

Big Data Made Easy A Working To The Complete Hadoop Toolset

Learn to use Apache Pig to develop lightweight big data applications easily and quickly. This book shows you many optimization techniques and covers every context where Pig is used in big data analytics. Beginning Apache Pig shows you how Pig is easy to learn and requires relatively little time to develop big data applications. The book is divided into four parts: the complete features of Apache Pig; integration with other tools; how to solve complex business problems; and optimization of tools. You'll discover topics such as MapReduce and why it cannot meet every business need; the features of Pig Latin such as data types for each load, store, joins, groups, and ordering; how Pig workflows can be created; submitting Pig jobs using Hue; and working with Oozie. You'll also see how to extend the framework by writing UDFs and custom load, store, and filter functions. Finally you'll cover different optimization techniques such as gathering statistics about a Pig script, joining strategies, parallelism, and the role of data formats in good performance. What You Will Learn

- Use all the features of Apache Pig
- Integrate Apache Pig with other tools
- Extend Apache Pig
- Optimize Pig Latin code
- Solve different use cases for Pig Latin

Who This Book Is For All levels of IT professionals: architects, big data enthusiasts, engineers, developers, and big data administrators. This book takes away the fear of working with, analyzing, and visualizing data. Understand the key concepts involved with data analytics while working with real-world business examples. You are introduced to two fantastic tools to cleanse and analyze data (KNIME) and visualize your insights (Microsoft Power BI), but the principles from this ...

The financial services technology industry is booming and promises to change the way we manage our money online, disrupting the current landscape of the industry. Understanding fintech's many facets is the key to navigating the complex nuances of this global industry. Fintech in a Flash is a comprehensive guide to the future of banking and insurance. It discusses an array of hot topics such as online payments, crowdfunding, challenger banks, online insurance, digital lending, big data, and digital commerce. The author provides easy to understand explanations of the 14 main areas of fintech and their future, and insight into the main fintech hubs in the world and the so-called unicorns, fintech firms that have made it past a \$1 billion valuation. He breaks down the key concepts of fintech in a way that will help you understand every aspect so that you can take advantage of new technologies. This detailed guide is your go-to source for everything you need to confidently navigate the ever-changing scene of this booming industry.

"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

Where To Download Big Data Made Easy A Working To The Complete Hadoop Toolset

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 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! The best-selling author of Big Data is back, this time with a unique and in-depth insight into how specific companies use big data. Big data is on the tip of everyone's tongue. Everyone understands its power and importance, but many fail to grasp the actionable steps and resources required to utilise it effectively. This book fills the knowledge gap by showing how major companies are using big data every day, from an up-close, on-the-ground perspective. From technology, media and retail, to sport teams, government agencies and financial institutions, learn the actual strategies and processes being used to learn about customers, improve manufacturing, spur innovation, improve safety and so much more. Organised for easy dip-in navigation, each chapter follows the same structure to give you the information you need quickly. For each company profiled, learn what data was used, what problem it solved and the processes put it place to make it practical, as well as the technical details, challenges and lessons learned from each unique scenario. Learn how predictive analytics helps Amazon, Target, John Deere and Apple understand their customers Discover how big data is behind the success of Walmart, LinkedIn, Microsoft and more Learn how big data is changing medicine, law enforcement, hospitality, fashion, science and banking Develop your own big data strategy by accessing additional reading materials at the end of each chapter

This book unravels the mystery of Big Data computing and its power to transform business operations. The approach it uses will be helpful to any professional who must present a case for realizing Big Data computing solutions or to those who could be involved in a Big Data computing project. It provides a framework that enables business and technical managers to make optimal decisions necessary for the successful migration to Big Data computing environments and applications within their organizations.

In Data Sketches, Nadieh Bremer and Shirley Wu document the deeply creative process behind 24 unique data visualization projects, and they combine this with powerful technical insights which reveal the mindset behind coding creatively. Exploring 12 different themes – from the Olympics to Presidents & Royals and from Movies to Myths & Legends – each pair of visualizations explores different technologies and forms, blurring the boundary between visualization as an exploratory tool and an artform in its own right. This beautiful book provides an intimate, behind-the-scenes account of all 24 projects and shares the authors’ personal

Where To Download Big Data Made Easy A Working To The Complete Hadoop Toolset

notes and drafts every step of the way. The book features: Detailed information on data gathering, sketching, and coding data visualizations for the web, with screenshots of works-in-progress and reproductions from the authors' notebooks Never-before-published technical write-ups, with beginner-friendly explanations of core data visualization concepts Practical lessons based on the data and design challenges overcome during each project Full-color pages, showcasing all 24 final data visualizations This book is perfect for anyone interested or working in data visualization and information design, and especially those who want to take their work to the next level and are inspired by unique and compelling data-driven storytelling. Big Data Made Easy A Working Guide to the Complete Hadoop Toolset Apress

Big Data Analytics with R and Hadoop is a tutorial style book that focuses on all the powerful big data tasks that can be achieved by integrating R and Hadoop. This book is ideal for R developers who are looking for a way to perform big data analytics with Hadoop. This book is also aimed at those who know Hadoop and want to build some intelligent applications over Big data with R packages. It would be helpful if readers have basic knowledge of R.

