

Instant Mapreduce Patterns Hadoop Essentials How To Perera Srinath

Right here, we have countless book **Instant Mapreduce Patterns Hadoop Essentials How To Perera Srinath** and collections to check out. We additionally manage to pay for variant types and then type of the books to browse. The all right book, fiction, history, novel, scientific research, as competently as various extra sorts of books are readily clear here.

As this Instant Mapreduce Patterns Hadoop Essentials How To Perera Srinath , it ends taking place physical one of the favored ebook Instant Mapreduce Patterns Hadoop Essentials How To Perera Srinath collections that we have. This is why you remain in the best website to look the amazing book to have.

Mastering Tableau - David Baldwin 2016-12-06
Master the intricacies of Tableau to create effective data visualizations About This Book
Arm yourself with an arsenal of advanced chart types and geocoding to efficiently and

engagingly present information Map a grid over a network node diagram and use that grid to demonstrate loads, processing time, and more in Tableau Integrate R with Tableau by utilizing R functions, libraries, and saved models Who This

Book Is For If you are a business analyst without developer-level programming skills, then this book is for you. You are expected to have at least a fundamental understanding of Tableau and basic knowledge of joins, however SQL knowledge is not assumed. You should have basic computer skills, including at least moderate Excel proficiency. What You Will Learn Create a worksheet that can display the current balance for any given period in time Recreate a star schema from in a data warehouse in Tableau Combine level of detail calculations with table calculations, sets, and parameters Create custom polygons to build filled maps for area codes in the USA Visualize data using a set of analytical and advanced charting techniques Know when to use Tableau instead of PowerPoint Build a dashboard and export it to PowerPoint In Detail Tableau has emerged as one of the most popular Business Intelligence solutions in recent times, thanks to its powerful and interactive data visualization capabilities.

This book will empower you to become a master in Tableau by exploiting the many new features introduced in Tableau 10.0. You will embark on this exciting journey by getting to know the valuable methods of utilizing advanced calculations to solve complex problems. These techniques include creative use of different types of calculations such as row-level, aggregate-level, and more. You will discover how almost any data visualization challenge can be met in Tableau by getting a proper understanding of the tool's inner workings and creatively exploring possibilities. You'll be armed with an arsenal of advanced chart types and techniques to enable you to efficiently and engagingly present information to a variety of audiences through the use of clear, efficient, and engaging dashboards. Explanations and examples of efficient and inefficient visualization techniques, well-designed and poorly designed dashboards, and compromise options when Tableau consumers will not embrace data

visualization will build on your understanding of Tableau and how to use it efficiently. By the end of the book, you will be equipped with all the information you need to create effective dashboards and data visualization solutions using Tableau. Style and approach This book takes a direct approach, to systematically evolve to more involved functionalities such as advanced calculation, parameters & sets, data blending and R integration. This book will help you gain skill in building visualizations previously beyond your capacity.

Hadoop in Action - Chuck Lam 2010-11-30
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.

Kafka: The Definitive Guide - Neha Narkhede
2017-08-31

Every enterprise application creates data, whether it's log messages, metrics, user activity, outgoing messages, or something else. And how to move all of this data becomes nearly as important as the data itself. If you're an application architect, developer, or production engineer new to Apache Kafka, this practical guide shows you how to use this open source streaming platform to handle real-time data feeds. Engineers from Confluent and LinkedIn who are responsible for developing Kafka explain how to deploy production Kafka clusters, write reliable event-driven microservices, and build scalable stream-processing applications with this platform. Through detailed examples,

you'll learn Kafka's design principles, reliability guarantees, key APIs, and architecture details, including the replication protocol, the controller, and the storage layer. Understand publish-subscribe messaging and how it fits in the big data ecosystem. Explore Kafka producers and consumers for writing and reading messages Understand Kafka patterns and use-case requirements to ensure reliable data delivery Get best practices for building data pipelines and applications with Kafka Manage Kafka in production, and learn to perform monitoring, tuning, and maintenance tasks Learn the most critical metrics among Kafka's operational measurements Explore how Kafka's stream delivery capabilities make it a perfect source for stream processing systems

HBase - Lars George 2011-09-05

"HBase: The Definitive Guide" provides the details for evaluating this high-performance, non-relational database, or putting it into practice right away. HBase's adoption rate is

beginning to climb, and IT executives are asking pointed questions about this high-capacity database.

