

Big Data Analytics Beyond Hadoop Real Time Applic

This handbook offers comprehensive coverage of recent advancements in Big Data technologies and related paradigms. Chapters are authored by international leading experts in the field, and have been reviewed and revised for maximum reader value. The volume consists of twenty-five chapters organized into four main parts. Part one covers the fundamental concepts of Big Data technologies in mechanisms, data models, storage models, programming models and programming platforms. It also dives into the details of implementing Big SQL query engines and big stream processing systems. Part Two focuses on the semantic aspects of Big Data management including data integration and exploratory ad hoc analysis in addition to structured querying and pattern matching techniques. Part three provides a comprehensive overview of large scale graph processing. It covers the most recent research in large scale graph processing platforms, introducing several scalable graph querying and mining mechanisms in domains such as social networks. Part Four details novel applications that have been made possible by the rapid emergence of Big Data technologies such as Internet-of-Things (IoT), Cognitive Systems. All parts of the book discuss open research problems, including potential opportunities, that have arisen from the rapid progress of Big Data technologies and the associated increasing requirements of application domains. Designed for researchers, IT professionals and graduate students, this book is a timely contribution to the growing Big Data field. Big Data has been recognized as one of the top 10 technologies that will have a major contribution and impact on the various fields of science and varies aspect of the human society over the coming decades. Therefore, the content in this book will be an essential tool to help readers understand the development and future of the field.

Unique perspective on the big data analytics phenomenon for both business and IT professionals The availability of Big Data, low-cost commodity hardware and new information management and analytics software has produced a unique moment in the history of business. The convergence of these trends means that we have the capabilities required to analyze astonishing data sets quickly and accurately. This book provides a history. These capabilities are neither theoretical nor trivial. They represent a genuine leap forward and a clear opportunity to realize enormous gains in terms of efficiency, productivity, revenue and profitability. The Age of Big Data is here, and these are truly revolutionary times. This timely book looks at cutting-edge companies supporting an exciting new generation of business analytics. Learn how to use Big Data and how they are impacting the business world (Risk, Marketing, Healthcare, Financial Services, etc.) Explains this new technology and how companies can use them effectively to gather the data that they need and glean critical insights Explores relevant topics such as data privacy, data visualization, unstructured data, crowd sourcing data scientists, cloud computing for big data, and much more Building Big Data Applications helps data managers and their organizations make the most of unstructured data with an existing data warehouse. It provides readers with what they need to know to make sense of how Big Data fits into the world of Data Warehousing. Readers will learn about infrastructure options and integration and come away with a solid understanding on how to leverage what Big Data can do. This book includes a wide range of use cases that will help data managers visualize reference architectures in the context of specific industries (healthcare, big oil, transportation, software, etc.). Explores various ways to leverage Big Data by effectively integrating it into the data warehouse Includes real-world case studies which clearly demonstrate Big Data technologies Provides insights on how to build a Big Data infrastructure and integrate newer infrastructure matching data processing workloads and requirements

This book constitutes the refereed proceedings of the First International Conference on Big Data Analytics, BDA 2012, held in New Delhi, India, in December 2012. The 5 regular papers and 5 short papers presented were carefully reviewed and selected from 42 submissions. The volume also contains two tutorial papers in the section perspectives on big data analytics. The regular contributions are: data analytics applications; knowledge discovery through information extraction; and data models in analytics.

Data Analytics with Hadoop

Applied Big Data Analytics in Operations Management

Data Intensive Computing Applications for Big Data

A Roadmap for Usage and Exploitation of Big Data in Europe

Real-Time Applications with Storm, Spark, and More Hadoop Alternatives

Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem

How Cloudiness Keeps Changing Our Life, Economy and Technology

This book provides a collection of concepts, algorithms, and techniques that effectively harness the power of Spatial Network Big Data. Reading this book is a first step towards understanding the immense challenges and novel applications of SNBD database systems. This book explores these challenges via investigating scalable graph-based query processing strategies and I/O efficient storage architectures. This book is for researchers, engineers with a particular interest in network database models, network query processing, and physical storage models.

