

sample hive tables used:

hive data:

01.31.05,user5,www.apache.com,6,12.23.28.37
07.23.24,user6,www.pig.com,12,16.34.21.109
05.16.16,user7,www.mapred.com,18,18.46.38.35
09.18.47,user8,www.hdfs.com,64,28.52.89.102

hivepage:

10.31.05,user1,www.apache.com,6,12.23.28.37
12.41.24,user2,www.ubuntu.com,8,16.34.21.109
05.32.16,user3,www.amazon.com,34,18.46.38.35
09.31.47,user4,www.hadoop.com,78,28.52.89.102

hivepage1:

10.31.05,user1,www.apache.com,6,12.23.28.37
12.41.24,user2,www.ubuntu.com,8,16.34.21.109
05.32.16,user3,www.amazon.com,34,18.46.38.35
09.31.47,user4,www.hadoop.com,78,28.52.89.102

hivepage2:

10.31.05,user1,www.apache.com,6,12.23.28.37
12.41.24,user2,www.ubuntu.com,8,16.34.21.109
05.32.16,user3,www.amazon.com,34,18.46.38.35
09.31.47,user4,www.hadoop.com,78,28.52.89.102

hivepage3:

10.31.05,user1,www.apache.com,6,12.23.28.37
12.41.24,user2,www.ubuntu.com,8,16.34.21.109
05.32.16,user3,www.amazon.com,34,18.46.38.35
09.31.47,user4,www.hadoop.com,78,28.52.89.102

hivecmplx:

John Doe,100000.0,Mary Smith:Todd Jones,Federal Taxes->2:State Taxes->05:Insurance->1,1 Michigan Ave:Chicago:IL:60600
Mary Smith,80000.0,Bill King,Federal Taxes->2:State Taxes->05:Insurance->1,100 Ontario St:Chicago:IL:60601
Todd Jones,70000.0,Siemon Jones:Stuart Bill,Federal Taxes->15:State Taxes->03:Insurance->1,200 Chicago Ave:Oak Park:IL:60700
Bill King,60000.0,petterson Mary:Bishop Ray,Federal Taxes->15:State Taxes->03:Insurance->1,300 Obscure Dr:Obscuria:IL:60100

=====

hive commands:

hive> show databases;

o/p:

OK

default

Time taken: 8.035 seconds, Fetched: 1 row(s)

creating and dropping a database:

create:

create database test;

show databases;

o/p:

OK

default

test

Time taken: 0.076 seconds, Fetched: 2 row(s)

you can browse the file system and see what is created

/user/hive/warehouse (this is by default; but it is configurable)

dropping Database:

drop database test;

o/p:

OK

Time taken: 0.686 seconds

=====

loading data from local FS:

creating table:

```
CREATE TABLE page_view_from_local(viewTime STRING, userid STRING,
page_url STRING,
pageNo INT,ip STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES
TERMINATED BY '\n';
```

loading data:

```
LOAD DATA LOCAL INPATH '/home/geetha/pgms/hivepage' INTO TABLE
page_view_from_local;
```

This will just copy the data from local directory

o/p:

Copying data from file:/home/geetha/pgms/hivepage

Copying file: file:/home/geetha/pgms/hivepage

Loading data to table default.page_view_from_local

```
Table default.page_view_from_local stats: [num_partitions: 0, num_files:
1, num_rows: 0, total_size: 180, raw_data_size: 0]
OK
```

```
load data into hdfs:
-----
```

```
hive> dfs -copyFromLocal /home/geetha/pgms/hivedata /usr/hadoop-
inst/hadoop-1.0.3/hivedata;
```

```
creating table:
-----
```

```
CREATE TABLE page_view_from_hdfs(viewTime STRING, userid STRING, page_url
STRING,
pageNo INT,ip STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES
TERMINATED BY '\n';
```

```
loading data:
-----
```

```
LOAD DATA INPATH '/usr/hadoop-inst/hadoop-1.0.3/hivedata' INTO TABLE
page_view_from_hdfs;
```

This will MOVE the data from HDFS directory to hive ware house

```
Copying data from file:/home/geetha/pgms/hivedata
Copying file: file:/home/geetha/pgms/hivedata
Loading data to table default.page_view_from_hdfs
Table default.page_view_from_hdfs stats: [num_partitions: 0, num_files:
1, num_rows: 0, total_size: 176, raw_data_size: 0]
OK
```

```
show command:
-----
```

```
hive> show tables;
```

this will show all the tables that were created till now

=====

```
select cmd:
-----
```

```
select * from page_view_from_local;
```

```
10.31.05    user1 www.apache.com    6      12.23.28.37
12.41.24    user2 www.ubuntu.com    8      16.34.21.109
05.32.16    user3 www.amazon.com    34     18.46.38.35
09.31.47    user4 www.hadoop.com    78     28.52.89.102
Time taken: 0.48 seconds, Fetched: 4 row(s)
```

=====

external tables:

```
dfs -rmr /usr/hadoop-inst/hadoop-1.0.3/hivesample;
```

```
dfs -copyFromLocal /home/geetha/pgms/hivedata /usr/hadoop-inst/hadoop-1.0.3/hivedir;
```

```
CREATE EXTERNAL TABLE page_view_external(viewTime STRING, userid STRING,
page_url STRING,
pageNo INT,ip STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES
TERMINATED BY '\n'
LOCATION '/usr/hadoop-inst/hadoop-1.0.3/hivedir';
```

```
hive> select * form page_view_external;
```

