HIVE – HBASE CONNECTIVITY

Step -1 : Create a Hive table ,which acts as stage table to hold data for operation

hive > CREATE TABLE sales_order_csv
```
   (
   order_number int,
   qty_ordered int,
   price_each int,
   order_line_number int,
   sales int,
   status STRING,
   qtr_id int,
   month_id int,
   year_id int,
   product_line STRING,
   msrp int,
   product_code STRING,
   phone string,
   city STRING,
   state string,
   postal_code STRING,
   country STRING,
   territory STRING,
   contact_lname STRING,
   contact_fname STRING,
   dealsize STRING
   )
   ROW FORMAT DELIMITED
   FIELDS TERMINATED BY ','
   TBLPROPERTIES("skip.header.line.count"="1");
```

```
    >        )
    >        ROW FORMAT DELIMITED
    >        FIELDS TERMINATED BY ','
    >        TBLPROPERTIES("skip.header.line.count"="1");
OK
Time taken: 2.482 seconds
hive> show tables
    > ;
OK
sales_order_csv
Time taken: 0.219 seconds, Fetched: 1 row(s)
```

Step -2 : Load data from csv file to stage table created in Step-1

$hive > load data local inpath '/tmp/sales_order_data.csv' into table sales_order_csv;

```
hive> load data local inpath '/tmp/sales_order_data.csv' into table sales_order_
csv;
Loading data to table default.sales_order_csv
Table default.sales_order_csv stats: [numFiles=1, totalSize=360233]
OK
Time taken: 0.988 seconds
```

Step -3 : Create Hbase table with column families
(Order number will act as row_key, hence it is not considered in column family while creating table)

hbase(main):013:0> create 'sales_order_data_hbase','qty_ordered','price_each','order_line_number','sales','status','qtr_id','month_id','year_id','product_line','msrp','product_code','phone','city','state','postal_code','country','territory','contact_lname','contact_fname','dealsize'

Step -4 : Create Hive table with Hbase SerDe properties

CREATE EXTERNAL TABLE IF NOT EXISTS sales_order_hive_hbase_connect1
(
key int,
qty_ordered int,
price_each int,
order_line_number int,
sales int,
status STRING,
qtr_id int,
month_id int,
year_id int,
product_line STRING,
msrp int,
product_code STRING,
phone string,
city STRING,
state string,
postal_code STRING,
country STRING,
territory STRING,

```
contact_lname STRING,
contact_fname STRING,
dealsize STRING
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
WITH SERDEPROPERTIES
("hbase.columns.mapping"=" :key,
qty_ordered:qty_ordered,
price_each:price_each,
order_line_number:order_line_number,
sales:sales,
status:status,
qtr_id:qtr_id,
month_id:month_id,
year_id:year_id,
product_line:product_line,
msrp:msrp,
product_code:product_code,
phone:phone,
city:city,
state:state,
postal_code:postal_code,
country:country,
territory:territory,
contact_lname:contact_lname,
contact_fname:contact_fname,
dealsize:dealsize")
TBLPROPERTIES("hbase.table.name"="sales_order_data_hbase");
```

Step -5 : Insert data from stage table to Hive table with SerDe properties

hive > Insert into table sales_order_hive_hbase_connect select * from sales_order_csv;

```
hive> Insert into table sales_order_hive_hbase_connect1 select * from sales_order_csv
    > ;
Query ID = cloudera_20221128082727_6df2f5eb-2b73-42d5-a6b1-ff278f5cf397
Total jobs = 1
Launching Job 1 out of 1
Number of reduce tasks is set to 0 since there's no reduce operator
Starting Job = job_1669648232524_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1669648232524_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job  -kill job_1669648232524_0001
Hadoop job information for Stage-0: number of mappers: 1; number of reducers: 0
2022-11-28 08:28:46,545 Stage-0 map = 0%,  reduce = 0%
2022-11-28 08:29:23,150 Stage-0 map = 100%,  reduce = 0%, Cumulative CPU 6.27 sec
MapReduce Total cumulative CPU time: 6 seconds 270 msec
Ended Job = job_1669648232524_0001
MapReduce Jobs Launched:
Stage-Stage-0: Map: 1   Cumulative CPU: 6.27 sec   HDFS Read: 375245 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 6 seconds 270 msec
OK
Time taken: 89.353 seconds
hive>
```

