# Display Schema Name

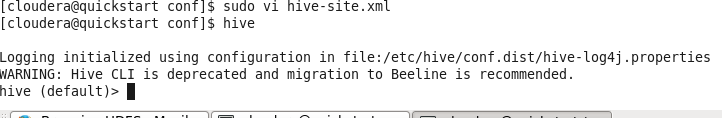
To display the schema name in hive prompt, you can use

set hive.cli.print.current.db=true; but this is applicable for current session as you exit this will lost

To make it permanent we need to edit /etc/hive/conf/hive-site.xml file

Just add below property

|  |
| --- |
| <property>  <name>hive.cli.print.current.db</name>  <value>true</value>  </property> |



You can create .hiverc file in /etc/hive/conf folder and add property

|  |
| --- |
| set hive.cli.print.current.db=true |

Priority => command line set > .hiverc >hive-site.xml

# Print Header

To check the create script of any table

|  |
| --- |
| show create table <table\_name>; |

To print the header value in result ( column name in case of select statement) add below property in hive-site.xml

|  |
| --- |
| <property>  <name>hive.cli.print.header</name>  <value>true</value>  </property> |

# Drop all tables

|  |
| --- |
| hive -e 'use testdb;show tables' | xargs -I '{}' hive -e 'use testdb;drop table {}' |

# Drop database

|  |
| --- |
| DROP DATABASE IF EXISTS testdb CASCADE; |

# Create parquet table from text table

|  |
| --- |
| CREATE TABLE Carrier\_Claims\_Parquet LIKE Carrier\_Claims STORED AS PARQUET; |

# Insert data in partition table from another table

## Insert using where clause (static partition)

Here we need to define the exact partition value and also need to define all columns in select clause except partition column that we need to define in where criteria

|  |
| --- |
| insert into Beneficiary\_Summary\_PARQUET PARTITION (year=2008) select DESYNPUF\_ID,BENE\_BIRTH\_DT,BENE\_DEATH\_DT,BENE\_SEX\_IDENT\_CD,BENE\_RACE\_CD,BENE\_ESRD\_IND,SP\_STATE\_CODE,BENE\_COUNTY\_CD,BENE\_HI\_CVRAGE\_TOT\_MONS,BENE\_SMI\_CVRAGE\_TOT\_MONS,BENE\_HMO\_CVRAGE\_TOT\_MONS,PLAN\_CVRG\_MOS\_NUM,SP\_ALZHDMTA,SP\_CHF,SP\_CHRNKIDN,SP\_CNCR,SP\_COPD,SP\_DEPRESSN,SP\_DIABETES,SP\_ISCHMCHT,SP\_OSTEOPRS,SP\_RA\_OA,SP\_STRKETIA,MEDREIMB\_IP,BENRES\_IP,PPPYMT\_IP,MEDREIMB\_OP,BENRES\_OP,PPPYMT\_OP,MEDREIMB\_CAR,BENRES\_CAR,PPPYMT\_CAR from Beneficiary\_Summary where year=2008; |

## Insert all partition column without where clause (dynamic partition)

Instead of define each partition in multiple insert query we can make a single insert query without defining the exact value, see below query

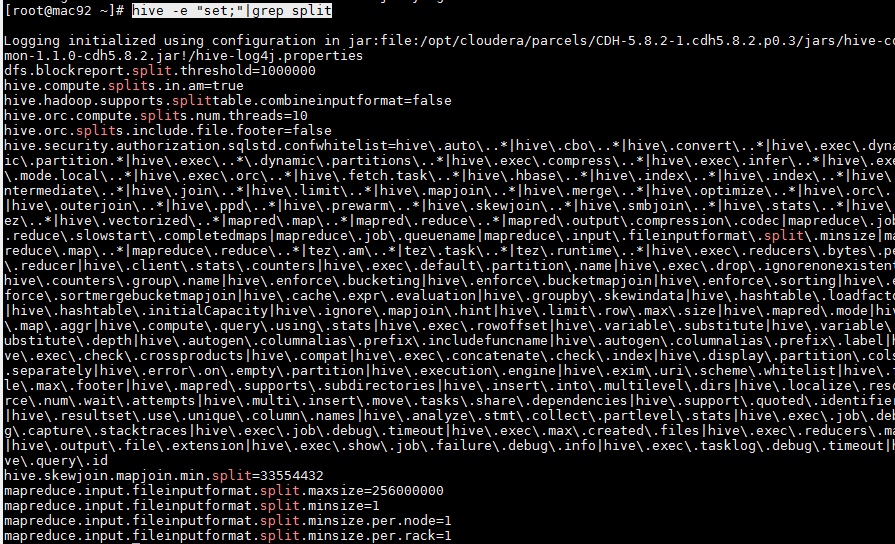
set hive.exec.dynamic.partition.mode=nonstrict;

