

Hadoop Essentials A Quantitative Approach

This book takes you on a fantastic journey to discover the attributes of big data using Apache Hive. Key Features Grasp the skills needed to write efficient Hive queries to analyze the Big Data Discover how Hive can coexist and work with other tools within the Hadoop ecosystem Uses practical, example-oriented scenarios to cover all the newly released features of Apache Hive 2.3.3 Book Description In this book, we prepare you for your journey into big data by firstly introducing you to backgrounds in the big data domain, alongwith the process of setting up and getting familiar with your Hive working environment. Next, the book guides you through discovering and transforming the values of big data with the help of examples. It also hones your skills in using the Hive language in an efficient manner. Toward the end, the book focuses on advanced topics, such as performance, security, and extensions in Hive, which will guide you on exciting adventures on this worthwhile big data journey. By the end of the book, you will be familiar with Hive and able to work efficiently to find solutions to big data problems What you will learn Create and set up the Hive environment Discover how to use Hive's definition language to describe data Discover interesting data by joining and filtering datasets in Hive Transform data by using Hive sorting, ordering, and functions Aggregate and sample data in different ways Boost Hive query performance and enhance data security in Hive Customize Hive to your needs by using user-defined functions and integrate it with other tools Who this book is for If you are a data analyst, developer, or simply someone who wants to quickly get started with Hive to explore and analyze Big Data in Hadoop, this is the book for you. Since Hive is an SQL-like language, some previous experience with SQL will be useful to get the most out of this book.

A practical guide to implementing your enterprise data lake using Lambda Architecture as the base About This Book Build a full-fledged data lake for your organization with popular big data technologies using the Lambda architecture as the base Delve into the big data technologies required to meet modern day business strategies A highly practical guide to implementing enterprise data lakes with lots of examples and real-world use-cases Who This Book Is For Java developers and architects who would like to implement a data lake for their enterprise will find this book useful. If you want to get hands-on experience with the Lambda Architecture and big data technologies by implementing a practical solution using these technologies, this book will also help you. What You Will Learn Build an enterprise-level data lake using the relevant big data technologies Understand the core of the Lambda architecture and how to apply it in an enterprise Learn the technical details around Sqoop and its functionalities Integrate Kafka with Hadoop components to acquire enterprise data Use flume with streaming technologies for stream-based processing Understand stream-based processing with reference to Apache Spark Streaming Incorporate Hadoop components and know the advantages they provide for enterprise data lakes Build fast, streaming, and high-performance applications using Elasticsearch Make your data ingestion process consistent across various data formats with configurability Process your data to derive intelligence using machine learning algorithms In Detail The term "Data Lake" has recently emerged as a prominent term in the big data industry. Data scientists can make use of it in deriving meaningful insights that can be used by businesses to redefine or transform the way they operate. Lambda architecture is also emerging as one of the very eminent patterns in the big data landscape, as it not only helps to derive useful information from historical data but also correlates real-time data to enable business to take critical decisions. This book tries to bring these two important aspects — data lake and lambda architecture—together. This book is divided into three main sections. The first introduces you to the concept of data lakes, the importance of data lakes in enterprises, and getting you up-to-speed with the Lambda architecture. The second section delves into the principal components of building a data lake using the Lambda architecture. It introduces you to popular big data technologies such as Apache Hadoop, Spark, Sqoop, Flume, and Elasticsearch. The third section is a highly practical demonstration of putting it all together, and shows you how an enterprise data lake can be implemented, along with several real-world use-cases. It also shows you how other peripheral components can be added to the lake to make it more efficient. By the end of this book, you will be able to choose the right big data technologies using the lambda architectural patterns to build your enterprise data lake. Style and approach The book takes a pragmatic approach, showing ways to leverage big data technologies and lambda architecture to build an enterprise-level data lake.

This book presents machine learning models and algorithms to address big data classification problems. Existing machine learning techniques like the decision tree (a hierarchical approach), random forest (an ensemble hierarchical approach), and deep learning (a layered approach) are highly suitable for the system that can handle such problems. This book helps readers, especially students and newcomers to the field of big data and machine learning, to gain a quick understanding of the techniques and technologies; therefore, the theory, examples, and programs (Matlab and R) presented in this book have been simplified, hardcoded, repeated, or spaced for improvements. They provide vehicles to test and understand the complicated concepts of various topics in the field. It is expected that the readers adopt these programs to experiment with the examples, and then modify or write their own programs toward advancing their knowledge for solving more complex and challenging problems. The presentation format of this book focuses on simplicity, readability, and dependability so that both undergraduate and graduate students as well as new researchers, developers, and practitioners in this field can easily trust and grasp the concepts, and learn them effectively. It has been written to reduce the mathematical complexity and help the vast majority of readers to understand the topics and get interested in the field. This book consists of four parts, with the total of 14 chapters. The first part mainly focuses on the topics that are needed to help analyze and understand data and big data. The second part covers the topics that can explain the systems required for processing big data. The third part presents the topics required to understand and select machine learning techniques to classify big data. Finally, the fourth part concentrates on the topics that explain the scaling-up machine learning, an important solution for modern big data problems. The definitive guide for statisticians and data scientists who understand the advantages of becoming proficient in both R and Python The first book of its kind, Python for R Users: A Data Science Approach makes it easy for R programmers to code in Python and Python users to program in R. Short on theory and long on actionable analytics, it provides readers with a detailed comparative introduction and overview of both languages and features concise tutorials with command-by-command translations—complete with sample code—of R to Python and Python to R. Following an introduction to both languages, the author cuts to the chase with step-by-step coverage of the full range of pertinent programming features and functions, including data input, data inspection/data

quality, data analysis, and data visualization. Statistical modeling, machine learning, and data mining—including supervised and unsupervised data mining methods—are treated in detail, as are time series forecasting, text mining, and natural language processing.

