Data Analysis using Hadoop: Module 4, Lesson 1  
Getting Started with HDP 2.5 Hands-On Lab

## Overview

In this lab, you will provision an HDP 2.5 cluster. You will run a hive shell on the cluster and Hadoop HDFS commands.

## Objectives

In this hands-on lab you will learn how to:

* Create and manage an HDP 2.5 cluster.
* Run HDFS commands on an HDP cluster.
* Run and interact with a Hive shell.

## Prerequisites

The following are required to complete this hands-on lab:

* A Microsoft Azure subscription
* A web browser
* PuTTY

**Note:** The Azure portal is continually improved and changed. The steps in this exercise reflect the user interface of the Microsoft Azure portal at the time of writing, but may not match the latest design of portal.

## Exercises

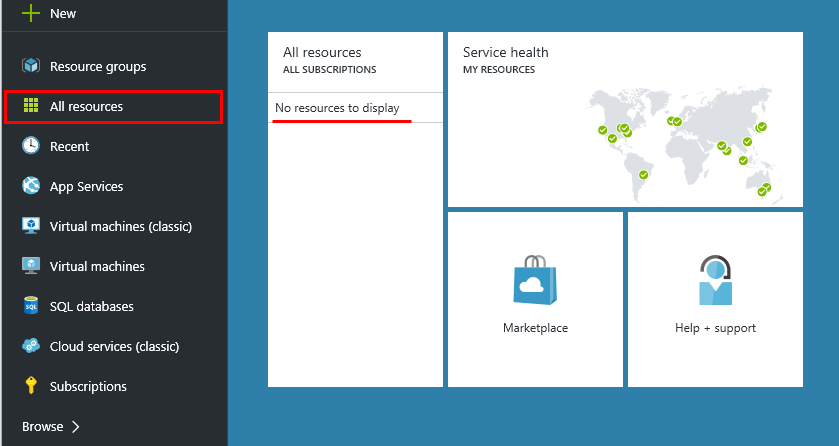
This hands-on lab includes the following exercises:

* Exercise 1: Provisioning and configuring an HDP 2.5.
* Exercise 2: Browser Cluster Storage in cluster & Run hive shell.

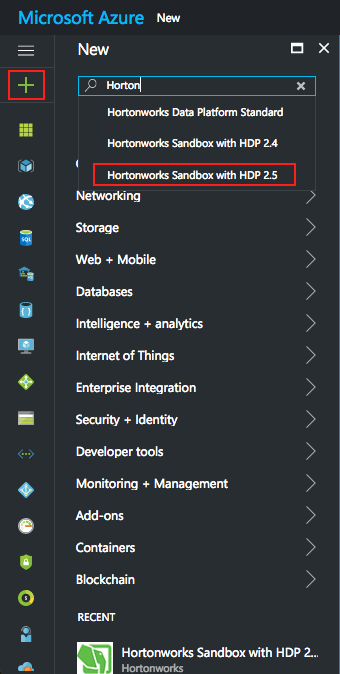
## Exercise 1: Provisioning and configuring an HDP 2.5

