How to Intergrate HBase & Hive

|  |  |
| --- | --- |
| **Objective** | Understand how HBase and Hive integrate |
| **Location of Files** | ~/materials/data |
| **Expected Outcome** | You will complete data storage in HBase from Hive table data |
| **Before You Begin** | Go into your lab instance terminal |

**HBase and Hive Integration - Part 1:**

**Build a Hive table for HBase**

1. Start hive shell:

$ hive

2. Create a HBase table

hive>

CREATE TABLE hbase\_test\_1(key int, value string)

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

WITH SERDEPROPERTIES ("hbase.columns.mapping" = ":key,cf1:val")

TBLPROPERTIES ("hbase.table.name" = "test\_1");

Results displayed:

OK

Time taken: 11.565 seconds

3. Validate the table in Hive

hive>

describe hbase\_test\_1;

  Check on the output:

OK

key int from deserializer

value string from deserializer

Time taken: 1.795 seconds, Fetched: 2 row(s)

4. Exit Hive:

hive>

exit;

**Validate the table in HBase**

1. Enter HBase command line interface

$ hbase shell

2. Use LIST command to check:

hbase(main):005:0> list

You should see the test\_1 table on the list:

TABLE

...

test\_1

...

3. Validate the table in HBase

hbase(main):007:0> describe "test\_1"

You should see the below results:

DESCRIPTION ENABLED

'test\_1', {NAME => 'cf1', DATA\_BLOCK\_ENCODING => 'NONE', BLOOMFILTER => 'ROW', REPLICATI true ON\_SCOPE => '0', VERSIONS => '1', COMPRESSION => 'NONE', MIN\_VERSIONS => '0', TTL => '2147483647', KEEP\_DELETED\_CELLS => 'false', BLOCKSIZE => '65536', IN\_MEMORY => 'false', BLOCKCACHE => 'true'}

1 row(s) in 0.0970 seconds

**Question: How is the command used different from before?**

<double quotes used here>

4. Use SCAN to find if data exists:

hbase(main):008:0> scan 'test\_1'

  Observe the results displayed:

ROW COLUMN+CELL

0 row(s) in 0.0060 seconds

  As it shows, there is no data in the table

<pstyle="padding-left: 30px;">2.5 Exit HBase shell

hbase(main):024:0> quit

**Populate the table in Hive with Tez**

1. Start Hive shell

$ hive

2. Run below command to set Hive engine as Tez:

hive>

set hive.execution.engine=tez;

3. Execute below HiveQL to populate the table

hive>

INSERT OVERWRITE TABLE hbase\_test\_1 SELECT volume, symbol FROM stock\_price\_fn WHERE volume>10000000;

Results displayed here:

Query ID = root\_20140830123030\_3fee9010- e712-4c44-89ec-1261c220e424

Total jobs = 1

Launching Job 1 out of 1

Status: Running (application id: application\_1409394057604\_0003)

Map 1: -/-

Map 1: 0/1

Map 1: 0/1

Map 1: 0/1

Map 1: 0/1

Map 1: 0/1

Map 1: 0/1

Map 1: 1/1

Status: Finished successfully

OK

Time taken: 24.005 seconds

4. Now check the data in Hive:

hive>

select \* from hbase\_test\_1;

  The results shown here:

OK

16298100 FNFG

16982800 FNFG

23728300 FNFG

26681300 FNFG

Time taken: 0.342 seconds, Fetched: 4 row(s)

Question: does the above job require a MR or Tez execution?

<no>

5. Exit Hive session:

hive>

exit;

**Verify the table data in HBase**

1. Enter HBase session

$ hbase shell

2. Use SCAN to view the data

hbase(main):009:0> scan 'test\_1'

Below are the results shown:

ROW COLUMN+CELL

16298100 column=cf1:val, timestamp=1409427074706, value=FNFG

16982800 column=cf1:val, timestamp=1409427074706, value=FNFG

23728300 column=cf1:val, timestamp=1409427074706, value=FNFG

26681300 column=cf1:val, timestamp=1409427074706, value=FNFG

4 row(s) in 0.1110 seconds

3. Exit hbase:

hbase(main):002:0> exit

**END**