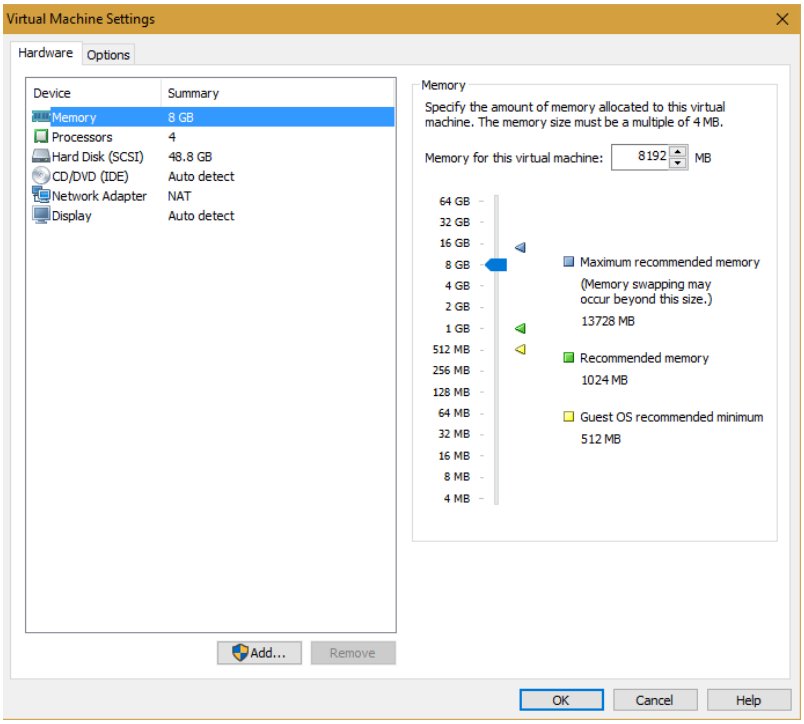
- Change the name or storage location of the virtual machine here you are creating if you want. Make sure that the path you are specifying has at least 50Gb of free space available. Click import when you are done.



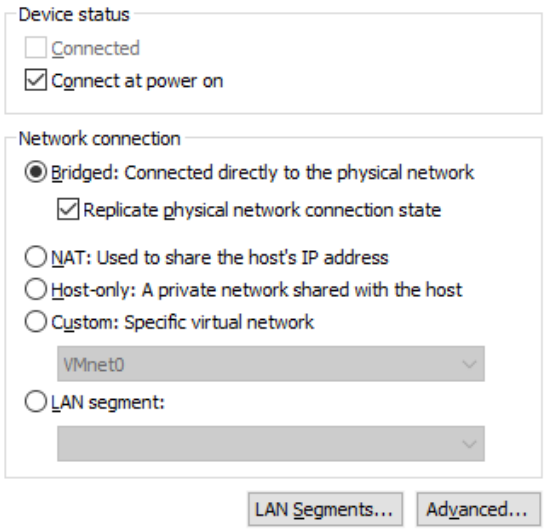
- When the Virtual machine is done importing its window will appear and it will be added on the virtual machine on the left side.



