



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

# 共享变量

Spark重要的功能特性之一是共享变量。通常情况下，Spark的算子使用外部变量，该外部变量的值会被复制到节点执行的每个task，每个task只能操作自己复制的变量副本。如果能让节点的所有task的能够共享该变量，Spark提供了两种方式：

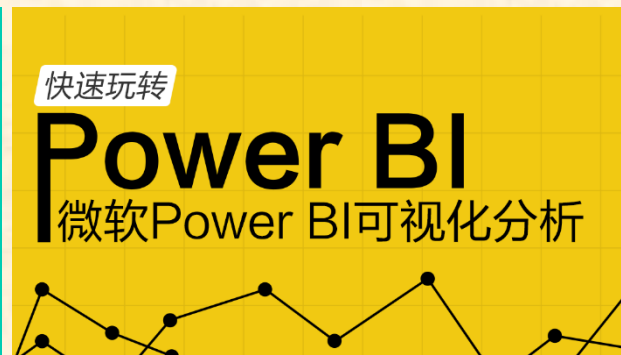
1.广播变量（Broadcast Variables）：将使用的变量仅仅为每个节点复制一份，更大的用处是优化性能，减少网络传输和内存消耗。

调用SparkContext的broadcast()方法，来针对某个变量创建广播变量，每个节点可以使用广播变量的value()方法获取值。

2.累加变量（Accumulators）：让多个task共同操作变量，主要可以进行累加操作，不能读取值，只有driver能读取值。



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