Shalaj Shukla

This document described how to copy the data between two hive cluster using hive IMPORT and Export commands

Copy Data between hive Cluster

Contents

[Copy Data from one hive instance to other 2](#_Toc465847348)

[Export Table 2](#_Toc465847349)

[Distcp to copy data 2](#_Toc465847350)

[Import Table 3](#_Toc465847351)

[Issue log 4](#_Toc465847352)

# Copy Data from one hive instance to other

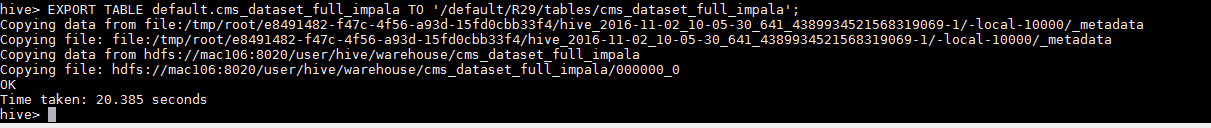
Since version 0.8, Hive supports EXPORT and IMPORT features that allows you to export the metadata as well as the data for the corresponding table to a directory in HDFS, which can then be imported back to another database or Hive instance

Here we are trying to copying data from 172.27.155.106 to 172.27.155.78

## Export Table

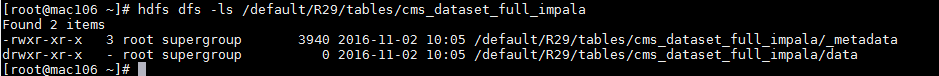
First enter into hive prompt in 172.27.155.106 and issue following command

|  |
| --- |
| EXPORT TABLE default.cms\_dataset\_full\_impala TO '/default/R29/tables/cms\_dataset\_full\_impala'; |



This command will load data as well as metadata into specified directory

Check whether folder structure created on hdfs



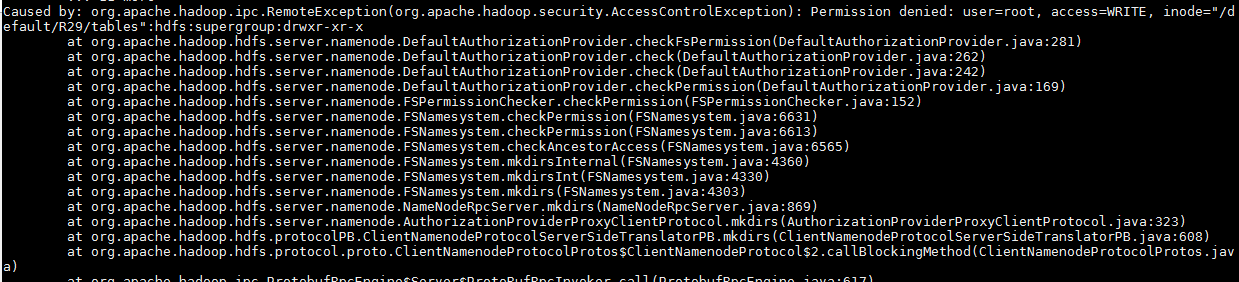
To export partitioned table you need to specify partition values

|  |
| --- |
| EXPORT TABLE default.dar\_site\_wise partition (sampled="unsampled", dt="2016-06-30",path="pagepath",exclude\_paid\_traffic="1",us\_desk\_tablet="1",site="US") TO '/default/R29/tables/dar\_site\_wise'; |

## Distcp to copy data

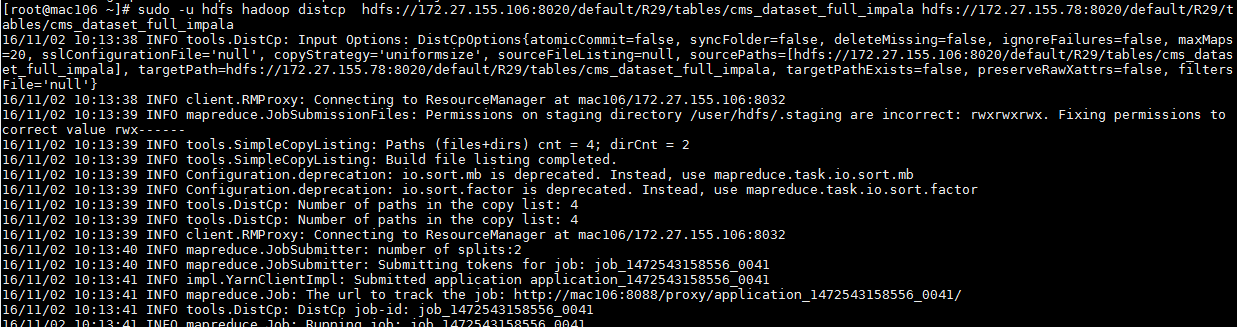
The next step is to copy the data from 172.27.155.106 to to 172.27.155.78 you can use the “distcp” command from Hadoop:

Here we are executing this command with user hdfs as we are getting below permission issue while executing it with root user

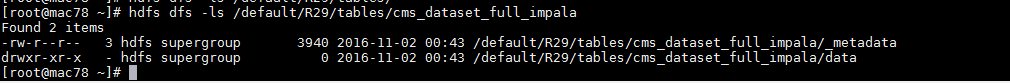


|  |
| --- |
| sudo -u hdfs hadoop distcp hdfs://172.27.155.106:8020/default/R29/tables/cms\_dataset\_full\_impala hdfs://172.27.155.78:8020/default/R29/tables/cms\_dataset\_full\_impala |

*Note: you can run command without sudo if there is no permission issue*



Now check in 172.27.155.78, if files are successfully copied



## Import Table

Now enter into hive prompt in 172.27.155.78 and issue import command

|  |
| --- |
| IMPORT EXTERNAL TABLE R29.cms\_dataset\_full\_impala FROM '/default/R29/tables/cms\_dataset\_full\_impala'; |

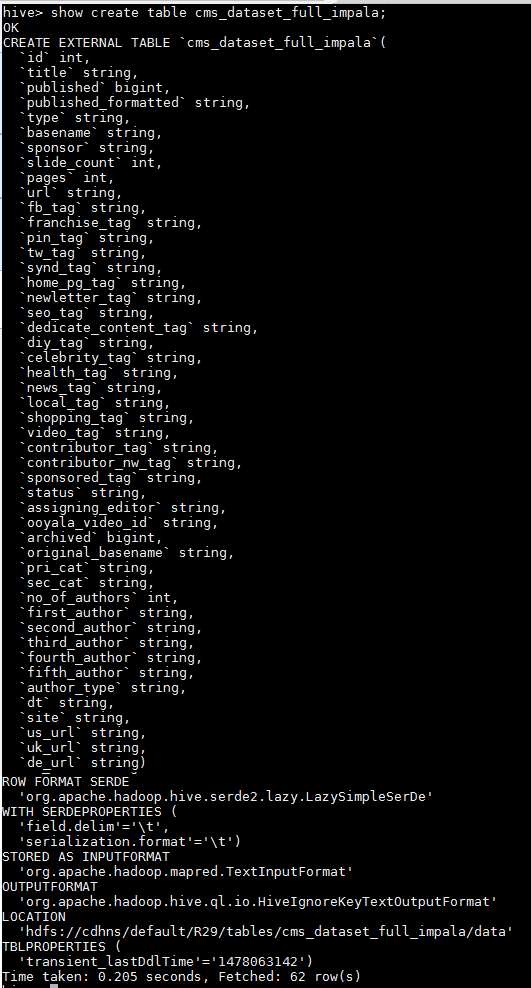


Since it is an External Table we are using External keyword else you need not to use it.