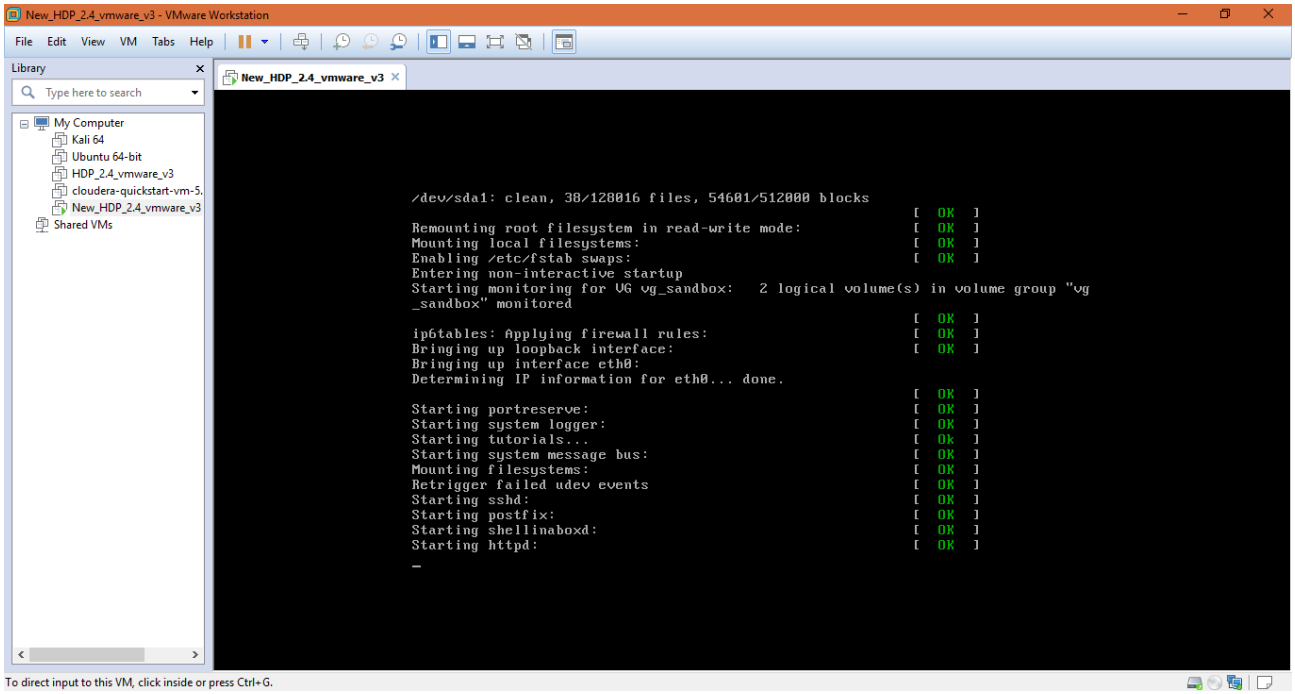
- Click on edit virtual machine settings. The settings window should appear with the default virtual machine settings:



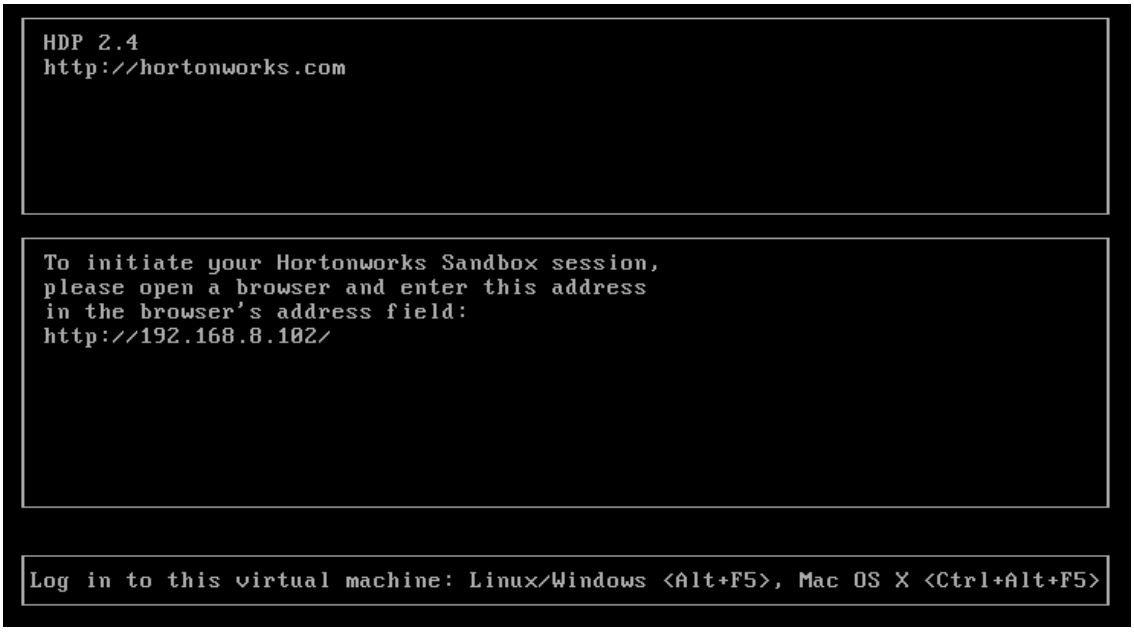
- On the memory tab set the slider to the maximum recommended memory pointer.
- On the Network Adapter tab change the settings to bridged and allow replication of physical connection state. The settings should look as follows:



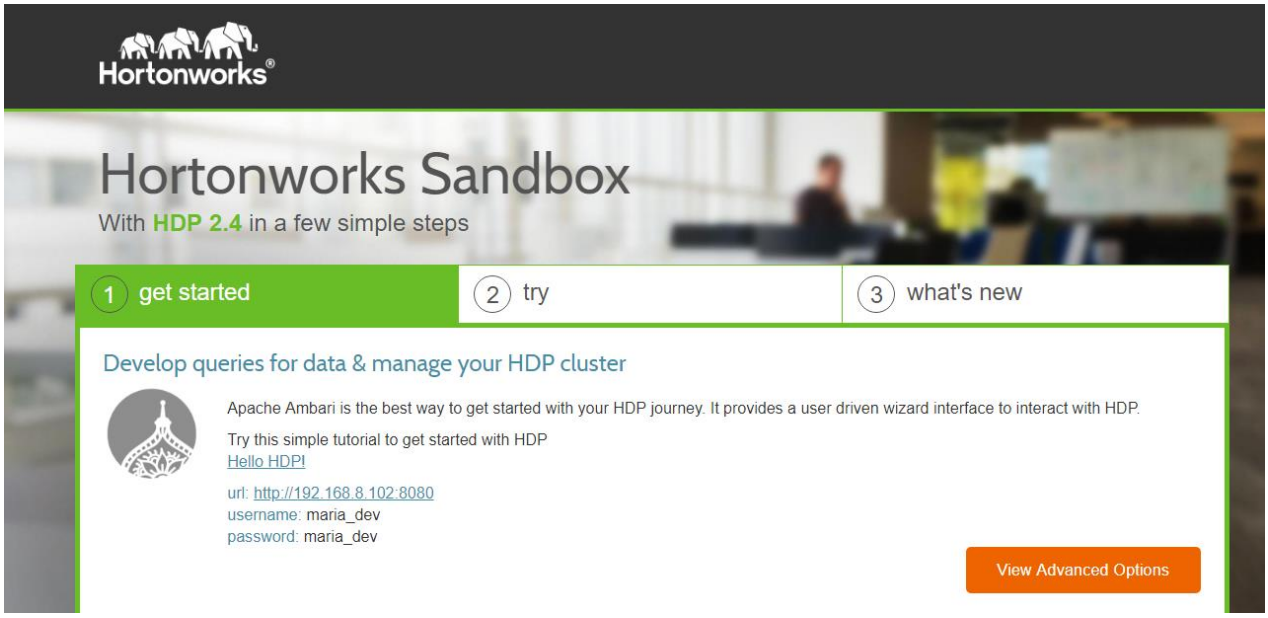
- Click OK and click Power on this virtual machine.
- The machine should start booting up. Normally this should take around 5 minutes but on older devices this step can take around 15 minutes as well.



- The Hortonworks Sandbox virtual machine has been set up. After the machine has completed boot up the following welcome screen should appear. The IP address can be different for each user.



- Go to your favorite browser (Google Chrome or Mozilla Firefox are recommended but this will work in all browsers. On the address bar type the IP address you see on your virtual machine's welcome screen and press enter. The Hortonworks Sandbox welcome page should appear on a successful installation.



If the welcome page does not appear then please refer to the section 4 of this document for some quick troubleshooting.

