hive> create table ManagedTable\_Examp(id int,name string,sal int,address string) row format delimited fields terminated by '\t' lines terminated by '\n' stored as TEXTFILE;

OK

Time taken: 1.092 seconds

hive> DESC MANAGEDTABLE\_EXAMP;

OK

id int

name string

sal int

address string

Time taken: 0.454 seconds

=========================================================

DATA INSERTION TRACE

---------------------

hive> load data local inpath '/home/Gopal/HIVE-Related/input4Manged.txt' into table managedtable\_examp;

Copying data from file:/home/Gopal/HIVE-Related/input4Manged.txt

Copying file: file:/home/Gopal/HIVE-Related/input4Manged.txt

Loading data to table default.managedtable\_examp

OK

Time taken: 1.267 seconds

hive> select \* from managedtable\_examp;

OK

10 Gopal 10000 Hyd

11 Vijay 21000 Hyd

13 prakash 23000 Bag

14 Raghu 34234 Pune

15 Maha 12345 Chennai

16 Kaasi 45000 Bangalore

17 Viswanath 46789 Pune

18 Radha 78000 Gujarath

Time taken: 0.544 seconds

hive>

===========================================================

DEFAULT LOCATION OF THE MANAGED TABLES TRACE

--------------------------------------------

root@ubuntu:/home/Gopal/HIVE-Related# hadoop fs -ls /user/hive/warehouse;

Found 3 items

drwxr-xr-x - root supergroup 0 2012-12-03 06:09 /user/hive/warehouse/firsttable

drwxr-xr-x - root supergroup 0 2012-12-03 23:43 /user/hive/warehouse/managedtable\_examp

drwxr-xr-x - root supergroup 0 2012-12-03 06:00 /user/hive/warehouse/sample

root@ubuntu:/home/Gopal/HIVE-Related#

============================================================

EXTERNAL TABLES RELATED

----------------------

hive> create external table externalTable\_examp(id int,name string,sal int,address string) row format delimited fields terminated by '\t' lines terminated by '\n' stored as textfile location '/user/Gopal/ExternalLocal/';

OK

Time taken: 17.391 seconds

hive> desc externalTable\_examp;

OK

id int

name string

sal int

address string

Time taken: 10.68 seconds

hive> load data local inpath '/home/Gopal/HIVE-Related/input4External.txt' into table externalTable\_examp;

Copying data from file:/home/Gopal/HIVE-Related/input4External.txt

Copying file: file:/home/Gopal/HIVE-Related/input4External.txt

Loading data to table default.externaltable\_examp

OK

Time taken: 15.484 seconds

hive> select \* from externaltable\_examp;

OK

10 Gopal 10000 Hyd

11 Vijay 21000 Hyd

13 prakash 23000 Bag

14 Raghu 34234 Pune

15 Maha 12345 Chennai

16 Kaasi 45000 Bangalore

17 Viswanath 46789 Pune

18 Radha 78000 Gujarath

Time taken: 2.927 seconds

hive>

===============================================================================

EXTERNAL TABLE DATA INSERTION FROM HDFS MODE:

-------------------------------------------

hive> create external table Extern(id int,name string,sal int,address string) row format delimited fields terminated by '\t' lines terminated by '\n' stored as textfile location '/user/Gopal/Extern/';

OK

Time taken: 2.272 seconds

hive> desc extern;

OK

id int

name string

sal int

address string

Time taken: 7.263 seconds

hive> create external table Extern(id int,name string,sal int,address string) row format delimited fields terminated by '\t' lines terminated by '\n' stored as textfi load data inpath '/user/Gopal/InputPath4External/input4External.txt' into table extern;

FAILED: Error in semantic analysis: Line 1:17 Invalid path '/user/Gopal/InputPath4External/input4External.txt': No files matching path hdfs://localhost:8020/user/Gopal/InputPath4External/input4External.txt

hive> load data inpath '/user/Gopal/InputPath4External/input4External.txt\_copy\_1' into table extern;

Loading data to table default.extern

OK

Time taken: 7.795 seconds

hive> select \* from extern;