Hadoop 2 Quick-Start Guide - Douglas

Eadline 2015-10-28

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

Blockchain Quick Reference - Brenn Hill

2018-08-10

Understand the Blockchain revolution and get to grips with Ethereum, Hyperledger Fabric, and Corda. Key Features
Resolve common challenges and problems faced in the Blockchain domain
Study architecture, concepts, terminologies, and Dapps
Make smart choices using Blockchain for personal and business investments
Book Description
Blockchain Quick Reference takes you through the electrifying world of blockchain technology and is designed for those who want

to polish their existing knowledge regarding the various pillars of the blockchain ecosystem. This book is your go-to guide, teaching you how to apply principles and ideas for making your life and business better. You will cover the architecture, Initial Coin Offerings (ICOs), tokens, smart contracts, and terminologies of the blockchain technology, before studying how they work. All you need is a curious mind to get started with blockchain technology. Once you have grasped the basics, you will explore components of Ethereum, such as ether tokens, transactions, and smart contracts, in order to build simple Dapps. You will then move on to learning why Solidity is used specifically for Ethereum-based projects, followed by exploring different types of blockchain with easy-to-follow examples. All this will help you tackle challenges and problems. By the end of this book, you will not only have solved current and future problems relating to blockchain technology but will also be able to build efficient decentralized

applications. What you will learn Understand how blockchain architecture components work Acquaint yourself with cryptography and the mechanics behind blockchain Apply consensus protocol to determine the business sustainability Understand what ICOs and crypto-mining are and how they work Create cryptocurrency wallets and coins for transaction mechanisms Understand the use of Ethereum for smart contract and DApp development Who this book is for Blockchain Quick Reference is for you if you are a developer who wants to get well-versed with blockchain and its associated concepts and terminologies. You will explore the working mechanism of a decentralized application with the help of examples. Business leaders and blockchain enthusiasts will also find this book useful, as it will help you effectively address challenges and make better personal and business investments.

Hadoop: The Definitive Guide - Tom White
2012-05-10

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

Data Structures And Algorithms - Shi-kuo Chang 2003-09-29

This is an excellent, up-to-date and easy-to-use text on data structures and algorithms that is intended for undergraduates in computer science and information science. The thirteen chapters, written by an international group of experienced teachers, cover the fundamental concepts of algorithms and most of the important data structures as well as the concept of interface design. The book contains many examples and diagrams. Whenever appropriate, program codes are included to facilitate learning. This book is supported by an international group of authors who are experts

on data structures and algorithms, through its website at www.cs.pitt.edu/~jung/GrowingBook/, so that both teachers and students can benefit from their expertise.

Python for Finance - Yuxing Yan 2014-04-25
A hands-on guide with easy-to-follow examples to help you learn about option theory, quantitative finance, financial modeling, and time series using Python. Python for Finance is perfect for graduate students, practitioners, and application developers who wish to learn how to utilize Python to handle their financial needs. Basic knowledge of Python will be helpful but knowledge of programming is necessary.

[Troubleshooting Ubuntu Server](#) - Skanda Bhargav 2015-09-25
Make life at the office easier for server administrators by helping them build resilient Ubuntu server systems About This Book Tackle the issues you come across in keeping your Ubuntu server up and running Build server

machines and troubleshoot cloud computing related issues using Open Stack Discover tips and best practices to be followed for minimum maintenance of Ubuntu Server 3 Who This Book Is For This book is for a vast audience of Linux system administrators who primarily work on Debian-based systems and spend long hours trying fix issues with the enterprise server. Ubuntu is already one of the most popular OSES and this book targets the most common issues that most administrators have to deal with. With the right tools and definite solutions, you will be able to keep your Ubuntu servers in the pink of health. What You Will Learn Deploy packages and their dependencies with repositories Set up your own DNS and network for Ubuntu Server Authenticate and validate users and their access to various systems and services Maintain, monitor, and optimize your server resources and avoid tremendous load Get to know about processes, assigning and changing priorities, and running processes in background Optimize

your shell with tools and provide users with an improved shell experience Set up separate environments for various services and run them safely in isolation Understand, build, and deploy OpenStack on your Ubuntu Server In Detail Ubuntu is becoming one of the favorite Linux flavors for many enterprises and is being adopted to a large extent. It supports a wide variety of common network systems and the use of standard Internet services including file serving, e-mail, Web, DNS, and database management. A large scale use and implementation of Ubuntu on servers has given rise to a vast army of Linux administrators who battle it out day in and day out to make sure the systems are in the right frame of operation and pre-empt any untoward incidents that may result in catastrophes for the businesses using it. Despite all these efforts, glitches and bugs occur that affect Ubuntu server's network, memory, application, and hardware and also generate cloud computing related issues using

OpenStack. This book will help you end to end. Right from setting up your new Ubuntu Server to learning the best practices to host OpenStack without any hassles. You will be able to control the priority of jobs, restrict or allow access users to certain services, deploy packages, tackle issues related to server effectively, and reduce downtime. Also, you will learn to set up OpenStack, and manage and monitor its services while tuning the machine with best practices. You will also get to know about Virtualization to make services serve users better. Chapter by chapter, you will learn to add new features and functionalities and make your Ubuntu server a full-fledged, production-ready system. Style and approach This book contains topic-by-topic discussion in an easy-to-understand language with loads of examples to help you take care of Ubuntu Server. Plenty of screenshots will guide you through a step-by-step approach.

