

---

# Apache Spark Tutorial Machine Learning Article Datacamp

---

Learning Spark  
Examples and Case Studies  
Apache Spark Graph Processing  
Big Data Analytics with Spark  
Learning Spark SQL  
A Practitioner's Guide to Using Spark for Large Scale Data Analysis  
Apache Spark 2 for Beginners  
Mastering Structured Streaming and Spark Streaming  
Building a Deep Learning Model with TensorFlow  
Practical Examples in Apache Spark and Neo4j  
Data processing, ML algorithms, smart analytics, and more  
With Resilient Distributed Datasets, Spark SQL, Structured Streaming and Spark Machine Learning library  
The Complete Guide to Large-Scale Analysis and Modeling  
Apache Spark in 24 Hours, Sams Teach Yourself  
Mastering Spark with R  
Machine Learning with Apache Spark Quick Start Guide  
Machine Learning with Spark  
Over 60 recipes for implementing big data processing and analytics using Apache Spark and Python  
A Problem-Solution Approach with PySpark2  
Apache Spark 2.x Machine Learning Cookbook  
Patterns for Learning from Data at Scale  
Build powerful models with cognitive machine learning and artificial intelligence  
Beginning Apache Spark 2  
Stream Processing with Apache Spark  
Deep Learning Pipeline  
Spark GraphX in Action  
Practical Apache Spark  
PySpark Cookbook  
Natural Language Processing with Spark NLP  
Learning Spark  
Scala:Applied Machine Learning  
Hands-On Data Science and Python Machine Learning  
Learning to Understand Text at Scale  
Machine Learning with Spark  
Using the Scala API  
Big Data Processing Made Simple  
Learning PySpark  
Advanced Analytics with Spark  
Lightning-Fast Big Data Analysis  
Uncover patterns, derive actionable insights, and learn from big data using MLlib

*Apache Spark Tutorial Machine  
Learning Article Datacamp*

Downloaded from <ftp.wtvq.com> by guest

---

## HERRERA ELLEN

---

### *Learning Spark* Apress

Build data-intensive applications locally and deploy at scale using the combined powers of Python and Spark 2.0 About This Book Learn why and how you can efficiently use Python to process data and build machine learning models in Apache Spark 2.0 Develop and deploy efficient, scalable real-time Spark solutions Take your understanding of using Spark with Python to the next level with this jump start guide Who This Book Is For If you are a Python developer who wants to learn about the Apache Spark 2.0 ecosystem, this book is for you. A firm understanding of Python is expected to get the best out of the book. Familiarity with Spark would be useful, but is not mandatory. What You Will Learn Learn about Apache Spark and the Spark 2.0 architecture Build and interact with Spark DataFrames using Spark SQL Learn how to solve graph and deep learning problems using GraphFrames and

TensorFrames respectively Read, transform, and understand data and use it to train machine learning models Build machine learning models with MLlib and ML Learn how to submit your applications programmatically using spark-submit Deploy locally built applications to a cluster In Detail Apache Spark is an open source framework for efficient cluster computing with a strong interface for data parallelism and fault tolerance. This book will show you how to leverage the power of Python and put it to use in the Spark ecosystem. You will start by getting a firm understanding of the Spark 2.0 architecture and how to set up a Python environment for Spark. You will get familiar with the modules available in PySpark. You will learn how to abstract data with RDDs and DataFrames and understand the streaming capabilities of PySpark. Also, you will get a thorough overview of machine learning capabilities of PySpark using ML and MLlib, graph processing using GraphFrames, and polyglot persistence using Blaze. Finally, you will learn how to deploy your applications to the cloud using the spark-submit command. By

the end of this book, you will have established a firm understanding of the Spark Python API and how it can be used to build data-intensive applications. Style and approach This book takes a very comprehensive, step-by-step approach so you understand how the Spark ecosystem can be used with Python to develop efficient, scalable solutions. Every chapter is standalone and written in a very easy-to-understand manner, with a focus on both the hows and the whys of each concept.

*Examples and Case Studies* "O'Reilly Media, Inc."

