

Practical Hadoop By Example Nyoug

Essentials of Forensic Accounting Essentials of Forensic Accounting is an authoritative resource covering a comprehensive range of forensic accounting topics. As a foundation review, a reference book, or as preparation for the Certification in Financial Forensics (CFF®) Exam, this publication will provide thoughtful and insightful examination of the key themes in this field, including: Professional responsibilities and practice management Fundamental forensic knowledge including laws, courts, and dispute resolution Specialized forensic knowledge such as bankruptcy, insolvency, reorganization, and valuation Through illustrative examples, cases, and explanations, this book makes abstract concepts come to life to help you understand and successfully navigate this complex area.

Audience and media analytics is more important now than ever, and this latest volume in the cutting-edge BEA Electronic Media Research Series collects some of the top scholars working with big data and analytics today. These chapters describe the development and help define media analytics as an academic discipline and professional practice. Understanding audiences is integral to creating and distributing media messages and the study of media analytics requires knowing a range of skills including research methods, the necessary tools available, familiarity with statistical procedures, and a mindset to provide insights and apply findings. This book summarizes the insights of analytics practitioners regarding the current state of legacy media analysis and social media analytics. Topics covered include the evolution of media technologies, the teaching of media measurement and analytics, the transition taking place in media research, and the use of media analytics to answer meaningful questions, drive content creation, and engage with audiences.

Tackle the real-world complexities of modern machine learning with innovative, cutting-edge, techniques About This Book Fully-coded working examples using a wide range of machine learning libraries and tools, including Python, R, Julia, and Spark Comprehensive practical solutions taking you into the future of machine learning Go a step further and integrate your machine learning projects with Hadoop Who This Book Is For This book has been created for data scientists who want to see machine learning in action and explore its real-world application. With guidance on everything from the fundamentals of machine learning and predictive analytics to the latest innovations set to lead the big data revolution into the future, this is an unmissable resource for anyone dedicated to tackling current big data challenges. Knowledge of programming (Python and R) and mathematics is advisable if you want to get started immediately. What You Will Learn Implement a wide range of algorithms and techniques for tackling complex data Get to grips with some of the most powerful languages in data science, including R, Python, and Julia Harness the capabilities of Spark and Hadoop to manage and process data successfully Apply the appropriate machine learning technique to address real-world problems Get acquainted with Deep learning and find out how neural networks are being used at the cutting-edge of machine learning Explore the future of machine learning and dive deeper into polyglot persistence, semantic data, and more In Detail Finding meaning in increasingly larger and more complex datasets is a growing demand of the modern world. Machine learning and predictive analytics have become the most important approaches to uncover data gold mines. Machine learning uses complex algorithms to make improved predictions of outcomes based on historical patterns and the behaviour of data sets. Machine learning can deliver dynamic insights into trends, patterns, and relationships within data, immensely valuable to business growth and development. This book explores an extensive range of machine learning techniques uncovering hidden tricks and tips for several types of data using practical and real-world examples. While machine learning can be highly theoretical, this book offers a refreshing hands-on approach without losing sight of the underlying principles. Inside, a full exploration of the various algorithms gives you high-quality guidance so you can begin to see just how effective machine learning is at tackling contemporary challenges of big data. This is the only book you need to implement a whole suite of open source tools, frameworks, and languages in machine learning. We will cover the leading data science languages, Python and R, and the underrated but powerful Julia, as well as a range of other big data platforms including Spark, Hadoop, and Mahout. Practical Machine Learning is an essential resource for the modern data scientists who want to get to grips with its real-world application. With this book, you will not only learn the fundamentals of machine learning but dive deep into the complexities of real world data before moving on to using Hadoop and its wider ecosystem of tools to process and manage your structured and unstructured data. You will explore different machine learning techniques for both supervised and unsupervised learning: from decision trees to Naive Bayes classifiers and linear and clustering methods, you will learn strategies for a truly advanced approach to the statistical analysis of data. The book also explores the cutting-edge advancements in machine learning, with worked examples and guidance on deep learning and reinforcement learning, providing you with practical demonstrations and samples that help take the theory-and mystery-out of even the most advanced machine learning methodologies. Style and approach A practical data science tutorial designed to give you an insight into the practical application of machine learning, this book takes you through complex concepts and tasks in an accessible way. Featuring information on a wide range of data science techniques, Practical Machine Learning is a comprehensive data science resource.

This book describes current problems in data science and Big Data. Key topics are data classification, Graph Cut, the Laplacian Matrix, Google Page Rank, efficient algorithms, hardness of problems, different types of big data, geometric data structures, topological data processing, and various learning methods. For unsolved problems such as incomplete data relation and reconstruction, the book includes possible solutions and both statistical and computational methods for data analysis. Initial chapters focus on exploring the properties of incomplete data sets and partial-connectedness among data points or data sets. Discussions also cover the completion problem of Netflix matrix: machine learning method on massive data sets: image segmentation and video search. This book introduces software tools for data science and Big Data such MapReduce, Hadoop, and Spark. This book contains three parts. The first part explores the fundamental tools of data science. It includes basic graph theoretical methods, statistical and AI methods for massive data sets. In second part, chapters focus on the procedural treatment of data science problems including machine learning methods, mathematical image and video processing, topological data analysis, and statistical methods. The final section provides case studies on special topics in variational learning, manifold learning, business and financial data recovery, geometric search, and computing models. Mathematical Problems in Data Science is a valuable resource for researchers and professionals working in data science, information systems and networks. Advanced-level students studying computer science, electrical engineering and mathematics will also find the content helpful.