MapReduce Design Patterns - Donald Miner
2012-11-21

Until now, design patterns for the MapReduce framework have been scattered among various research papers, blogs, and books. This handy guide brings together a unique collection of valuable MapReduce patterns that will save you time and effort regardless of the domain, language, or development framework you're using. Each pattern is explained in context, with pitfalls and caveats clearly identified to help you avoid common design mistakes when modeling your big data architecture. This book also provides a complete overview of MapReduce that explains its origins and implementations, and why design patterns are so important. All code examples are written for Hadoop.

Summarization patterns: get a top-level view by summarizing and grouping data
Filtering patterns: view data subsets such as records generated from one user
Data organization patterns: reorganize data to work with other systems, or to make MapReduce analysis easier
Join patterns: analyze different datasets together

to discover interesting relationships
Metapatterns: piece together several patterns to solve multi-stage problems, or to perform several analytics in the same job
Input and output patterns: customize the way you use Hadoop to load or store data
"A clear exposition of MapReduce programs for common data processing patterns—this book is indispensable for anyone using Hadoop." --Tom White, author of Hadoop: The Definitive Guide

Trino: The Definitive Guide - Matt Fuller

2021-04-14

Perform fast interactive analytics against different data sources using the Trino high-performance distributed SQL query engine. With this practical guide, you'll learn how to conduct analytics on data where it lives, whether it's Hive, Cassandra, a relational database, or a proprietary data store. Analysts, software engineers, and production engineers will learn how to manage, use, and even develop with Trino. Initially developed by Facebook, open

source Trino is now used by Netflix, Airbnb, LinkedIn, Twitter, Uber, and many other companies. Matt Fuller, Manfred Moser, and Martin Traverso show you how a single Trino query can combine data from multiple sources to allow for analytics across your entire organization. Get started: Explore Trino's use cases and learn about tools that will help you connect to Trino and query data
Go deeper: Learn Trino's internal workings, including how to connect to and query data sources with support for SQL statements, operators, functions, and more
Put Trino in production: Secure Trino, monitor workloads, tune queries, and connect more applications; learn how other organizations apply Trino

Instant MapReduce Patterns - Hadoop Essentials How-to - Srinath Perera 2013

"MapReduce is a technology that enables users to process large datasets and Hadoop is an implementation of MapReduce." This book "is a concise introduction to Hadoop and

programming with MapReduce. It is aimed to get you started and give you an overall feel for programming with Hadoop providing you with a well-grounded foundation to understand and solve all of your MapReduce problems as needed"--Cover.

Practical Data Science with Hadoop and Spark -
Ofer Mendeleevitch 2016-12-08

The Complete Guide to Data Science with Hadoop—For Technical Professionals, Businesspeople, and Students Demand is soaring for professionals who can solve real data science problems with Hadoop and Spark. Practical Data Science with Hadoop® and Spark is your complete guide to doing just that. Drawing on immense experience with Hadoop and big data, three leading experts bring together everything you need: high-level concepts, deep-dive techniques, real-world use cases, practical applications, and hands-on tutorials. The authors introduce the essentials of data science and the modern Hadoop ecosystem, explaining how

Hadoop and Spark have evolved into an effective platform for solving data science problems at scale. In addition to comprehensive application coverage, the authors also provide useful guidance on the important steps of data ingestion, data munging, and visualization. Once the groundwork is in place, the authors focus on specific applications, including machine learning, predictive modeling for sentiment analysis, clustering for document analysis, anomaly detection, and natural language processing (NLP). This guide provides a strong technical foundation for those who want to do practical data science, and also presents business-driven guidance on how to apply Hadoop and Spark to optimize ROI of data science initiatives. Learn What data science is, how it has evolved, and how to plan a data science career How data volume, variety, and velocity shape data science use cases Hadoop and its ecosystem, including HDFS, MapReduce, YARN, and Spark Data importation with Hive

and Spark Data quality, preprocessing, preparation, and modeling Visualization: surfacing insights from huge data sets Machine learning: classification, regression, clustering, and anomaly detection Algorithms and Hadoop tools for predictive modeling Cluster analysis and similarity functions Large-scale anomaly detection NLP: applying data science to human language

Briggs - Barry Briggs 2016-01-07

How do you start? How should you build a plan for cloud migration for your entire portfolio? How will your organization be affected by these changes? This book, based on real-world cloud experiences by enterprise IT teams, seeks to provide the answers to these questions. Here, you'll see what makes the cloud so compelling to enterprises; with which applications you should start your cloud journey; how your organization will change, and how skill sets will evolve; how to measure progress; how to think about security, compliance, and business buy-in; and

how to exploit the ever-growing feature set that the cloud offers to gain strategic and competitive advantage.

