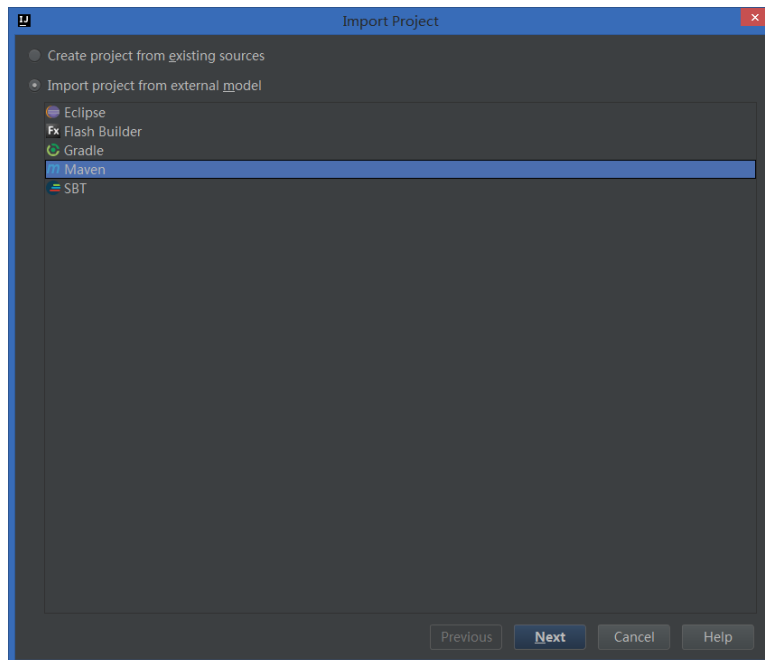
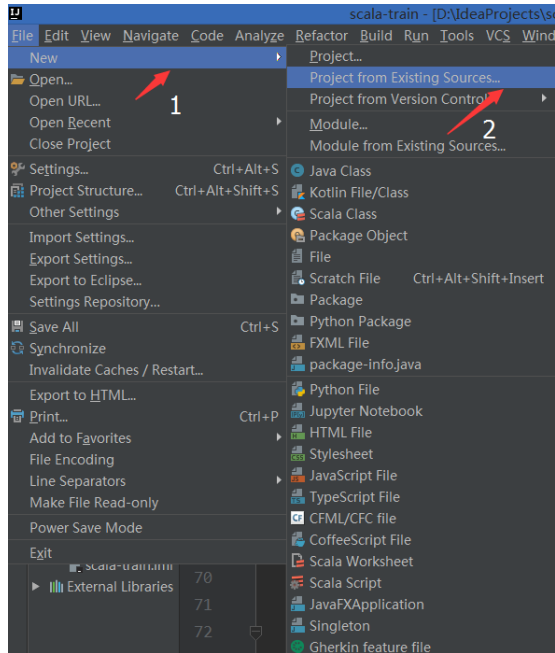
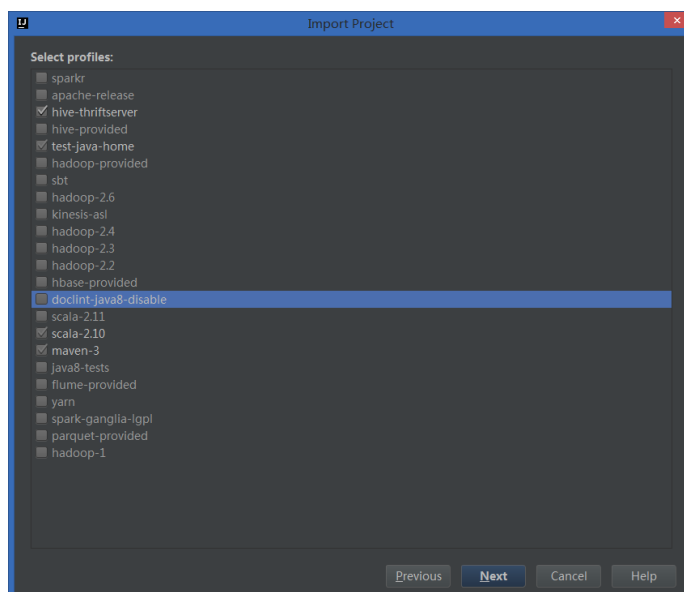
