

Sales_Data_Assignment

29 October 2022 17:16

1. Create table in Hbase

```
login as: cloudera
cloudera@192.168.0.102's password:
Last login: Sat Oct 29 04:39:23 2022 from 192.168.0.104
[cloudera@quickstart ~]$ hbase shell
2022-10-29 04:41:53,656 INFO [main] Configuration.deprecation: hadoop.na
b is deprecated. Instead, use io.native.lib.available
HBase Shell; enter 'help<RETURN>' for list of supported commands.
Type "exit<RETURN>" to leave the HBase Shell
Version 1.2.0-cdh5.13.0, rUnknown, Wed Oct 4 11:16:18 PDT 2017

hbase(main):001:0> create 'sales','product_detail','person_detail'
0 row(s) in 30.3540 seconds
```

2. Create Hive table

```
cloudera@quickstart:~
hive> create external table sales_data(order_number int,quantity_ordered int,price_each float,orderline_number int,sales float,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 int,country string,territory string,last_name string,first_name string,deal_size string) row format delimited fields terminated by ',' STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
> with serdeproperties ("hbase.columns.mapping"=":key,product_detail:quantity_ordered,product_detail:price_each,product_detail:orderline_number,product_detail:sales,product_detail:status,product_detail:qtr_id,product_detail:month_id,product_detail:year_id,product_detail:product_line,product_detail:msrp,product_detail:product_code,person_detail:phone,person_detail:city,person_detail:state,person_detail:postal_code,person_detail:country,person_detail:territory,person_detail:last_name,person_detail:first_name,person_detail:deal_size") TBLPROPERTIES("hbase.table.name"="sales");
OK
Time taken: 18.57 seconds
hive>
```

3. Create staging table in Hive

```
cloudera@quickstart:~
hive> create external table sales_data_stag(order_number int,quantity_ordered int,price_each float,orderline_number int,sales float,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 int,country string,territory string,last_name string,first_name string,deal_size string) row format delimited fields terminated by ',' tblproperties("skip.header.line.count"="1");
OK
Time taken: 0.594 seconds
hive>
```

4. Loading data into staging table

```
hive> load data local inpath 'file:///home/cloudera/hive_data/sales_data.csv' into table sales_data_stag;
Loading data to table default.sales_data_stag
Table default.sales_data_stag stats: [numFiles=2, numRows=0, totalSize=720466, rawDataSize=0]
OK
Time taken: 1.001 seconds
hive> set hive.cli.print.header=true;
hive> select * from sales_data_stag limit 5;
OK
sales_data_stag.order_number sales_data_stag.quantity_ordered sales_data_stag.price_each sales_data_stag.orderline_number sales_data_stag.sales sales_data_stag.stat
s sales_data_stag.qtr_id sales_data_stag.month_id sales_data_stag.year_id sales_data_stag.product_line sales_data_stag.msrp sales_data_stag.product_code sales_data_s
ag.phone sales_data_stag.city sales_data_stag.state sales_data_stag.postal_code sales_data_stag.country sales_data_stag.territory sales_data_stag.last_name sales
_data_stag.first_name sales_data_stag.deal_size
I0107 30 95.7 2 2871.0 Shipped 1 2 2003 Motorcycles 95 $10_1678 2125557818 NYC NY 10022 USA NA Yu Kwai Smal
I0121 34 81.35 5 2765.9 Shipped 2 5 2003 Motorcycles 95 $10_1678 26.47.1555 Reims 51100 France EMEA Henriot Paul Smal
I0134 41 94.74 2 3884.34 Shipped 3 7 2003 Motorcycles 95 $10_1678 +33 1 46 62 7555 Paris 75508 France EMEA Da Cunha
aniel Medium
I0145 45 83.26 6 3746.7 Shipped 3 8 2003 Motorcycles 95 $10_1678 6265557265 Pasadena CA 90003 USA NA Young Juli
Medium
I0159 49 100.0 14 5205.27 Shipped 4 10 2003 Motorcycles 95 $10_1678 6505551386 San Francisco CA NULL USA NA Brown Juli
Medium
Time taken: 0.677 seconds, Fetched: 5 row(s)
hive>
```