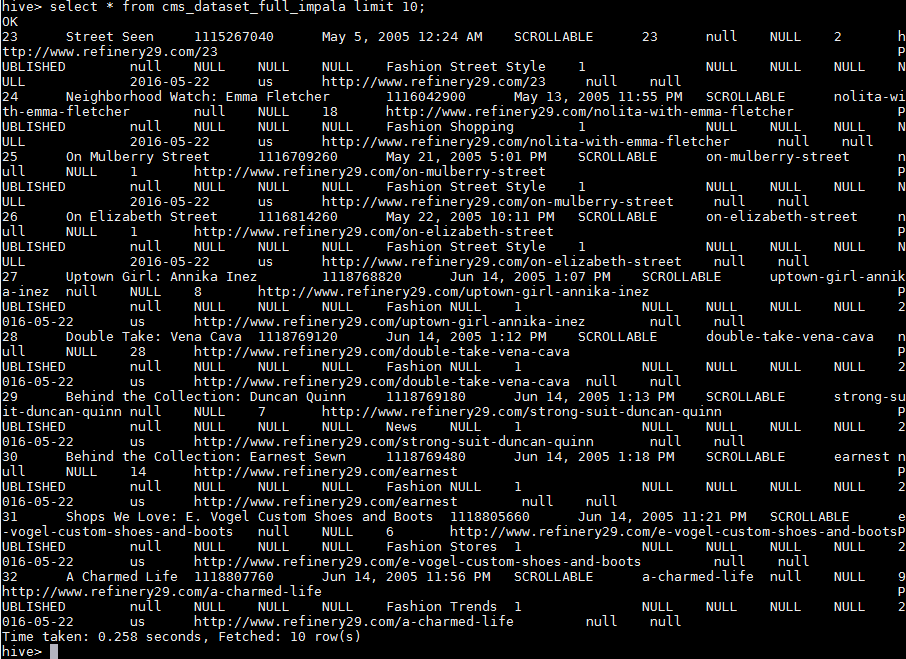
*Note: please drop table if already exist*

You can check the create table script using below command

|  |
| --- |
| show create table cms\_dataset\_full\_impala; |



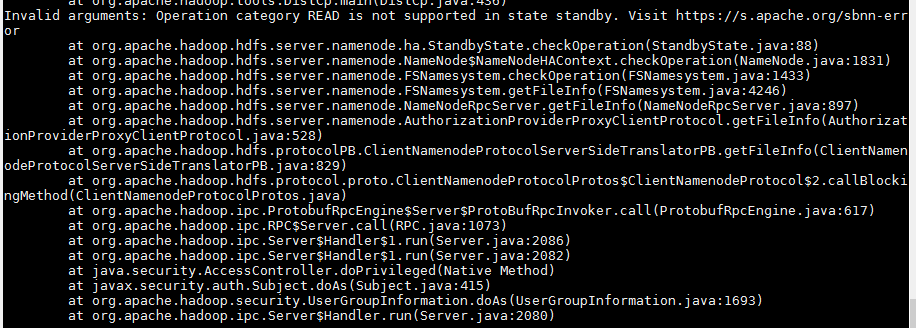
Now check Data



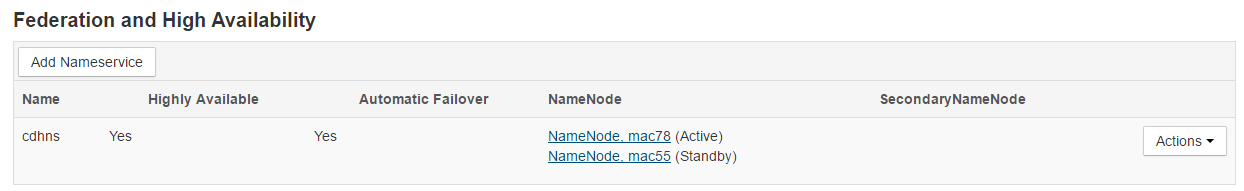
## Issue log

1. While issuing distcp command we are getting below issue

|  |
| --- |
| sudo -u hdfs hadoop distcp -overwrite hdfs://172.27.155.106:8020/default/R29/tables/cms\_dataset\_full\_impala hdfs://172.27.155.55:8020/default/R29/tables/cms\_dataset\_full\_impala |



This issue occurs because we are trying to copying data over standby node, this issue can occur if you configured high availability, so check the Active and Standby Node in Cloudera Manager UI and change the IP accordingly



2. During Import you can get following issue



This issue occurs because we are importing External table without using External command

So change command accordingly

|  |
| --- |
| IMPORT EXTERNAL TABLE R29.cms\_dataset\_full\_impala FROM '/default/R29/tables/cms\_dataset\_full\_impala'; |

