

1) First create a table in h_base

```
hbase(main):010:0> create
```

```
'my_sales','order','sales','product','info','location','contact_details','deal_details'
```

```
0 row(s) in 1.2740 seconds
```

```
=> Hbase::Table - my_sales
```

```
hbase(main):011:0>
```

```
hbase(main):005:0> create 'my_sales','order','sales','product','info','location'
,'contact_details','deal_details'
0 row(s) in 2.6160 seconds

=> Hbase::Table - my_sales
hbase(main):006:0> █
```

2) Now create a landing table in Hive

```
create external table my_hive_sales_landing(
```

```
key int,
```

```
ORDERNUMBER int,
```

```
QUANTITYORDERED int,
```

```
PRICEEACH float,
```

```
ORDERLINENUMBER int,
```

```
SALE float,
```

```
STATUS string,
```

```
QUATER int,
```

```
MONTH int,
```

```
YEAR int,
```

```
PRODUCTNAME string,
```

```
MRP int,
```

```
PRODUCTCODE string,
```

PHONE string,

CITY string,

STATE string,

POSTALCODE string,

COUNTRY string,

TERRITORY string,

CONTACTLASTNAME string,

CONTACTFIRSTNAME string,

DEALSIZE string

)

row format delimited

fields terminated by ','

tblproperties("skip.header.line.count"="1");

```

Time taken: 0.267 seconds
hive> create external table my_hive_sales_landing(
  > key int,
  > ORDERNUMBER int,
  > QUANTITYORDERED int,
  > PRICEEACH float,
  > ORDERLINENUMBER int,
  > SALE float,
  > STATUS string,
  > QUATER int,
  > MONTH int,
  > YEAR int,
  > PRODUCTNAME string,
  > MRP int,
  > PRODUCTCODE string,
  > PHONE string,
  > CITY string,
  > STATE string,
  > POSTALCODE string,
  > COUNTRY string,
  > TERRITORY string,
  > CONTACTLASTNAME string,
  > CONTACTFIRSTNAME string,
  > DEALSIZE string
  > )
  > row format delimited
  > fields terminated by ','
  > tblproperties("skip.header.line.count"="1");
OK
Time taken: 0.736 seconds
hive> 

```

3)load the data from local to landing table

load data local inpath '/home/cloudera/sales_order_date.csv' into table my_hive_sales_landing

```

hive> load data local inpath '/home/cloudera/sales_order_date.csv' into table my_hive_sales_landing;
Loading data to table assignment.my_hive_sales_landing
Table assignment.my_hive_sales_landing stats: [numFiles=1, totalSize=357411]
OK
Time taken: 0.885 seconds
hive> 

```

4)select * from my_hive_sales_landing limit 5;

```

hive> select * from my_hive_sales_landing limit 5;
OK
my_hive_sales_landing.key      my_hive_sales_landing.ordernumber  my_hive_sales_landing.quantityordered  my_hive_sales_landing.priceeach  my_hive_sales_landing.orderlinenumber  my_hive_sales_landing.sale  my_hive_sales_landing.status  my_hive_sales_landing.quater  my_hive_sales_landing.month  my_hive_sales_landing.year  my_hive_sales_landing.productname  my_hive_sales_landing.mrp  my_hive_sales_landing.productcode  my_hive_sales_landing.phone  my_hive_sales_landing.city  my_hive_sales_landing.state  my_hive_sales_landing.contactlastname  my_hive_sales_landing.contactfirstname  my_hive_sales_landing.dealsize
10107 30 95 2.0 2871 NULL 1 2 2003 NULL 95 NULL 2125557818 NYC NY 10022 USA NA Yu Kwai Small NULL
10121 34 81 5.0 2765 NULL 2 5 2003 NULL 95 NULL 26.47.1555 Reims 51100 France EMEA Henriot Paul Small NULL
10134 41 94 2.0 3884 NULL 3 7 2003 NULL 95 NULL +33 1 46 62 7555 Paris 75508 France EMEA Da Cunha Daniel Medium
10145 45 83 6.0 3746 NULL 3 8 2003 NULL 95 NULL 6265557265 Pasadena CA 90003 USA NA Young Julie Medium NULL
10159 49 100 14.0 5205 NULL 4 10 2003 NULL 95 NULL 6505551386 San Francisco CA USA NA Brown Julie Medium NULL
Time taken: 0.385 seconds, Fetched: 5 row(s)
hive> 

