Creating Lookup Table

Command to create the Lookup Table

Here we are going to create HBase table for Card Lookup with Hive support (Hive-Hbase integration).

1. Create an empty HBase table with the following command:
   create 'card_lookup', 'lkp_info'

2. Create external Hive table with HBase storage handler and cluster it into 8 buckets with card_id for joins optimization.
   CREATE EXTERNAL TABLE IF NOT EXISTS card_lookup(
   card_id BIGINT,
   ucl DOUBLE,
   postcode INT,
   transaction_dt TIMESTAMP,
   credit_score INT
)
   CLUSTERED BY (card_id)INTO 8 BUCKETS
   STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
   WITH SERDEPROPERTIES ('hbase.columns.mapping'=':key, lkp_info:ucl, lkp_info:postcode, lkp_info:transaction_dt, lkp_info:credit_score')
   TBLPROPERTIES ('hbase.table.name' = 'card_lookup');

All the above steps are wrapped into a single shell script (creates empty HBase table and calls Hive script – **card_lookup_ddl.hql** to execute the above create and insert statements) with name **create_lookup_nosql.sh** placed in the path:
/home/hadoop/cred_financials_data/script/create_lookup_nosql.sh

```
[hadoop@ip-172-31-80-45 script]$ cat create_lookup_nosql.sh
#!/bin/bash
# Create Hive and HBase tables of Card Lookup

echo "create 'card_lookup', 'lkp_info'" | hbase shell -n

hive -f /home/hadoop/cred_financials_data/script/card_lookup_ddl.hql
[hadoop@ip-172-31-80-45 script]$
```

```
[hadoop@ip-172-31-80-45 script]$ cat card_lookup_ddl.hql
USE cred_financials_data;

-- Enforce Hive Bucketing
SET hive.enforce.bucketing=true;

-- Create Card Lookup Hive table (External) with HBase Integration
CREATE EXTERNAL TABLE IF NOT EXISTS card_lookup(
card_id BIGINT,
ucl DOUBLE,
postcode INT,
transaction_dt TIMESTAMP,
credit_score INT
)
CLUSTERED BY (card_id)INTO 8 BUCKETS
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
WITH SERDEPROPERTIES ('hbase.columns.mapping'=':key, lkp_info:ucl, lkp_info:postcode, lkp_info:transaction_dt, lkp_info:credit_score')
TBLPROPERTIES ('hbase.table.name' = 'card_lookup');

exit;
[hadoop@ip-172-31-80-45 script]$
```

Execute the shell script with the below command:

/home/hadoop/cred_financials_data/script/create_lookup_nosql.sh

Command to see the table created

Hive: -

use cred_financials_data;

describe formatted card_lookup;

```
hive> describe formatted card_lookup;
OK
col_name        data_type       comment
# col_name              data_type               comment

card_id                 bigint
ucl                     double
postcode                int
transaction_dt          timestamp
credit_score            int

# Detailed Table Information
Database:               cred_financials_data
Owner:                  hadoop
CreateTime:             Sat Dec 17 19:30:06 UTC 2022
LastAccessTime:         UNKNOWN
Retention:              0
Location:               hdfs://ip-172-31-80-45.ec2.internal:8020/user/hive/warehouse/cred_financials_data.db/card_lookup
Table Type:             EXTERNAL_TABLE
Table Parameters:
        COLUMN_STATS_ACCURATE   {\"BASIC_STATS\":\"true\"}
        EXTERNAL                TRUE
        hbase.table.name        card_lookup
        numFiles                0
        numRows                 0
        rawDataSize             0
        storage_handler         org.apache.hadoop.hive.hbase.HBaseStorageHandler
        totalSize               0
        transient_lastDdlTime   1671305406

# Storage Information
SerDe Library:          org.apache.hadoop.hive.hbase.HBaseSerDe
InputFormat:            null
OutputFormat:           null
Compressed:             No
Num Buckets:            8
Bucket Columns:         [card_id]
Sort Columns:           []
Storage Desc Params:
        hbase.columns.mapping   :key, lkp_info:ucl, lkp_info:postcode, lkp_info:transaction_dt, lkp_info:credit_score
        serialization.format    1
Time taken: 0.03 seconds, Fetched: 38 row(s)
```

HBase: -

list ‘card.*’

```
hbase(main):007:0> list 'card.*'
TABLE
card_lookup
card_transactions
2 row(s) in 0.0080 seconds
```

describe ‘card_lookup’

```
hbase(main):008:0> describe 'card_lookup'
Table card_lookup is ENABLED
card_lookup
COLUMN FAMILIES DESCRIPTION
{NAME => 'lkp_info', BLOOMFILTER => 'ROW', VERSIONS => '1', IN_MEMORY => 'false', KEEP_DELETED_CELLS => 'FALSE', DATA_BLOCK_ENCODING => 'NONE', TTL => 'FOREVER', COM
PRESSION => 'NONE', MIN_VERSIONS => '0', BLOCKCACHE => 'true', BLOCKSIZE => '65536', REPLICATION_SCOPE => '0'}
1 row(s) in 0.0210 seconds
```