OK

10 Gopal 10000 Hyd

11 Vijay 21000 Hyd

13 prakash 23000 Bag

14 Raghu 34234 Pune

15 Maha 12345 Chennai

16 Kaasi 45000 Bangalore

17 Viswanath 46789 Pune

18 Radha 78000 Gujarath

Time taken: 8.868 seconds

hive>

====================================================================================

DIFFRENCE BETWEEN 'MANAGED TABLE VS EXTERNAL TABLE'

---------------------------------------------------

"When an external table is dropped , only the table get dropped but the data associated with the extenal table still persists.See the below example, even after dropping the external table called "extern", the data in its metastore(external location '/user/Gopal/Extern') still exists."

hive> drop table extern;

OK

Time taken: 1.685 seconds

hive> select \* from extern;

FAILED: Error in semantic analysis: Line 1:14 Table not found extern

hive>

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

root@ubuntu:/home/Gopal/HIVE-Related# hadoop fs -ls /user/Gopal/

Found 19 items

drwxr-xr-x - root supergroup 0 2012-12-04 06:33 /user/Gopal/Extern

drwxr-xr-x - root supergroup 0 2012-12-04 06:19 /user/Gopal/ExternalLocal

-rw-r--r-- 1 root supergroup 140 2012-11-18 04:02 /user/Gopal/Input.txt

drwxr-xr-x - root supergroup 0 2012-12-04 06:33 /user/Gopal/InputPath4External

drwxrwxr-x - root supergroup 0 2012-11-18 00:24 /user/Gopal/Inputdir

drwxr-xr-x - root supergroup 0 2012-11-18 04:09 /user/Gopal/MROutPut

drwxr-xr-x - root supergroup 0 2012-11-18 04:11 /user/Gopal/MapRed

drwxr-xr-x - root supergroup 0 2012-11-19 04:25 /user/Gopal/MapRedOut

drwxr-xr-x - root supergroup 0 2012-11-18 04:12 /user/Gopal/MapRedOutput

drwxr-xr-x - root supergroup 0 2012-11-18 02:05 /user/Gopal/OutputPig

drwxr-xr-x - root supergroup 0 2012-11-28 09:33 /user/Gopal/Sqoop-Out

drwxr-xr-x - root supergroup 0 2012-11-28 09:36 /user/Gopal/Sqoop-Output1

-rw-r--r-- 1 root supergroup 1953 2012-11-18 02:00 /user/Gopal/passwd

drwxr-xr-x - root supergroup 0 2012-12-02 03:50 /user/Gopal/sundayTest

drwxr-xr-x - root supergroup 0 2012-12-02 05:47 /user/Gopal/sundayTestforFORMAT

drwxr-xr-x - root supergroup 0 2012-12-02 04:28 /user/Gopal/sundayTestforJOIN

drwxr-xr-x - root supergroup 0 2012-12-02 04:34 /user/Gopal/sundayTestforLEFTOUTERJOIN

drwxr-xr-x - root supergroup 0 2012-12-02 04:21 /user/Gopal/sundayTestforSEL

drwxr-xr-x - root supergroup 0 2012-12-02 06:09 /user/Gopal/sundayTestforSIMPLEWHERECON

root@ubuntu:/home/Gopal/HIVE-Related# hadoop fs -ls /user/Gopal/Extern

Found 1 items

-rw-r--r-- 1 root supergroup 174 2012-12-04 06:23 /user/Gopal/Extern/input4External.txt\_copy\_1

root@ubuntu:/home/Gopal/HIVE-Related# hadoop fs -cat /user/Gopal/Extern/input4External.txt\_copy\_1

10 Gopal 10000 Hyd

11 Vijay 21000 Hyd

13 prakash 23000 Bag

14 Raghu 34234 Pune

15 Maha 12345 Chennai

16 Kaasi 45000 Bangalore

17 Viswanath 46789 Pune

18 Radha 78000 Gujarath

root@ubuntu:/home/Gopal/HIVE-Related#

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

TRACE FOR Managed Table dropping

-------------------------------

hive> desc sample;

OK

name string

Time taken: 6.55 seconds

hive>

Before dropping the managed table called "sample"

root@ubuntu:/home/Gopal/HIVE-Related# hadoop fs -ls /user/hive/warehouse

Found 3 items

drwxr-xr-x - root supergroup 0 2012-12-04 00:00 /user/hive/warehouse/firsttable

drwxr-xr-x - root supergroup 0 2012-12-03 23:43 /user/hive/warehouse/managedtable\_examp

drwxr-xr-x - root supergroup 0 2012-12-03 06:00 /user/hive/warehouse/sample

root@ubuntu:/home/Gopal/HIVE-Related#

After dropping the managed table called "sample"

hive> drop table sample

> ;

