

LAB 05 – HBase in Azure HDInsight

Paso 1: Pre-Requisitos

1. Clúster de HDInsight con Apache HBase.

Paso 2: Creación de tablas e inserción de datos

Puede usar SSH para conectarse a los clústeres de HBase y, después, usar el shell de Apache HBase para crear tablas de HBase e insertar y consultar datos.

Para la mayoría de las personas, los datos aparecen en formato tabular:

Datos tabulares de Apache HBase de HDInsight.

Para usar el shell de HBase, siga estos pasos:

1. Use el comando **ssh** para conectarse al clúster de HBase. Modifique el comando siguiente: reemplace **CLUSTERNAME** por el nombre del clúster y, luego, escriba el comando:

ssh sshuser@CLUSTERNAME-ssh.azurehdinsight.net

2. Use el comando **hbase shell** para iniciar el shell interactivo de HBase. Escriba el siguiente comando en la conexión SSH:

hbase shell

3. Use el comando **create** para crear una tabla de HBase con dos familias de columnas. Los nombres de columna y tabla distinguen mayúsculas de minúsculas. Escriba el comando siguiente:

create 'Contacts', 'Personal', 'Office'

4. Use el comando **list** para mostrar todas las tablas de HBase. Escriba el comando siguiente:

list



5. Use el comando **put** para insertar valores en una columna especificada en una fila especificada de una tabla determinada. Escriba los siguientes comandos:

```
put 'Contacts', '1000', 'Personal:Name', 'John Dole'
```

```
put 'Contacts', '1000', 'Personal:Phone', '1-425-000-0001'
```

```
put 'Contacts', '1000', 'Office:Phone', '1-425-000-0002'
```

```
put 'Contacts', '1000', 'Office:Address', '1111 San Gabriel Dr.'
```

6. Use el comando **scan** para buscar y devolver los datos de tabla **Contacts**. Escriba el comando siguiente:

```
scan 'Contacts'
```

7. Use el comando **get** para capturar el contenido de una fila. Escriba el comando siguiente:

```
get 'Contacts', '1000'
```

Observará que los resultados son similares al uso del comando scan porque solo hay una fila.

8. Use el comando **exit** para detener el shell interactivo de HBase. Escriba el comando siguiente:

Exit

Para cargar datos de forma masiva en la tabla HBase de contactos

1. Se puede encontrar un archivo de datos de ejemplo en un contenedor de blobs público,
<wasb://hbasecontacts@hdtutorialdata.blob.core.windows.net/contacts.txt>.
El contenido del archivo de datos es:



8396	Calvin Raji	230-555-0191	230-555-0191	5415 San Gabriel Dr.
16600	Karen Wu	646-555-0113	230-555-0192	9265 La Paz
4324	Karl Xie	508-555-0163	230-555-0193	4912 La Vuelta
16891	Jonn Jackson	674-555-0110	230-555-0194	40 Ellis St.
3273	Miguel Miller	397-555-0155	230-555-0195	6696 Anchor Drive
3588	Osa Agbonile	592-555-0152	230-555-0196	1873 Lion Circle
10272	Julia Lee	870-555-0110	230-555-0197	3148 Rose Street
4868	Jose Hayes	599-555-0171	230-555-0198	793 Crawford Street
4761	Caleb Alexander	670-555-0141	230-555-0199	4775 Kentucky Dr.
16443	Terry Chander	998-555-0171	230-555-0200	771 Northridge Drive

Este procedimiento usa la tabla HBase **Contacts** que creó en el último procedimiento.

- Desde la conexión SSH abierta, ejecute el siguiente comando para transformar el archivo de datos en StoreFiles y almacenarlo en una ruta de acceso relativa especificada por **Dimporttsv.bulk.output**.

```
hbase org.apache.hadoop.hbase.mapreduce.ImportTsv -
Dimporttsv.columns="HBASE_ROW_KEY,Personal:Name,Personal:Phone,Office:Phone,Office:Address" -Dimporttsv.bulk.output="/example/data/storeDataFileOutput"
Contacts
wasb://hbasecontacts@hdtutorialdata.blob.core.windows.net/contacts.txt
```

- Ejecute el siguiente comando para cargar los datos desde **/example/data/storeDataFileOutput** en la tabla de HBase:

```
hbase org.apache.hadoop.hbase.mapreduce.LoadIncrementalHFiles
/example/data/storeDataFileOutput Contacts
```

- Puede abrir el shell de HBase y usar el comando **scan** para mostrar el contenido de la tabla.

Paso 3: Usar Apache Hive para consultar Apache HBase

Puede consultar datos en las tablas de HBase mediante Apache Hive. En esta sección, creará una tabla de Hive que se asigna a la tabla de HBase y la usará para consultar los datos de la tabla de HBase.



1. En la conexión SSH abierta, use el siguiente comando para iniciar Beeline:

```
beeline -u 'jdbc:hive2://localhost:10001/;transportMode=http' -n admin
```

2. Ejecute el siguiente script de HiveQL para crear una tabla de Hive que se asigne a la tabla de HBase. Asegúrese de haber creado la tabla de ejemplo a la que se hace referencia anteriormente en este artículo mediante el shell de HBase antes de ejecutar esta instrucción.

```
CREATE EXTERNAL TABLE hbasecontacts(rowkey STRING, name STRING,  
homephone STRING, officephone STRING, officeaddress STRING)
```

```
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
```

```
WITH SERDEPROPERTIES ('hbase.columns.mapping' =  
'key,Personal:Name,Personal:Phone,Office:Phone,Office:Address')
```

```
TBLPROPERTIES ('hbase.table.name' = 'Contacts');
```

3. Ejecute el siguiente script de HiveQL para consultar los datos de la tabla de HBase:

```
SELECT count(rowkey) AS rk_count FROM hbasecontacts;
```

4. Para salir de Beeline, use **!exit**
5. Para cerrar la conexión SSH, use **exit**