Discover how graph algorithms can help you leverage the relationships within your data to develop more intelligent solutions and enhance your machine learning models. You'll learn how graph analytics are uniquely suited to unfold complex structures and reveal difficult-to-find patterns lurking in your data. Whether you are trying to build dynamic network models or forecast real-world behavior, this book illustrates how graph algorithms deliver value—from finding vulnerabilities and bottlenecks to detecting communities and improving machine learning predictions. This practical book walks you through hands-on examples of how to use graph algorithms in Apache Spark and Neo4j—two of the most common choices for graph analytics. Also included: sample code and tips for over 20 practical graph algorithms that cover optimal pathfinding, importance through centrality, and community detection. Learn how graph analytics vary from conventional statistical analysis Understand how classic graph algorithms work, and how they are applied Get guidance on which algorithms to use for different types of questions Explore algorithm examples with working code and sample datasets from Spark and Neo4j See how connected feature extraction can increase machine learning accuracy and precision Walk through creating an ML workflow for link prediction combining Neo4j and Spark

*Apache Spark Graph Processing* Simon and Schuster

Develop applications for the big data landscape with Spark and Hadoop. This book also explains the role of Spark in developing scalable machine learning and analytics applications with Cloud technologies. *Beginning Apache Spark 2* gives you an introduction to Apache Spark and shows you how to work with it. Along the way, you'll discover resilient distributed datasets (RDDs); use Spark SQL for structured data; and learn stream processing and build real-time applications with Spark Structured Streaming. Furthermore, you'll learn the fundamentals of Spark ML for machine learning and much more. After you read this book, you will have the fundamentals to become proficient in using Apache Spark and know when and how to apply it to your big data applications. What You Will Learn Understand Spark unified data processing platform How to run Spark in Spark Shell or Databricks Use and manipulate RDDs Deal with structured data using Spark SQL through its operations and advanced functions Build real-time applications using Spark Structured Streaming Develop intelligent applications with the Spark Machine Learning library Who This Book Is For Programmers and developers active in big data, Hadoop, and Java but who are new to the Apache Spark platform.

*Big Data Analytics with Spark* Apress

Summary Spark GraphX in Action starts out with an overview of Apache Spark and the GraphX graph processing API. This example-based tutorial then teaches you how to configure GraphX and how to use it interactively. Along the way, you'll collect practical techniques for enhancing applications and applying machine learning algorithms to graph data. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology GraphX is a powerful graph processing API for the Apache Spark analytics engine that lets you draw insights from large datasets. GraphX

gives you unprecedented speed and capacity for running massively parallel and machine learning algorithms. About the Book Spark GraphX in Action begins with the big picture of what graphs can be used for. This example-based tutorial teaches you how to use GraphX interactively. You'll start with a crystal-clear introduction to building big data graphs from regular data, and then explore the problems and possibilities of implementing graph algorithms and architecting graph processing pipelines. Along the way, you'll collect practical techniques for enhancing applications and applying machine learning algorithms to graph data. What's Inside Understanding graph technology Using the GraphX API Developing algorithms for big graphs Machine learning with graphs Graph visualization About the Reader Readers should be comfortable writing code. Experience with Apache Spark and Scala is not required. About the Authors Michael Malak has worked on Spark applications for Fortune 500 companies since early 2013. Robin East has worked as a consultant to large organizations for over 15 years and is a data scientist at Worldpay. Table of Contents PART 1 SPARK AND GRAPHS Two important technologies: Spark and graphs GraphX quick start Some fundamentals PART 2 CONNECTING VERTICES GraphX Basics Built-in algorithms Other useful graph algorithms Machine learning PART 3 OVER THE ARC The missing algorithms Performance and monitoring Other languages and tools [Learning Spark SQL](#) Packt Publishing Ltd

This book covers the fundamentals of machine learning with Python in a concise and dynamic manner. It covers data mining and large-scale machine learning using Apache Spark. About This Book Take your first steps in the world of data science by understanding the tools and techniques of data analysis Train efficient Machine Learning models in Python using the supervised and unsupervised learning methods Learn how to use Apache Spark for processing Big Data efficiently Who This Book Is For If you are a budding data scientist or a data analyst who wants to analyze and gain actionable insights from data using Python, this book is for you. Programmers with some experience in Python who want to enter the lucrative world of Data Science will also find this book to be very useful, but you don't need to be an expert Python coder or mathematician to get the most from this book. What You Will Learn Learn how to clean your data and ready it for analysis Implement the popular clustering and regression methods in Python Train efficient machine learning models using decision trees and random forests Visualize the results of your analysis using Python's Matplotlib library Use Apache Spark's MLlib package to perform machine learning on large datasets In Detail Join Frank Kane, who worked on Amazon and IMDb's machine learning algorithms, as he guides you on your first steps into the world of data science. Hands-On Data Science and Python Machine Learning gives you the tools that you need to understand and explore the core topics in the field, and the confidence and practice to build and analyze your own machine learning models. With the help of interesting and easy-to-follow practical examples, Frank Kane explains potentially complex topics such as Bayesian methods and K-means clustering in a way that anybody can understand them. Based on Frank's successful data science course, Hands-On Data Science and Python Machine Learning empowers you to conduct data analysis and perform efficient machine learning using Python. Let Frank help you unearth the value in your data using the various data mining and data analysis techniques available in Python, and to develop efficient predictive models to predict future results. You will also learn how to perform large-scale machine learning on Big Data using Apache Spark. The book covers preparing your data for analysis, training machine learning models, and visualizing the final data analysis. Style and

