

Hadoop Real World Solutions Cookbook

SQL Server 2017 Integration Services Cookbook Dancing with Qubits Apache Kafka Cookbook Practical Hive Apache Karaf Cookbook Hadoop Real World Solutions Cookbook Getting Started with Kudu Hadoop Real-World Solutions Cookbook Hadoop Operations Hadoop Application Architectures Hadoop: Data Processing and Modelling Hadoop: The Definitive Guide Optimizing Hadoop for MapReduce R Data Analysis Cookbook Data Algorithms Learning Spark Hadoop MapReduce Cookbook Python: Real World Machine Learning Tableau 2019.x Cookbook Cloudera Administration Handbook Big Data Analytics with R and Hadoop Hadoop Operations and Cluster Management Cookbook Hadoop Real-World Solutions Cookbook Second Edition The Car Hacker's Handbook Learning Hadoop 2 Apache Hadoop 3 Quick Start Guide Data Just Right Hadoop Real-World Solutions Cookbook Professional Hadoop Solutions Apache Hive Cookbook Learning Apache Flink Data-intensive Text Processing with MapReduce Hadoop: The Definitive Guide Robotic Process Automation Projects R for Data Science Cookbook Big Data Analytics Bad Data Handbook Mastering DynamoDB Mastering Hadoop 3 Foundations for Architecting Data Solutions

SQL Server 2017 Integration Services Cookbook

If you have interest in DynamoDB and want to know what DynamoDB is all about and become proficient in using it, this is the book for you. If you are an intermediate user who wishes to enhance your knowledge of DynamoDB, this book is aimed at you. Basic familiarity with programming, NoSQL, and cloud computing concepts would be helpful.

Dancing with Qubits

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.

Apache Kafka Cookbook

If you are ready to dive into the MapReduce framework for processing large datasets, this practical book takes you step by step through the algorithms and tools you need to build distributed MapReduce applications with Apache Hadoop or Apache Spark. Each chapter provides a recipe for solving a massive computational problem, such as building a recommendation system. You’ll learn how to implement the appropriate MapReduce solution with code that you can use in your projects. Dr. Mahmoud Parsian covers basic design patterns, optimization techniques, and data mining and machine learning solutions for problems in bioinformatics, genomics, statistics, and social network analysis. This book also includes an overview of MapReduce, Hadoop, and Spark. Topics include: Market basket analysis for a large set of transactions Data mining algorithms (K-means, KNN, and Naive Bayes) Using huge genomic data to sequence DNA and RNA Naive Bayes theorem and Markov chains for data and market prediction Recommendation algorithms and pairwise document similarity Linear regression, Cox regression, and Pearson correlation Allelic frequency and mining DNA Social network analysis (recommendation systems, counting triangles, sentiment analysis)

Practical Hive

Modern cars are more computerized than ever. Infotainment and navigation systems, Wi-Fi, automatic software updates, and other innovations aim to make driving more convenient. But vehicle technologies haven’t kept pace with today’s more hostile security environment, leaving millions vulnerable to attack. The Car Hacker’s Handbook will give you a deeper understanding of the computer systems and embedded software in modern vehicles. It begins by examining vulnerabilities

and providing detailed explanations of communications over the CAN bus and between devices and systems. Then, once you have an understanding of a vehicle's communication network, you'll learn how to intercept data and perform specific hacks to track vehicles, unlock doors, glitch engines, flood communication, and more. With a focus on low-cost, open source hacking tools such as Metasploit, Wireshark, Kayak, can-utils, and ChipWhisperer, The Car Hacker's Handbook will show you how to:

- Build an accurate threat model for your vehicle
- Reverse engineer the CAN bus to fake engine signals
- Exploit vulnerabilities in diagnostic and data-logging systems
- Hack the ECU and other firmware and embedded systems
- Feed exploits through infotainment and vehicle-to-vehicle communication systems
- Override factory settings with performance-tuning techniques
- Build physical and virtual test benches to try out exploits safely

If you're curious about automotive security and have the urge to hack a two-ton computer, make The Car Hacker's Handbook your first stop.

Apache Karaf Cookbook

Making Big Data Work: Real-World Use Cases and Examples, Practical Code, Detailed Solutions Large-scale data analysis is now vitally important to virtually every business. Mobile and social technologies are generating massive datasets; distributed cloud computing offers the resources to store and analyze them; and professionals have radically new technologies at their command, including NoSQL databases. Until now, however, most books on "Big Data" have been little more than business polemics or product catalogs. Data Just Right is different: It's a completely practical and indispensable guide for every Big Data decision-maker, implementer, and strategist. Michael Manooch, a former Google engineer and data hacker, writes for professionals who need practical solutions that can be implemented with limited resources and time. Drawing on his extensive experience, he helps you focus on building applications, rather than infrastructure, because that's where you can derive the most value. Manooch shows how to address each of today's key Big Data use cases in a cost-effective way by combining technologies in hybrid solutions. You'll find expert approaches to managing massive datasets, visualizing data, building data pipelines and dashboards, choosing tools for statistical analysis, and more. Throughout, the author demonstrates techniques using many of today's leading data analysis tools, including Hadoop, Hive, Shark, R, Apache Pig, Mahout, and Google BigQuery. Coverage includes Mastering the four guiding principles of Big Data success—and avoiding common pitfalls Emphasizing collaboration and avoiding problems with siloed data Hosting and sharing multi-terabyte datasets efficiently and economically "Building for infinity" to support rapid growth Developing a NoSQL Web app with Redis to collect crowd-sourced data Running distributed queries over massive datasets with Hadoop, Hive, and Shark Building a data dashboard with Google BigQuery Exploring large datasets with advanced visualization Implementing efficient pipelines for transforming immense amounts of data Automating complex processing with Apache Pig and the Cascading Java library Applying machine learning to classify, recommend, and predict incoming information Using R to perform statistical analysis on massive datasets Building highly efficient analytics workflows with Python and Pandas Establishing sensible purchasing strategies: when to build, buy, or outsource Previewing emerging trends and