Big Data in Practice

Principles of Data Wrangling

Apache Hadoop YARN

How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results

Data Analytics: Principles, Tools, and Practices

Beyond the Basics Using SAS, Third Edition

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

A handy reference guide for data analysts and data scientists to help to obtain value from big data analytics using Spark on Hadoop clusters About This Book This book is based on the latest 2.0 version of Apache Spark and 2.7 version of Hadoop integrated with most commonly used tools. Learn all Spark stack components including latest topics such as DataFrames, DataSets, GraphFrames, Structured Streaming, DataFrame based ML Pipelines and SparkR. Integrations with frameworks such as HDFS, YARN and tools such as Jupyter, Zeppelin, NiFi, Mahout, HBase Spark Connector, GraphFrames, H2O and Hivemall. Who This Book Is For Though this book is primarily aimed at data analysts and data scientists, it will also help architects, programmers, and practitioners. Knowledge of either Spark or Hadoop would be beneficial. It is assumed that you have basic programming background in Scala, Python, SQL, or R programming with basic Linux experience. Working experience within big data environments is not mandatory. What You Will Learn Find out and implement the tools and techniques of big data analytics using Spark on Hadoop clusters with wide variety of tools used with Spark and Hadoop Understand all the Hadoop and Spark ecosystem components Get to know all the Spark components: Spark Core, Spark SQL, DataFrames, DataSets, Conventional and Structured Streaming, MLlib, ML Pipelines and GraphX See batch and real-time data analytics using Spark Core, Spark SQL, and Conventional and Structured Streaming Get to grips with data science and machine learning using MLlib, ML Pipelines, H2O, Hivemall, GraphX, SparkR and Hivemall. In Detail Big Data Analytics book aims at providing the fundamentals of Apache Spark and Hadoop. All Spark components – Spark Core, Spark SQL, DataFrames, Data sets, Conventional Streaming, Structured Streaming, MLlib, GraphX and Hadoop core components – HDFS, MapReduce and Yarn are explored in greater depth with implementation examples on Spark + Hadoop clusters. It is moving away from MapReduce to Spark. So, advantages of Spark over MapReduce are explained at great depth to reap benefits of in-memory speeds. DataFrames API, Data Sources API and new Data set API are explained for building Big Data analytical applications. Real-time data analytics using Spark Streaming with Apache Kafka and HBase is covered to help building streaming applications. New Structured streaming concept is explained with an IOT (Internet of Things) use case. Machine learning techniques are covered using MLlib, ML Pipelines and SparkR and Graph Analytics are covered with GraphX and GraphFrames components of Spark. Readers will also get an opportunity to get started with web based notebooks such as Jupyter, Apache Zeppelin and data flow tool Apache NiFi to analyze and visualize data. Style and approach This step-by-step pragmatic guide will make life easy no matter what your level of experience. You will deep dive into Apache Spark on Hadoop clusters through ample exciting real-life examples. Practical tutorial explains data science in simple terms to help programmers and data analysts get started with Data Science

Big data solutions enable us to change how we do business by exploiting previously unused sources of information in ways that were not possible just a few years ago. In IBM® Smarter Planet® terms, big data helps us to change the way that the world works. The purpose of this IBM Redpaper™ publication is to consider the performance and capacity implications of big data solutions, which must be taken into account for them to be viable. This paper describes the benefits that big data approaches can provide. We then cover performance and capacity considerations for creating big data solutions. We conclude with what this means for big data solutions, both now and in the future. Intended readers for this paper include decision-makers, consultants, and IT architects.

A practical guide to obtaining, transforming, exploring, and analyzing data using Python, MongoDB, and Apache Spark About This Book Learn to use various data analysis tools and algorithms to classify, cluster, visualize, simulate, and forecast your data Apply Machine Learning algorithms to different kinds of data such as social networks, time series, and images A hands-on guide to understanding the nature of data and how to turn it into insight Who This Book Is For This book is for developers who want to implement data analysis and data-driven algorithms in a practical way. It is also suitable for those without a background in data analysis or data processing. Basic knowledge of Python programming, statistics, and linear algebra is assumed. What You Will Learn Acquire, format, and visualize your data Build an image-similarity search engine Generate meaningful visualizations anyone can understand Get started with analyzing social network graphs Find out how to implement sentiment text analysis Install data analysis tools such as Pandas, MongoDB, and Apache Spark Get to grips with Apache Spark Implement machine learning algorithms such as classification or forecasting In Detail Beyond buzzwords like Big Data or Data Science, there are a great opportunities to innovate in many businesses using data analysis to get data-driven products. Data analysis involves asking many questions about data in order to discover insights and generate value for a product or a service. This book explains the basic data algorithms without the theoretical jargon, and you'll get hands-on turning data into insights using machine learning techniques. We will perform data-driven innovation processing for several types of data such as text, Images, social network graphs, documents, and time series, showing you how to implement large data processing with MongoDB and Apache Spark. Style and approach This is a hands-on guide to data analysis and data processing. The concrete examples are explained with simple code and accessible data.