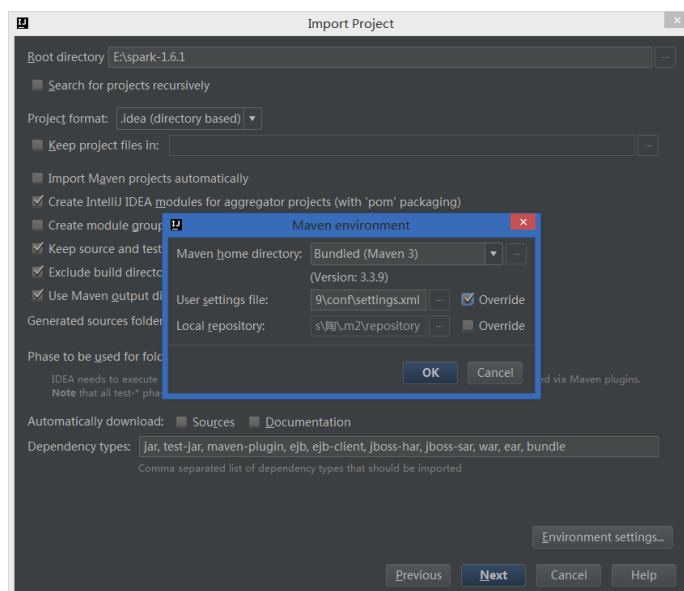
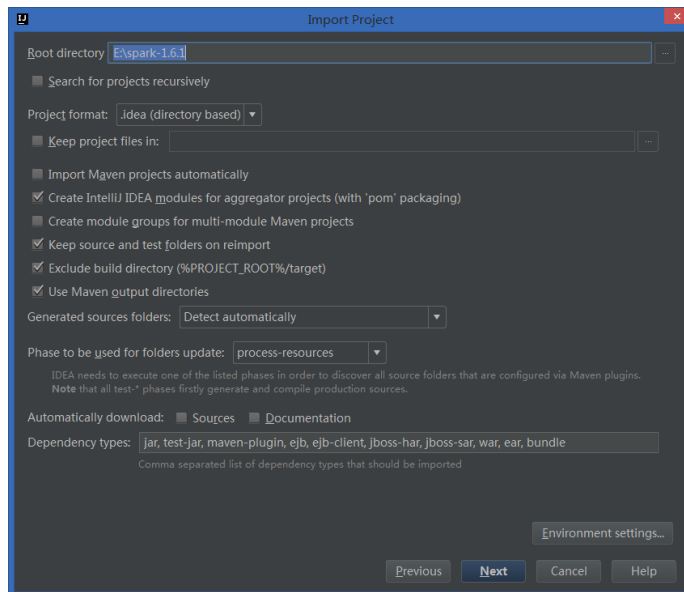