Now that we have set up the Sandbox we now need to set up Putty for easy access to the sandbox command line and set up the admin credentials for the Ambari service. Refer to section 2 for Putty installation and section 3 for admin password setup.

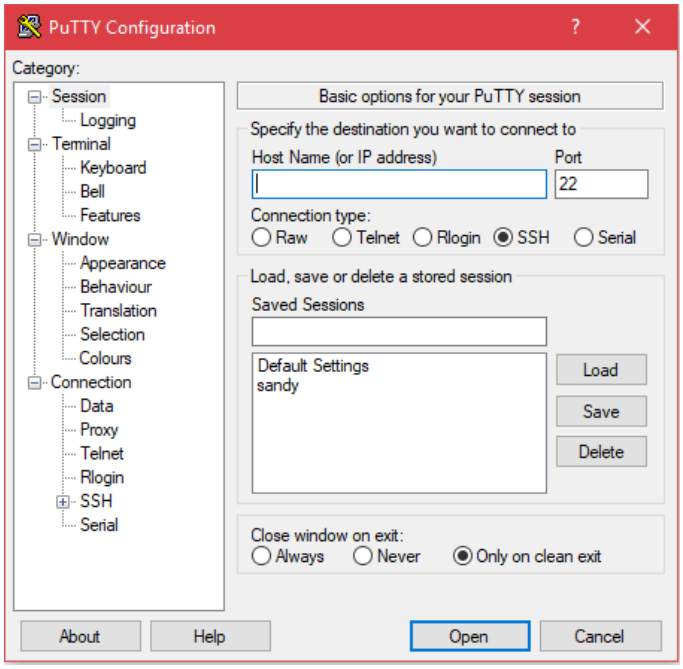
2. Setting up Putty

In this section we will install Putty and access the Sandbox command line with it. The sandbox command line accessed through Putty is also accessible through the browser. Just enter your sandbox IP address in the browser with the port 4200 as follows:
Your-ip-address:4200

For example
192.168.8.102:4200

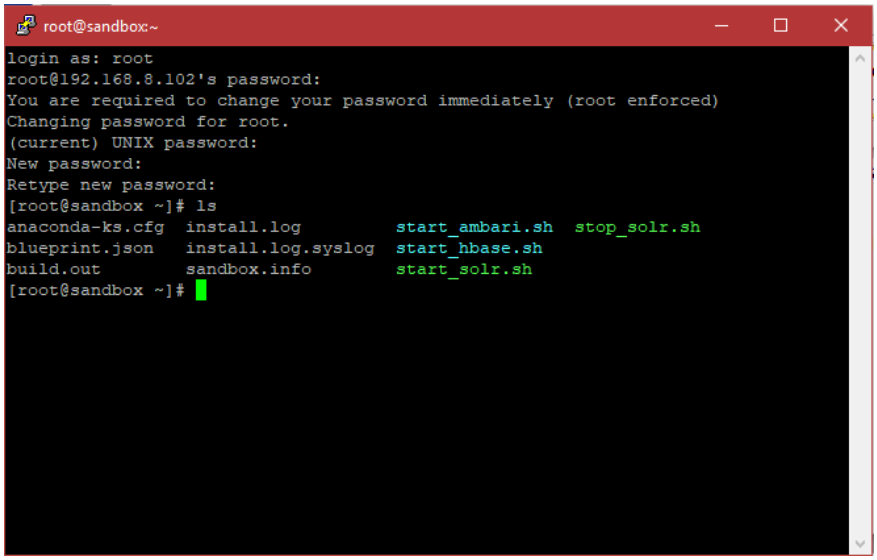
Hence, this section is skippable but it is recommended to set up Putty to give you a feel for the professional environment and due to the fact that Putty is faster at the tasks.

- Download the Putty installer from the following link:
  https://www.chiark.greenend.org.uk/~sgtatham/putty/latest.html
  After the installer has downloaded, run it and install Putty by following the wizard.
- Open Putty. The following session screen will show up.



Make sure the Hortonworks Sandbox virtual machine is running before continuing with the procedure.

