

## Contents

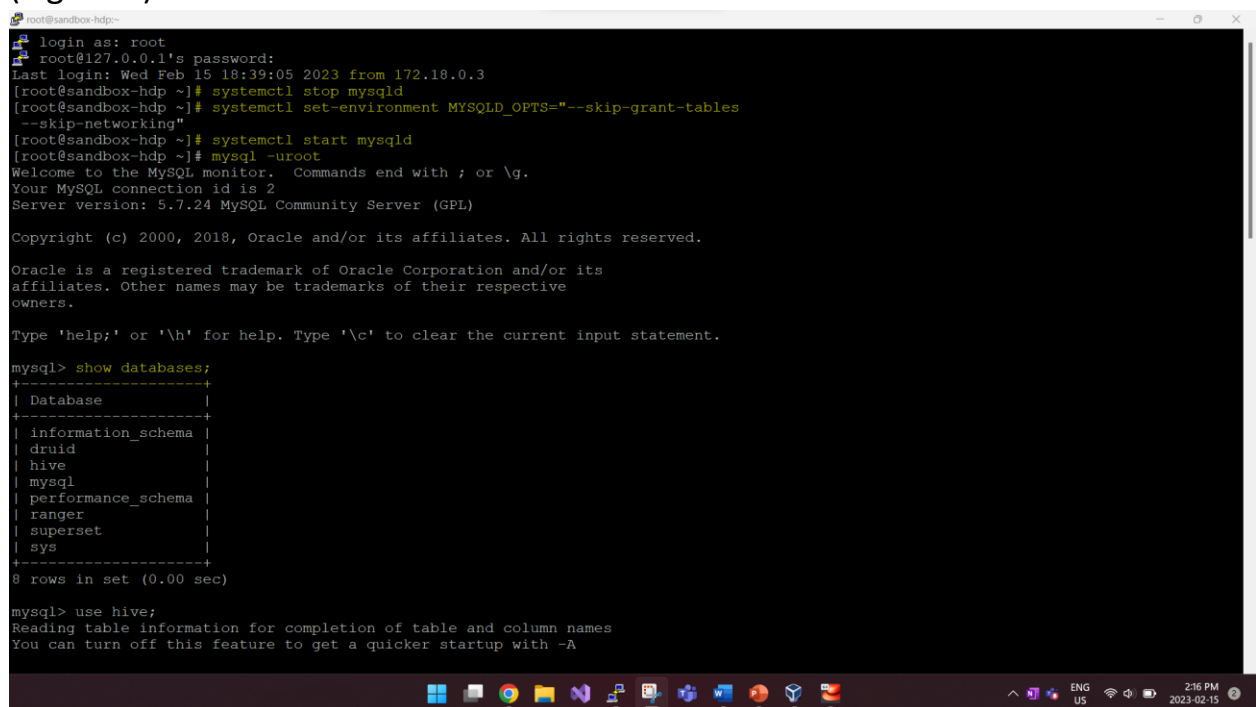
Assignment 3- Hive & Sqoop.....	1
<b>Objective:</b> Use Sqoop to transfer data to MySQL from Hadoop or vice versa .....	1
Import a table from MySQL to HDFS (Choose the parameter as you wish) .....	1
2 Demonstrate export command. ....	6
Task 2: Week 6: Hive.....	8

## Assignment 3- Hive & Sqoop

**Objective:** Use Sqoop to transfer data to MySQL from Hadoop or vice versa

Import a table from MySQL to HDFS (Choose the parameter as you wish)

Well, importing a table from **MYSQL** to **HDFS**, I had to start MYSQL with **systemctl start mysqld** then I logged with root (**MySQL -uroot**). After that I checked databases in mysql where I got hive databases which used for table. (Figure 1)



```
root@sandbox-hdp:~# login as: root
root@127.0.0.1's password:
Last login: Wed Feb 15 18:39:05 2023 from 172.18.0.3
[root@sandbox-hdp ~]# systemctl stop mysqld
[root@sandbox-hdp ~]# systemctl set-environment MYSQLD_OPTS="--skip-grant-tables
--skip-networking"
[root@sandbox-hdp ~]# systemctl start mysqld
[root@sandbox-hdp ~]# mysql -uroot
Welcome to the MySQL monitor.  Commands end with ; or \g.
Your MySQL connection id is 2
Server version: 5.7.24 MySQL Community Server (GPL)

Copyright (c) 2000, 2018, Oracle and/or its affiliates. All rights reserved.

Oracle is a registered trademark of Oracle Corporation and/or its
affiliates. Other names may be trademarks of their respective
owners.

Type 'help;' or '\h' for help. Type '\c' to clear the current input statement.

mysql> show databases;
+-----+
| Database |
+-----+
| information_schema |
| druid      |
| hive      |
| mysql     |
| performance_schema |
| ranger    |
| superset  |
| sys       |
+-----+
8 rows in set (0.00 sec)

mysql> use hive;
Reading table information for completion of table and column names
You can turn off this feature to get a quicker startup with -A
```