Although interest in machine learning has reached a high point, lofty expectations often scuttle projects before they get very far. How can machine learning—especially deep neural networks—make a real difference in your organization? This hands-on guide not only provides the most practical information available on the subject, but also helps you get started building efficient deep learning networks. Authors Adam Gibson and Josh Patterson provide theory on deep learning before introducing their open-source DeepLearning4j (DL4J) library for developing production-class workflows. Through real-world examples, you'll learn methods and strategies for training deep network architectures and running deep learning workflows on Spark and Hadoop with DL4J. Dive into machine learning concepts in general, as well as deep learning in particular Understand how deep networks evolved from neural network fundamentals Explore the major deep network architectures, including Convolutional and Recurrent Learn how to map specific deep networks to the right problem Walk through the fundamentals of tuning general neural networks and specific deep network architectures Use vectorization techniques for different data types with DataVec, DL4J's workflow tool Learn how to use DL4J natively on Spark and Hadoop

Hadoop: The Definitive Guide

Unlocking Hadoop for Your Relational Database

Expert Hadoop 2 Administration

Application of Intelligent Systems in Multi-modal Information Analytics

A Beginner's Guide to Storytelling with Data

Practical Data Analysis

Apache Spark for the Enterprise: Setting the Business Free

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

Practical SQL is an approachable and fast-paced guide to SQL (Structured Query Language), the standard programming language for defining, organizing, and exploring data in relational databases. The book focuses on using SQL to find the story your data tells, with the popular open-source database PostgreSQL and the pgAdmin interface as its primary tools. You'll first cover the fundamentals of databases and the SQL language, then build skills by analyzing data from the U.S. Census and other federal and state government agencies. With exercises and real-world examples in each chapter, this book will teach even those who have never programmed before all the tools necessary to build powerful databases and access information quickly and efficiently. You'll learn how to: - Create databases and related tables using your own data - Define the right data types for your information - Aggregate, sort, and filter data to find patterns - Use basic math and advanced statistical functions - Identify errors in data and clean them up - Import and export data using delimited text files - Write queries for geographic information systems (GIS) - Create advanced queries and automate tasks Learning SQL doesn't have to be dry and complicated. Practical SQL delivers clear examples with an easy-to-follow approach to teach you the tools you need to build and manage your own databases. This book uses PostgreSQL, but the SQL syntax is applicable to many database applications, including Microsoft SQL Server and MySQL.

PROC SQL: Beyond the Basics Using SAS®, Third Edition, is a step-by-step, example-driven guide that helps readers master the language of PROC SQL. Packed with analysis and examples illustrating an assortment of PROC SQL options, statements, and clauses, this book not only covers all the basics, but it also offers extensive guidance on complex topics such as set operators and correlated subqueries. Programmers at all levels will appreciate Kirk Laffler's easy-to-follow examples, clear explanations, and handy tips to extend their knowledge of PROC SQL. This third edition explores new and powerful features in SAS® 9.4, including topics such as: IFC and IFN functions nearest neighbor processing the HAVING clause indexes It also features two completely new chapters on fuzzy matching and data-driven programming. Delving into the workings of PROC SQL with greater analysis and discussion, PROC SQL: Beyond the Basics Using SAS®, Third Edition, explores this powerful database language using discussion and numerous real-world examples.

This is the eBook of the printed book and may not include any media, website access codes, or print supplements that may come packaged with the bound book. The Comprehensive, Up-to-Date Apache Hadoop Administration Handbook and Reference "Sam Alapati has worked with production Hadoop clusters for six years. His unique depth of experience has enabled him to write the go-to resource for all administrators looking to spec, size, expand, and secure production Hadoop clusters of any size." —Paul Dix, Series Editor In Expert Hadoop® Administration, leading Hadoop administrator Sam R. Alapati brings together authoritative knowledge for creating, configuring, securing, managing, and optimizing production Hadoop clusters in any environment. Drawing on his experience with large-scale Hadoop administration, Alapati integrates action-oriented advice with carefully researched explanations of both problems and solutions. He covers an unmatched range of topics and offers an unparalleled collection of realistic examples. Alapati demystifies complex Hadoop environments, helping you understand exactly what happens behind the scenes when you administer your cluster. You'll gain unprecedented insight as you walk through building clusters from scratch and configuring high availability, performance, security, encryption, and other key attributes. The high-value administration skills you learn here will be indispensable no matter what Hadoop distribution you use or what Hadoop applications you run. Understand Hadoop's architecture from an administrator's standpoint Create simple and fully distributed clusters Run MapReduce and Spark applications in a Hadoop cluster Manage and protect Hadoop data and high availability Work with HDFS commands, file permissions, and storage management Move data, and use YARN to allocate resources and schedule jobs Manage job workflows with Oozie and Hue Secure, monitor, log, and optimize Hadoop Benchmark and troubleshoot Hadoop