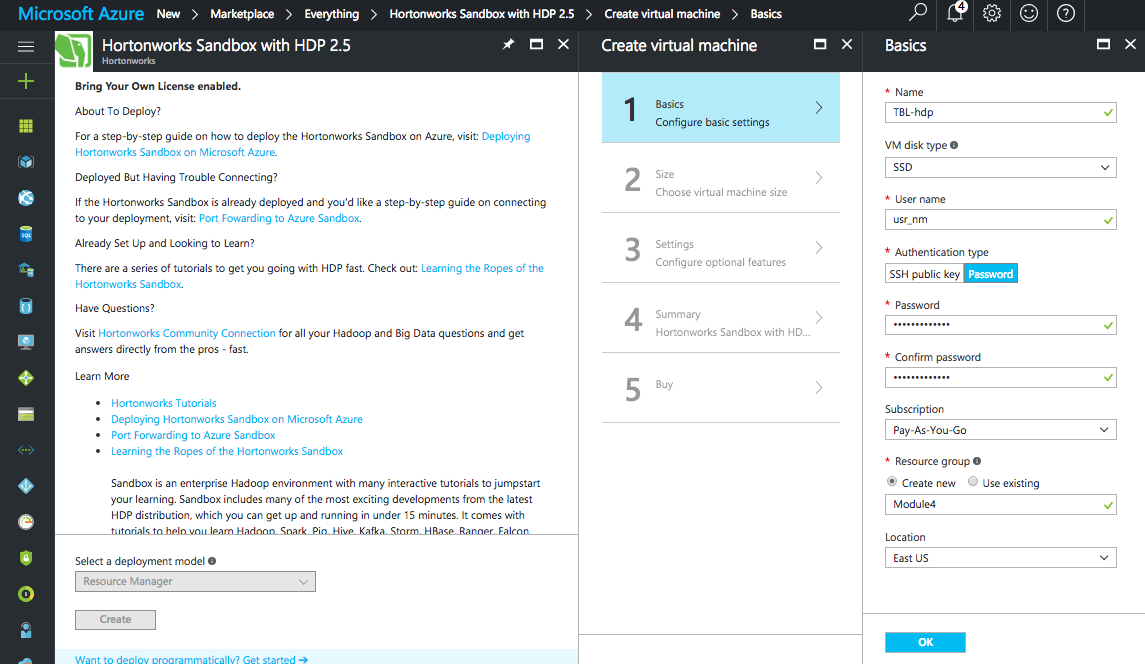
The first task you have to perform is to provision an HDP 2.5 cluster for Hive.

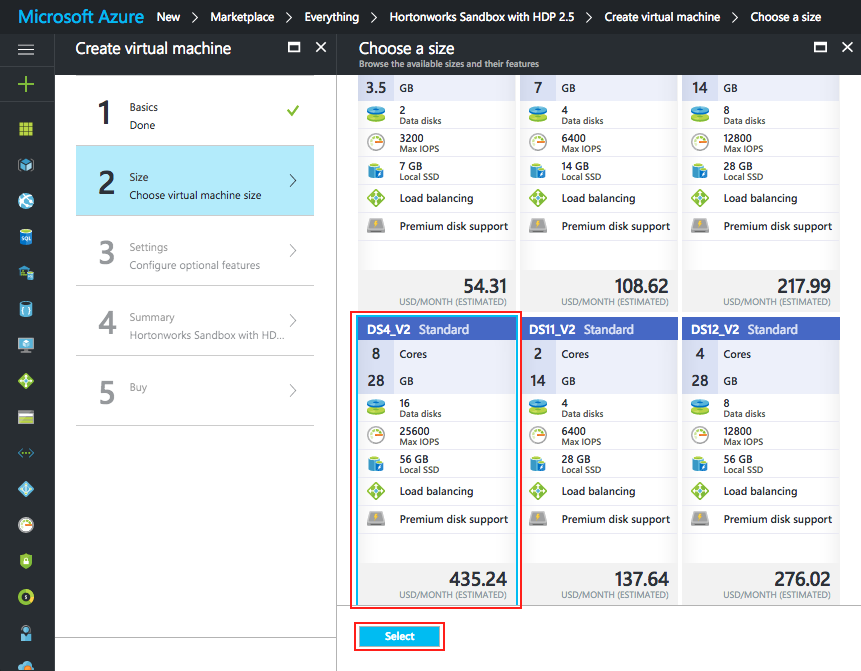
1. In a web browser, navigate to <http://portal.azure.com>. Sign into the portal using your subscription.
2. In the Azure portal, click “All resources”, and verify that there are no existing HDInsight clusters in your subscription.