**Figure 1** Start Mysql and logged with root user

I used **ROLES** table, and Firstly I noticed that what's in a table, used **select \* from ROLES**. (Figure 4)

```
root@sandbox-hdp:~#
You can turn off this feature to get a quicker startup with -A

Database changed
mysql> show tables;
+-----+
| Tables in hive |
+-----+
| AUX_TABLE      |
| BUCKETING_COLS |
| CDS             |
| COLUMNS_V2     |
| COMPACTION_QUEUE |
| COMPLETED_COMPACTIONS |
| COMPLETED_TXN_COMPONENTS |
| CTLG           |
| DATABASE_PARAMS |
| DBS            |
| DB_PRIVS       |
| DELEGATION_TOKENS |
| FUNCS          |
| FUNC_RU        |
| GLOBAL_PRIVS   |
| HIVE_LOCKS     |
| IDX            |
| INDEX_PARAMS   |
| I_SCHEMA       |
| KEY_CONSTRAINTS |
| MASTER_KEYS    |
| MATERIALIZATION_REBUILD_LOCKS |
| METASTORE_DB_PROPERTIES |
| MIN_HISTORY_LEVEL |
| MV_CREATION_METADATA |
| MV_TABLES_USED |
| NEXT_COMPACTION_QUEUE_ID |
| NEXT_LOCK_ID   |
| NEXT_TXN_ID    |
| NEXT_WRITE_ID  |
| NOTIFICATION_LOG |
| NOTIFICATION_SEQUENCE |
+-----+
```

**Figure 2** Hive table details

```
root@sandbox-hdp:~#
mysql> select * from ROLES;
+-----+
| PART_PRIVS |
| REPL_TXN_MAP |
| ROLES      |
| ROLE_MAP   |
| RUNTIME_STATS |
| SCHEMA_VERSION |
| SDS        |
| SD_PARAMS  |
| SEQUENCE_TABLE |
| SERDES     |
| SERDE_PARAMS |
| SKEWED_COL_NAMES |
| SKEWED_COL_VALUE_LOC_MAP |
| SKEWED_STRING_LIST |
| SKEWED_STRING_LIST_VALUES |
| SKEWED_VALUES |
| SORT_COLS  |
| TABLE_PARAMS |
| TAB_COL_STATS |
| TBLS       |
| TBL_COL_PRIVS |
| TBL_PRIVS   |
| TXNS        |
| TXN_COMPONENTS |
| TXN_TO_WRITE_ID |
| TXN_WRITE_NOTIFICATION_LOG |
| TYPES       |
| TYPE_FIELDS |
| VERSION     |
| WM_MAPPING  |
| WM_POOL     |
| WM_POOL_TO_TRIGGER |
| WM_RESOURCEPLAN |
| WM_TRIGGER  |
| WRITE_SET   |
+-----+
75 rows in set (0.00 sec)

mysql>
```

**Figure 3** Hive table details

```
root@sandbox-hdp~  
| PART_COL_STATS  
| PART_PRIVS  
| REFL_TXN_MAP  
| ROLES  
| ROLE_MAP  
| RUNTIME_STATS  
| SCHEMA_VERSION  
| SDS  
| SD_PARAMS  
| SEQUENCE_TABLE  
| SEARDS  
| SERIE_PARAMS  
| SKEWED_COL_NAMES  
| SKEWED_COL_VALUE_LOC_MAP  
| SKEWED_STRING_LIST  
| SKEWED_STRING_LIST_VALUES  
| SKEWED_VALUES  
| SORT_COLS  
| TABLE_PARAMS  
| TAB_COL_STATS  
| TBLS  
| TBL_COL_PRIVS  
| TBL_PRIVS  
| TXNS  
| TXN_COMPONENTS  
| TXN_TO_WRITE_ID  
| TXN_WRITE_NOTIFICATION_LOG  
| TYPES  
| TYPE_FIELDS  
| VERSION  
| WM_MAPPING  
| WM_POOL  
| WM_POOL_TO_TRIGGER  
| WM_RESOURCEPLAN  
| WM_TRIGGER  
| WRITE_SET  
-----  
75 rows in set (0.01 sec)  
  
mysql> select * from ROLES;  
+-----+-----+-----+-----+  
| ROLE_ID | CREATE TIME | OWNER_NAME | ROLE_NAME |  
+-----+-----+-----+-----+  
| 1 | 1543513948 | admin | admin |  
| 2 | 1543513948 | public | public |  
+-----+-----+-----+-----+  
2 rows in set (0.00 sec)  
  
mysql> 
```