approach This comprehensive book is a perfect blend of theory and hands-on code examples in Python which can be used for your reference at any time.

[A Practitioner's Guide to Using Spark for Large Scale Data Analysis](#) "O'Reilly Media, Inc."

Apache Spark is a fast, scalable, and flexible open source distributed processing engine for big data systems and is one of the most active open source big data projects to date. In just 24 lessons of one hour or less, Sams Teach Yourself Apache Spark in 24 Hours helps you build practical Big Data solutions that leverage Spark's amazing speed, scalability, simplicity, and versatility. This book's straightforward, step-by-step approach shows you how to deploy, program, optimize, manage, integrate, and extend Spark—now, and for years to come. You'll discover how to create powerful solutions encompassing cloud computing, real-time stream processing, machine learning, and more. Every lesson builds on what you've already learned, giving you a rock-solid foundation for real-world success. Whether you are a data analyst, data engineer, data scientist, or data steward, learning Spark will help you to advance your career or embark on a new career in the booming area of Big Data. Learn how to

- Discover what Apache Spark does and how it fits into the Big Data landscape
- Deploy and run Spark locally or in the cloud
- Interact with Spark from the shell
- Make the most of the Spark Cluster Architecture
- Develop Spark applications with Scala and functional Python
- Program with the Spark API, including transformations and actions
- Apply practical data engineering/analysis approaches designed for Spark
- Use Resilient Distributed Datasets (RDDs) for caching, persistence, and output
- Optimize Spark solution performance
- Use Spark with SQL (via Spark SQL) and with NoSQL (via Cassandra)
- Leverage cutting-edge functional programming techniques
- Extend Spark with streaming, R, and Sparkling Water
- Start building Spark-based machine learning and graph-processing applications
- Explore advanced messaging technologies, including Kafka
- Preview and prepare for Spark's next generation of innovations

Instructions walk you through common questions, issues, and tasks; Q-and-As, Quizzes, and Exercises build and test your knowledge; "Did You Know?" tips offer insider advice and shortcuts; and "Watch Out!" alerts help you avoid pitfalls. By the time you're finished, you'll be comfortable using Apache Spark to solve a wide spectrum of Big Data problems.

[Apache Spark 2 for Beginners](#) O'Reilly Media

Build, process and analyze large-scale graph data effectively with Spark About This Book Find solutions for every stage of data processing from loading and transforming graph data to Improve the scalability of your graphs with a variety of real-world applications with complete Scala code. A concise guide to processing large-scale networks with Apache Spark. Who This Book Is For This book is for data scientists and big data developers who want to learn the processing and analyzing graph datasets at scale. Basic programming experience with Scala is assumed. Basic knowledge of Spark is assumed. What You Will Learn Write, build and deploy Spark applications with the Scala Build Tool. Build and analyze large-scale network datasets Analyze and transform graphs using RDD and graph-specific operations Implement new custom graph operations tailored to specific needs. Develop iterative and efficient graph algorithms using message aggregation and Pregel abstraction Extract subgraphs and use it to discover common clusters Analyze graph data and solve various data science problems using real-world datasets. In Detail Apache Spark is the next standard of open-source cluster-computing engine for processing big data. Many practical computing problems concern large graphs, like the Web graph and various social networks. The scale of these graphs - in