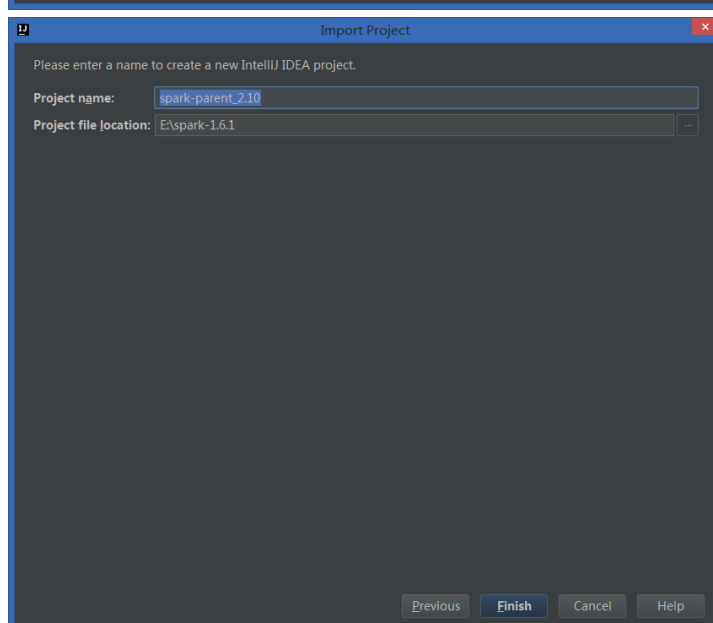
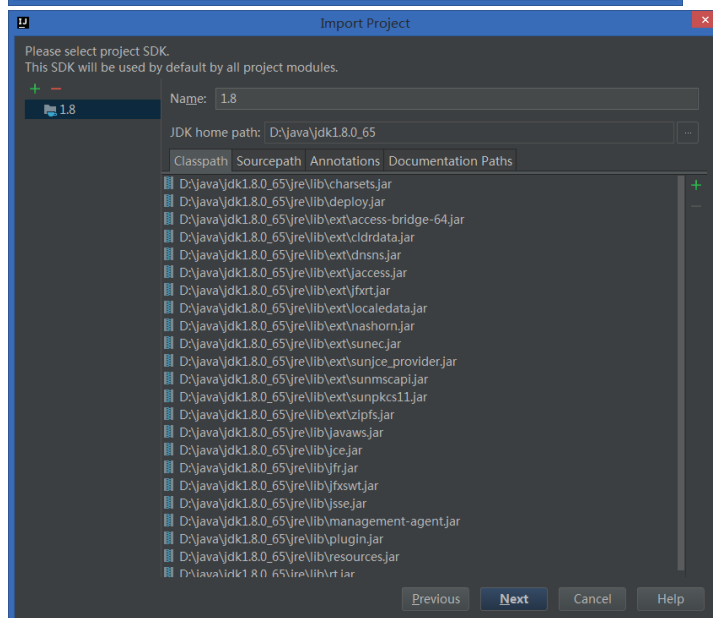
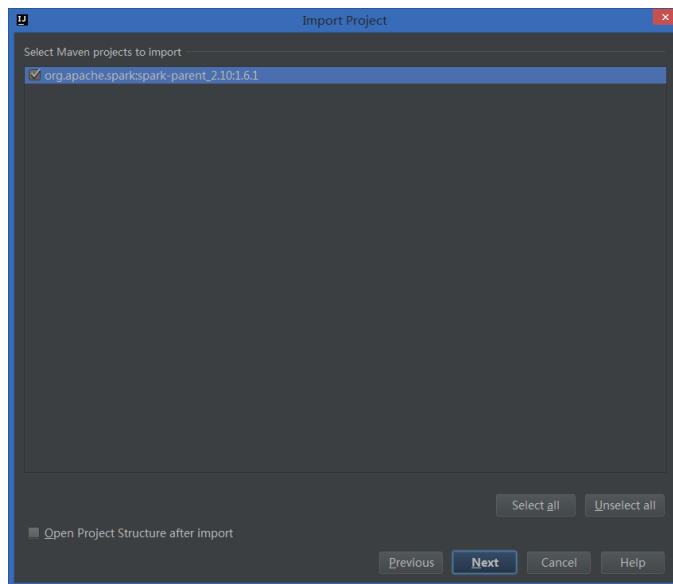