OK

Time taken: 4.74 seconds

hive>

root@ubuntu:/home/Gopal/HIVE-Related# hadoop fs -ls /user/hive/warehouse

Found 2 items

drwxr-xr-x - root supergroup 0 2012-12-04 00:00 /user/hive/warehouse/firsttable

drwxr-xr-x - root supergroup 0 2012-12-03 23:43 /user/hive/warehouse/managedtable\_examp

root@ubuntu:/home/Gopal/HIVE-Related#

NOTE: /user/hive/warehouse/sample is not there now after dropping the table

===============================================================

PARTITIONS RELATED:

---------------------

hive> drop table logdata; create table logData(logid int,logerror string) PARTITIONED BY (timeStampDT string,country string) row format delimited fields terminated by '|' lines terminated by '\n' stored as textfile;

OK

Time taken: 1.946 seconds

OK

Time taken: 0.1 seconds

hive>

create table logData(logid int,logerror string) PARTITIONED BY (timeStampDT string,country string) row format delimited fields terminated by '|' lines terminated by '\n' stored as textfile;

hive> DESC LOGDATA;

OK

logid int

logerror string

timestampdt string

country string

Time taken: 0.147 seconds

LOADING THE DATA IN THE PARTITIONED TABLE

------------------------------------------

hive> load data local inpath '/home/Gopal/HIVE-Related/input4Parti.txt' into table logdata PARTITION(timeStampDT='2012-12-05' , country='INDIA');

Copying data from file:/home/Gopal/HIVE-Related/input4Parti.txt

Copying file: file:/home/Gopal/HIVE-Related/input4Parti.txt

Loading data to table default.logdata partition (timestampdt=2012-12-05, country=INDIA)

OK

Time taken: 0.466 seconds

hive>

root@ubuntu:/home/Gopal/HIVE-Related# cat input4Parti.txt

1001|Network Related

1002|Router Related

1003|Switch Related

1004|Hub Related

1005|Machine IP Related

root@ubuntu:/home/Gopal/HIVE-Related#

hive> SELECT \* FROM LOGDATA;

OK

1001 Network Related 2012-12-05 INDIA

1002 Router Related 2012-12-05 INDIA

1003 Switch Related 2012-12-05 INDIA

1004 Hub Related 2012-12-05 INDIA

1005 Machine IP Related 2012-12-05 INDIA

Time taken: 0.235 seconds

hive>

hive> load data local inpath '/home/Gopal/HIVE-Related/input4Parti.txt' into table logdata PARTITION(timeStampDT='2012-12-04' , country='USA');

Copying data from file:/home/Gopal/HIVE-Related/input4Parti.txt

Copying file: file:/home/Gopal/HIVE-Related/input4Parti.txt

Loading data to table default.logdata partition (timestampdt=2012-12-04, country=USA)

OK

Time taken: 0.388 seconds

hive> SELECT \* FROM LOGDATA;

OK

1001 Network Related 2012-12-04 USA

1002 Router Related 2012-12-04 USA

1003 Switch Related 2012-12-04 USA

1004 Hub Related 2012-12-04 USA

1005 Machine IP Related 2012-12-04 USA

1001 Network Related 2012-12-05 INDIA

1002 Router Related 2012-12-05 INDIA

1003 Switch Related 2012-12-05 INDIA

1004 Hub Related 2012-12-05 INDIA

1005 Machine IP Related 2012-12-05 INDIA

Time taken: 0.35 seconds

hive>

hive> load data local inpath '/home/Gopal/HIVE-Related/input4Parti.txt' into table logdata PARTITION(timeStampDT='2012-12-06' , country='CHINA');

Copying data from file:/home/Gopal/HIVE-Related/input4Parti.txt

Copying file: file:/home/Gopal/HIVE-Related/input4Parti.txt

Loading data to table default.logdata partition (timestampdt=2012-12-06, country=CHINA)