```

5)Now create the actual table

```
create external table my_hive_sales_staging(  
key int,  
ORDERNUMBER int,  
QUANTITYORDERED int,  
PRICEEACH float,  
ORDERLINENUMBER int,  
SALE float,  
STATUS string,  
QUATER int,  
MONTH int,  
YEAR int,  
PRODUCTNAME string,  
MRP int,  
PRODUCTCODE string,  
PHONE string,  
CITY string,  
STATE string,  
POSTALCODE string,  
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,order:ORDERNUMBER,order:QUANTITYORDERED,order:PRICEEACH,order:ORDERLINENUMBER,sales:SALE,sales:STATUS,sales:QUATER,sales:MONTH,sales:YEAR,product:PRODUCTNAME,product:MRP,product:PRODUCTCODE,info:PHONE,info:CITY,info:STATE,info:POSTALCODE,info:COUNTRY,info:TERRITORY,contact_details:CONTACTLASTNAME,contact_details:CONTACTFIRSTNAME,deal_details:DEALSIZE")

TBLPROPERTIES("hbase.table.name"="my_sales");

```
hive> create external table my_hive_sales_staging(
  > key int,
  > ORDERNUMBER int,
  > QUANTITYORDERED int,
  > PRICEEACH float,
  > ORDERLINENUMBER int,
  > SALE float,
  > STATUS string,
  > QUATER int,
  > MONTH int,
  > YEAR int,
  > PRODUCTNAME string,
  > MRP int,
  > PRODUCTCODE string,
  > PHONE string,
  > CITY string,
  > STATE string,
  > POSTALCODE string,
  > 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,order:ORDERNUMBER,order:QUANTITYORDERED,order:PRICEEACH,order:ORDERLINENUMBER,sales:SALE,sales:STATUS,sales:QUATER,sales:MONTH,sales:YEAR,product:PRODUCTNAME,product:MRP,product:PRODUCTCODE,info:PHONE,info:CITY,info:STATE,info:POSTALCODE,info:COUNTRY,info:TERRITORY,contact_details:CONTACTLASTNAME,contact_details:CONTACTFIRSTNAME,deal_details:DEALSIZE")
  > TBLPROPERTIES("hbase.table.name"="my_sales");
OK
Time taken: 1.052 seconds
hive> []
```

5)Now we need to load the data from landing table to staging table

insert into table my_hive_sales_staging select * from my_hive_sales_landing

```
hive> insert into table my_hive_sales_staging select * from my_hive_sales_landing;
Query ID = c1oudera_20221109195656_8c5c09b4-7503-4ef3-9a9c-113874af68b9
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_1668051658294_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1668051658294_0001/
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1668051658294_0001
Hadoop job information for Stage-0: number of mappers: 17 number of reducers: 0
2022-11-09 19:57:18,909 Stage-0 map = 0%, reduce = 0%
2022-11-09 19:57:59,030 Stage-0 map = 100%, reduce = 0%, Cumulative CPU 18.3 sec
MapReduce Total Cumulative CPU time: 18 seconds 300 msec
Ended Job = job_1668051658294_0001
MapReduce Jobs Launched:
Stage-Stage-0: Map: 1 Cumulative CPU: 18.3 sec HDFS Read: 372296 HDFS Write: 0 SUCCESS
Total MapReduce CPU Time Spent: 18 seconds 300 msec
OK
my_hive_sales_landing.key      my_hive_sales_landing.ordernumber  my_hive_sales_landing.quantityordered  my_hive_sales_landing.priceeach  my_hive_sales_landing.orderlinenumber  my_hive_sales_landing.sale  my_hive_sales_landing.status  my_hive_sales_landing.quater  my_hive_sales_landing.month  my_hive_sales_landing.year  my_hive_sales_landing.productname  my_hive_sales_landing.mrp  my_hive_sales_landing.productcode  my_hive_sales_landing.phone  my_hive_sales_landing.city  my_hive_sales_landing.state  my_hive_sales_landing.postalcode  my_hive_sales_landing.country  my_hive_sales_landing.territory  my_hive_sales_landing.contactlastname  my_hive_sales_landing.contactfirstname  my_hive_sales_landing.dealsize
Time taken: 80.588 seconds
hive> []
```

6)select * from my_hive_sales_staging limit 5;

