hive> show tables;

OK

Time taken: 1.401 seconds

hive> create database ourfirstdatabase;

hive> create table first\_table(firstname string,lastname string,employeeid int);

hive> describe database extended ourfirstdatabase;

OK

ourfirstdatabase hdfs://localhost:9000/user/hive/warehouse/ourfirstdatabase.db hduser USER

Time taken: 0.066 seconds, Fetched: 1 row(s)

hive> ALTER DATABASE ourfirstdatabase SET DBPROPERTIES (‘creator'='Varsha','created\_for'='Hive DDL');

OK

Time taken: 0.27 seconds

hive> describe database extended ourfirstdatabase;

OK

ourfirstdatabase hdfs://localhost:9000/user/hive/warehouse/ourfirstdatabase.db hduser USER {creator=Varsha, created\_for=Hive DDL}

Time taken: 0.182 seconds, Fetched: 1 row(s)

hive> insert into first\_table(firstname,lastname,employeeid) values("varsha","Khandekar",23);

hive> select \* from first\_table;

OK

varsha Khandekar 23

Time taken: 0.599 seconds, Fetched: 1 row(s)

hive> exit;

hduser@ubuntu:~$ cd /usr/local/hbase/bin

hduser@ubuntu:/usr/local/hbase/bin$ ./start-hbase.sh

localhost: starting zookeeper, logging to /usr/local/hbase/bin/../logs/hbase-hduser-zookeeper-ubuntu.out

starting master, logging to /usr/local/hbase/logs/hbase-hduser-master-ubuntu.out

starting regionserver, logging to /usr/local/hbase/logs/hbase-hduser-1-regionserver-ubuntu.out

hduser@ubuntu:/usr/local/hbase/bin$ ./hbase shell

hbase(main):001:0> list

TABLE

0 row(s) in 0.7760 seconds

=> []

hbase(main):002:0> create 'flightinfo','flightdetails'

0 row(s) in 3.0820 seconds

=> Hbase::Table – flightinfo

hbase(main):004:0> describe 'flightinfo'

hbase(main):005:0> count 'flightinfo'

hbase(main):006:0> count 'flightinfo', CACHE=>1000 # count no of rows

hbase(main):010:0> alter 'flightinfo',{name=>'source'},{name=>'dest'},{name=>'dept\_time'},{name=>'rate'}

hbase(main):015:0> put 'flightinfo', '1','flightdetails:source','Pune'

0 row(s) in 0.4220 seconds

hbase(main):016:0> put 'flightinfo', '2','flightdetails:source','Mumbai'

0 row(s) in 0.0310 seconds

hbase(main):017:0> put 'flightinfo', '3','flightdetails:source','Mumbai'

0 row(s) in 0.0200 seconds

hbase(main):018:0> put 'flightinfo', '3','flightdetails:source','Bengluru'

0 row(s) in 0.0140 seconds

hbase(main):019:0> put 'flightinfo', '4','flightdetails:source','Chennai'

0 row(s) in 0.0330 seconds

hbase(main):020:0> put 'flightinfo', '5','flightdetails:source','Delhi'

0 row(s) in 0.0520 seconds

hbase(main):021:0> put 'flightinfo', '1','flightdetails:dest','Delhi'

0 row(s) in 0.0380 seconds

hbase(main):022:0> put 'flightinfo', '2','flightdetails:dest','Chennai'

0 row(s) in 0.0230 seconds

hbase(main):023:0> put 'flightinfo', '3','flightdetails:dest','kolkatta'

0 row(s) in 0.0400 seconds

hbase(main):024:0> put 'flightinfo', '4','flightdetails:dest','Nagpur'

0 row(s) in 0.0280 seconds

hbase(main):025:0> put 'flightinfo', '5','flightdetails:dest','Mumbai'

0 row(s) in 0.0100 seconds

hbase(main):026:0> put 'flightinfo', '1','flightdetails:dept\_time','2:00 am'

0 row(s) in 0.0340 seconds

hbase(main):027:0> put 'flightinfo', '2','flightdetails:dept\_time','3:30 am'

0 row(s) in 0.0320 seconds

hbase(main):028:0> put 'flightinfo', '3','flightdetails:dept\_time','6:30 am'

0 row(s) in 0.0260 seconds

hbase(main):029:0> put 'flightinfo', '4','flightdetails:dept\_time','9:30 am'

0 row(s) in 0.0640 seconds

hbase(main):030:0> put 'flightinfo', '5','flightdetails:dept\_time','11:30 am'