6. Loading data into hive table

```
hive> insert into sales_data select * from sales_data_stag;
Query ID = cloudera_20221029054141_e73a7b52-e1aa-43c0-8b54-70bf76982313
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_1667042792462_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667042792462_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1667042792462_0001
Hadoop job information for Stage-0: number of mappers: 1; number of reducers: 0
2022-10-29 05:43:14,455 Stage-0 map = 0%, reduce = 0%
2022-10-29 05:43:57,500 Stage-0 map = 50%, reduce = 0%, Cumulative CPU 14.26 sec
2022-10-29 05:44:02,657 Stage-0 map = 100%, reduce = 0%, Cumulative CPU 18.74 sec
MapReduce Total cumulative CPU time: 18 seconds 740 msec
Ended Job = job_1667042792462_0001
MapReduce Jobs Launched:
Stage-Stage-0: Map: 1 Cumulative CPU: 18.74 sec HDFS Read: 735315 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 18 seconds 740 msec
OK
sales_data_stag.order_number sales_data_stag.quantity_ordered sales_data_stag.price_each sales_data_stag.orderline_number sales_data_stag.sales sales_data_stag.statu
s sales_data_stag.qtr_id sales_data_stag.month_id sales_data_stag.year_id sales_data_stag.product_line sales_data_stag.msrp sales_data_stag.product_code sales_data_st
ag.phone sales_data_stag.city sales_data_stag.state sales_data_stag.postal_code sales_data_stag.country sales_data_stag.territory sales_data_stag.last_name sales
_data_stag.first_name sales_data_stag.deal_size
Time taken: 158.049 seconds
```

7. Cheking data into Hive table

```

cloudera@quickstart~
hive> select * from sales_data limit 5;
OK
sales_data.order_number sales_data.quantity_ordered sales_data.price each sales_data.orderline_number sales_data.sales sales_data.status sales_data.qtr_id sales
_data.month_id sales_data.year_id sales_data.product_line sales_data.msrp sales_data.product_code sales_data.phone sales_data.city sales_data.state sales_data.postal_cod
e sales_data.country sales_data.territory sales_data.last_name sales_data.first_name sales_data.deal_size sales_data.city sales_data.state sales_data.postal_cod
10100 49 34.47 1 1689.03 Shipped 1 1 2003 Vintage Cars 41 S24_3969 6035558647 Nashua NH 62005 USA NA Young Valarie Small
10101 46 53.76 2 2472.96 Shipped 1 1 2003 Vintage Cars 44 S24_2022 +49 69 66 90 2555 Frankfurt 60528 Germany EMEA Keite
1 Roland Small
10102 41 50.14 1 2055.74 Shipped 1 1 2003 Vintage Cars 53 S18_1367 2125551500 NYC NY 10022 USA NA Frick Michael Small
10103 42 100.0 6 4460.82 Shipped 1 1 2003 Classic Cars 101 S700_2824 07-98 9555 Stavern 4110 Norway EMEA Bergulfsen Jonas
Medium
10104 32 53.31 2 1705.92 Shipped 1 1 2003 Trains 58 S50_1514 (91) 555 94 44 Madrid 28034 Spain EMEA Freyre Diego Small
Time taken: 0.562 seconds, Fetched: 5 row(s)
hive>

```

8. Checking data into Hbase table (limit to 5 row only)

```

hbase(main):005:0> scan 'sales' , {COLUMNS=>['product_detail:product_line'], STARTROW=>'10100' , ENDROW=>'10104'}
ROW
COLUMN+CELL
10100 column=product_detail:product_line, timestamp=1667047441621, value=Vintage Cars
10101 column=product_detail:product_line, timestamp=1667047437493, value=Vintage Cars
10102 column=product_detail:product_line, timestamp=1667047434221, value=Vintage Cars
10103 column=product_detail:product_line, timestamp=1667047441621, value=Classic Cars
4 row(s) in 0.0710 seconds

hbase(main):006:0>

```