Creating data base in hive:

create database sales_order_hive;

```
hive> create database sales_order_hive;
OK
Time taken: 3.322 seconds
hive> use sales_order_hive;
OK
Time taken: 0.107 seconds
```

using data base:

```
use sales_order_hive;
```

```
hive> use sales_order_hive;
OK
Time taken: 0.107 seconds
```

crating stage table in hive:

```
CREATE EXTERNAL TABLE IF not exists sales_order_hive_stage (

ORDERNUMBER int,

QUANTITYORDERED int,

PRICEEACH float,

ORDERLINENUMBER int,

SALES float,

STATUS string,

QTR_ID int,

MONTH_ID int,

YEAR_ID int,

PRODUCTLINE string,

MSRP int,

PRODUCTCODE varchar(30),

PHONE varchar(30),
```

```
CITY string,

STATE string,

POSTALCODE varchar(30),

COUNTRY string,

TERRITORY string,

CONTACTLASTNAME string,

CONTACTFIRSTNAME string,

DEALSIZE string

)

ROW FORMAT DELIMITED

FIELDS TERMINATED BY ','

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

```
hive> CREATE EXTERNAL TABLE IF not exists sales_order_hive_stage
   > ORDERNUMBER int,
   > QUANTITYORDERED int,
    > PRICEEACH float,
   > ORDERLINENUMBER int,
   > SALES float,
   > STATUS string,
   > QTR ID int,
   > MONTH ID int,
    > YEAR ID int,
   > PRODUCTLINE string,
   > MSRP int,
   > PRODUCTCODE varchar(30),
   > PHONE varchar(30),
   > CITY string,
   > STATE string,
    > POSTALCODE varchar(30),
   > COUNTRY string,
   > TERRITORY string,
   > CONTACTLASTNAME string,
   > CONTACTFIRSTNAME string,
   > DEALSIZE string
    > ROW FORMAT DELIMITED
   > FIELDS TERMINATED BY ','
   > tblproperties("skip.header.line.count"="1")
Time taken: 0.176 seconds
```

Loading data into table:

load data local inpath '/home/cloudera/Downloads/sales_order_data.csv' into table sales_order_hive_stage;

```
hive> load data local inpath '/home/cloudera/Downloads/sales_order_data.csv' into table sales_order_hive_stage;
Loading data to table sales_order_hive.sales_order_hive_stage
Table sales_order_hive.sales_order_hive_stage stats: [numFiles=1, numRows=0, totalSize=360233, rawDataSize=0]
OK
Time taken: 1.492 seconds
```

Cross verifying the data from resource:

select * from sales_order_hive_stage limit 10;

```
from sales_order_hive_stage limit
                   2871.0 Shipped 1
                                                                          S10 1678
                   2765.9 Shipped 2
                                                                          S10 1678
                                                                                        26.47.1555 Reims
                                                                                                                    51100 France EMEA
                                                                          S10 1678
                    3884.34 Shipped 3
                                                                                                                                 France
                    Medium
3746.7 Shipped 3
                                                                                                      Pasadena
                                                                                                      Burlingame
                   2497.77 Shipped 4
                                                                                                                    59000 France EMEA
Martine Small
                                                                                                                  N 5804 Norway EMEA
```

select coubt(*) from sales_order_hive_stage;

```
FAILED: SemanticException [Error 10011]: Line 1:7 Invalid function 'coubt' hive> select count(*) from sales_order_hive_stage;
Query ID = cloudera_20221104091818_7745dcb4-a7la-4f6b-906d-1ff1215af6e8
 Total jobs = 1
 Number of reduce tasks determined at compile time: 1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
  set hive.exec.reducers.max=<number>
In order to set a constant number of reducers:
  set mapreduce.job.reduces=<number>
 Starting Job = job_1667576221039_0001, Tracking URL = http://quickstart.cloudera:8088/proxy/application_1667576221039_0001
Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_1667576221039_0001
Hadoop job information for Stage-1: number of mappers: 1; number of reducers: 1
2022-11-04 09:19:27,749 Stage-1 map = 0%, reduce = 0%
2022-11-04 09:19:43,718 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 2.25 sec
2022-11-04 09:20:00,057 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 4.74 sec
MapReduce Total cumulative CPU time: 4 seconds 740 msec
 Ended Job = job_1667576221039_0001
MapReduce Jobs Launched:
Stage-Stage-1: Map: 1 Reduce: 1 Cumulative CPU: 4.74 sec HDFS Read: 370705 HDFS Write: 5 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 740 msec
Time taken: 65.955 seconds, Fetched: 1 row(s)
```

Creating table in Hbase:

create

'sales_order_hbase','ORDERNUMBER','QUANTITYORDERED','PRICEEACH','ORDERLINENUMBER','SALES','
STATUS','QTR_ID','MONTH_ID','YEAR_ID','PRODUCTLINE','MSRP','PRODUCTCODE','PHONE','CITY','STATE'
,'POSTALCODE','COUNTRY','TERRITORY','CONTACTLASTNAME','CONTACTFIRSTNAME','DEALSIZE'

Creating connectivity table from hive to hbase:

```
CREATE EXTERNAL TABLE sales_order_hive_main
ORDERNUMBER int,
QUANTITYORDERED int,
PRICEEACH float,
ORDERLINENUMBER int,
SALES float,
STATUS string,
QTR_ID int,
MONTH_ID int,
YEAR_ID int,
PRODUCTLINE string,
MSRP int,
PRODUCTCODE varchar(30),
PHONE varchar(30),
CITY string,
STATE string,
POSTALCODE varchar(30),
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,QUANTITYORDERED:QUANTITYORDERED,PRICEEACH:PRICEEACH;ORD ERLINENUMBER: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"="sales_order_hbase_1");

Hbase table details:

describe 'sales order hbase'

```
Date main 1031100 describe 'Bales order hase'
Table sales order hase is ENBEDD

sales order has order
```

21 row(s) in 0.1560 seconds

Copying data from stage table to main table in hive:

insert into sales_order_hive_main select * from sales_order_hive_stage;

```
hive> insert into sales_order_hive_main select * from sales_order_hive_stage;

Query ID = cloudera_20221104102727_5df5d79e-2355-4b8b-a9lb-4602bbl62ad3

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_l667576221039_0002, Tracking URL = http://quickstart.cloudera:8088/proxy/application_l667576221039_0002/

Kill Command = /usr/lib/hadoop/bin/hadoop job -kill job_l667576221039_0002

Hadoop job information for Stage-0: number of mappers: 1; number of reducers: 0

2022-11-04 10:28:25,106 Stage-0 map = 0%, reduce = 0%

2022-11-04 10:29:05,943 Stage-0 map = 100%, reduce = 0%, Cumulative CPU 9.54 sec

MapReduce Total cumulative CPU time: 9 seconds 540 msec

Ended Job = job_l667576221039_0002

MapReduce Jobs Launched:

Stage-Stage-0: Map: 1 Cumulative CPU: 9.54 sec HDFS Read: 375411 HDFS Write: 0 SUCCESS

Total MapReduce CPU Time Spent: 9 seconds 540 msec

OK

Time taken: 81.334 seconds
```

Checking data inside hive main table:

select * from sales order hive main limit 10;

				1689.03	Shipped			Vintage Cars		S24_3969	6035558	647	Nashua	NH		USA	NA	
	Valarie 46			2472.96	Shipped		2003	Vintage Cars	44	S24 2022	+49 69	66 90 25		Frankfu	rt		60528	
			Roland	Small														
	41			2055.74	Shipped		2003	Vintage Cars		S18 1367	2125551	500	NYC	NY	10022	USA	NA	
k	Michael	Small																
				4460.82	Shipped		2003	Classic Cars		S700_2824	07-98 9	555	Stavern		4110	Norway	EMEA	
ulfse		Jonas	Medium															
		53.31		1705.92	Shipped		2003	Trains 58	S50_151	4 (91	1) 555 94 44	Madrid		28034	Spain	EMEA	Freyr	
	Small	56.50							GGO 001		10 0555				1704			
	25 Jytte	56.78 Small		1419.5	Snipped		2003	Ships 54	S72_321		12 3555	Kobenha	vn		1734	Denmark	EMEA	
	48		15	2526.72	Shinned		2003	Planes 49	572 125	3 039	5-640555	Bergamo		24100	Italy	EMEA	Rovel	á
	Small	02.01		2020172	Diiipped		2000	1141100 15	D,12_110		0 010000	Dergamo		21200	10011	211211	1.07-01	Ī
	20	92.9		1858.0	Shipped		2003	Motorcycles	99	532 1374	2125557	818	NYC	NY	10022	USA	NA	
	Kwai	Small																
					Shipped			Motorcycles		S50_4713		55 3587	Makati			1227 MM	1 Phili	p
	Japan		Arnold	Small														
	29				Shipped		2003	Classic Cars		S24_2972	2155559	857	Philade	lphia	PA	71270	USA	
	29 Hernande	32.1 z	6 Rosa					Classic Cars			2155559	857	Philade	lphia		712		70 USA

Checking data inside hbase table:

scan 'sales_order_hbase_1'