convergences in scalable data technologies and the evolving role of the Data Scientist

Hadoop Real World Solutions Cookbook

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

Getting Started with Kudu

Easy, hands-on recipes to help you understand Hive and its integration with frameworks that are used widely in today's big data world About This Book Grasp a complete reference of different Hive topics. Get to know the latest recipes in development in Hive including CRUD operations Understand Hive internals and integration of Hive with different frameworks used in today's world. Who This Book Is For The book is intended for those who want to start in Hive or who have basic understanding of Hive framework. Prior knowledge of basic SQL command is also required What You Will Learn Learn different features and offering on the latest Hive Understand the working and structure of the Hive internals Get an insight on the latest development in Hive framework Grasp the concepts of Hive Data Model Master the key concepts like Partition, Buckets and Statistics Know how to integrate Hive with other frameworks such as Spark, Accumulo, etc In Detail Hive was developed by Facebook and later open sourced in Apache community. Hive provides SQL like interface to run queries on Big Data frameworks. Hive provides SQL like syntax also called as HiveQL that includes all SQL capabilities like analytical functions which are the need of the hour in today's Big Data world. This book provides you easy installation steps with different types of metastores supported by Hive. This book has simple and easy to learn recipes for configuring Hive clients and services. You would also learn different Hive optimizations including Partitions and Bucketing. The book also covers the source code explanation of latest Hive version. Hive Query Language is being used by other frameworks including spark. Towards the end you will cover integration of Hive with these frameworks. Style and approach Starting with

the basics and covering the core concepts with the practical usage, this book is a complete guide to learn and explore Hive offerings.

Hadoop Real-World Solutions Cookbook

Hadoop: The Definitive Guide helps you harness the power of your data. Ideal for processing large datasets, the Apache Hadoop framework is an open source implementation of the MapReduce algorithm on which Google built its empire. This comprehensive resource demonstrates how to use Hadoop to build reliable, scalable, distributed systems: programmers will find details for analyzing large datasets, and administrators will learn how to set up and run Hadoop clusters. Complete with case studies that illustrate how Hadoop solves specific problems, this book helps you: Use the Hadoop Distributed File System (HDFS) for storing large datasets, and run distributed computations over those datasets using MapReduce Become familiar with Hadoop's data and I/O building blocks for compression, data integrity, serialization, 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 Use Pig, a high-level query language for large-scale data processing Take advantage of HBase, Hadoop's database for structured and semi-structured data Learn ZooKeeper, a toolkit of coordination primitives for building distributed systems If you have lots of data -- whether it's gigabytes or petabytes -- Hadoop is the perfect solution. Hadoop: The Definitive Guide is the most thorough book available on the subject. "Now you have the opportunity to learn about Hadoop from a master-not only of the technology, but also of common sense and plain talk."-- Doug Cutting, Hadoop Founder, Yahoo!

Hadoop Operations

Discover the definitive guide to crafting lightning-fast data processing for distributed systems with Apache Flink About This Book Build your expertise in processing real-time data with Apache Flink and its ecosystem Gain insights into the working of all components of Apache Flink such as FlinkML, Gelly, and Table API filled with real world use cases Exploit Apache Flink's capabilities like distributed data streaming, in-memory processing, pipelining and iteration operators to improve performance. Solve real world big-data problems with real time in-memory and disk-based processing capabilities of Apache Flink. Who This Book Is For Big data developers who are looking to process batch and real-time data on distributed systems. Basic knowledge of Hadoop and big data is assumed. Reasonable knowledge of Java or Scala is expected. What You Will Learn Learn how to build end to end real time analytics projects Integrate with existing big data stack and utilize existing infrastructure Build predictive analytics applications using FlinkML Use graph library to perform graph querying and search. Understand Flink's - "Streaming First" architecture to implementing real streaming applications Learn Flink Logging and Monitoring best practices in order to efficiently design your data pipelines Explore the detailed processes to deploy Flink

cluster on Amazon Web Services(AWS) and Google Cloud Platform (GCP). In Detail With the advent of massive computer systems, organizations in different domains generate large amounts of data on a real-time basis. The latest entrant to big data processing, Apache Flink, is designed to process continuous streams of data at a lightning fast pace. This book will be your definitive guide to batch and stream data processing with Apache Flink. The book begins with introducing the Apache Flink ecosystem, setting it up and using the DataSet and DataStream API for processing batch and streaming datasets. Bringing the power of SQL to Flink, this book will then explore the Table API for querying and manipulating data. In the latter half of the book, readers will get to learn the remaining ecosystem of Apache Flink to achieve complex tasks such as event processing, machine learning, and graph processing. The final part of the book would consist of topics such as scaling Flink solutions, performance optimization and integrating Flink with other tools such as Elasticsearch. Whether you want to dive deeper into Apache Flink, or want to investigate how to get more out of this powerful technology, you'll find everything you need inside. Style and approach This book is a comprehensive guide that covers advanced features of the Apache Flink, and communicates them with a practical understanding of the underlying concepts for how, when, and why to use them.