Beginning Apache Pig - Balaswamy Vaddeman 2016-12-10

Learn to use Apache Pig to develop lightweight big data applications easily and quickly. This book shows you many optimization techniques and covers every context where Pig is used in big data analytics. Beginning Apache Pig shows you how Pig is easy to learn and requires relatively little time to develop big data applications. The book is divided into four parts: the complete features of Apache Pig; integration with other tools; how to solve complex business problems; and optimization of tools. You'll discover topics such as MapReduce and why it cannot meet every business need; the features of Pig Latin such as data types for each load, store, joins, groups, and ordering; how Pig workflows can be created; submitting Pig jobs using Hue; and working with Oozie. You'll also see how to

extend the framework by writing UDFs and custom load, store, and filter functions. Finally you'll cover different optimization techniques such as gathering statistics about a Pig script, joining strategies, parallelism, and the role of data formats in good performance. What You Will Learn • Use all the features of Apache Pig • Integrate Apache Pig with other tools • Extend Apache Pig • Optimize Pig Latin code • Solve different use cases for Pig Latin Who This Book Is For All levels of IT professionals: architects, big data enthusiasts, engineers, developers, and big data administrators

Mastering Cloud Computing - Rajkumar Buyya
2013-04-05

Mastering Cloud Computing is designed for undergraduate students learning to develop cloud computing applications. Tomorrow's applications won't live on a single computer but will be deployed from and reside on a virtual server, accessible anywhere, any time. Tomorrow's application developers need to

understand the requirements of building apps for these virtual systems, including concurrent programming, high-performance computing, and data-intensive systems. The book introduces the principles of distributed and parallel computing underlying cloud architectures and specifically focuses on virtualization, thread programming, task programming, and map-reduce programming. There are examples demonstrating all of these and more, with exercises and labs throughout. Explains how to make design choices and tradeoffs to consider when building applications to run in a virtual cloud environment Real-world case studies include scientific, business, and energy-efficiency considerations

NoSQL Distilled - Pramod J. Sadalage 2013
'NoSQL Distilled' is designed to provide you with enough background on how NoSQL databases work, so that you can choose the right data store without having to trawl the whole web to do it. It won't answer your questions definitively, but it

should narrow down the range of options you have to consider.

Cloudera Administration Handbook - Rohit Menon 2014-07-18

An easy-to-follow Apache Hadoop administrator's guide filled with practical screenshots and explanations for each step and configuration. This book is great for administrators interested in setting up and managing a large Hadoop cluster. If you are an administrator, or want to be an administrator, and you are ready to build and maintain a production-level cluster running CDH5, then this book is for you.

Hadoop Essentials - Shiva Achari 2015-04-29

If you are a system or application developer interested in learning how to solve practical problems using the Hadoop framework, then this book is ideal for you. This book is also meant for Hadoop professionals who want to find solutions to the different challenges they come across in their Hadoop projects.

Big Data Processing with Apache Spark - Srin Penchikala 2018-03-13

Apache Spark is a popular open-source big-data processing framework that's built around speed, ease of use, and unified distributed computing architecture. Not only it supports developing applications in different languages like Java, Scala, Python, and R, it's also hundred times faster in memory and ten times faster even when running on disk compared to traditional data processing frameworks. Whether you are currently working on a big data project or interested in learning more about topics like machine learning, streaming data processing, and graph data analytics, this book is for you. You can learn about Apache Spark and develop Spark programs for various use cases in big data analytics using the code examples provided. This book covers all the libraries in Spark ecosystem: Spark Core, Spark SQL, Spark Streaming, Spark ML, and Spark GraphX.

Instant Mapreduce Patterns - Hadoop Essentials

How-To - Srinath Perera 2013-05-22

Filled with practical, step-by-step instructions and clear explanations for the most important and useful tasks. This is a Packt Instant How-to guide, which provides concise and clear recipes for getting started with Hadoop. This book is for big data enthusiasts and would-be Hadoop programmers. It is also meant for Java programmers who either have not worked with Hadoop at all, or who know Hadoop and MapReduce but are not sure how to deepen their understanding.

Hadoop in Practice - Alex Holmes 2014-09-29

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

Practical Data Analysis Cookbook - Tomasz Drabas 2016-04-29

Over 60 practical recipes on data exploration and analysis About This Book Clean dirty data, extract accurate information, and explore the relationships between variables Forecast the output of an electric plant and the water flow of American rivers using pandas, NumPy, Statsmodels, and scikit-learn Find and extract the most important features from your dataset using the most efficient Python libraries Who This Book Is For If you are a beginner or intermediate-level professional who is looking to solve your day-to-day, analytical problems with Python, this book is for you. Even with no prior programming and data analytics experience, you will be able to finish each recipe and learn while doing so. What You Will Learn Read, clean, transform, and store your data using Pandas and OpenRefine Understand your data and explore the relationships between variables using Pandas and D3.js Explore a variety of techniques

