

Hortonworks

Setup Guide for HDP Developer: Custom YARN Applications

Revision 2

Hortonworks University

Overview

The Hortonworks Training Course that you are attending is taught using a virtual machine (VM) for the lab environment. Before attending the course, you must complete the steps in this document. Here is an overview of the steps you will perform:

- Download the virtual machine file. The VM file is in a folder online that we will share with you. The file is very large (~ 4GB) and may take several hours to download if you have a slower Internet connection.
- You must install **VMWare Player** to open the VM file on Windows. On Mac, you can download a free trial of **VMWare Fusion** at <http://www.vmware.com/products/fusion/>.
- You must start the virtual machine after it is imported into VMWare to verify that it can run on your machine's hardware and software.
- It is possible that you will need to configure the BIOS on your Windows machine to enable virtualization.
- **THE FOLLOWING SETUP INSTRUCTIONS MUST BE COMPLETED AT LEAST 24 HOURS PRIOR TO THE START OF THE CLASS!**

IMPORTANT: It is critical for the success of the training that prior to the start of the class each student's computer has the VM file downloaded, imported into VMWare, and verified that the virtual machine starts on the student's computer and a MapReduce can execute successfully.

IMPORTANT: If any of the following steps fail, or if you have any issues or questions, please send an email to training-support@hortonworks.com.

System Requirements

The following is the recommended minimal system requirements:

- Reasonably powerful **x86/amd64 hardware**. Any **recent** Intel or AMD processor should do.
- RAM: At least **8 GB**
- Available Hard-disk: at least **50 GB**
- **64-bit** OS (Windows 7, Windows 8 and Mac OSX)
- At least 800 x 600 display
- VMWare Player for Windows (or VMWare Fusion for Mac)
- **Virtualization must be enabled on BIOS** for Windows based machine
 - Procedures to turn on virtualization settings in your computer's BIOS vary depending on the BIOS manufacturer. Check the information that came with your computer or go to the computer manufacturer's website.
 - Alternatively, to find out whether your computer's CPU is capable of hardware-assisted virtualization, you can download and run the Hardware-Assisted Virtualization Detection Tool (<http://www.microsoft.com/en-us/download/details.aspx?id=592>) provided by Microsoft.
- Internet connection must be available on the system.

Complete the Following Steps

Objective:	Install the necessary software for this course, and also import the VMWare image into VMWare Player and start the virtual machine successfully.
Before You Begin:	Make sure your hardware meets the requirements listed above.

Perform the following steps at least one day prior to the start of the course:

Step 1: Install VMWare Player

1.1. On Windows, download the latest 64-bit version of VMWare Player:

<http://www.vmware.com/support/download-player.html>

On a Mac, download a free trial of VMWare Fusion at

<http://www.vmware.com/products/fusion/>.

1.2. Run the downloaded installation file.

Step 2: Download the VM

2.1. Download the VM here. This file is about 4GB and may take a couple hours to download, depending on your internet connection:

Western US: <http://tinyurl.com/basewestrev2>

Eastern US: <http://tinyurl.com/baseeastrev2>

Europe: <http://tinyurl.com/baseeuroperev2>

Asia: <http://tinyurl.com/baseasiarev2>

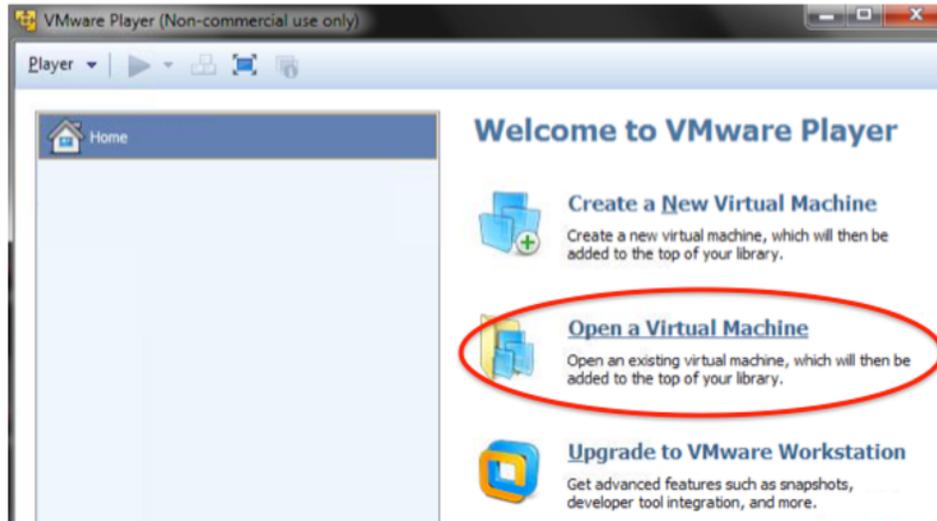
Step 3: Unzip the Virtual Machine

3.1. Unzip the file you just downloaded, which may take a couple of minutes.

NOTE: Some versions of the built-in Windows extraction tool have been known to cause issues. You may want to unzip the file using a third-party tool like 7-zip.

Step 4: Open the VM with VMWare Player

4.1. Start the VMWare Player application and select the option to **Open a Virtual Machine**:



4.2. Browse to the folder you just unzipped, which should be named **HWX_classroom_base_2.1**.

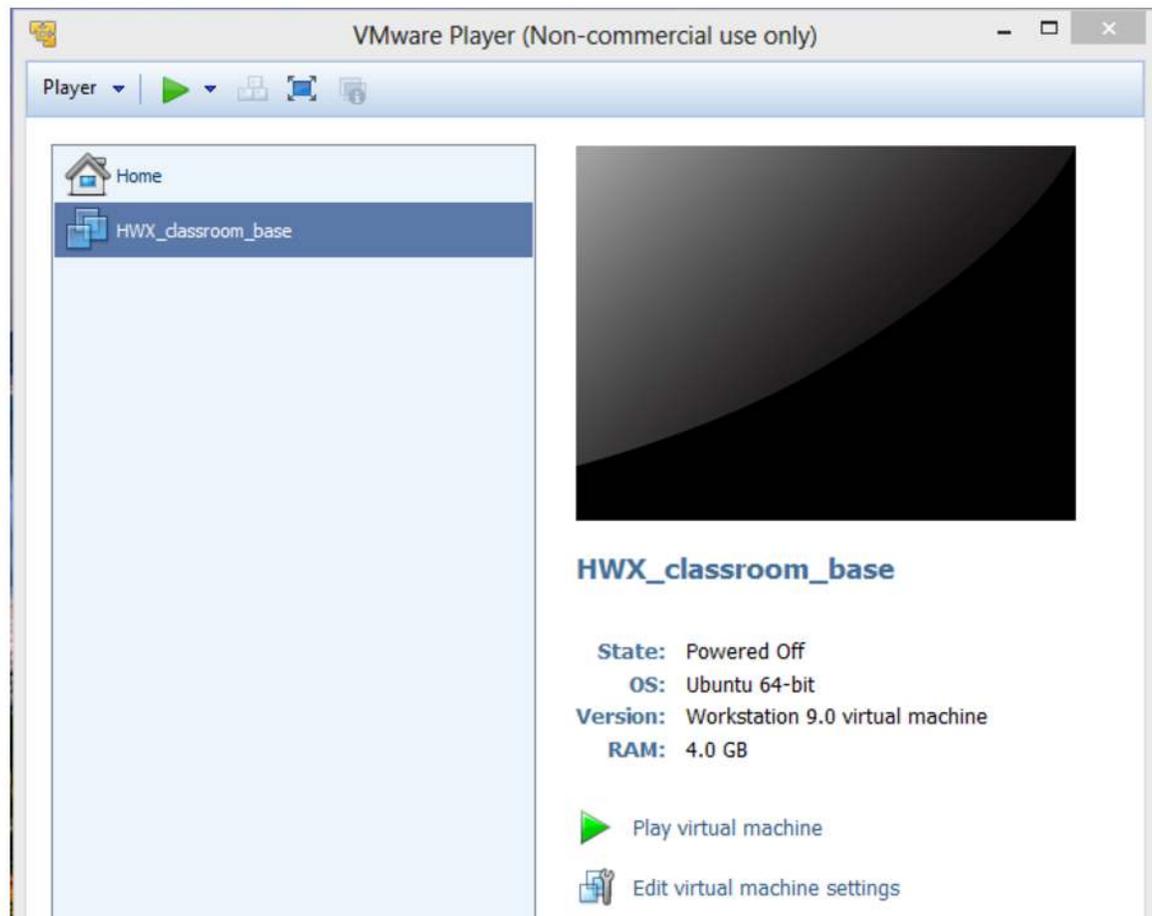
4.3. Within the **HWX_classroom_base_2.1.vmware** folder, select the file **HWX_classroom_2.1_RevN.vmx**.