Is the Brexit vote successful big data politics or the end of democracy? Why do airlines overbook, and why do banks get it wrong so often? How does big data enable Netflix to forecast a hit, CERN to find the Higgs boson and medics to discover if red wine really is good for you? And how are companies using big data to benefit from smart meters, use advertising that spies on you and develop the gig economy, where workers are managed by the whim of an algorithm? The volumes of data we now access can give unparalleled abilities to make predictions, respond to customer demand and solve problems. But Big Brother's shadow hovers over it. Though big data can set us free and enhance our lives, it has the potential to create an underclass and a totalitarian state. With big data ever-present, you can't afford to ignore it. Acclaimed science writer Brian Clegg - a habitual early adopter of new technology (and the owner of the second-ever copy of Windows in the UK) - brings big data to life.

Are you interested to get into the programming world? Do you want to learn and understand Python and Machine Learning? Python Machine Learning for Beginners is the guide for you. Python Machine Learning for Beginners is the ultimate guide for beginners looking to learn and understand how Python programming works. Python Machine Learning for Beginners is split up into easy to learn chapters that will help guide the readers through the early stages of Python programming. It's this thought out and systematic approach to learning which makes Python Machine Learning for Beginners such a sought-after resource for those that want to learn about Python programming and about Machine Learning using an object-oriented programming approach. Inside Python Machine Learning for Beginners you will discover: An introduction to Machine Learning The main concepts of Machine Learning The basics of Python for beginners Machine Learning with Python Data Processing, Analysis, and Visualizations Case studies and much more! Throughout the book, you will learn the basic concepts behind Python programming which is designed to introduce you to Python programming. You will learn about getting started, the keywords and statements, data types and type conversion. Along with different examples, there are also exercises to help ensure that the information sinks in. You will find this book an invaluable tool for starting and mastering Machine Learning using Python. Once you complete Python Machine Learning for Beginners, you will be more than prepared to take on any Python programming. Scroll back up to the top of this page and hit BUY IT NOW to get your copy of Python Machine Learning for Beginners! You won't regret it!

Data Visualization Made Simple is a practical guide to the fundamentals, strategies, and real-world cases for data visualization, an essential skill required in today's information-rich world. With foundations rooted in statistics, psychology, and computer science, data visualization offers practitioners in almost every field a coherent way to share findings from original research, big data, learning analytics, and more. In nine appealing chapters, the book:

Where To Download Big Data Made Easy A Working To The Complete Hadoop Toolset

examines the role of data graphics in decision-making, sharing information, sparking discussions, and inspiring future research; scrutinizes data graphics, deliberates on the messages they convey, and looks at options for design visualization; and includes cases and interviews to provide a contemporary view of how data graphics are used by professionals across industries Both novices and seasoned designers in education, business, and other areas can use this book's effective, linear process to develop data visualization literacy and promote exploratory, inquiry-based approaches to visualization problems.

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.

Build next-generation Artificial Intelligence systems with Java Key Features Implement AI techniques to build smart applications using Deeplearning4j Perform big data analytics to derive quality insights using Spark MLlib Create self-learning systems using neural networks, NLP, and reinforcement learning Book Description In this age of big data, companies have larger amount of consumer data than ever before, far more than what the current technologies can ever hope to keep up with. However, Artificial Intelligence closes the gap by moving past human limitations in order to analyze data. With the help of Artificial Intelligence for big data, you will learn to use Machine Learning algorithms such as k-means, SVM, RBF, and regression to perform advanced data analysis. You will understand the current status of Machine and Deep Learning techniques to work on Genetic and Neuro-Fuzzy algorithms. In

Where To Download Big Data Made Easy A Working To The Complete Hadoop Toolset

addition, you will explore how to develop Artificial Intelligence algorithms to learn from data, why they are necessary, and how they can help solve real-world problems. By the end of this book, you'll have learned how to implement various Artificial Intelligence algorithms for your big data systems and integrate them into your product offerings such as reinforcement learning, natural language processing, image recognition, genetic algorithms, and fuzzy logic systems. What you will learn Manage Artificial Intelligence techniques for big data with Java Build smart systems to analyze data for enhanced customer experience Learn to use Artificial Intelligence frameworks for big data Understand complex problems with algorithms and Neuro-Fuzzy systems Design stratagems to leverage data using Machine Learning process Apply Deep Learning techniques to prepare data for modeling Construct models that learn from data using open source tools Analyze big data problems using scalable Machine Learning algorithms Who this book is for This book is for you if you are a data scientist, big data professional, or novice who has basic knowledge of big data and wish to get proficiency in Artificial Intelligence techniques for big data. Some competence in mathematics is an added advantage in the field of elementary linear algebra and calculus.