to classify and cluster outbound marketing campaign calls data of a bank using Pandas, mlp, NumPy, and Statsmodels Reduce the dimensionality of your dataset and extract the most important features with pandas, NumPy, and mlp Predict the output of a power plant with regression models and forecast water flow of American rivers with time series methods using pandas, NumPy, Statsmodels, and scikit-learn Explore social interactions and identify fraudulent activities with graph theory concepts using NetworkX and Gephi Scrape Internet web pages using urllib and BeautifulSoup and get to know natural language processing techniques to classify movies ratings using NLTK Study simulation techniques in an example of a gas station with agent-based modeling In Detail Data analysis is the process of systematically applying statistical and logical techniques to describe and illustrate, condense and recap, and evaluate data. Its importance has been most visible in the sector of information and communication

technologies. It is an employee asset in almost all economy sectors. This book provides a rich set of independent recipes that dive into the world of data analytics and modeling using a variety of approaches, tools, and algorithms. You will learn the basics of data handling and modeling, and will build your skills gradually toward more advanced topics such as simulations, raw text processing, social interactions analysis, and more. First, you will learn some easy-to-follow practical techniques on how to read, write, clean, reformat, explore, and understand your data—arguably the most time-consuming (and the most important) tasks for any data scientist. In the second section, different independent recipes delve into intermediate topics such as classification, clustering, predicting, and more. With the help of these easy-to-follow recipes, you will also learn techniques that can easily be expanded to solve other real-life problems such as building recommendation engines or predictive models.

In the third section, you will explore more advanced topics: from the field of graph theory through natural language processing, discrete choice modeling to simulations. You will also get to expand your knowledge on identifying fraud origin with the help of a graph, scrape Internet websites, and classify movies based on their reviews. By the end of this book, you will be able to efficiently use the vast array of tools that the Python environment has to offer. Style and approach This hands-on recipe guide is divided into three sections that tackle and overcome real-world data modeling problems faced by data analysts/scientist in their everyday work. Each independent recipe is written in an easy-to-follow and step-by-step fashion.

Ethereum Cookbook - Manoj P R 2018-08-31
Mine Ether, deploy smart contracts, tokens, and ICOs, and manage security vulnerabilities of Ethereum Key Features Build end-to-end decentralized Ethereum apps using Truffle, Web3, and Solidity Explore various solution-

based recipes to build smart contracts and foolproof decentralized applications Develop decentralized marketplaces from scratch, build wallets, and manage transactions Book Description Ethereum and Blockchain will change the way software is built for business transactions. Most industries have been looking to leverage these new technologies to gain efficiencies and create new business models and opportunities. The Ethereum Cookbook covers various solutions such as setting up Ethereum, writing smart contracts, and creating tokens, among others. You'll learn about the security vulnerabilities, along with other protocols of Ethereum. Once you have understood the basics, you'll move on to exploring various design decisions and tips to make your application scalable and secure. In addition to this, you'll work with various Ethereum packages such as Truffle, Web3, and Ganache. By the end of this book, you'll have comprehensively grasped the Ethereum principles and ecosystem. What you

will learn Efficiently write smart contracts in Ethereum Build scalable distributed applications and deploy them Use tools and frameworks to develop, deploy, and test your application Use block explorers such as Etherscan to find a specific transaction Create your own tokens, initial coin offerings (ICOs), and games Understand various security flaws in smart contracts in order to avoid them Who this book is for The Ethereum Cookbook is for you if you are a software engineer, Blockchain developer, or research scientist who wants to build smart contracts, develop decentralized applications, and facilitate peer-to-peer transaction. It is assumed that you are familiar with Blockchain concepts and have sound knowledge of JavaScript.

Expert Hadoop 2 Administration - Sam R. Alapati 2016-11-29

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

Tableau Cookbook - Recipes for Data Visualization - Shweta Sankhe-Savale

2016-12-26

Create beautiful data visualizations and interactive dashboards with Tableau About This Book Delve into the features and functionalities of Tableau from the ground up with this step-by-step guide that has over 50 "follow-me" recipes Build rich visualizations to effectively highlight the underlying trends and patterns in your data Build beautiful interactive dashboards and storyboards to stitch your visualizations together and tell a story Who This Book Is For This book is for anyone who wishes to use Tableau. It will be of use to both beginners who want to learn Tableau from scratch and to more seasoned users who simply want a quick reference guide. This book is a ready reckoner guide for you. The book will be such that both new & existing Tableau users who don't know, or can't recall how to perform different Tableau tasks can use the book and be benefited from it. What You Will Learn Get to grips with the Tableau workspace and terminologies and understand what data