**Figure 4** Inside ROLES table

I used *Hadoop command* to check all directories where I got *user directory* then I made a new *subdirectory* in directory which is *sandeep*. I used subdirectory to *import table* from *databases* to *hdfs*. (Figure 5)

```
hdfs@sandbox-hdp:~$ hdfs dfs -ls /user
Found 16 items
drwxr-xr-x - admin hdfs 0 2018-11-29 17:28 /user/admin
drwxr-xr-x - ambari-ga hdfs 0 2018-11-29 17:28 /user/ambari-ga
drwxr-xr-x - amy_ds hdfs 0 2018-11-29 17:28 /user/amy_ds
drwxr-xr-x - anonymous hdfs 0 2018-11-29 17:28 /user/anonymous
drwxr-xr-x - druid hadoop 0 2018-11-29 19:01 /user/druid
drwxr-xr-x - hbase hdfs 0 2018-11-29 17:40 /user/hbase
drwxr-xr-x - hdfs hdfs 0 2018-11-29 19:21 /user/hdfs
drwxr-xr-x - hive hdfs 0 2023-02-19 09:20 /user/hive
drwxr-xr-x - livy hdfs 0 2018-11-29 17:55 /user/livy
drwxr-xr-x - maria_dev hdfs 0 2018-11-29 17:28 /user/maria_dev
drwxr-xr-x - oozie hdfs 0 2018-11-29 19:06 /user/oozie
drwxr-xr-x - raj_ops hdfs 0 2018-11-29 17:28 /user/raj_ops
drwxr-xr-x - root hdfs 0 2018-11-29 17:28 /user/root
drwxr-xr-x - spark hdfs 0 2018-11-29 17:57 /user/spark
drwxr-xr-x - zeppelin hdfs 0 2018-11-29 17:50 /user/zeppelin
hdfs@sandbox-hdp:~$ hdfs dfs -mkdir /user/sandeep
mkdir: permission denied: user=root, access=write, inode="/user":hdfs:hdfs:drwxr-xr-x
hdfs@sandbox-hdp:~$ sudo -iu hdfs
hdfs@sandbox-hdp:~$ hdfs dfs -mkdir /user/sandeep
hdfs@sandbox-hdp:~$ hdfs dfs -ls /user
Found 16 items
drwxr-xr-x - admin hdfs 0 2018-11-29 17:28 /user/admin
drwxr-xr-x - ambari-ga hdfs 0 2018-11-29 17:28 /user/ambari-ga
drwxr-xr-x - amy_ds hdfs 0 2018-11-29 17:28 /user/amy_ds
drwxr-xr-x - anonymous hdfs 0 2018-11-29 17:28 /user/anonymous
drwxr-xr-x - druid hadoop 0 2018-11-29 19:01 /user/druid
drwxr-xr-x - hbase hdfs 0 2018-11-29 17:40 /user/hbase
drwxr-xr-x - hdfs hdfs 0 2018-11-29 19:21 /user/hdfs
drwxr-xr-x - hive hdfs 0 2023-02-19 09:20 /user/hive
drwxr-xr-x - livy hdfs 0 2018-11-29 17:55 /user/livy
drwxr-xr-x - maria_dev hdfs 0 2018-11-29 17:28 /user/maria_dev
drwxr-xr-x - oozie hdfs 0 2018-11-29 19:06 /user/oozie
drwxr-xr-x - raj_ops hdfs 0 2018-11-29 17:28 /user/raj_ops
drwxr-xr-x - root hdfs 0 2018-11-29 17:28 /user/root
drwxr-xr-x - hdfs hdfs 0 2023-02-19 09:50 /user/sandeep
drwxr-xr-x - spark hdfs 0 2018-11-29 17:57 /user/spark
drwxr-xr-x - zeppelin hdfs 0 2018-11-29 17:50 /user/zeppelin
hdfs@sandbox-hdp:~$ hdfs dfs -ls /user/sandeep
Warning: /usr/hdp/3.0.1.0-187/accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/usr/hdp/3.0.1.0-187/hadoop/lib/slf4j-log4j12-1.7.25.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/usr/hdp/3.0.1.0-187/hive/lib/log4j-slf4j-impl-2.10.0.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
23/02/19 09:54:05 INFO sqoop.Sqoop: Running Sqoop version: 1.4.7.3.0.1.0-187
23/02/19 09:54:05 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/02/19 09:54:05 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
```

Figure 5 made a subdirectory with sandeep name

After importing a table, I used **ls command** to check there is a **sqoop** directory where I got it and used again **ls command** to check what's in a table, I got all content. (Figure 6)

```
hdfs@sandbox-hdp-
FILE: Number of read operations=0
FILE: Number of large read operations=0
FILE: Number of write operations=0
HDFS: Number of bytes read=742
HDFS: Number of bytes written=27710
HDFS: Number of read operations=42
HDFS: Number of large read operations=0
HDFS: Number of write operations=14
Job Counters
  Launched map tasks=7
  Other local map tasks=7
  Total time spent by all maps in occupied slots (ms)=293240
  Total time spent by all reduces in occupied slots (ms)=0
  Total time spent by all map tasks (ms)=73310
  Total vcore-milliseconds taken by all map tasks=73310
  Total megabyte-milliseconds taken by all map tasks=75069440
Map-Reduce Framework
  Map input records=88
  Map output records=88
  Input split bytes=742
  Spilled Records=0
  Failed Shuffles=0
  Merged Map outputs=0
  GC time elapsed (ms)=933
  CPU time spent (ms)=10290
  Physical memory (bytes) snapshot=1585328128
  Virtual memory (bytes) snapshot=19886223104
  Total committed heap usage (bytes)=999817216
  Peak Map Physical memory (bytes)=243437568
  Peak Map Virtual memory (bytes)=2855944192
File Input Format Counters
  Bytes Read=0
File Output Format Counters
  Bytes Written=27710
23/02/19 09:55:36 INFO mapreduce.ImportJobBase: Transferred 27.0605 KB in 75.7074 seconds (366.0145 bytes/sec)
23/02/19 09:55:36 INFO mapreduce.ImportJobBase: Retrieved 88 records.
(hdfs@sandbox-hdp ~)$ hdfs dfs -ls /user/sandeep/
Found 1 items
drwxr-xr-x  - hdfs hdfs          0 2023-02-19 09:55 /user/sandeep/sqoop
(hdfs@sandbox-hdp ~)$ hdfs dfs -ls /user/sandeep/sqoop/
Found 8 items
-rw-r--r--  1 hdfs hdfs          0 2023-02-19 09:55 /user/sandeep/sqoop/_SUCCESS
-rw-r--r--  1 hdfs hdfs       1000 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00000
-rw-r--r--  1 hdfs hdfs       1033 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00001
-rw-r--r--  1 hdfs hdfs       2519 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00002
-rw-r--r--  1 hdfs hdfs       2015 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00003
-rw-r--r--  1 hdfs hdfs        879 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00004
-rw-r--r--  1 hdfs hdfs       1081 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00005
-rw-r--r--  1 hdfs hdfs       1083 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00006
```

