

Apache Spark Machine Learning Blueprints

Build interactive dashboards and storytelling reports at scale with the cloud-native BI tool - Amazon QuickSight, including embedded analytics and ML-powered insights Key Features Understand how to set up Amazon QuickSight, manage data sources, and build and share dashboards Learn the advanced features of Amazon Quicksight to develop interactive and embedded dashboards Manage and monitor dashboards using the QuickSight API and other AWS services such as Amazon CloudTrail Book Description The adoption of cloud-native BI tools, like Amazon QuickSight, enables organizations to gather insights from data at scale. This book is a practical guide to performing simple-to-advanced tasks with Amazon QuickSight. You'll begin by learning QuickSight's fundamental concepts and how to configure data sources. Next, you'll be introduced to the main analysis-building functionality of QuickSight to develop visuals and dashboards. The book will also demonstrate how to develop and share interactive dashboards with parameters and on-screen controls. Advanced filtering options with URL actions will then be covered, before learning how to set up alerts and scheduled reports. Later, you'll explore the Insights visual type in QuickSight using both existing insights and by building custom insights. Further chapters will show you how to add machine learning insights such as forecasting capabilities, analyzing time series data, adding narratives, and outlier detection to your dashboards. You'll also explore patterns to automate operations and look closer into the API actions that allow us to control settings. Finally, you'll learn advanced topics such as embedded dashboards and multitenancy. By the end of this book, you'll be well-versed with QuickSight's BI and analytics functionalities that will help you create BI apps with ML capabilities. What you will learn Understand the wider AWS analytics ecosystem and how QuickSight fits within it Set up and configure data sources with Amazon QuickSight Include custom controls and add interactivity to your BI application using parameters Add ML insights such as forecasting, anomaly detection, and narratives Explore patterns to automate operations using QuickSight APIs Create interactive dashboards and storytelling with Amazon QuickSight Design an embedded multi-tenant analytics architecture Focus on data permissions and how to manage Amazon QuickSight operations Who this book is for This book is for business intelligence (BI) developers and data analysts who are looking to create interactive dashboards using data from Lake House on AWS with Amazon QuickSight. This book will also be useful for anyone who wants to learn Amazon QuickSight in depth using practical examples. You will need to be familiar with general data visualization concepts, however, no prior experience with Amazon QuickSight is required.

If you are ready to dive into the MapReduce framework for processing large datasets, this practical book takes you step by step through the algorithms and tools you need to build distributed MapReduce applications with Apache Hadoop or Apache Spark. Each chapter provides a recipe for solving a massive computational problem, such as building a recommendation system. You'll learn how to implement the appropriate MapReduce solution with code that you can use in your projects. Dr. Mahmoud Parsian covers basic design patterns, optimization techniques, and data mining and machine learning solutions for problems in bioinformatics, genomics, statistics, and social network analysis. This book also includes an overview of MapReduce, Hadoop, and Spark. Topics include: Market basket analysis for a large set of transactions Data mining algorithms (K-means, KNN, and Naive Bayes) Using huge genomic data to sequence DNA and RNA Naive Bayes theorem and Markov chains for data and market prediction Recommendation algorithms and pairwise document similarity Linear regression, Cox regression, and Pearson correlation Allelic frequency and mining DNA Social network analysis (recommendation systems, counting triangles, sentiment analysis)

Before you can build analytics tools to gain quick insights, you first need to know how to process data in real time. With this practical guide, developers familiar with Apache Spark will learn how to put this in-memory framework to use for streaming data. You'll discover how Spark enables you to write streaming jobs in almost the same way you write batch jobs. Authors Gerard Maas and François Garillot help you explore the theoretical underpinnings of Apache Spark. This comprehensive guide features two sections that compare and contrast the streaming APIs Spark now supports: the original Spark Streaming library and the newer Structured Streaming API. Learn fundamental stream processing concepts and examine different streaming architectures Explore Structured Streaming through practical examples; learn different aspects of stream processing in detail Create and operate streaming jobs and applications with Spark Streaming; integrate Spark Streaming with other Spark APIs Learn advanced Spark Streaming techniques, including approximation algorithms and machine learning algorithms Compare Apache Spark to other stream processing projects, including Apache Storm, Apache Flink, and Apache Kafka Streams

Turning text into valuable information is essential for businesses looking to gain a competitive advantage. With recent improvements in natural language processing (NLP), users now have many options for solving complex challenges. But it's not always clear which NLP tools or libraries would work for a business's needs, or which techniques you should use and in what order. This practical book provides data scientists and developers with blueprints for best practice solutions to common tasks in text analytics and natural language processing. Authors Jens Albrecht, Sidharth Ramachandran, and Christian Winkler provide real-world case studies and detailed code examples in Python to help you get started quickly. Extract data from APIs and web pages Prepare textual data for statistical analysis and machine learning Use machine learning for classification, topic modeling, and summarization Explain AI models and classification results Explore and visualize semantic similarities with word embeddings Identify customer sentiment in product reviews Create a knowledge graph based on named entities and their relations

In this book readers will find technological discussions on the existing and emerging technologies across the different stages of the big data value chain. They will learn about

legal aspects of big data, the social impact, and about education needs and requirements. And they will discover the business perspective and how big data technology can be exploited to deliver value within different sectors of the economy. The book is structured in four parts: Part I “The Big Data Opportunity” explores the value potential of big data with a particular focus on the European context. It also describes the legal, business and social dimensions that need to be addressed, and briefly introduces the European Commission’s BIG project. Part II “The Big Data Value Chain” details the complete big data lifecycle from a technical point of view, ranging from data acquisition, analysis, curation and storage, to data usage and exploitation. Next, Part III “Usage and Exploitation of Big Data” illustrates the value creation possibilities of big data applications in various sectors, including industry, healthcare, finance, energy, media and public services. Finally, Part IV “A Roadmap for Big Data Research” identifies and prioritizes the cross-sectorial requirements for big data research, and outlines the most urgent and challenging technological, economic, political and societal issues for big data in Europe. This compendium summarizes more than two years of work performed by a leading group of major European research centers and industries in the context of the BIG project. It brings together research findings, forecasts and estimates related to this challenging technological context that is becoming the major axis of the new digitally transformed business environment.

