Data Analysis using Hadoop: Module 4, Lesson 3  
Getting Started with HDInsight Hands-On Lab

## Overview

In this lab, you will provision an HDInsight cluster. You will then run a hive shell on the cluster. Finally you will explore the Ambari dashboard for your new cluster. See [Module 4 Lesson 3](https://github.com/MSFTImagine/computerscience/tree/master/Complimentary%20Course%20Content/Module4/Lessons) slides for more information on this lab.

## Objectives

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

* Create and manage an HDInsight cluster.
* Connect to an HDInsight cluster using PuTTY.
* View the Ambari dashboard for an HDInsight cluster.

## Prerequisites

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

* A Microsoft Azure subscription
* A web browser
* PuTTY

Note:

You should have already completed the [Module 4 Lesson 1 Lab](https://github.com/MSFTImagine/computerscience/tree/master/Complimentary%20Course%20Content/Module4/Labs) on provisioning an HDP 2.4 cluster.

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:

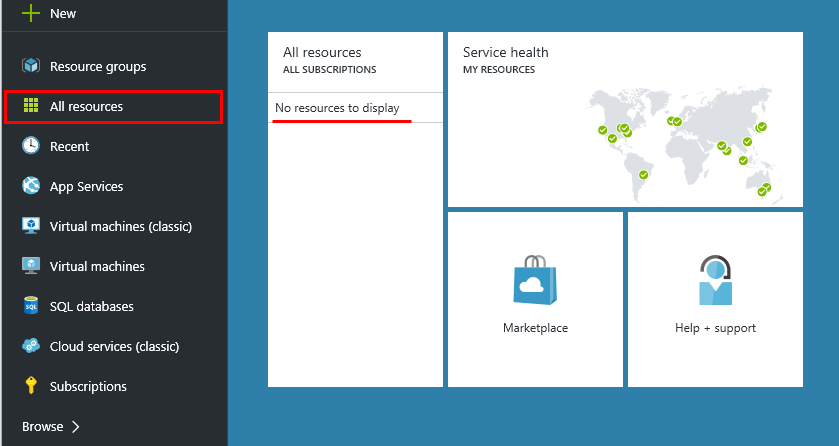
* Exercise 1: Provisioning and configuring an HDInsight Cluster
* Exercise 2: Connect to HDInsight Cluster using PuTTY
* Exercise 3: View the Cluster Dashboard

## Exercise 1: Provisioning and configuring an HDInsight Cluster

The first task you will perform is provisioning an HDInsight cluster for Hive.

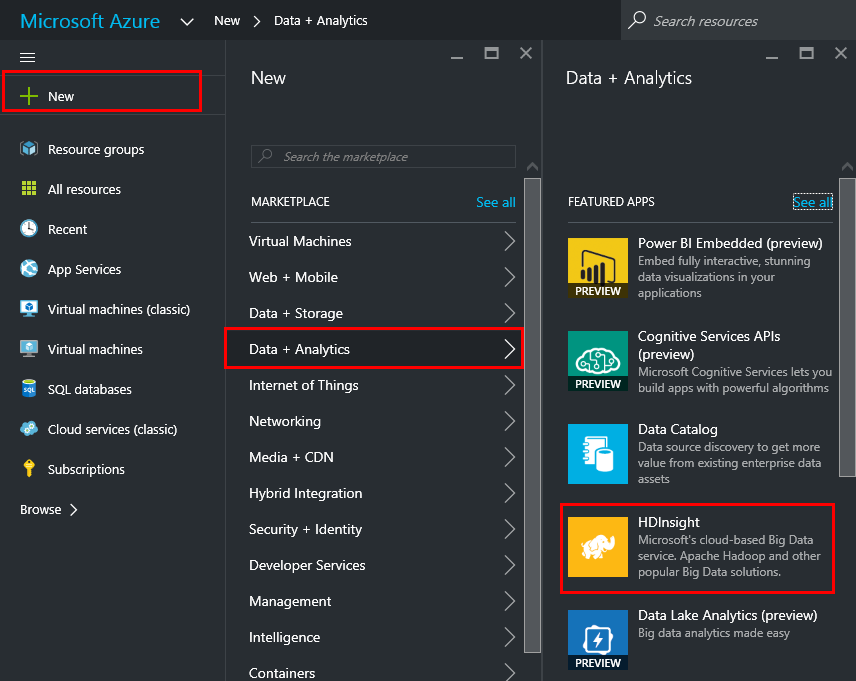
1. In a web browser, navigate to <http://portal.azure.com>. Sign into the portal using your subscription.