Hadoop Application Architectures

A comprehensive guide to mastering the most advanced Hadoop 3 concepts Key Features Get to grips with the newly introduced features and capabilities of Hadoop 3 Crunch and process data using MapReduce, YARN, and a host of tools within the Hadoop ecosystem Sharpen your Hadoop skills with real-world case studies and code Book Description Apache Hadoop is one of the most popular big data solutions for distributed storage and for processing large chunks of data. With Hadoop 3, Apache promises to provide a high-performance, more fault-tolerant, and highly efficient big data processing platform, with a focus on improved scalability and increased efficiency. With this guide, you'll understand advanced concepts of the Hadoop ecosystem tool. You'll learn how Hadoop works internally, study advanced concepts of different ecosystem tools, discover solutions to real-world use cases, and understand how to secure your cluster. It will then walk you through HDFS, YARN, MapReduce, and Hadoop 3 concepts. You'll be able to address common challenges like using Kafka efficiently, designing low latency, reliable message delivery Kafka systems, and handling high data volumes. As you advance, you'll discover how to address major challenges when building an enterprise-grade messaging system, and how to use different stream processing systems along with Kafka to fulfil your enterprise goals. By the end of this book, you'll have a complete understanding of how components in the Hadoop ecosystem are effectively integrated to implement a fast and reliable data pipeline, and you'll be equipped to tackle a range of real-world problems in data pipelines. What you will learn Gain an in-depth understanding of distributed computing using Hadoop 3 Develop enterprise-grade applications using Apache Spark, Flink, and more Build scalable and high-performance Hadoop data pipelines with security, monitoring, and data governance Explore batch data processing patterns and how to model data in Hadoop Master best practices for enterprises using, or planning to use, Hadoop 3 as a data platform Understand security aspects of Hadoop, including

authorization and authentication Who this book is for If you want to become a big data professional by mastering the advanced concepts of Hadoop, this book is for you. You'll also find this book useful if you're a Hadoop professional looking to strengthen your knowledge of the Hadoop ecosystem. Fundamental knowledge of the Java programming language and basics of Hadoop is necessary to get started with this book.

Hadoop: Data Processing and Modelling

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. You are expected to be familiar with the Unix/Linux command-line interface and have some experience with the Java programming language. Familiarity with Hadoop would be a plus.

Hadoop: The Definitive Guide

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!

Optimizing Hadoop for MapReduce

Realistic, simple code examples to solve problems at scale with Hadoop and related technologies.

R Data Analysis Cookbook

A handy reference guide for data analysts and data scientists to help to obtain value from big data analytics using Spark on Hadoop clusters About This Book This book is based on the latest 2.0 version of Apache Spark and 2.7 version of Hadoop integrated with most commonly used tools. Learn all Spark stack components including latest topics such as DataFrames, DataSets, GraphFrames, Structured Streaming, DataFrame based ML Pipelines and SparkR. Integrations with frameworks such as HDFS, YARN and tools such as Jupyter, Zeppelin, NiFi, Mahout, HBase Spark Connector, GraphFrames, H2O and Hivemall. Who This Book Is For Though this book is primarily aimed at data analysts and data scientists, it will also help architects, programmers, and practitioners. Knowledge of either Spark or Hadoop would be beneficial. It is assumed that you have basic programming background in Scala, Python, SQL, or R programming with basic Linux experience. Working experience within big data environments is not mandatory. What You Will Learn Find out and implement the tools and techniques of big data analytics using Spark on Hadoop clusters with wide variety of tools used with Spark and Hadoop Understand all the Hadoop and Spark ecosystem components Get to know all the Spark components: Spark Core, Spark SQL, DataFrames, DataSets, Conventional and Structured Streaming, MLLib, ML Pipelines and Graphx See batch and real-time data analytics using Spark Core, Spark SQL, and Conventional and Structured Streaming Get to grips with data science and machine learning using MLLib, ML Pipelines, H2O, Hivemall, Graphx, SparkR and Hivemall. In Detail Big Data Analytics book aims at providing the fundamentals of Apache Spark and Hadoop. All Spark components - Spark Core, Spark SQL,