This open access book is part of the LAMBDA Project (Learning, Applying, Multiplying Big Data Analytics), funded by the European Union, GA No. 809965. Data Analytics involves applying algorithmic processes to derive insights. Nowadays it is used in many industries to allow organizations and companies to make better decisions as well as to verify or disprove existing theories or models. The term Data Analytics involves intelligence, statistics, reasoning, data mining, knowledge discovery, and others. The goal of this book is to introduce some of the definitions, methods, tools, frameworks, and solutions for big data processing, starting from the process of information extraction and knowledge representation, via knowledge processing and analytics to visualization, sense-making, and practical applications. Each chapter provides a processing chain, with a specific focus on understanding Enterprise Knowledge Graphs, Semantic Big Data Architectures, and Smart Data Analytics solutions. This book is addressed to graduate students from technical disciplines, to professional audiences following continuous education short courses, and to researchers from diverse areas following self-study courses. Basic skills in computer science, data science, and statistics are assumed. The book is intended for practitioners and researchers alike. "The chapters in this volume offer useful case studies, technical roadmaps, lessons learned, and a few prescriptions to 'do this, avoid that.'" —From the Foreword by Joe LaCugna, Ph.D., Enterprise Analytics and Business Intelligence, Starbucks Coffee Company With the growing barrage of "big data," it becomes vitally important for organizations to make sense of this data and information in a timely manner.

Research shows that organizations that use business analytics to guide their decision making are more productive and experience higher returns on equity. Big Data and Business Analytics helps you quickly grasp the trends and techniques of big data and business analytics to make your organization more competitive. Packed with case studies, this book assembles insights from some of the leading minds in the field of government, not-for-profit organizations, and academia, they share valuable perspectives on big data domains such as cybersecurity, marketing, emergency management, healthcare, finance, and transportation. Understand the trends, potential, and challenges associated with big data and business analytics Get an overview of machine learning, advanced statistical techniques, and other predictive analytics techniques. This book provides a comprehensive overview of Big Data/Insights & Analytics via case studies of Fortune 100 companies, government agencies, universities, and not-for-profits Big data problems are complex. This book shows you how to go from being data-rich to insight-rich, improving your decision making and creating competitive advantage. Author Jay Liebowitz recently had an article published in The World Financial Review. www.worldfinancialreview.com Use the latest data mining best practices to enable timely, actionable, evidence-based decision making throughout your organization! Real-World Data Mining demystifies current best practices, showing how to use data mining to uncover hidden patterns and correlations, and leverage these to improve all aspects of business performance. Drawing on extensive experience as a researcher, practitioner, and educator, the author provides a comprehensive overview of concepts, techniques and applications. Without compromising either simplicity or clarity, he provides enough technical depth to help readers truly understand how data mining technologies work. Coverage includes: processes, methods, techniques, tools, and metrics; the role and management of data; text and web mining; sentiment analysis; and Big Data integration. Throughout, Delen's concepts and examples (of both successes and failures), as well as simple, hands-on tutorials. Real-World Data Mining will be valuable to professionals on analytics teams; professionals seeking certification in the field; and undergraduate or graduate students in any analytics program: concentrations, certificate-based, or degree-based.

Queries and Storage Methods

Moving beyond MapReduce and Batch Processing with Apache Hadoop 2

Internet of Things and Big Data Analytics Toward Next-Generation Intelligence

Data Governance Principles for Big Data Analytics

Concepts, Paradigms and Solutions

Advances on Broadband and Wireless Computing, Communication and Applications

Hands-on techniques to implement enterprise analytics and machine learning using Hadoop, Spark, NoSQL and R

PEEK "UNDER THE HOOD" OF BIG DATA ANALYTICS The world of big data analytics grows ever more complex. And while many people can work superficially with specific frameworks, far fewer understand the fundamental principles of large-scale, distributed data processing systems and how they operate. In Foundations of Data Intensive Applications: Large Scale Data Analytics under the Hood, renowned big-data experts and computer scientists Drs. Supun Kamburugamuve and Saliya Ekanayake deliver a practical guide to applying the principles of big data to software development for optimal performance. The authors discuss foundational components of large-scale data systems and walk readers through the major software design decisions that define performance, application type, and usability. You???ll learn how to recognize and solve performance problems in your applications resulting in performance and distributed operation issues, diagnose them, and effectively eliminate them by relying on the bedrock big data principles explained within. Moving beyond individual frameworks and APIs for data processing, this book unlocks the theoretical ideas that operate under the hood of every big data processing system. Ideal for data scientists, data architects, dev-ops engineers, and developers, Foundations of Data Intensive Applications: Large Scale Data Analytics under the Hood shows readers how to: Identify the foundations of large-scale, distributed data processing systems Make major software design decisions that optimize performance Diagnose performance problems and distributed operation issues Understand state-of-the-art research in big data Explain and use the major big data frameworks and understand what underpins them Use big data analytics in the real world to solve practical problems