Knowledge in its pure state is tacit in nature—difficult to formalize and communicate—but can be converted into codified form and shared through both social interactions and the use of IT-based applications and systems. Even though there seems to be considerable synergies between the resulting huge data and the convertible knowledge, there is still a debate on how the increasing amount of data captured by corporations could improve decision making and foster innovation through effective knowledge-sharing practices. Big Data and Knowledge Sharing in Virtual Organizations provides innovative insights into the influence of big data analytics and artificial intelligence and the tools, methods, and techniques for knowledge-sharing processes in virtual organizations. The content within this publication examines cloud computing, machine learning, and knowledge sharing. It is designed for government officials and organizations, policymakers, academicians, researchers, technology developers, and students.

Big Data is a big topic, based on simple principles. Guided by leading expert in the field, David Stephenson, you will be amazed at how you can transform your company, and significantly improve KPIs across a broad range of business units and applications. Find out how an ecommerce company avoided two million product returns per year, how a newspaper saw triple-digit annual growth in digital subscriptions, how researchers in England learned to better detect pending cardiovascular problems, and how AI programs taught themselves to win games using techniques that even their human programmers didn't understand, all thanks to big data. Find out also how one company realized it could swap a million dollar hardware system with a twenty thousand dollar replacement. With simple and straightforward chapters that allow you to map examples onto your own business, Big Data Demystified will help you:

- Know which data is most useful to collect now and why it's important to start collecting that data as soon as possible.
- Understand big data and data science and how they can help you reach your business goals and gain competitive advantage.
- Use big data to understand where you are now and how you can improve in the future.
- Understand factors in choosing a big data system, including whether to go with cloud-based solutions.
- Construct your big data team in a way that supports an effective strategy and helps make your business more data-driven.

BIG DATA MAKES A BIG DIFFERENCE "Read this book! It is an essential guide to using data in a practical way that drives results." Ian McHenry, CEO Beyond Pricing "This is the book we've been missing: big data explained without the complexity." Marc Salomon, Professor in Decision Sciences and Dean at University of Amsterdam Business School "Big Data for the rest of us! I have never come across a book that is so full of practical advice, actionable examples and helpful explanations. Read this one book and start executing Big Data at your workplace tomorrow!" Tobias Wann CEO at @Leisure Group

Where To Download Big Data Made Easy A Working To The Complete Hadoop Toolset

Are You Actively Analyzing the Data Surrounding Your Business? Keep Reading to Learn Why You Should Be.. You may be the owner of a business, or someone who actively participates in the day to day operations of a business. We will go ahead and assume that your business is operating at a profit and you are happy with the direction it is going. As someone in this situation you might ask yourself, "Why do I need Data Analysis anyways?". I'll tell you why, one simple reason. You are leaving money on the table. Let's put it this way.. you are doing good, but wouldn't you rather be doing great? Wouldn't you rather have the ability to predict how the consumers in your target market are going to be behaving a year from now? Five years from now? This is where Data Analysis comes in. Many people realize the need to pay attention to data in their business, but have no clue where to start. With the help of this book you will be better able to understand the importance of the data surrounding your business and exactly what to do with it. A Preview of What You Will Learn The Importance of Data in Business Exactly How to Handle and Manage Big Data Real World Examples of Data Science Benefiting Businesses Ways Data Can Be Used to Mitigate Risks The Entire Process of Data Analytics Much, much more! Take charge of your business today and buy this book!

This handbook covers a wide range of topics related to the collection, processing, analysis, and use of geospatial data in their various forms. This handbook provides an overview of how spatial computing technologies for big data can be organized and implemented to solve real-world problems. Diverse subdomains ranging from indoor mapping and navigation over trajectory computing to earth observation from space, are also present in this handbook. It combines fundamental contributions focusing on spatio-textual analysis, uncertain databases, and spatial statistics with application examples such as road network detection or colocation detection using GPUs. In summary, this handbook gives an essential introduction and overview of the rich field of spatial information science and big geospatial data. It introduces three different perspectives, which together define the field of big geospatial data: a societal, governmental, and governance perspective. It discusses questions of how the acquisition, distribution and exploitation of big geospatial data must be organized both on the scale of companies and countries. A second perspective is a theory-oriented set of contributions on arbitrary spatial data with contributions introducing into the exciting field of spatial statistics or into uncertain databases. A third perspective is taking a very practical perspective to big geospatial data, ranging from chapters that describe how big geospatial data infrastructures can be implemented and how specific applications can be implemented on top of big geospatial data. This would include for example, research in historic map data, road network extraction, damage estimation from remote sensing imagery, or the analysis of spatio-textual collections and social media. This multi-disciplinary approach makes the book unique. This handbook can be used as a reference for undergraduate students, graduate students and researchers focused on big geospatial data. Professionals can use this book, as well as practitioners facing big collections of geospatial data.