With the intriguing development of technologies in several industries, along with the advent of ubiquitous computational resources, there are now ample opportunities to develop innovative computational technologies in order to solve a wide range of issues concerning uncertainty, imprecision, and vagueness in various real-life problems. The challenge of blending modern computational techniques with traditional computing methods has inspired researchers and academics alike to focus on developing innovative computational techniques. In the near future, computational techniques may provide vital solutions by effectively using evolving technologies such as computer vision, natural language processing, deep learning, machine learning, scientific computing, and computational vision. A vast number of intelligent computational algorithms are emerging, along with increasing computational power, which has significantly expanded the potential for developing intelligent applications. These proceedings of the International Conference on Inventive Computation Technologies [ICICT 2019] cover innovative computing applications in the areas of data mining, big data processing, information management, and security.

Healthcare and technology are at a convergence point where significant changes are poised to take place. The vast and complex requirements of medical record keeping, coupled with stringent patient privacy laws, create an incredibly unwieldy maze of health data needs. While the past decade has seen giant leaps in AI, machine learning, wearable technologies, and data mining capacities that have enabled quantities of data to be accumulated, processed, and shared around the globe. Transforming Healthcare with Big Data and AI examines the crossroads of these two fields and looks to the future of leveraging advanced technologies and developing data ecosystems to the healthcare field. This book is the product of the Transforming Healthcare with Data conference, held at the University of Southern California. Many speakers and digital healthcare industry leaders contributed multidisciplinary expertise to chapters in this work. Authors’ backgrounds range from data scientists, healthcare experts, university professors, and digital healthcare entrepreneurs. If you have an understanding of data technologies and are interested in the future of Big Data and A.I. in healthcare, this book will provide a wealth of insights into the new landscape of healthcare.

A practical guide to help you tackle different real-time data processing and analytics problems using the best tools for each scenario About This Book Learn about the various challenges in real-time data processing and use the right tools to overcome them This book covers popular tools and frameworks such as Spark, Flink, and Apache Storm to solve all your distributed processing problems A practical guide filled with examples, tips, and tricks to help you perform efficient Big Data processing in real-time Who This Book Is For If you are a Java developer who would like to be equipped with all the tools required to devise an end-to-end practical solution on real-time data streaming, then this book is for you. Basic knowledge of real-time processing would be helpful, and knowing the fundamentals of Maven, Shell, and Eclipse would be great. What You Will Learn Get an introduction to the established real-time stack Understand the key integration of all the components Get a thorough understanding of the basic building blocks for real-time solution designing Garnish the search and visualization aspects for your real-time solution Get conceptually and practically acquainted with real-time analytics Be well equipped to apply the knowledge and create your own solutions In Detail With the rise of Big Data, there is an increasing need to process large amounts of data continuously, with a shorter turnaround time. Real-time data processing involves continuous input, processing and output of data, with the condition that the time required for processing is as short as possible. This book covers the majority of the existing and evolving open source technology stack for real-time processing and analytics. You will get to know about all the real-time solution aspects, from the source to the presentation to persistence. Through this practical book, you'll be equipped with a clear understanding of how to solve challenges on your own. We'll cover topics such as how to set up components, basic executions, integrations, advanced use cases, alerts, and monitoring. You'll be exposed to the popular tools used in real-time processing today such as Apache Spark, Apache Flink, and Storm. Finally, you will put your knowledge to practical use by implementing all of the techniques in the form of a practical, real-world use case. By the end of this book, you will have a solid understanding of all the aspects of real-time data processing and analytics, and will know how to deploy the solutions in production environments in the best possible manner. Style and Approach In this practical guide to real-time analytics, each chapter begins with a basic high-level concept of the topic, followed by a practical, hands-on implementation of each concept, where you can see the working and execution of it. The book is written in a DIY style, with plenty of practical use cases, well-explained code examples, and relevant screenshots and diagrams.

The first textbook to teach students how to build data analytic solutions on large data sets using cloud-based technologies. This is the first textbook to teach students how to build data analytic solutions on large data sets (specifically in Internet of Things applications) using cloud-based technologies for data storage, transmission and mashup, and AI techniques to analyze this data. This textbook is designed to train college students to master modern cloud computing systems in operating principles, architecture design, machine learning algorithms, programming models and software tools for big data mining, analytics, and cognitive applications. The book will be suitable for use in one-semester computer science or electrical engineering courses on cloud computing, machine learning, cloud programming, cognitive computing, or big data science. The book will also be very useful as a reference for professionals who want to work in cloud computing and data science. Cloud and Cognitive Computing begins with two introductory chapters on fundamentals of cloud computing, data science, and adaptive computing that lay the foundation for the rest of the book. Subsequent chapters cover topics including cloud architecture, mashup services, virtual machines, Docker containers, mobile clouds, IoT and AI, inter-cloud mashups, and cloud performance and benchmarks, with a focus on Google's Brain Project, DeepMind, and X-Lab programs, IBKai HwangM SyNapse, Bluemix programs, cognitive initiatives, and neurocomputers. The book then covers machine learning algorithms and cloud programming software tools and application development, applying the tools in machine learning, social media, deep learning, and cognitive applications. All cloud systems are illustrated with big data and cognitive application examples.

Use Hadoop to solve business problems by learning from a rich set of real-life case studies
About This Book* Learn to solve real-world business problems using Hadoop and other big data technologies*
Power-packed with seven case studies to get you going with Hadoop for Business Intelligence* Analyze sensor data using Hadoop and develop a data visualization dashboard
Who This Book Is For
This book is designed for those who are interested in exploring the capabilities of Hadoop. This book assumes that you have basic knowledge of Hadoop and its ecosystem of tools. Basic knowledge of Java and scripting is expected
What You Will Learn* Use Sqoop and Hive to build your customer a 360 degree view* Use Spark and Hadoop to build a fraud detection system* See how customer master data and transaction streams can be used to build a fraud detection system* Familiarize yourself on how machine-learning libraries can be used with Hadoop to solve business problems* Collect data through GPS sensors to build dashboards using visualization* Get to grips with building a Hadoop-based Data Lake for large enterprises
In Detail
If you have a basic understanding of Hadoop and want to put your knowledge to use to build fantastic business solutions, then this book is for you. You will learn to build six real-life, end-to-end applications using Hadoop in its ecosystem. Each chapter will cover the steps required to build an application end to end. We start off by discussing various industry use cases and the associated business cases. We also discuss common architectural patterns. Next, you'll learn how to import and utilize structured and semi-structured data to build a 360-degree view of the customer. We'll also build a classification model based upon historical data using a machine learning library. We then move on to use an analytics model to predict the likelihood of a certain event. Through these real-world cases, you'll take your Hadoop learning to the next level. You will be able to try out the solutions explained in the book and use the knowledge gained to extend them further in your own problem space.

