Assigning HBase Resources in HDInsight

Lab 6 - Getting Started with HBase

Overview

In this lab, you will provision an HDInsight HBase cluster. You will then create an HBase table and use it to store data.

What You’ll Need

To complete the labs, you will need the following:

* A web browser
* A Microsoft account
* A Microsoft Azure subscription
* A Microsoft Windows computer containing: o Microsoft Visual Studio/VS code

o The lab doc for this course

**Note**: To set up the required environment for the lab, follow the instructions in the **Setup** document forthis course. Specifically, you must have signed up for an Azure subscription.

Provisioning an HDInsight HBase Cluster

The first task you must perform is to provision an HDInsight HBase cluster.

**Note**: The Microsoft Azure portal is continually improved in response to customer feedback. The steps inthis exercise reflect the user interface of the Microsoft Azure portal at the time of writing, but may not match the latest design of the portal exactly.

Provision an HDInsight Cluster

1. In a web browser, navigate to [http://portal.azure.com,](http://portal.azure.com/) and if prompted, sign in using the Microsoft account that is associated with your Azure subscription.
2. In the Microsoft Azure portal, add a new HDInsight cluster with the following settings:
   * **Cluster Name**:*Enter a unique name (and make a note of it!)*
   * **Cluster Type**: HBase 1.1.2(HDI 3.6) [As of writing this Lab]
   * **Operating System**:*Choose the latest version of Linux Server*
   * **Subscription**:*Your Azure subscription*
   * **Resource Group**:*Create a new resource group with a unique name*
   * **Credentials**:
     + **Cluster Login Username**:*Enter a user name of your choice (and make a note of it!)*
     + **Cluster Login Password**:*Enter and confirm a strong password (and make a note of it!)*
     + **Enable Remote Desktop:** Yes
     + **Expires on**:*Select the date a week from now*
     + **SSH Username:** *Enter another user name of your choice (and make a**note of it!)*
     + **SSH Password:** *Enter and confirm a strong password (and make a note**of it!)*
   * **Data Source**:
     + **Selection Method**: From all subscriptions
     + **Create a new storage account**:*Enter a unique name for your storage account (and**make a note of it!)*
     + **Choose Default Container**:*The name of your cluster*
     + **Location**:*Select any available region*.
   * **Optional Configuration**:
     + **HDInsight Version**:*Select the most recent version available*
     + **Virtual Network**: Not Configured
     + **External Metastores**: Not Configured
     + **Script Actions**: Not Configured
     + **Azure Storage Keys**: Not Configured
   * **Node Pricing Tiers**:
     + **Number of Worker nodes**: 1
     + **Worker Nodes Pricing Tier**:*Leave the default selection*
     + **Head Node Pricing Tier**:*Leave the default selection*
   * **Pin to Startboard**:*Not selected*
3. In the Azure portal, view **Notifications** to verify that deployment has started. Then wait for the cluster to be deployed (this can take a long time – often 30 minutes or more. Now may be a good time to go and have a cup of coffee!)

**Note**: As soon as an HDInsight cluster is running, the credit in your Azure subscription will start to becharged. The free-trial subscription includes a credit limit of approximately $100 (or local equivalent) that you can spend over a period of 30 days, which is enough to complete the labs in this course as long as clusters are deleted when not in use. If you decide not to complete this lab, follow the instructions in the *Clean Up* procedure at the end of the lab to delete your cluster in order to avoid using your Azure credit unnecessarily.

Creating an HBase Table

Now that you have provisioned an HDInsight HBase cluster, you can create HBase tables and store data in them.

Open a Remote Desktop Connection to the Cluster

1. In the Azure portal, browse to the HBase cluster you just created.
2. Open a SSH connection to the cluster using the SSH username and password you specified when provisioning the cluster.
3. When the SSH connection connects, open the **Hadoop Command Line** console and view the syntax documentation for the Hadoop command line tool.