|  |
| --- |
| insert into Beneficiary\_Summary\_PARQUET PARTITION (year) select \* from Beneficiary\_Summary ; |

# Check any configuration parameter in hive

From linux file system run below command to check the configuration parameter which contain split

|  |
| --- |
| hive -e "set;"|grep split |



# Connect hive with beeline

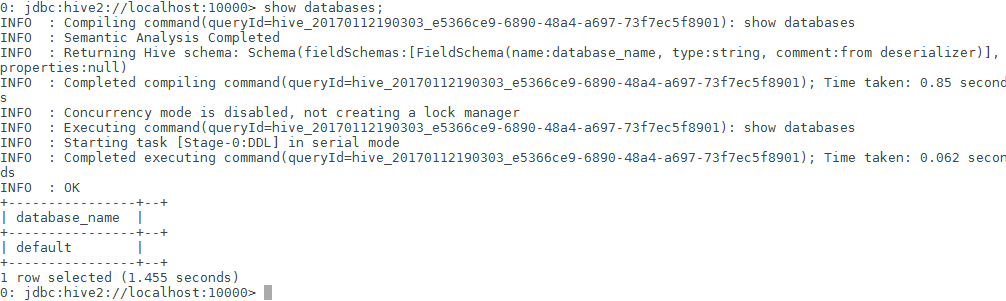
Just type beeline on terminal and click enter

!connect jdbc:hive2://localhost:10000

|  |
| --- |
| beeline> !connect jdbc:hive2://localhost:10000  Connecting to jdbc:hive2://localhost:10000  Enter username for jdbc:hive2://localhost:10000: db\_hive  Enter password for jdbc:hive2://localhost:10000: \*\*\*\*\*\*\*\*  Connected to: Apache Hive (version 1.1.0-cdh5.9.0)  Driver: Hive JDBC (version 1.1.0-cdh5.9.0)  Transaction isolation: TRANSACTION\_REPEATABLE\_READ  0: jdbc:hive2://localhost:10000> |

Or to connect without user name and password use below command

|  |
| --- |
| beeline -u jdbc:hive2://mac127:10000 |



!q to quit beeline

# Permission denied at time of creating External table

|  |
| --- |
| hive (esplus)> CREATE EXTERNAL TABLE ApplicationSwitchData  > (  > json string  > )  > LOCATION '/esplus/tables/applicationswitchdata';  FAILED: Execution Error, return code 1 from org.apache.hadoop.hive.ql.exec.DDLTask. MetaException(message:java.security.AccessControlException: Permission denied: user=root, access=WRITE, inode="/esplus/tables/applicationswitchdata":hdfs:hdfs:drwxr-xr-x  at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.check(FSPermissionChecker.java:319)  at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkPermission(FSPermissionChecker.java:219)  at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkPermission(FSPermissionChecker.java:190)  at org.apache.hadoop.hdfs.server.namenode.FSDirectory.checkPermission(FSDirectory.java:1771)  at org.apache.hadoop.hdfs.server.namenode.FSDirectory.checkPermission(FSDirectory.java:1755)  at org.apache.hadoop.hdfs.server.namenode.FSDirectory.checkPathAccess(FSDirectory.java:1729)  at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkAccess(FSNamesystem.java:8348)  at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.checkAccess(NameNodeRpcServer.java:1978)  at org.apache.hadoop.hdfs.protocolPB.ClientNamenodeProtocolServerSideTranslatorPB.checkAccess(ClientNamenodeProtocolServerSideTranslatorPB.java:1443)  at org.apache.hadoop.hdfs.protocol.proto.ClientNamenodeProtocolProtos$ClientNamenodeProtocol$2.callBlockingMethod(ClientNamenodeProtocolProtos.java)  at org.apache.hadoop.ipc.ProtobufRpcEngine$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine.java:616)  at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:969)  at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2151)  at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2147)  at java.security.AccessController.doPrivileged(Native Method)  at javax.security.auth.Subject.doAs(Subject.java:422)  at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1657)  at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2145) |

