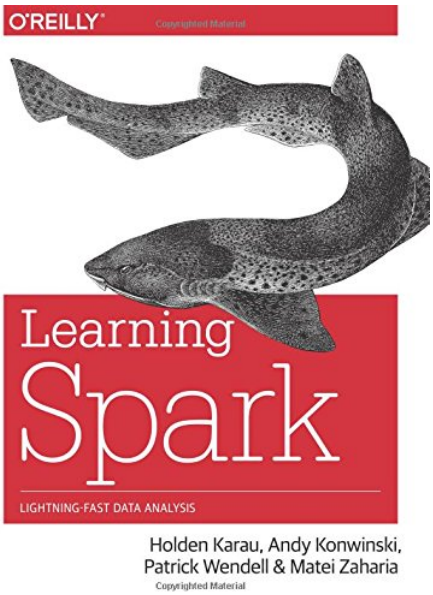
1U11b [Free PDF File] Learning Spark: Lightning-Fast Big Data Analysis



Data in all domains is getting bigger. How can you work with it efficiently? Recently updated for Spark 1.3, this book introduces Apache Spark, the open source cluster computing system that makes data analytics fast to write and fast to run. With Spark, you can tackle big datasets quickly through simple APIs in Python, Java, and Scala. This edition includes new information on Spark SQL, Spark Streaming, setup, and Maven coordinates.Written by the developers of Spark, this book will have data scientists and engineers up and running in no time. You'll learn how to express parallel jobs with just a few lines of code, and cover applications from simple batch jobs to stream processing and machine learning.Quickly dive into Spark capabilities such as distributed datasets, in-memory caching, and the interactive shellLeverage Spark's powerful built-in libraries, including Spark SQL, Spark Streaming, and MLlibUse one programming paradigm instead of mixing and matching tools like Hive, Hadoop, Mahout, and StormLearn how to deploy interactive, batch, and streaming applicationsConnect to data sources including HDFS, Hive, JSON, and S3Master advanced topics like data partitioning and shared variables

Data in all domains is getting bigger. How can you work with it efficiently? Recently updated for Spark 1.3, this book introduces Apache Spark, the open source cluster computing system that makes data analytics fast to write and fast to run. With Spark, you can tackle big datasets quickly through simple APIs in Python, Java, and Scala. This edition includes new information on Spark SQL, Spark Streaming, setup, and Maven coordinates.Written by the developers of Spark, this book will have data scientists and engineers up and running in no time. You'll learn how to express parallel jobs with just a few lines of code, and cover applications from simple batch jobs to stream processing and machine learning.Quickly dive into Spark capabilities such as distributed datasets, in-memory caching, and the interactive shellLeverage Spark's powerful built-in libraries, including Spark SQL, Spark Streaming, and MLlibUse one programming paradigm instead of mixing and matching tools like Hive, Hadoop, Mahout, and StormLearn how to deploy interactive, batch, and streaming applicationsConnect to data sources including HDFS, Hive, JSON, and S3Master advanced topics like data partitioning and shared variables