In the Azure portal, click “All resources”, and verify that there are no existing HDInsight clusters in your subscription.



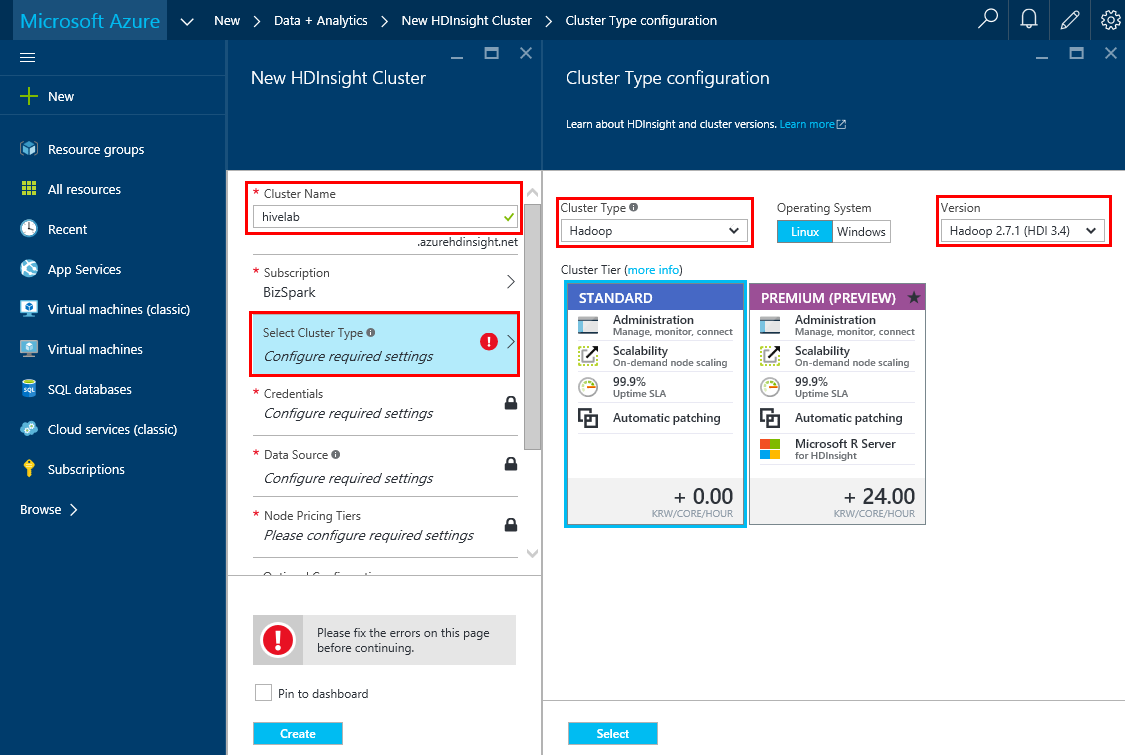
1. In the Hub menu (on the left edge), click New (indicated by a +).

Under “Data + Analytics,” click on HDInsight. Then use New HDInsight Cluster section to create a new cluster.



1. Create a new cluster with configuration value. The Cluster Name is unique name for cluster. You have to remember this name. Then, click the “Select cluster type” for more detailed information. Select the value below as:

* Cluster type : Hadoop
* Cluster Operating System : Linux
* Version : Choose the latest version of Hadoop. Ex) 2.7.1 (HDI 3.4)
* Cluster Tier : Standard
* And, click the “select” button.