Learn to design and build Virtual Reality experiences, applications, and games in Unreal Engine 4 through a series of practical, hands-on projects that teach you to create controllable avatars, user interfaces, and more. Key Features
Learn about effective VR design and develop virtual reality games and applications for every VR platform
Build essential features for VR such as player locomotion and interaction, 3D user interfaces, and 360 media players
Learn about multiplayer networking and how to extend the engine using plugins and asset packs
Book Description
Unreal Engine 4 (UE4) is a powerful tool for developing VR games and applications. With its visual scripting language, Blueprint, and built-in support for all major VR headsets, it's a perfect tool for designers, artists, and engineers to realize their visions in VR. This book will guide you step-by-step through a series of projects that teach essential concepts and techniques for VR development in UE4. You will begin by learning how to think about (and design for) VR and then proceed to set up a development environment. A series of practical projects follows, taking you through essential VR concepts. Through these exercises, you'll learn how to set up UE4 projects that run effectively in VR, how to build player locomotion schemes, and how to use hand controllers to interact with the world. You'll then move on to create user interfaces in 3D space, use the editor's VR mode to build environments directly in VR, and profile/optimize worlds you've built. Finally, you'll explore more advanced topics, such as displaying stereo media in VR, networking in Unreal, and using plugins to extend the engine. Throughout, this book focuses on creating a deeper understanding of why the relevant tools and techniques work as they do, so you can use the techniques and concepts learned here as a springboard for further learning and exploration in VR. What you will learn
Understand design principles and concepts for building VR applications
Set up your development environment with Unreal Blueprints and C++
Create a player character with several locomotion schemes
Evaluate and solve performance problems in VR to maintain high frame rates
Display mono and stereo videos in VR
Extend Unreal Engine's capabilities using various plugins
Who this book is for
This book is for anyone interested in learning to develop Virtual Reality games and applications using UE4. Developers new to UE4 will benefit from hands-on projects that guide readers through clearly-explained steps, while both new and experienced developers will learn crucial principles and techniques for VR development in UE4. How do you start? How should you build a plan for cloud migration for your entire portfolio? How will your organization be affected by these changes? This book, based on real-world cloud experiences by enterprise IT teams, seeks to provide the answers to these questions. Here, you'll see what makes the cloud so compelling to enterprises; with which applications you should start your cloud journey; how your organization will change, and how skill sets will evolve; how to measure progress; how to think about security, compliance, and business buy-in; and how to exploit the ever-growing feature set that the cloud offers to gain strategic and competitive advantage.

The Official, Full-Color Guide to Developing Interactive Visualizations, Animations, and Renderings with Unreal Engine 4
Unreal Engine 4 (UE4) was created to develop video games, but it has gone viral among architecture, science, engineering, and medical visualization communities. UE4's stunning visual quality, cutting-edge toolset, unbeatable price (free!), and unprecedented ease of use redefines the state of the art and has turned the gaming, film, and visualization industries on their heads. Unreal Engine 4 for Design Visualization delivers the knowledge visualization professionals need to leverage UE4's immense power. World-class UE4 expert Tom Shannon introduces Unreal Engine 4's components and technical concepts, mentoring you through the entire process of building outstanding visualization content—all with realistic, carefully documented, step-by-step sample projects. Shannon answers the questions most often asked about UE4 visualization, addressing issues ranging from data import and processing to lighting, advanced materials, and rendering. He reveals important ways in which UE4 works differently from traditional rendering systems, even when it uses similar terminology. Throughout, he writes from the perspective of visualization professionals in architecture, engineering, or science—not gaming. Understand UE4's components and development environment
Master UE4's pipeline from source data to delivered application
Recognize and adapt to the differences between UE4 and traditional visualization and rendering techniques
Achieve staggering realism with UE4's Physically Based Rendering (PBR) Materials, Lighting, and Post-Processing pipelines
Create production-ready Materials with the interactive real-time Material Editor
Quickly set up projects, import massive datasets, and populate worlds with accurate visualization data
Develop bright, warm lighting for architectural visualizations
Create pre-rendered animations with Sequencer
Use Blueprints Visual Scripting to create complex interactions without writing a single line of code
Work with (and around) UE4's limitations and leveraging its advantages to achieve your vision
All UE4 project files and 3ds Max source files, plus additional resources and links, are available at the book's companion website.

This book provides a comprehensive survey of techniques, technologies and applications of Big Data and its analysis. The Big Data phenomenon is increasingly impacting all sectors of business and industry, producing an emerging new information ecosystem. On the applications front, the book offers detailed descriptions of various application areas for Big Data Analytics in the important domains of Social Semantic Web Mining, Banking and Financial Services, Capital Markets, Insurance, Advertisement, Recommendation Systems, Bio-Informatics, the IoT and Fog Computing, before delving into issues of security and privacy. With regard to machine learning techniques, the book presents all the standard algorithms for learning – including supervised, semi-supervised and unsupervised techniques such as clustering and reinforcement learning techniques to perform collective Deep Learning. Multi-layered and nonlinear learning for Big Data are also covered. In turn, the book highlights real-life case studies on successful implementations of Big Data Analytics at large IT companies such as Google, Facebook, LinkedIn and Microsoft. Multi-sectorial case studies on domain-based companies such as Deutsche Bank, the power provider Opower, Delta Airlines and a Chinese City Transportation application represent a valuable addition. Given its comprehensive coverage of Big Data Analytics, the book offers a unique resource for undergraduate and graduate students, researchers, educators and IT professionals alike.

