# Display Schema Name

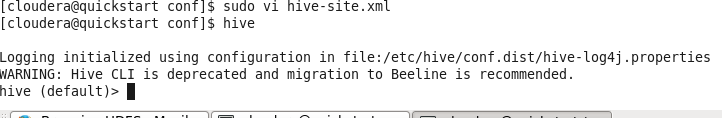
To display the schema name in hive prompt, you can use

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;  set hive.exec.dynamic.partition = true; |

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

Or

|  |
| --- |
| From Beneficiary\_Summary insert into Beneficiary\_Summary\_PARQUET PARTITION (year) |

# Insert into local directory from hive

We can load data into local file system directory from hive table

Syntax

|  |
| --- |
| Standard syntax:  INSERT OVERWRITE [LOCAL] DIRECTORY directory1  [ROW FORMAT row\_format] [STORED AS file\_format] (Note: Only available starting with Hive 0.11.0)  SELECT ... FROM ...    Hive extension (multiple inserts):  FROM from\_statement  INSERT OVERWRITE [LOCAL] DIRECTORY directory1 select\_statement1  [INSERT OVERWRITE [LOCAL] DIRECTORY directory2 select\_statement2] ...      row\_format  : DELIMITED [FIELDS TERMINATED BY char [ESCAPED BY char]] [COLLECTION ITEMS TERMINATED BY char]  [MAP KEYS TERMINATED BY char] [LINES TERMINATED BY char]  [NULL DEFINED AS char] (Note: Only available starting with Hive 0.13) |

Example:

|  |
| --- |
| hive> INSERT OVERWRITE LOCAL DIRECTORY '/root/tempdir' select \* from temptableapplication limit 10;  Query ID = root\_20170222090707\_5ed481dc-4242-4edc-ab3b-4eaac2fa5313  Total jobs = 1  Launching Job 1 out of 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\_1484894548299\_0047, Tracking URL = http://mac55:8088/proxy/application\_1484894548299\_0047/  Kill Command = /opt/cloudera/parcels/CDH-5.8.2-1.cdh5.8.2.p0.3/lib/hadoop/bin/hadoop job -kill job\_1484894548299\_0047  Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 1  2017-02-22 09:08:07,490 Stage-1 map = 0%, reduce = 0%  2017-02-22 09:08:39,424 Stage-1 map = 33%, reduce = 0%, Cumulative CPU 29.43 sec  2017-02-22 09:08:54,757 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 38.28 sec  2017-02-22 09:09:54,949 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 50.15 sec  2017-02-22 09:09:55,967 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 50.5 sec  2017-02-22 09:10:01,065 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 51.42 sec  MapReduce Total cumulative CPU time: 51 seconds 420 msec  Ended Job = job\_1484894548299\_0047  Copying data to local directory /root/tempdir  MapReduce Jobs Launched:  Stage-Stage-1: Map: 3 Reduce: 1 Cumulative CPU: 51.42 sec HDFS Read: 132172178 HDFS Write: 4500 SUCCESS  Total MapReduce CPU Time Spent: 51 seconds 420 msec  OK  Time taken: 132.949 seconds |

