Lab_4

Name: Karim Gamal Mahmoud Mohamed

ID : 21KGMM

- **Create a Customers table in Hbase with two column families namely CustomerInfo and ContactInfo with row key as customer_id**

First, we need to Create a Customers table in Hbase with two column families namely CustomerInfo and ContactInfo

```
Took 2.5767 seconds
hbase(main):003:0> create 'Customers', 'CustomerInfo', 'ContactInfo'
Created table Customers
Took 2.6067 seconds
=> Hbase::Table - Customers
hbase(main):004:0>
```

- Describe Customers Table after create it.

```
=> Hbase::Table - Customers
hbase(main):004:0> describe 'Customers'
Table Customers is ENABLED
Customers
COLUMN FAMILIES DESCRIPTION
{NAME => 'ContactInfo', VERSIONS => '1', EVICT_BLOCKS_ON_CLOSE => 'false', NEW_V
ERSION_BEHAVIOR => 'false', KEEP_DELETED_CELLS => 'FALSE', CACHE_DATA_ON_WRITE =
> 'false', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0',
 REPLICATION_SCOPE => '0', BLOOMFILTER => 'ROW', CACHE_INDEX_ON_WRITE => 'false'
, IN_MEMORY => 'false', CACHE_BLOOMS_ON_WRITE => 'false', PREFETCH_BLOCKS_ON_OPE
N => 'false', COMPRESSION => 'NONE', BLOCKCACHE => 'true', BLOCKSIZE => '65536'}
{NAME => 'CustomerInfo', VERSIONS => '1', EVICT_BLOCKS_ON_CLOSE => 'false', NEW_
VERSION_BEHAVIOR => 'false', KEEP_DELETED_CELLS => 'FALSE', CACHE_DATA_ON_WRITE
=> 'false', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', MIN_VERSIONS => '0'
, REPLICATION_SCOPE => '0', BLOOMFILTER => 'ROW', CACHE_INDEX_ON_WRITE => 'false
', IN_MEMORY => 'false', CACHE_BLOOMS_ON_WRITE => 'false', PREFETCH_BLOCKS_ON_OP
EN => 'false', COMPRESSION => 'NONE', BLOCKCACHE => 'true', BLOCKSIZE => '65536'
}
2 row(s)
Took 1.1352 seconds
hbase(main):005:0>
```

- **From the given MYSQL database named 'retail_db' within VM, import the data of Customers table into the Hbase Customers table.**

- for contactInfo column family

- ContactInfo (column family) with these columns customer_street, customer_city, customer_state, customer_zip

```
[osboxes@quickstart-bigdata ~]$ sqoop import \
> --connect jdbc:mysql://192.168.159.128/retail_db \
> --username root --password bigdata --table customers \
> --columns "customer_id,customer_street, customer_city, customer_state, customer_zipcode" \
> --hbase-table Customers \
> --column-family ContactInfo \
> --hbase-row-key customer_id
```

- Imported successfully:

```
22/06/28 04:57:32 INFO mapreduce.ImportJobBase: Transferred 0 bytes in 205.7784 seconds (0 bytes/sec)
22/06/28 04:57:32 INFO mapreduce.ImportJobBase: Retrieved 12435 records.
[osboxes@quickstart-bigdata ~]$
```