**Figure 6** used **ls** command to check the directory and got sqoop

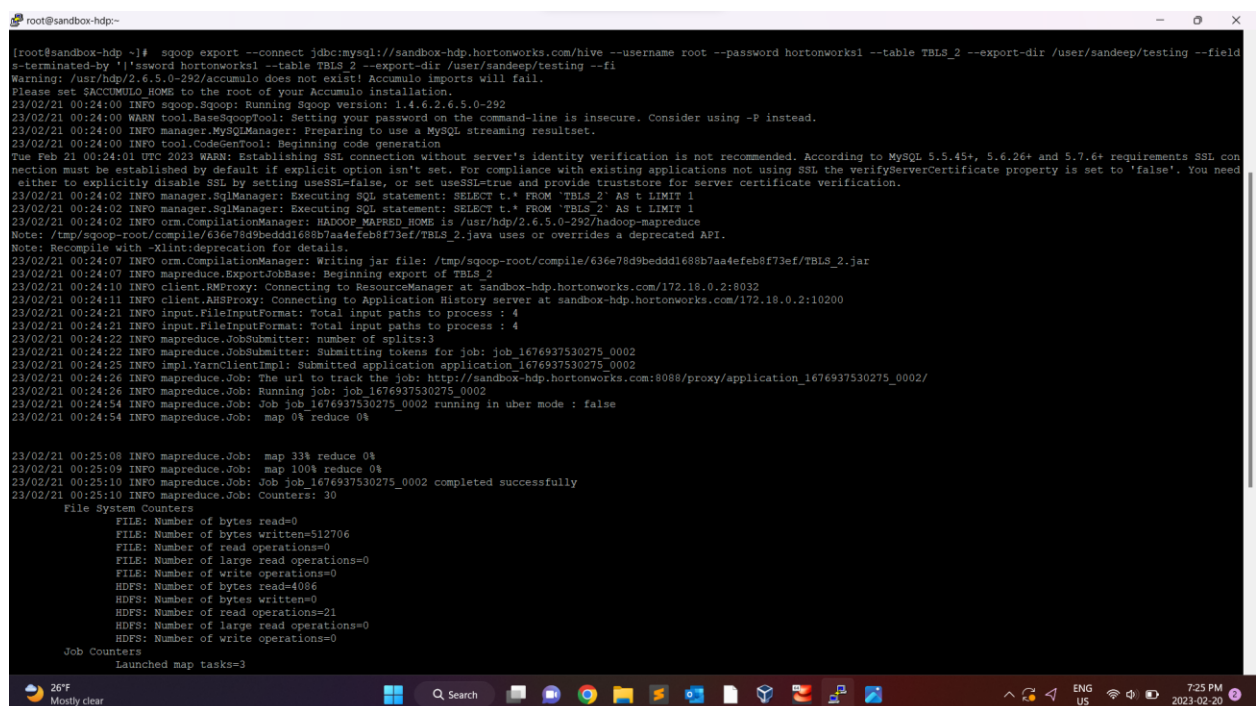
and used again **ls** command to check what's in a table, I got all content which I showed in Figure 7. I have checked the content in **part-m-0000**.

```
[root@sandbox-hdp ~]$ hdfs dfs -ls /user/sandeep/sqoop/
Found 8 items
-rw-r--r--  1 hdfs hdfs          0 2023-02-19 09:55 /user/sandeep/sqoop/_SUCCESS
-rw-r--r--  1 hdfs hdfs       1000 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00000
-rw-r--r--  1 hdfs hdfs       1033 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00001
-rw-r--r--  1 hdfs hdfs       2519 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00002
-rw-r--r--  1 hdfs hdfs       2015 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00003
-rw-r--r--  1 hdfs hdfs        879 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00004
-rw-r--r--  1 hdfs hdfs       1081 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00005
-rw-r--r--  1 hdfs hdfs       1083 2023-02-19 09:55 /user/sandeep/sqoop/part-m-00006
[root@sandbox-hdp ~]$ hdfs dfs -cat /user/sandeep/sqoop/part-m-00000
11543513905|6|0|hive|USER|0|1|bucketing cols|EXTERNAL_TABLE|null|null|false|0
211543513987|6|0|hive|USER|0|2|cols|EXTERNAL_TABLE|null|null|false|0
311543513989|6|0|hive|USER|0|3|columns v2|EXTERNAL_TABLE|null|null|false|0
411543513991|6|0|hive|USER|0|4|database_params|EXTERNAL_TABLE|null|null|false|0
511543513993|6|0|hive|USER|0|5|dbs|EXTERNAL_TABLE|null|null|false|0
611543513995|6|0|hive|USER|0|6|db_privs|EXTERNAL_TABLE|null|null|false|0
711543513996|6|0|hive|USER|0|7|global_privs|EXTERNAL_TABLE|null|null|false|0
811543513998|6|0|hive|USER|0|8|partitions|EXTERNAL_TABLE|null|null|false|0
911543514000|6|0|hive|USER|0|9|partition_keys|EXTERNAL_TABLE|null|null|false|0
1011543514002|6|0|hive|USER|0|10|partition_key_vals|EXTERNAL_TABLE|null|null|false|0
1111543514003|6|0|hive|USER|0|11|partition_params|EXTERNAL_TABLE|null|null|false|0
1211543514005|6|0|hive|USER|0|12|part_col_privs|EXTERNAL_TABLE|null|null|false|0
1311543514007|6|0|hive|USER|0|13|part_privs|EXTERNAL_TABLE|null|null|false|0
[root@sandbox-hdp ~]$
```

