OMB No. 7314159520362

Hadoop Real World Solutions Cookbook

Top 10 books To Learn Hadoop In 2021 | Best Books For Hadoop Beginners | Hadoop Training | Edureka Hadoop In 5 Minutes | What Is Hadoop? | Introduction To Hadoop | Hadoop Explained |Simplilearn Hadoop In Real World - Advanced Course Best Data Science Books for Beginners Big Data Engineering Full Course Part 1 | 17 Hours I've Read Over 100 Books on Python. Here are the Top 3 How I Became A Data Scientist (No CS Degree, No Bootcamp) Starting a Career in Data Science (10 Thing I Wish I Knew…) Books every software engineer should read in 2024. How to Become a Data Scientist in 2024? (complete roadmap) Apache Hadoop \u0026 Big Data 101: The Basics Hadoop Bigdata Real time Project Explanation What is Big Data? Introduction to Big Data | Hadoop Tutorial for Beginners | Hadoop [Part 1] 11 years later ❤ @shrads Hadoop Developer In Real World : Course Structure Don't make eye contact Hadoop Developer In Real World : Understanding Big Data Problem Data Engineering Cookbook 2022 Hadoop Developer In Real World : HDFS - Why Another Filesystem? How much does a LEAD ANALYST make? How much does a DATA ENGINEER make? I've read 40 programming books. Top 5 you must read. How much does a SOFTWARE ENGINEER make? Data Science Cook Book Collection for £12.99. A Great Humble Bundle

Teradata Cookbook
Hadoop Operations and Cluster Management Cookbook
Big Data Analytics with R and Hadoop
Hadoop Beginner's Guide
Hadoop Real-World Solutions Cookbook Second Edition
MapReduce Design Patterns
R Data Analysis Cookbook
Apache Hadoop 3 Quick Start Guide
Hadoop Operations
Applied Text Analysis with Python
Practical Data Science Cookbook
Hadoop: Data Processing and Modelling

Snowflake Cookbook
Data Algorithms
Clojure in Action
SQL Server 2017 Integration Services Cookbook
Hadoop in Action
Proceedings of ICETIT 2019
NoSQL Distilled
Hadoop Real World Solutions Cookbook
Mastering Hadoop 3
R Data Analysis Cookbook, Second Edition

CANTU WALLS

Teradata Cookbook "O'Reilly Media, Inc."
Over 90 hands-on recipes to design Internet scalable web and mobile applications with Amazon DynamoDB About This Book Construct top-notch mobile and web applications with the Internet scalable NoSQL database and host it on cloud Integrate your applications with other AWS services like AWS EMR, AWS S3, AWS Redshift, and AWS CloudSearch etc. in order to achieve a one-stop application stack Step-by-step implementation guide that provides real-world use with hands-on recipes Who This Book Is For This book is intended for those who have a basic understanding of AWS services and want to take their knowledge to the next level by getting their hands dirty with coding recipes in DynamoDB. What You Will Learn Design DynamoDB tables to achieve high read and write throughput Discover best practices like caching, exponential back-offs and auto-retries, storing large items in AWS S3, storing compressed data etc. Effectively use DynamoDB Local in order to make your development smooth and cost effective Implement cost effective best practices to reduce the burden of DynamoDB charges Create and maintain secondary indexes to support improved data access Integrate various other AWS services like AWS EMR, AWS CloudSearch, AWS Pipeline etc. with DynamoDB In Detail AWS DynamoDB is an excellent example of a production-ready NoSQL database. In recent years, DynamoDB has been able to attract many customers because of its features like high-availability, reliability and infinite scalability. DynamoDB can be easily integrated with massive data crunching tools like Hadoop /EMR, which is an essential part of this data-driven world and hence it is widely accepted. The cost and time-efficient design makes DynamoDB stand out amongst its peers. The design of DynamoDB is so neat and clean that it has inspired many NoSQL databases to simply follow it. This book will get your hands on some engineering best practices DynamoDB engineers

use, which can be used in your day-to-day life to build robust and scalable applications. You will start by operating with DynamoDB tables and learn to manipulate items and manage indexes. You will also discover how to easily integrate applications with other AWS services like EMR, S3, CloudSearch, RedShift etc. A couple of chapters talk in detail about how to use DynamoDB as a backend database and hosting it on AWS ElasticBean. This book will also focus on security measures of DynamoDB as well by providing techniques on data encryption, masking etc. By the end of the book you'll be adroit in designing web and mobile applications using DynamoDB and host it on cloud. Style and approach An easy-to-follow guide, full of real-world examples, which takes you through the world of DynamoDB following a step-by-step, problem-solution based approach.