- Enter the IP address of the virtual machine in the Host Name bar and click open. A command line should appear prompting you for credentials. Use the following credentials to log in:
  Username: root
  UNIX Password: hadoop
- Since this is the first time you are accessing the command line you will be prompted to change your password. The above password is the UNIX password. Set your own password and make sure to remember it.
- After password has been set run the ls command (type ls and press enter) in the command line to verify successful connection with the sandbox.



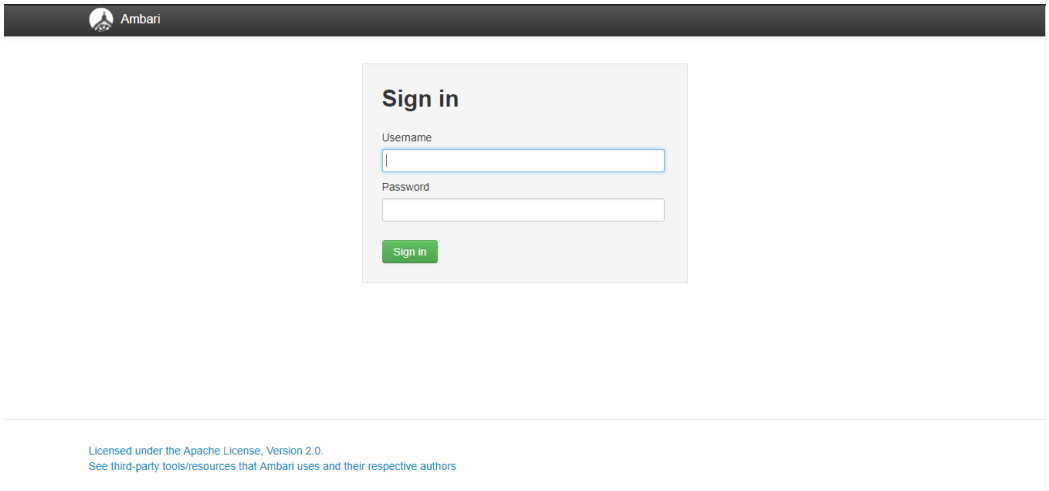
3. Changing Ambari Admin Password

Make sure the virtual machine is running and Putty is set up before doing the following procedure.