DataFrames, Data sets, Conventional Streaming, Structured Streaming, MLlib, Graphx and Hadoop core components - HDFS, MapReduce and Yarn are explored in greater depth with implementation examples on Spark + Hadoop clusters. It is moving away from MapReduce to Spark. So, advantages of Spark over MapReduce are explained at great depth to reap benefits of in-memory speeds. DataFrames API, Data Sources API and new Data set API are explained for building Big Data analytical applications. Real-time data analytics using Spark Streaming with Apache Kafka and HBase is covered to help building streaming applications. New Structured streaming concept is explained with an IOT (Internet of Things) use case. Machine learning techniques are covered using MLlib, ML Pipelines and SparkR and Graph Analytics are covered with GraphX and GraphFrames components of Spark. Readers will also get an opportunity to get started with web based notebooks such as Jupyter, Apache Zeppelin and data flow tool Apache NiFi to analyze and visualize data. Style and approach This step-by-step pragmatic guide will make life easy no matter what your level of experience. You will deep dive into Apache Spark on Hadoop clusters through ample exciting real-life examples. Practical tutorial explains data science in simple terms to help programmers and data analysts get started with Data Science

Data Algorithms

While many companies ponder implementation details such as distributed processing engines and algorithms for data analysis, this practical book takes a much wider view of big data development, starting with initial planning and moving diligently toward execution. Authors Ted Malaska and Jonathan Seidman guide you through the major components necessary to start, architect, and develop successful big data projects. Everyone from CIOs and COOs to lead architects and developers will explore a variety of big data architectures and applications, from massive data pipelines to web-scale applications. Each chapter addresses a piece of the software development life cycle and identifies patterns to maximize long-term success throughout the life of your project. Start the planning process by considering the key data project types Use guidelines to evaluate and select data management solutions Reduce risk related to technology, your team, and vague requirements Explore system interface design using APIs, REST, and pub/sub systems Choose the right distributed storage system for your big data system Plan and implement metadata collections for your data architecture Use data pipelines to ensure data integrity from source to final storage Evaluate the attributes of various engines for processing the data you collect

Learning Spark

Fast data ingestion, serving, and analytics in the Hadoop ecosystem have forced developers and architects to choose solutions using the least common denominator—either fast analytics at the cost of slow data ingestion or fast data ingestion at the cost of slow analytics. There is an answer to this problem. With the Apache Kudu column-oriented data store, you can

easily perform fast analytics on fast data. This practical guide shows you how. Begun as an internal project at Cloudera, Kudu is an open source solution compatible with many data processing frameworks in the Hadoop environment. In this book, current and former solutions professionals from Cloudera provide use cases, examples, best practices, and sample code to help you get up to speed with Kudu. Explore Kudu's high-level design, including how it spreads data across servers Fully administer a Kudu cluster, enable security, and add or remove nodes Learn Kudu's client-side APIs, including how to integrate Apache Impala, Spark, and other frameworks for data manipulation Examine Kudu's schema design, including basic concepts and primitives necessary to make your project successful Explore case studies for using Kudu for real-time IoT analytics, predictive modeling, and in combination with another storage engine

Hadoop MapReduce Cookbook

Solve specific problems using individual self-contained code recipes, or work through the book to develop your capabilities. This book is packed with easy-to-follow code and commands used for illustration, which makes your learning curve easy and quick. If you are a Hadoop cluster system administrator with Unix/Linux system management experience and you are looking to get a good grounding in how to set up and manage a Hadoop cluster, then this book is for you. It's assumed that you will have some experience in Unix/Linux command line already, as well as being familiar with network communication basics.

Python: Real World Machine Learning

Helping developers become more comfortable and proficient with solving problems in the Hadoop space. People will become more familiar with a wide variety of Hadoop related tools and best practices for implementation. Hadoop Real-World Solutions Cookbook will teach readers how to build solutions using tools such as Apache Hive, Pig, MapReduce, Mahout, Giraph, HDFS, Accumulo, Redis, and Ganglia. Hadoop Real-World Solutions Cookbook provides in depth explanations and code examples. Each chapter contains a set of recipes that pose, then solve, technical challenges, and can be completed in any order. A recipe breaks a single problem down into discrete steps that are easy to follow. The book covers (un)loading to and from HDFS, graph analytics with Giraph, batch data analysis using Hive, Pig, and MapReduce, machine learning approaches with Mahout, debugging and troubleshooting MapReduce, and columnar storage and retrieval of structured data using Apache Accumulo. Hadoop Real-World Solutions Cookbook will give readers the examples they need to apply Hadoop technology to their own problems.

Tableau 2019.x Cookbook

This book is intended for developers who have some familiarity with Apache Karaf and who want a quick reference for practical, proven tips on how to perform common tasks such as configuring Pax modules deployed in Apache Karaf, Extending HttpService with Apache Karaf. You should have working knowledge of Apache karaf, as the book provides a deeper understanding of the capabilities of Apache Karaf.

Cloudera Administration Handbook

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.