some cases billions of vertices, trillions of edges - poses challenges to their efficient processing. Apache Spark GraphX API combines the advantages of both data-parallel and graph-parallel systems by efficiently expressing graph computation within the Spark data-parallel framework. This book will teach the user to do graphical programming in Apache Spark, apart from an explanation of the entire process of graphical data analysis. You will journey through the creation of graphs, its uses, its exploration and analysis and finally will also cover the conversion of graph elements into graph structures. This book begins with an introduction of the Spark system, its libraries and the Scala Build Tool. Using a hands-on approach, this book will quickly teach you how to install and leverage Spark interactively on the command line and in a standalone Scala program. Then, it presents all the methods for building Spark graphs using illustrative network datasets. Next, it will walk you through the process of exploring, visualizing and analyzing different network characteristics. This book will also teach you how to transform raw datasets into a usable form. In addition, you will learn powerful operations that can be used to transform graph elements and graph structures. Furthermore, this book also teaches how to create custom graph operations that are tailored for specific needs with efficiency in mind. The later chapters of this book cover more advanced topics such as clustering graphs, implementing graph-parallel iterative algorithms and learning methods from graph data. Style and approach A step-by-step guide that will walk you through the key ideas and techniques for processing big graph data at scale, with practical examples that will ensure an overall understanding of the concepts of Spark.

[Mastering Structured Streaming and Spark Streaming](#) Packt Pub Limited

If you want to build an enterprise-quality application that uses natural language text but aren't sure where to begin or what tools to use, this practical guide will help get you started. Alex Thomas, principal data scientist at Wisecube, shows software engineers and data scientists how to build scalable natural language processing (NLP) applications using deep learning and the Apache Spark NLP library. Through concrete examples, practical and theoretical explanations, and hands-on exercises for using NLP on the Spark processing framework, this book teaches you everything from basic linguistics and writing systems to sentiment analysis and search engines. You'll also explore special concerns for developing text-based applications, such as performance. In four sections, you'll learn NLP basics and building blocks before diving into application and system building: Basics: Understand the fundamentals of natural language processing, NLP on Apache Spark, and deep learning Building blocks: Learn techniques for building NLP applications—including tokenization, sentence segmentation, and named-entity recognition—and discover how and why they work Applications: Explore the design, development, and experimentation process for building your own NLP applications Building NLP systems: Consider options for productionizing and deploying NLP models, including which human languages to support

[Building a Deep Learning Model with TensorFlow](#) Packt Publishing Ltd

Build your own pipeline based on modern TensorFlow approaches rather than outdated engineering concepts. This book shows you how to build a deep learning pipeline for real-life TensorFlow projects. You'll learn what a pipeline is and how it works so you can build a full application easily and rapidly. Then troubleshoot and overcome basic TensorFlow obstacles to easily create functional apps and deploy well-trained models. Step-by-step and example-oriented instructions help you understand each step of the deep learning pipeline while you apply the most

straightforward and effective tools to demonstrative problems and datasets. You'll also develop a deep learning project by preparing data, choosing the model that fits that data, and debugging your model to get the best fit to data all using Tensorflow techniques. Enhance your skills by accessing some of the most powerful recent trends in data science. If you've ever considered building your own image or text-tagging solution or entering a Kaggle contest, Deep Learning Pipeline is for you! What You'll Learn Develop a deep learning project using data Study and apply various models to your data Debug and troubleshoot the proper model suited for your data Who This Book Is For Developers, analysts, and data scientists looking to add to or enhance their existing skills by accessing some of the most powerful recent trends in data science. Prior experience in Python or other TensorFlow related languages and mathematics would be helpful.

[Practical Examples in Apache Spark and Neo4j](#) Lulu.com

Leverage Scala and Machine Learning to study and construct systems that can learn from data About This Book\* Explore a broad variety of data processing, machine learning, and genetic algorithms through diagrams, mathematical formulation, and updated source code in Scala\* Take your expertise in Scala programming to the next level by creating and customizing AI applications\* Experiment with different techniques and evaluate their benefits and limitations using real-world applications in a tutorial style Who This Book Is For If you're a data scientist or a data analyst with a fundamental knowledge of Scala who wants to learn and implement various Machine learning techniques, this book is for you. All you need is a good understanding of the Scala programming language, a basic knowledge of statistics, a keen interest in Big Data processing, and this book! What You Will Learn\* Build dynamic workflows for scientific computing\* Leverage open source libraries to extract patterns from time series\* Write your own classification, clustering, or evolutionary algorithm\* Perform relative performance tuning and evaluation of Spark\* Master probabilistic models for sequential data\* Experiment with advanced techniques such as regularization and kernelization\* Dive into neural networks and some deep learning architecture\* Apply some basic multiarm-bandit algorithms\* Solve big data problems with Scala parallel collections, Akka actors, and Apache Spark clusters\* Apply key learning strategies to a technical analysis of financial markets In Detail The discovery of information through data clustering and classification is becoming a key differentiator for competitive organizations. Machine learning applications are everywhere, from self-driving cars, engineering design, logistics, manufacturing, and trading strategies, to detection of genetic anomalies. The book is your one stop guide that introduces you to the functional capabilities of the Scala programming language that are critical to the creation of machine learning algorithms such as dependency injection and implicits. You start by learning data preprocessing and filtering techniques. Following this, you'll move on to unsupervised learning techniques such as clustering and dimension reduction, followed by probabilistic graphical models such as Naive Bayes, hidden Markov models and Monte Carlo inference. Further, it covers the discriminative algorithms such as linear, logistic regression with regularization, kernelization, support vector machines, neural networks, and deep learning. You'll move on to evolutionary computing, multibandit algorithms, and reinforcement learning. Finally, the book includes a comprehensive overview of parallel computing in Scala and Akka followed by a description of Apache Spark and its ML library. With updated codes based on the latest version of Scala and comprehensive examples, this book will ensure that you have more than just a solid fundamental knowledge in machine