Get Started Fast with Apache Hadoop® 2, YARN, and Today's Hadoop Ecosystem With Hadoop 2.x and YARN, Hadoop moves beyond MapReduce to become practical for virtually any type of data processing. Hadoop 2.x and the Data Lake concept represent a radical shift away from conventional approaches to data usage and storage. Hadoop 2.x installations offer unmatched scalability and breakthrough extensibility that supports new and existing Big Data analytics processing methods and models. Hadoop® 2 Quick-Start Guide is the first easy, accessible guide to Apache Hadoop 2.x, YARN, and the modern Hadoop ecosystem. Building on his unsurpassed experience teaching Hadoop and Big Data, author Douglas Eadline covers all the basics you need to know to install and use Hadoop 2 on personal computers or servers, and to navigate the powerful technologies that complement it. Eadline concisely introduces and explains every key Hadoop 2 concept, tool, and service, illustrating each with a simple "beginning-to-end" example and identifying trustworthy, up-to-date resources for learning more. This guide is ideal if you want to learn about Hadoop 2 without getting mired in technical details. Douglas Eadline will bring you up to speed quickly, whether you're a user, admin, devops specialist, programmer, architect, analyst, or data scientist. Coverage Includes Understanding what Hadoop 2 and YARN do, and how they improve on Hadoop 1 with MapReduce Understanding Hadoop-based Data Lakes versus RDBMS Data Warehouses Installing Hadoop 2 and core services on Linux machines, virtualized sandboxes, or clusters Exploring the Hadoop Distributed File System (HDFS) Understanding the essentials of MapReduce and YARN application programming Simplifying programming and data movement with Apache Pig, Hive, Sqoop, Flume, Oozie, and HBase Observing application progress, controlling jobs, and managing workflows Managing Hadoop efficiently with Apache Ambari—including recipes for HDFS to NFSv3 gateway, HDFS snapshots, and YARN configuration Learning basic Hadoop 2 troubleshooting, and installing Apache Hue and Apache Spark Takes an in-depth look at the financial value of big data analytics and offers tools and best practices for working with big data. Once the domain of large on-line retailers such as eBay and Amazon, big data is now accessible by businesses of all sizes and across industries.

This book highlights state-of-the-art research on big data and the Internet of Things (IoT), along with related areas to ensure efficient and Internet-compatible IoT systems. It not only discusses big data security and privacy challenges, but also energy-efficient approaches to improving virtual machine placement in cloud computing environments. Big data and the Internet of Things (IoT) are ultimately two sides of the same coin, yet extracting, analyzing and managing IoT data poses a serious challenge. Accordingly, proper analytics infrastructures/platforms should be used to analyze IoT data. Information technology (IT) allows people to upload, retrieve, store and collect information, which ultimately forms big data. The use of big data analytics has grown tremendously in just the past few years. At the same time, the IoT has entered the public consciousness, sparking people's imaginations as to what a fully connected world can offer. Further, the book discusses the analysis of real-time big data to derive actionable intelligence in enterprise applications in several domains, such as in industry and agriculture. It explores possible automated solutions in daily life, including structures for smart cities and automated home systems based on IoT technology, as well as health care systems that manage large amounts of data (big data) to improve clinical decisions. The book addresses the security and privacy of the IoT and big data technologies, while also revealing the impact of IoT technologies on several scenarios in smart cities design. Intended as a comprehensive introduction, it offers in-depth analysis and provides scientists, engineers and professionals the latest techniques, frameworks and strategies used in IoT and big data technologies.

Big Data, Big Analytics

Streamlined Enterprise Data Management and Analysis

Applied Business Analytics and Decision Making

Big Data Analytics with Hadoop 3

Research Anthology on Big Data Analytics, Architectures, and Applications

Hadoop in Practice

Building Big Data Applications

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

Cloud computing is becoming the next revolution in the IT industry; providing central storage for internet data and services that have the potential to bring data transmission performance, security and privacy, data deluge, and inefficient architecture to the next level. Enabling the New Era of Cloud Computing: Data Security, Transfer, and Management discusses cloud computing as an emerging technology and its critical role in the IT industry upgrade and economic development in the future. This book is an essential resource for business decision makers, technology investors, architects and engineers, and cloud consumers interested in the cloud computing future.

There is an ongoing data explosion conspiring that will make previous creations, collections, and storage of data look trivial. Big Data, Mining, and Analytics: Components of Strategic Decision Making ties together big data, data mining, and analytics to explain how readers can leverage them to extract valuable insights from their data. Facilitati

The book ' Data Intensive Computing Applications for Big Data ' discusses the technical concepts of big data, data intensive computing through machine learning, soft computing and parallel computing paradigms. It brings together researchers to report their latest results or progress in the development of the above mentioned areas. Since there are few books on this specific subject, the editors aim to provide a common platform for researchers working in this area to exhibit their novel findings. The book is intended as a reference work for advanced undergraduates and graduate students, as well as multidisciplinary, interdisciplinary and transdisciplinary research workers and scientists on the subjects of big data and cloud/parallel and distributed computing, and explains didactically many of the core concepts of these approaches for practical applications. It is organized into 24 chapters providing a comprehensive overview of big data analysis using parallel computing and addresses the complete data science workflow in the cloud, as well as dealing with privacy issues and the challenges faced in a data-intensive cloud computing environment. The book explores both fundamental and high-level concepts, and will serve as a manual for those in the industry, while also helping beginners to understand the basic and advanced aspects of big data and cloud computing.