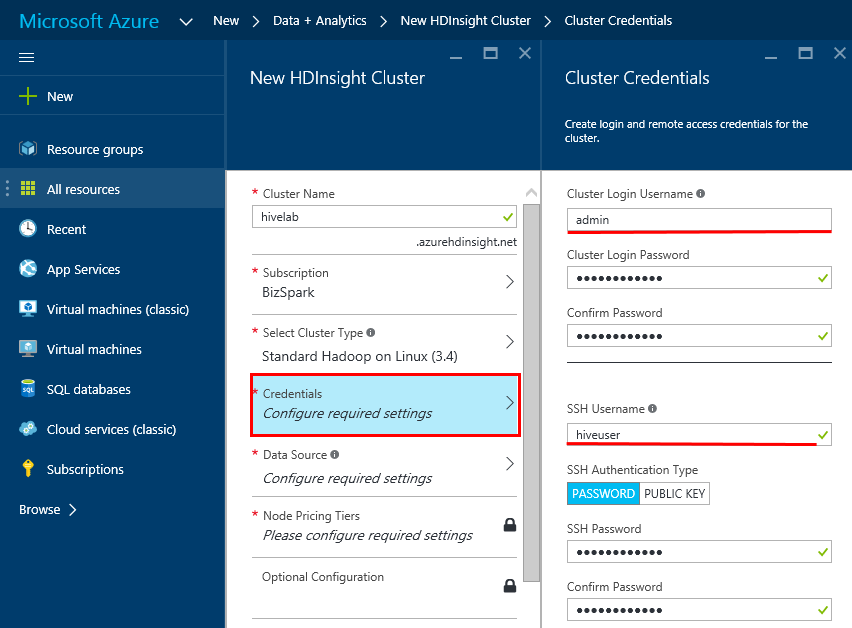


1. Click the “Credentials” section.

Enter a user name of your choice (the default is “admin.” Your SSH Username is different than your Azure username).

Enter and confirm a strong password (the password must be at least 10 characters in length and must contain at least one digit, one non-alphanumeric, and one upper or lower case letter).

Click the “select” button.

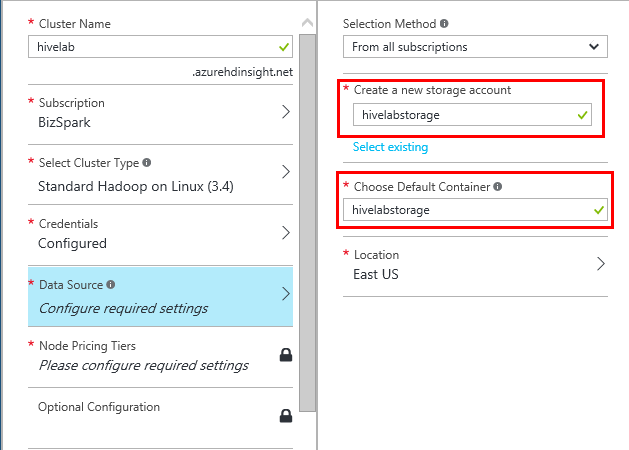


1. Click the “Data Source” section.

Enter a unique name consisting of lower-case letters and numbers only in storage account and choose default container.

Select any available region in Location field.

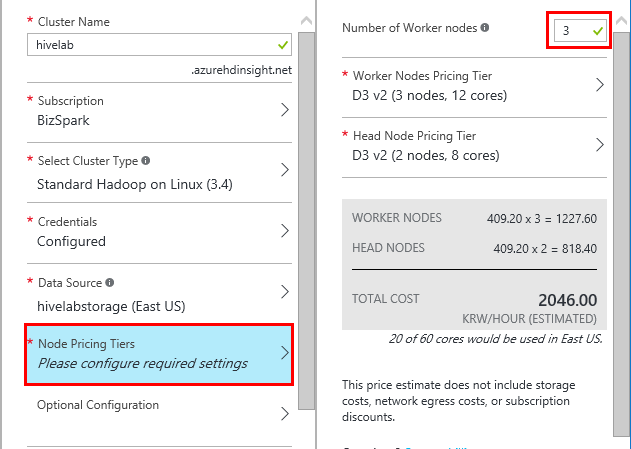
Click the “select” button.