It's time to dispel the myth that machine learning is difficult. Grokking Machine Learning teaches you how to apply ML to your projects using only standard Python code and high school-level math. No specialist knowledge is required to tackle the hands-on exercises using readily-available machine learning tools In Grokking Machine Learning, expert machine learning engineer Luis Serrano introduces the most valuable ML techniques and teaches you how to make them work for you. Practical examples illustrate each new concept to ensure you're grokking as you go. You'll build models for spam detection, language analysis, and image recognition as you lock in each carefully-selected skill. Packed with easy-to-follow Python-based exercises and mini-projects, this book sets you on the path to becoming a machine learning expert. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications.

Go beyond the basics and unleash the full power of QGIS 3.4 and 3.6 with practical, step-by-step examples Key Features One-stop solution to all of your GIS needs Master QGIS by learning about database integration, and geoprocessing tools Learn about the new and updated Processing toolbox and perform spatial analysis Book Description QGIS is an open source solution to GIS and widely used by GIS professionals all over the world. It is the leading alternative to proprietary GIS software. Although QGIS is described as intuitive, it is also, by default, complex. Knowing which tools to use and how to apply them is essential to producing valuable deliverables on time. Starting with a refresher on the QGIS basics and getting you acquainted with the latest QGIS 3.6 updates, this book will take you all the way through to teaching you how to create a spatial database and a GeoPackage. Next, you will learn how to style raster and vector data by choosing and managing different colors. The book will then focus on processing raster and vector data. You will be then taught advanced applications, such as creating and editing vector data. Along with that, you will also learn about the newly updated Processing Toolbox, which will help you develop the advanced data visualizations. The book will then explain to you the graphic modeler, how to create QGIS plugins with PyQGIS, and how to integrate Python analysis scripts with QGIS. By the end of the book, you will understand how to work with all aspects of QGIS and will be ready to use it for any type of GIS work. What you will learn Create and manage a spatial database Get to know advanced techniques to style GIS data Prepare both vector and raster data for processing Add heat maps, live layer effects, and labels to your maps Master LAsTools and GRASS integration with the Processing Toolbox Edit and repair topological data errors Automate workflows with batch processing and the QGIS Graphical Modeler Integrate Python scripting into your data processing workflows Develop your own QGIS plugins Who this book is for If you are a GIS professional, a consultant, a student, or perhaps a fast learner who wants to go beyond the basics of QGIS, then this book is for you. It will prepare you to realize the full potential of QGIS.

Summary Introducing Data Science teaches you how to accomplish the fundamental tasks that occupy data scientists. Using the Python language and common Python libraries, you'll experience firsthand the challenges of dealing with data at scale and gain a solid foundation in data science. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Many companies need developers with data science skills to work on projects ranging from social media marketing to machine learning. Discovering what you need to learn to begin a career as a data scientist can seem bewildering. This book is designed to help you get started. About the Book Introducing Data Science Introducing Data Science explains vital data science concepts and teaches you how to accomplish the fundamental tasks that occupy data scientists. You'll explore data visualization, graph databases, the use of NoSQL, and the data science process. You'll use the Python language and common Python libraries as you experience firsthand the challenges of dealing with data at scale. Discover how Python allows you to gain insights from data sets so big that they need to be stored on multiple machines, or from data moving so quickly that no single machine can handle it. This book gives you hands-on experience with the most popular Python data science libraries, Scikit-learn and StatsModels. After reading this book, you'll have the solid foundation you need to start a career in data science. What's Inside Handling large data Introduction to machine learning Using Python to work with data Writing data science algorithms About the Reader This book assumes you're comfortable reading code in Python or a similar language, such as C, Ruby, or JavaScript. No prior experience with data science is required. About the Authors Davy Cielen, Arno D. B. Meysman, and Mohamed Ali are the founders and managing partners of Optimately and Maiton, where they focus on developing data science projects and solutions in various sectors. Table of Contents Data science in a big data world The data science process Machine learning Handling large data on a single computer First steps in big data Join the NoSQL movement The rise of graph databases Text mining and text analytics Data visualization to the end user

Develop a range of cutting-edge machine learning projects with Apache Spark using this actionable guide About This Book Customize Apache Spark and R to fit your analytical needs in customer research, fraud detection, risk analytics, and recommendation engine development Develop a set of practical Machine Learning applications that can be implemented in real-life projects A comprehensive, project-based guide to improve and refine your predictive models for practical implementation Who This Book Is For If you are a data scientist, a data analyst, or an R and SPSS user with a good understanding of machine learning concepts, algorithms, and techniques, then this is the book for you. Some basic understanding of Spark and its core elements and application is required. What You Will Learn Set up Apache Spark for machine learning and discover its impressive processing power Combine Spark and R to unlock detailed business insights essential for decision making Build machine learning systems with Spark that can detect fraud and analyze financial risks Build predictive models focusing on customer scoring and service ranking Build a recommendation systems using SPSS on Apache Spark Tackle parallel computing and find out how it can support your machine learning projects Turn open data and communication data into actionable insights by making use of various forms of machine learning In Detail There's a reason why Apache Spark has become one of the most popular tools in Machine Learning – its ability to handle huge datasets at an impressive speed means you can be much more responsive to the data at your disposal. This book shows you Spark at its very best, demonstrating how to connect it with R and unlock maximum value not only from the tool but also from your data. Packed with a range of project "blueprints" that demonstrate some of the most interesting challenges that Spark can help you tackle, you'll find out how to use Spark notebooks and access, clean, and join different datasets before putting your knowledge into practice with some real-world projects, in which you will see how Spark Machine Learning can help you with everything from fraud detection to analyzing customer attrition. You'll also find out how to build a recommendation engine using Spark's parallel computing powers. Style and approach This book offers a step-by-step approach to setting up Apache Spark, and use other analytical tools with it to process Big Data and build machine learning projects. The initial chapters focus more on the theory aspect of machine learning with Spark, while each of the later chapters focuses on

building standalone projects using Spark.