Create an HBase Table

**Note**: The commands in this procedure arecase-sensitive.

1. In the Hadoop Command Line console window, enter the following command to change the current directory to the HBase installation directory:

cd %HBASE\_HOME%\bin

1. Enter the following command to start the HBase shell. hbase shell
2. Enter the following command to create a table named **Stocks** with two column families named **Current** and **Closing**.

create 'Stocks', 'Current', 'Closing'

1. Enter the following command to insert a field for a record with the key **ABC** and a value of **97.3** for a column named **Price** in the **Current** column family.

put 'Stocks', 'ABC', 'Current:Price', '97.3'

1. Enter the following command to insert a field for record **ABC** and a value of **95.7** for a column named **Price** in the **Closing** column family.

put 'Stocks', 'ABC', 'Closing:Price', '95.7'

1. Enter the following command to return all rows from the table. scan 'Stocks'
2. Verify that the output shows the two values you entered for the row ABC, as shown here:

ROW

ABC

ABC

COLUMN+CELL column=Closing:Price, column=Current:Price,

timestamp=*nnn*, value=95.7 timestamp=*nnn*, value=97.3

1. Enter the following command to insert a field for record **ABC** and a value of **Up** for a column named **Status** in the **Current** column family.

put 'Stocks', 'ABC', 'Current:Status', 'Up'

1. Enter the following command to return the values for row ABC. get 'Stocks', 'ABC'
2. Verify that the output shows the values of all cells for row ABC, as shown here:

COLUMNCELL

Closing:Pricetimestamp=*nnn*, value=95.7

Current:Pricetimestamp=*nnn*, value=97.3

Current:Statustimestamp=*nnn*, value=Up

1. Enter the following command to set the **Price** column in the **Current** column family of row **ABC** to **99.1**.

put 'Stocks', 'ABC', 'Current:Price', '99.1'

1. Enter the following command to return the values for row ABC.

get 'Stocks', 'ABC'

1. Verify that the output shows the updated values of all cells for row ABC, as shown here:

COLUMNCELL

Closing:Pricetimestamp=*nnn*, value=95.7

Current:Pricetimestamp=*nnn*, value=99.1

Current:Statustimestamp=*nnn*, value=Up

1. Note the **timestamp** value for the **Current:Price** cell. Then enter the following command to retrieve the previous version of the cell value by replacing ***nnn-1*** with the timestamp for **Current:Price** minus 1 (for example, if the timestamp for **Current:Price** in the results above is144012345678, replace ***nnn-1*** with 144012345677.)

get 'Stocks', 'ABC', {TIMERANGE=>[0,***nnn-1***]}

1. Verify that the output shows previous **Current:Price** value, as shown here:

COLUMNCELL

Closing:Pricetimestamp=*nnn*, value=95.7

Current:Pricetimestamp=*nnn*, value=97.3

Current:Statustimestamp=*nnn*, value=Up

1. Enter the following command to delete the **Status** column in the **Current** column family of row **ABC**.

delete 'Stocks', 'ABC', 'Current:Status'

1. Enter the following command to return the values for row ABC. get 'Stocks', 'ABC'
2. Verify that the **Current:Status** cell has been deleted as shown here:

COLUMNCELL

Closing:Pricetimestamp=*nnn*, value=95.7

Current:Pricetimestamp=*nnn*, value=99.1

1. Enter the following command to exit the HBase shell and return to the Hadoop command line. quit
2. Minimize the remote desktop window (you will return to the Hadoop Command Line later.)

Sample Commands for HBase Table

1. Enter the following command to start the HBase shell. hbase shell
2. Enter the following command to return all rows from the table. scan 'Stocks'
3. Enter the following command to return only the **Current:Price** column for each row: scan 'Stocks', {COLUMNS => 'Current:Price'}

...

1. Enter the following command to return only the first three rows: scan 'Stocks', {LIMIT => 3}

9. Minimize the remote desktop window.

Clean Up

Now that you have finished using HBase, you can delete your cluster and the associated storage account. This ensures that you avoid being charged for cluster resources when you are not using them. If you are using a trial Azure subscription that includes a limited free credit value, deleting the cluster maximizes your credit and helps to prevent using it all before the free trial period has ended.

Delete the Resource Group

1. Close the browser tab containing the HDInsight Query Console if it is open.
2. In the Azure portal, view your **Resource groups** and select the resource group you created for your cluster. This resource group contains your cluster and the associated storage account.
3. In the blade for your resource group, click **Delete**. When prompted to confirm the deletion, enter the resource group name and click **Delete**.
4. Wait for a notification that your resource group has been deleted.
5. Close the browser.