

Introduction	4
Section 1: An Introduction to Machine Learning in Apache Spark	5
New Features in MLlib in Spark 1.0	6
Statistics Functionality in Spark 1.1	9
Scalable Collaborative Filtering with Spark MLlib	13
ML Pipelines: A New High-Level API for MLlib	16
New Features in Machine Learning Pipelines in Spark 1.4	21
Improved Frequent Pattern Mining in Spark 1.5: Association Rules and Sequential Patterns	26
MLlib Highlights in Spark 1.6	29
Section 2: Advanced Topics: Clustering, Trees, Graph Processing, and More	32
Distributing the Singular Value Decomposition with Spark	33
Efficient Similarity Algorithm Now in Spark, Thanks to Twitter	35
Scalable Decision Trees in MLlib	37
Random Forests and Boosting in MLlib	41
Introducing Streaming k-means in Spark 1.2	47
Topic modeling with LDA: MLlib meets GraphX	51
Large Scale Topic Modeling: Improvements to LDA on Spark	56
Introducing GraphFrames	59
Section 3: Working with R on Apache Spark	63
Announcing SparkR: R on Spark	64
Generalized Linear Models in SparkR and R Formula Support in MLlib	67
Introducing R Notebooks in Databricks	71
Conclusion	75