1. Under “Pricing”, set the Number of Worker Nodes to 3.

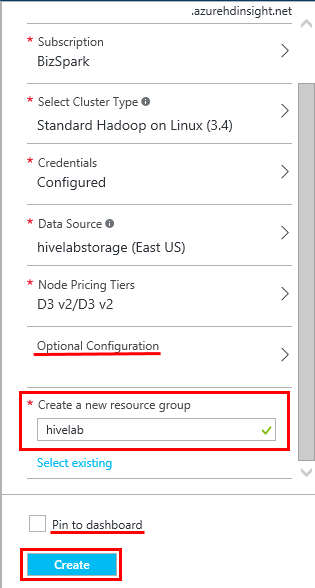
Worker Nodes size and Head Node size, should auto populate.

Click the “select” button.



1. Below “Create a new resource group” enter a unique name for your new Resource Group. Unselect “Pin to dashboard”

Click the “Create” button.



1. After you have clicked “Create,” wait for the cluster to be provisioned. (This can take a while, click on the bell to view progress/status).

When the cluster has been provisioned you will receive a notification in the Azure Portal and the credit in your subscription will start being charged.

At the end of the lab, remember to delete your cluster in order to avoid unnecessarily using your Azure credit.

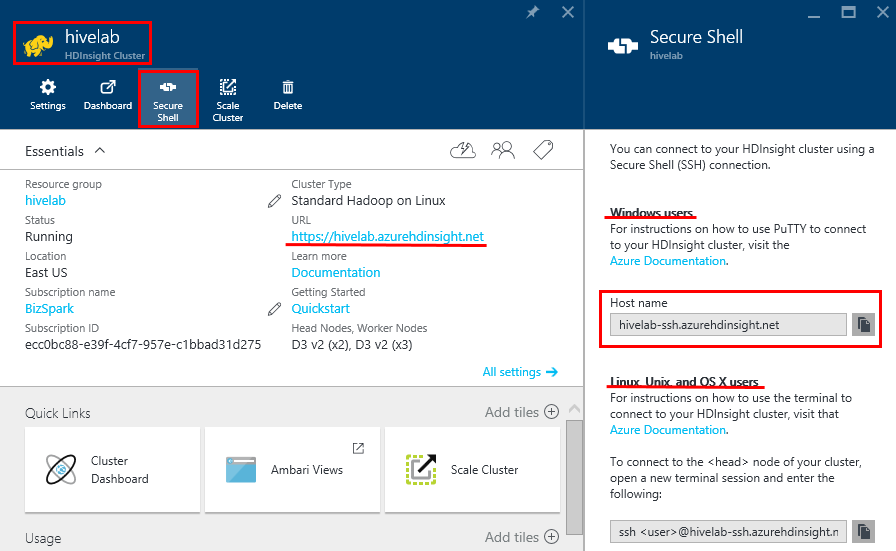
## Exercise 2: Connecting to HDInsight using PuTTY.

Now that you have provisioned an HDInsight cluster, you can connect to it using PuTTY and use PuTTY to work with your cluster.

1. In the Azure portal, enter the “All Resources” dashboard from the navigation pane on the left.

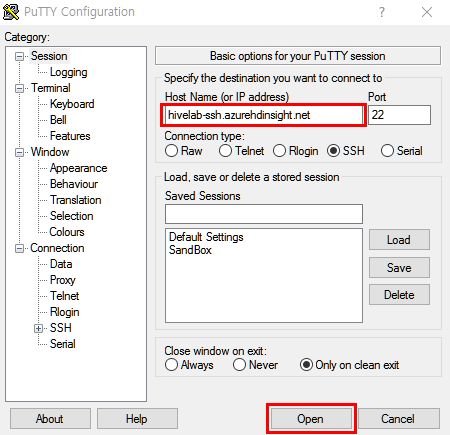
Select the HDInsight Cluster you just provisioned (it should be named hivelab). Enter the Secure Shell dashboard by clicking on Secure Shell as shown below.