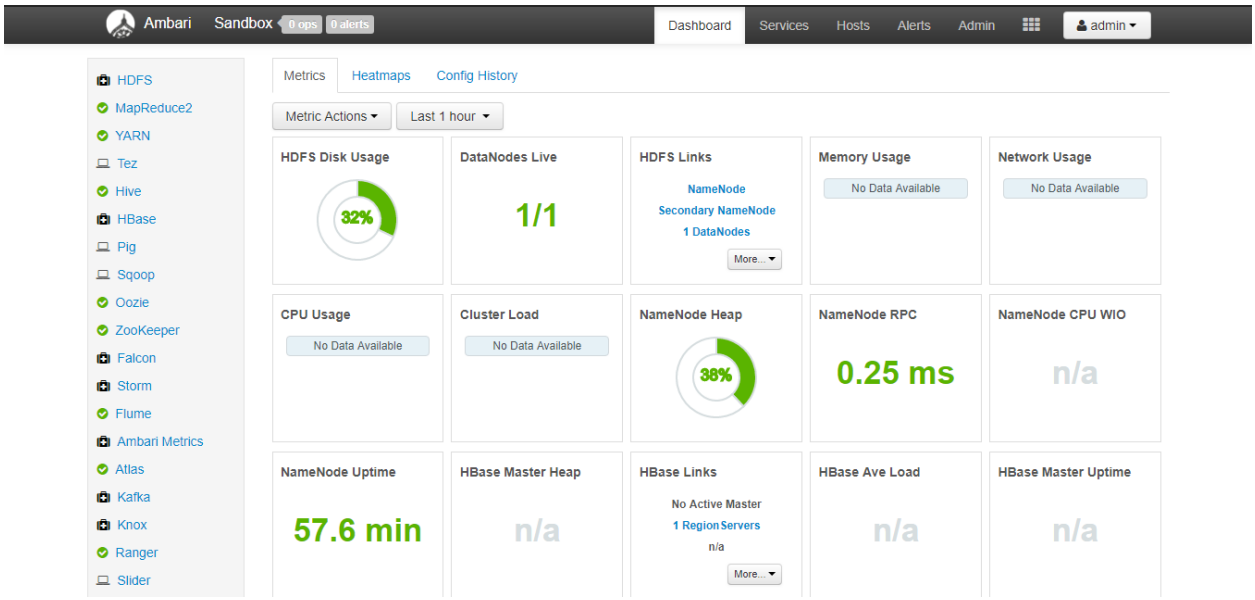
- On the Sandbox command line you are accessing via Putty run the following command **ambari-admin-password-reset**
- Set the new password for the admin account and make sure to remember it. After setting up the password the ambari server will restart for the changes to take effect.

```
[root@sandbox ~]# ambari-admin-password-reset
Please set the password for admin:
Please retype the password for admin:

The admin password has been set.
Restarting ambari-server to make the password change effective...

Using python  /usr/bin/python2
Restarting ambari-server
Using python  /usr/bin/python2
Stopping ambari-server
Ambari Server stopped
Using python  /usr/bin/python2
Starting ambari-server
Ambari Server running with administrator privileges.
Organizing resource files at /var/lib/ambari-server/resources...
Server PID at: /var/run/ambari-server/ambari-server.pid
Server out at: /var/log/ambari-server/ambari-server.out
Server log at: /var/log/ambari-server/ambari-server.log
Waiting for server start.....................
Ambari Server 'start' completed successfully.
[root@sandbox ~]#
```

- Go to your browser and enter your sandbox IP address with the port 8080 as follows: **Your-ip-address:8080**
  For example: 192.168.8.102:8080
- This will open the Ambari login screen.



- For the admin account the username will be admin and the password will be the password you set earlier. After logging in you will see the main Ambari dashboard.



The setup of Hortonworks Sandbox is now complete.

It is recommended to follow this tutorial to familiarize yourself with Hortonworks Sandbox and its various components
https://hortonworks.com/tutorial/hadoop-tutorial-getting-started-with-hdp/

4. Common Problems

This section is a guide to solving commonly occurring problems during the installation of Sandbox. This is not a complete guide and you may or may not find the solution to your problem here. This section only documents problems commonly faced and how to solve them. If you do not find the answer to your issue here then Google is your best friend.

4.1. Checking if Sandbox is accessible

This step is vital to locating your specific problem and finding the solution so make sure to do this step before any other troubleshooting.

- To check if the sandbox is accessible open the Windows command prompt and run the command:
  ping your-ip-address
  For example **ping 192.168.8.102**

If there is a successful reply then the sandbox is running. There can be cases of Request Timeout and Host Unreachable as well which can be due to various reasons. Some of these will be discussed in this section.

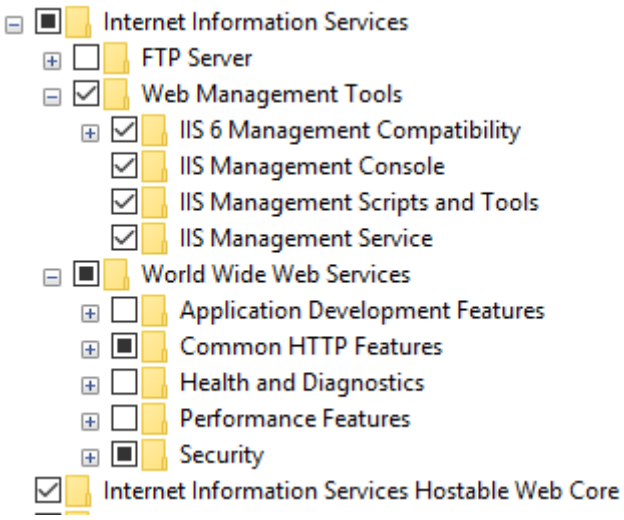
4.2. Ambari not opening in browser but sandbox is accessible

This subsection documents the issue if the sandbox is accessible but the browser UI provided by Ambari is not accessible.