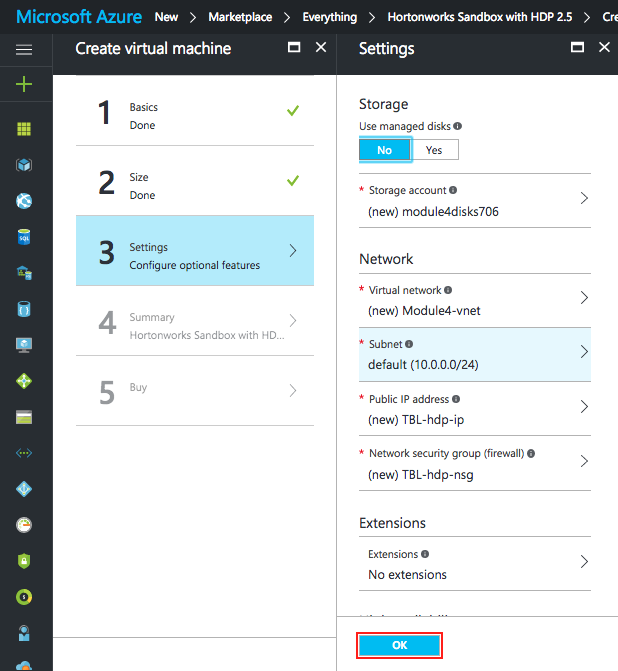
1. In the menu on the left edge, click New (indicated by a +), and search for Hortonworks Sandbox with HDP2.5.

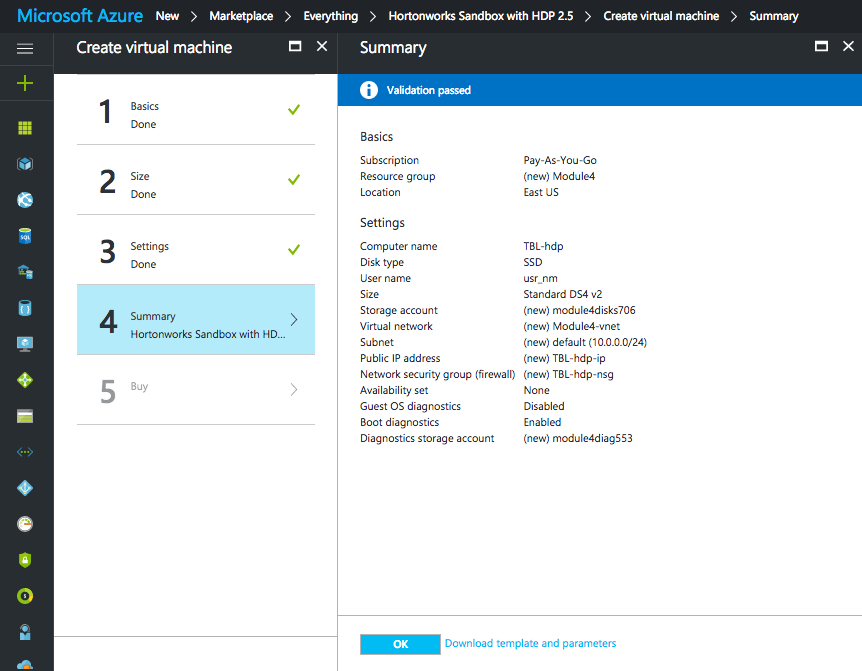


1. Then use New HDP 2.5 Cluster section to create a new cluster, and click “Create”.
2. Enter a value for each of the fields. This is the name you want to reference the machine. Enter a user name of your choice. Enter a strong password. The password must be at least 12 characters in length and must contain at least one digit, one non-alphanumeric, and one upper or lower case letter. Username is the name of the user account used to log into the machine.
3. Enter a unique name in “create a new resource group” for Resource Group. And click the “OK”
4. In size, choose virtual machine size, and click “Select”