OK

```
01.31.05    user5 www.apache.com    6      12.23.28.37
07.23.24    user6 www.pig.com12     16.34.21.109
05.16.16    user7 www.mapred.com    18     18.46.38.35
09.18.47    user8 www.hdfs.com      64     28.52.89.102
Time taken: 0.405 seconds, Fetched: 4 row(s)
```

when some of data is missing NULL values will be stored at that position

=====

//Select queries

```
SELECT * FROM page_view_internal_from_local;
```

```
SELECT * FROM page_view_internal_from_hdfs;
```

```
SELECT * FROM page_view_external;
```

=====

Creating partitions:

```
CREATE TABLE page_view_partitioned(viewTime STRING, userid STRING,
page_url STRING, pageNo
INT,ip STRING) PARTITIONED BY (dt STRING, country STRING) ROW FORMAT
DELIMITED FIELDS
TERMINATED BY ',' LINES TERMINATED BY '\n';
```

```
hive> show partitions page_view_partitioned;
```

o/p:

dt=2012-12-11/country=US

loading data:

//Insert into Partitioned Table from files. Actual Partition directories are created now

```
LOAD DATA LOCAL INPATH '/home/geetha/pgms/hivepage1' INTO TABLE
page_view_partitioned PARTITION(dt='2012-12-11', country='US');
```

```
LOAD DATA LOCAL INPATH '/home/geetha/pgms/hivepage2' INTO TABLE
page_view_partitioned PARTITION(dt='2012-12-11', country='INDIA');
```

```
LOAD DATA LOCAL INPATH '/home/geetha/pgms/hivepage3' INTO TABLE
page_view_partitioned PARTITION(dt='2012-12-12' , country='US');
```

```
LOAD DATA LOCAL INPATH '/home/geetha/pgms/hivepage4' INTO TABLE
page_view_partitioned PARTITION(dt='2012-12-12' , country='INDIA');
```

sample directory structure created would be as :

```
/user/hive/warehouse/page_view_partitioned/dt=2012-12-11/country=US
```

```
SELECT FROM PARTITION:
```

```
select * from page_view_partitioned where country='US';
```

```
select * from page_view_partitioned where country='US' and dt='2012-12-11';
```

=====

//Inserting Data from Hive Tables

//Inserting Data from Hive Tables to other Hive Tables

```
CREATE TABLE page_view_dest(viewTime STRING, userid STRING, page_url
STRING, pageNo INT,ip
STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY
'\n';
```

below commands used for further analysis with HIVE:

```
INSERT INTO TABLE page_view_dest SELECT * FROM page_view_from_local;
```

```
INSERT OVERWRITE TABLE page_view_dest SELECT * FROM page_view_from_local;
```

below commands used for further analysis may be with pig or M/R:

```
INSERT INTO DIRECTORY '/usr/hadoop-inst/hadoop-1.0.3/hivedir/hivedata'
SELECT * FROM page_view_from_local;
```

```
INSERT OVERWRITE DIRECTORY '/usr/hadoop-inst/hadoop-1.0.3/hivedir/hivedata'
SELECT * FROM page_view_from_local;
```

```
INSERT INTO LOCAL DIRECTORY '/home/geetha/pgms/hivedata' SELECT * FROM
page_view_from_local;
```

```
INSERT OVERWRITE LOCAL DIRECTORY '/home/geetha/pgms/hivedata' SELECT *
FROM page_view_from_local;
```

=====

creating buckets:

```
CREATE TABLE user(id INT, name STRING)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n';
```

```
LOAD DATA LOCAL INPATH '/home/geetha/pgms/users' INTO TABLE user;
```

```
hive> select * from user;
```

```
CREATE TABLE bucketed_users (id INT, name STRING)
CLUSTERED BY (id) INTO 4 BUCKETS;
```

check if bucketing option is enabled in your hive shell :

```
set hive.enforce.bucketing;
```

If false,run command : set hive.enforce.bucketing=true;

```
INSERT OVERWRITE TABLE bucketed_users
SELECT * FROM user;
```

o/p : 4 seperate files will be created.

sampling can be done using the command TABLESAMPLE

```
SELECT * FROM bucketed_users
TABLESAMPLE(BUCKET 1 OUT OF 4 ON id);
```

o/p :

0 Nat

4 Ann

=====

complex queries:

```
CREATE TABLE employees (
name STRING,
salary FLOAT,
subordinates ARRAY<STRING>,
deductions MAP<STRING, INT>,
address STRUCT<street:STRING, city:STRING, state:STRING, zip:INT>
)
```

```
ROW FORMAT DELIMITED
```

```
FIELDS TERMINATED BY ','
```

```
COLLECTION ITEMS TERMINATED BY ':'
```

```
MAP KEYS TERMINATED BY '->'
```

```
LINES TERMINATED BY '\n'  
STORED AS TEXTFILE;
```

```
LOAD DATA LOCAL INPATH '/home/geetha/pgms/hivecmplx' INTO TABLE  
employees;
```

```
SELECT name, salary FROM employees;
```

```
SELECT name, subordinates FROM employees;
```

```
SELECT name, deductions FROM employees;
```

```
SELECT name, address FROM employees;
```

```
SELECT name, subordinates[0] FROM employees;
```

```
SELECT name, deductions["State Taxes"] FROM employees;
```

```
SELECT name, address.city FROM employees;
```