



Habits of Effective Sqoop Users

Kate Ting, Customer Operations Engineer
kate@cloudera.com

cloudera

Halp! Sqoop doesn't work!

Now what?

Agenda

- First Things First
- Common Problems
- MySQL
 - Connection Failure
 - Importing into Hive
- Oracle
 - Case-Sensitive Catalog Query Errors
 - Sqoop Export Failing
- Effective Sqoop Habits

Agenda

- **First Things First**
- Common Problems
- MySQL
 - Connection Failure
 - Importing into Hive
- Oracle
 - Case-Sensitive Catalog Query Errors
 - Sqoop Export Failing
- Effective Sqoop Habits

First Things First

Save time by providing this upfront:

- Versions: Sqoop, Hadoop, OS, JDBC
- Run with `--verbose` flag then attach log
- Sqoop command including options-file
- Expected output vs. actual output
- Table definition
- Input data set that triggers problem
- Hadoop task logs
- Check permissions on input files
- Divide and conquer
 - e.g. Import that creates and populates a Hive table is failing
 - First, do the import alone
 - Second, create a Hive table without the import using the `create-hive-table` tool

Common Problems

Agenda

- First Things First
- Common Problems
- **MySQL**
 - Connection Failure
 - Importing into Hive
- Oracle
 - Case-Sensitive Catalog Query Errors
 - Sqoop Export Failing
- Effective Sqoop Habits

MySQL: Connection Failure

```
java.lang.RuntimeException: java.lang.RuntimeException:  
com.mysql.jdbc.exceptions.jdbc4.CommunicationsException: Communications link failure
```

The last packet sent successfully to the server was 0 milliseconds ago. The driver has not received any packets from the server.

```
at com.cloudera.sqoop.mapreduce.db.DBInputFormat.setConf(DBInputFormat.java:164 )  
at org.apache.hadoop.util.ReflectionUtils.setConf(ReflectionUtils.java:62)  
at org.apache.hadoop.util.ReflectionUtils.newInstance(ReflectionUtils.java:117)  
at org.apache.hadoop.mapred.MapTask.runNewMapper(MapTask.java:606)  
at org.apache.hadoop.mapred.MapTask.run(MapTask.java:323)  
at org.apache.hadoop.mapred.Child$4.run(Child.java:270)  
at java.security.AccessController.doPrivileged(Native Method)  
at javax.security.auth.Subject.doAs(Subject.java:396)  
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1127)  
at org.apache.hadoop.mapred.Child.main(Child.java:264)
```


MySQL: Connection Failure

- Problem: Communications Link Failure caused by incorrect permissions.
- Solution:
 - Verify that you can connect to the database from the node where you are running Sqoop:
 - `$ mysql --host=<IP Address> --database=test --user=<username> --password=<password>`
 - Add the network port for the server to your my.cnf file
 - Set up a user account to connect via Sqoop. Grant permissions to the user to access the database over the network:
 - Log into MySQL as root `mysql -u root -p<ThisIsMyPassword>`
 - Issue the following command: `mysql> grant all privileges on test.* to 'testuser'@'%' identified by 'testpassword'`

MySQL: Importing into Hive

- Troubleshooting tips:
 - Look at `/tmp/${user}/hive.log`
 - Identifies exceptions during the load
 - Look at `/user/hive/warehouse`
 - View contents of the imported data

Agenda

- First Things First
- Common Problems
- MySQL
 - Connection Failure
 - Importing into Hive
- **Oracle**
 - Case-Sensitive Catalog Query Errors
 - Sqoop Export Failing
- Effective Sqoop Habits

Oracle: Case-Sensitive Catalog Query Errors

```
INFO manager.OracleManager: Time zone has been set to GMT
DEBUG manager.SqlManager: Using fetchSize for next query: 1000
INFO manager.SqlManager: Executing SQL statement:
SELECT t.* FROM addlabel_pris t WHERE 1=0
DEBUG manager.OracleManager$ConnCache: Caching
released connection for jdbc:oracle:thin:
ERROR sqoop.Sqoop: Got exception running Sqoop:
java.lang.NullPointerException
java.lang.NullPointerException
at com.cloudera.sqoop.hive.TableDefWriter.getCreateTableStmt(TableDefWriter.java:148)
at com.cloudera.sqoop.hive.HiveImport.importTable(HiveImport.java:187)
at com.cloudera.sqoop.tool.ImportTool.importTable(ImportTool.java:362)
at com.cloudera.sqoop.tool.ImportTool.run(ImportTool.java:423)
at com.cloudera.sqoop.Sqoop.run(Sqoop.java:144)
at org.apache.hadoop.util.ToolRunner.run(ToolRunner.java:65)
at com.cloudera.sqoop.Sqoop.runSqoop(Sqoop.java:180)
at com.cloudera.sqoop.Sqoop.runTool(Sqoop.java:219)
at com.cloudera.sqoop.Sqoop.runTool(Sqoop.java:228)
at com.cloudera.sqoop.Sqoop.main(Sqoop.java:237)
```

Oracle: Case-Sensitive Catalog Query Errors

- Problem: NPE caused by using the wrong case for the user name and table name.
- Solution: Always specify the user and table names in upper case (unless it was created with mixed/lower case within quotes).

Oracle: Sqoop Export Failing

INFO mapred.JobClient: Running job: job_201109231340_0785

INFO mapred.JobClient: map 0% reduce 0%

INFO mapred.JobClient: Task Id :

attempt_201109231340_0785_m_000000_0, Status : FAILED

java.lang.NullPointerException

at

com.cloudera.sqoop.mapreduce.db.DataDrivenDBRecordReader.getSelectQuery(DataDrivenDBRecordReader.java:87)

at com.cloudera.sqoop.mapreduce.db.DBRecordReader.nextKeyValue(DBRecordReader.java:225)

at org.apache.hadoop.mapred.MapTask

\$NewTrackingRecordReader.nextKeyValue(MapTask.java:455)

at org.apache.hadoop.mapreduce.MapContext.nextKeyValue(MapContext.java:67)

Oracle: Sqoop Export Failing

- Problem: `IllegalArgumentException` caused by not non-owner trying to connect to the table.
- Solution: Prefix the table name with the schema, for example `SchemaName.OracleTableName`.

Agenda

- First Things First
- Common Problems
- MySQL
 - Connection Failure
 - Importing into Hive
- Oracle
 - Case-Sensitive Catalog Query Errors
 - Sqoop Export Failing
- **Effective Sqoop Habits**

Effective Sqoop Habits

- Do create an empty export table.
- Don't use the same table for both import and export.

Effective Sqoop Habits

- Do use `--escaped-by` option during import and `--input-escaped-by` during export.
- Do use `fields-terminated-by` during import and `input-fields-terminated-by` during export.
- Don't reverse them.

Effective Sqoop Habits

- Do specify the direct mode option (`--direct`), if you use the direct connector.
- Don't specify the query, if you use the direct connector.

How Do You Eat an Elephant?

- One bite at a time
 - Versions
 - Verbose flag
 - Console log
 - Exact command, etc
- Sqoop Troubleshooting Guide
 - <http://archive.cloudera.com/cdh/3/sqoop/SqoopUserGuide.html#troubleshooting>