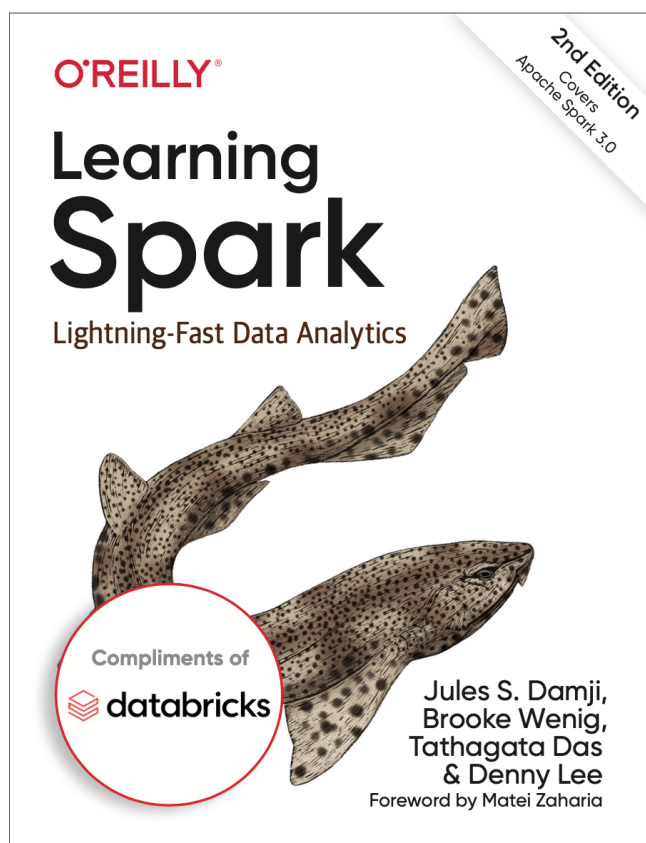


[Learning Spark, 2nd Edition 可以免费下载了](#)

《Learning Spark, 2nd Edition》这本书是由 O'Reilly Media 出版社于2020年7月出版的，作者包括 Jules S. Damji, Brooke Wenig, Tathagata Das, Denny Lee。



如果想及时了解 Spark、Hadoop 或者 Hbase 相关的文章，欢迎关注微信公共帐号：iteblog_hadoop

图书介绍

第二版已更新包含了 Spark 3.0 的一些东西，本书向数据工程师和数据科学家展示了 Spark 中结构化和统一性为何如此重要。本书分为12章，介绍了 Apache Spark 统一分析引擎的原有；Spark Structured APIs；在 Spark SQL 和 DataFrames 中如何使用内置和外置的数据源；Spark SQL and Datasets 介绍；优化 Spark 应用程序；Structured Streaming 介绍；使用 Apache Spark 构建可靠的数据湖；机器学习相关知识；最后一章介绍了 Apache Spark 3.0 的东西。通过本书可以比较全面的了解 Spark 的核心概念等。具体如下：

- 学习 Python, SQL, Scala 或 Java 高层次 Structured APIs : DataFrames 和 Datasets
- 理解 Spark 算子以及 SQL 引擎

- 通过 Spark 配置和 Spark UI 检查、调优和调试 Spark 算子
- 使用: JSON, Parquet, CSV, Avro, ORC, Hive, S3, 或 Kafka 数据源
- 使用 Structured Streaming 对批数据和流数据进行数据分析
- 使用开源的 Delta Lake 和 Spark构建可靠的数据管道
- 使用 Mllib 开发机器学习管道并使用 MLflow 生产模型

下载方式

关注 [过往记忆大数据](#) 微信公众号，并回复 [Learning_Spark2](#) 获取。

本博客文章除特别声明，全部都是原创！
转载本文请加上：转载自过往记忆 (<https://www.iteblog.com/>)
本文链接: [【】](#) ()