- Features a quick-learning format with concise tutorials and actionable analytics
- Provides command-by-command translations of R to Python and vice versa
- Incorporates Python and R code throughout to make it easier for readers to compare and contrast features in both languages
- Offers numerous comparative examples and applications in both programming languages
- Designed for use for practitioners and students that know one language and want to learn the other
- Supplies slides useful for teaching and learning either software on a companion website

Python for R Users: A Data Science Approach is a valuable working resource for computer scientists and data scientists that know R and would like to learn Python or are familiar with Python and want to learn R. It also functions as textbook for students of computer science and statistics. A. Ohri is the founder of Decisionstats.com and currently works as a senior data scientist. He has advised multiple startups in analytics off-shoring, analytics services, and analytics education, as well as using social media to enhance buzz for analytics products. Mr. Ohri's research interests include spreading open source analytics, analyzing social media manipulation with mechanism design, simpler interfaces for cloud computing, investigating climate change and knowledge flows. His other books include R for Business Analytics and R for Cloud Computing.

Health Communication

5th International Conference, SIMBig 2018, Lima, Peru, September 3 – 5, 2018, Proceedings

IIENC 2020

Machine Learning Algorithms

Data Science for Economics and Finance

An End-to-end Approach

Management Information Systems

As the Internet becomes increasingly interconnected with modern society, the transition to online business has developed into a prevalent form of commerce. While there exist various advantages and disadvantages to online business, it plays a major role in contemporary business methods. Improving E-Commerce Web Applications Through Business Intelligence Techniques provides emerging research on the core areas of e-commerce web applications. While highlighting the use of data mining, search engine optimization, and online marketing to advance online business, readers will learn how the role of online commerce is becoming more prevalent in modern business. This book is an important resource for vendors, website developers, online customers, and scholars seeking current research on the development and use of e-commerce.

Ready to unlock the power of your data? With this comprehensive guide, you'll learn how to build and maintain reliable, scalable, distributed systems with Apache Hadoop. This book is ideal for programmers looking to analyze datasets of any size, and for administrators who want to set up and run Hadoop clusters. You'll find illuminating case studies that demonstrate how Hadoop is used to solve specific problems. This third edition covers recent changes to Hadoop, including material on the new MapReduce API, as well as MapReduce 2 and its more flexible execution model (YARN). Store large datasets with the Hadoop Distributed File System (HDFS) Run distributed computations with MapReduce Use Hadoop's data and I/O building blocks for compression, data integrity, serialization (including Avro), and persistence Discover common pitfalls and advanced features for writing real-world MapReduce programs Design, build, and administer a dedicated Hadoop cluster—or run Hadoop in the cloud Load data from relational databases into HDFS, using Sqoop Perform large-scale data processing with the Pig query language Analyze datasets with Hive, Hadoop's data warehousing system Take advantage of HBase for structured and semi-structured data, and ZooKeeper for building distributed systems

Management Information Systems provides comprehensive and integrative coverage of essential new technologies, information system applications, and their impact on business models and managerial decision-making in an exciting and interactive manner. The twelfth edition focuses on the major changes that have been made in information technology over the past two years, and includes new opening, closing, and Interactive Session cases.

Written as a tutorial to explore and understand the power of R for machine learning. This practical guide that covers all of the need to know topics in a very systematic way. For each machine learning approach, each step in the process is detailed, from preparing the data for analysis to evaluating the results. These steps will build the knowledge you need to apply them to your own data science tasks. Intended for those who want to learn how to use R's machine learning capabilities and gain insight from your data. Perhaps you already know a bit about machine learning, but have never used R; or perhaps you know a little R but are new to machine learning. In either case, this book will get you up and running quickly. It would be helpful to have a bit of familiarity with basic programming concepts, but no prior experience is required.

Discovering, Analyzing, Visualizing and Presenting Data

Hadoop Essentials

Methodologies and Applications

Special Collection

Python Data Science Essentials

Machine Learning with R

A Quantitative Approach

The guide to targeting and leveraging business opportunities using big data & analytics By leveraging big data & analytics, businesses create the potential to better understand, manage, and strategically exploiting the complex dynamics of customer behavior. *Analytics in a Big Data World* reveals how to tap into the powerful tool of data analytics to create a strategic advantage and identify new business opportunities. Designed to be an accessible resource, this essential book does not include exhaustive coverage of all analytical techniques, instead focusing on analytics techniques that really provide added value in business environments. The book draws on author Bart Baesens' expertise on the topics of big data, analytics and its applications in e.g. credit risk, marketing, and fraud to provide a clear roadmap for organizations that want to use data analytics to their advantage, but need a good starting point. Baesens has conducted extensive research on big data, analytics, customer relationship management, web analytics, fraud detection, and credit risk management, and uses this experience to bring clarity to a complex topic. Includes

numerous case studies on risk management, fraud detection, customer relationship management, and web analytics Offers the results of research and the author's personal experience in banking, retail, and government Contains an overview of the visionary ideas and current developments on the strategic use of analytics for business Covers the topic of data analytics in easy-to-understand terms without an undo emphasis on mathematics and the minutiae of statistical analysis For organizations looking to enhance their capabilities via data analytics, this resource is the go-to reference for leveraging data to enhance business capabilities.

Health Communication: Strategies and Skills for a New Era provides a practical process model for developing a health communication intervention. The book also explores exposure to media and how it shapes our conceptions of health and illness.