Copy the Host name (which should be hivelab-ssh.azurehdinsight.net) to the clipboard.

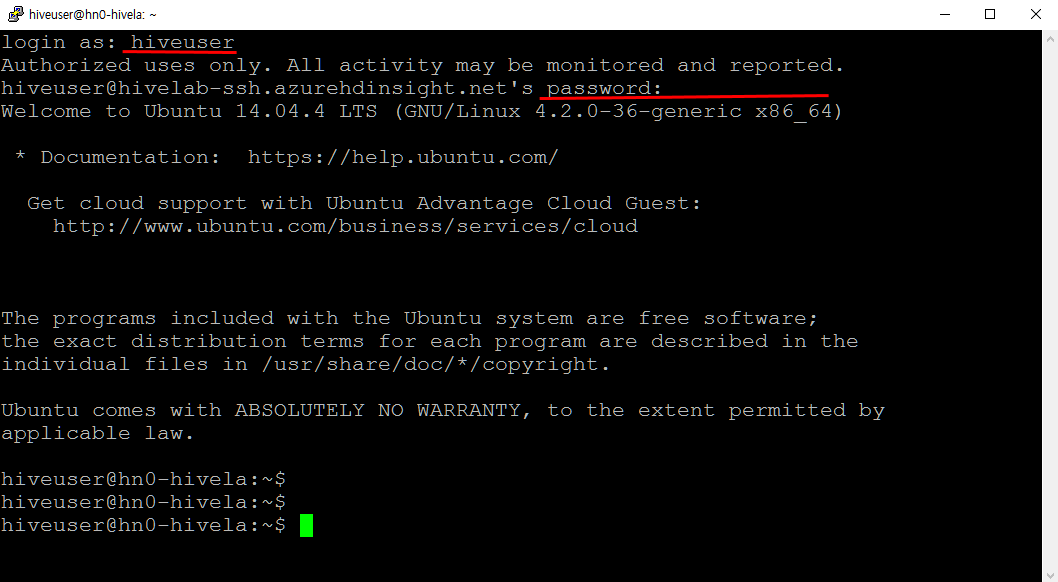


1. Open PuTTY and paste the host name into the Host Name field.

Then click Open. You can now connect to HDInsight using your PuTTY terminal.



1. In PuTTY, enter the SSH username and password you specified when provisioning the cluster.



1. In the PuTTY console, enter the following command to use hive shell.

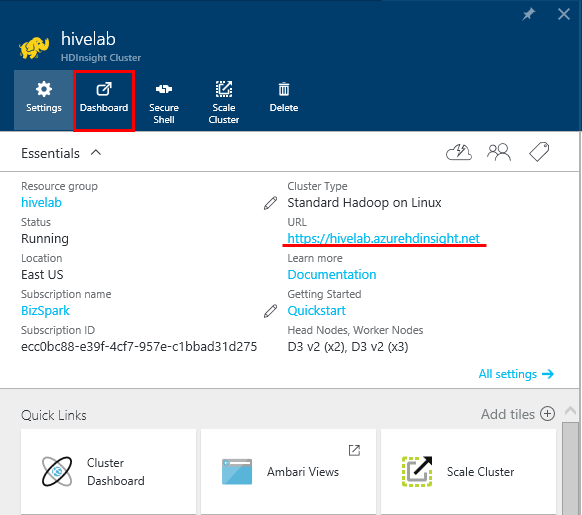
$ hive

You will practice more hive shell in the next lab ([Module 4 Lesson 9 Lab](https://github.com/MSFTImagine/computerscience/tree/master/Complimentary%20Course%20Content/Module4/Labs)).