**Figure 7** used ls command to check the whole content inside sqoop

## 2 Demonstrate export command.

I showed a command to export the content from **hdfs to mysql** databases. But before exporting the data from **HDFS to MySql**, I needed to create a table in **mysql** where I had to export all content from **hdfs**. I created a table which is **TBLS\_2** from texting directory After creating a table I used export command which I showed in Figure 8



```
[root@sandbox-hdp ~]# sqoop export --connect jdbc:mysql://sandbox-hdp.hortonworks.com/hive --username root --password hortonworks1 --table TBLS_2 --export-dir /user/sandeep/testing --fields-terminated-by '|'
Warning: /usr/hdp/2.6.5.0-292/accumulo does not exist: Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
23/02/21 00:24:00 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6.2.6.5.0-292
23/02/21 00:24:00 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/02/21 00:24:00 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/02/21 00:24:00 INFO tool.CodeGenTool: Beginning code generation
Tue Feb 23 00:24:01 UTC 2023 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL con
nection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need
either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
23/02/21 00:24:02 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'TBLS_2' AS t LIMIT 1
23/02/21 00:24:02 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'TBLS_2' AS t LIMIT 1
23/02/21 00:24:02 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/hdp/2.6.5.0-292/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/636e78d9beddd1688b7aa4efeb8f73ef/TBLS_2.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/02/21 00:24:07 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/636e78d9beddd1688b7aa4efeb8f73ef/TBLS_2.jar
23/02/21 00:24:07 INFO mapreduce.ExportJobBase: Beginning export of TBLS_2
23/02/21 00:24:10 INFO client.RMProxy: Connecting to ResourceManager at sandbox-hdp.hortonworks.com/172.18.0.2:8032
23/02/21 00:24:11 INFO client.AHSProxy: Connecting to Application History server at sandbox-hdp.hortonworks.com/172.18.0.2:10200
23/02/21 00:24:21 INFO input.FileInputFormat: Total input paths to process : 4
23/02/21 00:24:21 INFO input.FileInputFormat: Total input paths to process : 4
23/02/21 00:24:22 INFO mapreduce.JobSubmitter: number of splits:3
23/02/21 00:24:22 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1676937530275_0002
23/02/21 00:24:25 INFO impl.YarnClientImpl: Submitted application application_1676937530275_0002
23/02/21 00:24:26 INFO mapreduce.Job: The url to track the job: http://sandbox-hdp.hortonworks.com:8088/proxy/application_1676937530275_0002/
23/02/21 00:24:26 INFO mapreduce.Job: Running job: job_1676937530275_0002
23/02/21 00:24:54 INFO mapreduce.Job: Job job_1676937530275_0002 running in uber mode : false
23/02/21 00:24:54 INFO mapreduce.Job: map 0% reduce 0%

23/02/21 00:25:08 INFO mapreduce.Job: map 33% reduce 0%
23/02/21 00:25:09 INFO mapreduce.Job: map 100% reduce 0%
23/02/21 00:25:10 INFO mapreduce.Job: Job job_1676937530275_0002 completed successfully
23/02/21 00:25:10 INFO mapreduce.Job: Counters: 30

File System Counters
  FILE: Number of bytes read=0
  FILE: Number of bytes written=512706
  FILE: Number of read operations=0
  FILE: Number of large read operations=0
  FILE: Number of write operations=0
  HDFS: Number of bytes read=4086
  HDFS: Number of bytes written=0
  HDFS: Number of read operations=21
  HDFS: Number of large read operations=0
  HDFS: Number of write operations=0

Job Counters
  Launched map tasks=3
```

**Figure 8** Export command

Export Command> **sqoop export --connect jdbc:mysql://sandbox-hdp.hortonworks.com/hive --username root --password hortonworks1 --table TBLS\_2 --table TBLS\_2 --export-dir /user/Sandeep/testing --fields-terminated-by '|'**

I used **hive** databases to send all content from hdfs to mysql where I used **TBLS\_2** table.

```
root@sandbox-hdp-
terminated-by '|' -num-mappers 7^C
[root@sandbox-hdp ~]# ^C
[root@sandbox-hdp ~]#
[root@sandbox-hdp ~]#
[root@sandbox-hdp ~]#
[root@sandbox-hdp ~]# sqoop-import --connect jdbc:mysql://sandbox-hdp.hortonworks.com/hive --
username root -password hortonworks1 --table TBLS --target-dir /user/sandeeep/testing --fields
-terminated-by '|'
Warning: /usr/hdp/2.6.5.0-292/accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
23/02/21 00:15:40 INFO sqoop.Sqoop: Running Sqoop version: 1.4.6.2.6.5.0-292
23/02/21 00:15:40 WARN tool.BaseSqoopTool: Setting your password on the command-line is insecure. Consider using -P instead.
23/02/21 00:15:41 INFO manager.MySQLManager: Preparing to use a MySQL streaming resultset.
23/02/21 00:15:41 INFO tool.CodeGenTool: Beginning code generation
Tue Feb 21 00:15:42 UTC 2023 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL con
nection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need
either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
23/02/21 00:15:43 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'TBLS' AS t LIMIT 1
23/02/21 00:15:43 INFO manager.SqlManager: Executing SQL statement: SELECT t.* FROM 'TBLS' AS t LIMIT 1
23/02/21 00:15:43 INFO orm.CompilationManager: HADOOP_MAPRED_HOME is /usr/hdp/2.6.5.0-292/hadoop-mapreduce
Note: /tmp/sqoop-root/compile/2ed6ec23a59188cafa1262be3543f680/TBLS.java uses or overrides a deprecated API.
Note: Recompile with -Xlint:deprecation for details.
23/02/21 00:15:49 INFO orm.CompilationManager: Writing jar file: /tmp/sqoop-root/compile/2ed6ec23a59188cafa1262be3543f680/TBLS.jar
23/02/21 00:15:49 WARN manager.MySQLManager: It looks like you are importing from mysql.
23/02/21 00:15:49 WARN manager.MySQLManager: This transfer can be faster! Use the --direct
23/02/21 00:15:49 WARN manager.MySQLManager: option to exercise a MySQL-specific fast path.
23/02/21 00:15:49 INFO manager.MySQLManager: Setting zero RANTIME behavior to convertToNull (mysql)
23/02/21 00:15:49 INFO mapreduce.ImportJobBase: Beginning import of TBLS
23/02/21 00:15:54 INFO client.RMProxy: Connecting to ResourceManager at sandbox-hdp.hortonworks.com/172.18.0.2:8032
23/02/21 00:15:54 INFO client.AHSProxy: Connecting to Application History server at sandbox-hdp.hortonworks.com/172.18.0.2:10200
Tue Feb 21 00:16:03 UTC 2023 WARN: Establishing SSL connection without server's identity verification is not recommended. According to MySQL 5.5.45+, 5.6.26+ and 5.7.6+ requirements SSL con
nection must be established by default if explicit option isn't set. For compliance with existing applications not using SSL the verifyServerCertificate property is set to 'false'. You need
either to explicitly disable SSL by setting useSSL=false, or set useSSL=true and provide truststore for server certificate verification.
23/02/21 00:16:03 INFO db.DBInputFormat: Using read committed transaction isolation
23/02/21 00:16:03 INFO db.DataDrivenBInputFormat: BoundingValsQuery: SELECT MIN('TBL_ID'), MAX('TBL_ID') FROM 'TBLS'
23/02/21 00:16:03 INFO db.IntegerSplitter: Split size: 10; Num splits: 4 from: 1 to: 41
23/02/21 00:16:03 INFO mapreduce.JobSubmitter: number of splits:4
23/02/21 00:16:03 INFO mapreduce.JobSubmitter: Submitting tokens for job: job_1676937530275_0001
23/02/21 00:16:05 INFO impl.YarnClientImpl: Submitted application application_1676937530275_0001
23/02/21 00:16:05 INFO mapreduce.Job: The url to track the job: http://sandbox-hdp.hortonworks.com:8088/proxy/application_1676937530275_0001/
23/02/21 00:16:05 INFO mapreduce.Job: Running job: job_1676937530275_0001

23/02/21 00:16:46 INFO mapreduce.Job: Job job_1676937530275_0001 running in uber mode : false
23/02/21 00:16:46 INFO mapreduce.Job: map 0% reduce 0%
```