Build, implement and scale distributed deep learning models for large-scale datasets About This Book Get to grips with the deep learning concepts and set up Hadoop to put them to use Implement and parallelize deep learning models on Hadoop's YARN framework A comprehensive tutorial to distributed deep learning with Hadoop Who This Book Is For If you are a data scientist who wants to learn how to perform deep learning on Hadoop, this is the book for you. Knowledge of the basic machine learning concepts and some understanding of Hadoop is required to make the best use of this book. What You Will Learn Explore Deep Learning and various models associated with it Understand the challenges of implementing distributed deep learning with Hadoop and how to overcome it Implement Convolutional Neural Network (CNN) with deeplearning4j Delve into the implementation of Restricted Boltzmann Machines (RBM) Understand the mathematical explanation for implementing Recurrent Neural Networks (RNN) Get hands on practice of deep learning and their implementation with Hadoop. In Detail This book will teach you how to deploy large-scale dataset in deep neural networks with Hadoop for optimal performance. Starting with understanding what deep learning is, and what the various models associated with deep neural networks are, this book will then show you how to set up the Hadoop environment for deep learning. In this book, you will also learn how to overcome the challenges that you face while implementing distributed deep learning with large-scale unstructured datasets. The book will also show you how you can implement and parallelize the widely used deep learning models such as Deep Belief Networks, Convolutional Neural Networks, Recurrent Neural Networks, Restricted Boltzmann Machines and autoencoder using the popular deep learning library deeplearning4j. Get in-depth mathematical explanations and visual representations to help you understand the design and implementations of Recurrent Neural network and Denoising AutoEncoders with deeplearning4j. To give you a more practical perspective, the book will also teach you the implementation of large-scale video processing, image processing and natural language processing on Hadoop. By the end of this book, you will know how to deploy various deep neural networks in distributed systems using Hadoop. Style and approach This book takes a comprehensive, step-by-step approach to implement efficient deep learning models on Hadoop. It starts from the basics and builds the readers' knowledge as they strengthen their understanding of the concepts. Practical examples are included in every step of the way to supplement the theory.

No need to spend hours ploughing through endless data – let Spark, one of the fastest big data processing engines available, do the hard work for you. Key Features Get up and running with Apache Spark and Python Integrate Spark with AWS for real-time analytics Apply processed data streams to machine learning APIs of Apache Spark Book Description Processing big data in real time is challenging due to scalability, information consistency, and fault-tolerance. This book teaches you how to use Spark to make your overall analytical workflow faster and more efficient. You'll explore all core concepts and tools within the Spark ecosystem, such as Spark Streaming, the Spark Streaming API, machine learning extension, and structured streaming. You'll begin by learning data processing fundamentals using Resilient Distributed Datasets (RDDs), SQL, Datasets, and Dataframes APIs. After grasping these fundamentals, you'll move on to using Spark Streaming APIs to consume data in real time from TCP sockets, and integrate Amazon Web Services (AWS) for stream consumption. By the end of this book, you'll not only have understood how to use machine learning extensions and structured streams but you'll also be able to apply Spark in your own upcoming big data projects. What you will learn Write your own Python programs that can interact with Spark Implement data stream consumption using Apache Spark Recognize common operations in Spark to process known data streams Integrate Spark streaming with Amazon Web Services (AWS) Create a collaborative filtering model with the movielens dataset Apply processed data streams to Spark machine learning APIs Who this book is for Data Processing with Apache Spark is for you if you are a software engineer, architect, or IT professional who wants to explore distributed systems and big data analytics. Although you don't need any knowledge of Spark, prior experience of working with Python is recommended.

With this practical book, AI and machine learning practitioners will learn how to successfully build and deploy data science projects on Amazon Web Services. The Amazon AI and machine learning stack unifies data science, data engineering, and application development to help level up your skills. This guide shows you how to build and run pipelines in the cloud, then integrate the results into applications in minutes instead of days. Throughout the book, authors Chris Fregly and Antje Barth demonstrate how to reduce cost and improve performance. Apply the Amazon AI and ML stack to real-world use cases for natural language processing, computer vision, fraud detection, conversational devices, and more Use automated machine learning to implement a specific subset of use cases with SageMaker Autopilot Dive deep into the complete model development lifecycle for a BERT-based NLP use case including data ingestion, analysis, model training, and deployment Tie everything together into a repeatable machine learning operations pipeline Explore real-time ML, anomaly detection, and streaming analytics on data streams with Amazon Kinesis and Managed Streaming for Apache Kafka Learn security best practices for data science projects and workflows including identity and access management, authentication, authorization, and more

Learn the basics of Data Science through an easy to understand conceptual framework and immediately practice using RapidMiner platform. Whether you are brand new to data science or working on your tenth project, this book will show you how to analyze data, uncover hidden patterns and relationships to aid important decisions and predictions. Data Science has become an essential tool to extract value from data for any organization that collects, stores and processes data as part of its operations. This book is ideal for business users, data analysts, business analysts, engineers, and analytics professionals and for anyone who works with data. You'll be able to: Gain the necessary knowledge of different data science techniques to extract value from data. Master the concepts and inner workings of 30 commonly used powerful data science algorithms. Implement step-by-step data science process using using RapidMiner, an open source GUI based data science platform Data Science techniques covered: Exploratory data analysis, Visualization, Decision trees, Rule induction, k-nearest neighbors, Naïve Bayesian classifiers, Artificial neural networks, Deep learning, Support vector machines, Ensemble models, Random forests, Regression, Recommendation engines, Association analysis, K-Means and Density based clustering, Self organizing maps, Text mining, Time series forecasting, Anomaly detection, Feature selection and more... Contains fully updated content on data science, including tactics on how to mine business data for information Presents simple explanations for over twenty powerful data science techniques Enables the practical use of data science algorithms without the need for programming Demonstrates processes with practical use cases Introduces each algorithm or technique and explains the workings of a data science algorithm in plain language Describes the commonly used setup options for the open source tool RapidMiner

