
Hadoop Operations

A Guide for Developers and Administrators

Hadoop Operations

How to Install, Deploy, and Optimize Hadoop in a Virtualized Architecture

Big Data Using Hadoop and Hive

Hadoop: The Definitive Guide

Big Data and Hadoop

Big Data Analytics with R and Hadoop

Hadoop For Dummies

Architecting Modern Data Platforms

Storage and Analysis at Internet Scale

Third TPC Technology Conference, TPCTC 2011, Seattle, WA, USA, August 29- September 3, 2011. Revised Selected Papers

An Introduction for Data Scientists

Architecting Modern Data Platforms

Hadoop in 24 Hours, Sams Teach Yourself

Space Operations

Data Analytics with Hadoop

Big data processing at scale to unlock unique business insights

An Introduction for Data Scientists

Cloudera Administration Handbook

Unlocking Hadoop for Your Relational Database

Apache Sqoop Cookbook

The First Step towards Hadoop Administration and Management

Mastering Hadoop

Hadoop Application Architectures

Hadoop Operations

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

Managing Spark, YARN, and MapReduce

Big Data Forensics - Learning Hadoop Investigations

Beginning Apache Hadoop Administration

Professional Hadoop Solutions

A Guide for Developers and Administrators

Applied Big Data Analytics in Operations Management

Building Effective Algorithms and Analytics for Hadoop and Other Systems

Mastering Hadoop 3

Hadoop: Data Processing and Modelling

Hadoop 2.x Administration Cookbook

15th International Symposium, SSTD 2017, Arlington, VA, USA, August 21 - 23, 2017, Proceedings

Data Analytics with Hadoop

VIRGINIA MILLS

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

Summarization patterns: get a top-level view by summarizing and grouping data
Filtering patterns: view data subsets such as records generated from one user
Data organization patterns: reorganize data to work with other systems, or to make MapReduce analysis easier
Join patterns: analyze different datasets together to discover interesting relationships
Metapatterns: piece together several patterns to solve multi-stage problems, or to perform several analytics in the same job
Input and output patterns: customize the way you use Hadoop to load or store data
"A clear exposition of MapReduce programs for common data processing patterns—this book is indispensable for anyone using Hadoop." --Tom White, author of *Hadoop: The Definitive Guide*

Hadoop Operations Packt Publishing Ltd

Explore big data concepts, platforms, analytics, and their applications using the power of Hadoop 3
Key Features
Learn Hadoop 3 to build effective big data analytics solutions on-premise and on cloud
Integrate Hadoop with other big data tools such as R, Python, Apache Spark, and Apache Flink
Exploit big data using Hadoop 3 with real-world examples
Book Description
Apache Hadoop is the most popular platform for big data processing, and can be combined with a host of other big data tools to build powerful analytics solutions. *Big Data Analytics with Hadoop 3* shows you how to do just that, by providing insights into the software as well as its benefits with the help of practical

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

How to Install, Deploy, and Optimize Hadoop in a Virtualized Architecture "O'Reilly Media, Inc."

Integrating data from multiple sources is essential in the age of big data, but it can be a challenging and time-consuming task. This handy cookbook provides dozens of ready-to-use recipes for using Apache Sqoop, the command-line interface application that optimizes data transfers between relational databases and Hadoop. Sqoop is both powerful and bewildering, but with this cookbook's problem-solution-discussion format, you'll quickly learn how to deploy and then apply Sqoop in your environment. The authors provide MySQL, Oracle, and PostgreSQL database examples on GitHub that you can easily adapt for SQL Server, Netezza, Teradata, or other relational systems. Transfer data from a single database table into your Hadoop ecosystem
Keep table

data and Hadoop in sync by importing data incrementally
Import data from more than one database table
Customize transferred data by calling various database functions
Export generated, processed, or backed-up data from Hadoop to your database
Run Sqoop within Oozie, Hadoop's specialized workflow scheduler
Load data into Hadoop's data warehouse (Hive) or database (HBase)
Handle installation, connection, and syntax issues common to specific database vendors
Big Data Using Hadoop and Hive "O'Reilly Media, Inc."
Discover how Apache Hadoop can unleash the power of your data. This comprehensive resource shows you how to build and maintain reliable, scalable, distributed systems with the Hadoop framework -- an open source implementation of MapReduce, the algorithm on which Google built its empire. Programmers will find details for analyzing datasets of any size, and administrators will learn how to set up and run Hadoop clusters. This revised edition covers recent changes to Hadoop, including new features such as Hive, Sqoop, and Avro. It also provides illuminating case studies that illustrate how Hadoop is used to solve specific problems. Looking to get the most out of your data? This is your book. Use the Hadoop Distributed File System (HDFS) for storing large datasets, then run distributed computations over those datasets with 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
Analyze datasets with Hive, Hadoop's data warehousing system
Take advantage of HBase, Hadoop's database for structured and semi-structured data
Learn ZooKeeper, a toolkit of coordination primitives for building distributed systems
"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, Cloudera
Hadoop: The Definitive Guide Springer Nature
This is the eBook of the printed book and may not include any media, website access codes, or print supplements that may come packaged with the bound book. The Comprehensive, Up-to-Date Apache Hadoop Administration Handbook and Reference
"Sam Alapati has worked with production Hadoop clusters for six