Hadoop Operations and Cluster Management Cookbook

Packt Publishing Ltd

Over 90 hands-on recipes to help you learn and master the intricacies of Apache Hadoop 2.X, YARN, Hive, Pig, Oozie, Flume, Sqoop, Apache Spark, and MahoutAbout This Book- Implement outstanding Machine Learning use cases on your own analytics models and processes.- Solutions to common problems when working with the Hadoop ecosystem.- Step-by-step implementation of end-to-end big data use cases.Who This Book Is ForReaders who have a basic knowledge of big data systems and want to advance their knowledge with hands-on recipes.What You Will Learn- Installing and maintaining Hadoop 2.X cluster and its ecosystem.- Write advanced Map Reduce programs and understand design patterns.- Advanced Data Analysis using the Hive, Pig, and Map Reduce programs.- Import and export data from various sources using Sqoop and Flume.- Data storage in various file formats such as Text, Sequential, Parquet, ORC, and RC Files.- Machine learning principles with libraries such as Mahout- Batch and Stream data processing using Apache SparkIn DetailBig data is the current requirement. Most organizations produce huge amount of data every day. With the arrival of Hadoop-like tools, it has become easier for everyone to solve big data problems with great efficiency and at minimal cost. Grasping Machine Learning techniques will help you greatly in building predictive models and using this data to make the right decisions for your organization.Hadoop Real World Solutions Cookbook gives readers insights into learning and mastering big data via recipes. The book not only clarifies most big data tools in the market but also provides best practices for using them. The book provides recipes that are based on the latest versions of Apache Hadoop 2.X, YARN, Hive, Pig, Sqoop, Flume, Apache Spark, Mahout and many more such ecosystem tools. This real-world-solution cookbook is packed with handy recipes you can apply to your own everyday issues. Each chapter provides in-depth recipes that can be referenced easily. This book provides detailed practices on the latest technologies such as YARN and Apache Spark. Readers will be able to consider themselves as big data experts on completion of this book.This guide is an invaluable tutorial if you are planning to implement a big data warehouse for your business.Style and approachAn easy-to-follow guide that walks you through world of big data. Each tool in the Hadoop ecosystem is explained in detail and the recipes are placed in such a manner that readers can implement them sequentially. Plenty of reference links are provided for advanced reading.

Big Data Analytics with R and Hadoop Packt Publishing Ltd
Get to grips with building and productionizing end-to-end big data solutions in Azure and learn best practices for working with large datasets Key FeaturesIntegrate with Azure Synapse Analytics, Cosmos DB, and Azure HDInsight Kafka Cluster to scale and analyze your projects and build pipelinesUse Databricks SQL to run ad hoc queries on your data lake and create dashboardsProductionize a solution using CI/CD for deploying notebooks and Azure Databricks Service to various environmentsBook Description Azure Databricks is a unified collaborative platform for performing scalable analytics in an interactive environment. The Azure Databricks Cookbook provides recipes to get hands-on with the analytics process, including ingesting data from various batch and streaming sources and building a modern data warehouse. The book starts by teaching you how to create an Azure Databricks instance within the Azure portal, Azure CLI, and ARM templates. You'll work through clusters in Databricks and explore recipes for ingesting data from sources, including files, databases, and streaming sources such as Apache Kafka and EventHub. The book will help you explore all the features supported by Azure Databricks for building powerful end-to-end data pipelines. You'll also find out how to build a modern data warehouse by using Delta tables and Azure Synapse Analytics. Later, you'll learn how to write ad hoc queries and extract meaningful insights from the data lake by creating visualizations and dashboards with Databricks SQL. Finally, you'll deploy and productionize a data pipeline as well as deploy notebooks and Azure Databricks service using continuous integration and continuous delivery (CI/CD). By the end of this Azure book, you'll be able to use Azure Databricks to streamline different processes involved in building data-driven apps. What you will learnRead and write data from and to various Azure resources and file formatsBuild a modern data warehouse with Delta Tables and Azure Synapse AnalyticsExplore jobs, stages, and tasks and see how Spark lazy evaluation worksHandle concurrent transactions and learn performance optimization in Delta tablesLearn Databricks SQL and create real-time dashboards in Databricks SQLIntegrate Azure DevOps for version control, deploying, and productionizing solutions with CI/CD pipelinesDiscover how to use RBAC and ACLs to restrict data accessBuild end-to-end data processing pipeline for near real-time data analyticsWho this book is for This recipe-based book is for data scientists, data engineers, big data professionals, and machine learning engineers who want to perform data analytics on their applications. Prior experience of working with Apache Spark and Azure is necessary to get the most out of this book.