0 row(s) in 0.0300 seconds

hbase(main):032:0> put 'flightinfo', '1','flightdetails:rate','3000'

0 row(s) in 0.1140 seconds

hbase(main):033:0> put 'flightinfo', '2','flightdetails:rate','4000'

0 row(s) in 0.0250 seconds

hbase(main):034:0> put 'flightinfo', '3','flightdetails:rate','2500'

0 row(s) in 0.0420 seconds

hbase(main):035:0> put 'flightinfo', '4','flightdetails:rate','5500'

0 row(s) in 0.0070 seconds

hbase(main):036:0> put 'flightinfo', '5','flightdetails:rate','3500'

0 row(s) in 0.0160 seconds

hbase(main):031:0> scan 'flightinfo'

Drop table

To drop table , first disable table.

hbase(main):039:0> create 'samtable','cfamily'

0 row(s) in 1.2830 seconds

hbase(main):041:0> disable 'samtable'

0 row(s) in 2.4540 seconds

hbase(main):042:0> drop 'samtable'

0 row(s) in 1.4340 seconds

hbase(main):004:0> get 'flightinfo',1

COLUMN CELL

flightdetails:dept\_t timestamp=1552558651479, value=2:00 am

ime

flightdetails:dest timestamp=1552558474321, value=Delhi

flightdetails:rate timestamp=1552559105615, value=3000

flightdetails:source timestamp=1552558323593, value=Pune

4 row(s) in 0.1770 seconds

hbase(main):007:0> get 'flightinfo',1,COLUMN=>'flightdetails:source'

1. Creating an external Hive table to connect to the HBase for Customer Information Table

create Customer Information Table

hbase(main):002:0> create 'CustomerInformation','ContactInfo','CustomerName'

Add Data in CustomerInformation

hbase(main):004:0> put 'CustomerInformation','1','ContactInfo:phno','9011092397'

0 row(s) in 0.3770 seconds

hbase(main):005:0> put 'CustomerInformation','1','ContactInfo:city','pune'

0 row(s) in 0.0200 seconds

hbase(main):006:0> put 'CustomerInformation','1','ContactInfo:phno','9011092396'

0 row(s) in 0.0330 seconds

hbase(main):007:0> put 'CustomerInformation','1','ContactInfo:phno','9011092397'

0 row(s) in 0.0240 seconds

hbase(main):008:0> put 'CustomerInformation','2','ContactInfo:phno','9011092396'

0 row(s) in 0.0230 seconds

hbase(main):009:0> put 'CustomerInformation','3','ContactInfo:phno','9011092398'

0 row(s) in 0.0570 seconds

hbase(main):010:0> put 'CustomerInformation','4','ContactInfo:phno','9011092399'

0 row(s) in 0.0400 seconds

hbase(main):011:0> put 'CustomerInformation','5','ContactInfo:phno','9011092395'

0 row(s) in 0.0210 seconds

hbase(main):012:0> put 'CustomerInformation','2','ContactInfo:city','mumbai'

0 row(s) in 0.0310 seconds

hbase(main):013:0> put 'CustomerInformation','3','ContactInfo:city','Delhi'

0 row(s) in 0.0280 seconds

hbase(main):014:0> put 'CustomerInformation','4','ContactInfo:city','newyork'

0 row(s) in 0.0300 seconds

hbase(main):015:0> put 'CustomerInformation','5','ContactInfo:city','chennai'

-0 row(s) in 0.0270 se

hbase(main):012:0> put 'CustomerInformation','2','ContactInfo:city','mumbai'

0 row(s) in 0.0310 seconds

hbase(main):013:0> put 'CustomerInformation','3','ContactInfo:city','Delhi'

0 row(s) in 0.0280 seconds

hbase(main):014:0> put 'CustomerInformation','4','ContactInfo:city','newyork'

0 row(s) in 0.0300 seconds

hbase(main):015:0> put 'CustomerInformation','5','ContactInfo:city','chennai'

0 row(s) in 0.0270 seconds

hbase(main):016:0> put 'CustomerInformation','1','CustomerName:fname','varsha'

0 row(s) in 0.0460 seconds

hbase(main):017:0> put 'CustomerInformation','2','CustomerName:fname','yogita'

0 row(s) in 0.0250 seconds

hbase(main):018:0> put 'CustomerInformation','3','CustomerName:fname','Gauri'

0 row(s) in 0.0410 seconds

hbase(main):019:0> put 'CustomerInformation','4','CustomerName:fname','Nivi'