Check if file is loaded into local direcotry

|  |
| --- |
| [root@mac55 ~]# cd /root/tempdir/  [root@mac55 tempdir]# ls -l  total 8  -rw-r--r-- 1 root root 4500 Feb 22 09:10 000000\_0  [root@mac55 tempdir]# cat 000000\_0  61006D00690074007000610077004300540032002D004100500050002D0039003000390031003000300031004100410030003200370032003900360042003100370032002E00320037002E0036002E00310035003500430059004200410047004500f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_5mstsc.exeC:\Windows\system32\mstsc.exeMicrosoft® Windows® Operating SystemMicrosoft CorporationRemote Desktop ConnectionRemote Desktop Connection0.02016-05-25 18:01:53.8482016-05-25 18:00:002016-05-25  61006D00690074007000610077004300540032002D004100500050002D0039003000390031003000300031004100410030003200370032003900360042003100370032002E00320037002E0036002E00310035003500430059004200410047004500f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_2Explorer.EXEC:\Windows\Explorer.EXEMicrosoft® Windows® Operating SystemMicrosoft CorporationWindows ExplorerSearch Pane0.02016-05-25 18:01:43.6132016-05-25 18:00:002016-05-25  61006D00690074007000610077004300540032002D004100500050002D0039003000390031003000300031004100410030003200370032003900360042003100370032002E00320037002E0036002E00310035003500430059004200410047004500f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_19mstsc.exeC:\Windows\system32\mstsc.exeMicrosoft® Windows® Operating SystemMicrosoft CorporationRemote Desktop ConnectionRemote Desktop Connection0.02016-05-25 18:05:39.9152016-05-25 18:00:002016-05-25  61006D00690074007000610077004300540032002D004100500050002D0039003000390031003000300031004100410030003200370032003900360042003100370032002E00320037002E0036002E00310035003500430059004200410047004500f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_16mstsc.exeC:\Windows\system32\mstsc.exeMicrosoft® Windows® Operating SystemMicrosoft CorporationRemote Desktop ConnectionRemote Desktop Connection0.02016-05-25 18:03:34.3032016-05-25 18:00:002016-05-25  61006D00690074007000610077004300540032002D004100500050002D0039003000390031003000300031004100410030003200370032003900360042003100370032002E00320037002E0036002E00310035003500430059004200410047004500f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_13mstsc.exeC:\Windows\system32\mstsc.exeMicrosoft® Windows® Operating SystemMicrosoft CorporationRemote Desktop ConnectionRemote Desktop Connection0.02016-05-25 18:03:05.5372016-05-25 18:00:002016-05-25  61006D00690074007000610077004300540032002D004100500050002D0039003000390031003000300031004100410030003200370032003900360042003100370032002E00320037002E0036002E00310035003500430059004200410047004500f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_10mstsc.exeC:\Windows\system32\mstsc.exeMicrosoft® Windows® Operating SystemMicrosoft CorporationRemote Desktop ConnectionRemote Desktop Connection0.02016-05-25 18:02:47.3962016-05-25 18:00:002016-05-25  61006D00690074007000610077004300540032002D004100500050002D0039003000390031003000300031004100410030003200370032003900360042003100370032002E00320037002E0036002E00310035003500430059004200410047004500f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_7mstsc.exeC:\Windows\system32\mstsc.exeMicrosoft® Windows® Operating SystemMicrosoft CorporationRemote Desktop ConnectionRemote Desktop Connection0.02016-05-25 18:02:19.4262016-05-25 18:00:002016-05-25  61006D00690074007000610077004300540032002D004100500050002D0039003000390031003000300031004100410030003200370032003900360042003100370032002E00320037002E0036002E00310035003500430059004200410047004500f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_4mstsc.exeC:\Windows\system32\mstsc.exeMicrosoft® Windows® Operating SystemMicrosoft CorporationRemote Desktop ConnectionRemote Desktop Connection0.02016-05-25 18:01:53.8012016-05-25 18:00:002016-05-25  61006D00690074007000610077004300540032002D004100500050002D0039003000390031003000300031004100410030003200370032003900360042003100370032002E00320037002E0036002E00310035003500430059004200410047004500f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_1Explorer.EXEC:\Windows\Explorer.EXEMicrosoft® Windows® Operating SystemMicrosoft CorporationWindows ExplorerStart menu0.02016-05-25 18:01:33.412016-05-25 18:00:002016-05-25  61006D00690074007000610077004300540032002D004100500050002D0039003000390031003000300031004100410030003200370032003900360042003100370032002E00320037002E0036002E00310035003500430059004200410047004500f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_0StartWindowsSwitchStartWindowsSwitchStartWindowsSwitchStartWindowsSwitchStartWindowsSwitchStartWindowsSwitch0.02016-05-25 18:01:33.412016-05-25 18:00:002016-05-25 |