sources you can connect Learn to create basic charts like bar chart, stacked bar, pie chart, line chart, area chart, tree map & word cloud Go even further with more advanced visualizations such as scatter plot, box & whiskers plot, dual axis, bullet chart, Histograms, Maps, etc Use pre-defined calculation and change its scope and direction to affect outcome Learn to define Parameters and call them into parametric calculations that provide outcomes based on user inputs Build Dashboards and use Actions to link multiple sheets on the dashboard Connect to multiple data sources using Data Blending, Multiple Table Join within the same data source as well as across data sources, Custom SQL and learn to work with data Extracts Compute statistical trends, build forecasting models and use Reference lines for benchmarking In Detail Data is everywhere and everything is data! Visualization of data allows us to bring out the underlying trends and patterns inherent in the data and gain insights that enable faster and

smarter decision making. Tableau is one of the fastest growing and industry leading Business Intelligence platforms that empowers business users to easily visualize their data and discover insights at the speed of thought. Tableau is a self-service BI platform designed to make data visualization and analysis as intuitive as possible. Creating visualizations with simple drag-and-drop, you can be up and running on Tableau in no time. Starting from the fundamentals such as getting familiarized with Tableau Desktop, connecting to common data sources and building standard charts; you will walk through the nitty gritty of Tableau such as creating dynamic analytics with parameters, blended data sources, and advanced calculations. You will also learn to group members into higher levels, sort the data in a specific order & filter out the unnecessary information. You will then create calculations in Tableau & understand the flexibility & power they have and go on to building story-boards and

share your insights with others. Whether you are just getting started or whether you need a quick reference on a "how-to" question, This book is the perfect companion for you Style and approach This cookbook takes a step-by-step approach and the text systematically evolves to cover more involved functionalities. Every recipe includes illustrative screenshots which provide a detailed visual resource for each step.

Data Analysis and Business Modeling with Excel 2013 - David Rojas 2015-10-27

Manage, analyze, and visualize data with Microsoft Excel 2013 to transform raw data into ready to use information About This Book Create formulas to help you analyze and explain findings Develop interactive spreadsheets that will impress your audience and give them the ability to slice and dice data A step-by-step guide to learn various ways to model data for businesses with the help of Excel 2013 Who This Book Is For If you want to start using Excel 2013 for data analysis and business modeling and

enhance your skills in the data analysis life cycle then this book is for you, whether you're new to Excel or experienced. What You Will Learn Discover what Excel formulas are all about and how to use them in your spreadsheet development Identify bad data and learn cleaning strategies Create interactive spreadsheets that engage and appeal to your audience Leverage Excel's powerful built-in tools to get the median, maximum, and minimum values of your data Build impressive tables and combine datasets using Excel's built-in functionality Learn the powerful scripting language VBA, allowing you to implement your own custom solutions with ease In Detail Excel 2013 is one of the easiest to use data analysis tools you will ever come across. Its simplicity and powerful features has made it the go to tool for all your data needs. Complex operations with Excel, such as creating charts and graphs, visualization, and analyzing data make it a great tool for managers, data scientists, financial data

analysts, and those who work closely with data. Learning data analysis and will help you bring your data skills to the next level. This book starts by walking you through creating your own data and bringing data into Excel from various sources. You'll learn the basics of SQL syntax and how to connect it to a Microsoft SQL Server Database using Excel's data connection tools. You will discover how to spot bad data and strategies to clean that data to make it useful to you. Next, you'll learn to create custom columns, identify key metrics, and make decisions based on business rules. You'll create macros using VBA and use Excel 2013's shiny new macros. Finally, at the end of the book, you'll be provided with useful shortcuts and tips, enabling you to do efficient data analysis and business modeling with Excel 2013. Style and approach This is a step-by-step guide to performing data analysis and business modelling with Excel 2013, complete with examples and tips.

Programming Hive - Edward Capriolo

2012-09-26

Describes the features and functions of Apache Hive, the data infrastructure for Hadoop.

[Apache Mahout Essentials](#) - Jayani Withanawasam 2015-06-19

Apache Mahout is a scalable machine learning library with algorithms for clustering, classification, and recommendations. It empowers users to analyze patterns in large, diverse, and complex datasets faster and more scalably. This book is an all-inclusive guide to analyzing large and complex datasets using Apache Mahout. It explains complicated but very effective machine learning algorithms simply, in relation to real-world practical examples. Starting from the fundamental concepts of machine learning and Apache Mahout, this book guides you through Apache Mahout's implementations of machine learning techniques including classification, clustering, and recommendations. During this exciting walkthrough, real-world applications, a diverse

range of popular algorithms and their implementations, code examples, evaluation strategies, and best practices are given for each technique. Finally, you will learn vdata visualization techniques for Apache Mahout to bring your data to life.