learning with Scala. Style and approach This book is designed as a tutorial with hands-on exercises using technical analysis of financial markets and corporate data. The approach of each chapter is such that it allows you to understand key concepts easily.

[Data processing, ML algorithms, smart analytics, and more](#) Apress

Learn about the fastest-growing open source project in the world, and find out how it revolutionizes big data analytics About This Book Exclusive guide that covers how to get up and running with fast data processing using Apache Spark Explore and exploit various possibilities with Apache Spark using real-world use cases in this book Want to perform efficient data processing at real time? This book will be your one-stop solution. Who This Book Is For This guide appeals to big data engineers, analysts, architects, software engineers, even technical managers who need to perform efficient data processing on Hadoop at real time. Basic familiarity with Java or Scala will be helpful. The assumption is that readers will be from a mixed background, but would be typically people with background in engineering/data science with no prior Spark experience and want to understand how Spark can help them on their analytics journey. What You Will Learn Get an overview of big data analytics and its importance for organizations and data professionals Delve into Spark to see how it is different from existing processing platforms Understand the intricacies of various file formats, and how to process them with Apache Spark. Realize how to deploy Spark with YARN, MESOS or a Stand-alone cluster manager. Learn the concepts of Spark SQL, SchemaRDD, Caching and working with Hive and Parquet file formats Understand the architecture of Spark MLlib while discussing some of the off-the-shelf algorithms that come with Spark. Introduce yourself to the deployment and usage of SparkR. Walk through the importance of Graph computation and the graph processing systems available in the market Check the real world example of Spark by building a recommendation engine with Spark using ALS. Use a Telco data set, to predict customer churn using Random Forests. In Detail Spark juggernaut keeps on rolling and getting more and more momentum each day. Spark provides key capabilities in the form of Spark SQL, Spark Streaming, Spark ML and Graph X all accessible via Java, Scala, Python and R. Deploying the key capabilities is crucial whether it is on a Standalone framework or as a part of existing Hadoop installation and configuring with Yarn and Mesos. The next part of the journey after installation is using key components, APIs, Clustering, machine learning APIs, data pipelines, parallel programming. It is important to understand why each framework component is key, how widely it is being used, its stability and pertinent use cases. Once we understand the individual components, we will take a couple of real life advanced analytics examples such as 'Building a Recommendation system', 'Predicting customer churn' and so on. The objective of these real life examples is to give the reader confidence of using Spark for real-world problems. Style and approach With the help of practical examples and real-world use cases, this guide will take you from scratch to building efficient data applications using Apache Spark. You will learn all about this excellent data processing engine in a step-by-step manner, taking one aspect of it at a time. This highly practical guide will include how to work with data pipelines, dataframes, clustering, SparkSQL, parallel programming, and such insightful topics with the help of real-world use cases.

**With Resilient Distributed Datasets, Spark SQL, Structured Streaming and Spark Machine Learning library** Academic Press