```
hive> select * from my_hive_sales_staging limit 5;
OK
my_hive_sales_staging.key      my_hive_sales_staging.ordernumber  my_hive_sales_staging.quantityordered  my_hive_sales_staging.priceseach  my_hive_sales_staging.orderlinenumber  my_hive_sales_staging.sale  my_hive_sales_staging.status  my_hive_sales_staging.quater  my_hive_sales_staging.month  my_hive_sales_staging.year  my_hive_sales_staging.productname  my_hive_sales_staging.mrp  my_hive_sales_staging.productcode  my_hive_sales_staging.phone  my_hive_sales_staging.city  my_hive_sales_staging.state  my_hive_sales_staging.postalcode  my_hive_sales_staging.country  my_hive_sales_staging.territory  my_hive_sales_staging.contactlastname  my_hive_sales_staging.contactfirstname  my_hive_sales_staging.dealsize
1      10107  30      95.7      2      2871.0  Shipped  1      2      2003      Motorcycles      95      S10_1678      2125557818      NYC      NY      10022      USA      NA      Yu      KwaiS
mall
10     10211  41      100.0     14      4708.44 Shipped  1      1      2004      Motorcycles      95      S10_1678      (1) 47.55.6555      Paris      75016      France      EMEA      Perrier Domin
ique
100    10229  26      100.0     3      5868.2  Shipped  4      11     2004      Motorcycles      193     S10_4698      2125557818      NYC      NY      10022      USA      NA      Yu      KwaiM
edium
1000   10232  22      100.0     6      3606.02 Shipped  1      3      2004      Vintage Cars      136     S18_3140      (198) 555-8888      Cowes      Isle of Wight      PO31 7PJ      UK      EMEA
Bennett Helen Medium
1001   10248  32      100.0     12     3802.56 Cancelled  2      5      2004      Vintage Cars      136     S18_3140      2125557818      NYC      NY      10022      USA      NA      Yu      K
wai Medium
Time taken: 0.119 seconds, Fetched: 5 row(s)
```

7)scan 'my_sales',{'LIMIT'=5}