3. During executing ***select count(\*) from ga\_dm\_impala*** getting below issue

|  |
| --- |
| Caused by: org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.AccessControlException): Permission denied: user=root, access=WRITE, inode="/user":hdfs:supergroup:drwxr-xr-x  at org.apache.hadoop.hdfs.server.namenode.DefaultAuthorizationProvider.checkFsPermission(DefaultAuthorizationProvider.java:281)  at org.apache.hadoop.hdfs.server.namenode.DefaultAuthorizationProvider.check(DefaultAuthorizationProvider.java:262)  at org.apache.hadoop.hdfs.server.namenode.DefaultAuthorizationProvider.check(DefaultAuthorizationProvider.java:242)  at org.apache.hadoop.hdfs.server.namenode.DefaultAuthorizationProvider.checkPermission(DefaultAuthorizationProvider.java:169)  at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkPermission(FSPermissionChecker.java:152)  at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkPermission(FSNamesystem.java:6631)  at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkPermission(FSNamesystem.java:6613)  at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkAncestorAccess(FSNamesystem.java:6565)  at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirsInternal(FSNamesystem.java:4360)  at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirsInt(FSNamesystem.java:4330)  at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.mkdirs(FSNamesystem.java:4303)  at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.mkdirs(NameNodeRpcServer.java:869)  at org.apache.hadoop.hdfs.server.namenode.AuthorizationProviderProxyClientProtocol.mkdirs(AuthorizationProviderProxyClientProtocol.java:323) |

After debugging into the details, found out the to read the count() the user needs to read the blocks on the HDFS. Since the user didn’t have needed privileges, we are getting the error. The issue was resolved using below

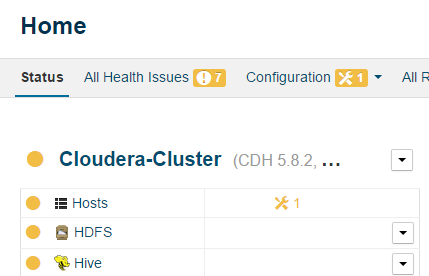
|  |
| --- |
| [root@mac78 ~]# sudo -u hdfs hadoop fs -mkdir /user/root  [root@mac78 ~]# sudo -u hdfs hadoop fs -chown root:root /user/root |

OR

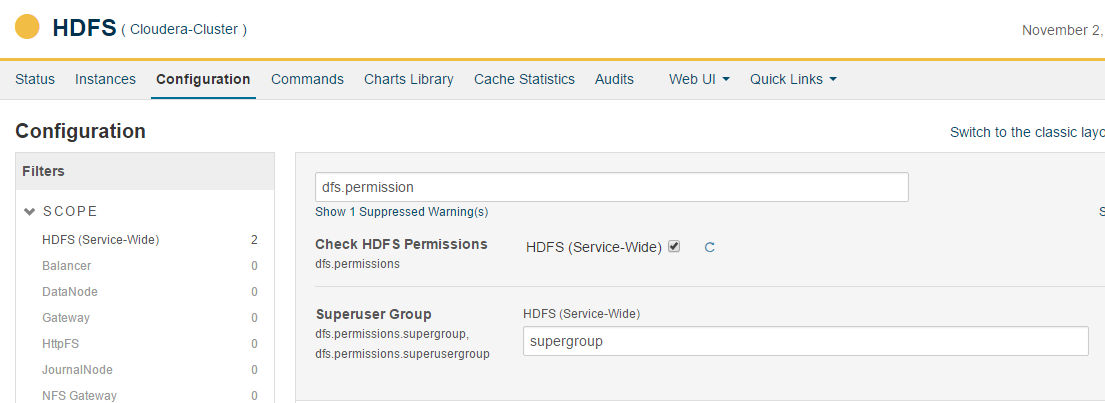
There is one more way to solve this issue, you can update/add below property in /etc/Hadoop/conf/hdfs-site.xml file

|  |
| --- |
| <property> <name>dfs.permissions</name> <value>false</value> </property> |

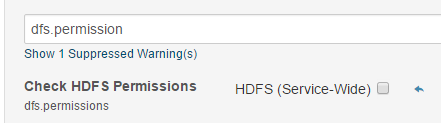
If you are using Cloudera-manager click on HDFS on home page



Click on Configuration



Search **dfs.permission** and deselect check box



Click on **Save Changes** and redeploy it