Additionally we can define Row Format and Field separator to load the data in directory

|  |
| --- |
| hive> INSERT OVERWRITE LOCAL DIRECTORY '/root/tempdir'  > ROW FORMAT DELIMITED  > FIELDS TERMINATED BY ','  > SELECT \* FROM temptableapplication LIMIT 10;  Query ID = root\_20170222091414\_bdef1157-15c4-4943-a9be-fa5e0ce5999e  Total jobs = 1  Launching Job 1 out of 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\_1484894548299\_0049, Tracking URL = http://mac55:8088/proxy/application\_1484894548299\_0049/  Kill Command = /opt/cloudera/parcels/CDH-5.8.2-1.cdh5.8.2.p0.3/lib/hadoop/bin/hadoop job -kill job\_1484894548299\_0049  Hadoop job information for Stage-1: number of mappers: 3; number of reducers: 1  2017-02-22 09:14:55,736 Stage-1 map = 0%, reduce = 0%  2017-02-22 09:15:10,525 Stage-1 map = 33%, reduce = 0%, Cumulative CPU 28.08 sec  2017-02-22 09:15:12,577 Stage-1 map = 67%, reduce = 0%, Cumulative CPU 32.54 sec  2017-02-22 09:15:31,008 Stage-1 map = 100%, reduce = 0%, Cumulative CPU 42.8 sec  2017-02-22 09:15:36,135 Stage-1 map = 100%, reduce = 100%, Cumulative CPU 43.71 sec  MapReduce Total cumulative CPU time: 43 seconds 710 msec  Ended Job = job\_1484894548299\_0049  Copying data to local directory /root/tempdir  MapReduce Jobs Launched:  Stage-Stage-1: Map: 3 Reduce: 1 Cumulative CPU: 43.71 sec HDFS Read: 132172238 HDFS Write: 4500 SUCCESS  Total MapReduce CPU Time Spent: 43 seconds 710 msec  OK  Time taken: 59.292 seconds |