Hadoop Beginner's Guide

Packt Publishing Ltd
'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.
Hadoop Real-World Solutions Cookbook Second Edition CRC Press
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

MapReduce Design Patterns

Packt Publishing Ltd

Learn to solve challenging data science problems by building powerful machine learning models using Python About This Book Understand which algorithms to use in a given context with the help of this exciting recipe-based guide This practical tutorial tackles real-world computing problems through a rigorous and effective approach Build state-of-the-art models and develop personalized recommendations to perform machine learning at scale Who This Book Is For This Learning Path is for Python programmers who are looking to use machine learning algorithms to create real-world applications. It is ideal for Python professionals who want to work with large and complex datasets and Python developers and analysts or data scientists who are looking to add to their existing skills by accessing some of the most powerful recent trends in data science. Experience with Python, Jupyter Notebooks, and command-line execution together with a good level of mathematical knowledge to understand the concepts is expected. Machine learning basic knowledge is also expected. What You Will Learn Use predictive modeling and apply it to real-world problems Understand how to perform market segmentation using unsupervised learning Apply your new-found skills to solve real problems, through clearly-explained code for every technique and test Compete with top data scientists by gaining a practical and theoretical understanding of cutting-edge deep learning algorithms Increase predictive accuracy with deep learning and scalable data-handling techniques Work with modern state-of-the-art large-scale machine learning techniques Learn to use Python code to implement a range of machine learning algorithms and techniques In Detail Machine learning is increasingly spreading in the modern data-driven world. It is used extensively across many fields such as search engines, robotics, self-driving cars, and more. Machine learning is transforming the way we understand and interact with the world around us. In the first module, Python Machine Learning Cookbook, you will learn how to perform various machine learning tasks using a wide variety of machine learning algorithms to solve real-world problems and use Python to implement these algorithms. The

second module, Advanced Machine Learning with Python, is designed to take you on a guided tour of the most relevant and powerful machine learning techniques and you'll acquire a broad set of powerful skills in the area of feature selection and feature engineering. The third module in this learning path, Large Scale Machine Learning with Python, dives into scalable machine learning and the three forms of scalability. It covers the most effective machine learning techniques on a map reduce framework in Hadoop and Spark in Python. This Learning Path will teach you Python machine learning for the real world. The machine learning techniques covered in this Learning Path are at the forefront of commercial practice. This Learning Path combines some of the best that Packt has to offer in one complete, curated package. It includes content from the following Packt products: Python Machine Learning Cookbook by Prateek Joshi Advanced Machine Learning with Python by John Hearty Large Scale Machine Learning with Python by Bastiaan Sjardin, Alberto Boschetti, Luca Massaron Style and approach This course is a smooth learning path that will teach you how to get started with Python machine learning for the real world, and develop solutions to real-world problems. Through this comprehensive course, you'll learn to create the most effective machine learning techniques from scratch and more!