Using a life stages and environments approach, the book touches on the patient role and how we 'hear' information from health care providers as well as guidance on how to be a thoughtful consumer of health information.

This book constitutes revised selected papers from the 25th International Symposium on Graph Drawing and Network Visualization, GD 2017, held in Boston, MA, USA, in September 2017. The 34 full and 9 short papers presented in this volume were carefully reviewed and selected from 87 submissions. Also included in this book are 2 abstracts of keynote presentations, 16 poster abstracts, and 1 contest report. The papers are organized in topical sections named: straight-line representations; obstacles and visibility; topological graph theory; orthogonal representations and book embeddings; evaluations; tree drawings; graph layout designs; point-set embeddings; special representations; and beyond planarity.

"This text should be required reading for everyone in contemporary business." --Peter Woodhull, CEO, Modus21 "The one book that clearly describes and links Big Data concepts to business utility." --Dr. Christopher Starr, PhD "Simply, this is the best Big Data book on the market!" --Sam Rostam, Cascadian IT Group "...one of the most contemporary approaches I've seen to Big Data fundamentals..." --Joshua M. Davis, PhD

The Definitive Plain-English Guide to Big Data for Business and Technology Professionals

Big Data Fundamentals provides a pragmatic, no-nonsense introduction to Big Data. Best-selling IT author Thomas Erl and his team clearly explain key Big Data concepts, theory and terminology, as well as fundamental technologies and techniques. All coverage is supported with case study examples and numerous simple diagrams. The authors begin by explaining how Big Data can propel an organization forward by solving a spectrum of previously intractable business problems. Next, they demystify key analysis techniques and technologies and show how a Big Data solution environment can be built and integrated to offer competitive advantages.

Discovering Big Data's fundamental concepts and what makes it different from previous forms of data analysis and data science

Understanding the business motivations and drivers behind Big Data adoption, from operational improvements through innovation

Planning strategic, business-driven Big Data initiatives Addressing considerations such as data management, governance, and security

Recognizing the 5 "V" characteristics of datasets in Big Data environments: volume, velocity, variety, veracity, and value

Clarifying Big Data's relationships with OLTP, OLAP, ETL, data warehouses, and data marts Working with Big Data in structured, unstructured, semi-structured, and metadata formats

Increasing value by integrating Big Data resources with corporate performance monitoring Understanding how Big Data leverages distributed and parallel processing Using NoSQL and other technologies to meet Big Data's distinct data processing requirements

Leveraging statistical approaches of quantitative and qualitative analysis Applying computational analysis methods, including machine learning

TEXT ANALYTICS WITH SAS

Managing the Digital Firm

Data Science and Big Data Analytics

A Problem-Solver's Guide to Building Real-World Intelligent Systems

A Hands-On Introduction to Hacking

Data Lake for Enterprises

Developing Enterprise Applications With Spring

Dig deep into the data with a hands-on guide to machine learning with updated examples and more! Machine Learning: Hands-On for Developers and Technical Professionals provides hands-on instruction and fully-coded working examples for the most common machine learning techniques used by developers and technical professionals. The book contains a breakdown of each ML variant, explaining how it works and how it is used within certain industries, allowing readers to incorporate the presented techniques into their own work as they follow along. A core tenant of machine learning is a strong focus on data preparation, and a full exploration of the various types of learning algorithms illustrates how the proper tools can help any developer extract information and insights from existing data. The book includes a full complement of Instructor's Materials to facilitate use in the classroom, making this resource useful for students and as a professional reference. At its core, machine learning is a mathematical, algorithm-based technology that forms the basis of historical data mining and modern big data science. Scientific analysis of big data requires a working knowledge of machine learning, which forms predictions based on known properties learned from training data. Machine Learning is an accessible, comprehensive guide for the non-mathematician, providing clear guidance that allows readers to: Learn the languages of machine learning including Hadoop, Mahout, and Weka Understand decision trees, Bayesian networks, and artificial neural networks Implement Association Rule, Real Time, and Batch learning Develop a strategic plan for safe, effective, and efficient machine learning By learning to construct a system that can learn from data, readers can increase their utility across industries. Machine learning sits at the core of deep dive data analysis and visualization, which is increasingly in demand as companies discover the goldmine hiding in their existing data. For the tech professional involved in data science, Machine Learning: Hands-On for Developers and Technical Professionals provides the skills and techniques required to dig deeper.

Master the essential skills needed to recognize and solve complex problems with machine learning and deep learning. Using real-world examples that leverage the popular Python machine learning ecosystem, this book is your perfect companion for learning the art and science of machine learning to become a successful practitioner. The concepts, techniques, tools, frameworks, and methodologies used in this book will teach you how to think,

design, build, and execute machine learning systems and projects successfully. Practical Machine Learning with Python follows a structured and comprehensive three-tiered approach packed with hands-on examples and code. Part 1 focuses on understanding machine learning concepts and tools. This includes machine learning basics with a broad overview of algorithms, techniques, concepts and applications, followed by a tour of the entire Python machine learning ecosystem. Brief guides for useful machine learning tools, libraries and frameworks are also covered. Part 2 details standard machine learning pipelines, with an emphasis on data processing analysis, feature engineering, and modeling. You will learn how to process, wrangle, summarize and visualize data in its various forms. Feature engineering and selection methodologies will be covered in detail with real-world datasets followed by model building, tuning, interpretation and deployment. Part 3 explores multiple real-world case studies spanning diverse domains and industries like retail, transportation, movies, music, marketing, computer vision and finance. For each case study, you will learn the application of various machine learning techniques and methods. The hands-on examples will help you become familiar with state-of-the-art machine learning tools and techniques and understand what algorithms are best suited for any problem. Practical Machine Learning with Python will empower you to start solving your own problems with machine learning today! What You'll Learn Execute end-to-end machine learning projects and systems Implement hands-on examples with industry standard, open source, robust machine learning tools and frameworks Review case studies depicting applications of machine learning and deep learning on diverse domains and industries Apply a wide range of machine learning models including regression, classification, and clustering. Understand and apply the latest models and methodologies from deep learning including CNNs, RNNs, LSTMs and transfer learning. Who This Book Is For IT professionals, analysts, developers, data scientists, engineers, graduate students