Check data

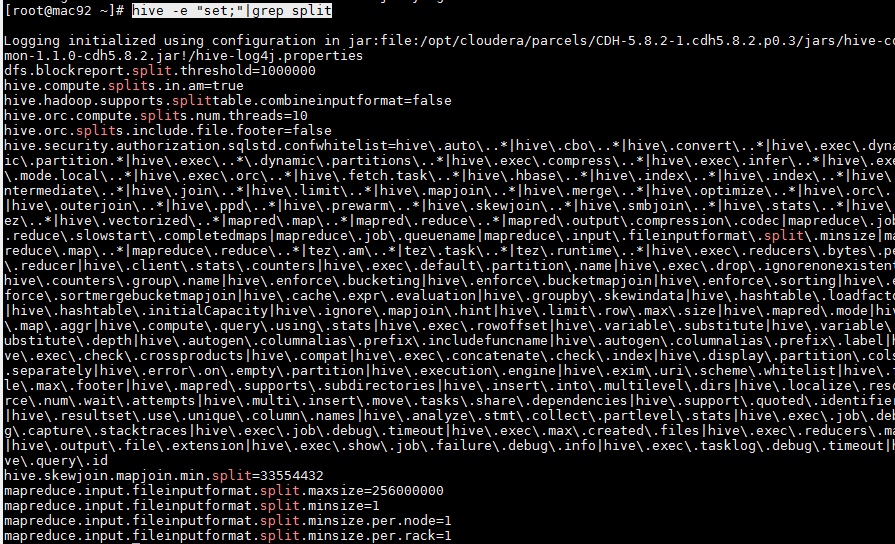
|  |
| --- |
| [root@mac55 tempdir]# cat 000000\_0  61006D0069007400700061007700,4300540032002D004100500050002D003900300039003100,300030003100410041003000320037003200390036004200,3100370032002E00320037002E0036002E00310035003500,430059004200410047004500,f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_5,mstsc.exe,C:\Windows\system32\mstsc.exe,Microsoft® Windows® Operating System,Microsoft Corporation,Remote Desktop Connection,Remote Desktop Connection,0.0,2016-05-25 18:01:53.848,2016-05-25 18:00:00,2016-05-25  61006D0069007400700061007700,4300540032002D004100500050002D003900300039003100,300030003100410041003000320037003200390036004200,3100370032002E00320037002E0036002E00310035003500,430059004200410047004500,f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_2,Explorer.EXE,C:\Windows\Explorer.EXE,Microsoft® Windows® Operating System,Microsoft Corporation,Windows Explorer,Search Pane,0.0,2016-05-25 18:01:43.613,2016-05-25 18:00:00,2016-05-25  61006D0069007400700061007700,4300540032002D004100500050002D003900300039003100,300030003100410041003000320037003200390036004200,3100370032002E00320037002E0036002E00310035003500,430059004200410047004500,f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_19,mstsc.exe,C:\Windows\system32\mstsc.exe,Microsoft® Windows® Operating System,Microsoft Corporation,Remote Desktop Connection,Remote Desktop Connection,0.0,2016-05-25 18:05:39.915,2016-05-25 18:00:00,2016-05-25  61006D0069007400700061007700,4300540032002D004100500050002D003900300039003100,300030003100410041003000320037003200390036004200,3100370032002E00320037002E0036002E00310035003500,430059004200410047004500,f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_16,mstsc.exe,C:\Windows\system32\mstsc.exe,Microsoft® Windows® Operating System,Microsoft Corporation,Remote Desktop Connection,Remote Desktop Connection,0.0,2016-05-25 18:03:34.303,2016-05-25 18:00:00,2016-05-25  61006D0069007400700061007700,4300540032002D004100500050002D003900300039003100,300030003100410041003000320037003200390036004200,3100370032002E00320037002E0036002E00310035003500,430059004200410047004500,f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_13,mstsc.exe,C:\Windows\system32\mstsc.exe,Microsoft® Windows® Operating System,Microsoft Corporation,Remote Desktop Connection,Remote Desktop Connection,0.0,2016-05-25 18:03:05.537,2016-05-25 18:00:00,2016-05-25  61006D0069007400700061007700,4300540032002D004100500050002D003900300039003100,300030003100410041003000320037003200390036004200,3100370032002E00320037002E0036002E00310035003500,430059004200410047004500,f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_10,mstsc.exe,C:\Windows\system32\mstsc.exe,Microsoft® Windows® Operating System,Microsoft Corporation,Remote Desktop Connection,Remote Desktop Connection,0.0,2016-05-25 18:02:47.396,2016-05-25 18:00:00,2016-05-25  61006D0069007400700061007700,4300540032002D004100500050002D003900300039003100,300030003100410041003000320037003200390036004200,3100370032002E00320037002E0036002E00310035003500,430059004200410047004500,f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_7,mstsc.exe,C:\Windows\system32\mstsc.exe,Microsoft® Windows® Operating System,Microsoft Corporation,Remote Desktop Connection,Remote Desktop Connection,0.0,2016-05-25 18:02:19.426,2016-05-25 18:00:00,2016-05-25  61006D0069007400700061007700,4300540032002D004100500050002D003900300039003100,300030003100410041003000320037003200390036004200,3100370032002E00320037002E0036002E00310035003500,430059004200410047004500,f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_4,mstsc.exe,C:\Windows\system32\mstsc.exe,Microsoft® Windows® Operating System,Microsoft Corporation,Remote Desktop Connection,Remote Desktop Connection,0.0,2016-05-25 18:01:53.801,2016-05-25 18:00:00,2016-05-25  61006D0069007400700061007700,4300540032002D004100500050002D003900300039003100,300030003100410041003000320037003200390036004200,3100370032002E00320037002E0036002E00310035003500,430059004200410047004500,f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_1,Explorer.EXE,C:\Windows\Explorer.EXE,Microsoft® Windows® Operating System,Microsoft Corporation,Windows Explorer,Start menu,0.0,2016-05-25 18:01:33.41,2016-05-25 18:00:00,2016-05-25  61006D0069007400700061007700,4300540032002D004100500050002D003900300039003100,300030003100410041003000320037003200390036004200,3100370032002E00320037002E0036002E00310035003500,430059004200410047004500,f0b9c075-72aa-46b3-af68-9253df8d66b1\_Client\_0,StartWindowsSwitch,StartWindowsSwitch,StartWindowsSwitch,StartWindowsSwitch,StartWindowsSwitch,StartWindowsSwitch,0.0,2016-05-25 18:01:33.41,2016-05-25 18:00:00,2016-05-25 |