What is Big Data, and why should you care? Big data knows where you've been and who your friends are. It knows what you like and what makes you angry. It can predict what you'll buy, where you'll be the victim of crime and when you'll have a heart attack. Big data knows you better than you know yourself, or so it claims. But how well do you know big data? You've probably seen the phrase in newspaper headlines, at work in a marketing meeting, or on a fitness-tracking gadget. But can you understand it without being a Silicon Valley nerd who writes computer programs for fun? Yes. Yes, you can. Timandra Harkness writes comedy, not computer code. The only programmes she makes are on the radio. If you can read a newspaper you can read this book. Starting with the basics – what IS data? And what makes it big? – Timandra takes you on a whirlwind tour of how people are using big data today: from science to smart cities, business to politics, self-quantification to the Internet of Things. Finally, she asks the big questions about where it's taking us; is it too big for its boots, or does it think

Where To Download Big Data Made Easy A Working To The Complete Hadoop Toolset

too small? Are you a data point or a human being? Will this book be full of rhetorical questions? No. It also contains puns, asides, unlikely stories and engaging people, inspiring feats and thought-provoking dilemmas. Leaving you armed and ready to decide what you think about one of the decade's big ideas: big data.

This book presents unique techniques to conquer different Big Data processing and analytics challenges using Hadoop. Practical examples are provided to boost your understanding of Big Data concepts and their implementation. By the end of the book, you will have all the knowledge and skills you need to become a true Big Data expert.

Less than 0.5 per cent of all data is currently analysed and used. However, business leaders and managers cannot afford to be unconcerned or sceptical about data. Data is revolutionizing the way we work and it is the companies that view data as a strategic asset that will survive and thrive. Bernard Marr's Data Strategy is a must-have guide to creating a robust data strategy. Explaining how to identify your strategic data needs, what methods to use to collect the data and, most importantly, how to translate your data into organizational insights for improved business decision-making and performance, this is essential reading for anyone aiming to leverage the value of their business data and gain competitive advantage. Packed with case studies and real-world examples, advice on how to build data competencies in an organization and crucial coverage of how to ensure your data doesn't become a liability, Data Strategy will equip any organization with the tools and strategies it needs to profit from big data, analytics and the Internet of Things.

Examine the techniques and Java tools supporting the growing field of data science About This Book Your entry ticket to the world of data science with the stability and power of Java Explore, analyse, and visualize your data effectively using easy-to-follow examples Make your Java applications more capable using machine learning Who This Book Is For This book is for Java developers who are comfortable developing applications in Java. Those who now want to enter the world of data science or wish to build intelligent applications will find this book ideal.

Aspiring data scientists will also find this book very helpful. What You Will Learn Understand the nature and key concepts used in the field of data science Grasp how data is collected, cleaned, and processed Become comfortable with key data analysis techniques See specialized analysis techniques centered on machine learning Master the effective visualization of your data Work with the Java APIs and techniques used to perform data analysis In Detail Data science is concerned with extracting knowledge and insights from a wide variety of data sources to analyse patterns or predict future behaviour. It draws from a wide array of disciplines including statistics, computer science, mathematics, machine learning, and data mining. In this book, we cover the important data science concepts and how they are supported by Java, as well as the often statistically challenging techniques, to provide you with an understanding of their purpose and application. The book starts with an introduction of data science, followed by the basic data science tasks of data collection, data cleaning, data analysis, and data visualization. This is followed by a discussion of statistical techniques and more advanced topics including machine learning, neural networks, and deep learning. The next section examines the major categories of data analysis including text, visual, and audio data, followed by a discussion of resources that support parallel implementation. The final chapter illustrates an in-depth data science problem and provides a comprehensive, Java-based solution. Due to the nature of the topic, simple examples of techniques are presented early followed by a more detailed treatment later in the book. This permits a more natural introduction to the techniques and concepts presented in the book. Style and approach This book follows a tutorial approach, providing examples of each of the major concepts covered. With a step-by-step instructional style, this book covers various facets of data science and will get you up and running quickly.