Delivering the Promise of Big Data and Data Science

Managing Spark, YARN, and MapReduce

New Age Education Models

Data Lake for Enterprises

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

Mahout in Action

Moving Beyond MapReduce and Batch Processing with Apache Hadoop 2

Cookbook recipes demonstrate Hadoop in action and then explain the concepts behind the code. This book is ideal for developers who wish to have a better understanding of Hadoop application development and associated tools, and developers who understand Hadoop conceptually but want practical examples of real world applications.

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 is written for Windows professionals who are familiar with PowerShell and want to learn to build, operate, and administer their Windows workloads in the Microsoft cloud. Pro PowerShell for Microsoft Azure is packed with practical examples and scripts, with easy-to-follow explanations for a wide range of day-to-day needs and essential administration tasks. Author Sherif Talaat begins by explaining the fundamental concepts behind the Microsoft Azure platform and how to get started configuring it through PowerShell. Readers will find out how to deploy, configure and manage the various components of the Azure platform, from storage and virtual networks to Azure Web Sites, HDInsight clusters and the Azure SQL Database. Workload automation, scheduling and resource management are covered in depth to help build efficiency in everyday tasks, and administrators will gain full control over Azure identity and access rights using Azure Active Directory and Rights Management Services. Put your PowerShell skills to good use and ensure that your applications and data are available anywhere at any time, with Pro PowerShell for Microsoft Azure. What You'll Learn Create and manage virtual networks and VPNs using PowerShell. Configure and maintain Azure Storage accounts, blobs, and containers. Provision and manage a redundant Windows or Linux server. Deploy and configure your sites in the cloud using Microsoft Azure Web Sites. Provision Apache Hadoop clusters in the cloud using Azure HDInsight. Deploy, configure and manage a Microsoft Azure SQL Database. Protect and secure identities and resources with Azure Active Directory and Azure Rights Management Services. Who This Book Is For This book is for the intermediate to advanced Windows professional who is ready to make the leap to the cloud.

Integrating data from multiple sources is essential in the age of big data, but it can be a challenging and time-consuming task. This handy cookbook provides dozens of ready-to-use recipes for using Apache Sqoop, the command-line interface application that optimizes data transfers between relational databases and Hadoop. Sqoop is both powerful and bewildering, but with this cookbook's problem-solution-discussion format, you'll quickly learn how to deploy and then apply Sqoop in your environment. The authors provide MySQL, Oracle, and PostgreSQL database examples on GitHub that you can easily adapt for SQL Server, Netezza, Teradata, or other relational systems. Transfer data from a single database table into your Hadoop ecosystem Keep table data and Hadoop in sync by importing data incrementally Import data from more than one database table Customize transferred data by calling various database functions Export generated, processed, or backed-up data from Hadoop to your database Run Sqoop within Oozie, Hadoop's specialized workflow scheduler Load data into Hadoop's data warehouse (Hive) or database (HBase) Handle installation, connection, and syntax issues common to specific database vendors

Web Data Management

Practical Machine Learning with Python

Hadoop Real World Solutions Cookbook

Essentials of Forensic Accounting

Theoretical and Practical Methods

Consumer-Driven Technologies in Healthcare: Breakthroughs in Research and Practice

Performance and Capacity Implications for Big Data

Hadoop in Action teaches readers how to use Hadoop and write MapReduce programs. The intended readers are programmers, architects, and project managers who have to process large amounts of data offline. Hadoop in Action will lead the reader from obtaining a copy of Hadoop to setting it up in a cluster and writing data analytic programs. The book begins by making the basic idea of Hadoop and MapReduce easier to grasp by applying the default Hadoop installation to a few easy-to-follow tasks, such

as analyzing changes in word frequency across a body of documents. The book continues through the basic concepts of MapReduce applications developed using Hadoop, including a close look at framework components, use of Hadoop for a variety of data analysis tasks, and numerous examples of Hadoop in action. Hadoop in Action will explain how to use Hadoop and present design patterns and practices of programming MapReduce. MapReduce is a complex idea both conceptually and in its implementation, and Hadoop users are challenged to learn all the knobs and levers for running Hadoop. This book takes you beyond the mechanics of running Hadoop, teaching you to write meaningful programs in a MapReduce framework. This book assumes the reader will have a basic familiarity with Java, as most code examples will be written in Java. Familiarity with basic statistical concepts (e.g. histogram, correlation) will help the reader appreciate the more advanced data processing examples. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book.