4.2.1. Opening Ambari for the first time

If this is the first time you are accessing ambari after installing Sandbox and it is not opening then it is possible that the Internet Information Service (IIS) of your device is not enabled.

- Go to the Control Panel and click the Programs and Features option.
- Click Turn Windows Features On or Off on the left side of the window. A dialog will appear:
- Go to Internet Information Services option and click the plus next to it. Make sure the settings are as follows:



- Click OK and restart your device.

4.2.2. Normal Ambari Crash

If the ambari service was running fine before the sudden inaccessibility then it is possible that the ambari service has crashed. This issue can be resolved by restarting the Ambari server and Ambari agent.

- On the Sandbox command line run the following commands in the same sequence:
  service ambari-server restart

```
[root@sandbox ~]# service ambari-server restart
Using python  /usr/bin/python2
Restarting ambari-server
Using python  /usr/bin/python2
Stopping ambari-server
Ambari Server stopped
Using python  /usr/bin/python2
Starting ambari-server
Ambari Server running with administrator privileges.
Organizing resource files at /var/lib/ambari-server/resources...
Server PID at: /var/run/ambari-server/ambari-server.pid
Server out at: /var/log/ambari-server/ambari-server.out
Server log at: /var/log/ambari-server/ambari-server.log
Waiting for server start.....................
Ambari Server 'start' completed successfully.
```

service ambari-agent restart

```
[root@sandbox ~]# service ambari-agent restart
Verifying Python version compatibility...
Using python  /usr/bin/python2
Found ambari-agent PID: 1448
Stopping ambari-agent
Removing PID file at /var/run/ambari-agent/ambari-agent.pid
ambari-agent successfully stopped
Verifying Python version compatibility...
Using python  /usr/bin/python2
Checking for previously running Ambari Agent...
Starting ambari-agent
Verifying ambari-agent process status...
Ambari Agent successfully started
Agent PID at: /var/run/ambari-agent/ambari-agent.pid
Agent out at: /var/log/ambari-agent/ambari-agent.out
Agent log at: /var/log/ambari-agent/ambari-agent.log
```

- After both the commands go to the browser and access ambari again. It will work successfully this time.

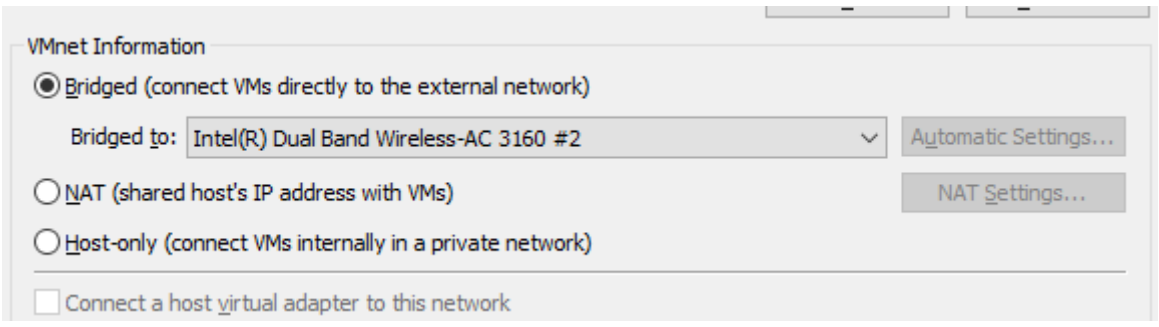
4.3. Sandbox inaccessible. Host Unreachable

If the sandbox is running on the VMware but is inaccessible through command line and ambari and returns a request timed out while pinging then it is possible that this is a network issue and there can be a few reasons for this.

4.3.1. Bridging fault

It is possible that the VMware bridged connection service has failed or has been set to some other configuration than required. This usually occurs when there are multiple cards configured to the bridging service.