Step -6 : Inspecting data

1). Hive table

Hive > Select * from sales_order_hive_hbase_connect limit 10;

```
hive> select * from sales_order_hive_hbase_connect1 limit 10;
OK
10100   49      34      1       1689    Shipped 1       1       2003    Vintage Cars    41      S24_3969     6035558647       Nashua  NH      62005   USA N
A       Young   Valarie Small
10101   46      53      2       2472    Shipped 1       1       2003    Vintage Cars    44      S24_2022     +49 69 66 90 2555         Frankfurt        6
0528    Germany EMEA    Keitel  Roland  Small
10102   41      50      1       2055    Shipped 1       1       2003    Vintage Cars    53      S18_1367     2125551500       NYC     NY      10022   USA N
A       Frick   Michael Small
10103   42      100     6       4460    Shipped 1       1       2003    Classic Cars    101     S700_2824    07-98 9555       Stavern          4110    Norwa
y       EMEA    Bergulfsen      Jonas   Medium
10104   32      53      2       1705    Shipped 1       1       2003    Trains  58      S50_1514        (91) 555 94 44 Madrid            28034   Spain   EMEAF
reyre   Diego   Small
10105   25      56      8       1419    Shipped 1       2       2003    Ships   54      S72_3212        31 12 3555       Kobenhavn                1734    Denma
rk      EMEA    Petersen        Jytte   Small
10106   48      52      15      2526    Shipped 1       2       2003    Planes  49      S72_1253        035-640555       Bergamo          24100   Italy   EMEAR
ovelli  Giovanni        Small
10107   20      92      8       1858    Shipped 1       2       2003    Motorcycles     99      S32_1374     2125557818       NYC     NY      10022   USA N
A       Yu      Kwai    Small
10108   34      82      14      2821    Shipped 1       3       2003    Motorcycles     81      S50_4713     +63 2 555 3587 Makati City              1227
MM      Philippines     Japan   Cruz    Arnold  Small
10109   29      32      6       930     Shipped 1       3       2003    Classic Cars    37      S24_2972     2155559857       Philadelphia     PA      71270
USA     NA      Hernandez       Rosa    Small
Time taken: 0.189 seconds, Fetched: 10 row(s)
hive>
```

2). Hbase table – before loading data

Hbase > scan 'sales_order_data_hbase'

```
=> Hbase::Table - sales_order_data_hbase
hbase(main):019:0> scan 'sales_order_data_hbase'
ROW                             COLUMN+CELL
0 row(s) in 0.2580 seconds

hbase(main):020:0>
```

3). Hbase table – After loading data

Hbase > scan 'sales_order_data_hbase'

```
 10425                        column=phone:phone, timestamp=1669652961409, value=40.67.8555
 10425                        column=postal_code:postal_code, timestamp=1669652961409, value=44000
 10425                        column=price_each:price_each, timestamp=1669652961409, value=100
 10425                        column=product_code:product_code, timestamp=1669652961409, value=S50_1392
 10425                        column=product_line:product_line, timestamp=1669652961409, value=Trucks and Buses
 10425                        column=qtr_id:qtr_id, timestamp=1669652961409, value=2
 10425                        column=qty_ordered:qty_ordered, timestamp=1669652961409, value=18
 10425                        column=sales:sales, timestamp=1669652961409, value=1895
 10425                        column=state:state, timestamp=1669652961409, value=
 10425                        column=status:status, timestamp=1669652961409, value=In Process
 10425                        column=territory:territory, timestamp=1669652961409, value=EMEA
 10425                        column=year_id:year_id, timestamp=1669652961409, value=2005
307 row(s) in 9.1850 seconds

hbase(main):021:0>
```

4). Loading sample data in Hbase table

Hbase > put 'sales_order_data_hbase','11111','qty_ordered: qty_ordered','10'

```
0 row(s) in 0.3350 seconds

hbase(main):044:0> put 'sales_order_data_hbase','11111','qty_ordered:qty_ordered','10'
0 row(s) in 0.0060 seconds

hbase(main):045:0>
```

5). Accessing hbase table data from Hive

Hive > Select * from sales_order_hive_hbase_connect where key = 11111;

```
hive> select * from sales_order_hive_hbase_connect1 where key = 11111;
OK
11111   10      NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULL    NULLN
ULL
Time taken: 0.362 seconds, Fetched: 1 row(s)
hive>
```