1. In settings, you use the default setting for cluster. And, click the “Select” button.



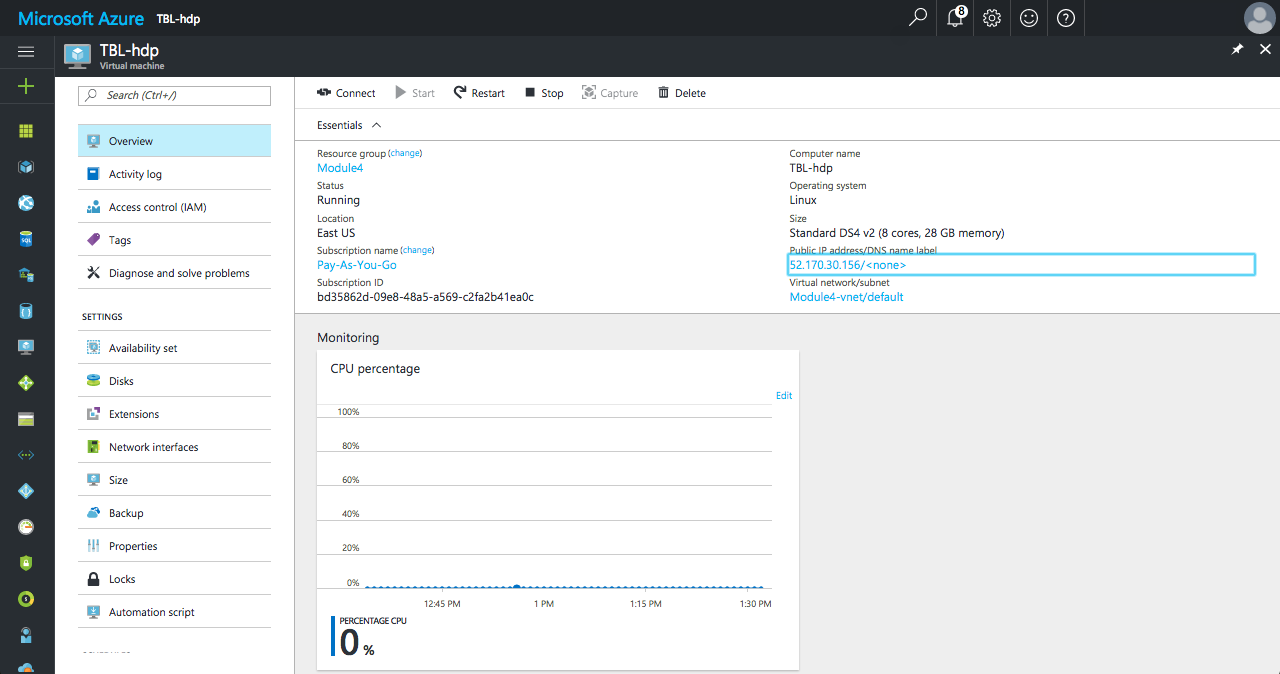
1. In Summary, check your configuration value for validation. If validation passed, click the “OK”. 
2. In “Buy”, you can review the cost for your HDP 2.5 cluster. Then, click the “Purchase” button. (This will utilize the free trial subscription you set up in [Module 1 Lesson 1 Lab](https://github.com/MSFTImagine/computerscience/tree/master/Instructor-Led/Labs/Module1))
3. After you have clicked “Purchase”, wait for the cluster to be provisioned (This can take a while, click on the bell [on the top navigation bar of the Azure portal] to view progress/status). When the cluster has been provisioned you will receive a notification in the Azure Portal.

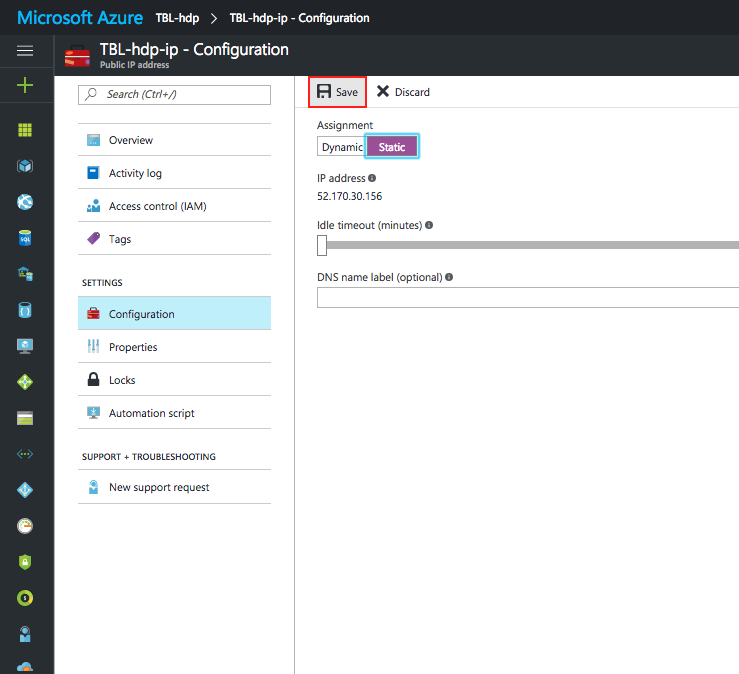
## Exercise 2: View the Cluster Dashboard and browser Cluster Storage in cluster

Now that you’ve created HDP 2.5 for your cluster, you can use it to work with the cluster.