“This book is a splendid and valuable addition to this subject. The whole book is well written

Where To Download Big Data Made Easy A Working To The Complete Hadoop Toolset

and I have no hesitation to recommend that this can be adapted as a textbook for graduate courses in Business Intelligence and Data Mining.” Dr. Edi Shivaji, Des Moines, Iowa “As a complete novice to this area just starting out on a MBA course I found the book incredibly useful and very easy to follow and understand. The concepts are clearly explained and make it an easy task to gain an understanding of the subject matter.” -- Mr. Craig Domoney, South Africa. Business Intelligence and Data Mining is a conversational and informative book in the exploding area of Business Analytics. Using this book, one can easily gain the intuition about the area, along with a solid toolset of major data mining techniques and platforms. This book can thus be gainfully used as a textbook for a college course. It is also short and accessible enough for a busy executive to become a quasi-expert in this area in a couple of hours. Every chapter begins with a case-let from the real world, and ends with a case study that runs across the chapters.

Make informed decisions using data analytics, machine learning, and data visualizations
Key Features: Take raw data and transform it to add value to your organization
Learn the art of telling stories with your data to engage with your audience
Apply machine learning algorithms to your data with a few clicks of a button
Book Description: Data analytics has become a necessity in modern business, and skills such as data visualization, machine learning, and digital storytelling are now essential in every field. If you want to make sense of your data and add value with informed decisions, this is the book for you. Data Analytics Made Easy is an accessible guide to help you start analyzing data and quickly apply these skills to your work. It focuses on how to generate insights from your data at the click of a few buttons, using the popular tools KNIME and Microsoft Power BI. The book introduces the concepts of data analytics and shows you how to get your data ready and apply ML algorithms. Implement a full predictive analytics solution with KNIME and assess its level of accuracy. Create impressive visualizations with Microsoft Power BI and learn the greatest secret in successful analytics - how to tell a story with your data. You'll connect the dots on the various stages of the data-to-insights process and gain an overview of alternative tools, including Tableau and H2O Driverless AI. By the end of this book, you will have learned how to implement machine learning algorithms and sell the results to your customers without writing a line of code.
What You Will Learn: Understand the potential of data and its impact on any business
Influence business decisions with effective data storytelling when delivering insights
Use KNIME to import, clean, transform, combine data feeds, and automate recurring workflows
Learn the basics of machine learning and AutoML to add value to your organization
Build, test, and validate simple supervised and unsupervised machine learning models with KNIME
Use Power BI and Tableau to build professional-looking and business-centric visuals and dashboards
Who this book is for: Whether you are working with data experts or want to find insights in your business' data, you'll find this book an effective way to add analytics to your skill stack. No previous math, statistics, or computer science knowledge is required. Big Data Analytics Made Easy is a must-read for everybody as it explains the power of Analytics in a simple and logical way along with an end to end code in

Where To Download Big Data Made Easy A Working To The Complete Hadoop Toolset

R. Even if you are a novice in Big Data Analytics, you will still be able to understand the concepts explained in this book. If you are already working in Analytics and dealing with Big Data, you will still find this book useful, as it covers exhaustive Data Mining Techniques, which are considered to be Advanced topics. It covers Machine Learning concepts and provides in-depth knowledge on unsupervised as well as supervised Learning, which is very important for decision-making. The toughest Data Analytics concepts are made simpler, It features examples from all the domains so that the reader gets connected to the book easily. This book is like a personal trainer that will help you master the Art of Data Science.

Find the right big data solution for your business or organization Big data management is one of the major challenges facing business, industry, and not-for-profit organizations. Data sets such as customer transactions for a mega-retailer, weather patterns monitored by meteorologists, or social network activity can quickly outpace the capacity of traditional data management tools. If you need to develop or manage big data solutions, you'll appreciate how these four experts define, explain, and guide you through this new and often confusing concept. You'll learn what it is, why it matters, and how to choose and implement solutions that work. Effectively managing big data is an issue of growing importance to businesses, not-for-profit organizations, government, and IT professionals. Authors are experts in information management, big data, and a variety of solutions. Explains big data in detail and discusses how to select and implement a solution, security concerns to consider, data storage and presentation issues, analytics, and much more. Provides essential information in a no-nonsense, easy-to-understand style that is empowering. Big Data For Dummies cuts through the confusion and helps you take charge of big data solutions for your organization. Many corporations are finding that the size of their data sets are outgrowing the capability of their systems to store and process them. The data is becoming too big to manage and use with traditional tools. The solution: implementing a big data system. As Big Data Made Easy: A Working Guide to the Complete Hadoop Toolset shows, Apache Hadoop offers a scalable, fault-tolerant system for storing and processing data in parallel. It has a very rich toolset that allows for storage (Hadoop), configuration (YARN and ZooKeeper), collection (Nutch and Solr), processing (Storm, Pig, and Map Reduce), scheduling (Oozie), moving (Sqoop and Avro), monitoring (Chukwa, Ambari, and Hue), testing (Big Top), and analysis (Hive). The problem is that the Internet offers IT pros wading into big data many versions of the truth and some outright falsehoods born of ignorance. What is needed is a book just like this one: a wide-ranging but easily understood set of instructions to explain where to get Hadoop tools, what they can do, how to install them, how to configure them, how to integrate them, and how to use them successfully. And you need an expert who has worked in this area for a decade—someone just like author and big data expert Mike Frampton. Big Data Made Easy approaches the problem of managing massive data sets from a

