

## Spark 1.1.0中使用Hive注意事项

Spark 1.1.0中兼容大部分Hive特性，我们可以在Spark中使用Hive。但是默认的Spark发行版本并没有将Hive相关的依赖打包进spark-assembly-1.1.0-hadoop2.2.0.jar文件中，官方对此的说明是：

Spark SQL also supports reading and writing data stored in Apache Hive. However, since Hive has a large number of dependencies, it is not included in the default Spark assembly

所以，如果你直接使用官方提供的预编译好的Spark发行版，将会出现以下的错误：

```
Exception in thread "main" java.lang.NoClassDefFoundError:
    org/apache/spark/sql/hive/api/java/JavaHiveContext
at SparkHiveTest.main(SparkHiveTest.java:20)
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$.launch(SparkSubmit.scala:328)
at org.apache.spark.deploy.SparkSubmit$.main(SparkSubmit.scala:75)
at org.apache.spark.deploy.SparkSubmit.main(SparkSubmit.scala)
Caused by: java.lang.ClassNotFoundException:
    org.apache.spark.sql.hive.api.java.JavaHiveContext
at java.net.URLClassLoader$1.run(URLClassLoader.java:202)
at java.security.AccessController.doPrivileged(Native Method)
at java.net.URLClassLoader.findClass(URLClassLoader.java:190)
at java.lang.ClassLoader.loadClass(ClassLoader.java:307)
at java.lang.ClassLoader.loadClass(ClassLoader.java:248)
... 8 more
```

解决办法就是重新编译Spark。可以用下面的两种方法解决：

### 一、如果你是Maven用户，可以用下面办法重新编译

```
$ mvn clean package -DskipTests -Pyarn -Phadoop-2.2 -Phive
```

## 二、如果你是SBT用户，可以用下面办法重新编译

```
$sbt/sbt -Phive assembly/assembly
```

用上面的命令编译完的Spark中的lib包将会多出现以下三个jar：

```
-rw-r--r-- 1 wyp iteblog 337012 Sep 25 16:28 datanucleus-api-jdo-3.2.1.jar  
-rw-r--r-- 1 wyp iteblog 1801810 Sep 25 16:28 datanucleus-core-3.2.2.jar  
-rw-r--r-- 1 wyp iteblog 1769726 Sep 25 16:28 datanucleus-rdbms-3.2.1.jar
```

而且在spark-assembly-1.1.0-hadoop2.2.0.jar中也有Hive相关的类，比如JavaHiveContext。  
这时候，你再去运行程序，在日志中你可以看到以下一行的输出：

```
Spark assembly has been built with Hive, including Datanucleus jars on classpath
```

这就说明你的Spark已经将Hive相关的包打进去了。

**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：[【】（）](#)