years. His unique depth of experience has enabled him to write the go-to resource for all administrators looking to spec, size, expand, and secure production Hadoop clusters of any size.” —Paul Dix, Series Editor In *Expert Hadoop® Administration*, leading Hadoop administrator Sam R. Alapati brings together authoritative knowledge for creating, configuring, securing, managing, and optimizing production Hadoop clusters in any environment. Drawing on his experience with large-scale Hadoop administration, Alapati integrates action-oriented advice with carefully researched explanations of both problems and solutions. He covers an unmatched range of topics and offers an unparalleled collection of realistic examples. Alapati demystifies complex Hadoop environments, helping you understand exactly what happens behind the scenes when you administer your cluster. You’ll gain unprecedented insight as you walk through building clusters from scratch and configuring high availability, performance, security, encryption, and other key attributes. The high-value administration skills you learn here will be indispensable no matter what Hadoop distribution you use or what Hadoop applications you run. Understand Hadoop’s architecture from an administrator’s standpoint Create simple and fully distributed clusters Run MapReduce and Spark applications in a Hadoop cluster Manage and protect Hadoop data and high availability Work with HDFS commands, file permissions, and storage management Move data, and use YARN to allocate resources and schedule jobs Manage job workflows with Oozie and Hue Secure, monitor, log, and optimize Hadoop Benchmark and troubleshoot Hadoop

Big Data and Hadoop Springer

Ready to use statistical and machine-learning techniques across large data sets? This practical guide shows you why the Hadoop ecosystem is perfect for the job. Instead of deployment, operations, or software development usually associated with distributed computing, you’ll focus on particular analyses you can build, the data warehousing techniques that Hadoop provides, and higher order data workflows this framework can produce. Data scientists and analysts will learn how to perform a wide range of techniques, from writing MapReduce and Spark applications with Python to using advanced modeling and data management with Spark MLlib, Hive, and HBase. You’ll also learn about the analytical processes and data systems available to

build and empower data products that can handle—and actually require—huge amounts of data. Understand core concepts behind Hadoop and cluster computing Use design patterns and parallel analytical algorithms to create distributed data analysis jobs Learn about data management, mining, and warehousing in a distributed context using Apache Hive and HBase Use Sqoop and Apache Flume to ingest data from relational databases Program complex Hadoop and Spark applications with Apache Pig and Spark DataFrames Perform machine learning techniques such as classification, clustering, and collaborative filtering with Spark’s MLlib

Big Data Analytics with R and Hadoop Packt Publishing Ltd Offers information on the architecture and data design necessary to create Hadoop-based enterprise applications.

Hadoop For Dummies CreateSpace

Apache Spark is a fast, scalable, and flexible open source distributed processing engine for big data systems and is one of the most active open source big data projects to date. In just 24 lessons of one hour or less, Sams Teach Yourself Apache Spark in 24 Hours helps you build practical Big Data solutions that leverage Spark’s amazing speed, scalability, simplicity, and versatility. This book’s straightforward, step-by-step approach shows you how to deploy, program, optimize, manage, integrate, and extend Spark—now, and for years to come. You’ll discover how to create powerful solutions encompassing cloud computing, real-time stream processing, machine learning, and more. Every lesson builds on what you’ve already learned, giving you a rock-solid foundation for real-world success. Whether you are a data analyst, data engineer, data scientist, or data steward, learning Spark will help you to advance your career or embark on a new career in the booming area of Big Data. Learn how to • Discover what Apache Spark does and how it fits into the Big Data landscape • Deploy and run Spark locally or in the cloud • Interact with Spark from the shell • Make the most of the Spark Cluster Architecture • Develop Spark applications with Scala and functional Python • Program with the Spark API, including transformations and actions • Apply practical data engineering/analysis approaches designed for Spark • Use Resilient Distributed Datasets (RDDs) for caching, persistence, and output • Optimize Spark solution performance • Use Spark with SQL (via Spark SQL) and with NoSQL (via Cassandra) •

Leverage cutting-edge functional programming techniques • Extend Spark with streaming, R, and Sparkling Water • Start building Spark-based machine learning and graph-processing applications • Explore advanced messaging technologies, including Kafka • Preview and prepare for Spark’s next generation of innovations Instructions walk you through common questions, issues, and tasks; Q-and-As, Quizzes, and Exercises build and test your knowledge; "Did You Know?" tips offer insider advice and shortcuts; and "Watch Out!" alerts help you avoid pitfalls. By the time you're finished, you'll be comfortable using Apache Spark to solve a wide spectrum of Big Data problems.