Where To Download Big Data Made Easy A Working To The Complete Hadoop Toolset

systems perspective, and it explains the roles for each project (like architect and tester, for example) and shows how the Hadoop toolset can be used at each system stage. It explains, in an easily understood manner and through numerous examples, how to use each tool. The book also explains the sliding scale of tools available depending upon data size and when and how to use them. Big Data Made Easy shows developers and architects, as well as testers and project managers, how to:

- Store big data
- Configure big data
- Process big data
- Schedule processes
- Move data among SQL and NoSQL systems
- Monitor data
- Perform big data analytics
- Report on big data processes and projects
- Test big data systems

Big Data Made Easy also explains the best part, which is that this toolset is free. Anyone can download it and—with the help of this book—start to use it within a day. With the skills this book will teach you under your belt, you will add value to your company or client immediately, not to mention your career.

Go ahead, be skeptical about big data. The author was—at first. When the term “big data” first came on the scene, bestselling author Tom Davenport (Competing on Analytics, Analytics at Work) thought it was just another example of technology hype. But his research in the years that followed changed his mind. Now, in clear, conversational language, Davenport explains what big data means—and why everyone in business needs to know about it. Big Data at Work covers all the bases: what big data means from a technical, consumer, and management perspective; what its opportunities and costs are; where it can have real business impact; and which aspects of this hot topic have been oversold. This book will help you understand:

- Why big data is important to you and your organization
- What technology you need to manage it
- How big data could change your job, your company, and your industry
- How to hire, rent, or develop the kinds of people who make big data work
- The key success factors in implementing any big data project
- How big data is leading to a new approach to managing analytics

With dozens of company examples, including UPS, GE, Amazon, United Healthcare, Citigroup, and many others, this book will help you seize all opportunities—from improving decisions, products, and services to strengthening customer relationships. It will show you how to put big data to work in your own organization so that you too can harness the power of this ever-evolving new resource.

Hone your analytic talents and become part of the next big thing Getting a Big Data Job For Dummies is the ultimate guide to landing a position in one of the fastest-growing fields in the modern economy. Learn exactly what “big data” means, why it's so important across all industries, and how you can obtain one of the most sought-after skill sets of the decade. This book walks you through the process of identifying your ideal big data job, shaping the perfect resume, and nailing the interview, all in one easy-to-read guide. Companies from all industries, including finance, technology, medicine, and defense, are harnessing massive amounts of data to reap a competitive advantage. The demand for big data professionals is growing every year, and experts forecast an estimated 1.9 million

Where To Download Big Data Made Easy A Working To The Complete Hadoop Toolset

additional U.S. jobs in big data by 2015. Whether your niche is developing the technology, handling the data, or analyzing the results, turning your attention to a career in big data can lead to a more secure, more lucrative career path. Getting a Big Data Job For Dummies provides an overview of the big data career arc, and then shows you how to get your foot in the door with topics like: The education you need to succeed The range of big data career path options An overview of major big data employers A plan to develop your job-landing strategy Your analytic inclinations may be your ticket to long-lasting success. In a highly competitive job market, developing your data skills can create a situation where you pick your employer rather than the other way around. If you're ready to get in on the ground floor of the next big thing, Getting a Big Data Job For Dummies will teach you everything you need to know to get started today.

Explores the idea of big data, which refers to our new found ability to crunch vast amounts of information, analyze it instantly, and draw profound and surprising conclusions from it.