Big Data Analytics with R and Hadoop

Harness the power of SQL Server 2017 Integration Services to build your data integration solutions with ease About This

Book Acquaint yourself with all the newly introduced features in SQL Server 2017 Integration Services Program and extend your packages to enhance their functionality This detailed, step-by-step guide covers everything you need to develop efficient data integration and data transformation solutions for your organization Who This Book Is For This book is ideal for software engineers, DW/ETL architects, and ETL developers who need to create a new, or enhance an existing, ETL implementation with SQL Server 2017 Integration Services. This book would also be good for individuals who develop ETL solutions that use SSIS and are keen to learn the new features and capabilities in SSIS 2017. What You Will Learn Understand the key components of an ETL solution using SQL Server 2016-2017 Integration Services Design the architecture of a modern ETL solution Have a good knowledge of the new capabilities and features added to Integration Services Implement ETL solutions using Integration Services for both on-premises and Azure data Improve the performance and scalability of an ETL solution Enhance the ETL solution using a custom framework Be able to work on the ETL solution with many other developers and have common design paradigms or techniques Effectively use scripting to solve complex data issues In Detail SQL Server Integration Services is a tool that facilitates data extraction, consolidation, and loading options (ETL), SQL Server coding enhancements, data warehousing, and customizations. With the help of the recipes in this book, you'll gain complete hands-on experience of SSIS 2017 as well as the 2016 new features, design and development improvements including SCD, Tuning, and Customizations. At the start, you'll learn to install and set up SSIS as well other SQL Server resources to make optimal use of this Business Intelligence tools. We'll begin by taking you through the new features in SSIS 2016/2017 and implementing the necessary features to get a modern scalable ETL solution that fits the modern data warehouse. Through the course of chapters, you will learn how to design and build SSIS data warehouses packages using SQL Server Data Tools. Additionally, you'll learn to develop SSIS packages designed to maintain a data warehouse using the Data Flow and other control flow tasks. You'll also be demonstrated many recipes on cleansing data and how to get the end result after applying different transformations. Some real-world scenarios that you might face are also covered and how to handle various issues that you might face when designing your packages. At the end of this book, you'll get to know all the key concepts to perform data integration and transformation. You'll have explored on-premises Big Data integration processes to create a classic data warehouse, and will know how to extend the toolbox with custom tasks and transforms. Style and approach This cookbook follows a problem-solution approach and tackles all kinds of data integration scenarios by using the capabilities of SQL Server 2016 Integration Services. This book is well supplemented with screenshots, tips, and tricks. Each recipe focuses on a particular task and is written in a very easy-to-follow manner.

Hadoop Operations and Cluster Management Cookbook

Over 50 hands-on recipes to efficiently administer, maintain, and use your Apache Kafka installation About This Book Quickly configure and manage your Kafka cluster Learn how to use the Apache Kafka cluster and connect it with tools for big data processing A practical guide to monitor your Apache Kafka installation Who This Book Is For If you are a

programmer or big data engineer using or planning to use Apache Kafka, then this book is for you. This book has several recipes which will teach you how to effectively use Apache Kafka. You need to have some basic knowledge of Java. If you don't know big data tools, this would be your stepping stone for learning how to consume the data in these kind of systems.

What You Will Learn

- Learn how to configure Kafka brokers for better efficiency
- Explore how to configure producers and consumers for optimal performance
- Set up tools for maintaining and operating Apache Kafka
- Create producers and consumers for Apache Kafka in Java
- Understand how Apache Kafka can be used by several third party system for big data processing, such as Apache Storm, Apache Spark, Hadoop, and more
- Monitor Apache Kafka using tools like graphite and Ganglia

In Detail

This book will give you details about how to manage and administer your Apache Kafka Cluster. We will cover topics like how to configure your broker, producer, and consumer for maximum efficiency for your situation. Also, you will learn how to maintain and administer your cluster for fault tolerance. We will also explore tools provided with Apache Kafka to do regular maintenance operations. We shall also look at how to easily integrate Apache Kafka with big data tools like Hadoop, Apache Spark, Apache Storm, and Elasticsearch.

Style and approach Easy-to-follow, step-by-step recipes explaining from start to finish how to accomplish real-world tasks.

Hadoop Real-World Solutions Cookbook Second Edition

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.

The Car Hacker's Handbook

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.

Learning Hadoop 2

Data is bigger, arrives faster, and comes in a variety of formats—and it all needs to be processed at scale for analytics or machine learning. But how can you process such varied workloads efficiently? Enter Apache Spark. Updated to include Spark 3.0, this second edition shows data engineers and data scientists why structure and unification in Spark matters. Specifically, this book explains how to perform simple and complex data analytics and employ machine learning algorithms. Through step-by-step walk-throughs, code snippets, and notebooks, you'll be able to: Learn Python, SQL, Scala, or Java high-level Structured APIs Understand Spark operations and SQL Engine Inspect, tune, and debug Spark operations with Spark configurations and Spark UI Connect to data sources: JSON, Parquet, CSV, Avro, ORC, Hive, S3, or Kafka Perform analytics on batch and streaming data using Structured Streaming Build reliable data pipelines with open source Delta Lake and Spark Develop machine learning pipelines with MLlib and productionize models using MLflow

Apache Hadoop 3 Quick Start Guide

Over 80 recipes to help you breeze through your data analysis projects using R About This Book Analyse your data using the popular R packages like ggplot2 with ready-to-use and customizable recipes Find meaningful insights from your data and generate dynamic reports A practical guide to help you put your data analysis skills in R to practical use Who This Book Is For This book is for data scientists, analysts and even enthusiasts who want to learn and implement the various data analysis techniques using R in a practical way. Those looking for quick, handy solutions to common tasks and challenges in data analysis will find this book to be very useful. Basic knowledge of statistics and R programming is assumed. What You