Develop large-scale distributed data processing applications

using Spark 2 in Scala and Python About This Book- This book offers an easy introduction to the Spark framework published on the latest version of Apache Spark 2- Perform efficient data processing, machine learning and graph processing using various Spark components- A practical guide aimed at beginners to get them up and running with Spark Who This Book Is For If you are an application developer, data scientist, or big data solutions architect who is interested in combining the data processing power of Spark from R, and consolidating data processing, stream processing, machine learning, and graph processing into one unified and highly interoperable framework with a uniform API using Scala or Python, this book is for you. What You Will Learn- Get to know the fundamentals of Spark 2 and the Spark programming model using Scala and Python- Know how to use Spark SQL and DataFrames using Scala and Python- Get an introduction to Spark programming using R- Perform Spark data processing, charting, and plotting using Python- Get acquainted with Spark stream processing using Scala and Python- Be introduced to machine learning using Spark MLlib- Get started with graph processing using the Spark GraphX- Bring together all that you've learned and develop a complete Spark application In Detail Spark is one of the most widely-used large-scale data processing engines and runs extremely fast. It is a framework that has tools that are equally useful for application developers as well as data scientists. This book starts with the fundamentals of Spark 2 and covers the core data processing framework and API, installation, and application development setup. Then the Spark programming model is introduced through real-world examples followed by Spark SQL programming with DataFrames. An introduction to SparkR is covered next. Later, we cover the charting and plotting features of Python in conjunction with Spark data processing. After that, we take a look at Spark's stream processing, machine learning, and graph processing libraries. The last chapter combines all the skills you learned from the preceding chapters to develop a real-world Spark application. By the end of this book, you will have all the knowledge you need to develop efficient large-scale applications using Apache Spark. Style and approach Learn about Spark's infrastructure with this practical tutorial. With the help of real-world use cases on the main features of Spark we offer an easy introduction to the framework.

*The Complete Guide to Large-Scale Analysis and Modeling* Packt Publishing Ltd

Apache Spark is amazing when everything clicks. But if you haven't seen the performance improvements you expected, or still don't feel confident enough to use Spark in production, this practical book is for you. Authors Holden Karau and Rachel Warren demonstrate performance optimizations to help your Spark queries run faster and handle larger data sizes, while using fewer resources. Ideal for software engineers, data engineers, developers, and system administrators working with large-scale data applications, this book describes techniques that can reduce data infrastructure costs and developer hours. Not only will you gain a more comprehensive understanding of Spark, you'll also learn how to make it sing. With this book, you'll explore: How Spark SQL's new interfaces improve performance over SQL's RDD data structure The choice between data joins in Core Spark and Spark SQL Techniques for getting the most out of standard RDD transformations How to work around performance issues in Spark's key/value pair paradigm Writing high-performance Spark code without Scala or the JVM How to test for functionality and performance when applying suggested improvements Using Spark MLlib and Spark ML machine learning libraries Spark's Streaming components and external community packages  
**Apache Spark in 24 Hours, Sams Teach Yourself** Packt

Publishing

If you're like most R users, you have deep knowledge and love for statistics. But as your organization continues to collect huge amounts of data, adding tools such as Apache Spark makes a lot of sense. With this practical book, data scientists and professionals working with large-scale data applications will learn how to use Spark from R to tackle big data and big compute problems. Authors Javier Luraschi, Kevin Kuo, and Edgar Ruiz show you how to use R with Spark to solve different data analysis problems. This book covers relevant data science topics, cluster computing, and issues that should interest even the most advanced users. Analyze, explore, transform, and visualize data in Apache Spark with R Create statistical models to extract information and predict outcomes; automate the process in production-ready workflows Perform analysis and modeling across many machines using distributed computing techniques Use large-scale data from multiple sources and different formats with ease from within Spark Learn about alternative modeling frameworks for graph processing, geospatial analysis, and genomics at scale Dive into advanced topics including custom transformations, real-time data processing, and creating custom Spark extensions

Mastering Spark with R "O'Reilly Media, Inc."

Combine advanced analytics including Machine Learning, Deep Learning Neural Networks and Natural Language Processing with modern scalable technologies including Apache Spark to derive actionable insights from Big Data in real-time Key Features Make a hands-on start in the fields of Big Data, Distributed Technologies and Machine Learning Learn how to design, develop and interpret the results of common Machine Learning algorithms Uncover hidden patterns in your data in order to derive real actionable insights and business value Book Description Every person and every organization in the world manages data, whether they realize it or not. Data is used to describe the world around us and can be used for almost any purpose, from analyzing consumer habits to fighting disease and serious organized crime. Ultimately, we manage data in order to derive value from it, and many organizations around the world have traditionally invested in technology to help process their data faster and more efficiently. But we now live in an interconnected world driven by mass data creation and consumption where data is no longer rows and columns restricted to a spreadsheet, but an organic and evolving asset in its own right. With this realization comes major challenges for organizations: how do we manage the sheer size of data being created every second (think not only spreadsheets and databases, but also social media posts, images, videos, music, blogs and so on)? And once we can manage all of this data, how do we derive real value from it? The focus of Machine Learning with Apache Spark is to help us answer these questions in a hands-on manner. We introduce the latest scalable technologies to help us manage and process big data. We then introduce advanced analytical algorithms applied to real-world use cases in order to uncover patterns, derive actionable insights, and learn from this big data. What you will learn Understand how Spark fits in the context of the big data ecosystem Understand how to deploy and configure a local development environment using Apache Spark Understand how to design supervised and unsupervised learning models Build models to perform NLP, deep learning, and cognitive services using Spark ML libraries Design real-time machine learning pipelines in Apache Spark Become familiar with advanced techniques for processing a large volume of data by applying machine learning algorithms Who this book is for This book is aimed at Business Analysts, Data Analysts and Data Scientists who wish to make a hands-on start in order to take advantage of