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

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

This volume represents the proceedings of the 7th International Conference on Innovation, Communication and Engineering (ICICE 2018), which was held in P.R. China, November 9-14, 2018. The conference aimed to provide an integrated communication platform for researchers in a wide range of fields including information technology, communication science, applied mathematics, computer science, advanced material science, and engineering. Hopefully, the conference and resulting proceedings will enhance interdisciplinary collaborations between science and engineering technologists in academia and industry within this unique international network.

Applications of Big Data in Healthcare

Breakthroughs in Research and Practice

A Complete Guide for Advanced Data Analytics Using the Latest Trends, Tools, and Technologies (English Edition)

Proceedings of the 7th International Conference on Innovation, Communication and Engineering (ICICE 2018), November 9-14, 2018, Hangzhou, China

The Golden Age of Data

Engineering Innovation and Design

Innovation and Reform in the 21st century

Building models is a small part of the story when it comes to deploying machine learning applications. The entire process involves developing, orchestrating, deploying, and running scalable and portable machine learning workloads--a process Kubeflow makes much easier. This practical book shows data scientists, data engineers, and platform architects how to plan and execute a Kubeflow project to make their Kubernetes workflows portable and scalable. Authors Josh Patterson, Michael Katzenellenbogen, and Austin Harris demonstrate how this open source platform orchestrates workflows by managing machine learning pipelines. You'll learn how to plan and execute a Kubeflow platform that can support workflows from on-premises to cloud providers including Google, Amazon, and Microsoft. Dive into Kubeflow architecture and learn best practices for using the platform Understand the process of planning your Kubeflow deployment Install Kubeflow on an existing on-premises Kubernetes cluster Deploy Kubeflow on Google Cloud Platform step-by-step from the command line Use the managed Amazon Elastic Kubernetes Service (EKS) to deploy Kubeflow on AWS Deploy and manage Kubeflow across a network of Azure cloud data centers around the world Use KFServing to develop and deploy machine learning models

Applications of Big Data in Healthcare: Theory and Practice begins with the basics of Big Data analysis and introduces the tools, processes and procedures associated with Big Data analytics. The book unites healthcare with Big Data analysis and uses the advantages of the latter to solve the problems faced by the former. The authors present the challenges faced by the healthcare industry, including capturing, storing, searching, sharing and analyzing data. This book illustrates the challenges in the applications of Big Data and suggests ways to overcome them, with a primary emphasis on data repositories, challenges, and concepts for data scientists, engineers and clinicians. The applications of Big Data have grown tremendously within the past few years and its growth can not only be attributed to its competence to handle large data streams but also to its abilities to find insights from complex, noisy, heterogeneous, longitudinal and voluminous data. The main objectives of Big Data in the healthcare sector is to come up with ways to provide personalized healthcare to patients by taking into account the enormous amounts of already existing data. Provides case studies that illustrate the business processes underlying the use of big data and deep learning health analytics to improve health care delivery Supplies readers with a foundation for further specialized study in clinical analysis and data management Includes links to websites, videos, articles and other online content to expand and support the primary learning objectives for each major section of the book

Summary Hadoop in Practice, Second Edition provides over 100 tested, instantly useful techniques that will help you conquer big data, using Hadoop. This revised new edition covers changes and new features in the Hadoop core architecture, including MapReduce 2. Brand new chapters cover YARN and integrating Kafka, Impala, and Spark SQL with Hadoop. You'll also get new and updated techniques for Flume, Sqoop, and Mahout, all of which have seen major new versions recently. In short, this is the most practical, up-to-date coverage of Hadoop available anywhere. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Book It's always a good time to upgrade your Hadoop skills! Hadoop in Practice, Second Edition provides a collection of 104 tested, instantly useful techniques for analyzing real-time streams, moving data securely, machine learning, managing large-scale clusters, and taming big data using Hadoop. This completely revised edition covers changes and new features in Hadoop core, including MapReduce 2 and YARN. You'll pick up hands-on best practices for integrating Spark, Kafka, and Impala with Hadoop, and get new and updated techniques for the latest versions of Flume, Sqoop, and Mahout. In short, this is the most practical, up-to-date coverage of Hadoop available. Readers need to know a programming language like Java and have basic familiarity with Hadoop. What's Inside Thoroughly updated for Hadoop 2 How to write YARN applications Integrate real-time technologies like Storm, Impala, and Spark Predictive analytics using Mahout and RR Readers need to know a programming language like Java and have basic familiarity with Hadoop. About the Author Alex Holmes works on tough big-data problems. He is a software engineer, author, speaker, and blogger specializing in large-scale Hadoop projects. Table of Contents PART 1 BACKGROUND AND FUNDAMENTALS Hadoop in a heartbeat Introduction to YARN PART 2 DATA LOGISTICS Data serialization—working with text and beyond Organizing and optimizing data in HDFS Moving data into and out of Hadoop PART 3 BIG DATA PATTERNS Applying MapReduce patterns to big data Utilizing data structures and algorithms at scale Tuning, debugging, and testing PART 4 BEYOND MAPREDUCE SQL on Hadoop Writing a YARN application Analytics is increasingly an integral part of day-to-day operations at today's leading businesses, and transformation is also occurring through huge growth in mobile and digital channels. Enterprise organizations are attempting to leverage analytics in new ways and transition existing analytics capabilities to respond with more flexibility while making the most efficient use of highly valuable data science skills. The recent growth and adoption of Apache Spark as an analytics framework and platform is very timely and helps meet these challenging demands. The Apache Spark environment on IBM z/OS® and Linux on IBM z Systems™ platforms allows this analytics framework to run on the same enterprise platform as the originating sources of data and transactions that feed it. If most of the data that will be used for Apache Spark analytics, or the most sensitive or quickly changing data is originating on z/OS, then an Apache Spark z/OS based environment will be the optimal choice for performance, security, and governance. This IBM® Redpaper™ publication explores the enterprise analytics market, use of Apache Spark on IBM z Systems™ platforms, integration between Apache Spark and other enterprise data sources, and case studies and examples of what can be achieved with Apache Spark in enterprise environments. It is of interest to data scientists, data engineers, enterprise architects, or anybody looking to better understand how to combine an analytics framework and platform on enterprise systems.