This book is designed to teach businesspeople, students, and others core statistical concepts and applications. It begins with absolute core principles and takes you through an overview of statistics, data and data collection, an introduction to SAS, and basic statistics (descriptive statistics and basic associational statistics). It provides an overview of statistical modeling, effect size, statistical significance and power testing, basics of linear regression, introduction to comparison of means, basics of chi-square tests for categories, extrapolating statistics to business outcomes, and some topical issues in statistics, such as big data, simulation, machine learning, and data warehousing. It teaches the core ideas of statistics through methods such as careful, intuitive written explanations, easy-to-follow diagrams, step-by-step technique implementation, and interesting metaphors. --

Convert the promise of big data into real world results There is so much buzz around big data. We all need to know what it is and how it works - that much is obvious. But is a basic understanding of the theory enough to hold your own in strategy meetings? Probably. But what will set you apart from the rest is actually knowing how to USE big data to get solid, real-world business results - and putting that in place to improve performance. Big Data will give you a clear understanding, blueprint, and step-by-step approach to building your own big data strategy. This is a well-needed practical introduction to actually putting the topic into practice. Illustrated with numerous real-world examples from a cross section of companies and organisations, Big Data will take you through the five steps of the SMART model: Start with Strategy, Measure Metrics and Data, Apply Analytics, Report Results, Transform. Discusses how companies need to clearly define what it is they need to know Outlines how companies can collect relevant data and measure the metrics that will help them answer their most important business questions Addresses how the results of big data analytics can be visualised and communicated to ensure key decisions-makers understand them Includes many high-profile case studies from the author's work with some of the world's best known brands Longlisted for the National Book Award New York Times Bestseller A former Wall Street quant sounds an alarm on the mathematical models that pervade modern life -- and threaten to rip apart our social fabric We live in the age of the algorithm. Increasingly, the decisions that affect

Where To Download Big Data Made Easy A Working To The Complete Hadoop Toolset

our lives--where we go to school, whether we get a car loan, how much we pay for health insurance--are being made not by humans, but by mathematical models. In theory, this should lead to greater fairness: Everyone is judged according to the same rules, and bias is eliminated. But as Cathy O'Neil reveals in this urgent and necessary book, the opposite is true. The models being used today are opaque, unregulated, and uncontested, even when they're wrong. Most troubling, they reinforce discrimination: If a poor student can't get a loan because a lending model deems him too risky (by virtue of his zip code), he's then cut off from the kind of education that could pull him out of poverty, and a vicious spiral ensues. Models are propping up the lucky and punishing the downtrodden, creating a "toxic cocktail for democracy." Welcome to the dark side of Big Data. Tracing the arc of a person's life, O'Neil exposes the black box models that shape our future, both as individuals and as a society. These "weapons of math destruction" score teachers and students, sort resumes, grant (or deny) loans, evaluate workers, target voters, set parole, and monitor our health. O'Neil calls on modelers to take more responsibility for their algorithms and on policy makers to regulate their use. But in the end, it's up to us to become more savvy about the models that govern our lives. This important book empowers us to ask the tough questions, uncover the truth, and demand change. -- Longlist for National Book Award (Non-Fiction) -- Goodreads, semi-finalist for the 2016 Goodreads Choice Awards (Science and Technology) -- Kirkus, Best Books of 2016 -- New York Times, 100 Notable Books of 2016 (Non-Fiction) -- The Guardian, Best Books of 2016 -- WBUR's "On Point," Best Books of 2016: Staff Picks -- Boston Globe, Best Books of 2016, Non-Fiction

"This book introduces you to R, RStudio, and the tidyverse, a collection of R packages designed to work together to make data science fast, fluent, and fun. Suitable for readers with no previous programming experience"--

Big Data For Beginners! The Ultimate Beginners Crash Course To Understanding And Interpreting Big Data! Are You Ready To Learn How To Understand SMART Big Data, Data Mining & Data Analytics For improved Business Performance, Life Decisions & More? If So You've Come To The Right Place - Regardless Of How Little Experience You May Have! Here's A Preview Of What Big Data For Beginners! Contains... A Conundrum Called 'Big Data' How To Understand Big Data Better What Can Big Data Do For You? Understanding The Analytics (And The Importance) The Obstacles And Importance Of The Big Data Situation We're In A Closer Look At Key Big Data Challenges Generating Business Value through Data Mining And Much, Much More! Order Your Copy Now And Let's Get Started!

Big Data Systems encompass massive challenges related to data diversity, storage mechanisms, and requirements of massive computational power. Further, capabilities of big data systems also vary with respect to type of problems. For instance, distributed memory systems are not recommended for iterative algorithms. Similarly, variations in big data systems also exist related to consistency and fault tolerance. The purpose of this book is to provide a detailed explanation of big data systems. The book covers various topics including Networking, Security, Privacy, Storage, Computation, Cloud Computing, NoSQL and NewSQL systems, High Performance Computing, and Deep Learning. An illustrative and practical approach has been adopted in which theoretical topics have been aided by well-explained programming and illustrative examples. Key Features: Introduces concepts and evolution of Big Data technology. Illustrates examples for thorough understanding. Contains programming examples for hands on development. Explains a variety of topics including NoSQL Systems, NewSQL systems, Security, Privacy, Networking, Cloud, High Performance Computing, and Deep Learning. Exemplifies widely used big data technologies such as Hadoop and Spark. Includes discussion on case studies and open issues. Provides end of chapter questions for enhanced learning. "Big Data, Big Dupe" is a little book about a big bunch of nonsense. The story of David and Goliath inspires us to hope that something little, when armed with truth, can topple something

