

Exporting data from Hive tables to HBase tables

- 1) Create a table in HBase:

```
hbase(main):004:0> create 'HBase_Online_Retail','cf'
0 row(s) in 0.8940 seconds
```

- 2) Create a table in Hive:

```
[cloudera@quickstart ~]$ hive
Logging initialized using configuration in file:/etc/hive/conf.dist/hive-log4j.properties
WARNING: Hive CLI is deprecated and migration to Beeline is recommended.
hive> create table OnlineRetail(
  > invoiceno string,
  > stockcode int,
  > description string,
  > quantity int,
  > invoicedate string,
  > unitprice int,
  > custid int,
  > country string)
  > row format delimited
  > fields terminated by ','
  > stored as textfile;
OK
Time taken: 1.593 seconds
```

- 3) Load data in the Hive table

```
hive> load data local inpath '/home/cloudera/OnlineRetail.csv' into table OnlineRetail;
Loading data to table default.onlineretail
Table default.onlineretail stats: [numFiles=1, totalSize=46133213]
OK
Time taken: 3.019 seconds
```

4) Create external table in Hive for connecting to the HBase table

```
hive> create external table hbaseonlineretail(  
  >   invoiceno string,  
  >   stockcode int,  
  >   description string,  
  >   quantity int,  
  >   invoicedate string,  
  >   unitprice int,  
  >   custid int,  
  >   country string) stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'  
  >   with serdeproperties(  
  >     'hbase.columns.mapping' = ':key,cf:stockcode,cf:description,cf:quantity,cf:invoicedate,cf:unitprice,cf:custid,cf:country')  
  >   tblproperties('hbase.table.name' = 'HBase_Online_Retail');  
OK  
Time taken: 11.632 seconds
```

5) Load data from Hive table to external Hive-HBase table

Insert into hbaseonlineretail select * from OnlineRetail;

```
hive> insert into hbaseonlineretail select * from OnlineRetail  
> ;  
Query ID = cloudera_20240510040808_a8d99cf9-8f71-49e0-8a49-854ab22d5eb4  
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_1715338982966_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1715338982966_0001/  
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1715338982966_0001  
Hadoop job information for Stage-0: number of mappers: 1; number of reducers: 0  
2024-05-10 04:09:11,772 Stage-0 map = 0%, reduce = 0%  
2024-05-10 04:10:12,901 Stage-0 map = 0%, reduce = 0%, Cumulative CPU 24.91 sec  
2024-05-10 04:10:31,430 Stage-0 map = 100%, reduce = 0%, Cumulative CPU 37.77 sec  
MapReduce Total cumulative CPU time: 37 seconds 770 msec  
Ended Job = job_1715338982966_0001  
MapReduce Jobs Launched:  
Stage-Stage-0: Map: 1 Cumulative CPU: 37.77 sec HDFS Read: 46144904 HDFS Write: 0 SUCCESS  
Total MapReduce CPU Time Spent: 37 seconds 770 msec  
OK  
Time taken: 135.995 seconds
```

6) Scan from HBase table

```
hbase(main):001:0> scan 'HBase_Online_Retail'  
ROW  
536365 COLUMN+CELL  
column=cf:country, timestamp=1715339379511, value=United Kingdom  
536365 column=cf:custid, timestamp=1715339379511, value=17850  
536365 column=cf:description, timestamp=1715339379511, value=GLASS STAR FROSTED T-LIGHT HOLDER  
536365 column=cf:invoicedate, timestamp=1715339379511, value=01-12-2010 08:26  
536365 column=cf:quantity, timestamp=1715339379511, value=6  
536365 column=cf:stockcode, timestamp=1715339379511, value=21730  
536365 column=cf:unitprice, timestamp=1715339379511, value=4  
536366 column=cf:country, timestamp=1715339379511, value=United Kingdom  
536366 column=cf:custid, timestamp=1715339379511, value=17850  
536366 column=cf:description, timestamp=1715339379511, value=HAND WARMER RED POLKA DOT  
536366 column=cf:invoicedate, timestamp=1715339379511, value=01-12-2010 08:28  
536366 column=cf:quantity, timestamp=1715339379511, value=6  
536366 column=cf:stockcode, timestamp=1715339379511, value=22632  
536366 column=cf:unitprice, timestamp=1715339379511, value=1  
536367 column=cf:country, timestamp=1715339379511, value=United Kingdom  
536367 column=cf:custid, timestamp=1715339379511, value=13047  
536367 column=cf:description, timestamp=1715339379511, value=DOORMAT NEW ENGLAND  
536367 column=cf:invoicedate, timestamp=1715339379511, value=01-12-2010 08:34  
536367 column=cf:quantity, timestamp=1715339379511, value=4  
536367 column=cf:stockcode, timestamp=1715339379511, value=48187  
536367 column=cf:unitprice, timestamp=1715339379511, value=7
```

7) Queries:

- create index idx on table OnlineRetail(invoiceno) as "compact" with deferred rebuild;
- select sum(quantity*price) as total_sales, avg(quantity*price) as avg_sales from OnlineRetail;
- select invoiceno, sum(quantity*price) as total_cost from OnlineRetail group by invoiceno order by total_cost desc limit 1;
- select customerid, sum(quantity*price) as total_cost from OnlineRetail group by customerid order by total_cost desc limit 1;
- select country, sum(quantity*price) as max_sale from OnlineRetail group by country order by max_sale desc limit 1;
- select country, sum(quantity*price) as min_sale from OnlineRetail group by country order by min_sale asc limit 1;