Turning Big Data into Big Money

Enabling the New Era of Cloud Computing: Data Security, Transfer, and Management

Proceedings of the 13th International Conference on Broadband and Wireless Computing, Communication and Applications (BWCCA-2018)

Knowledge Graphs and Big Data Processing

First International Conference, BDA 2012, New Delhi, India, December 24-26, 2012, Proceedings

Apache Hadoop YARN

Large Scale Data Analytics under the Hood

Master alternative Big Data technologies that can do what Hadoop can't: real-time analytics and iterative machine learning. When most technical professionals think of Big Data analytics today, they think of Hadoop. But there are many cutting-edge applications that Hadoop isn't well suited for, especially real-time analytics and contexts requiring the use of iterative machine learning algorithms.

Fortunately, several powerful new technologies have been developed specifically for use cases such as these. Big Data Analytics Beyond Hadoop is the first guide specifically designed to help you take the next steps beyond Hadoop. Dr. Vijay Srinivas Agneeswaran introduces the breakthrough Berkeley Data Analysis Stack (BDAS) in detail, including its motivation, design, architecture, Mesos cluster management, performance, and more. He presents realistic use cases and up-to-date example code for: Spark, the next generation in-memory computing technology from UC Berkeley Storm, the parallel real-time Big Data analytics technology from Twitter GraphLab, the next-generation graph processing paradigm from CMU and the University of Washington (with comparisons to alternatives such as Pregel and Piccolo) Halo also offers architectural and design guidance and code sketches for scaling machine learning algorithms to Big Data, and then realizing them in real-time. He concludes by previewing emerging trends, including real-time video analytics, SDNs, and even Big Data governance, security, and privacy issues. He identifies intriguing startups and new research possibilities, including BDAS extensions and cutting-edge model-driven analytics. Big Data Analytics Beyond Hadoop is an indispensable resource for everyone who wants to reach the cutting edge of Big Data analytics, and stay there: practitioners, architects, programmers, data scientists, researchers, startup entrepreneurs, and advanced students.

In this book readers will find technological discussions on the existing and emerging technologies across the different stages of the big data value chain. They will learn about legal aspects of big data, the social impact, and about education needs and requirements. And they will discover the business perspective and how big data technology can be exploited to deliver value within different sectors of the

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

“ This book is a critically needed resource for the newly released Apache Hadoop 2.0, highlighting YARN as the significant breakthrough that broadens Hadoop beyond the MapReduce paradigm. ” —From the Foreword by Raymie Stata, CEO of AltitScale The Insider ’ s Guide to Building Distributed, Big Data Applications with Apache Hadoop™ YARN Apache Hadoop is helping drive the Big Data revolution. Now, its data processing has been completely overhauled: Apache Hadoop YARN provides resource management at data center scale and easier ways to create distributed applications that process petabytes of data. And now in Apache Hadoop™ YARN, two Hadoop technical leaders show you how to develop new applications and adapt existing code to fully leverage these revolutionary advances. YARN project founder Arun Murthy and project lead Vinod Kumar Vavilapalli demonstrate how YARN increases scalability and cluster utilization, enables new programming models and services, and opens new options beyond Java and batch processing. They walk you through the entire YARN project lifecycle, from installation through deployment. You ’ ll find many examples drawn from the authors ’ cutting-edge experience—first as Hadoop ’ s earliest developers and implementers at Yahoo! and now as Hortonworks developers moving the platform forward and helping customers succeed with it. Coverage includes YARN ’ s goals, design, architecture, and components—how it expands the Apache Hadoop ecosystem Exploring YARN on a single node Administering YARN clusters and Capacity Scheduler Running existing MapReduce applications Developing a large-scale clustered YARN application Discovering new open source frameworks that run under YARN