Architecting Modern Data Platforms "O'Reilly Media, Inc."

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. *Storage and Analysis at Internet Scale* Simon and Schuster 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.

Third TPC Technology Conference, TPCTC 2011, Seattle, WA, USA, August 29- September 3, 2011. Revised Selected Papers "O'Reilly Media, Inc."

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

Until recently, Hadoop deployments existed on hardware owned and run by organizations. Now, of course, you can acquire the computing resources and network connectivity to run Hadoop clusters in the cloud. But there's a lot more to deploying Hadoop to the public cloud than simply renting machines. This hands-on guide shows developers and systems administrators familiar with Hadoop how to install, use, and manage cloud-born clusters efficiently. You'll learn how to architect clusters that work with cloud-provider features—not just to avoid pitfalls, but also to take full advantage of these services. You'll also compare the Amazon, Google, and Microsoft clouds, and learn how to set up clusters in each of them. Learn how Hadoop clusters run in the cloud, the problems they can help you solve, and their potential drawbacks. Examine the common concepts of cloud providers, including compute capabilities, networking and security, and storage. Build a functional Hadoop cluster on cloud infrastructure, and learn what the major providers require. Explore use cases for high availability, relational data with Hive, and complex analytics with Spark. Get patterns and practices for running cloud clusters, from designing for price and security to dealing with maintenance.

Architecting Modern Data Platforms "O'Reilly Media, Inc."

Get ready to unlock the power of your data. With the fourth edition of 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. Using Hadoop 2 exclusively, author Tom White presents new chapters on YARN and several Hadoop-related projects such as Parquet, Flume, Crunch, and Spark. You'll learn about recent changes to Hadoop, and explore new case studies on Hadoop's role in healthcare systems and genomics data processing. Learn fundamental components such as MapReduce, HDFS, and YARN. Explore MapReduce in depth, including steps for developing applications with it. Set up and maintain a Hadoop cluster running HDFS and MapReduce on YARN. Learn two data formats: Avro for data serialization and Parquet for nested data. Use data ingestion tools such as Flume (for streaming data) and Sqoop (for bulk data transfer). Understand how high-level data processing tools like Pig, Hive, Crunch, and Spark work with Hadoop. Learn the HBase distributed database and the ZooKeeper distributed configuration service.

Hadoop in 24 Hours, Sams Teach Yourself "O'Reilly Media, Inc."

Introduction Data warehousing is a success, judging by its 25 year history of use across all industries. Business intelligence met the needs it was designed for: to give non-technical people within the organization access to important, shared data. During the same period that data warehousing and BI matured, the automation and instrumenting of almost all processes and activities changed the data landscape in most companies. Where there were only a few applications and minimal monitoring 25 years ago, there is ubiquitous computing and data available about every activity today. Data warehouses have not been able to keep up with business demands for new sources of information, new types of data, more complex analysis and greater speed. Companies can put this data to use in countless ways, but for most it remains uncollected or unused, locked away in silos within IT. There has been a gradual maturing of data use in organizations. In the early days of BI it was enough to provide access to core financial and customer transactions. Better access enabled process changes, and these led to the need for more data and more varied uses of information. These changes put increasing strain on information processing and delivery capabilities that were designed under assumptions of stability and common use. Most companies now have a backlog of new data and analysis requests that BI groups