0 row(s) in 0.0260 seconds

hbase(main):020:0> put 'CustomerInformation','5','CustomerName:fname','Pranav'

0 row(s) in 0.0380 seconds

hbase(main):021:0> put 'CustomerInformation','1','CustomerName:lname','Khandekar'

0 row(s) in 0.0320 seconds

hbase(main):022:0> put 'CustomerInformation','2','CustomerName:lname','Wagh'

0 row(s) in 0.0210 seconds

hbase(main):023:0> put 'CustomerInformation','3','CustomerName:lname','Dhumal'

0 row(s) in 0.0460 seconds

hbase(main):024:0> put 'CustomerInformation','4','CustomerName:lname','Deshmukh'

0 row(s) in 0.0180 seconds

hbase(main):025:0> put 'CustomerInformation','5','CustomerName:lname','Pawar'

Display table

hbase(main):026:0> scan 'CustomerInformation'

Display particular row

hbase(main):027:0> get 'CustomerInformation','1'

COLUMN CELL

ContactInfo:city timestamp=1553171225541, value=pune

ContactInfo:phno timestamp=1553171271318, value=9011092397

CustomerName:fname timestamp=1553171555484, value=varsha

CustomerName:lname timestamp=1553171650435, value=Khandekar

4 row(s) in 0.1160 seconds

Open hive terminal

hive> create external table hive\_hbase\_table(key int,phno string,city string,fname string,lname string)STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler' WITH SERDEPROPERTIES("hbase.columns.mapping" = ":key,ContactInfo:phno,ContactInfo:city,CustomerName:fname,CustomerName:lname") TBLPROPERTIES ("hbase.table.name" = "CustomerInformation");

OK

Time taken: 3.785 seconds

hive> select \* from hive\_hbase\_table;

OK

1 9011092397 pune varsha Khandekar

2 9011092396 mumbai yogita Wagh

3 9011092398 Delhi Gauri Dhumal

4 9011092399 newyork Nivi Deshmukh

5 9011092395 chennai Pranav Pawar

Time taken: 0.176 seconds, Fetched: 5 row(s)

hive> select city from hive\_hbase\_table;

OK

pune

mumbai

Delhi

newyork

chennai

Time taken: 0.416 seconds, Fetched: 5 row(s)

hive> select city from hive\_hbase\_table where fname="varsha";

OK

pune

Time taken: 0.585 seconds, Fetched: 1 row(s)

1. Download data set of 2007 & 2008 from http://stat-computing.org/dataexpo/2009/the-data.html

hive> CREATE TABLE IF NOT EXISTS FlightInfo1\_2007 ( Year SMALLINT, Month TINYINT, DayofMonth TINYINT,DayOfWeek TINYINT,DepTime SMALLINT, CRSDepTime SMALLINT, ArrTime SMALLINT,CRSArrTime SMALLINT, UniqueCarrier STRING, FlightNum STRING, TailNum STRING, ActualElapsedTime SMALLINT, CRSElapsedTime SMALLINT, AirTime SMALLINT, ArrDelay SMALLINT, DepDelay SMALLINT,Origin STRING, Dest STRING,Distance INT, TaxiIn SMALLINT, TaxiOut SMALLINT, Cancelled SMALLINT, CancellationCode STRING, Diverted SMALLINT,WeatherDelay SMALLINT ,LateAircraftDelay SMALLINT) COMMENT 'Flight InfoTable'ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE TBLPROPERTIES ('creator'='PSB ', 'created\_at'='Tues Dec 5 3:00:00 EDT 2017');

OK

hive> load data local inpath '/home/hduser/Downloads/2007.csv' into table flightinfo2007;

Loading data to table default.flightinfo2007

Table default.flightinfo2007 stats: [numFiles=1, totalSize=0]

OK

Time taken: 1.371 second

hive> CREATE TABLE IF NOT EXISTS FlightInfo1\_2008 LIKE FlightInfo1\_2007;

hive> load data local inpath '/home/hduser/Downloads/2008.csv' into table FlightInfo1\_2008;

hive>CREATE TABLE IF NOT EXISTS myFlightInfo1 (Year SMALLINT, DontQueryMonth TINYINT, DayofMonth TINYINT, DayOfWeek TINYINT, DepTime SMALLINT, ArrTime SMALLINT,UniqueCarrier STRING, FlightNum STRING,AirTime SMALLINT, ArrDelay SMALLINT, DepDelay SMALLINT,Origin STRING, Dest STRING, Cancelled SMALLINT,CancellationCode STRING) COMMENT 'Flight InfoTable' PARTITIONED BY(Month TINYINT)ROW FORMAT DELIMITED FIELDS TERMINATED BY ','LINES TERMINATED BY '\n'STORED AS RCFILE TBLPROPERTIES ('creator'='Varsha','created\_at'='Fri March 15 14:24:19 EDT 2019');