Proceedings of CSI 2015

Theory and Practice

Discovering, Analyzing, Visualizing and Presenting Data

Apache Sqoop Cookbook

The Enterprise Big Data Lake

The 4th International Conference on Multi-modal Information Analytics (ICMIA 2022), Volume 1

Practical SQL

This volume comprises the select proceedings of the annual convention of the Computer Society of India. Divided into 10 topical volumes, the proceedings present papers on state-of-the-art research, surveys, and succinct reviews. The volumes cover diverse topics ranging from communications networks to big data analytics, and from system architecture to cyber security. This volume focuses on Big Data Analytics. The contents of this book will be useful to researchers and students alike.

Summary Mahout in Action is a hands-on introduction to machine learning with Apache Mahout. Following real-world examples, the book presents practical use cases and then illustrates how Mahout can be applied to solve them. Includes a free audio- and video-enhanced ebook. About the Technology A computer system that learns and adapts as it collects data can be really powerful. Mahout, Apache's open source machine learning project, captures the core algorithms of recommendation systems, classification, and clustering in ready-to-use, scalable libraries. With Mahout, you can immediately apply to your own projects the machine learning techniques that drive Amazon, Netflix, and others. About this Book This book covers machine learning using Apache Mahout. Based on experience with real-world applications, it introduces practical use cases and illustrates how Mahout can be applied to solve them. It places particular focus on issues of scalability and how to apply these techniques against large data sets using the Apache Hadoop framework. This book is written for developers familiar with Java -- no prior experience with Mahout is assumed. Owners of a Manning eBook purchased anywhere in the world can download a free eBook from manning.com at any time. They can do so multiple times and in any or all formats available (PDF, ePub or Kindle). To do so, customers must register their printed copy on Manning's site by creating a user account and then following instructions printed on the eBook registration insert at the front of the book. What's Inside Use group data to make individual recommendations Find logical clusters within your data Filter and refine with on-the-fly classification Free audio and video extras Table of Contents Meet Apache Mahout PART 1 RECOMMENDATIONS Introducing recommenders Representing recommender data Making recommendations Taking recommenders to production Distributing recommendation computations PART 2 CLUSTERING Introduction to clustering Representing data Clustering algorithms in Mahout Evaluating and improving clustering quality Taking clustering to production Real-world applications of clustering PART 3 CLASSIFICATION Introduction to classification Training a classifier Evaluating and tuning a classifier Deploying a classifier Case study: Shop It To Me Rethinking, Redesigning, and Restructuring of higher education is the need of the hour. This book intends to enlighten its readers about the 21st-century education models for 21st-century learners. Such models would include skill-oriented and technology-based teaching and learning. This book suggests various routes for student entrepreneurs to access resources throughout their college life to ultimately achieve their goals. It also talks about the effectiveness of e-learning tools in education, emerging teaching techniques, and methodologies by understanding the education models of different countries and benchmarking the best practices. The book also highlights the importance and advantage of open-source learning, mobile learning, and the role of creativity and its relevance with technology.

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!

Practical SQL, 2nd Edition

Practical Techniques for Data Preparation

Big Data Analytics

PROC SQL

Kubeflow Operations Guide

Country Life

Deep Learning

The Internet and World Wide Web have revolutionized access to information. Users now store information across multiple platforms from personal computers to smartphones and websites. As a consequence, data management concepts, methods and techniques are increasingly focused on distribution concerns. Now that information largely resides in the network, so do the tools that process this information. This book explains the foundations of XML with a focus on data distribution. It covers the many facets of distributed data management on the Web, such as description logics, that are already emerging in today's data integration applications and herald tomorrow's semantic Web. It also introduces the machinery used to manipulate the unprecedented amount of data collected on the Web. Several 'Putting into Practice' chapters describe detailed practical applications of the technologies and techniques. The book will serve as an introduction to the new, global, information systems for Web professionals and master's level courses.

