

TAJO-176

hyunsik

2013-09-10

Contents

Chapter 1

Root issue TAJO-176

1.1 Summary

Implement Tajo JDBC Driver

1.2 Description

Tajo is required to be integrated with legacy BI or OLAP tools. For this, Tajo should provide a JDBC driver.

1.3 Attachments

1. [TAJO-176_2.patch](#)
2. [TAJO-176_3.patch](#)
3. [TAJO-176.patch](#)

1.4 Commits

1. Commit **342fd47** by **Jihoon Son** (2013-11-18): TAJO-176: Implement Tajo JDBC Driver. (Keuntae Park via jihoon)

1.5 Comments

1. **hjkim:** I think that jdbc package should be separated from the other packages. So how about create sub project named "tajo-jdbc".
2. **jihoonson:** Currently, we try to support other types of metadata stores not only RDBMSs (TAJO-16).
So, if we create a new sub project for the jdbc package, we shall need to create sub projects for each metadata store.
Please explain why we separate the jdbc package from the catalog package.

Thanks,
Jihoon

3. **hjkim:** I think TAJO-16 is about Catalog Server. This issue(TAJO-176) is about Tajo JDBC Driver.
Tajo JDBC Driver is a module that acts as a client. Existing applications that use SQL easily run query to the Tajo using Tajo JDBC Driver.
4. **jihoonson:** Hyoungjun, you are right.
It was my misunderstand.
+1 for your idea.
5. **sirpkt:** I've uploaded the patch for the issue
 1. It implements read and metadata related part of jdbc interface
 - TajoDriver, TajoConnection, TajoResultSet, TajoStatement, TajoPreparedStatement
 - TajoDatabaseMetaData, TajoResultSetMetaData
 - renaming ResultSetImpl to TajoResultSet
 2. TestCase
 - TestTajoJdbc, TestResultSet
 3. Usage
 - Driver Class: org.apache.tajo.jdbc.TajoDriver
 - Connection URL: jdbc:tajo://<tajo master host>:<tajo master client service port>
 - example: Class.forName("org.apache.tajo.jdbc.TajoDriver").newInstance();
 - Connection conn = DriverManager.getConnection("jdbc:tajo://127.0.0.1:26002");
 4. modification in other modules
 - At JDBC connection close time, TajoClient connection should be also closed together, so every request from TajoClient to TajoMaster and QueryMaster is modified to close rpc connection after the request
 5. TODO
 - For jar distribution (like tajo-jdbc.jar), it needs to handle library dependency because TajoClient has dependency on (hdfs-core, hadoop-common, hadoop-auth, yarn-core, tajo-*)
 - sub project separation (maybe as 'tajo-jdbc') - currently implemented in tajo-core-backend
6. **jihoonson:** There are some points which should be discussed.
 - * The first is the different handling of unimplemented functions. While some unimplemented functions return a meaningless value such as 0, false, or "", other functions throw an exception. It will be better to use the same way to handle the unimplemented functions.
 - * There are *too many* wrong exception messages. For example, getMaxColumnsInGroupBy() throws SQLFeatureNotSupportedException("getMaxColumnsInGroupBy") instead of SQLFeatureNotSupportedException("commit").
 - * In this patch, RpcConnectionPool is closed when TajoClient.close() is called. I'm not sure that this is a valid action. If there are two or more TajoClients, what will happen?
7. **hyunsik:** Great job!

In overall, the patch looks great for me. The extensive unit tests are also nice.

In addition to the points that Jihoon mentioned, there are additional things to be improved.

* java.sql.DatabaseMetaData and java.sql.Driver have additional methods in JDK 7. So, the current patch cannot be compiled and work in java 7. It can be solved by overriding

java.sql.DatabaseMetaData and java.sql.Driver of JDK7 without @Override annotation.

* There are many unused imports. You can find them in TajoDriver, TajoResultSetMetaData, TajoResultSetBase, TestTajoJdbc, and TajoResultSet classes.