Machine Learning Models and Algorithms for Big Data Classification - Shan Suthaharan 2015-10-20

This book presents machine learning models and algorithms to address big data classification problems. Existing machine learning techniques like the decision tree (a hierarchical approach), random forest (an ensemble hierarchical approach), and deep learning (a layered approach) are highly suitable for the system that can handle such problems. This book helps readers, especially students and newcomers to the field of big data and machine learning, to gain a quick understanding of the techniques and technologies; therefore, the theory, examples, and programs (Matlab and R)

presented in this book have been simplified, hardcoded, repeated, or spaced for improvements. They provide vehicles to test and understand the complicated concepts of various topics in the field. It is expected that the readers adopt these programs to experiment with the examples, and then modify or write their own programs toward advancing their knowledge for solving more complex and challenging problems. The presentation format of this book focuses on simplicity, readability, and dependability so that both undergraduate and graduate students as well as new researchers, developers, and practitioners in this field can easily trust and grasp the concepts, and learn them effectively. It has been written to reduce the mathematical complexity and help the vast majority of readers to understand the topics and get interested in the field. This book consists of four parts, with the total of 14 chapters. The first part mainly focuses on the topics that are needed to help analyze and understand data and big data. The

second part covers the topics that can explain the systems required for processing big data. The third part presents the topics required to understand and select machine learning techniques to classify big data. Finally, the fourth part concentrates on the topics that explain the scaling-up machine learning, an important solution for modern big data problems.

Hadoop MapReduce Cookbook - Srinath Perera
2013-01-01

Individual self-contained code recipes. Solve specific problems using individual recipes, or work through the book to develop your capabilities. If you are a big data enthusiast and striving to use Hadoop to solve your problems, this book is for you. Aimed at Java programmers with some knowledge of Hadoop MapReduce, this is also a comprehensive reference for developers and system admins who want to get up to speed using Hadoop.

Big Data Analytics with R and Hadoop - Vignesh

Prajapati 2013-11-25

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

Kafka Streams in Action - Bill Bejeck 2018-08-29
Summary Kafka Streams in Action teaches you everything you need to know to implement stream processing on data flowing into your Kafka platform, allowing you to focus on getting more from your data without sacrificing time or effort. Foreword by Neha Narkhede, Cocreator of Apache Kafka Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Not all stream-based applications

require a dedicated processing cluster. The lightweight Kafka Streams library provides exactly the power and simplicity you need for message handling in microservices and real-time event processing. With the Kafka Streams API, you filter and transform data streams with just Kafka and your application. About the Book Kafka Streams in Action teaches you to implement stream processing within the Kafka platform. In this easy-to-follow book, you'll explore real-world examples to collect, transform, and aggregate data, work with multiple processors, and handle real-time events. You'll even dive into streaming SQL with KSQL! Practical to the very end, it finishes with testing and operational aspects, such as monitoring and debugging. What's inside Using the KStreams API Filtering, transforming, and splitting data Working with the Processor API Integrating with external systems About the Reader Assumes some experience with distributed systems. No knowledge of Kafka or

streaming applications required. About the Author Bill Bejeck is a Kafka Streams contributor and Confluent engineer with over 15 years of software development experience. Table of Contents PART 1 - GETTING STARTED WITH KAFKA STREAMS Welcome to Kafka Streams Kafka quicklyPART 2 - KAFKA STREAMS DEVELOPMENT Developing Kafka Streams Streams and state The KTable API The Processor APIPART 3 - ADMINISTERING KAFKA STREAMS Monitoring and performance Testing a Kafka Streams applicationPART 4 - ADVANCED CONCEPTS WITH KAFKA STREAMS Advanced applications with Kafka StreamsAPPENDIXES Appendix A - Additional configuration information Appendix B - Exactly once semantics
Deep Learning with Hadoop - Dipayan Dev
2017-02-20
Build, implement and scale distributed deep learning models for large-scale datasets About This Book Get to grips with the deep learning

concepts and set up Hadoop to put them to use
Implement and parallelize deep learning models
on Hadoop's YARN framework A comprehensive
tutorial to distributed deep learning with
Hadoop Who This Book Is For If you are a data
scientist who wants to learn how to perform
deep learning on Hadoop, this is the book for
you. Knowledge of the basic machine learning
concepts and some understanding of Hadoop is
required to make the best use of this book. What
You Will Learn Explore Deep Learning and
various models associated with it Understand
the challenges of implementing distributed deep
learning with Hadoop and how to overcome it
Implement Convolutional Neural Network (CNN)
with deeplearning4j Delve into the
implementation of Restricted Boltzmann
Machines (RBM) Understand the mathematical
explanation for implementing Recurrent Neural
Networks (RNN) Get hands on practice of deep
learning and their implementation with Hadoop.
In Detail This book will teach you how to deploy