Where To Download Big Data Made Easy A Working To The Complete Hadoop Toolset

big that is a lie. This is the author's hope. While others have written about the dangers of Big Data, Stephen Few reveals the deceit that belies its illusory nature. If "data is the new oil," Big Data is the new snake oil. It isn't real. It's a marketing campaign that has distracted us for years from the real and important work of deriving value from data.

Data collection, processing, analysis, and more About This Book Your entry ticket to the world of data science with the stability and power of Java Explore, analyse, and visualize your data effectively using easy-to-follow examples A highly practical course covering a broad set of topics - from the basics of Machine Learning to Deep Learning and Big Data frameworks. Who This Book Is For This course is meant for Java developers who are comfortable developing applications in Java, and now want to enter the world of data science or wish to build intelligent applications. Aspiring data scientists with some understanding of the Java programming language will also find this book to be very helpful. If you are willing to build efficient data science applications and bring them in the enterprise environment without changing your existing Java stack, this book is for you! What You Will Learn Understand the key concepts of data science Explore the data science ecosystem available in Java Work with the Java APIs and techniques used to perform efficient data analysis Find out how to approach different machine learning problems with Java Process unstructured information such as natural language text or images, and create your own search Learn how to build deep neural networks with DeepLearning4j Build data science applications that scale and process large amounts of data Deploy data science models to production and evaluate their performance In Detail Data science is concerned with extracting knowledge and insights from a wide variety of data sources to analyse patterns or predict future behaviour. It draws from a wide array of disciplines including statistics, computer science, mathematics, machine learning, and data mining. In this course, we cover the basic as well as advanced data science concepts and how they are implemented using the popular Java tools and libraries. The course starts with an introduction of data science, followed by the basic data science tasks of data collection, data cleaning, data analysis, and data visualization. This is followed by a discussion of statistical techniques and more advanced topics including machine learning, neural networks, and deep learning. You will examine the major categories of data analysis including text, visual, and audio data, followed by a discussion of resources that support parallel implementation. Throughout this course, the chapters will illustrate a challenging data science problem, and then go on to present a comprehensive, Java-based solution to tackle that problem. You will cover a wide range of topics – from classification and regression, to dimensionality reduction and clustering, deep learning and working with Big Data. Finally, you will see the different ways to deploy the model and evaluate it in production settings. By the end of this course, you will be up and running with various facets of data science using Java, in no time at all. This course contains premium content from two of our recently published popular titles: Java for Data Science Mastering Java for Data Science Style and approach This course follows a tutorial approach, providing examples of each of the concepts covered. With a step-by-step instructional style, this book covers various facets of data science and will get you up and running quickly.

A perfect guide to speed up the predicting power of machine learning algorithms Key Features Design, discover, and create dynamic, efficient features for your machine learning application Understand your data in-depth and derive astonishing data insights with the help of this Guide Grasp powerful feature-engineering techniques and build machine learning systems Book Description Feature engineering is the most important step in creating powerful machine learning systems. This book will take you through the entire feature-engineering journey to make your machine learning much more systematic and effective. You will start with understanding your data—often the success of your ML models depends on how you leverage different feature types, such as continuous, categorical, and more, You will learn when to

Where To Download Big Data Made Easy A Working To The Complete Hadoop Toolset

include a feature, when to omit it, and why, all by understanding error analysis and the acceptability of your models. You will learn to convert a problem statement into useful new features. You will learn to deliver features driven by business needs as well as mathematical insights. You'll also learn how to use machine learning on your machines, automatically learning amazing features for your data. By the end of the book, you will become proficient in Feature Selection, Feature Learning, and Feature Optimization. What you will learn Identify and leverage different feature types Clean features in data to improve predictive power Understand why and how to perform feature selection, and model error analysis Leverage domain knowledge to construct new features Deliver features based on mathematical insights Use machine-learning algorithms to construct features Master feature engineering and optimization Harness feature engineering for real world applications through a structured case study Who this book is for If you are a data science professional or a machine learning engineer looking to strengthen your predictive analytics model, then this book is a perfect guide for you. Some basic understanding of the machine learning concepts and Python scripting would be enough to get started with this book.

[Copyright: 44dfb02b8984f59b3f18e40ee7d4b800](#)