8. **sirpkt:** Thank you for the kind review, Jihoon and Hyunsik

I've uploaded new patch for the comments:

1. Same way to handle the unimplemented functions
 - Not supported functions throw exception
 - All methods that report supporting capabilities of database functions in DatabaseMetaData return true or false
2. Correcting wrong exception method
 - Exception specifies the name of the method not supported
3. Pool close
 - Concurrent execution of multiple TajoClients does not matter, because TajoClient does not release the connection to the pool but just closes it
 - ** Please, check the test cases in testMultipleConnections and testMultipleConnectionsSequentialClose of TestTajoJdbc
4. Supporting jdk7
 - Implement all the additional methods in jdk7 and replace "@Override annotation" with "//JDK1.7" comment
5. Unused imports
 - Remove all
6. rebase

9. **jihoonson:** Thanks for your comment.
Would you give me a more detailed description of your implementation?
Also, please remove commented out codes in TajoClient.

10. **sirpkt:** Sorry for the ambiguous explanation, Jihoon

I added more explanation:

1. At close() of Connection object in JDBC Driver, it should release all the resource that the object has
2. Connection uses TajoClient object, which creates netty client object through RpcConnectionPool to connect TajoMaster, QueryMaster, and Worker. After the request completion, however, netty connection is still active.
3. To solve the above problem(2), close() of Connection calls close() of TajoClient, which again calls close() of RpcConnectionPool to release all the resource
4. Problem of the above approach(3) is that RpcConnectionPool is implemented as singleton and multiple TajoClients may share it. So, instead of calling releaseConnection() of RpcConnectionPool, calling closeConnection(), which avoids sharing problem.
5. for 4, withRetries() method of ServerCallable is also changed to have two options, release or close
6. It may be better that TajoClient creates netty client by itself instead of using Rpc-

ConnectionPool

And I've uploaded new patch that removes commented out codes in TajoClient

11. **jihoonson:** Thanks for your detailed explanation.
I understand, and agree with you on the point described in 6.
But, I also agree with that this is a good temporal solution.

+1 for your patch.
12. **hyunsik:** +1

Jihoon,
Could you commit this patch?
13. **jihoonson:** Sure, I'm going to commit, now.
14. **jihoonson:** I've just committed it.
Thanks for your contribution, Keuntae!
15. **sirpkt:** Great!!
Thank you for your kind review, Jihoon :)
16. **hyunsik:** Fixing it as resolved.
17. **hudson:** SUCCESS: Integrated in Tajo-trunk-postcommit #557 (See [<https://builds.apache.org/job/Tajo-trunk-postcommit/557/>])
TAJO-176: Implement Tajo JDBC Driver. (Keuntae Park via jihoon) (jihoonson: <https://git-wip-us.apache.org/repos/asf?p=incubator-tajo.git&a=commit&h=342fd47ffd2a30f4941256ac4a5bc95707004599>)
* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/client/TajoClient.java
* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/client/ResultSetMetaDataImpl.java
* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/jdbc/TajoDatabaseMetaData.java
* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/jdbc/TajoDriver.java
* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/jdbc/TajoConnection.java
* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/jdbc/TajoResultSetMetaData.java
* CHANGES.txt
* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/jdbc/TajoResultSetBase.java
* tajo-rpc/src/main/java/org/apache/tajo/rpc/ServerCallable.java
* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/jdbc/TestResultSet.java
* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/query/TestGroupByQuery.java
* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/jdbc/TestTajoJdbc.java
* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/jdbc/TajoStatement.java
* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/client/ResultSetImpl.java
* tajo-rpc/src/main/java/org/apache/tajo/rpc/NettyClientBase.java
* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/query/TestResultSetImpl.java
* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/jdbc/TajoResultSet.java
* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/jdbc/MetaDataTuple.java
* tajo-rpc/src/main/java/org/apache/tajo/rpc/RpcConnectionPool.java
* tajo-core/tajo-core-backend/src/test/java/org/apache/tajo/engine/query/TestSortQuery.java
* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/jdbc/TajoPreparedStatement.java
* tajo-core/tajo-core-backend/src/main/java/org/apache/tajo/jdbc/TajoMetaDataResultSet.java

1.6 Pull requests

No pull requests