OK

Time taken: 0.494 seconds

hive> SELECT \* FROM LOGDATA;

OK

1001 Network Related 2012-12-04 USA

1002 Router Related 2012-12-04 USA

1003 Switch Related 2012-12-04 USA

1004 Hub Related 2012-12-04 USA

1005 Machine IP Related 2012-12-04 USA

1001 Network Related 2012-12-05 INDIA

1002 Router Related 2012-12-05 INDIA

1003 Switch Related 2012-12-05 INDIA

1004 Hub Related 2012-12-05 INDIA

1005 Machine IP Related 2012-12-05 INDIA

1001 Network Related 2012-12-06 CHINA

1002 Router Related 2012-12-06 CHINA

1003 Switch Related 2012-12-06 CHINA

1004 Hub Related 2012-12-06 CHINA

1005 Machine IP Related 2012-12-06 CHINA

Time taken: 0.393 seconds

hive>

TO SHOW THE PARTITIONS INFORMATION IN THE TABLE LOGDATA

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

hive> show partitions logdata;

OK

timestampdt=2012-12-04/country=USA

timestampdt=2012-12-05/country=INDIA

timestampdt=2012-12-06/country=CHINA

Time taken: 0.098 seconds

hive>

=====================================

ALTER TABLE

-----------

hive> show tables;

OK

bucketed\_users

externaltable\_examp

firsttable

logdata

managedtable\_examp

users

Time taken: 78.99 seconds

hive> alter table firsttable rename to secondtable;

OK

Time taken: 1.869 seconds

hive> desc firsttable;

FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask

hive> desc secondtable;

OK

id int

name string

sal int

Time taken: 0.135 seconds

hive> altertable secondtable add columns(address string,state string);

FAILED: Parse Error: line 1:0 cannot recognize input near 'altertable' 'secondtable' 'add'

hive> alter table secondtable add columns(address string,state string);

OK

Time taken: 0.179 seconds

hive> desc secondtable;

OK

id int

name string

sal int

address string

state string

Time taken: 0.142 seconds

hive>

==========================================

CTAS(CREATE TABLE AS SELECT)

----------------------------

hive> desc managedtable\_examp;

OK

id int

name string

sal int

address string

Time taken: 0.536 seconds

hive> CREATE TABLE MERGE\_TAB AS SELECT ID,NAME,SAL,ADDRESS FROM MANAGEDTABLE\_EXAMP;

Total MapReduce jobs = 2

Launching Job 1 out of 2

Number of reduce tasks is set to 0 since there's no reduce operator

Starting Job = job\_201212050653\_0001, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job\_201212050653\_0001

Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job\_201212050653\_0001

2012-12-06 06:34:59,492 Stage-1 map = 0%, reduce = 0%

2012-12-06 06:35:08,144 Stage-1 map = 100%, reduce = 0%

2012-12-06 06:35:11,214 Stage-1 map = 100%, reduce = 100%

Ended Job = job\_201212050653\_0001

Ended Job = 968272562, job is filtered out (removed at runtime).

Moving data to: hdfs://localhost:8020/tmp/hive-root/hive\_2012-12-06\_06-33-18\_728\_5071704976780310772/-ext-10001

Moving data to: hdfs://localhost:8020/user/hive/warehouse/merge\_tab

8 Rows loaded to hdfs://localhost:8020/tmp/hive-root/hive\_2012-12-06\_06-33-18\_728\_5071704976780310772/-ext-10000

OK

Time taken: 113.462 seconds

hive> SELECT \* FROM MERGE\_TAB;

OK

10 Gopal 10000 Hyd

11 Vijay 21000 Hyd

13 prakash 23000 Bag

14 Raghu 34234 Pune

15 Maha 12345 Chennai

16 Kaasi 45000 Bangalore

17 Viswanath 46789 Pune

18 Radha 78000 Gujarath

Time taken: 0.296 seconds

hive>

==================================================

SECOND EXAMPLE:ON CTAS WITH JOIN CONDITION

------------------------------------------

hive> create table Merge\_of\_Two as select id,name,sal from managedtable\_examp m JOIN externaltable\_examp e ON (m.id=e.id);

FAILED: Error in semantic analysis: Column id Found in more than One Tables/Subqueries

hive> create table Merge\_of\_Two as select m.id,m.name,e.sal from managedtable\_examp m JOIN externaltable\_examp e ON (m.id=e.id);