Spanning the multi-disciplinary scope of information technology, the Encyclopedia of Information Systems and Technology draws together comprehensive coverage of the inter-related aspects of information systems and technology. The topics covered in this encyclopedia encompass internationally recognized bodies of knowledge, including those of The IT BOK, the Chartered Information Technology Professionals Program, the International IT Professional Practice Program (British Computer Society), the Core Body of Knowledge for IT Professionals (Australian Computer Society), the International Computer Driving License Foundation (European Computer Driving License Foundation), and the Guide to the Software Engineering Body of Knowledge. Using the universally recognized definitions of IT and information systems from these recognized bodies of knowledge, the encyclopedia brings together the information that students, practicing professionals, researchers, and academicians need to keep their knowledge up to date. Also Available Online This Taylor & Francis encyclopedia is also available through online subscription, offering a variety of extra benefits for researchers, students, and librarians, including: Citation tracking and alerts Active reference linking Saved searches and marked lists HTML and PDF format options Contact Taylor and Francis for more information or to inquire about subscription options and print/online combination packages. US: (Tel) 1.888.318.2367; (E-mail) e-reference@taylorandfrancis.com International: (Tel) +44 (0) 20 7017 6062; (E-mail) online.sales@tandf.co.uk

Encyclopedia of Information Systems and Technology - Two Volume Set

Hadoop 2 Quick-Start Guide

Big Data Management

Spatial Network Big Databases

Learning Spark

Pro Hadoop Data Analytics

Big Data Analytics Beyond Hadoop

Master alternative Big Data technologies that can do what Hadoop can't: real-time analytics and iterative machine learning. When most technical professionals think of Big Data analytics today, they think of Hadoop. But there are many cutting-edge applications that Hadoop isn't well suited for, especially real-time analytics and contexts requiring the use of iterative machine learning algorithms. Fortunately, several powerful new technologies have been developed specifically for use cases such as these. Big Data Analytics Beyond Hadoop is the first guide specifically designed to help you take the next steps beyond Hadoop. Dr. Vijay Srinivas Agneeswaran introduces the breakthrough Berkeley Data Analysis Stack (BDAS) in detail, including its motivation, design, architecture, Mesos cluster management, performance, and more. He presents realistic use cases and up-to-date example code for:Spark, the next generation in-memory computing technology from UC BerkeleyStorm, the parallel real-time Big Data analytics technology from TwitterGraphLab, the next-generation graph processing paradigm from CMU and the University of Washington (with comparisons to alternatives such as Pregel and Piccolo)Halo also offers architectural and design guidance and code sketches for scaling machine learning algorithms to Big Data, and then realizing them in real-time. He concludes by previewing emerging trends, including real-time video analytics, SDNs, and even Big Data governance, security, and privacy issues. He identifies intriguing startups and new research possibilities, including BDAS extensions and cutting-edge model-driven analytics. Big Data Analytics Beyond Hadoop is an indispensable resource for everyone who wants to reach the cutting edge of Big Data analytics, and stay there: practitioners, architects, programmers, data scientists, researchers, startup entrepreneurs, and advanced students.

Learn advanced analytical techniques and leverage existing tool kits to make your analytic applications more powerful, precise, and efficient. This book provides the right combination of architecture, design, and implementation information to create analytical systems that go beyond the basics of classification, clustering, and recommendation. Pro Hadoop Data Analytics emphasizes best practices to ensure coherent, efficient development. A complete example system will be developed using standard third-party components that consist of the tool kits, libraries, visualization and reporting code, as well as support glue to provide a working and extensible end-to-end system. The book also highlights the importance of end-to-end, flexible, configurable, high-performance data pipeline systems with analytical components as well as appropriate visualization results. You'll discover the importance of mix-and-match or hybrid systems, using different analytical components in one application. This hybrid approach will be prominent in the examples. What You'll Learn Build big data analytic systems with the Hadoop ecosystem Use libraries, tool kits, and algorithms to make development easier and more effective Apply metrics to measure performance and efficiency of components and systems Connect to standard relational databases, noSQL data sources, and more Follow case studies with example components to create your own systems Who This Book Is For Software engineers, architects, and data scientists with an interest in the design and implementation of big data analytical systems using Hadoop, the Hadoop ecosystem, and other associated technologies.

This book presents on the latest research findings, and innovative research methods and development techniques related to the emerging areas of broadband and wireless computing from both theoretical and practical perspectives. Information networking is evolving rapidly with various kinds of networks with different characteristics emerging and being integrated into heterogeneous networks. As a result, a number of interconnection problems can occur at different levels of the communicating entities and communication networks' hardware and software design. These networks need to manage an increasing usage demand, provide support for a significant number of services, guarantee their QoS, and optimize the network resources. The success of all-IP networking and wireless technology has changed the way of life for people around the world, and the advances in electronic integration and wireless communications will pave the way for access to the wireless networks on the fly. This in turn means that all electronic devices will be able to exchange the information with each other in a ubiquitous way whenever necessary.