Will Learn Acquire, format and visualize your data using R Using R to perform an Exploratory data analysis Introduction to machine learning algorithms such as classification and regression Get started with social network analysis Generate dynamic reporting with Shiny Get started with geospatial analysis Handling large data with R using Spark and MongoDB Build Recommendation system- Collaborative Filtering, Content based and Hybrid Learn real world dataset examples- Fraud Detection and Image Recognition In Detail Data analytics with R has emerged as a very important focus for organizations of all kinds. R enables even those with only an intuitive grasp of the underlying concepts, without a deep mathematical background, to unleash powerful and detailed examinations of their data. This book will show you how you can put your data analysis skills in R to practical use, with recipes catering to the basic as well as advanced data analysis tasks. Right from acquiring your data and preparing it for analysis to the more complex data analysis techniques, the book will show you how you can implement each technique in the best possible manner. You will also visualize your data using the popular R packages like ggplot2 and gain hidden insights from it. Starting with implementing the basic data analysis concepts like handling your data to creating basic plots, you will master the more advanced data analysis techniques like performing cluster analysis, and generating effective analysis reports and visualizations. Throughout the book, you will get to know the common problems and obstacles you might encounter while implementing each of the data analysis techniques in R, with ways to overcoming them in the easiest possible way. By the end of this book, you will have all the knowledge you need to become an expert in data analysis with R, and put your skills to test in real-world scenarios. Style and Approach Hands-on recipes to walk through data science challenges using R Your one-stop solution for common and not-so-common pain points while performing real-world problems to execute a series of tasks. Addressing your common and not-so-common pain points, this is a book that you must have on the shelf

Data Just Right

Learn RPA by building business solutions such as ERP and CRM automation, software robots, and intelligent process automation from scratch Key Features Use popular RPA tools Automation Anywhere A2019 and UiPath, for real-world task automation Build automation solutions for domains such as System Administration, Finance, HR, Supply Chain, and Customer Relations Extend your RPA capabilities by implementing Intelligent process automation with APIs and AI Book Description Robotic Process automation helps businesses to automate monotonous tasks that can be performed by machines. This project-based guide will help you progress through easy to more advanced RPA projects. You'll learn the principles of RPA and how to architect solutions to meet the demands of business automation, along with exploring the most popular RPA tools - UiPath and Automation Anywhere. In the first part, you'll learn how to use UiPath by building a simple helpdesk ticket system. You'll then automate CRM systems by integrating Excel data with UiPath. After this, the book will guide you through building an AI-based social media moderator using Google Cloud Vision API. In the second part, you'll learn about Automation Anywhere's latest Cloud RPA platform (A2019) by creating projects such as an automated ERP

administration system, an AI bot for order and invoice processing, and an automated emergency notification system for employees. Later, you'll get hands-on with advanced RPA tasks such as invoking APIs, before covering complex concepts such as Artificial Intelligence (AI) and machine learning in automation to take your understanding of RPA to the next level. By the end of the book, you'll have a solid foundation in RPA with experience in building real-world projects. What you will learn

- Explore RPA principles, techniques, and tools using an example-driven approach
- Understand the basics of UiPath by building a helpdesk ticket generation system
- Automate read and write operations from Excel in a CRM system using UiPath
- Build an AI-based social media moderator platform using Google Cloud Vision API with UiPath
- Explore how to use Automation Anywhere by building a simple sales order processing system
- Build an automated employee emergency reporting system using Automation Anywhere
- Test your knowledge of building an automated workflow through fun exercises

Who this book is for This RPA book is for enterprise application developers, software developers, business analysts, or any professional who wants to implement RPA across various domains of the business. The book assumes some understanding of enterprise systems. Computer programming experience will also be beneficial.

Hadoop Real-World Solutions Cookbook

Dive into the world of SQL on Hadoop and get the most out of your Hive data warehouses. This book is your go-to resource for using Hive: authors Scott Shaw, Ankur Gupta, David Kjerrumgaard, and Andreas Francois Vermeulen take you through learning HiveQL, the SQL-like language specific to Hive, to analyze, export, and massage the data stored across your Hadoop environment. From deploying Hive on your hardware or virtual machine and setting up its initial configuration to learning how Hive interacts with Hadoop, MapReduce, Tez and other big data technologies, Practical Hive gives you a detailed treatment of the software. In addition, this book discusses the value of open source software, Hive performance tuning, and how to leverage semi-structured and unstructured data. What You Will Learn

- Install and configure Hive for new and existing datasets
- Perform DDL operations
- Execute efficient DML operations
- Use tables, partitions, buckets, and user-defined functions
- Discover performance tuning tips and Hive best practices

Who This Book Is For Developers, companies, and professionals who deal with large amounts of data and could use software that can efficiently manage large volumes of input. It is assumed that readers have the ability to work with SQL.