This is the eBook of the printed book and may not include any media, website access codes, or print supplements that may come packaged with the bound book. The definitive guide to next generation digital measurement; Indispensable insight for building high-value digital experiences! Helps you capture the knowledge you need to deliver deep personalization at scale Reflects today's latest insights into digital behavior and consumer psychology For every digital marketer, analyst, and executive who wants to improve performance To win at digital, you must capture the right data, quickly transform it into the right knowledge, and use them both to deliver deep personalization at scale. Conventional digital metrics simply aren't up to the task. Now, Gary Angel shows how to reinvent digital measurement so it delivers all you need to create richer, more compelling digital experiences. Angel shows how to transform "raw facts" about digital behavior into meaningful knowledge about your visitors... what they were trying to accomplish... how well you helped them... how you can personalize and optimize their digital experiences from now on... how you can use measurement to provide deep personalization at scale.

The data lake is a daring new approach for harnessing the power of big data technology and providing convenient self-service capabilities. But is it right for your company? This book is based on discussions with practitioners and executives from more than a hundred organizations, ranging from data-driven companies such as Google, LinkedIn, and Facebook, to governments and traditional corporate enterprises. You'll learn what a data lake is, why enterprises need one, and how to build one successfully with the best practices in this book. Alex Gorelik, CTO and founder of Waterline Data, explains why old systems and processes can no longer support data needs in the enterprise. Then, in a collection of essays about data lake implementation, you'll examine data lake initiatives, analytic projects, experiences, and best practices from data experts working in various industries. Get a succinct introduction to data warehousing, big data, and data science Learn various paths enterprises take to build a data lake Explore how to build a self-service model and best practices for providing analysts access to the data Use different methods for architecting your data lake Discover ways to implement a data lake from experts in different industries Ready to unlock the power of your data? With this comprehensive guide, you'll learn how to build and maintain reliable, scalable, distributed systems with Apache Hadoop. This book is ideal for programmers looking to analyze datasets of any size, and for administrators who want to set up and run Hadoop clusters. You'll find illuminating case studies that demonstrate how Hadoop is used to solve specific problems. This third edition covers recent changes to Hadoop, including material on the new MapReduce API, as well as MapReduce 2 and its more flexible execution model (YARN). Store large datasets with the Hadoop Distributed File System (HDFS) Run distributed computations with MapReduce Use Hadoop's data and I/O building blocks for compression, data integrity, serialization (including Avro), and persistence Discover common pitfalls and advanced features for writing real-world MapReduce programs Design, build, and administer a dedicated Hadoop cluster—or run Hadoop in the cloud Load data from relational databases into HDFS, using Sqoop Perform large-scale data processing with the Pig query language Analyze datasets with Hive, Hadoop's data warehousing system Take advantage of HBase for structured and semi-structured data, and ZooKeeper for building distributed systems

Measuring the Digital World

Data Science and Big Data Analytics

Pro PowerShell for Microsoft Azure

Media Analytics in Study & Practice

Hadoop in Practice

Hadoop in Action

Using Digital Analytics to Drive Better Digital Experiences

Analyze data like a pro, even if you're a beginner. Practical SQL is an approachable and fast-paced guide to SQL (Structured Query Language), the standard programming language for defining, organizing, and exploring data in relational databases. Anthony DeBarros, a journalist and data analyst, focuses on using SQL to find the story within your data. The examples and code use the open-source database PostgreSQL and its companion pgAdmin interface, and the concepts you learn will apply to most database management systems, including MySQL, Oracle, SQLite, and others.* You'll first cover the fundamentals of databases and the SQL language, then build skills by analyzing data from real-world datasets such as US Census demographics, New York City taxi rides, and earthquakes from US Geological Survey. Each chapter includes exercises and examples that teach even those who have never programmed before all the tools necessary to build powerful databases and access information quickly and efficiently. You'll learn how to: • Create databases and related tables using your own data • Aggregate, sort, and filter data to find patterns • Use functions for basic math and advanced statistical operations • Identify errors in data and clean them up • Analyze spatial data with a geographic information system (PostGIS) • Create advanced queries and automate tasks This updated second edition has been thoroughly revised to reflect the latest in SQL features, including additional advanced query techniques for wrangling data. This edition also has two new chapters: an expanded set of instructions on for setting up your system plus a chapter on using PostgreSQL with the popular JSON data interchange format. Learning SQL doesn't have to be dry and complicated. Practical SQL delivers clear examples with an easy-to-follow approach to teach you the tools you need to build and manage your own databases. * Microsoft SQL Server employs a variant of the language called T-SQL, which is not covered by Practical SQL.

