

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：

1、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)）：<https://www.iteblog.com/archives/1668>

2、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)：<https://www.iteblog.com/archives/1682>和[《Spark 2.0介绍：SparkSession创建和使用相关API》](https://www.iteblog.com/archives/1673)：<https://www.iteblog.com/archives/1673>）；

3、简单以及性能更好的accumulator API；

4、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（1）、不相关的标量子查询(Unrelated Scalar Subqueries)；（2）、相关的标量子查询(Correlated Scalar Subqueries)；（3）、NOT IN谓词子查询（在WHERE/HAVING语句中）；（4）、IN谓词子查询中（在WHERE/HAVING语句中）；（5）、(NOT) EXISTS谓词子查询中（在WHERE/HAVING语句中）。

5、支持视图规范化；

6、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>



优秀人才不缺工作机会，只缺适合自己的好机会。但是他们往往没有精力从海量机会中找到最适

合的那个。

100offer 会对平台上的人才和企业进行严格筛选，让「最好的人才」和「最好的公司」相遇。

注册 100offer，谈谈你对下一份工作的期待。一周内，收到 5-10 个满足你要求的好机会！

本博客文章除特别声明，全部都是原创！

禁止个人和公司转载本文、谢谢理解：过往记忆（<https://www.iteblog.com/>）

本文链接:【】（）