hive> INSERT OVERWRITE TABLE myflightinfo1 PARTITION (Month=1) SELECT Year, Month, DayofMonth, DayOfWeek, DepTime,ArrTime, UniqueCarrier,FlightNum, AirTime, ArrDelay, DepDelay, Origin,Dest, Cancelled,CancellationCode FROM flightinfo1\_2008 WHERE Month=1;

hive> FROM FlightInfo1\_2008 INSERT INTO TABLE myflightinfo1 PARTITION (Month=2) SELECT Year, Month, DayofMonth, DayOfWeek, DepTime,ArrTime, UniqueCarrier, FlightNum, AirTime, ArrDelay, DepDelay, Origin, Dest, Cancelled,CancellationCode WHERE Month=2 INSERT INTO TABLE myflightinfo1 PARTITION (Month=12) SELECT Year, Month, DayofMonth, DayOfWeek, DepTime,ArrTime, UniqueCarrier, FlightNum,AirTime, ArrDelay, DepDelay, Origin, Dest, Cancelled,CancellationCode WHERE Month=12;

hive> CREATE TABLE myflightinfo2007 AS SELECT Year, Month, DepTime, ArrTime, FlightNum,Origin, Dest FROM FlightInfo1\_2007 WHERE (Month = 7 AND DayofMonth = 3) AND (Origin='JFK' AND Dest='ORD');

hive> SELECT \* FROM myFlightInfo2007;

OK

2007 7 700 834 5447 JFK ORD

2007 7 1633 1812 5469 JFK ORD

2007 7 1905 2100 5492 JFK ORD

2007 7 1453 1624 4133 JFK ORD

2007 7 1810 1956 4392 JFK ORD

2007 7 643 759 903 JFK ORD

2007 7 939 1108 907 JFK ORD

2007 7 1313 1436 915 JFK ORD

\*2007 7 1617 1755 917 JFK ORD

2007 7 2002 2139 919 JFK ORD

Time taken: 0.714 seconds, Fetched: 10 row(s)

hive> CREATE TABLE myflightinfonew132008 AS SELECT Year, Month, DepTime, ArrTime, FlightNum,Origin, Dest FROM FlightInfo1\_2008 WHERE (Month = 7 AND DayofMonth = 3) AND (Origin='JFK' AND Dest='ORD');

hive> select \* from myflightinfonew132008

hive>SELECT m8.Year, m8.Month, m8.FlightNum, m8.Origin, m8.Dest, m7.Year, m7.Month, m7.FlightNum, m7.Origin, m7.Dest FROM myFlightinfonew132008 m8 JOIN myFlightinfo2007 m7 ON m8.FlightNum=m7.FlightNum;

hive> SELECT m8.FlightNum,m8.Origin,m8.Dest,m7.FlightNum,m7.Origin,m7.Dest FROM myFlightinfonew132008 m8 FULL OUTER JOIN myFlightinfo2007 m7 ON m8.FlightNum=m7.FlightNum;

create index

hive> CREATE INDEX f08\_index ON TABLE flightinfo1\_2008 (Origin) AS 'COMPACT' WITH DEFERRED REBUILD;

OK

Time taken: 1.788 seconds

hive> SHOW INDEXES ON FlightInfo1\_2008;

OK

f08\_index flightinfo1\_2008 origin default\_\_flightinfo1\_2008\_f08\_index\_\_ compact

Time taken: 0.34 seconds, Fetched: 1 row(s)

hive>

hive> SELECT DayOfWeek, AVG(DepDelay) FROM flightinfo1\_2008 GROUP BY DayOfWeek;

hive> CREATE VIEW avgdepdelay AS SELECT DayOfWeek, AVG(DepDelay) FROM flightinfo1\_2008 GROUP BY DayOfWeek;

hive> select \* from avgdepdelay;

OK

output

NULL NULL

3 8.289761053658728

6 8.645680904903614

1 10.269990244459473

4 9.772897177836702

7 11.568973392595312

2 8.97689712068735

5 12.158036387869656

Time taken: 10.636 seconds, Fetched: 8 row(s)