Spark SQL Debug

IntelliJ IDEA 导入 Spark 源码

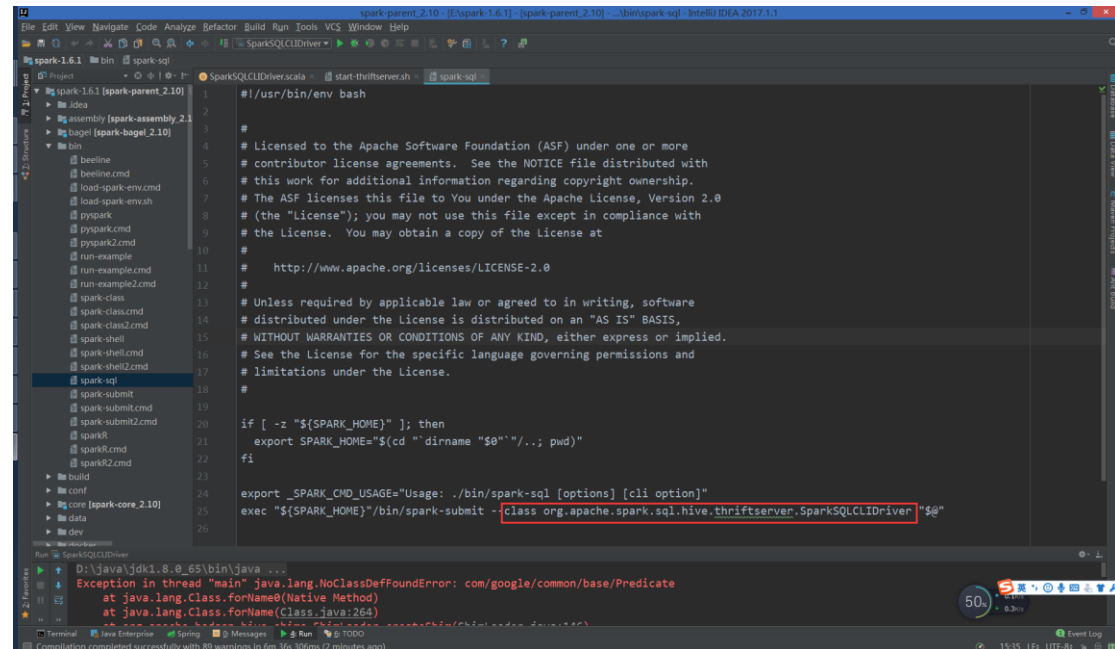






观察发现 Spark SQL 入口类

Spark SQL 进行 debug 的入口类为：SparkSQLCLIDriver.scala
通过 bin/spark-sql 进行观察可以知道：

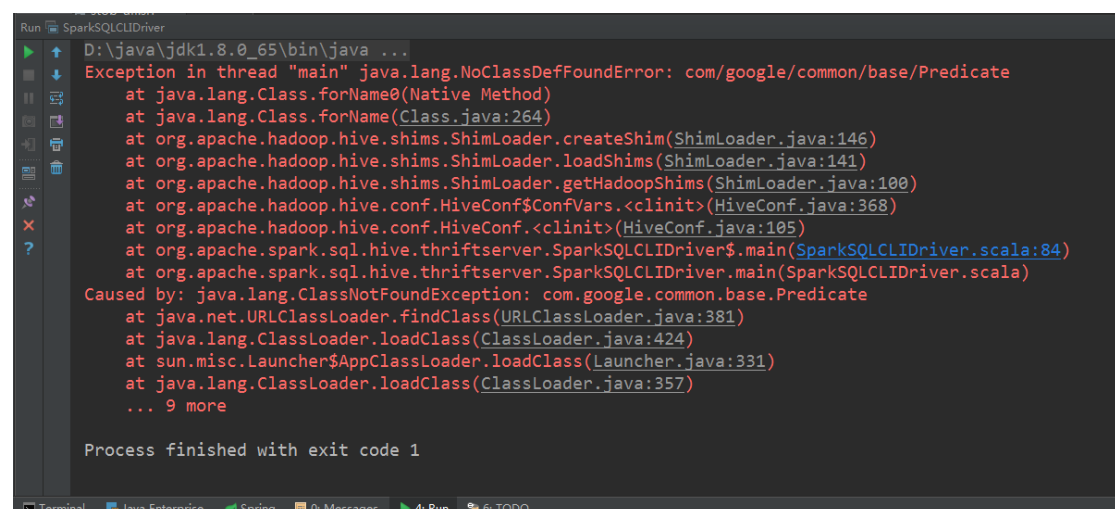


```
1 #!/usr/bin/env bash
2
3 #
4 # Licensed to the Apache Software Foundation (ASF) under one or more
5 # contributor license agreements. See the NOTICE file distributed with
6 # this work for additional information regarding copyright ownership.
7 # The ASF licenses this file to You under the Apache License, Version 2.0
8 # (the "License"); you may not use this file except in compliance with
9 # the License. You may obtain a copy of the License at
10 #
11 # http://www.apache.org/licenses/LICENSE-2.0
12 #
13 # Unless required by applicable law or agreed to in writing, software
14 # distributed under the License is distributed on an "AS IS" BASIS,
15 # WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied.
16 # See the license for the specific language governing permissions and
17 # limitations under the License.
18 #
19
20 if [ -z "${SPARK_HOME}" ]; then
21   export SPARK_HOME="$(cd "$(dirname "$0")" && pwd)"
22 fi
23
24 export _SPARK_CMD_USAGE="Usage: ./bin/spark-sql [options] [cli option]"
25 exec "${SPARK_HOME}"/bin/spark-submit --class org.apache.spark.sql.hive.thriftserver.SparkSQLCLIDriver "$@"
```