Frank Kane's hands-on Spark training course, based on his bestselling Taming Big Data with Apache Spark and Python video, now available in a book. Understand and analyze large data sets using Spark on a single system or on a cluster. About This Book Understand how Spark can be distributed across computing clusters Develop and run Spark jobs efficiently using Python A hands-on tutorial by Frank Kane with over 15 real-world examples teaching you Big Data processing with Spark Who This Book Is For If you are a data scientist or data analyst who wants to learn Big Data processing using Apache Spark and Python, this book is for you. If you have some programming experience in Python, and want to learn how to process large amounts of data using Apache Spark, Frank Kane's Taming Big Data with Apache Spark and Python will also help you. What You Will Learn Find out how you can identify Big Data problems as Spark problems Install and run Apache Spark on your computer or on a cluster Analyze large data sets across many CPUs using Spark's Resilient Distributed Datasets Implement machine learning on Spark using the MLlib library Process continuous streams of data in real time using the Spark streaming module Perform complex network analysis using Spark's GraphX library Use Amazon's Elastic MapReduce service to run your Spark jobs on a cluster In Detail Frank Kane's Taming Big Data with Apache Spark and Python is your companion to learning Apache Spark in a hands-on manner. Frank will start you off by teaching you how to set up Spark on a single system or on a cluster, and you'll soon

move on to analyzing large data sets using Spark RDD, and developing and running effective Spark jobs quickly using Python. Apache Spark has emerged as the next big thing in the Big Data domain – quickly rising from an ascending technology to an established superstar in just a matter of years. Spark allows you to quickly extract actionable insights from large amounts of data, on a real-time basis, making it an essential tool in many modern businesses. Frank has packed this book with over 15 interactive, fun-filled examples relevant to the real world, and he will empower you to understand the Spark ecosystem and implement production-grade real-time Spark projects with ease. Style and approach Frank Kane's Taming Big Data with Apache Spark and Python is a hands-on tutorial with over 15 real-world examples carefully explained by Frank in a step-by-step manner. The examples vary in complexity, and you can move through them at your own pace.

Data is bigger, arrives faster, and comes in a variety of formats—and it all needs to be processed at scale for analytics or machine learning. But how can you process such varied workloads efficiently? Enter Apache Spark. Updated to include Spark 3.0, this second edition shows data engineers and data scientists why structure and unification in Spark matters. Specifically, this book explains how to perform simple and complex data analytics and employ machine learning algorithms. Through step-by-step walk-throughs, code snippets, and notebooks, you'll be able to: Learn Python, SQL, Scala, or Java high-level Structured APIs Understand Spark operations and SQL Engine Inspect, tune, and debug Spark operations with Spark configurations and Spark UI Connect to data sources: JSON, Parquet, CSV, Avro, ORC, Hive, S3, or Kafka Perform analytics on batch and streaming data using Structured Streaming Build reliable data pipelines with open source Delta Lake and Spark Develop machine learning pipelines with MLlib and productionize models using MLflow

Deep learning is changing everything. This machine-learning method has already surpassed traditional computer vision techniques, and the same is happening with NLP. If you're looking to bring deep learning into your domain, this practical book will bring you up to speed on key concepts using Facebook's PyTorch framework. Once author Ian Pointer helps you set up PyTorch on a cloud-based environment, you'll learn how use the framework to create neural architectures for performing operations on images, sound, text, and other types of data. By the end of the book, you'll be able to create neural networks and train them on multiple types of data. Learn how to deploy deep learning models to production Explore PyTorch use cases in companies other than Facebook Learn how to apply transfer learning to images Apply cutting-edge NLP techniques using a model trained on Wikipedia

Use Hadoop to solve business problems by learning from a rich set of real-life case studies About This Book Solve real-world business problems using Hadoop and other Big Data technologies Build efficient data lakes in Hadoop, and develop systems for various business cases like improving marketing campaigns, fraud detection, and more Power packed with six case studies to get you going with Hadoop for Business Intelligence Who This Book Is For If you are interested in building efficient business solutions using Hadoop, this is the book for you This book assumes that you have basic knowledge of Hadoop, Java, and any scripting language. What You Will Learn Learn about the evolution of Hadoop as the big data platform Understand the basics of Hadoop architecture Build a 360 degree view of your customer using Sqoop and Hive Build and run classification models on Hadoop using BigML Use Spark and Hadoop to build a fraud detection system Develop a churn detection system using Java and MapReduce Build an IoT-based data collection and visualization system Get to grips with building a Hadoop-based Data Lake for large enterprises Learn about the coexistence of NoSQL and In-Memory databases in the Hadoop ecosystem In Detail If you have a basic understanding of Hadoop and want to put your knowledge to use to build fantastic Big Data solutions for business, then this book is for you. Build six real-life, end-to-end solutions using the tools in the Hadoop ecosystem, and take your knowledge of Hadoop to the next level. Start off by understanding various business problems which can be solved using Hadoop. You will also get acquainted with the common architectural patterns which are used to build Hadoop-based solutions. Build a 360-degree view of the customer by working with different types of data, and build an efficient fraud detection system for a financial institution. You will also develop a system in Hadoop to improve the effectiveness of marketing campaigns. Build a churn detection system for a telecom company, develop an Internet of Things (IoT) system to monitor the environment in a factory, and build a data lake – all making use of the concepts and techniques mentioned in this book. The book covers other technologies and frameworks like Apache Spark, Hive, Sqoop, and more, and how they can be used in conjunction with Hadoop. You will be able to try out the solutions explained in the book and use the knowledge gained to extend them further in your own problem space. Style and approach This is an example-driven book where each chapter covers a single business problem and describes its solution by explaining the structure of a dataset and tools required to process it. Every project is demonstrated with a step-by-step approach, and explained in a very easy-to-understand manner.

Journalist Walls grew up with parents whose ideals and stubborn nonconformity were their curse and their salvation. Rex and Rose Mary and their four children lived like nomads, moving among Southwest desert towns, camping in the mountains. Rex was a charismatic, brilliant man who, when sober, captured his children's imagination, teaching them how to embrace life fearlessly. Rose Mary painted and wrote and couldn't stand the responsibility of providing for her family. When the money ran out, the Walls retreated to the dismal West Virginia mining town Rex had tried to escape. As the dysfunction escalated, the children had to fend for themselves, supporting one another as they found the resources and will to leave home. Yet Walls describes her parents with deep affection in this tale of unconditional love in a family that, despite its profound flaws, gave her the fiery determination to carve out a successful life. -- From publisher description.