R Data Analysis Cookbook

Packt Publishing Ltd

Let Hadoop For Dummies help harness the power of your data and rein in the information overload Big data has become big business, and companies and organizations of all sizes are struggling to find ways to retrieve valuable information from their massive data sets with becoming overwhelmed. Enter Hadoop and this easy-to-understand For Dummies guide. Hadoop For Dummies helps readers understand the value of big data, make a business case for using Hadoop, navigate the Hadoop ecosystem, and build and manage Hadoop applications and clusters. Explains the origins of Hadoop, its economic benefits, and its functionality and practical applications Helps you find your way around the Hadoop ecosystem, program MapReduce, utilize design patterns, and get your Hadoop cluster up and running quickly and easily Details how to use Hadoop applications for data mining, web analytics and personalization, large-scale text processing, data science, and problem-solving Shows you how to improve the value of your Hadoop cluster, maximize your investment in Hadoop, and avoid common pitfalls when building your Hadoop cluster From programmers challenged with building and maintaining affordable, scaleable data systems to administrators who must deal with huge volumes of information effectively and efficiently, this how-to has something to help you with Hadoop.

Apache Hadoop 3 Quick Start Guide Manning

Data management and analytics simplified with Teradata Key Features Take your understanding of Teradata to the next level and build efficient data warehousing applications for your organization Covers recipes on data handling, warehousing, advanced querying and the administrative tasks in Teradata. Contains practical solutions to tackle common (and not-so-common) problems you might encounter in your day to day activities Book Description Teradata is an enterprise software company that develops and sells its eponymous relational

database management system (RDBMS), which is considered to be a leading data warehousing solutions and provides data management solutions for analytics. This book will help you get all the practical information you need for the creation and implementation of your data warehousing solution using Teradata. The book begins with recipes on quickly setting up a development environment so you can work with different types of data structuring and manipulation function. You will tackle all problems related to efficient querying, stored procedure searching, and navigation techniques. Additionally, you'll master various administrative tasks such as user and security management, workload management, high availability, performance tuning, and monitoring. This book is designed to take you through the best practices of performing the real daily tasks of a Teradata DBA, and will help you tackle any problem you might encounter in the process. What you will learn Understand Teradata's competitive advantage over other RDBMSs. Use SQL to process data stored in Teradata tables. Leverage Teradata's available application utilities and parallelism to play with large datasets Apply various performance tuning techniques to optimize the queries. Acquire deeper knowledge and understanding of the Teradata Architecture. Easy steps to load, archive, restore data and implement Teradata protection features Gain confidence in running a wide variety of Data analytics and develop applications for the Teradata environment Who this book is for This book is for Database administrator's and Teradata users who are looking for a practical, one-stop resource to solve all their problems while handling their Teradata solution. If you are looking to learn the basic as well as the advanced tasks involved in Teradata querying or administration, this book will be handy. Some knowledge of relational database concepts will be helpful to get the best out of this book.

Hadoop Operations

"O'Reilly Media, Inc."
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
Applied Text Analysis with Python Springer Nature
Realistic, simple code examples to solve problems at scale with Hadoop and related technologies.
Practical Data Science Cookbook Packt Publishing Ltd

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.

Hadoop: Data Processing and Modelling Packt Publishing Ltd

This book presents high-quality, original contributions (both theoretical and experimental) on Information Security, Machine Learning, Data Mining and Internet of Things (IoT). It gathers papers presented at ICETIT 2019, the 1st International Conference on Emerging Trends in Information Technology, which was held in Delhi, India, in June 2019. This conference series represents a targeted response to the growing need for research that reports on and assesses the practical implications of IoT and network technologies, AI and machine learning, data analytics and cloud computing, security and privacy, and next generation computing technologies.

Snowflake Cookbook Packt Publishing Ltd

Over 100 hands-on recipes to effectively solve real-world data problems using the most popular R packages and techniques About This Book Gain insight into how data scientists collect, process, analyze, and visualize data using some of the most popular R packages Understand how to apply useful data analysis techniques in R for real-world applications An easy-to-follow guide to make the life of data scientist easier with the problems faced while performing data analysis Who This Book Is For This book is for those who are already familiar with the basic operation of R, but want to learn how to efficiently and effectively analyze real-world data problems using practical R packages. What You Will Learn Get to know the functional characteristics of R language Extract, transform, and load data from heterogeneous sources Understand how easily R can confront probability and statistics problems Get simple R instructions to quickly organize and manipulate large datasets Create professional data visualizations and interactive reports Predict user purchase behavior by adopting a classification approach Implement data