This textbook adopts a unique approach to helping developers and CS students learn Hadoop MapReduce programming fast in an easy-to-setup, virtual 4-node Linux YARN cluster on a Windows laptop. Rather than filled with disjointed, piecemeal code snippets to show Hadoop MapReduce programming features one at a time, it is designed to place your total Hadoop MapReduce programming learning process in a common application context of mining customer spending patterns ensconced in large volumes of credit card transaction record data. Precise, end-to-end procedures are given to help you set up your Hadoop MapReduce development environment quickly on Eclipse with Maven on Windows. Step-by-step procedures are also given on how to set up a four-node Linux cluster at minimum so that you can run your MapReduce programs not only in local but also in standalone and fully distributed mode on a real cluster. In fact, all MapReduce programs presented in the book have been tested and verified on such a Linux cluster. This textbook mainly focuses on teaching Hadoop MapReduce programming in a scientific, objective, quantitative approach. Rather than heavily relying on subjective, verbose (and sometimes even pompous) textual descriptions with sparse code snippets, this textbook uses Hadoop Java APIs, Hadoop configuration parameters, complete MapReduce programs and their execution logs and outputs to demonstrate how Hadoop MapReduce framework works and how to write MapReduce programs. Specifically, this text covers the following subjects: * Introduction to Hadoop * Setting up a Linux Hadoop Cluster * The Hadoop Distributed FileSystem * MapReduce Job Orchestration and Workflows * Basic MapReduce Programming * Advanced MapReduce Programming * Hadoop Streaming * Hadoop Administration No matter what role you play on your team, this text can help you gain truly applicable Hadoop skills in a most effective and efficient manner. The book can also be used as a supplementary textbook for a distributed computing or Hadoop course offered to upper-division CS students.

SAS software provides many different techniques to monitor in real time and investigate your data, and several groundbreaking papers have been written to demonstrate how to use these techniques. Topics covered illustrate the power of SAS solutions that are available as tools for fraud analytics, highlighting a variety of domains, including money laundering, financial crime, and terrorism. Also available free as a PDF from: sas.com/books.

Essential techniques to help you process, and get unique insights from, big data, 2nd Edition

Machine Learning Models and Algorithms for Big Data Classification

Practical Machine Learning with Python

Concepts, Drivers & Techniques

Apache Hive Essentials

Master the complete predictive analytics process, from problem definition to model deployment

Information Management and Big Data

Big Data is the biggest game-changing opportunity for marketing and sales since the Internet went mainstream almost 20 years ago. The data big bang has unleashed torrents of terabytes about everything from customer behaviors to weather patterns to demographic consumer shifts in emerging markets. This collection of articles, videos, interviews, and slideshares highlights the most important lessons for companies looking to turn data into above-market growth: Using analytics to identify valuable business opportunities from the data to drive decisions and improve marketing return on investment (MROI) Turning those insights into well-designed products and offers that delight customers Delivering those products and offers effectively to the marketplace. The goldmine of data represents a pivot-point moment for marketing and sales leaders. Companies that inject big data and analytics into their operations show productivity rates and profitability that are 5 percent to 6 percent higher than those of their peers. That's an advantage no company can afford to ignore.

This book constitutes the refereed proceedings of the 5th International Conference on Information Management and Big Data, SIMBig 2018, held in Lima, Peru, in September 2018. The 34 papers presented were carefully reviewed and selected from 101 submissions. The papers address issues such as data mining, artificial intelligence, Natural Language Processing, information retrieval, machine learning, web mining.

SAS Programming for R Users, based on the free SAS Education course of the same name, is designed for experienced R users who want to transfer their programming skills to SAS. Emphasis is on programming and not statistical theory or interpretation. You will learn how to write programs in SAS that replicate familiar functions and capabilities in R. This book covers a wide range of topics including the basics of the SAS programming language, how to import data, how to create new variables, random number generation, linear modeling, Interactive Matrix Language (IML), and many other SAS procedures. This book also explains how to write R code directly in the SAS code editor for seamless integration between the two tools. Exercises are provided at the end of each chapter so that you can test your knowledge and practice your programming skills.

Penetration testers simulate cyber attacks to find security weaknesses in networks, operating systems, and applications. Information security experts worldwide use penetration techniques to evaluate enterprise defenses. In Penetration Testing, security expert, researcher, and trainer Georgia Weidman introduces you to the core skills and techniques that every pentester needs. Using a virtual machine-based lab that includes Kali Linux and vulnerable operating systems, you'll run through a series of practical lessons with tools like Wireshark, Nmap, and Burp Suite. As you follow along with the labs and launch attacks, you'll experience the key stages of an actual assessment—including information gathering, finding exploitable vulnerabilities, gaining access to systems, post exploitation, and more. Learn how to: –Crack passwords and wireless network keys with brute-forcing and wordlists –Test web applications for vulnerabilities –Use the Metasploit Framework to launch exploits and write your own Metasploit modules –Automate social-engineering attacks –Bypass antivirus software –Turn access to one machine into total control of the enterprise in the post exploitation phase You'll even explore writing your own exploits. Then it's on to mobile hacking—Weidman's particular area of research—with her tool, the Smartphone Pentest Framework. With its collection of hands-on lessons that cover key tools and strategies, Penetration Testing is the introduction that every aspiring hacker needs.