Learn how to build recommender systems from one of Amazon's pioneers in the field. Frank Kane spent over nine years at Amazon, where he managed and led the development of many of Amazon's personalized product recommendation technologies. You've seen automated recommendations everywhere - on Netflix's home page, on YouTube, and on Amazon as these machine learning algorithms learn about your unique interests, and show the best products or content for you as an individual. These technologies have become central to the largest, most prestigious tech employers out there, and by understanding how they work, you'll become very valuable to them. This book is adapted from

Frank's popular online course published by Sundog Education, so you can expect lots of visual aids from its slides and a conversational, accessible tone throughout the book. The graphics and scripts from over 300 slides are included, and you'll have access to all of the source code associated with it as well. We'll cover tried and true recommendation algorithms based on neighborhood-based collaborative filtering, and work our way up to more modern techniques including matrix factorization and even deep learning with artificial neural networks. Along the way, you'll learn from Frank's extensive industry experience to understand the real-world challenges you'll encounter when applying these algorithms at large scale and with real-world data. This book is very hands-on; you'll develop your own framework for evaluating and combining many different recommendation algorithms together, and you'll even build your own neural networks using Tensorflow to generate recommendations from real-world movie ratings from real people. We'll cover:

- Building a recommendation engine
- Evaluating recommender systems
- Content-based filtering using item attributes
- Neighborhood-based collaborative filtering with user-based, item-based, and KNN CF
- Model-based methods including matrix factorization and SVD
- Applying deep learning, AI, and artificial neural networks to recommendations
- Session-based recommendations with recursive neural networks
- Scaling to massive data sets with Apache Spark machine learning, Amazon DSSTNE deep learning, and AWS SageMaker with factorization machines
- Real-world challenges and solutions with recommender systems
- Case studies from YouTube and Netflix
- Building hybrid, ensemble recommenders

This comprehensive book takes you all the way from the early days of collaborative filtering, to bleeding-edge applications of deep neural networks and modern machine learning techniques for recommending the best items to every individual user. The coding exercises for this book use the Python programming language. We include an intro to Python if you're new to it, but you'll need some prior programming experience in order to use this book successfully. We also include a short introduction to deep learning, Tensorflow, and Keras if you are new to the field of artificial intelligence, but you'll need to be able to understand new computer algorithms. Dive in, and learn about one of the most interesting and lucrative applications of machine learning and deep learning there is!

What is machine learning? -- Automating machine learning -- Specify business problem -- Acquire subject matter expertise -- Define prediction target -- Decide on unit of analysis -- Success, risk, and continuation -- Accessing and storing data -- Data integration -- Data transformations -- Summarization -- Data reduction and splitting -- Startup processes -- Feature understanding and selection -- Build candidate models -- Understanding the process -- Evaluate model performance -- Comparing model pairs -- Interpret model -- Communicate model insights -- Set up prediction system -- Document modeling process for reproducibility -- Create model monitoring and maintenance plan -- Seven types of target leakage in machine learning and an exercise -- Time-aware modeling -- Time-series modeling.

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

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.

Build modular applications that are easy to scale using the most powerful components and design patterns that React can offer you right now About This Book Dive into the core patterns and components of React.js in order to master your application's design Improve their debugging skills using the DevTools This book is packed with easy-to-follow examples that can be used to create reusable code and extensible designs Who This Book Is For If you want to increase your understanding of React and apply it to real-life application development, then this book is for you. What You Will Learn Write clean and maintainable code Create reusable components applying consolidated techniques Use React effectively in the browser and node Choose the right styling approach according to the needs of the applications Use server-side rendering to make applications load faster Build high-performing applications by optimizing components In Detail Taking a complete journey through the most valuable design patterns in React, this book demonstrates how to apply design patterns and best practices in real-life situations, whether that's for new or already existing projects. It will help you to make your applications more flexible, perform better, and easier to maintain – giving your workflow a huge boost when it comes to speed without reducing quality. We'll begin by understanding the internals of React before gradually moving on to writing clean and maintainable code. We'll build components that are reusable across the application, structure applications, and create forms that actually work. Then we'll style React components and optimize them to make applications faster and more responsive. Finally, we'll write tests effectively and you'll learn how to contribute to React and its ecosystem. By the end of the book, you'll be saved from a lot of trial and error and developmental headaches, and you will be on the road to becoming a React expert. Style and approach The design patterns in the book are explained using real-world, step-by-step examples. For each design pattern, there are hints about when to use it and when to look for something more suitable. This book can also be used as a practical guide, showing you how to leverage design patterns.

Definitive guide to lightning fast data processing for distributed systems with Apache Flink About This Book* Build your expertise in processing realtime data with Apache Flink and its ecosystem* Gain insights into the working of all components of Apache Flink such as FlinkML, Gelly, and Table API Filled with real world use cases,* Your guide to take advantage of Apache Flink for solving real world problems Who This Book Is For Big data developers who are looking to process batch and real-time data on distributed systems. Basic knowledge of Hadoop and big data is assumed. Reasonable knowledge of Java or Scala is expected. What You Will Learn* Learn how to build end to end real time analytics projects* Integrate with existing big data stack and utilize existing infrastructure.* Build predictive analytics applications using FlinkML* Use graph library to perform graph querying and search. In Detail With the advent of massive computer systems, organizations in different domains generate large amounts of data at a realtime basis. The latest entrant to big data processing, Apache Flink, is designed to process continuous streams of data at a lightning fast pace. This book will be your definitive guide to batch and stream data processing with Apache Flink. The book begins with introducing the Apache Flink ecosystem, setting it up and using the DataSet and DataStream API for processing batch and streaming datasets. Bringing the power of SQL to Flink, this book will then explore the Table API for querying and manipulating data. In the latter half of the book, readers will get to learn the remaining ecosystem of Apache Flink to achieve complex tasks such as event processing, machine learning, and graph processing. The final part of the book would consist of topics such as scaling Flink solutions, performance optimization and integrating Flink with other tools such as Elasticsearch. Whether you want to dive deeper into Apache Flink, or want to investigate how to get more out of this powerful technology, you'll find everything inside