4.4. Click the **Open** button.

4.5. If a dialog appears stating that the virtual machine may already be in use, click the **Take Ownership** button:



4.6. You should now see the VM in the VMWare Player library:

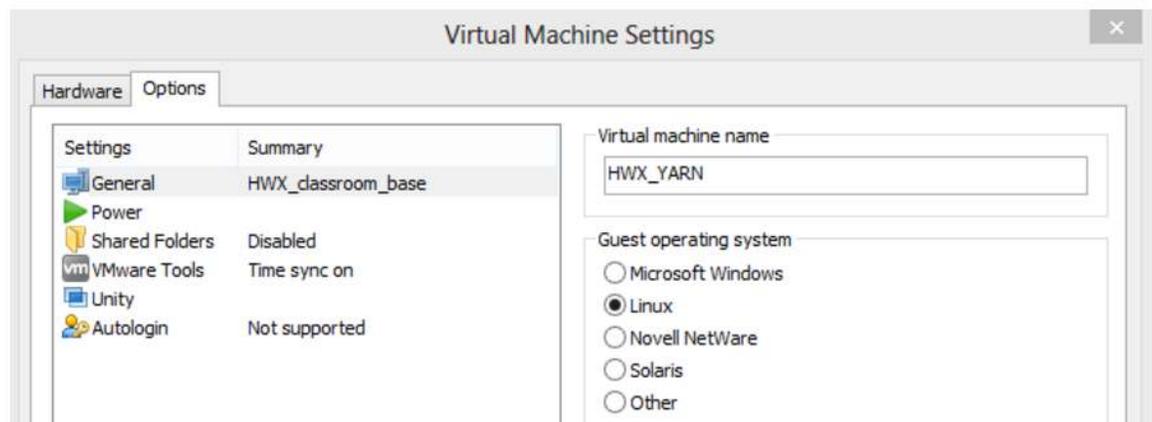


Step 5: Change the VM Name

5.1. Right-click on "HWX_classroom_base" in the list of VMs and select **Settings....**

5.2. Select the **Options** tab.

5.3. Change the **Virtual machine name** to **HWX_YARN**:

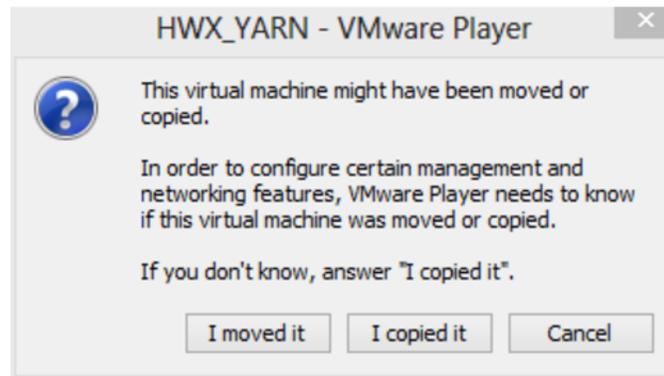


5.4. Select the **Hardware** tab. Make sure the VM is configured to use 4076MB of memory. If your hardware has 16GB or more of RAM, then feel free to increase the memory for this VM to 6 or 8GB. (This is not a requirement – just an option if your hardware has a large amount of physical memory.)