Total MapReduce jobs = 1

Launching Job 1 out of 1

Number of reduce tasks not specified. Estimated from input data size: 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 mapred.reduce.tasks=<number>

Starting Job = job\_201212050653\_0002, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job\_201212050653\_0002

Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job\_201212050653\_0002

2012-12-06 06:39:31,134 Stage-1 map = 0%, reduce = 0%

2012-12-06 06:40:31,727 Stage-1 map = 0%, reduce = 0%

2012-12-06 06:41:06,085 Stage-1 map = 50%, reduce = 0%

2012-12-06 06:41:09,195 Stage-1 map = 100%, reduce = 0%

2012-12-06 06:41:29,297 Stage-1 map = 100%, reduce = 100%

Ended Job = job\_201212050653\_0002

Moving data to: hdfs://localhost:8020/user/hive/warehouse/merge\_of\_two

8 Rows loaded to hdfs://localhost:8020/tmp/hive-root/hive\_2012-12-06\_06-39-16\_885\_7194275506457069858/-ext-10000

OK

Time taken: 137.224 seconds

hive> select \* from merge\_of\_two;

OK

10 Gopal 10000

11 Vijay 21000

13 prakash 23000

14 Raghu 34234

15 Maha 12345

16 Kaasi 45000

17 Viswanath 46789

18 Radha 78000

Time taken: 1.758 seconds

hive>

===============================================================

UDF Example:

=============

root@ubuntu:/home/Gopal/HIVE-Related# hive --auxpath /home/Gopal/HIVE-Related/strip.jar

Hive history file=/tmp/root/hive\_job\_log\_root\_201212060927\_736655719.txt

hive> add jar /home/Gopal/HIVE-Related/strip.jar;

Added /home/Gopal/HIVE-Related/strip.jar to class path

Added resource: /home/Gopal/HIVE-Related/strip.jar

hive> create temporary function strip as 'Strip';

OK

Time taken: 0.993 seconds

hive> select \* from users;

OK

100 Gopal

101 Prakash

102 Madhu

103 Ravi

104 Mohan

105 Mounika

Time taken: 41.893 seconds

hive> desc users;

OK

id int

name string

Time taken: 0.22 seconds

hive> select strip(' Gopal ') from users;

Total MapReduce 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\_201212050653\_0003, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job\_201212050653\_0003

Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job\_201212050653\_0003

2012-12-06 09:35:50,672 Stage-1 map = 0%, reduce = 0%

2012-12-06 09:36:00,100 Stage-1 map = 100%, reduce = 0%

2012-12-06 09:36:03,323 Stage-1 map = 100%, reduce = 100%

Ended Job = job\_201212050653\_0003

OK

Gopal

Gopal

Gopal

Gopal

Gopal

Gopal

Time taken: 42.184 seconds

hive>

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

hive> select strip('Gopal','al') from users;

Total MapReduce 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\_201212050653\_0005, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job\_201212050653\_0005

Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job\_201212050653\_0005

2012-12-06 09:39:52,899 Stage-1 map = 0%, reduce = 0%

2012-12-06 09:39:56,983 Stage-1 map = 100%, reduce = 0%

2012-12-06 09:40:00,017 Stage-1 map = 100%, reduce = 100%

Ended Job = job\_201212050653\_0005

OK

Gop

Gop

Gop

Gop

Gop

Gop

Time taken: 22.644 seconds

hive>

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

hive> select strip('Gopal','Go') from users;

Total MapReduce 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\_201212050653\_0006, Tracking URL = http://localhost:50030/jobdetails.jsp?jobid=job\_201212050653\_0006

Kill Command = /usr/lib/hadoop/bin/hadoop job -Dmapred.job.tracker=localhost:8021 -kill job\_201212050653\_0006

2012-12-06 09:41:12,175 Stage-1 map = 0%, reduce = 0%

2012-12-06 09:41:16,237 Stage-1 map = 100%, reduce = 0%

2012-12-06 09:41:19,266 Stage-1 map = 100%, reduce = 100%

Ended Job = job\_201212050653\_0006

OK

pal

pal

pal

pal

pal

pal

Time taken: 14.392 seconds