modern Big Data technologies combined with Advanced Analytics.

### **Machine Learning with Apache Spark Quick Start Guide** Apress

Access real-world documentation and examples for the Spark platform for building large-scale, enterprise-grade machine learning applications. The past decade has seen an astonishing series of advances in machine learning. These breakthroughs are disrupting our everyday life and making an impact across every industry. Next-Generation Machine Learning with Spark provides a gentle introduction to Spark and Spark MLlib and advances to more powerful, third-party machine learning algorithms and libraries beyond what is available in the standard Spark MLlib library. By the end of this book, you will be able to apply your knowledge to real-world use cases through dozens of practical examples and insightful explanations. What You Will Learn Be introduced to machine learning, Spark, and Spark MLlib 2.4.x Achieve lightning-fast gradient boosting on Spark with the XGBoost4J-Spark and LightGBM libraries Detect anomalies with the Isolation Forest algorithm for Spark Use the Spark NLP and Stanford CoreNLP libraries that support multiple languages Optimize your ML workload with the Alluxio in-memory data accelerator for Spark Use GraphX and GraphFrames for Graph Analysis Perform image recognition using convolutional neural networks Utilize the Keras framework and distributed deep learning libraries with Spark Who This Book Is For Data scientists and machine learning engineers who want to take their knowledge to the next level and use Spark and more powerful, next-generation algorithms and libraries beyond what is available in the standard Spark MLlib library; also serves as a primer for aspiring data scientists and engineers who need an introduction to machine learning, Spark, and Spark MLlib.

**Machine Learning with Spark** Packt Publishing Ltd  
Implement machine learning, cognitive services, and artificial intelligence solutions by leveraging Azure cloud technologies Key Features Learn advanced concepts in Azure ML and the Cortana Intelligence Suite architecture Explore ML Server using SQL Server and HDInsight capabilities Implement various tools in Azure to build and deploy machine learning models Book Description Implementing Machine learning (ML) and Artificial Intelligence (AI) in the cloud had not been possible earlier due to the lack of processing power and storage. However, Azure has created ML and AI services that are easy to implement in the cloud. Hands-On Machine Learning with Azure teaches you how to perform advanced ML projects in the cloud in a cost-effective way. The book begins by covering the benefits of ML and AI in the cloud. You will then explore Microsoft's Team Data Science Process to establish a repeatable process for successful AI development and implementation. You will also gain an understanding of AI technologies available in Azure and the Cognitive Services APIs to integrate them into bot applications. This book lets you explore prebuilt templates with Azure Machine Learning Studio and build a model using canned algorithms that can be deployed as web services. The book then takes you through a preconfigured series of virtual machines in Azure targeted at AI development scenarios. You will get to grips with the ML Server and its capabilities in SQL and HDInsight. In the concluding chapters, you'll integrate patterns with other non-AI services in Azure. By the end of this book, you will be fully equipped to implement smart cognitive actions in your models. What you will learn Discover the benefits of leveraging the cloud for ML and AI Use Cognitive Services APIs to build intelligent bots Build a model using canned algorithms from Microsoft and deploy it as a web service Deploy virtual machines in AI development scenarios Apply R, Python, SQL Server, and Spark in Azure Build

