

Apache Spark 2.0.1稳定版正式发布

今天凌晨（2016-10-05）Apache Spark 2.0.1稳定版正式发布。Apache Spark 2.0.1是一个维护版本，一共处理了300个Issues，推荐所有使用Spark 2.0.0的用户升级到此版本。Apache Spark 2.0为我们带来了许多新的功能：

- DataFrame和Dataset统一（可以参见[《Spark 2.0技术预览：更容易、更快速、更智能》](https://www.iteblog.com/archives/1668.html)）：<https://www.iteblog.com/archives/1668.html>
- SparkSession：一个新的切入点。（SparkSession的使用可以参见[《SparkSession：新的切入点》](https://www.iteblog.com/archives/1682.html)：<https://www.iteblog.com/archives/1682.html>和[《Spark 2.0介绍：SparkSession创建和使用相关API》](https://www.iteblog.com/archives/1673.html)：<https://www.iteblog.com/archives/1673.html>）；、
- 简单以及性能更好的accumulator API；
- Spark 2.0大幅提升了SQL功能，并支持SQL2003。Spark SQL现在可以运行所有的99 TPC-DS查询。支持ANSI-SQL和Hive SQL的内置SQL解析器；内置实现了DDL命令；支持子查询，包括
 - 不相关的标量子查询(Unrelated Scalar Subqueries)；
 - 相关的标量子查询(Correlated Scalar Subqueries)；
 - NOT IN谓词子查询（在WHERE/HAVING语句中）；
 - IN谓词子查询中（在WHERE/HAVING语句中）；
 - (NOT) EXISTS谓词子查询中（在WHERE/HAVING语句中）。
- 支持视图规范化；
- 其他性能的提升。

虽然有这么多新的特性和性能的提升，但是此前的Spark 2.0.0却不是稳定版，没多少人敢直接在生产环境下使用；今天发布的Spark 2.0.1终于可以让我们大胆地在生产环境下使用了，还不赶紧试试。

相关链接

- Apache Spark 2.0.1 Release Notes：[戳我](#)
- Apache Spark 2.0.1下载地址：[戳我](#)
- Apache Spark 2.0.1最新文档：[戳我](#)
- Spark 2.0相关文章：<https://www.iteblog.com/archives/tag/spark-2-0/>

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：[【】（）](#)