



专注于商业智能BI和大数据的垂直社区平台

# Spark初级课程

讲师：平常心

# HiveContext

Spark SQL 支持Hive数据的读写(前提是spark编译的时候通过-Phive和-Phive-thriftserver标记来启用hive支持)。通过创建HiveContext(HiveContext继承自SQLContext,所以除了使用sql()方法外,还有hql()方法,用Hive语法编译sql)。

先将hive-site.xml复制到\$SPARK\_HOME/conf, 然后引用spark-hive的jar包, 使用示例:

```
// sc is an existing SparkContext.  
val sqlContext = new org.apache.spark.sql.hive.HiveContext(sc)  
sqlContext.sql("CREATE TABLE IF NOT EXISTS src (key INT, value STRING)")  
sqlContext.sql("LOAD DATA LOCAL INPATH 'examples/src/main/resources/kv1.txt' INTO TABLE src")  
  
// Queries are expressed in HiveQL  
sqlContext.sql("FROM src SELECT key, value").collect().foreach(println)
```

# 数据写入Hive表

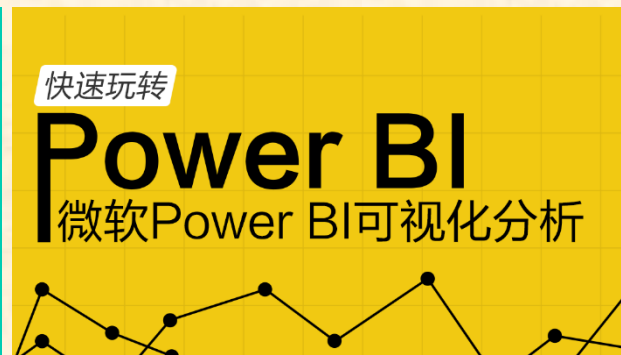
DataFrame通过write的saveAsTable将数据保存到Hive表中，不同于registerTempTable，saveAsTable是会将DataFrame中的数据物化到Hive表,属于Hive的内部表。registerTempTable只是注册一个临时的表，只要Spark Application停止，表就不存在。而saveAsTable创建的是物化的表，无论Spark Application重启或者停止，表都会一直存在。

调用HiveContext.table()方法，还可以直接针对Hive中的表，创建一个DataFrame。

注意SQLContext, HiveContext调用sql方法作用域，即使用HiveContext.sql创建的DataFrame,在之后的sql操作中能被使用，不能被SQLContext使用。同理， SQLContext.sql创建的DataFrame，在之后的sql操作中只供SQLContext使用。



更多商业智能BI和大数据精品视频尽在 [www.hellobi.com](http://www.hellobi.com)



BI、商业智能  
数据挖掘 大数据  
数据分析  
R Python  
机器学习  
Tableau  
QLIKVIEW  
Hive Hadoop  
BIWORK  
BAO胖子 seng  
曹浩 贝克汉姆