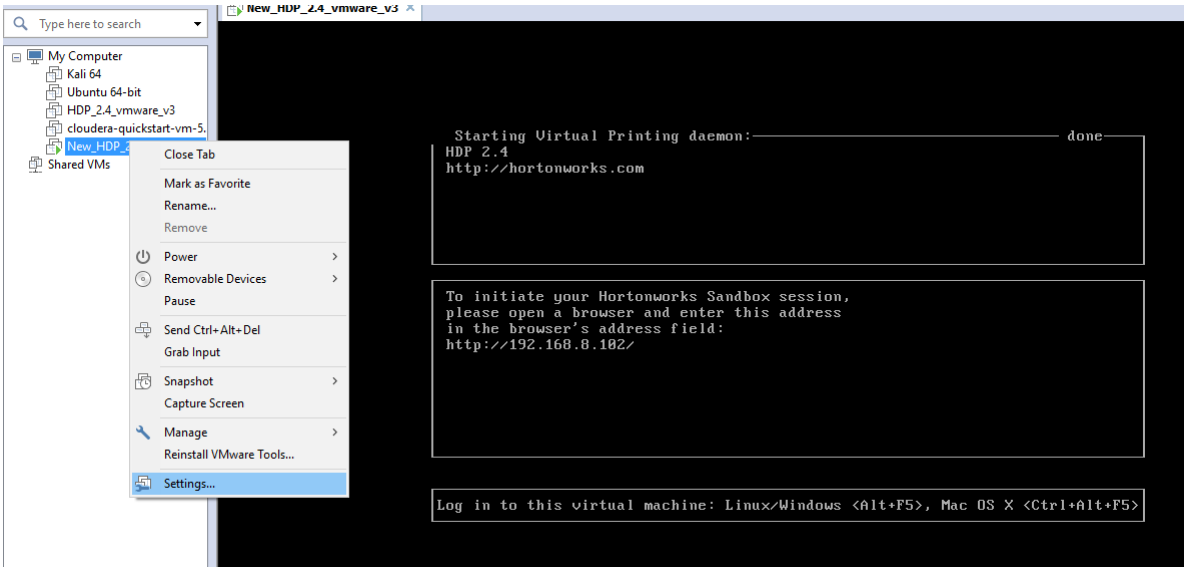
- Open the VMware workstation. Make sure the virtual machine is paused or shut down and that you have admin privileges on your device.
- On the VMware open the Edit menu and click Virtual Network Editor.
- Click change settings and give administrator access to the service.
- Select your network card from the dropdown menu under the Bridged option so that your setting looks something as below and click OK:



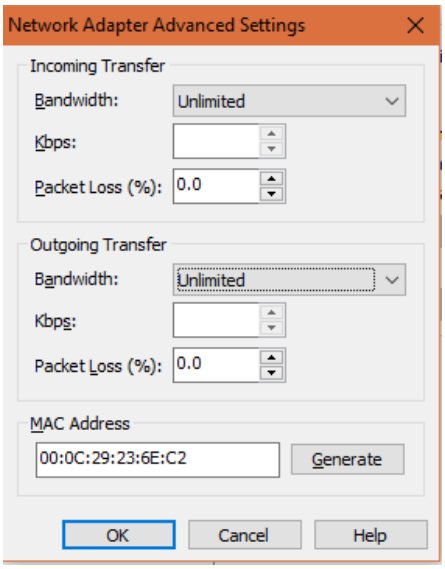
- Resume or start your sandbox service to see if this works.

The above procedure will work most of the time but there can be cases where it does not work. An additional step will be as follows:

- On the VMware workstation right click on your virtual machine and click the settings option



- On the network adapter tab click the Advanced button. A new dialog will appear similar to the one shown below.



- Change the bandwidth of the Incoming Transfer and Outgoing Transfer from Unlimited to Cable (100 mbps) and click OK.
- Ping the sandbox via command line again to see if it works otherwise you will need to restart the virtual machine.