large-scale dataset in deep neural networks with
Hadoop for optimal performance. Starting with
understanding what deep learning is, and what
the various models associated with deep neural
networks are, this book will then show you how
to set up the Hadoop environment for deep
learning. In this book, you will also learn how to
overcome the challenges that you face while
implementing distributed deep learning with
large-scale unstructured datasets. The book will
also show you how you can implement and
parallelize the widely used deep learning models
such as Deep Belief Networks, Convolutional
Neural Networks, Recurrent Neural Networks,
Restricted Boltzmann Machines and
autoencoder using the popular deep learning
library deeplearning4j. Get in-depth
mathematical explanations and visual
representations to help you understand the
design and implementations of Recurrent Neural
network and Denoising AutoEncoders with
deeplearning4j. To give you a more practical

perspective, the book will also teach you the implementation of large-scale video processing, image processing and natural language processing on Hadoop. By the end of this book, you will know how to deploy various deep neural networks in distributed systems using Hadoop. Style and approach This book takes a comprehensive, step-by-step approach to implement efficient deep learning models on Hadoop. It starts from the basics and builds the readers' knowledge as they strengthen their understanding of the concepts. Practical examples are included in every step of the way to supplement the theory.

HBase in Action - Amandeep Khurana

2012-11-01

Summary HBase in Action has all the knowledge you need to design, build, and run applications using HBase. First, it introduces you to the fundamentals of distributed systems and large scale data handling. Then, you'll explore real-world applications and code samples with just

enough theory to understand the practical techniques. You'll see how to build applications with HBase and take advantage of the MapReduce processing framework. And along the way you'll learn patterns and best practices. About the Technology HBase is a NoSQL storage system designed for fast, random access to large volumes of data. It runs on commodity hardware and scales smoothly from modest datasets to billions of rows and millions of columns. About this Book HBase in Action is an experience-driven guide that shows you how to design, build, and run applications using HBase. First, it introduces you to the fundamentals of handling big data. Then, you'll explore HBase with the help of real applications and code samples and with just enough theory to back up the practical techniques. You'll take advantage of the MapReduce processing framework and benefit from seeing HBase best practices in action. 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. What's Inside When and how to use HBase Practical examples Design patterns for scalable data systems Deployment, integration, and design Written for developers and architects familiar with data storage and processing. No prior knowledge of HBase, Hadoop, or MapReduce is required. Table of Contents PART 1 HBASE FUNDAMENTALS Introducing HBase Getting started Distributed HBase, HDFS, and MapReduce PART 2 ADVANCED CONCEPTS HBase table design Extending HBase with coprocessors Alternative HBase clients PART 3 EXAMPLE APPLICATIONS HBase by example: OpenTSDB Scaling GIS on HBase PART 4 OPERATIONALIZING HBASE Deploying HBase Operations

Hadoop Operations - Eric Sammer 2012-09-26

If you've been asked to maintain large and complex Hadoop clusters, this book is a must. Demand for operations-specific material has skyrocketed now that Hadoop is becoming the

de facto standard for truly large-scale data processing in the data center. Eric Sammer, Principal Solution Architect at Cloudera, shows you the particulars of running Hadoop in production, from planning, installing, and configuring the system to providing ongoing maintenance. Rather than run through all possible scenarios, this pragmatic operations guide calls out what works, as demonstrated in critical deployments. Get a high-level overview of HDFS and MapReduce: why they exist and how they work Plan a Hadoop deployment, from hardware and OS selection to network requirements Learn setup and configuration details with a list of critical properties Manage resources by sharing a cluster across multiple groups Get a runbook of the most common cluster maintenance tasks Monitor Hadoop clusters—and learn troubleshooting with the help of real-world war stories Use basic tools and techniques to handle backup and catastrophic failure

Essentials of Business Analytics -

Bhimasankaram Pochiraju 2019-07-10

This comprehensive edited volume is the first of its kind, designed to serve as a textbook for long-duration business analytics programs. It can also be used as a guide to the field by practitioners. The book has contributions from experts in top universities and industry. The editors have taken extreme care to ensure continuity across the chapters. The material is organized into three parts: A) Tools, B) Models and C) Applications. In Part A, the tools used by business analysts are described in detail. In Part B, these tools are applied to construct models used to solve business problems. Part C contains detailed applications in various functional areas of business and several case studies. Supporting material can be found in the appendices that develop the pre-requisites for the main text. Every chapter has a business orientation. Typically, each chapter begins with the description of business problems that are

transformed into data questions; and methodology is developed to solve these questions. Data analysis is conducted using widely used software, the output and results are clearly explained at each stage of development. These are finally transformed into a business solution. The companion website provides examples, data sets and sample code for each chapter.

Data Analytics with Hadoop - Benjamin Bengfort 2016-06

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 *Hadoop MapReduce v2 Cookbook - Second Edition* - Thilina Gunarathne 2015-02-25 If you are a Big Data enthusiast and wish to use Hadoop v2 to solve your problems, then this book is for you. This book is for Java programmers with little to moderate knowledge of Hadoop MapReduce. This is also a one-stop reference for developers and system admins who want to quickly get up to speed with using Hadoop v2. It would be helpful to have a basic knowledge of software development using Java and a basic working knowledge of Linux.