```
hbase(main):001:0> scan 'my_sales',{'LIMIT'=>5}
ROW COLUMN+CELL
1      column=contact details:CONTACTFIRSTNAME, timestamp=1668053778183, value=Kwai
1      column=contact details:CONTACTLASTNAME, timestamp=1668053778183, value=Yu
1      column=deal details:DEALSIZE, timestamp=1668053778183, value=Small
1      column=info:CITY, timestamp=1668053778183, value=NYC
1      column=info:COUNTRY, timestamp=1668053778183, value=USA
1      column=info:PHONE, timestamp=1668053778183, value=2125557818
1      column=info:POSTALCODE, timestamp=1668053778183, value=10022
1      column=info:STATE, timestamp=1668053778183, value=NY
1      column=info:TERRITORY, timestamp=1668053778183, value=NA
1      column=order:ORDERLINENUMBER, timestamp=1668053778183, value=2
1      column=order:ORDERNUMBER, timestamp=1668053778183, value=10107
1      column=order:PRICEEACH, timestamp=1668053778183, value=95.7
1      column=order:QUANTITYORDERED, timestamp=1668053778183, value=30
1      column=product:MRP, timestamp=1668053778183, value=95
1      column=product:PRODUCTCODE, timestamp=1668053778183, value=S10_1678
1      column=product:PRODUCTNAME, timestamp=1668053778183, value=Motorcycles
1      column=sales:MONTH, timestamp=1668053778183, value=2
1      column=sales:QUATER, timestamp=1668053778183, value=1
1      column=sales:SALE, timestamp=1668053778183, value=2871.0
1      column=sales:STATUS, timestamp=1668053778183, value=Shipped
1      column=sales:YEAR, timestamp=1668053778183, value=2003
10     column=contact details:CONTACTFIRSTNAME, timestamp=1668053778183, value=Dominique
10     column=contact details:CONTACTLASTNAME, timestamp=1668053778183, value=Perrier
10     column=deal details:DEALSIZE, timestamp=1668053778183, value=Medium
10     column=info:CITY, timestamp=1668053778183, value=Paris
10     column=info:COUNTRY, timestamp=1668053778183, value=France
10     column=info:PHONE, timestamp=1668053778183, value=(1) 47.55.6555
10     column=info:POSTALCODE, timestamp=1668053778183, value=75016
10     column=info:STATE, timestamp=1668053778183, value=
10     column=info:TERRITORY, timestamp=1668053778183, value=EMEA
10     column=order:ORDERLINENUMBER, timestamp=1668053778183, value=14
10     column=order:ORDERNUMBER, timestamp=1668053778183, value=10211
10     column=order:PRICEEACH, timestamp=1668053778183, value=100.0
10     column=order:QUANTITYORDERED, timestamp=1668053778183, value=41
10     column=product:MRP, timestamp=1668053778183, value=95
```

8)Adding values from H-base table

hbase(main):005:0> put 'my_sales','9999','order:ORDERNUMBER','11'

0 row(s) in 0.0440 seconds

hbase(main):006:0> put 'my_sales','9999','product:PRODUCTCODE','abc'

0 row(s) in 0.0120 seconds

hbase(main):007:0> put 'my_sales','9999','info:CITY','Guntur'

0 row(s) in 0.0130 seconds

```
hbase(main):008:0> put 'my_sales','9999','info:COUNTRY','India'
```

0 row(s) in 0.0140 seconds

```
hbase(main):007:0> get 'my_sales','9999'
COLUMN                                CELL
info:CITY                             timestamp=1668053986438, value=Guntur
info:COUNTRY                           timestamp=1668053994591, value=India
order:ORDERNUMBER                     timestamp=1668053938895, value=11
product:PRODUCTCODE                  timestamp=1668053948277, value=abc
4 row(s) in 0.0210 seconds
```

9)Checking whether the data is getting refelect in Hive

```
hive> select * from my_hive_sales_staging where key=9999;
OK
my_hive_sales_staging.key      my_hive_sales_staging.ordernumber  my_hive_sales_staging.quantityordered  my_hive_sales_staging.priceeach  my_hive_sales_staging.orderlinenumber  my_hive_sales_staging.sale  my_hive_sales_staging.status  my_hive_sales_staging.quater  my_hive_sales_staging.month  my_hive_sales_staging.year  my_hive_sales_staging.productname  my_hive_sales_staging.mrp  my_hive_sales_staging.productcode  my_hive_sales_staging.phone  my_hive_sales_staging.city  my_hive_sales_staging.state  my_hive_sales_staging.postalcode  my_hive_sales_staging.country  my_hive_sales_staging.territory  my_hive_sales_staging.contactlastname  my_hive_sales_staging.contactfirstname  my_hive_sales_staging.dealsize
9999      11      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL      NULL
Time taken: 0.134 seconds, Fetched: 1 row(s)
hive>
```