Professional Hadoop Solutions

Get expert guidance on architecting end-to-end data management solutions with Apache Hadoop. While many sources explain how to use various components in the Hadoop ecosystem, this practical book takes you through architectural considerations necessary to tie those components together into a complete tailored application, based on your particular use case. To reinforce those lessons, the book's second section provides detailed examples of architectures used in some of

the most commonly found Hadoop applications. Whether you're designing a new Hadoop application, or planning to integrate Hadoop into your existing data infrastructure, Hadoop Application Architectures will skillfully guide you through the process. This book covers: Factors to consider when using Hadoop to store and model data Best practices for moving data in and out of the system Data processing frameworks, including MapReduce, Spark, and Hive Common Hadoop processing patterns, such as removing duplicate records and using windowing analytics Giraph, GraphX, and other tools for large graph processing on Hadoop Using workflow orchestration and scheduling tools such as Apache Oozie Near-real-time stream processing with Apache Storm, Apache Spark Streaming, and Apache Flume Architecture examples for clickstream analysis, fraud detection, and data warehousing

Apache Hive Cookbook

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.

Learning Apache Flink

Explore the principles and practicalities of quantum computing* Discover how quantum computing works and delve into the math behind it with this quantum computing textbook* Learn how it may become the most important new computer technology of the century* Explore the inner workings of quantum computing technology to quickly process complex cloud data and solve problems* Book Description Quantum computing is making us change the way we think about computers. Quantum bits, a.k.a. qubits, can make it possible to solve problems that would otherwise be intractable with current computing technology. Dancing with Qubits is a quantum computing textbook that starts with an overview of why quantum computing is so different from classical computing and describes several industry use cases where it can have a major impact. From there it moves on to a fuller description of classical computing and the mathematical underpinnings necessary to understand such concepts as superposition, entanglement, and interference. Next up is circuits and algorithms, both basic and more sophisticated. It then nicely moves on to provide a survey of the physics and engineering ideas behind how quantum computing hardware is built. Finally, the book looks to the future and gives you guidance on understanding how further developments will affect you. Really understanding quantum computing requires a lot of math, and this book doesn't shy away from the necessary math concepts you'll need. Each topic is introduced and explained thoroughly, in clear English with helpful examples. What you will learn* See how quantum computing works, delve into the math behind it, what makes it different, and why it is so powerful with this quantum computing textbook* Discover the complex, mind-bending mechanics that underpin quantum systems* Understand the necessary concepts behind classical and quantum computing* Refresh and extend your grasp of essential mathematics, computing, and quantum theory* Explore the main applications of quantum computing to the fields of scientific computing, AI, and elsewhere* Examine a detailed overview of qubits, quantum circuits, and quantum algorithm Who this book is for Dancing with Qubits is a quantum computing textbook for those who want to deeply explore the inner workings of quantum computing. This entails some sophisticated mathematical exposition and is therefore best suited for those with a healthy interest in mathematics,

physics, engineering, and computer science.

Data-intensive Text Processing with MapReduce

What is bad data? Some people consider it a technical phenomenon, like missing values or malformed records, but bad data includes a lot more. In this handbook, data expert Q. Ethan McCallum has gathered 19 colleagues from every corner of the data arena to reveal how they've recovered from nasty data problems. From cranky storage to poor representation to misguided policy, there are many paths to bad data. Bottom line? Bad data is data that gets in the way. This book explains effective ways to get around it. Among the many topics covered, you'll discover how to: Test drive your data to see if it's ready for analysis Work spreadsheet data into a usable form Handle encoding problems that lurk in text data Develop a successful web-scraping effort Use NLP tools to reveal the real sentiment of online reviews Address cloud computing issues that can impact your analysis effort Avoid policies that create data analysis roadblocks Take a systematic approach to data quality analysis

Hadoop: The Definitive Guide

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.

Robotic Process Automation Projects

This book is an example-based tutorial that deals with Optimizing Hadoop for MapReduce job performance. If you are a Hadoop administrator, developer, MapReduce user, or beginner, this book is the best choice available if you wish to optimize your clusters and applications. Having prior knowledge of creating MapReduce applications is not necessary, but will help you better understand the concepts and snippets of MapReduce class template code.

R for Data Science Cookbook

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.

Big Data Analytics

Our world is being revolutionized by data-driven methods: access to large amounts of data has generated new insights and opened exciting new opportunities in commerce, science, and computing applications. Processing the enormous quantities of data necessary for these advances requires large clusters, making distributed computing paradigms more crucial than ever. MapReduce is a programming model for expressing distributed computations on massive datasets and an execution framework for large-scale data processing on clusters of commodity servers. The programming model provides an easy-to-understand abstraction for designing scalable algorithms, while the execution framework transparently handles many system-level details, ranging from scheduling to synchronization to fault tolerance. This book focuses on MapReduce algorithm design, with an emphasis on text processing algorithms common in natural language processing, information retrieval, and machine learning. We introduce the notion of MapReduce design patterns, which represent general reusable solutions to commonly occurring problems across a variety of problem domains. This book not only intends to help the reader "think in MapReduce", but also discusses limitations of the programming model as well. This volume is a printed version of a work that appears in the Synthesis Digital Library of Engineering and Computer Science. Synthesis Lectures provide concise, original presentations of important research and development topics, published quickly, in digital and print formats. For more information visit www.morganclaypool.com