踩坑及排错

在 SparkSQLCLIDriver.scala 运行

报错 1

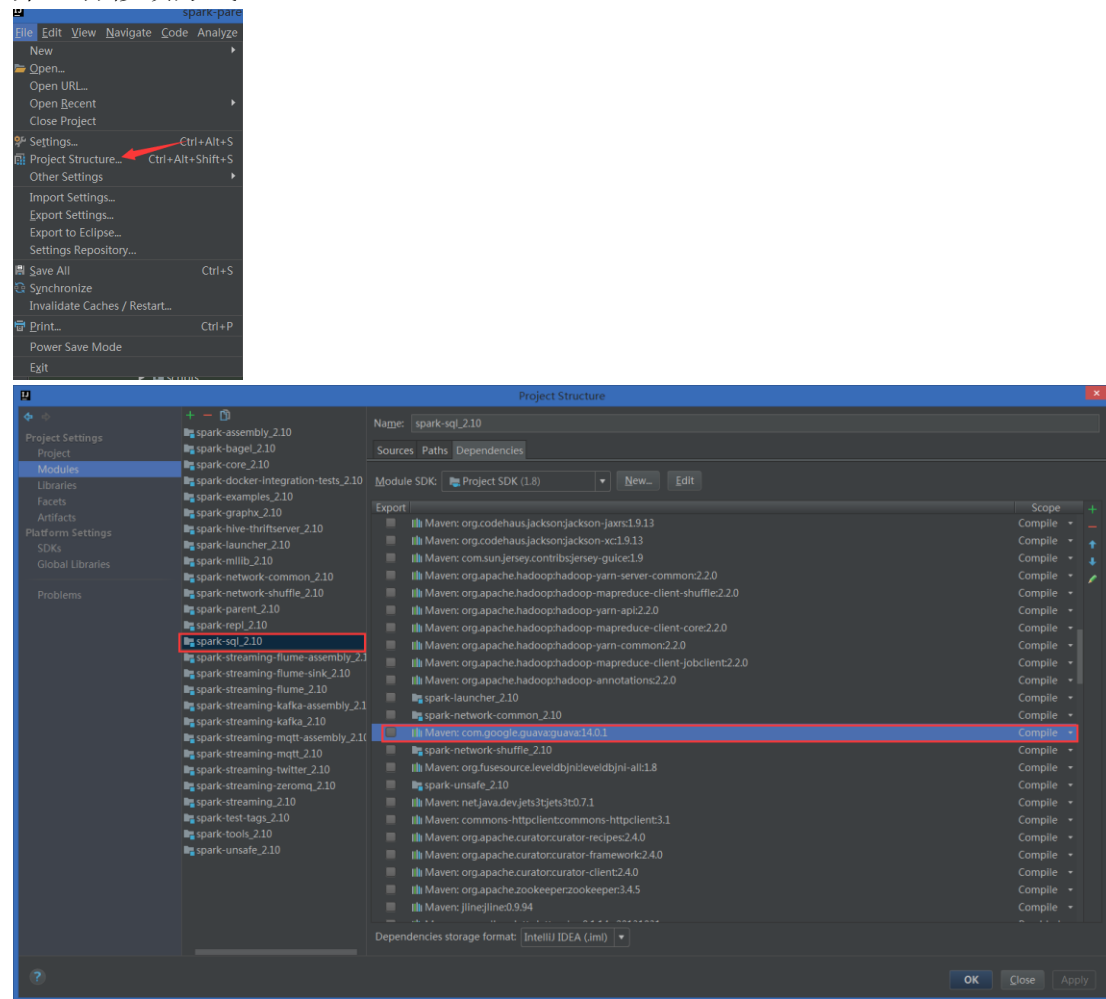


```
Run SparkSQLCLIDriver
D:\java\jdk1.8.0_65\bin\java ...
Exception in thread "main" java.lang.NoClassDefFoundError: com/google/common/base/Predicate
    at java.lang.Class.forName0(Native Method)
    at java.lang.Class.forName(Class.java:264)
    at org.apache.hadoop.hive.shims.ShimLoader.createShim(ShimLoader.java:146)
    at org.apache.hadoop.hive.shims.ShimLoader.loadShims(ShimLoader.java:141)
    at org.apache.hadoop.hive.shims.ShimLoader.getHadoopShims(ShimLoader.java:100)
    at org.apache.hadoop.hive.conf.HiveConf$ConfVars.<clinit>(HiveConf.java:368)