**Figure 9** Export command

```
root@sandbox-hdp:~#
23/02/21 00:24:26 INFO mapreduce.Job: Running job: job_1676937530275_0002
23/02/21 00:24:54 INFO mapreduce.Job: Job job_1676937530275_0002 running in uber mode : false
23/02/21 00:24:54 INFO mapreduce.Job: map 0% reduce 0%

23/02/21 00:25:08 INFO mapreduce.Job: map 33% reduce 0%
23/02/21 00:25:09 INFO mapreduce.Job: map 100% reduce 0%
23/02/21 00:25:10 INFO mapreduce.Job: Job job_1676937530275_0002 completed successfully
23/02/21 00:25:10 INFO mapreduce.Job: Counters: 30
  File System Counters
    FILE: Number of bytes read=0
    FILE: Number of bytes written=512706
    FILE: Number of read operations=0
    FILE: Number of large read operations=0
    FILE: Number of write operations=0
    HDFS: Number of bytes read=4086
    HDFS: Number of bytes written=0
    HDFS: Number of read operations=21
    HDFS: Number of large read operations=0
    HDFS: Number of write operations=0
  Job Counters
    Launched map tasks=3
    Data-local map tasks=3
    Total time spent by all maps in occupied slots (ms)=37664
    Total time spent by all reduces in occupied slots (ms)=0
    Total time spent by all map tasks (ms)=37664
    Total vcore-milliseconds taken by all map tasks=37664
    Total megabyte-milliseconds taken by all map tasks=9416000
  Map-Reduce Framework
    Map input records=38
    Map output records=38
    Input split bytes=732
    Spilled Records=0
    Failed Shuffles=0
    Merged Map outputs=0
    GC time elapsed (ms)=1547
    CPU time spent (ms)=11910
    Physical memory (bytes) snapshot=495316992
    Virtual memory (bytes) snapshot=5829050368
    Total committed heap usage (bytes)=127401984
  File Input Format Counters
    Bytes Read=0
  File Output Format Counters
    Bytes Written=0
23/02/21 00:25:10 INFO mapreduce.ExportJobBase: Transferred 3.9902 KB in 60.4251 seconds (67.6209 bytes/sec)
23/02/21 00:25:10 INFO mapreduce.ExportJobBase: Exported 38 records.
[root@sandbox-hdp ~]#
[root@sandbox-hdp ~]#
```

