Spark insertIntoJDBC找不到Mysql驱动解决方法

这里说明一点：本文提到的解决Spark insertIntoJDBC找不到Mysql驱动的方法是针对单机模式(也就是local模式)。在集群环境下，下面的方法是不行的。这是因为在分布式环境下，加载mysql驱动包存在一个Bug,1.3及以前的版本 --jars
分发的jar在executor端是通过Spark自身特化的classloader加载的。而JDBC driver manager使用的则是系统默认的classloader，因此无法识别。可行的方法之一是在所有 executor 节点上预先装好JDBC driver并放入默认的classpath。

不过Spark 1.4应该已经fix了这个问题，即 --jars 分发的 jar 也会纳入 YARN 的 classloader 范畴。

今天在使用Spark中DataFrame往Mysql中插入RDD，但是一直报出以下的异常次信息：

```
[itelbog@iteblog ~]$  bin/spark-submit --master local[2]
 --jars lib/mysql-connector-java-5.1.35.jar
 --class  spark.sparkToJDBC ./spark-test_2.10-1.0.jar

spark assembly has been built with Hive, including Datanucleus jars on classpath
Exception in thread "main" java.sql.SQLException: No suitable driver found for
jdbc:mysql://www.iteblog.com:3306/spark?user=root&password=123&useUnicode=
true&characterEncoding=utf8&autoReconnect=true
 at java.sql.DriverManager.getConnection(DriverManager.java:602)
 at java.sql.DriverManager.getConnection(DriverManager.java:207)
 at org.apache.spark.sql.DataFrame.createJDBCTable(DataFrame.scala:1189)
 at spark.SparkToJDBC$.toMysqlFromJavaBean(SparkToJDBC.scala:20)
 at spark.SparkToJDBC$.main(SparkToJDBC.scala:47)
 at spark.SparkToJDBC.main(SparkToJDBC.scala)
 at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
 at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:39)
 at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:25)
 at java.lang.reflect.Method.invoke(Method.java:597)
 at org.apache.spark.deploy.SparkSubmit$.org$apache$spark$deploy$SparkSubmit$
$runMain(SparkSubmit.scala:569)
 at org.apache.spark.deploy.SparkSubmit$.doRunMain$1(SparkSubmit.scala:166)
 at org.apache.spark.deploy.SparkSubmit$.submit(SparkSubmit.scala:189)
 at org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:110)
 at org.apache.spark.deploy.SparkSubmit.main(SparkSubmit.scala)
```

感觉很奇怪，我在启动作业的时候加了Mysql驱动啊在，怎么会出现这种异常呢？？经过查

找，发现在--jars参数里面加入Mysql是没有用的。通过查找，发现提交的作业可以通过加入--driver-class-path参数来设置driver的classpath，试了一下果然没有出现错误！

```
[itelbog@iteblog ~]$  bin/spark-submit --master local[2]
 --driver-class-path lib/mysql-connector-java-5.1.35.jar
 --class  spark.SparkToJDBC ./spark-test_2.10-1.0.jar
```

其实，我们还可以在spark安装包的conf/spark-env.sh通过配置SPARK_CLASSPATH来设置driver的环境变量，如下：

```
export SPARK_CLASSPATH=$SPARK_CLASSPATH:/iteblog/com/mysql-connector-java-5.1.35.jar
```

这样也可以解决上面出现的异常。但是，我们不能同时在conf/spark-env.sh里面配置SPARK_CLASSPATH和提交作业加上--driver-class-path参数，否则会出现以下异常：

```
[itelbog@iteblog ~]$  bin/spark-submit --master local[2]
 --driver-class-path lib/mysql-connector-java-5.1.35.jar
 --class  spark.SparkToJDBC ./spark-test_2.10-1.0.jar

Spark assembly has been built with Hive, including Datanucleus jars on classpath
Exception in thread "main" org.apache.spark.SparkException:
 Found both spark.driver.extraClassPath and SPARK_CLASSPATH. Use only the former.
 at org.apache.spark.SparkConf$$anonfun$validateSettings$6$$anonfun$apply
$7.apply(SparkConf.scala:339)
 at org.apache.spark.SparkConf$$anonfun$validateSettings$6$$anonfun$apply
$7.apply(SparkConf.scala:337)
 at scala.collection.immutable.List.foreach(List.scala:318)
 at org.apache.spark.SparkConf$$anonfun$validateSettings$6.apply(SparkConf.scala:337)
 at org.apache.spark.SparkConf$$anonfun$validateSettings$6.apply(SparkConf.scala:325)
 at scala.Option.foreach(Option.scala:236)
 at org.apache.spark.SparkConf.validateSettings(SparkConf.scala:325)
 at org.apache.spark.SparkContext.<init>(SparkContext.scala:197)
 at spark.SparkToJDBC$.main(SparkToJDBC.scala:41)
 at spark.SparkToJDBC.main(SparkToJDBC.scala)
 at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
 at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:39)
 at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:25)
 at java.lang.reflect.Method.invoke(Method.java:597)
 at org.apache.spark.deploy.SparkSubmit$.org$apache$spark$
  deploy$SparkSubmit$$runMain(SparkSubmit.scala:569)
```

```
at org.apache.spark.deploy.SparkSubmit$.doRunMain$1(SparkSubmit.scala:166)
at org.apache.spark.deploy.SparkSubmit$.submit(SparkSubmit.scala:189)
at org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:110)
at org.apache.spark.deploy.SparkSubmit.main(SparkSubmit.scala)
```

本博客文章除特别声明，全部都是原创！
原创文章版权归过往记忆大数据（过往记忆）所有，未经许可不得转载。
本文链接: 【】（）