"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." -- From the Amazon

A Complete Data Analytics Guide for Learners and Professionals. KEY FEATURES ● Learn Big Data, Hadoop Architecture, HBase, Hive and NoSQL Database. ● Dive into Machine Learning, its tools, and applications. ● Coverage of applications of Big Data, Data Analysis, and Business Intelligence. DESCRIPTION These days critical problem solving related to data and data sciences is in demand. Professionals who can solve real data science problems using data science tools are in demand. The book "Data Analytics: Principles, Tools, and Practices" can be considered a handbook or a guide for professionals who want to start their journey in the field of data science. The journey starts with the introduction of DBMS, RDBMS, NoSQL, and DocumentDB. The book introduces the essentials of data science and the modern ecosystem, including the important steps such as data ingestion, data munging, and visualization. The book covers the different types of analysis, different Hadoop ecosystem tools like Apache Spark, Apache Hive, R, MapReduce, and NoSQL Database. It also includes the different machine learning techniques that are useful for data analytics and how to visualize data with different graphs and charts. The book discusses useful tools and approaches for data analytics, supported by concrete code examples. After reading this book, you will be motivated to explore real data analytics and make use of the acquired knowledge on databases, BI/DW, data visualization, Big Data tools, and statistical science. WHAT YOU WILL LEARN ● Familiarize yourself with Apache Spark, Apache Hive, R,

MapReduce, and NoSQL Database. ● Learn to manage data warehousing with real time transaction processing. ● Explore various machine learning techniques that apply to data analytics. ● Learn how to visualize data using a variety of graphs and charts using real-world examples from the industry. ● Acquaint yourself with Big Data tools and statistical techniques for machine learning. WHO THIS BOOK IS FOR IT graduates, data engineers and entry-level professionals who have a basic understanding of the tools and techniques but want to learn more about how they fit into a broader context are encouraged to read this book. TABLE OF CONTENTS 1. Database Management System 2. Online Transaction Processing and Data Warehouse 3. Business Intelligence and its deeper dynamics 4. Introduction to Data Visualization 5. Advanced Data Visualization 6. Introduction to Big Data and Hadoop 7. Application of Big Data Real Use Cases 8. Application of Big Data 9. Introduction to Machine Learning 10. Advanced Concepts to Machine Learning 11. Application of Machine Learning

A key task that any aspiring data-driven organization needs to learn is data wrangling, the process of converting raw data into something truly useful. This practical guide provides business analysts with an overview of various data wrangling techniques and tools, and puts the practice of data wrangling into context by asking, "What are you trying to do and why?" Wrangling data consumes roughly 50-80% of an analyst's time before any kind of analysis is possible. Written by key executives at Trifacta, this book walks you through the wrangling process by exploring several factors—time, granularity, scope, and structure—that you need to consider as you begin to work with data. You'll learn a shared language and a comprehensive understanding of data wrangling, with an emphasis on recent agile analytic processes used by many of today's data-driven organizations. Appreciate the importance—and the satisfaction—of wrangling data the right way. Understand what kind of data is available Choose which data to use and at what level of detail Meaningfully combine multiple sources of data Decide how to distill the results to a size and shape that can drive downstream analysis

Mathematical Problems in Data Science

Data Science at the Command Line

Facing the Future with Time-Tested Tools

A Practitioner's Approach

Practical Machine Learning

This hands-on guide demonstrates how the flexibility of the command line can help you become a more efficient and productive data scientist. You'll learn how to combine small, yet powerful, command-line tools to quickly obtain, scrub, explore, and model your data. To get you started—whether you're on Windows, OS X, or Linux—author Jeroen Janssens introduces the Data Science Toolbox, an easy-to-install virtual environment packed with over 80 command-line tools. Discover why the command line is an agile, scalable, and extensible technology. Even if you're already comfortable processing data with, say, Python or R, you'll greatly improve your data science workflow by also leveraging the power of the command line. Obtain data from websites, APIs, databases, and spreadsheets Perform scrub operations on plain text, CSV, HTML/XML, and JSON Explore data, compute descriptive statistics, and create visualizations Manage your data science workflow using Drake Create reusable tools from one-liners and existing Python or R code Parallelize and distribute data-intensive pipelines using GNU Parallel Model data with dimensionality reduction, clustering, regression, and classification algorithms

The world of medical technologies is undergoing a sea change in the domain of consumer culture. Having a grasp on what appeals to consumers and how consumers are making purchasing decisions is essential to the success of any organization that thrives by offering a product or service. As such, it is vital to examine the consumer-centered aspects of medical technological developments that have a patient-centered focus and allow patients to take part in their own personal health and wellness.

Consumer-Driven Technologies in Healthcare: Breakthroughs in Research and Practice is a critical source of academic knowledge on the use of smartphones and other technological devices for cancer therapy, fitness and wellness, chronic disease monitoring, and other areas. The tracking of these items using technology has allowed consumers to take control of their own healthcare. Highlighting a range of pertinent topics such as clinical decision support systems, patient engagement, and electronic health records, this publication is an ideal reference source for doctors, nurse practitioners, hospital administrators, medical professionals, IT professionals, academicians, and researchers interested in advancing medical practice through technology.