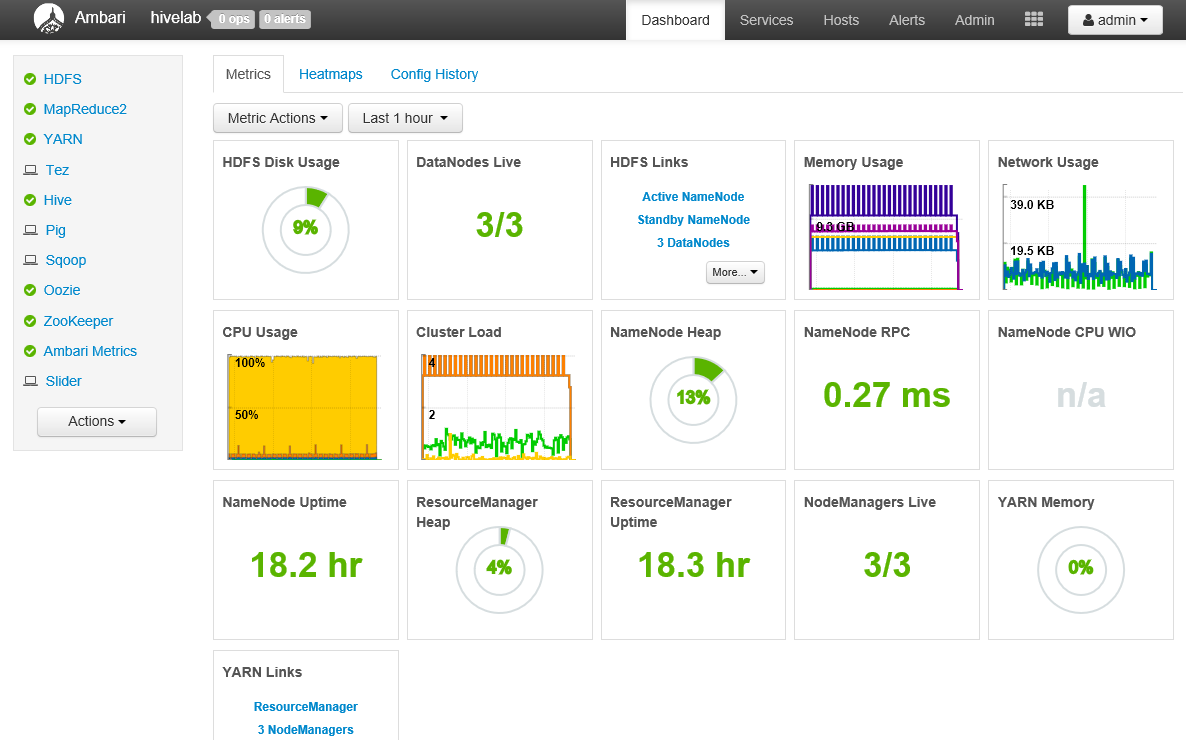
## Exercise 3: View the Cluster Dashboard in a Browser.

Now that you’ve connected to your cluster via PuTTY, you will use the Azure Portal in a web browser to access the Ambari dashboard for your HDInsight cluster.

1. Back in the Azure portal, click “Dashboard” under the HDInsight cluster section. When prompted, log in using the cluster username and password that you specified when provisioning the cluster (Exercise 1).



1. The dashboard is an Ambari web application, in which you can view and configure settings for your Hadoop cluster. Take a few minutes to look over the dashboard for your cluster.



\*\*\*Don’t forget to delete your cluster in order to avoid wasting your Azure credit\*\*\*(But it is highly recommended that you continue to the [Module 4 Lesson 6](https://github.com/MSFTImagine/computerscience/tree/master/Complimentary%20Course%20Content/Module4/Labs) lab so you don’t have to re-provision your cluster)

## Summary

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

* Create and manage an HDInsight cluster.
* Connect to an HDInsight cluster using PuTTY.
* View the Ambari dashboard for an HDInsight cluster.