Proceedings of Integrated Intelligence Enable Networks and Computing

Essentials of Business Analytics

Hadoop Beginner's Guide

Machine Learning

Graph Drawing and Network Visualization

An End-to-End Approach

Big Data

Create powerful data visualizations and unlock intelligent business insights with Tableau Key Features Explore all the latest Tableau 2019 features and redefine business analytics for your organization Create impressive data visualizations and interactive dashboards to get insights from your data Learn effective data storytelling to transform how your business leverages data and makes decisions Book Description Tableau is the gold standard of business intelligence and visual analytics tools in every industry. It enables rapid data visualization and interpretation with charts, graphs, dashboards, and much more. Updated with the latest features of Tableau, this book takes you from the foundations of the Tableau 2019 paradigm through to advanced topics. This third edition of the bestselling guide by Tableau Zen Master, Joshua Milligan, will help you come to grips with updated features, such as set actions and transparent views. Beginning with installation, you'll create your first visualizations with Tableau and then explore practical examples and advanced techniques. You'll create bar charts, tree maps, scatterplots, time series, and a variety of other visualizations. Next, you'll discover techniques to overcome challenges presented by data structure and quality and engage in effective data storytelling and decision making with business critical information. Finally, you'll be introduced to Tableau Prep, and learn how to use it to integrate and shape data for analysis. By the end of this book, you will be equipped to leverage the powerful features of Tableau 2019 for decision making. What you will learn Develop stunning visualizations that explain complexity with clarity Explore the exciting new features of Tableau Desktop and Tableau Prep Connect to various data sources to bring all your data together Uncover techniques to prep and structure your data for easy analysis Create and use calculations to solve problems and enrich analytics Master advanced topics such as sets, LOD calcs, and much more Enable smart decisions with clustering, distribution, and forecasting Share your data stories to build a culture of trust and action Who this book is for This Tableau book is for anyone who wants to understand data. If you're new to Tableau, don't worry. This book builds on the foundations to help you understand how Tableau really works and then builds on that knowledge with practical examples before moving on to advanced techniques. Working experience with databases will be useful but is not necessary to get the most out of this book.

This book presents best selected research papers presented at the First International Conference on Integrated Intelligence Enable Networks and Computing (IIENC 2020), held from May 25 to May 27, 2020, at the Institute of Technology, Gopeshwar, India (Government Institute of Uttarakhand Government and affiliated to Uttarakhand Technical University). The book includes papers in the field of intelligent computing. The book covers the areas of machine learning and robotics, signal processing and Internet of things, big data and renewable energy sources.

If you are a data analyst, developer, or simply someone who wants to use Hive to explore and analyze data in Hadoop, this is the book for you. Whether you are new to big data or an expert, with this book, you will be able to master both the basic and the advanced features of Hive. Since Hive is an SQL-like language, some previous experience with the SQL language and databases is useful to have a better understanding of this book.

Summary Hadoop in Practice, Second Edition provides over 100 tested, instantly useful techniques that will help you conquer big data, using Hadoop. This revised new edition covers changes and new features in the Hadoop core architecture, including MapReduce 2. Brand new chapters cover YARN and integrating Kafka, Impala, and Spark SQL with Hadoop. You'll also get new and updated techniques for Flume, Sqoop, and Mahout, all of which have seen major new versions recently. In short, this is the most practical, up-to-date coverage of Hadoop available anywhere. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Book It's always a good time to upgrade your Hadoop skills! Hadoop in Practice, Second Edition provides a collection of 104 tested, instantly useful techniques for analyzing real-time streams, moving data securely, machine learning, managing large-scale clusters, and taming big data using Hadoop. This completely revised edition covers changes and new features in Hadoop core, including MapReduce 2 and YARN. You'll pick up hands-on best practices for integrating Spark, Kafka, and Impala with Hadoop, and get new and updated techniques for the latest versions of Flume, Sqoop, and Mahout. In short, this is the most practical, up-to-date coverage of Hadoop available. Readers need to know a

programming language like Java and have basic familiarity with Hadoop. What's Inside Thoroughly updated for Hadoop 2 How to write YARN applications Integrate real-time technologies like Storm, Impala, and Spark Predictive analytics using Mahout and RR Readers need to know a programming language like Java and have basic familiarity with Hadoop. About the Author Alex Holmes works on tough big-data problems. He is a software engineer, author, speaker, and blogger specializing in large-scale Hadoop projects. Table of Contents PART 1 BACKGROUND AND FUNDAMENTALS Hadoop in a heartbeat Introduction to YARN PART 2 DATA LOGISTICS Data serialization—working with text and beyond Organizing and optimizing data in HDFS Moving data into and out of Hadoop PART 3 BIG DATA PATTERNS Applying MapReduce patterns to big data Utilizing data structures and algorithms at scale Tuning, debugging, and testing PART 4 BEYOND MAPREDUCE SQL on Hadoop Writing a YARN application