There is an easier way to build Hadoop applications. With this hands-on book, you'll learn how to use Cascading, the open source abstraction framework for Hadoop that lets you easily create and manage powerful enterprise-grade data processing applications—without having to learn the intricacies of MapReduce. Working with sample apps based on Java and other JVM languages, you'll quickly learn Cascading's streamlined approach to data processing, data filtering, and workflow optimization. This book demonstrates how this framework can help your business extract meaningful information from large amounts of distributed data. Start working on Cascading example projects right away Model and analyze unstructured data in any format, from any source Build and test applications with familiar constructs and reusable components Work with the Scalding and Cascalog Domain-Specific Languages Easily deploy applications to Hadoop, regardless of cluster location or data size Build workflows that integrate several big data frameworks and processes Explore common use cases for Cascading, including features and tools that support them Examine a case study that uses a dataset from the Open Data Initiative

Turning Big Data Into Big Money

New Horizons for a Data-Driven Economy

Digital Exhaust

Lightning-Fast Big Data Analysis

Real-World Data Mining

Inventing the Cloud Century

Big Data and Business Analytics

A vital component of any publishing project is the ethical dimensions, which can refer to varied categories of practice: from conducting a proper peer review to using proper citation in research. With the implementation of technology in research and publishing, it is important for today's researchers to address the standards of scientific research and publishing practices to avoid unethical behavior. Scholarly Ethics and Publishing: Breakthroughs in Research and Practice is an essential reference source that discusses various aspects of ethical values in academic settings including methods and tools to prevent and detect plagiarism, strategies for the principled gathering of data, and best practices for conducting and citing research. It also assists researchers in navigating the field of scholarly publishing through a careful analysis of multidisciplinary research topics and recent trends in the industry. Highlighting a range of pertinent topics such as academic writing, publication process, and research methodologies, this publication is an ideal reference source for researchers, graduate students, academicians, librarians, scholars, and industry-leading experts around the globe.

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.

Data analytics is core to business and decision making. The rapid increase in data volume, velocity and variety offers both opportunities and challenges. While open source solutions to store big data, like Hadoop, offer platforms for exploring value and insight from big data, they were not originally developed with data security and governance in mind. Big Data Management discusses numerous policies, strategies and recipes for managing big data. It addresses data security, privacy, controls and life cycle management offering modern principles and open source architectures for successful governance of big data. The author has collected best practices from the world's leading organizations that have successfully implemented big data platforms. The topics discussed cover the entire data management life cycle, data quality, data stewardship, regulatory considerations, data council, architectural and operational models are presented for successful management of big data. The book is a must-read for data scientists, data engineers and corporate leaders who are implementing big data platforms in their organizations.

Explore big data concepts, platforms, analytics, and their applications using the power of Hadoop 3 Key Features Learn Hadoop 3 to build effective big data analytics solutions on-premise and on cloud Integrate Hadoop with other big data tools such as R, Python, Apache Spark, and Apache Flink Exploit big data using Hadoop 3 with real-world examples Book Description Apache Hadoop is the most popular platform for big data processing, and can be combined with a host of other big data tools to build powerful analytics solutions. Big Data Analytics with Hadoop 3 shows you how to do just that, by providing insights into the software as well as its benefits with the help of practical examples. Once you have taken a tour of Hadoop 3's latest features, you will get an overview of HDFS, MapReduce, and YARN, and how they enable faster, more efficient big data processing. You will then move on to learning how to integrate Hadoop with the open source tools, such as Python and R, to analyze and visualize data and perform statistical computing on big data. As you get acquainted with all this, you will explore how to use Hadoop 3 with Apache Spark and Apache Flink for real-time data analytics and stream processing. In addition to this, you will understand how to use Hadoop to build analytics solutions on the cloud and an end-to-end pipeline to perform big data analysis using practical use cases. By the end of this book, you will be well-versed with the analytical capabilities of the Hadoop ecosystem. You will be able to build powerful solutions to perform big data analytics and get insight effortlessly. What you will learn Explore the new features of Hadoop 3 along with HDFS, YARN, and MapReduce Get well-versed with the analytical capabilities of Hadoop ecosystem using practical examples Integrate Hadoop with R and Python for more efficient big data processing Learn to use Hadoop with Apache Spark and Apache Flink for real-time data analytics Set up a Hadoop cluster on AWS cloud Perform big data analytics on AWS using Elastic Map Reduce Who this book is for Big Data Analytics with Hadoop 3 is for you if you are looking to build high-performance analytics solutions for your enterprise or business using Hadoop 3's powerful features, or you're new to big data analytics. A basic understanding of the Java programming language is required.