    at org.apache.hadoop.hive.conf.HiveConf.<clinit>(HiveConf.java:105)
    at org.apache.spark.sql.hive.thriftserver.SparkSQLCLIDriver$.main(SparkSQLCLIDriver.scala:84)
    at org.apache.spark.sql.hive.thriftserver.SparkSQLCLIDriver.main(SparkSQLCLIDriver.scala)
Caused by: java.lang.ClassNotFoundException: com.google.common.base.Predicate
    at java.net.URLClassLoader.findClass(URLClassLoader.java:381)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:424)
    at sun.misc.Launcher$AppClassLoader.loadClass(Launcher.java:331)
    at java.lang.ClassLoader.loadClass(ClassLoader.java:357)
    ... 9 more

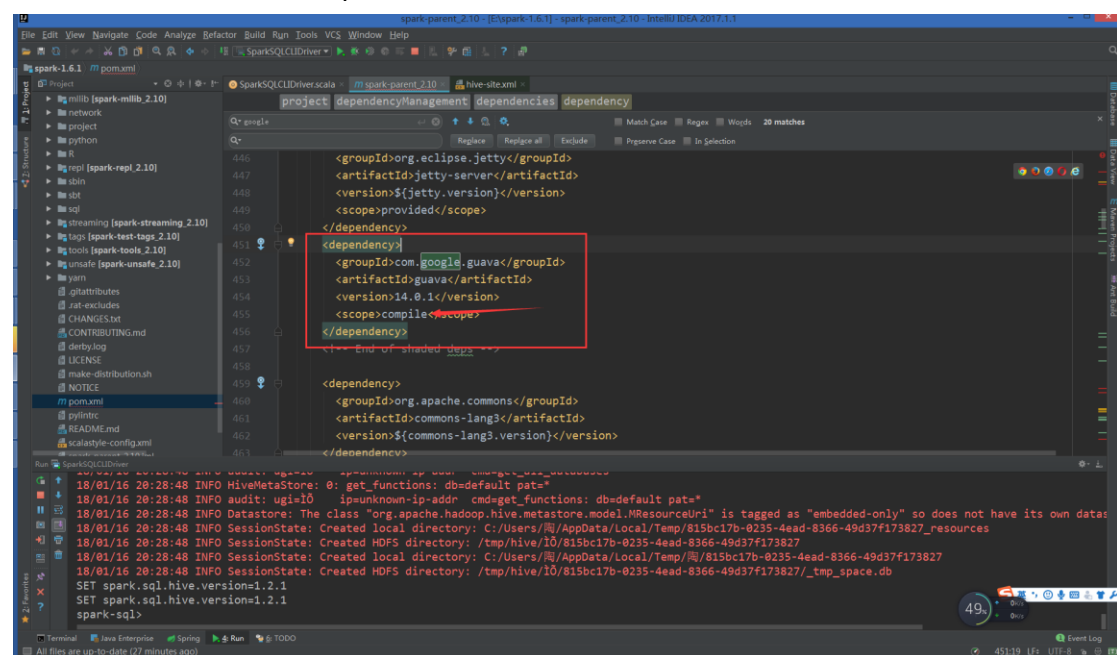
Process finished with exit code 1
```