**Figure 10** Exported all data from hdfs to mysql

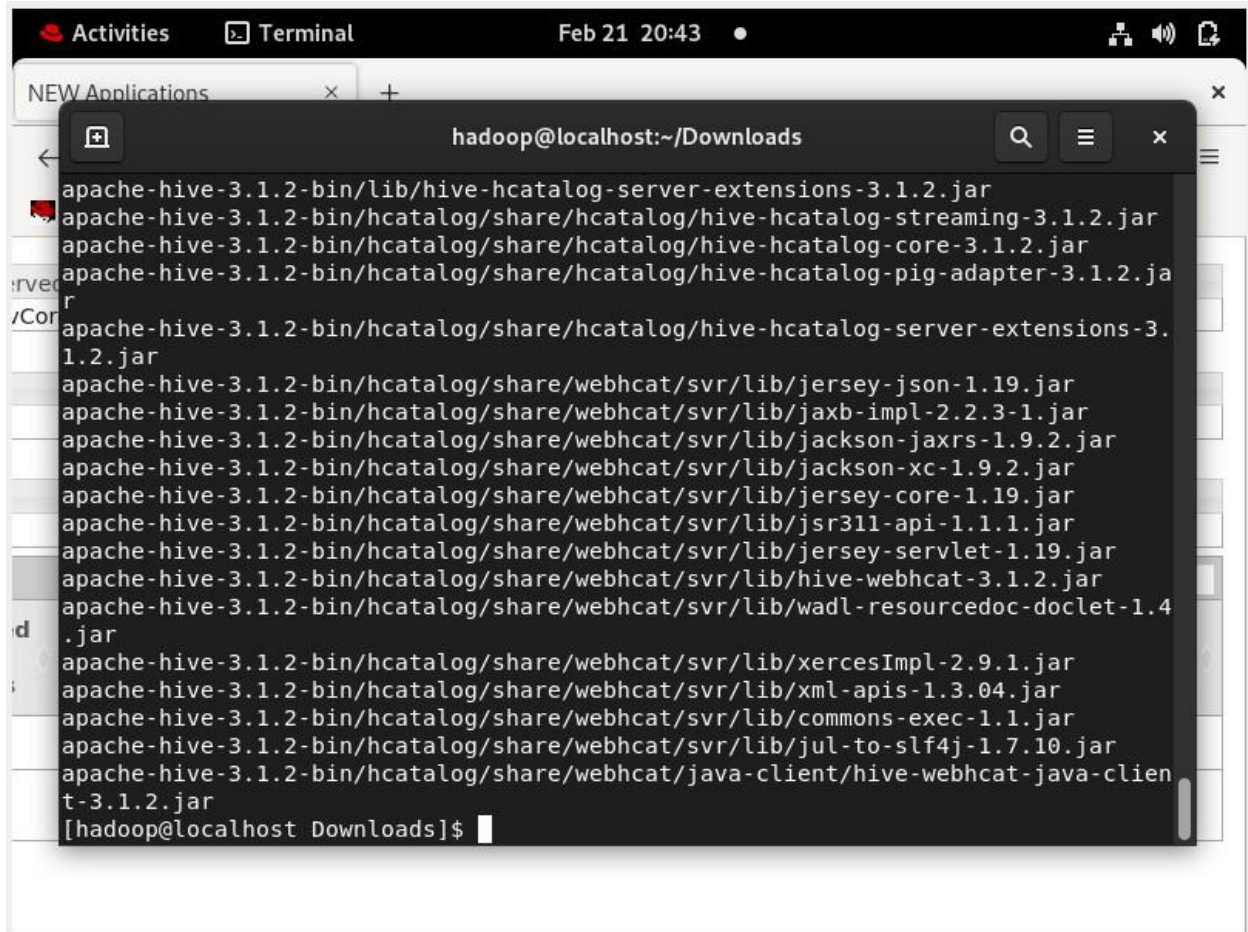
## Task 2: Week 6: Hive

Upload Book.xml in HDFS then create a table in Hive and parse it.

Download the Apache hive from the Apache website and extract the setup file.  
Using the tar command that is tar xzvf apache-hive-3.1.2-bin.tar.gz



```
Activities Terminal Feb 21 20:42
NEW Applications x +
hadoop@localhost:~/Downloads
Starting secondary namenodes [localhost:localhostdomain]
[hadoop@localhost ~]$ start-yarn.sh
Starting resourcemanager
Starting nodemanagers
[hadoop@localhost ~]$ cd Downloads
[hadoop@localhost Downloads]$ ls
apache-hive-3.1.2-bin.tar.gz
[hadoop@localhost Downloads]$ tar zxvf apache-hive-3.1.2-bin.tar.gz
apache-hive-3.1.2-bin/LICENSE
apache-hive-3.1.2-bin/NOTICE
apache-hive-3.1.2-bin/RELEASE_NOTES.txt
apache-hive-3.1.2-bin/binary-package-licenses/asm-LICENSE
apache-hive-3.1.2-bin/binary-package-licenses/com.google.protobuf-LICENSE
apache-hive-3.1.2-bin/binary-package-licenses/com.ibm.icu.icu4j-LICENSE
apache-hive-3.1.2-bin/binary-package-licenses/com.sun.jersey-LICENSE
apache-hive-3.1.2-bin/binary-package-licenses/com.thoughtworks.paranamer-LICENSE
apache-hive-3.1.2-bin/binary-package-licenses/javax.transaction.transaction-api-LICENSE
apache-hive-3.1.2-bin/binary-package-licenses/javolution-LICENSE
apache-hive-3.1.2-bin/binary-package-licenses/jline-LICENSE
apache-hive-3.1.2-bin/binary-package-licenses/NOTICE
apache-hive-3.1.2-bin/binary-package-licenses/org.abego.treelayout.core-LICENSE
apache-hive-3.1.2-bin/binary-package-licenses/org.antlr-LICENSE
apache-hive-3.1.2-bin/binary-package-licenses/org.antlr.antlr4-LICENSE
apache-hive-3.1.2-bin/binary-package-licenses/org.antlr.stringtemplate-LICENSE
```



2. Update the hive location in the bashrc file using export command.

Export HIVE Home=/usr/local/hive

Export PATH=\$PATH:\$HIVE HOME/bin

```
root@localhost:~  
GNU nano 5.6.1 /root/.bashrc Modified  
# .bashrc  
export HIVE_HOME=/usr/local/hive  
export PATH=$PATH:$HIVE_HOME/bin  
export CLASSPATH=$CLASSPATH:/usr/local/Hadoop/lib/*:.  
export CLASSPATH=$CLASSPATH:/usr/local/hive/lib/*:.  
  
# Source global definitions  
if [ -f /etc/bashrc ]; then  
    . /etc/bashrc  
fi  
  
# User specific environment  
if ! [[ "$PATH" =~ "$HOME/.local/bin:$HOME/bin:" ]]  
then  
    PATH="$HOME/.local/bin:$HOME/bin:$PATH"  
fi  
export PATH  
  
# Uncomment the following line if you don't like systemctl's auto-paging feature  
# export SYSTEMD_PAGER=
```