Big Data Analytics

Real-time Applications with Storm, Spark, and More Hadoop Alternatives

Big Data, Mining, and Analytics

Practical Big Data Analytics

What Everyone Should Know About Big Data, Digitization and Digitally Driven Innovation

Breakthroughs in Research and Practice

Leveraging Engagement, Behavior Change, and Technology for a Healthy Life

Will "Big Data" supercharge the economy, tyrannize us, or both? Data Exhaust is the definitive primer for everyone who wants to understand all the implications of Big Data, digitally driven innovation, and the accelerating Internet Economy. Renowned digital expert Dale Neef clearly explains: What Big Data really is, and what's new and different about it How Big Data works, and what you need to know about Big Data technologies Where the data is coming from: how Big Data integrates sources ranging from social media to machine sensors, smartphones to financial transactions How companies use Big Data analytics to gain a more nuanced, accurate picture of their customers, their own performance, and the newest trends How governments and individual citizens can also benefit from Big Data How to overcome obstacles to success with Big Data – including poor data that can magnify human error A realistic assessment of Big Data threats to employment and personal privacy, now and in the future Neef places the Big Data phenomenon where it belongs: in the context of the broader global shift to the Internet economy, with all that implies. By doing so, he helps businesses plan Big Data strategy more effectively – and helps citizens and policymakers identify sensible policies for preventing its misuse. By conservative estimate, the global Big Data market will soar past \$50 billion by 2018. But those direct expenses represent just the "tip of the iceberg," when it comes to Big Data's impact. Big Data is now of acute strategic interest for every organization that aims to succeed – and it is equally important to everyone else. Whoever you are, Data Exhaust tells you exactly what you need to know about Big Data – and what to do about it, too.

Unique insights to implement big data analytics and reap big returns to your bottom line Focusing on the business and financial value of big data analytics, respected technology journalist Frank J. Ohlhorst shares his insights on the newly emerging field of big data analytics in Big Data Analytics. This breakthrough book demonstrates the importance of analytics, defines the processes, highlights the tangible and intangible values and discusses how you can turn a business liability into actionable material that can be used to redefine markets, improve profits and identify new business opportunities. Reveals big data analytics as the next wave for businesses looking for competitive advantage Takes an in-depth look at the financial value of big data analytics Offers tools and best practices for working with big data Once the domain of large on-line retailers such as eBay and Amazon, big data is now accessible by businesses of all sizes and across industries. From how to mine the data your company collects, to the data that is available on the outside, Big Data Analytics shows how you can leverage big data into a key component in your business's growth strategy.

This book combines the three dimensions of technology, society and economy to explore the advent of today's cloud ecosystems as successors to older service ecosystems based on networks. Further, it describes the shifting of services to the cloud as a long-term trend that is still progressing rapidly.The book adopts a comprehensive perspective on the key success factors for the technology – compelling business models and ecosystems including private, public and national organizations. The authors explore the evolution of service ecosystems, describe the similarities and differences, and analyze the way they have created and changed industries. Lastly, based on the current status of cloud computing and related technologies like virtualization, the internet of things, fog computing, big data and analytics, cognitive computing and blockchain, the authors provide a revealing outlook on the possibilities of future technologies, the future of the internet, and the potential impacts on business and society.

Big Data represents a new era in data exploration and utilization, and IBM is uniquely positioned to help clients navigate this transformation. This book reveals how IBM is leveraging open source Big Data technology, infused with IBM technologies, to deliver a robust, secure, highly available, enterprise-class Big Data platform. The three defining characteristics of Big Data--volume, variety, and velocity--are discussed. You'll get a primer on Hadoop and how IBM is hardening it for the enterprise, and learn when to leverage IBM InfoSphere BigInsights (Big Data at rest) and IBM InfoSphere Streams (Big Data in motion) technologies. Industry use cases are also included in this practical guide. Learn how IBM hardens Hadoop for enterprise-class scalability and reliability Gain insight into IBM's unique in-motion and at-rest Big Data analytics platform Learn tips and tricks for Big Data use cases and solutions Get a quick Hadoop primer

Build highly effective analytics solutions to gain valuable insight into your big data

Emerging Business Intelligence and Analytic Trends for Today's Businesses

Handbook of Big Data Technologies

Principles and Paradigms

Data Security, Transfer, and Management

Enterprise Data Workflows with Cascading

Designing and Building Big Data Systems using the Hadoop Ecosystem

Society is now completely driven by data with many industries relying on data to conduct business or basic functions within the organization. With the efficiencies that big data bring to all institutions, data is continuously being collected and analyzed. However, data sets may be too complex for traditional data-processing, and therefore, different strategies must evolve to solve the issue. The field of big data works as a valuable tool for many different industries. The Research Anthology on Big Data Analytics, Architectures, and Applications is a complete reference source on big data analytics that offers the latest, innovative architectures and frameworks and explores a variety of applications within various industries. Offering an international perspective, the applications discussed within this anthology feature global representation. Covering topics such as advertising curricula, driven supply chain, and smart cities, this research anthology is ideal for data scientists, data analysts, computer engineers, software engineers, technologists, government officials, managers, CEOs, professors, graduate students, researchers, and academicians.