The Essential Guide to Data Science and its Applications

SAS Programming for R Users

Tools for Business Intelligence, data prep, and visual analytics, 3rd Edition

Principles and Paradigms

Hadoop 2 Essentials

Fraud Analytics with SAS

Python for R Users

This textbook adopts a unique approach to helping developers and CS students learn Hadoop MapReduce programming fast. Rather than filled with disjointed, piecemeal code snippets to show Hadoop MapReduce programming features one at a time, it is designed to place your total Hadoop MapReduce programming learning process in a common application context of mining customer spending patterns ensconced in large volumes of credit card transaction record data. Precise, end-to-end procedures are given to help you set up your Hadoop MapReduce development environment quickly on Eclipse with Maven on Mac OS X. Step-by-step procedures are also given on how to set up a four-node Linux cluster at minimum so that you can run your MapReduce programs not only in local mode on your Mac OS X machine but also in fully distributed mode on a real cluster. In fact, all MapReduce programs presented in the book have been tested and verified in local mode and on such a Linux cluster. This textbook mainly focuses on teaching Hadoop MapReduce programming in a scientific, objective, quantitative approach. Rather than heavily relying on subjective, verbose (and sometimes even pompous) textual descriptions with sparse code snippets, this textbook uses Hadoop Java APIs, Hadoop configuration parameters, complete MapReduce programs and their execution logs and outputs to demonstrate how Hadoop MapReduce framework works and how to write MapReduce programs. Specifically, this text covers the following subjects: *Introduction to Hadoop *Setting up a Linux Hadoop Cluster *The Hadoop Distributed FileSystem *MapReduce Job Orchestration and Workflows *Basic MapReduce Programming *Advanced MapReduce Programming *Hadoop Streaming *Hadoop Administration No matter what role you play on your team, this text can help you gain truly applicable Hadoop skills in a most effective and efficient manner. The book can also be used as a supplementary textbook for a distributed computing or Hadoop course offered to upper-division college CS students.

SAS Visual Analytics is a business intelligence and analytics platform that provides visual exploration and discovery, self-service analytics, and interactive reporting for organizations of all sizes. All organizations have a wide variety of users, and each user needs something different from data and analytics. SAS Visual Analytics allows everyone to easily discover and share powerful insights that inspire action. Several useful papers have been written to demonstrate how to use these techniques. We have carefully selected a handful of these from recent Global Forum contributions to introduce you to the topic and let you sample what each has to offer. Also available free as a PDF from sas.com/books.

Hadoop in Action teaches readers how to use Hadoop and write MapReduce programs. The intended readers are programmers, architects, and project managers who have to process large amounts of data offline. Hadoop in Action will lead the reader from obtaining a copy of Hadoop to setting it up in a cluster and writing data analytic programs. The book begins by making the basic idea of Hadoop and MapReduce easier to grasp by applying the default Hadoop installation to a few easy-to-follow tasks, such as analyzing changes in word frequency across a body of documents. The book continues through the basic concepts of MapReduce applications developed using Hadoop, including a close look at framework components, use of Hadoop for a variety of data analysis tasks, and numerous examples of Hadoop in action. Hadoop in Action will explain how to use Hadoop and present design patterns and practices of programming MapReduce. MapReduce is a complex idea both conceptually and in its implementation, and Hadoop users are challenged to learn all the knobs and levers for running Hadoop. This book takes you beyond the mechanics of running Hadoop, teaching you to write meaningful programs in a MapReduce framework. This book assumes the reader will have a basic familiarity with Java, as most code examples will be written in Java. Familiarity with basic statistical concepts (e.g. histogram,

correlation) will help the reader appreciate the more advanced data processing examples. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book.

This open access book covers the use of data science, including advanced machine learning, big data analytics, Semantic Web technologies, natural language processing, social media analysis, time series analysis, among others, for applications in economics and finance. In addition, it shows some successful applications of advanced data science solutions used to extract new knowledge from data in order to improve economic forecasting models. The book starts with an introduction on the use of data science technologies in economics and finance and is followed by thirteen chapters showing success stories of the application of specific data science methodologies, touching on particular topics related to novel big data sources and technologies for economic analysis (e.g. social media and news); big data models leveraging on supervised/unsupervised (deep) machine learning; natural language processing to build economic and financial indicators; and forecasting and nowcasting of economic variables through time series analysis. This book is relevant to all stakeholders involved in digital and data-intensive research in economics and finance, helping them to understand the main opportunities and challenges, become familiar with the latest methodological findings, and learn how to use and evaluate the performances of novel tools and frameworks. It primarily targets data scientists and business analysts exploiting data science technologies, and it will also be a useful resource to research students in disciplines and courses related to these topics. Overall, readers will learn modern and effective data science solutions to create tangible innovations for economic and financial applications.

Analytics in a Big Data World

Big Data, Analytics, and the Future of Marketing & Sales
bookdown

A Data Science Approach

Learning Tableau 2019

The Data Warehouse Toolkit

Hadoop in Action

Become an efficient data science practitioner by understanding Python's key concepts About This Book Quickly get familiar with data science using Python 3.5 Save time (and effort) with all the essential tools explained Create effective data science projects and avoid common pitfalls with the help of examples and hints dictated by experience Who This Book Is For If you are an aspiring data scientist and you have at least a working knowledge of data analysis and Python, this book will get you started in data science. Data analysts with experience of R or MATLAB will also find the book to be a comprehensive reference to enhance their data manipulation and machine learning skills. What You Will Learn Set up your data science toolbox using a Python scientific environment on Windows, Mac, and Linux Get data ready for your data science project Manipulate, fix, and explore data in order to solve data science problems Set up an experimental pipeline to test your data science hypotheses Choose the most effective and scalable learning algorithm for your data science tasks Optimize your machine learning models to get the best performance Explore and cluster graphs, taking advantage of interconnections and links in your data In Detail Fully expanded and upgraded, the second edition of Python Data Science Essentials takes you through all you need to know to succeed in data science using Python. Get modern insight into the core of Python data, including the latest versions of Jupyter notebooks, NumPy, pandas and scikit-learn. Look beyond the fundamentals with beautiful data visualizations with Seaborn and ggplot, web development with Bottle, and even the new frontiers of deep learning with Theano and TensorFlow. Dive into building your essential Python 3.5 data science toolbox, using a single-source approach that will allow to to work with Python 2.7 as well. Get to grips fast with data munging and preprocessing, and all the techniques you need to load, analyse, and process your data. Finally, get a complete overview of principal machine learning algorithms, graph analysis techniques, and all the visualization and deployment instruments that make it easier to present your results to an audience of both data science experts and business users. Style and approach The book is structured as a data science project. You will always benefit from clear code and simplified examples to help you understand the underlying mechanics and real-world datasets.