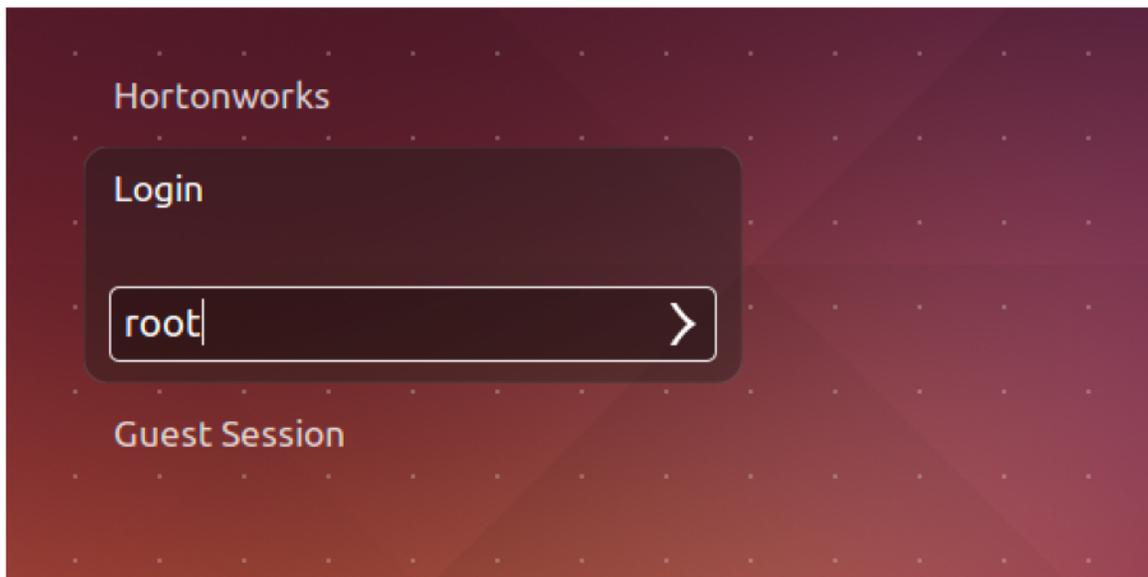
5.5. Click the **OK** button to close the **Virtual Machine Settings** dialog.

Step 6: Start the VM

6.1. Click the **Play virtual machine** link to start your **HWX_YARN** VM. If a dialog appears asking if you moved or copied the VM, click the “**I copied it**” button:



6.2. Wait for the VM to start. When the VM has started successfully, you should see a login screen:



6.3. Login as the **root** user. The password is **hadoop**. When you are logged in, desktop for the VM should display:



Step 7: Install the Course Files

7.1. Open a Terminal window by either clicking on the Terminal shortcut on the left-side toolbar, or pressing **Ctrl+Alt+T**.

7.2. Enter the following command to download and install the course lab files:

```
root@ubuntu:~# ./install_course.sh YARN_Rev2
```

A few programs are being downloaded and installed, so this may take a minute or two.

7.3. There are a couple of checks you can make to verify the course installed properly. First, run the following command:

```
docker images
```

You should see three of them:

```
hwx/hdp_node
```

```
hwx/hdp_node_base
centos
```

The **hwx/hdp_node** image is new, so if it is missing from the list of images then something failed during the build and you should try running the **install_course.sh** command again.

7.4. You should also see a list of subfolders in the **/root/yarn/labs** folder:

```
root@ubuntu:~# ls -la /root/yarn/labs/
```

Step 8: Run a Smoke Test

8.1. To verify that your machine has enough memory, try starting a Hadoop cluster on your VM. Enter the following command (as the **root** user), which starts up a 7 node Hadoop cluster in your VM:

```
root@ubuntu:~# yarn_cluster.sh
```

NOTE: If the above script is not found in your **PATH**, then the **install_course.sh** script likely failed. Try running it again and verify that **yarn_cluster.sh** is in your **/root/scripts** folder.