mining techniques to discover items that are frequently purchased together Group similar text documents by using various clustering methods In Detail This cookbook offers a range of data analysis samples in simple and straightforward R code, providing step-by-step resources and time-saving methods to help you solve data problems efficiently. The first section deals with how to create R functions to avoid the unnecessary duplication of code. You will learn how to prepare, process, and perform sophisticated ETL for heterogeneous data sources with R packages. An example of data manipulation is provided, illustrating how to use the “dplyr” and “data.table” packages to efficiently process larger data structures. We also focus on “ggplot2” and show you how to create advanced figures for data exploration. In addition, you will learn how to build an interactive report using the “ggvis” package. Later chapters offer insight into time series analysis on financial data, while there is detailed information on the hot topic of machine learning, including data classification, regression, clustering, association rule mining, and dimension reduction. By the end of this book, you will understand how to resolve issues and will be able to comfortably offer solutions to problems encountered while performing data analysis. Style and approach This easy-to-follow guide is full of hands-on examples of data analysis with R. Each topic is fully explained beginning with the core concept, followed by step-by-step practical examples, and concluding with detailed explanations of each concept used.

Data Algorithms

Packt Publishing Ltd

Unlock the power of your data with Hadoop 2.X ecosystem and its data warehousing techniques across large data sets About This Book Conquer the mountain of data using Hadoop 2.X tools The authors succeed in creating a context for Hadoop and its ecosystem Hands-on examples and recipes giving the bigger picture and helping you to master Hadoop 2.X data processing platforms Overcome the challenging data processing problems using this exhaustive course with Hadoop 2.X Who This Book Is For This course is for Java developers, who know scripting, wanting a career shift to Hadoop - Big Data segment of the IT industry. So if you are a novice in Hadoop or an expert, this book will make you reach the most advanced level in Hadoop 2.X. What You Will Learn Best practices for setup and configuration of Hadoop clusters, tailoring the system to the problem at hand Integration with relational databases, using Hive for SQL queries and Sqoop for data transfer Installing and maintaining Hadoop 2.X cluster and its ecosystem Advanced Data Analysis using the Hive, Pig, and Map Reduce programs Machine learning principles with libraries such as Mahout and Batch and Stream data processing using Apache Spark Understand the changes involved in the process in the move from Hadoop 1.0 to Hadoop 2.0 Dive into YARN and Storm and use YARN to integrate Storm with Hadoop Deploy Hadoop on Amazon Elastic MapReduce and Discover HDFS replacements and learn about HDFS Federation In Detail As Marc Andreessen has said “Data is eating the world,” which can be witnessed today being the age of Big Data, businesses are producing data in huge volumes every day and this rise in tide of data need to be organized and analyzed in a more secured way. With proper and effective use of Hadoop, you

can build new-improved models, and based on that you will be able to make the right decisions. The first module, Hadoop beginners Guide will walk you through on understanding Hadoop with very detailed instructions and how to go about using it. Commands are explained using sections called “What just happened” for more clarity and understanding. The second module, Hadoop Real World Solutions Cookbook, 2nd edition, is an essential tutorial to effectively implement a big data warehouse in your business, where you get detailed practices on the latest technologies such as YARN and Spark. Big data has become a key basis of competition and the new waves of productivity growth. Hence, once you get familiar with the basics and implement the end-to-end big data use cases, you will start exploring the third module, Mastering Hadoop. So, now the question is if you need to broaden your Hadoop skill set to the next level after you nail the basics and the advance concepts, then this course is indispensable. When you finish this course, you will be able to tackle the real-world scenarios and become a big data expert using the tools and the knowledge based on the various step-by-step tutorials and recipes. Style and approach This course has covered everything right from the basic concepts of Hadoop till you master the advance mechanisms to become a big data expert. The goal here is to help you learn the basic essentials using the step-by-step tutorials and from there moving toward the recipes with various real-world solutions for you. It covers all the important aspects of Hadoop from system designing and configuring Hadoop, machine learning principles with various libraries with chapters illustrated with code fragments and schematic diagrams. This is a compendious course to explore Hadoop from the basics to the most advanced techniques available in Hadoop 2.X.

Clojure in Action Packt Publishing Ltd

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.