Use below command to log into hive shell and create table again

|  |
| --- |
| sudo -u hdfs hive |

# File permission issue (No files matching path)

|  |
| --- |
| LOAD DATA LOCAL INPATH '/root/Esplus/newdatasets/13\_07\_2016/4\_1-7-2016\_1-44-1\_06096.txt' INTO TABLE esplus.ApplicationSwitchData;  FAILED: SemanticException Line 1:23 Invalid path ''/root/Esplus/newdatasets/13\_07\_2016/4\_1-7-2016\_1-44-1\_06096.txt'': No files matching path file:/root/Esplus/newdatasets/13\_07\_2016/4\_1-7-2016\_1-44-1\_06096.txt |

The file does exist at the specified location in the Linux file system and all users have read permission.

If its in the root folder only root can access it. Hive error messages can be pretty generic here and not distinguish between access rights and files exists. If you run beeline the read will be executed by the hive server and that one is running under the hive user so only he could access the data.

First give read write permission to folder

|  |
| --- |
| chmod -R 755 newdatasets/ |

Move Esplus folder to root (/)

|  |
| --- |
| mv Esplus / |

Change owner from root to hdfs

chown -R hdfs:hdfs Esplus

|  |
| --- |
| -rw-r--r-- 1 root root 12785 Dec 12 02:19 emp.java  drwxr-xr-x 5 hdfs hdfs 109 Jan 3 00:52 Esplus |

# Hdfs command in hive

|  |
| --- |
| Hive>dfs -ls |

# Unix command in hive

|  |
| --- |
| Hive> **!ls** |

# Move data from json serde table to normal table

Json serde Table use

|  |
| --- |
| CREATE EXTERNAL TABLE JsonTable\_raw (  data array<struct<  start\_date:string,  end\_date:string,  measures:struct<  Visitors:int,  Singlepagevisits:int  ) ROW FORMAT SERDE 'org.openx.data.jsonserde.JsonSerDe' |

Non serde table

|  |
| --- |
| CREATE EXTERNAL TABLE JsonTable (  start\_date string,  end\_date string,  Visitors int,  Singlepagevisits int  ) |

Insert query

|  |
| --- |
| INSERT INTO TABLE JsonTable  select cols.start\_date,cols.end\_date,cols.measures.Visitors,  cols.measures.Singlepagevisits  FROM JsonTable\_raw jt lateral view explode(jt.data) collection as cols; |

# Hive - External Table with Partitions

When the data files become huge (number and size) then we might need to use Partition to improve the efficiency of data processing.

|  |
| --- |
| CREATE TABLE user (  userId BIGINT,  type INT,  level TINYINT,  )  COMMENT 'User Infomation'  PARTITIONED BY (date String) |

data String is moved to PARTITIONED BY, when we need to load data into hive, partition must be assigned.

|  |
| --- |
| LOAD INPATH '/user/chris/data/testdata' OVERWRITE INTO TABLE user PARTITION (date='2012-02-22') |

After data is loaded, we can see a new folder named date=2010-02-22 is created inside /user/chris/warehouse/user/

So, how can we do it using external table?

Same as before, first declare the external table user, and assign the location.

|  |
| --- |
| CREATE EXTERNAL TABLE user (  userId BIGINT,  type INT,  level TINYINT,  date String  )  COMMENT 'User Infomation'  PARTITIONED BY (date String)  LOCATION '/user/chris/datastore/user/'; |

Then, create the folder **date=2010-02-22** inside **/user/chris/datastore/user/**

At last, put the data files of date 2010-02-22 into the folder, done.

But,

When we executes select \* from user; nothing appears.

Because when external table is declared, default table path is changed to specified location in hive metadata which contains in metastore, but about partition, nothing is changed, so, we must manually add those metadata.

|  |
| --- |
| ALTER TABLE user ADD PARTITION(date='2010-02-22'); |

Every time a new data=... folder (partition) is created, we must manually alter the table to add partition information.