Bad Data Handbook

The go-to guidebook for deploying Big Data solutions with Hadoop Today's enterprise architects need to understand how the Hadoop frameworks and APIs fit together, and how they can be integrated to deliver real-world solutions. This book is a practical, detailed guide to building and implementing those solutions, with code-level instruction in the popular Wrox tradition. It covers storing data with HDFS and Hbase, processing data with MapReduce, and automating data processing with Oozie. Hadoop security, running Hadoop with Amazon Web Services, best practices, and automating Hadoop processes in real time are also covered in depth. With in-depth code examples in Java and XML and the latest on recent additions to the Hadoop ecosystem, this complete resource also covers the use of APIs, exposing their inner workings and allowing architects and developers to better leverage and customize them. The ultimate guide for developers, designers, and architects who need to build and deploy Hadoop applications Covers storing and processing data with various technologies, automating data processing, Hadoop security, and delivering real-time solutions Includes detailed, real-world examples and code-level guidelines Explains when, why, and how to use these tools effectively Written by a team of Hadoop experts in the programmer-to-programmer Wrox style Professional Hadoop Solutions is the reference enterprise architects

and developers need to maximize the power of Hadoop.

Mastering DynamoDB

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

Mastering Hadoop 3

Perform advanced dashboard, visualization, and analytical techniques with Tableau Desktop, Tableau Prep, and Tableau Server Key Features Unique problem-solution approach to aid effective business decision-making Create interactive dashboards and implement powerful business intelligence solutions Includes best practices on using Tableau with modern cloud analytics services Book Description Tableau has been one of the most popular business intelligence solutions in recent times, thanks to its powerful and interactive data visualization capabilities. Tableau 2019.x Cookbook is full of useful recipes from industry experts, who will help you master Tableau skills and learn each aspect of Tableau's ecosystem. This book is enriched with features such as Tableau extracts, Tableau advanced calculations, geospatial analysis, and building dashboards. It will guide you with exciting data manipulation, storytelling, advanced filtering, expert visualization, and forecasting techniques using real-world examples. From basic functionalities of Tableau to complex deployment on Linux, you will cover it all. Moreover, you will learn advanced features of Tableau using R, Python, and various APIs. You will learn how to prepare data for analysis using the latest Tableau Prep. In the concluding chapters, you will learn how Tableau fits the modern world of analytics and works with modern data platforms such as Snowflake and Redshift. In addition, you will learn about the best practices of integrating Tableau with ETL using Matillion ETL. By the end of the book, you will be ready to tackle business intelligence challenges using Tableau's features. What you will learn Understand the basic and advanced skills of Tableau Desktop Implement best practices of visualization, dashboard, and storytelling Learn advanced analytics with the use of build in statistics Deploy the multi-node server on Linux and Windows Use Tableau with big data sources

such as Hadoop, Athena, and Spectrum Cover Tableau built-in functions for forecasting using R packages Combine, shape, and clean data for analysis using Tableau Prep Extend Tableau's functionalities with REST API and R/Python Who this book is for Tableau 2019.x Cookbook is for data analysts, data engineers, BI developers, and users who are looking for quick solutions to common and not-so-common problems faced while using Tableau products. Put each recipe into practice by bringing the latest offerings of Tableau 2019.x to solve real-world analytics and business intelligence challenges. Some understanding of BI concepts and Tableau is required.

Foundations for Architecting Data Solutions

A fast paced guide that will help you learn about Apache Hadoop 3 and its ecosystem Key Features Set up, configure and get started with Hadoop to get useful insights from large data sets Work with the different components of Hadoop such as MapReduce, HDFS and YARN Learn about the new features introduced in Hadoop 3 Book Description Apache Hadoop is a widely used distributed data platform. It enables large datasets to be efficiently processed instead of using one large computer to store and process the data. This book will get you started with the Hadoop ecosystem, and introduce you to the main technical topics, including MapReduce, YARN, and HDFS. The book begins with an overview of big data and Apache Hadoop. Then, you will set up a pseudo Hadoop development environment and a multi-node enterprise Hadoop cluster. You will see how the parallel programming paradigm, such as MapReduce, can solve many complex data processing problems. The book also covers the important aspects of the big data software development lifecycle, including quality assurance and control, performance, administration, and monitoring. You will then learn about the Hadoop ecosystem, and tools such as Kafka, Sqoop, Flume, Pig, Hive, and HBase. Finally, you will look at advanced topics, including real time streaming using Apache Storm, and data analytics using Apache Spark. By the end of the book, you will be well versed with different configurations of the Hadoop 3 cluster. What you will learn Store and analyze data at scale using HDFS, MapReduce and YARN Install and configure Hadoop 3 in different modes Use Yarn effectively to run different applications on Hadoop based platform Understand and monitor how Hadoop cluster is managed Consume streaming data using Storm, and then analyze it using Spark Explore Apache Hadoop ecosystem components, such as Flume, Sqoop, HBase, Hive, and Kafka Who this book is for Aspiring Big Data professionals who want to learn the essentials of Hadoop 3 will find this book to be useful. Existing Hadoop users who want to get up to speed with the new features introduced in Hadoop 3 will also benefit from this book. Having knowledge of Java programming will be an added advantage.

[ROMANCE](#) [ACTION & ADVENTURE](#) [MYSTERY & THRILLER](#) [BIOGRAPHIES & HISTORY](#) [CHILDREN'S](#) [YOUNG ADULT](#) [FANTASY](#)
[HISTORICAL FICTION](#) [HORROR](#) [LITERARY FICTION](#) [NON-FICTION](#) [SCIENCE FICTION](#)