Now data is comma separated

# 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

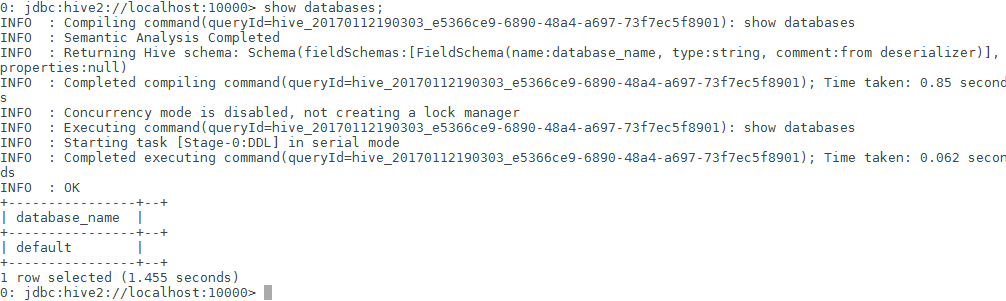
Just type beeline on terminal and click enter

!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 the 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.

# Replace Blank Value with Null

|  |
| --- |
| CREATE EXTERNAL TABLE person  (  name string,  title string,  birth\_year string  )  ROW FORMAT DELIMITED FIELDS TERMINATED BY '|'  LOCATION '/person'  **TBLPROPERTIES('serialization.null.format'='');** |

Set table properties like mentioned above, this will replace all blank values into null while inserting data into table

For more information refer below link

<https://abhijitsureshshingate.wordpress.com/2013/11/16/how-to-use-blank-as-null-in-hive/>

# Map phase stuck at 0%

I was trying to launch one query in which we have a join of two tables , hive converted this join to map side join and stuck at 0% and it is launching tasks again and again , we tried to increase container memory but issue persist

Please find below log for the same

|  |
| --- |
| Total jobs = 1  Execution log at: /tmp/root/root\_20170309060707\_2fb18bdd-b054-42b7-9075-c718ff55da4f.log  2017-03-09 06:07:49 Starting to launch local task to process map join; maximum memory = 1908932608  2017-03-09 06:07:55 Dump the side-table for tag: 1 with group count: 15 into file: file:/tmp/root/858bcac1-d9dd-484d-b51f-23c909351983/hive\_2017-03-09\_06-07-36\_752\_4756823141955389275-1/-local-10002/HashTable-Stage-4/MapJoin-mapfile01--.hashtable  2017-03-09 06:07:55 Uploaded 1 File to: file:/tmp/root/858bcac1-d9dd-484d-b51f-23c909351983/hive\_2017-03-09\_06-07-36\_752\_4756823141955389275-1/-local-10002/HashTable-Stage-4/MapJoin-mapfile01--.hashtable (760085 bytes)  2017-03-09 06:07:55 End of local task; Time Taken: 5.587 sec.  Execution completed successfully  MapredLocal task succeeded  Launching Job 1 out of 1  Number of reduce tasks is set to 0 since there's no reduce operator  Starting Job = job\_1488780768629\_0041, Tracking URL = http://c2esplusm1.espluscs1.i6.internal.cloudapp.net:8088/proxy/application\_1488780768629\_0041/  Kill Command = /opt/cloudera/parcels/CDH-5.7.0-1.cdh5.7.0.p0.45/lib/hadoop/bin/hadoop job -kill job\_1488780768629\_0041  Hadoop job information for Stage-4: number of mappers: 1; number of reducers: 0  2017-03-09 06:08:18,025 Stage-4 map = 0%, reduce = 0%  2017-03-09 06:09:18,968 Stage-4 map = 0%, reduce = 0%, Cumulative CPU 34.67 sec |

So to resolve this issue we change below property

set hive.auto.convert.join=false;

By doing so it is not converted into map side join and ran successfully