- Showing the customers table data by using 'scan'

```
12435 row(s)
Took 8.3491 seconds
hbase(main):002:0> scan 'Customers'
```

```
hbase(main):005:0> scan 'Customers', {LIMIT => 5}
ROW                    COLUMN+CELL
 1                     column=ContactInfo:customer_city, timestamp=1656372401722,
                        value=Brownsville
 1                     column=ContactInfo:customer_state, timestamp=1656372401722
                       , value=TX
 1                     column=ContactInfo:customer_street, timestamp=165637240172
                       2, value=6303 Heather Plaza
 1                     column=ContactInfo:customer_zipcode, timestamp=16563724017
                       22, value=78521
 10                    column=ContactInfo:customer_city, timestamp=1656372401722,
                        value=Stafford
 10                    column=ContactInfo:customer_state, timestamp=1656372401722
                       , value=VA
 10                    column=ContactInfo:customer_street, timestamp=165637240172
                       2, value=8598 Harvest Beacon Plaza
 10                    column=ContactInfo:customer_zipcode, timestamp=16563724017
                       22, value=22554
 100                   column=ContactInfo:customer_city, timestamp=1656372401722,
                        value=Caguas
 100                   column=ContactInfo:customer_state, timestamp=1656372401722
                       , value=PR
 100                   column=ContactInfo:customer_street, timestamp=165637240172
                       2, value=4110 Silent Pointe
 100                   column=ContactInfo:customer_zipcode, timestamp=16563724017
                       22, value=00725
 1000                  column=ContactInfo:customer_city, timestamp=1656372404524,
                        value=Freeport
 1000                  column=ContactInfo:customer_state, timestamp=1656372404524
                       , value=NY
 1000                  column=ContactInfo:customer_street, timestamp=165637240452
                       4, value=3627 Lost Butterfly Drive
 1000                  column=ContactInfo:customer_zipcode, timestamp=16563724045
                       24, value=11520
 10000                 column=ContactInfo:customer_city, timestamp=1656372446436,
                        value=Caguas
 10000                 column=ContactInfo:customer_state, timestamp=1656372446436
                       , value=PR
 10000                 column=ContactInfo:customer_street, timestamp=165637244643
                       6, value=9938 Sunny Loop
 10000                 column=ContactInfo:customer_zipcode, timestamp=16563724464
                       36, value=00725
5 row(s)
Took 0.5873 seconds
hbase(main):006:0>
```

- **From the given MYSQL database named 'retail_db' within VM, import the data of Customers table into the Hbase Customers table.**

- For CustomertInfo column family

- CustomerInfo (column family) with these columns customer_fname, customer_lname, customer_email, customer_password

```
[osboxes@quickstart-bigdata ~]$ sqoop import --connect jdbc:mysql://192.168.159.128/retail_db --userna
me root --password bigdata --table customers --columns "customer_id, customer_fname, customer_lname,cu
stomer_email, customer_password" --hbase-table Customers --column-family CustomerInfo --hbase-row-key
customer_id
```

- Imported successfully:

```
22/06/28 05:07:46 INFO mapreduce.ImportJobBase: Transferred 0 bytes in 146.6714 seconds (0 bytes/sec)
22/06/28 05:07:46 INFO mapreduce.ImportJobBase: Retrieved 12435 records.
[osboxes@quickstart-bigdata ~]$
```

- Showing the customers table data by using ‘scan’

```
hbase(main):006:0> scan 'Customers', {LIMIT => 5}
ROW                         COLUMN+CELL
 1                          column=ContactInfo:customer_city, timestamp=1656372401722, value=Brownsville
 1                          column=ContactInfo:customer_state, timestamp=1656372401722, value=TX
 1                          column=ContactInfo:customer_street, timestamp=1656372401722, value=6303 Heat
                            her Plaza
 1                          column=ContactInfo:customer_zipcode, timestamp=1656372401722, value=78521
 1                          column=CustomerInfo:customer_email, timestamp=1656373031303, value=XXXXXXXXX
 1                          column=CustomerInfo:customer_fname, timestamp=1656373031303, value=Richard
 1                          column=CustomerInfo:customer_lname, timestamp=1656373031303, value=Hernandez
 1                          column=CustomerInfo:customer_password, timestamp=1656373031303, value=XXXXXX
                            XXX
 10                         column=ContactInfo:customer_city, timestamp=1656372401722, value=Stafford
 10                         column=ContactInfo:customer_state, timestamp=1656372401722, value=VA
 10                         column=ContactInfo:customer_street, timestamp=1656372401722, value=8598 Harv
                            est Beacon Plaza
 10                         column=ContactInfo:customer_zipcode, timestamp=1656372401722, value=22554
 10                         column=CustomerInfo:customer_email, timestamp=1656373031303, value=XXXXXXXXX
 10                         column=CustomerInfo:customer_fname, timestamp=1656373031303, value=Melissa
 10                         column=CustomerInfo:customer_lname, timestamp=1656373031303, value=Smith
 10                         column=CustomerInfo:customer_password, timestamp=1656373031303, value=XXXXXX
                            XXX
 100                        column=ContactInfo:customer_city, timestamp=1656372401722, value=Caguas
 100                        column=ContactInfo:customer_state, timestamp=1656372401722, value=PR
 100                        column=ContactInfo:customer_street, timestamp=1656372401722, value=4110 Sile
                            nt Pointe
```

