

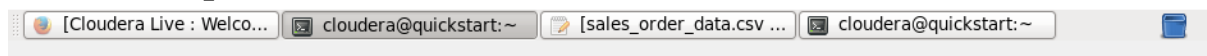
HBASE connecting with Hive - Assignment

1) Create a column family in hbase

```
create 'test_sales_order', {NAME => 'ordernumber',VERSIONS =>3}, {NAME =>
'quantityordered',VERSIONS =>3}, {NAME => 'priceeach',VERSIONS =>3}, {NAME =>
'orderlinenumber',VERSIONS =>3}, {NAME => 'sales',VERSIONS =>3}, {NAME => 'status',VERSIONS
=>3}, {NAME => 'qtr_id',VERSIONS =>3}, {NAME => 'month_id',VERSIONS =>3}, {NAME =>
'year_id',VERSIONS =>3}, {NAME => 'productline',VERSIONS =>3}, {NAME => 'msrp',VERSIONS
=>3}, {NAME => 'productcode',VERSIONS =>3}, {NAME => 'phone',VERSIONS =>3}, {NAME =>
'city',VERSIONS =>3}, {NAME => 'state',VERSIONS =>3}, {NAME => 'postalcode', VERSIONS =>3},
{NAME => 'country',VERSIONS =>3}, {NAME => 'territory',VERSIONS =>3}, {NAME =>
'contactlastname',VERSIONS =>3}, {NAME => 'contactfirstname',VERSIONS =>3}, {NAME =>
'dealsize',VERSIONS =>3}
```

```
hbase(main):016:0> create 'test_sales_order', {NAME => 'ordernumber',VERSIONS =>3}, {NAME => 'quantit
yordered',VERSIONS =>3}, {NAME => 'priceeach',VERSIONS =>3}, {NAME => 'orderlinenumber',VERSIONS =>3}
, {NAME => 'sales',VERSIONS =>3}, {NAME => 'status',VERSIONS =>3}, {NAME => 'qtr_id',VERSIONS =>3}, {
NAME => 'month_id',VERSIONS =>3}, {NAME => 'year_id',VERSIONS =>3}, {NAME => 'productline',VERSIONS =
>3}, {NAME => 'msrp',VERSIONS =>3}, {NAME => 'productcode',VERSIONS =>3}, {NAME => 'phone',VERSIONS =
>3}, {NAME => 'city',VERSIONS =>3}, {NAME => 'state',VERSIONS =>3}, {NAME => 'postalcode', VERSIONS =
>3}, {NAME => 'country',VERSIONS =>3}, {NAME => 'territory',VERSIONS =>3}, {NAME => 'contactlastname'
,VERSIONS =>3}, {NAME => 'contactfirstname',VERSIONS =>3}, {NAME => 'dealsize',VERSIONS =>3}
0 row(s) in 2.5040 seconds
```

```
=> Hbase::Table - test_sales_order
hbase(main):017:0> █
```



2) Create a schema in hive (storage handler)

- create external table if not exists hbase_sales_order
- (key string, ordernumber int, quantityordered int, priceeach float, orderlinenumber int, sales int, status string, qtr_id int, month_id int, year_id int, productline string, msrp int, productcode int, phone int, city string, state string, postalcode int, country string, territory string, contactlastname string, contactfirstname string, dealsize string)
- row format delimited
- fields terminated by ','
- stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
- with serdeproperties
- ("hbase.columns.mapping"=':key,ordernumber:ordernumber,quantityordered:qunatit yordered,priceeach:priceeach,orderlinenumber:orderlinenumber,sales:sales,status:sta tus,qtr_id:qtr_id,month_id:month_id,year_id:year_id,productline:productline,msrp:m srp,productcode:productcode,phone:phone,city:city,state:state,postalcode:postalcode ,country:country,territory:territory,contactlastname:contactlastname,contactfirstnam e:contactfirstname,dealsize:dealsize")
- tblproperties("hbase.table.name"="test_sales_order");

```
hive> create external table if not exists hbase_sales_order
> (key string, ordernumber int, quantityordered int, priceeach float, orderlinenumber int, sales int, status string, qtr_id int, month_id int, year_id int, productline string, msrp int, productcode int, phone int, city string, state string, postalcode int, country string, territory string, contactlastname string, contactfirstname string, dealsize string)
> row format delimited
> fields terminated by ','
> stored by 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
> with serdeproperties
> ("hbase.columns.mapping"=":key,ordernumber:ordernumber,quantityordered:quantityordered,priceeach:priceeach,orderlinenumber:orderlinenumber,sales:sales,status:status,qtr_id:qtr_id,month_id:month_id,year_id:year_id,productline:productline,msrp:msrp,productcode:productcode,phone:phone,city:city,state:state,postalcode:postalcode,country:country,territory:territory,contactlastname:contactlastname,contactfirstname:contactfirstname,dealsize:dealsize")
> tblproperties("hbase.table.name"="test_sales_order")
> ;
OK
Time taken: 3.89 seconds
```

Hive stage table:

- create external table if not exists hbase_stg_sales_order
- (key string, ordernumber int, quantityordered int, priceeach float, orderlinenumber int, sales int, status string, qtr_id int, month_id int, year_id int, productline string, msrp int, productcode int, phone int, city string, state string, postalcode int, country string, territory string, contactlastname string, contactfirstname string, dealsize string)
- row format delimited
- fields terminated by ','
- ;

Load data into stage table:

Load data local inpath '/home/cloudera/Desktop/sales_order_data.csv'
into table hbase_stg_sales_order;

Insert data from stage table to main hive table:

Insert into table hbase_sales_order select * from hbase_stg_sales_order;

```

hive> load data local inpath '/home/cloudera/Desktop/sales_order_data.csv' into
table hbase_stg_sales_order;
Loading data to table hive_class_b1.hbase_stg_sales_order
Table hive_class_b1.hbase_stg_sales_order stats: [numFiles=2, numRows=0, totalSize=720466, rawDataSize=0]
OK
Time taken: 1.523 seconds
hive> insert into table hbase_sales_order select * from hbase_stg_sales_order;
Query ID = cloudera_20221104074444_1938b017-8ebc-441c-aa0b-0859a3606586
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_1667559152311_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667559152311_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1667559152311_0001
Hadoop job information for Stage-0: number of mappers: 1; number of reducers: 0
2022-11-04 07:44:52,958 Stage-0 map = 0%, reduce = 0%
2022-11-04 07:45:15,184 Stage-0 map = 100%, reduce = 0%, Cumulative CPU 13.37 sec
MapReduce Total cumulative CPU time: 13 seconds 370 msec
Ended Job = job_1667559152311_0001
MapReduce Jobs Launched:
Stage-Stage-0: Map: 1 Cumulative CPU: 13.37 sec HDFS Read: 735589 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 13 seconds 370 msec
OK
Time taken: 55.877 seconds

```

Select city from hbase_sales_order;

```

hive> select city from hbase_sales_order;

```

OK

NH

NY

NY

PA

CA

CA

NY

Victoria

NV

Victoria

NY

MA

CT

CA

Cloudera Live : Welco... cloudera@quickstart:~ [sales_order_data.csv ... cloudera@quickstart:~

Check the table details in hbase:

Scan 'test_sales_order'

10424	column=contactlastname:contactlastname, timestamp=1667573113888, value=Diego
10424	column=country:country, timestamp=1667573113888, value=EMEA
10424	column=month_id:month_id, timestamp=1667573113888, value=2005
10424	column=orderlinenumber:orderlinenumber, timestamp=1667573113888, value=3722
10424	column=ordernumber:ordernumber, timestamp=1667573113888, value=46
10424	column=priceeach:priceeach, timestamp=1667573113888, value=1.0
10424	column=productline:productline, timestamp=1667573113888, value=101
10424	column=qtr_id:qtr_id, timestamp=1667573113888, value=5
10424	column=quantityordered:quantityordered, timestamp=1667573113888, value=80
10424	column=state:state, timestamp=1667573113888, value=28034
10424	column=status:status, timestamp=1667573113888, value=2
10424	column=territory:territory, timestamp=1667573113888, value=Freyre
10425	column=city:city, timestamp=1667573113888, value=
10425	column=contactfirstname:contactfirstname, timestamp=1667573113888, value=Small
10425	column=contactlastname:contactlastname, timestamp=1667573113888, value=Jane
10425	column=country:country, timestamp=1667573113888, value=EMEA
10425	column=month_id:month_id, timestamp=1667573113888, value=2005
10425	column=orderlinenumber:orderlinenumber, timestamp=1667573113888, value=1895
10425	column=ordernumber:ordernumber, timestamp=1667573113888, value=18
10425	column=priceeach:priceeach, timestamp=1667573113888, value=2.0
10425	column=productline:productline, timestamp=1667573113888, value=115
10425	column=qtr_id:qtr_id, timestamp=1667573113888, value=5
10425	column=quantityordered:quantityordered, timestamp=1667573113888, value=100
10425	column=state:state, timestamp=1667573113888, value=44000
10425	column=status:status, timestamp=1667573113888, value=2
10425	column=territory:territory, timestamp=1667573113888, value=Labruno
ORDERNUMBER	column=city:city, timestamp=1667573112767, value=STATE
ORDERNUMBER	column=contactfirstname:contactfirstname, timestamp=1667573112767, value=DEALSIZE
ORDERNUMBER	column=contactlastname:contactlastname, timestamp=1667573112767, value=CONTACTFIRSTNAME
ORDERNUMBER	column=country:country, timestamp=1667573112767, value=TERRITORY
ORDERNUMBER	column=productline:productline, timestamp=1667573112767, value=MSRP
ORDERNUMBER	column=state:state, timestamp=1667573112767, value=POSTALCODE
ORDERNUMBER	column=status:status, timestamp=1667573112767, value=QTR_ID
ORDERNUMBER	column=territory:territory, timestamp=1667573112767, value=CONTACTLASTNAME

308 row(s) in 4.1940 seconds