3. Update the hive-env.sh file

```
root@localhost:usr/local/hive/conf
mv: cannot move 'apache-hive-3.1.2-bin' to '/usr/local/hive': Permission denied
[hadoop@localhost Downloads]$ su -
Password:
[root@localhost ~]# mv apache-hive-3.1.2-bin /usr/local/hive
mv: cannot stat 'apache-hive-3.1.2-bin': No such file or directory
[root@localhost ~]# ~/.bashrc
-bash: /root/.bashrc: Permission denied
[root@localhost ~]# vi ~/.bashrc

[1]+  Stopped                  vi ~/.bashrc
[root@localhost ~]# vi ~/.bashrc
[root@localhost ~]# vi ~/.bashrc

[2]+  Stopped                  vi ~/.bashrc
[root@localhost ~]# nano ~/.bashrc
[root@localhost ~]# nano ~/.bashrc
[root@localhost ~]# source ~/.bashrc
[root@localhost ~]# cd $HIVE_HOME/conf
-bash: cd: /usr/local/hive/conf: No such file or directory
[root@localhost ~]# cd $HIVE_HOME/conf
[root@localhost conf]# cp hive-env.sh.template hive-env.sh
[root@localhost conf]# nano hive-env.sh
[root@localhost conf]# nano hive-env.sh
[root@localhost conf]#
```

Timed Out

Derby

4. Install the derby in the hive.

```
Activities Terminal Feb 21 21:20
NEW Applications x Hive - Installation x +
root@localhost:~
[2]+ Stopped vi ~/.bashrc
[root@localhost ~]# nano ~/.bashrc
[root@localhost ~]# nano ~/.bashrc
[root@localhost ~]# source ~/.bashrc
[root@localhost ~]# cd $HIVE_HOME/conf
-bash: cd: /usr/local/hive/conf: No such file or directory
[root@localhost ~]# cd $HIVE_HOME/conf
[root@localhost conf]# cp hive-env.sh.template hive-env.sh
[root@localhost conf]# nano hive-env.sh
[root@localhost conf]# nano hive-env.sh
[root@localhost conf]# cd ~
[root@localhost ~]# wget http://archive.apache.org/dist/db/derby/db-derby-10.4.2
.0/db-derby-10.4.2.0-bin.tar.gz
--2023-02-21 21:19:32-- http://archive.apache.org/dist/db/derby/db-derby-10.4.2
.0/db-derby-10.4.2.0-bin.tar.gz
Resolving archive.apache.org (archive.apache.org)... 2a01:4f8:172:2ec5::2, 138.2
01.131.134
Connecting to archive.apache.org (archive.apache.org)|2a01:4f8:172:2ec5::2|:80..
. connected.
HTTP request sent, awaiting response... 200 OK
Length: 18692878 (18M) [application/x-gzip]
Saving to: 'db-derby-10.4.2.0-bin.tar.gz'
d 8%[>] 1.47M 28.2KB/s eta 6m 53s
Timed Out db-derby-10.4.2.0-bin.tar.gz
```

5. Set the path of derby in the bashrc file using export command.

Export DERBY HOME /usr/local/derby

Export PATH \$PTH:\$DERBY HOME/bin



Activities Terminal Feb 21 21:47

NEW Applications x Hive - Installation x +

root@localhost:~

GNU nano 5.6.1 /root/.bashrc

```
# .bashrc
export HIVE_HOME=/usr/local/hive
export PATH=$PATH:$HIVE_HOME/bin
export CLASSPATH=$CLASSPATH:/usr/local/Hadoop/lib/*:.
export CLASSPATH=$CLASSPATH:/usr/local/hive/lib/*:.

export DERBY_HOME=/usr/local/derby
export PATH=$PATH:$DERBY_HOME/bin

export CLASSPATH=$CLASSPATH:$DERBY_HOME/lib/derby.jar:$DERBY_HOME/lib/derbytool
# Source global definitions
if [ -f /etc/bashrc ]; then
    . /etc/bashrc
fi

# User specific environment
if ! [[ "$PATH" =~ "$HOME/.local/bin:$HOME/bin:" ]]
then
    Read 42 lines ]
    ^G Help    ^O Write Out ^W Where Is  ^K Cut      ^T Execute  ^C Location
    ^X Exit    ^R Read File ^\ Replace   ^U Paste    ^J Justify  ^_ Go To Line
```

Timed Out \$ mkdir \$DERBY\_HOME/data

Right Ctrl

6. Create the book table in the hive.

```

harman@harman-VirtualBox: ~
login as: harman
harman@192.168.43.233's password:
Welcome to Ubuntu 22.04.2 LTS (GNU/Linux 5.8.0-43-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

933 updates can be installed immediately.
0 of these updates are security updates.
To see these additional updates run: apt list --upgradable

New release '22.04.1 LTS' available.
Run 'do-release-upgrade' to upgrade to it.

Last login: Wed Feb 22 05:55:33 2023 from 127.0.0.1
harman@harman-VirtualBox:~$ hivehive> create temporary table if not exists books_xml(line string);
OK
Time taken: 0.932 seconds
hive> load data local inpath 'books.xml' into table book_xml;
Loading data to table default.book_xml
OK
Time taken: 0.935 seconds
hive> Create temporary table if not exists books(
  > author string,
  > title array<string>,
  > genre array<string>,
  > price array<string>,
  > discount array<string>,
  > publishdate array<string>,
  > description array<string>,
  > row format delimited
  > fields terminated by '|';
OK
Time taken: 0.194 seconds

```

Name of students who has participated in the assignment.
Sandeep Kaur
Harmandeep Kaur
Gabriel Adeniyi
Sukhman Kaur