- **List top 10 records of NY and VA state from the Hbase table.**

- We need to import the following:

import org.apache.hadoop.hbase.filter.CompareFilter

import org.apache.hadoop.hbase.filter.SingleColumnValueFilter

import org.apache.hadoop.hbase.filter.SubstringComparator

import org.apache.hadoop.hbase.util.Bytes

```
hbase(main):007:0> import org.apache.hadoop.hbase.filter.CompareFilter
=> [Java::OrgApacheHadoopHbaseFilter::CompareFilter]
hbase(main):008:0> import org.apache.hadoop.hbase.filter.SingleColumnValueFilter
=> [Java::OrgApacheHadoopHbaseFilter::SingleColumnValueFilter]
hbase(main):009:0> import org.apache.hadoop.hbase.filter.SubstringComparator
=> [Java::OrgApacheHadoopHbaseFilter::SubstringComparator]
hbase(main):010:0> import org.apache.hadoop.hbase.util.Bytes
=> [Java::OrgApacheHadoopHbaseUtil::Bytes]
hbase(main):011:0>
```

```
hbase(main):012:0> scan 'Customers', { COLUMNS => 'ContactInfo:customer_state',LIMIT =>10, FILTER =>Sing
leColumnValueFilter.new(Bytes.toBytes('ContactInfo'),Bytes.toBytes('customer_state'),CompareFilter::Comp
areOp.valueOf('EQUAL'),SubstringComparator.new('NY'))}
ROW                     COLUMN+CELL
 1000                   column=ContactInfo:customer_state, timestamp=1656372404524, value=NY
 10002                  column=ContactInfo:customer_state, timestamp=1656372446436, value=NY
 10007                  column=ContactInfo:customer_state, timestamp=1656372446436, value=NY
 10008                  column=ContactInfo:customer_state, timestamp=1656372446436, value=NY
 10016                  column=ContactInfo:customer_state, timestamp=1656372446436, value=NY
 10019                  column=ContactInfo:customer_state, timestamp=1656372446436, value=NY
 10031                  column=ContactInfo:customer_state, timestamp=1656372446436, value=NY
 1005                   column=ContactInfo:customer_state, timestamp=1656372404524, value=NY
 10050                  column=ContactInfo:customer_state, timestamp=1656372446436, value=NY
 10095                  column=ContactInfo:customer_state, timestamp=1656372446436, value=NY
10 row(s)
Took 0.1705 seconds
hbase(main):013:0>
```

```
hbase(main):013:0> scan 'Customers', { COLUMNS => 'ContactInfo:customer_state',LIMIT =>10, FILTER =>Sing
leColumnValueFilter.new(Bytes.toBytes('ContactInfo'),Bytes.toBytes('customer_state'),CompareFilter::Comp
areOp.valueOf('EQUAL'),SubstringComparator.new('VA'))}
ROW                     COLUMN+CELL
 10                     column=ContactInfo:customer_state, timestamp=1656372401722, value=VA
 10130                  column=ContactInfo:customer_state, timestamp=1656372446436, value=VA
 10238                  column=ContactInfo:customer_state, timestamp=1656372446740, value=VA
 10259                  column=ContactInfo:customer_state, timestamp=1656372446740, value=VA
 10510                  column=ContactInfo:customer_state, timestamp=1656372446740, value=VA
 10540                  column=ContactInfo:customer_state, timestamp=1656372446740, value=VA
 10787                  column=ContactInfo:customer_state, timestamp=1656372446932, value=VA
 10870                  column=ContactInfo:customer_state, timestamp=1656372446932, value=VA
 10932                  column=ContactInfo:customer_state, timestamp=1656372446932, value=VA
 10939                  column=ContactInfo:customer_state, timestamp=1656372446932, value=VA
10 row(s)
Took 0.2516 seconds
hbase(main):014:0>
```

Conclusion:

In this lab, I used My SQL to manage a database (a retail database), then I presented tables, ran Hbase to create the tables "Customers," "CustomerInfo," and "ContactInfo," and last I imported data into HDFS from external datastores.

additionally, I demonstrated the records in the table Customer with limitations 5 in Hbase to demonstrate the entire process.

Going well, I finally listed the top 10 records for the states of NY and VA from the Hbase table.

once the three filters (CompareFilter, BinaryComparator, and SingleColumnValueFilter) have been installed.

Sources :

https://www.javatpoint.com/mysql-commands-cheat-sheet

https://www.projectpro.io/recipes/connect-database-and-import-data-into-hbase-sqoop

https://stackoverflow.com/questions/7256100/scan-with-filter-using-hbase-shell