SQL Server 2017 Integration Services Cookbook CreateSpace

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 indespensible for anyone using Hadoop." --Tom White, author of Hadoop: The Definitive Guide

Hadoop in Action

Packt Publishing Ltd

This book constitutes the proceedings of the 4th Conference on Creativity in Intellectual Technologies and Data Science, CIT&DS 2021, held in Volgograd, Russia, in September 2021. The 39 full papers, 7 short papers, and 2 keynote papers presented were carefully reviewed and selected from 182 submissions. The papers are organized in the following topical sections: Artificial intelligence and deep learning technologies: knowledge discovery in patent and open sources; open science semantic technologies; IoT and computer vision in knowledge-based control; Cyber-physical systems and big data-driven control: pro-active modeling in intelligent decision making support; design creativity in CASE/CAI/CAD/PDM; intelligent technologies in urban design and computing; Intelligent technologies in social engineering: data science in social networks analysis and cyber security; educational creativity and game-based learning; intelligent assistive technologies: software design and application.

Proceedings of ICETIT 2019 Manning Publications

Hadoop Real-World Solutions Cookbook Second EditionPackt Publishing

NoSQL Distilled Packt Publishing Ltd

This book provides multifaceted components and full practical perspectives of systems engineering and risk management in security and defense operations with a focus on infrastructure and manpower control systems, missile design, space technology, satellites, intercontinental ballistic missiles, and space security. While there are many existing selections of systems engineering and risk management textbooks, there is no existing work that connects systems engineering and risk management concepts to solidify its usability in the entire security and defense actions. With this book Dr. Anna M. Doro-on rectifies the current imbalance. She provides a comprehensive overview of systems engineering and risk management before moving to deeper practical engineering principles integrated with newly developed concepts and examples based on industry and government methodologies. The chapters also cover related points including design principles for defeating and deactivating improvised explosive devices and land mines and security measures against kinds of threats. The book is designed for systems engineers in practice, political risk professionals, managers, policy makers, engineers in other engineering fields, scientists, decision makers in industry and government and to serve as a reference work in systems engineering and risk management courses with focus on security and defense operations.

Hadoop Real World Solutions Cookbook Packt Publishing Ltd

Over 90 hands-on recipes to help you learn and master the intricacies of Apache Hadoop 2.X, YARN, Hive, Pig, Oozie, Flume, Sqoop, Apache Spark, and Mahout About This Book Implement

outstanding Machine Learning use cases on your own analytics models and processes. Solutions to common problems when working with the Hadoop ecosystem. Step-by-step implementation of end-to-end big data use cases. Who This Book Is For Readers who have a basic knowledge of big data systems and want to advance their knowledge with hands-on recipes. What You Will Learn Installing and maintaining Hadoop 2.X cluster and its ecosystem. Write advanced Map Reduce programs and understand design patterns. Advanced Data Analysis using the Hive, Pig, and Map Reduce programs. Import and export data from various sources using Sqoop and Flume. Data storage in various file formats such as Text, Sequential, Parquet, ORC, and RC Files. Machine learning principles with libraries such as Mahout Batch and Stream data processing using Apache Spark In Detail Big data is the current requirement. Most organizations produce huge amount of data every day. With the arrival of Hadoop-like tools, it has become easier for everyone to solve big data problems with great efficiency and at minimal cost. Grasping Machine Learning techniques will help you greatly in building predictive models and using this data to make the right decisions for your organization. Hadoop Real World Solutions Cookbook gives readers insights into learning and mastering big data via recipes. The book not only clarifies most big data tools in the market but also provides best practices for using them. The book provides recipes that are based on the latest versions of Apache Hadoop 2.X, YARN, Hive, Pig, Sqoop, Flume, Apache Spark, Mahout and many more such ecosystem tools. This real-world-solution cookbook is packed with handy recipes you can apply to your own everyday issues. Each chapter provides in-depth recipes that can be referenced easily. This book provides detailed practices on the latest technologies such as YARN and Apache Spark. Readers will be able to consider themselves as big data experts on completion of this book. This guide is an invaluable tutorial if you are planning to implement a big data warehouse for your business. Style and approach An easy-to-follow guide that walks you through world of big data. Each tool in the Hadoop ecosystem is explained in detail and the recipes are placed in such a manner that readers can implement them sequentially. Plenty of reference links are provided for advanced reading.