8.2. To verify all the components of the cluster are running, SSH to the **namenode**:

```
root@ubuntu:~# ssh namenode
```

The password is **hadoop**.

8.3. Enter the **jps** command. You should see the **NameNode** process running:

```
root@namenode:~
File Edit View Search Terminal Help
root@ubuntu:~# yarn_cluster.sh
Starting NameNode...
NameNode started at 172.17.0.2
Formating NameNode and creating initial folders in HDFS...
Starting ResourceManager...
ResourceManager running on 172.17.0.3
Starting a Hive/Oozie server...
Hive/Oozie running on 172.17.0.4
Starting 4 WorkerNodes...
Started node1 on IP 172.17.0.5
Started node2 on IP 172.17.0.6
Started node3 on IP 172.17.0.7
Started node4 on IP 172.17.0.8
Cluster is up and running!
root@ubuntu:~# ssh namenode
The authenticity of host 'namenode (172.17.0.2)' can't be established.
RSA key fingerprint is 0d:1f:22:0e:8a:94:d5:ae:4c:01:69:7e:af:8b:c6:7d.
Are you sure you want to continue connecting (yes/no)? yes
Warning: Permanently added 'namenode,172.17.0.2' (RSA) to the list of known hosts.
root@namenode's password:
[root@namenode ~]# jps
574 Jps
298 NameNode
[root@namenode ~]#
```

8.4. Enter the following command:

```
hdfs dfsadmin -report | more
```

You should see 4 Datanodes available in your cluster:

```
root@namenode:~  
File Edit View Search Terminal Help  
Configured Capacity: 404874444796 (377.07 GB)  
Present Capacity: 349492150471 (325.49 GB)  
DFS Remaining: 349462515712 (325.46 GB)  
DFS Used: 29634759 (28.26 MB)  
DFS Used%: 0.01%  
Under replicated blocks: 0  
Blocks with corrupt replicas: 0  
Missing blocks: 0  
-----  
Datanodes available: 4 (4 total, 0 dead)  
  
Live datanodes:  
Name: 172.17.0.8:50010 (node4)  
Hostname: node4  
Decommission Status : Normal  
Configured Capacity: 101218611199 (94.27 GB)  
DFS Used: 9676410 (9.23 MB)  
Non DFS Used: 13843315077 (12.89 GB)  
DFS Remaining: 87365619712 (81.37 GB)  
DFS Used%: 0.01%  
DFS Remaining%: 86.31%  
Configured Cache Capacity: 0 (0 B)  
--More--
```

8.5. Enter the following command:

```
yarn node -list -all
```

You should see 4 NodeManager nodes in your cluster.

Step 9: Cleanup the Smoke Test

9.1. Enter **exit** to logout of the **nodemanager** node:

```
[root@namenode ~]# exit  
logout  
Connection to namenode closed.  
root@ubuntu:~#
```

9.2. Enter the following command to stop and delete the cluster:

```
root@ubuntu:~# destroy_cluster.sh
```

The output will look similar to the following:

```
root@ubuntu: ~
File Edit View Search Terminal Help
[root@namenode ~]# exit
logout
Connection to namenode closed.
root@ubuntu:~# destroy_cluster.sh
9447c57db162
731d299e8d46
aab87034d05b
86b7954ed004
42689a838499
cc1aa27413bf
b04f97600a8c
9447c57db162
731d299e8d46
aab87034d05b
86b7954ed004
42689a838499
cc1aa27413bf
b04f97600a8c
Cluster successfully destroyed!
```

Step 10: Shutdown the VM

10.1. Click the “gear” icon in the upper-right corner of the desktop and select “Shutdown” to shutdown the virtual machine.

RESULT: You should now have the classroom VM downloaded and imported into VMWare Player (or Fusion). The VM should be able to start successfully and run a Hadoop cluster.

IMPORTANT: If any of these steps failed or if you have any issues or questions, please send an email to training-support@hortonworks.com.