Build machine and deep learning systems with the newly released TensorFlow 2 and Keras for the lab, production, and mobile devices Key Features Introduces and then uses TensorFlow 2 and Keras right from the start Teaches key machine and deep learning techniques Understand the fundamentals of deep learning and machine learning through clear explanations and extensive code samples Book Description Deep Learning with TensorFlow 2 and Keras, Second Edition teaches neural networks and deep learning techniques alongside TensorFlow (TF) and Keras. You'll learn how to write deep learning applications in the most powerful, popular, and scalable machine learning stack available. TensorFlow is the machine learning library of choice for professional applications, while Keras offers a simple and powerful Python API for accessing TensorFlow. TensorFlow 2 provides full Keras integration, making advanced machine learning easier and more convenient than ever before. This book also introduces neural networks with TensorFlow, runs through the main applications (regression, ConvNets (CNNs), GANs, RNNs, NLP), covers two working example apps, and then dives into TF in production, TF mobile, and using TensorFlow with AutoML. What you will learn Build machine learning and deep learning systems with TensorFlow 2 and the Keras API Use Regression analysis, the most popular approach to machine learning Understand ConvNets (convolutional neural networks) and how they are essential for deep learning systems such as image classifiers Use GANs (generative adversarial networks) to create new data that fits with existing patterns Discover RNNs (recurrent neural networks) that can process sequences of input intelligently, using one part of a sequence to correctly interpret another Apply deep learning to natural human language and interpret natural language texts to produce an appropriate response Train your models on the cloud and put TF to work in real environments Explore how Google tools can automate simple ML workflows without the need for complex modeling Who this book is for This book is for Python developers and data scientists who want to build machine learning and deep learning systems with TensorFlow. Whether or not you have done machine learning before, this book gives you the theory and practice required to use Keras, TensorFlow 2, and AutoML to build machine learning systems.

This book constitutes the refereed proceedings of the Second International Symposium on Benchmarking, Measuring, and Optimization, Bench 2019, held in Denver, CO, USA, in November 2019. The 20 full papers and 11 short papers presented were carefully reviewed and selected from 79 submissions. The papers are organized in topical sections named: Best Paper Session; AI Challenges on Cambircon using AIBench; AI Challenges on RISC-V using AIBench; AI Challenges on X86 using AIBench; AI Challenges on 3D Face Recognition using AIBench; Benchmark; AI and Edge; Big Data; Datacenter; Performance Analysis; Scientific Computing.

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.

A solution-based guide to put your deep learning models into production with the power of Apache Spark Key Features Discover practical recipes for distributed deep learning with Apache Spark Learn to use libraries such as Keras and TensorFlow Solve problems in order to train your deep learning models on Apache Spark Book Description With deep learning gaining rapid mainstream adoption in modern-day industries, organizations are looking for ways to unite popular big data tools with highly efficient deep learning libraries. As a result, this will help deep learning models train with higher efficiency and speed. With the help of the Apache Spark Deep Learning Cookbook, you'll work through specific recipes to generate outcomes for deep learning algorithms, without getting bogged down in theory. From setting up Apache Spark for deep learning to implementing types of neural net, this book tackles both common and not so common problems to perform deep learning on a distributed environment. In addition to this, you'll get access to deep learning code within Spark that can be reused to answer similar problems or tweaked to answer slightly different problems. You will also learn how to stream and cluster your data with Spark. Once you have got to grips with the basics, you'll explore how to implement and deploy deep learning models, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) in Spark, using popular libraries such as TensorFlow and Keras. By the end of the book, you'll have the expertise to train and deploy efficient deep learning models on Apache Spark. What you will learn Set up a fully functional Spark environment Understand practical machine learning and deep learning concepts Apply built-in machine learning libraries within Spark Explore libraries that are compatible with TensorFlow and Keras Explore NLP models such as Word2vec and TF-IDF on Spark Organize dataframes for deep learning evaluation Apply testing and training modeling to ensure accuracy Access readily available code that may be reusable Who this book is for If you're looking for a practical and highly useful resource for implementing efficiently distributed deep learning models with Apache Spark, then the Apache Spark Deep Learning Cookbook is for you. Knowledge of the core machine learning concepts and a basic understanding of the Apache Spark framework is required to get the best out of this book. Additionally, some programming knowledge in Python is a plus.

Get command of your organizational Big Data using the power of data science and analytics Key Features A perfect companion to boost your Big Data storing, processing, analyzing skills to help you take informed business decisions Work with the best tools such as Apache Hadoop, R, Python, and Spark for NoSQL platforms to perform massive online analyses Get expert tips on statistical inference, machine learning, mathematical modeling, and data visualization for Big Data Book Description Big Data analytics relates to the strategies used by organizations to collect, organize and analyze large amounts of data to uncover valuable business insights that otherwise cannot be analyzed through traditional systems. Crafting an enterprise-scale cost-efficient Big Data and machine learning solution to uncover insights and value from your organization's data is a challenge. Today, with hundreds of new Big Data systems, machine learning packages and BI Tools, selecting the right combination of technologies is an even greater challenge. This book will help you do that. With the help of this guide, you will be able to bridge the gap between the theoretical world of technology with the practical ground reality of building corporate Big Data and data science platforms. You will get hands-on exposure to Hadoop and Spark, build machine learning dashboards using R and R Shiny, create web-based apps using NoSQL databases such as MongoDB and even learn how to write R code for neural networks. By the end of the book, you will have a very clear and concrete understanding of what Big Data analytics means, how it drives revenues for organizations, and how you can develop your own Big Data analytics solution using different tools and methods articulated in this book. What you will learn - Get a 360-degree view into the world of Big Data, data science and machine learning - Broad range of technical and business Big Data analytics topics that caters to the interests of the technical experts as well as corporate IT executives - Get hands-on experience with industry-standard Big Data and machine learning tools such as Hadoop, Spark, MongoDB, KDB+ and R - Create production-grade machine learning BI Dashboards using R and R Shiny with step-by-step instructions - Learn how to combine open-source Big Data, machine learning and BI Tools to create low-cost business analytics applications - Understand corporate strategies for successful Big Data and data science projects - Go beyond general-purpose analytics to develop cutting-edge Big Data applications using emerging technologies Who this book is for The book is intended for existing and aspiring Big Data professionals who wish to become the go-to person in their organization when it comes to Big Data architecture, analytics, and governance. While no prior knowledge of Big Data or related technologies is assumed, it will be helpful to have some programming experience.

Apache Spark Machine Learning Blueprints Apache Spark Machine Learning Blueprints Packt Publishing Ltd

[Copyright: eae832ef888041630b71d0a879abefa3](https://www.packtpub.com/book/big-data-and-business-intelligence/9781617130416/9781617130416041630b71d0a879abefa3)