ESSENTIALS OF BUSINESS ANALYTICS, 2e can be used by students who have previously taken a course on basic statistical methods as well as students who have not had a prior course in statistics. The expanded material in the second edition of Essentials of Business Analytics also makes it amenable to a two-course sequence in business statistics and analytics. All statistical concepts contained in this textbook are presented from a business analytics perspective using practical business examples. Important Notice: Media content referenced within the product description or the product text may not be available in the ebook version.

Big Data: Principles and Paradigms captures the state-of-the-art research on the architectural aspects, technologies, and applications of Big Data. The book identifies potential future directions and technologies that facilitate insight into numerous scientific, business, and consumer applications. To help realize Big Data's full potential, the book addresses numerous challenges, offering the conceptual and technological solutions for tackling them. These challenges include life-cycle data management, large-scale storage, flexible processing infrastructure, data modeling, scalable machine learning, data analysis algorithms, sampling techniques, and privacy and ethical issues. Covers computational platforms supporting Big Data applications Addresses key principles underlying Big Data computing Examines key developments supporting next generation Big Data platforms Explores the challenges in Big Data computing and ways to overcome them Contains expert contributors from both academia and industry

bookdown: Authoring Books and Technical Documents with R Markdown presents a much easier way to write books and

technical publications than traditional tools such as LaTeX and Word. The bookdown package inherits the simplicity of syntax and flexibility for data analysis from R Markdown, and extends R Markdown for technical writing, so that you can make better use of document elements such as figures, tables, equations, theorems, citations, and references. Similar to LaTeX, you can number and cross-reference these elements with bookdown. Your document can even include live examples so readers can interact with them while reading the book. The book can be rendered to multiple output formats, including LaTeX/PDF, HTML, EPUB, and Word, thus making it easy to put your documents online. The style and theme of these output formats can be customized. We used books and R primarily for examples in this book, but bookdown is not only for books or R. Most features introduced in this book also apply to other types of publications: journal papers, reports, dissertations, course handouts, study notes, and even novels. You do not have to use R, either. Other choices of computing languages include Python, C, C++, SQL, Bash, Stan, JavaScript, and so on, although R is best supported. You can also leave out computing, for example, to write a fiction. This book itself is an example of publishing with bookdown and R Markdown, and its source is fully available on GitHub.

Authoring Books and Technical Documents with R Markdown

Hadoop: The Definitive Guide

Improving E-Commerce Web Applications Through Business Intelligence Techniques

The Complete Guide to Dimensional Modeling

An Introduction to the Methodology and its Applications

Visual Analytics with SAS Viya

Big Data Fundamentals

Data Science and Big Data Analytics is about harnessing the power of data for new insights. The book covers the breadth of activities and methods and tools that Data Scientists use. The content focuses on concepts, principles and practical applications that are applicable to any industry and technology environment, and the learning is supported and explained with examples that you can replicate using open-source software. This book will help you: Become a contributor on a data science team Deploy a structured lifecycle approach to data analytics problems Apply appropriate analytic techniques and tools to analyzing big data Learn how to tell a compelling story with data to drive business action Prepare for EMC Proven Professional Data Science Certification Corresponding data sets are available from the book's page at Wiley which you can find on the Wiley site by searching for the ISBN 9781118876138. Get started discovering, analyzing, visualizing, and presenting data in a meaningful way today!

This book adopts a unique approach to helping enterprise Java developers learn Spring 4 fast. Rather than filled with disjointed, piecemeal samples to show Spring features one at a time, it is designed to base your total Spring learning experience on a functioning, end-to-end integrated sample named SOBA (Secure Online Banking Application), which runs on any one of the three operating systems (Windows, Linux and Mac OS X), any one of the four Java App Servers (Tomcat, GlassFish, JBoss and WebLogic), and any one of the four RDBMS (MySQL, PostgreSQL, Oracle and SQL Server). The book also includes another standalone sample application named MyNotes, which is simpler than SOBA. Specifically, this book helps you learn the following latest Spring technologies: * Spring Core Framework * Spring MVC Web Framework * Spring Data Access Framework (JDBC and Hibernate) * Spring RESTful Web Services Framework * Spring Security Framework * Spring Transaction Management Framework * Spring Validation Framework * Spring Aspect Oriented Programming (AOP) Framework * Spring Testing * Spring Integration with EJB * Spring Web Flow Framework At the end of your learning experience with this book, you will gain truly applicable skills and will be able to start contributing to the success of your Spring-based enterprise application project immediately.