1. In the Azure Portal, click on Virtual Machines in the navigation pane on the left. Then click on “TBL-hdp”, your new Virtual Machine in the HDP 2.5 cluster section
2. Take note of the IP address, in this example it’s 52.170.30.156

Your virtual machine will have a different IP address than the example shown here.

1. Click on the public IP address.
2. Switch the assignment from dynamic to static and save the configuration.



1. Open a PowerShell window and edit your ~/.ssh/config file. Copy and paste the following into your config file:

Host azureSandbox

Port 22

User <your user name>

HostName <your IP>

LocalForward 8080 127.0.0.1:8080

LocalForward 8888 127.0.0.1:8888

LocalForward 9995 127.0.0.1:9995

LocalForward 9996 127.0.0.1:9996

LocalForward 8886 127.0.0.1:8886

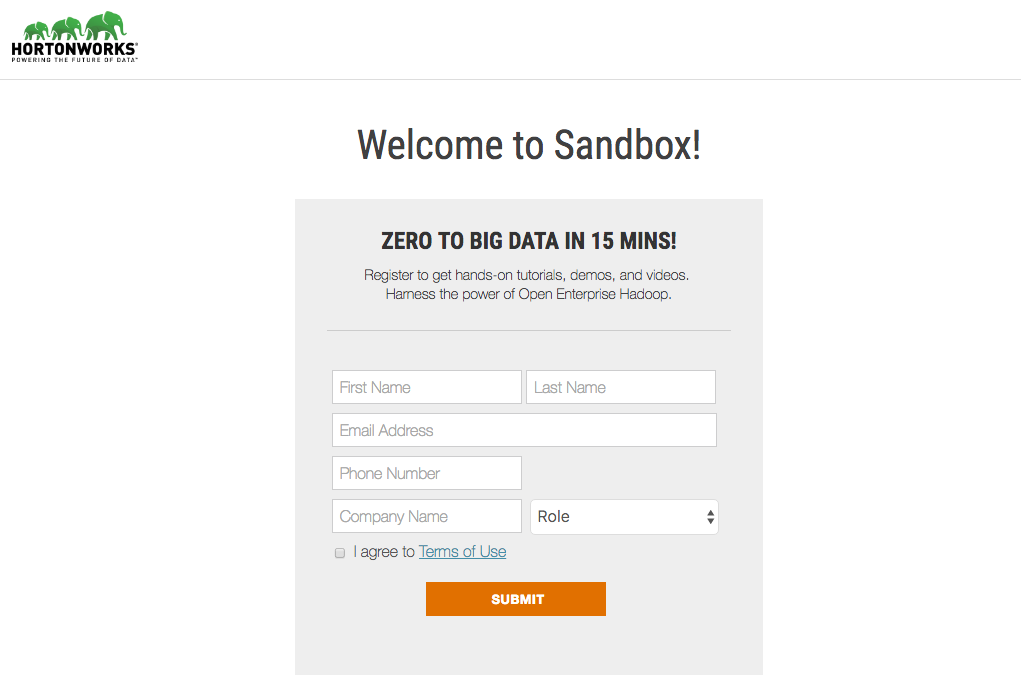
LocalForward 10500 127.0.0.1:10500

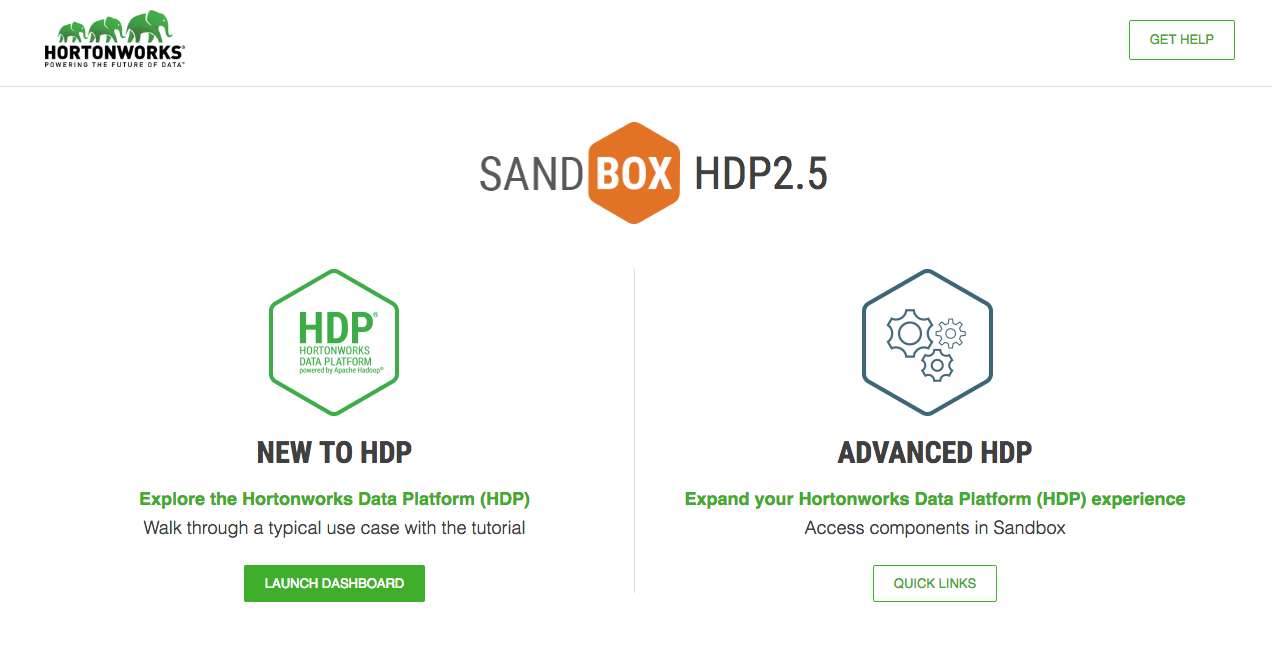
LocalForward 4200 127.0.0.1:4200

LocalForward 2222 127.0.0.1:2222

Save and exit the editor. Note that spacing and capitalization are important.

1. Run the command ssh azureSandbox in your PowerShell window and enter the password you created when you provisioned the cluster.
2. Open the browser tab and type in: localhost:8888

You should see the following page:

Fill out the form and click the submit button to proceed to the following page: 

Now you have the HDP 2.5 Sandbox up and running. If you are new to HDP, click on the link “Launch Dashboard” to get started. Explore the Ambari dashboard using

maria\_dev

as both the username and password.

1. Return to the PowerShell window where you have connected to the HDP cluster. Run the following command

ssh root@<your IP> -p 2222

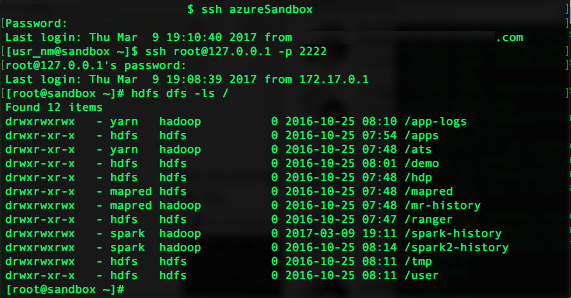
You will be prompted for a password, enter:

hadoop

Once you have successfully logged in, you will be immediately prompted to change the password.

After changing the password, enter the following command to check the contents of the HDFS directory.

$hdfs dfs -ls /

This will print the contents of the root directory of the HDFS file system. It should look something like this:

\*\*\*Don’t forget to delete your cluster in order to avoid wasting your Azure credit\*\*\*

## Summary

In this hands-on lab, you learned how to:

* Create and manage a virtual machine HDInsight
* Connect to HDInsight cluster via PowerShell
* Run HDFS commands

For more information on Sandbox and HDFS commands:

<https://hortonworks.com/hadoop-tutorial/learning-the-ropes-of-the-hortonworks-sandbox/>