and deploy deep learning solutions with CNTK, MMLSpark, and TensorFlow Implement model retraining in IoT, Streaming, and Blockchain solutions Explore best practices for integrating ML and AI functions with ADLA and logic apps Who this book is for If you are a data scientist or developer familiar with Azure ML and cognitive services and want to create smart models and make sense of data in the cloud, this book is for you. You'll also find this book useful if you want to bring powerful machine learning services into your cloud applications. Some experience with data manipulation and processing, using languages like SQL, Python, and R, will aid in understanding the concepts covered in this book *Over 60 recipes for implementing big data processing and analytics using Apache Spark and Python* Packt Publishing Ltd Work with Apache Spark using Scala to deploy and set up single-node, multi-node, and high-availability clusters. This book discusses various components of Spark such as Spark Core, DataFrames, Datasets and SQL, Spark Streaming, Spark MLlib, and R on Spark with the help of practical code snippets for each topic. Practical Apache Spark also covers the integration of Apache Spark with Kafka with examples. You'll follow a learn-to-do-by-yourself approach to learning - learn the concepts, practice the code snippets in Scala, and complete the assignments given to get an overall exposure. On completion, you'll have knowledge of the functional programming aspects of Scala, and hands-on expertise in various Spark components. You'll also become familiar with machine learning algorithms with real-time usage. What You Will Learn Discover the functional programming features of Scala Understand the complete architecture of Spark and its components Integrate Apache Spark with Hive and Kafka Use Spark SQL, DataFrames, and Datasets to process data using traditional SQL queries Work with different machine learning concepts and libraries using Spark's MLlib packages Who This Book Is For Developers and professionals who deal with batch and stream data processing.

*A Problem-Solution Approach with PySpark2* Packt Publishing Ltd Simplify machine learning model implementations with Spark About This Book Solve the day-to-day problems of data science with Spark This unique cookbook consists of exciting and intuitive numerical recipes Optimize your work by acquiring, cleaning, analyzing, predicting, and visualizing your data Who This Book Is For This book is for Scala developers with a fairly good exposure to and understanding of machine learning techniques, but lack practical implementations with Spark. A solid knowledge of machine learning algorithms is assumed, as well as hands-on experience of implementing ML algorithms with Scala. However, you do not need to be acquainted with the Spark ML libraries and ecosystem. What You Will Learn Get to know how Scala and Spark go hand-in-hand for developers when developing ML systems with Spark Build a recommendation engine that scales with Spark Find out how to build unsupervised clustering systems to classify data in Spark Build machine learning systems with the Decision Tree and Ensemble models in Spark Deal with the curse of high-dimensionality in big data using Spark Implement Text analytics for Search Engines in Spark Streaming Machine Learning System implementation using Spark In Detail Machine learning aims to extract knowledge from data, relying on fundamental concepts in computer science, statistics, probability, and optimization. Learning about algorithms enables a wide range of applications, from everyday tasks such as product recommendations and spam filtering to cutting edge applications such as self-driving cars and personalized medicine. You will gain hands-on experience of applying these principles using Apache Spark, a resilient cluster computing system well suited for large-scale machine learning tasks. This book begins with a quick overview of setting up the necessary IDEs to facilitate the

execution of code examples that will be covered in various chapters. It also highlights some key issues developers face while working with machine learning algorithms on the Spark platform. We progress by uncovering the various Spark APIs and the implementation of ML algorithms with developing classification systems, recommendation engines, text analytics, clustering, and learning systems. Toward the final chapters, we'll focus on building high-end applications and explain various unsupervised methodologies and challenges to tackle when implementing with big data ML systems. Style and approach This book is packed with intuitive recipes supported with line-by-line explanations to help you understand how to optimize your work flow and resolve problems when working with complex data modeling tasks and predictive algorithms. This is a valuable resource for data scientists and those working on large scale data projects.

*Apache Spark 2.x Machine Learning Cookbook* Packt Publishing Ltd

R and Data Mining introduces researchers, post-graduate students, and analysts to data mining using R, a free software

environment for statistical computing and graphics. The book provides practical methods for using R in applications from academia to industry to extract knowledge from vast amounts of data. Readers will find this book a valuable guide to the use of R in tasks such as classification and prediction, clustering, outlier detection, association rules, sequence analysis, text mining, social network analysis, sentiment analysis, and more. Data mining techniques are growing in popularity in a broad range of areas, from banking to insurance, retail, telecom, medicine, research, and government. This book focuses on the modeling phase of the data mining process, also addressing data exploration and model evaluation. With three in-depth case studies, a quick reference guide, bibliography, and links to a wealth of online resources, R and Data Mining is a valuable, practical guide to a powerful method of analysis. Presents an introduction into using R for data mining applications, covering most popular data mining techniques Provides code examples and data so that readers can easily learn the techniques Features case studies in real-world applications to help readers apply the techniques in their work