Data is arriving faster than you can process it and the overall volumes keep growing at a rate that keeps you awake at night. Hadoop can help you tame the data beast. Effective use of Hadoop however requires a mixture of programming, design, and system administration skills. "Hadoop Beginner's Guide" removes the mystery from Hadoop, presenting Hadoop and related technologies with a focus on building working systems and getting the job done, using cloud services to do so when it makes sense. From basic concepts and initial setup through developing applications and keeping the system running as the data grows, the book gives the understanding needed to effectively use Hadoop to solve real world problems. Starting with the basics of installing and configuring Hadoop, the book explains how to develop applications, maintain the system, and how to use additional products to integrate with other systems. While learning different ways to develop applications to run on Hadoop the book also covers tools such as Hive, Sqoop, and Flume that show how Hadoop can be integrated with relational databases and log collection. In addition to examples on Hadoop clusters on Ubuntu uses of cloud services such as Amazon, EC2 and Elastic MapReduce are covered.

Step-by-step guide to build high performing predictive applications Key FeaturesUse the Python data analytics ecosystem to implement end-to-end predictive analytics projectsExplore advanced predictive modeling algorithms with an emphasis on theory with intuitive explanationsLearn to deploy a predictive model's results as an interactive applicationBook Description Predictive analytics is an applied field that employs a variety of quantitative methods using data to make predictions. It involves much more than just throwing data onto a computer to build a model. This book provides practical coverage to help you understand the most important concepts of predictive analytics. Using practical, step-by-step examples, we build predictive analytics solutions while using cutting-edge Python tools and packages. The book's step-by-step approach starts by defining the problem and moves on to identifying relevant data. We will also be performing data preparation, exploring and visualizing relationships, building models, tuning, evaluating, and deploying model. Each stage has relevant practical examples and efficient Python code. You will work with models such as KNN, Random Forests, and neural networks using the most important libraries in Python's data science stack: NumPy, Pandas, Matplotlib, Seaborn, Keras, Dash, and so on. In addition to hands-on code examples, you will find intuitive explanations of the inner workings of the main techniques and algorithms used in predictive analytics. By the end of this book, you will be all set to build high-performance predictive analytics solutions using Python programming. What you

will learn Get to grips with the main concepts and principles of predictive analytics Learn about the stages involved in producing complete predictive analytics solutions Understand how to define a problem, propose a solution, and prepare a dataset Use visualizations to explore relationships and gain insights into the dataset Learn to build regression and classification models using scikit-learn Use Keras to build powerful neural network models that produce accurate predictions Learn to serve a model's predictions as a web application Who this book is for This book is for data analysts, data scientists, data engineers, and Python developers who want to learn about predictive modeling and would like to implement predictive analytics solutions using Python's data stack. People from other backgrounds who would like to enter this exciting field will greatly benefit from reading this book. All you need is to be proficient in Python programming and have a basic understanding of statistics and college-level algebra.

25th International Symposium, GD 2017, Boston, MA, USA, September 25-27, 2017, Revised Selected Papers

Hadoop in Practice

Hands-On Predictive Analytics with Python

Thinking with Examples for Effective Learning

Hands-On for Developers and Technical Professionals

Penetration Testing

Build strong foundation for entering the world of Machine Learning and data science with the help of this comprehensive guide About This Book Get started in the field of Machine Learning with the help of this solid, concept-rich, yet highly practical guide. Your one-stop solution for everything that matters in mastering the whats and whys of Machine Learning algorithms and their implementation. Get a solid foundation for your entry into Machine Learning by strengthening your roots (algorithms) with this comprehensive guide. Who This Book Is For This book is for IT professionals who want to enter the field of data science and are very new to Machine Learning. Familiarity with languages such as R and Python will be invaluable here. What You Will Learn Acquaint yourself with important elements of Machine Learning Understand the feature selection and feature engineering process Assess performance and error trade-offs for Linear Regression Build a data model and understand how it works by using different types of algorithm Learn to tune the parameters of Support Vector machines Implement clusters to a dataset Explore the concept of Natural Processing Language and Recommendation Systems Create a ML architecture from scratch. In Detail As the amount of data continues to grow at an almost incomprehensible rate, being able to understand and process data is becoming a key differentiator for competitive organizations. Machine learning applications are everywhere, from self-driving cars, spam detection, document search, and trading strategies, to speech recognition. This makes machine learning well-suited to the present-day era of Big Data and Data Science. The main challenge is how to transform data into actionable knowledge. In this book you will learn all the important Machine Learning algorithms that are commonly used in the field of data science. These algorithms can be used for supervised as well as unsupervised learning, reinforcement learning, and semi-supervised learning. A few famous algorithms that are covered in this book are Linear regression, Logistic Regression, SVM, Naive Bayes, K-Means, Random Forest, TensorFlow, and Feature engineering. In this book you will also learn how these algorithms work and their practical implementation to resolve your problems. This book will also introduce you to the Natural Processing Language and Recommendation systems, which help you run multiple algorithms simultaneously. On completion of the book you will have mastered selecting Machine Learning algorithms for clustering, classification, or regression based on for your problem. Style and approach An easy-to-follow, step-by-step guide that will help you get to grips with real -world applications of Algorithms for Machine Learning.

This comprehensive edited volume is the first of its kind, designed to serve as a textbook for long-duration business analytics programs. It can also be used as a guide to the field by practitioners. The book has contributions from experts in top universities and industry. The editors have taken extreme care to ensure continuity across the chapters. The material is organized into three parts: A) Tools, B) Models and C) Applications. In Part A, the tools used by business analysts are described in detail. In Part B, these tools are applied to construct models used to solve business problems. Part C contains detailed applications in various functional areas of business and several case studies. Supporting material can be found in the appendices that develop the pre-requisites for the main text. Every chapter has a business orientation. Typically, each chapter begins with the description of business problems that are transformed into data questions; and methodology is developed to solve these questions. Data analysis is conducted using widely used software, the output and results are clearly explained at each stage of development. These are finally transformed into a business solution. The companion website provides examples, data sets and sample code for each chapter.