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

Operations management is a tool by which companies can effectively meet customers' needs using the least amount of resources necessary. With the emergence of sensors and smart metering, big data is becoming an intrinsic part of modern operations management. Applied Big Data Analytics in Operations Management enumerates the challenges and creative solutions and tools to apply when using big data in operations management. Outlining revolutionary concepts and applications that help businesses predict customer behavior along with applications of artificial neural networks, predictive analytics, and opinion mining on business management, this comprehensive publication is ideal for IT professionals, software engineers, business professionals, managers, and students of management.

Data in all domains is getting bigger. How can you work with it efficiently? Recently updated for Spark 1.3, this book introduces Apache Spark, the open source cluster computing system that makes data analytics fast to write and fast to run. With Spark, you can tackle big datasets quickly through simple APIs in Python, Java, and Scala. This edition includes new information on Spark SQL, Spark Streaming, setup, and Maven coordinates. Written by the developers of Spark, this book will have data scientists and engineers up and running in no time. You'll learn how to express parallel jobs with just a few lines of code, and cover applications from simple batch jobs to stream processing and machine learning. Quickly dive into Spark capabilities such as distributed datasets, in-memory caching, and the interactive shell Leverage Spark's powerful built-in libraries, including Spark SQL, Spark Streaming, and MLlib Use one programming paradigm instead of mixing and matching tools like Hive, Hadoop, Mahout, and Storm Learn how to deploy interactive, batch, and streaming applications Connect to data sources including HDFS, Hive, JSON, and S3 Master advanced topics like data partitioning and shared variables

Multimedia Big Data Computing for IoT Applications

Big Data

Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data

Scholarly Ethics and Publishing: Breakthroughs in Research and Practice

An Introduction for Data Scientists

Components of Strategic Decision Making

Using Person-Centered Health Analytics to Live Longer

This book considers all aspects of managing the complexity of Multimedia Big Data Computing (MMBD) for IoT applications and develops a comprehensive taxonomy. It also discusses a process model that addresses a number of research challenges associated with MMBD, such as scalability, accessibility, reliability, heterogeneity, and Quality of Service (QoS) requirements, presenting case studies to demonstrate its application. Further, the book examines the layered architecture of MMBD computing and compares the life cycle of both big data and MMBD. Written by leading experts, it also includes numerous solved examples, technical descriptions, scenarios, procedures, and algorithms.

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 American way of producing health is failing. It continues to rank very low among developed countries on our most vital need...to live a long and healthy life. Despite the well-intentioned actions on the part of government, life sciences, and technology, the most important resource for achieving our full health potential is ourselves. This book is about how you can do so, and how others can help you. Dwight McNeill introduces person-centered health analytics (pchA) and shows how you can use it to master five everyday behaviors that cause and perpetuate most chronic diseases. Using Person-Centered Health Analytics to Live Longer combines deep insight, a comprehensive framework, and practical tools for living longer and healthier lives. It offers a clear path forward for both individuals and stakeholders, including providers, payers, health promotion companies, technology innovators, government, and analytics practitioners.

Ready to use statistical and machine-learning techniques across large data sets? This practical guide shows you why the Hadoop ecosystem is perfect for the job. Instead of deployment, operations, or software development usually associated with distributed computing, you'll focus on particular analyses you can build, the data warehousing techniques that Hadoop provides, and higher order data workflows this framework can produce. Data scientists and analysts will learn how to perform a wide range of techniques, from writing MapReduce and Spark applications with Python to using advanced modeling and data management with Spark MLlib, Hive, and HBase. You'll also learn about the analytical processes and data systems available to build and empower data products that can handle—and actually require—huge amounts of data. Understand core concepts behind Hadoop and cluster computing Use design patterns and parallel analytical algorithms to create distributed data analysis jobs Learn about data management, mining, and warehousing in a distributed context using Apache Hive and HBase Use Sqoop and Apache Flume to ingest data from relational databases Program complex Hadoop and Spark applications with Apache Pig and Spark DataFrames Perform machine learning techniques such as classification, clustering, and collaborative filtering with Spark's MLlib

Big Data Analytics: Systems, Algorithms, Applications

Foundations of Data Intensive Applications