解决方案

第 1 种修改方式:



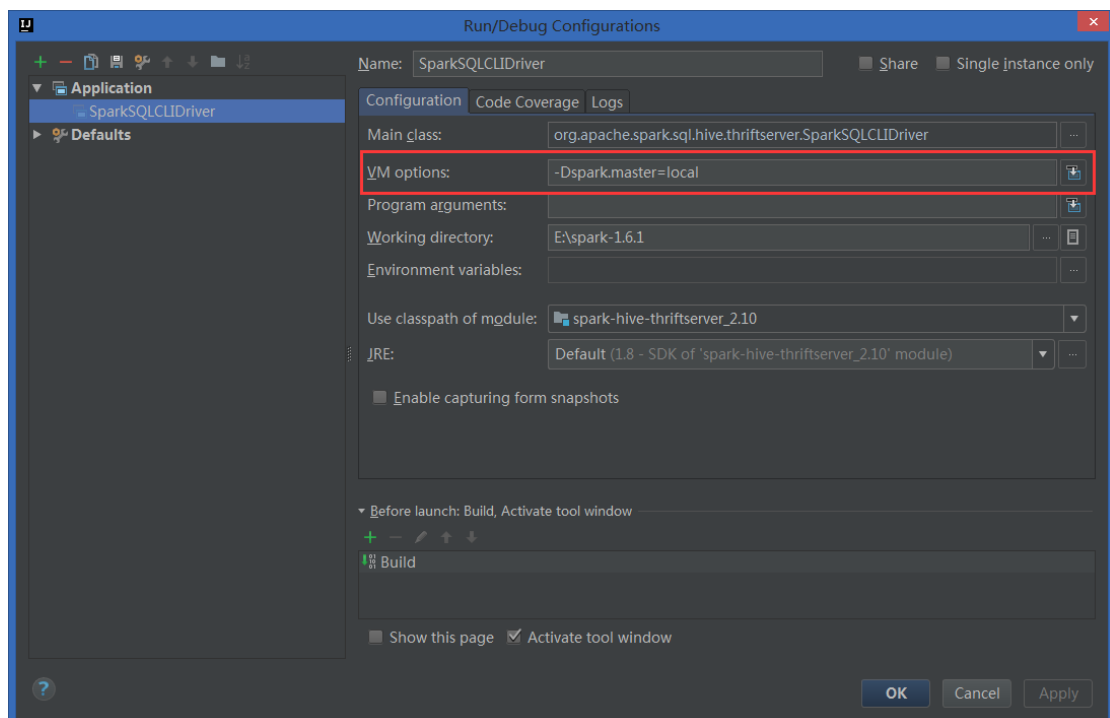
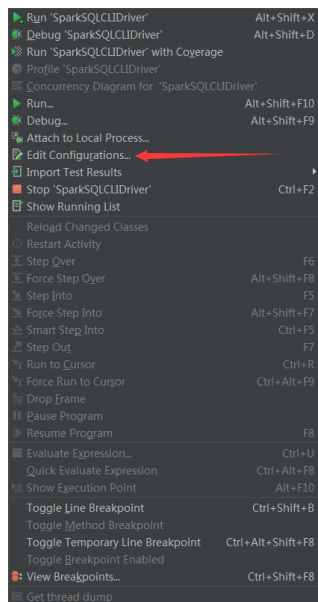
第 2 种修改方式，修改 pom.xml:



报错 2

```
D:\java\jdk1.8.0_65\bin\java ...
log4j:WARN No appenders could be found for logger (org.apache.hadoop.conf.Configuration).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
Using Spark's default log4j profile: org/apache/spark/log4j-defaults.properties
18/01/16 18:01:31 INFO SparkContext: Running Spark version 1.6.1
18/01/16 18:01:31 ERROR SparkContext: Error initializing SparkContext.
org.apache.spark.SparkException: A master URL must be set in your configuration
    at org.apache.spark.SparkContext.<init>(SparkContext.scala:401)
    at org.apache.spark.sql.hive.thriftserver.SparkSQLEnv$.init(SparkSQLEnv.scala:56)
    at org.apache.spark.sql.hive.thriftserver.SparkSQLCLIDriver.<init>(SparkSQLCLIDriver.scala:267)
```

解决方案:
设置 configuration



成功 Debug Spark SQL

The screenshot shows an IDE with the SparkSQLCLIDriver code open. The code is in Scala and includes comments in Chinese. The logs at the bottom show the execution of the driver, including the creation of the HiveMetaStore, the creation of the local directory, and the execution of the SQL query.

```
SparkSQLCLIDriver main(args: Array[String])
153 if (sessionState.database != null) {
154   SparkSQLEnv.hiveContext.runSqlHive(s"USE ${sessionState.database}")
155 }
156
157 // Execute -i init files (always in silent mode)
158 cli.processInitFiles(sessionState)
159
160 if (sessionState.execString != null) {
161   System.exit(cli.processLine(sessionState.execString))
162 }
163
164 try {
165   if (sessionState.fileName != null) {
166     System.exit(cli.processFile(sessionState.fileName))
167   }
168 } catch {
169   case e: FileNotFoundException =>
170     logError(s"Could not open input file for reading. (${e.getMessage})")
171     System.exit(3)
172 }
173
```

Logs:

```
18/01/16 20:28:48 INFO HiveMetaStore: 0: get_functions: db=default pate=
18/01/16 20:28:48 INFO audit: ugi=i0 ip=unknown-ip=addr cmd=get_functions: db=default pate=
18/01/16 20:28:48 INFO Datastore: The class "org.apache.hadoop.hive.metastore.model.HResource" is tagged as "embedded-only" so does not have its own datas
18/01/16 20:28:48 INFO SessionState: Created local directory: C:/Users/周/AppData/Local/Temp/815bc17b-0235-4ead-8366-49d37f173827_resources
18/01/16 20:28:48 INFO SessionState: Created local directory: C:/Users/周/AppData/Local/Temp/815bc17b-0235-4ead-8366-49d37f173827
18/01/16 20:28:48 INFO SessionState: Created local directory: C:/Users/周/AppData/Local/Temp/815bc17b-0235-4ead-8366-49d37f173827/_tmp_space.db
SET spark.sql.hive.version=1.2.1
SET spark.sql.hive.version=1.2.1
spark-sql>
```