are struggling to meet. Big data is not simply about growing data volumes - it's also about the fact that the data being collected today is different in ways that make it unwieldy for conventional databases and BI tools. Big data is also about new technologies that were developed to support the storage, retrieval and processing of this new data. The technologies originated in the world of web applications and internet-based companies, but they are now spreading into enterprise applications of all sorts. New technology coupled with new data enables new practices like real-time monitoring of operations across retail channels, supply chain practices at finer grain and faster speed, and analysis of customers at the level of individual activities and behaviors. Until recently, large scale data collection and analysis capabilities like these would have required a Wal-Mart sized investment, limiting them to large organizations. These capabilities are now available to all, regardless of company size or budget. This is creating a rush to adopt big data technologies. As the use of big data grows, the need for data management will grow. Many organizations already struggle to manage existing data. Big data adds complexity, which will only increase the challenge. The combination of new data and new technology requires new data management capabilities and processes to capture the promised long-term value. Wal-Mart handles more than a million customer transactions each hour and imports those into databases estimated to contain more than 2.5 petabytes of data. Radio frequency identification (RFID) systems used by retailers and others can generate 100 to 1,000 times the data of conventional bar code systems. Facebook handles more than 250 million photo uploads and the interactions of 800 million active users with more than 900 million objects (pages, groups, etc.) - each day. More than 5 billion people are calling, texting, tweeting and browsing on mobile phones worldwide. Organizations are inundated with data - terabytes and petabytes of it. To put it in context, 1 terabyte contains 2,000 hours of CD-quality music and 10 terabytes could store the entire US Library of Congress print collection. Exabytes, zettabytes and yottabytes definitely are on the horizon. Data is pouring in from every conceivable direction: from operational and transactional systems, from scanning and facilities management systems, from inbound and outbound customer contact points, from mobile media and the Web.

Space Operations KHANNA PUBLISHING

Bigdata is one of the most demanding markets in the IT sector. If you are an administrator or have a passion for knowing the internal configurations of Hadoop, then this book is for you. This book enables a professional to learn about Hadoop in terms of installation, configuration, and management. This book will help the reader to jumpstart with Hadoop frameworks, its eco-system components and slowly progress towards learning the administration part of Hadoop. The level of this book goes from beginner to intermediate with 70% hands-on exercises. Some of the techniques that you will learn include, • Installation and configuration of Hadoop cluster • Performing Hadoop Cluster Upgrade • Understanding and implementing HDFS Federation • Understanding and Implementing High Availability • Implementing HA on a Federated Cluster • Zookeeper CLI • Apache Hive Installation and Security • HBase Multi-master setup • Oozie installation, configuration and job submission • Setting up HDFS Quotas • Setting up HDFS NFS gateway • Understanding and implementing rolling upgrade and much more.

Data Analytics with Hadoop VMWare Press

This book constitutes the proceedings of the Third Technology Conference on Performance Evaluation and Benchmarking, TPCTC 2011, held in conjunction with the 37th International Conference on Very Large Data Bases, VLDB 2011, in Seattle, August/September 2011. The 12 full papers and 2 keynote papers were carefully selected and reviewed from numerous submissions. The papers present novel ideas and methodologies in performance evaluation, measurement, and characterization. *Big data processing at scale to unlock unique business insights* O'Reilly Media

There's a lot of information about big data technologies, but splicing these technologies into an end-to-end enterprise data platform is a daunting task not widely covered. With this practical book, you'll learn how to build big data infrastructure both on-premises and in the cloud and successfully architect a modern data platform. Ideal for enterprise architects, IT managers, application architects, and data engineers, this book shows you how to overcome the many challenges that emerge during Hadoop projects. You'll explore the vast landscape of tools available in the Hadoop and big data realm in a thorough technical primer before diving into: Infrastructure: Look at all component layers in a modern data platform, from the server to

the data center, to establish a solid foundation for data in your enterprise Platform: Understand aspects of deployment, operation, security, high availability, and disaster recovery, along with everything you need to know to integrate your platform with the rest of your enterprise IT Taking Hadoop to the cloud: Learn the important architectural aspects of running a big data platform in the cloud while maintaining enterprise security and high availability

Addison-Wesley Professional

Perform forensic investigations on Hadoop clusters with cutting-edge tools and techniques About This Book Identify, collect, and analyze Hadoop evidence forensically Learn about Hadoop's internals and Big Data file storage concepts A step-by-step guide to help you perform forensic analysis using freely available tools Who This Book Is For This book is meant for statisticians and forensic analysts with basic knowledge of digital forensics. They do not need to know Big Data Forensics. If you are an IT professional, law enforcement professional, legal professional, or a student interested in Big Data and forensics, this book is the perfect hands-on guide for learning how to conduct Hadoop forensic investigations. Each topic and step in the forensic process is described in accessible language. What You Will Learn Understand Hadoop internals and file storage Collect and analyze Hadoop forensic evidence Perform complex forensic analysis for fraud and other investigations Use state-of-the-art forensic tools Conduct interviews to identify Hadoop evidence Create compelling presentations of your forensic findings Understand how Big Data clusters operate Apply advanced forensic techniques in an investigation, including file carving, statistical analysis, and more In Detail Big Data forensics is an important type of digital investigation that involves the identification, collection, and analysis of large-scale Big Data systems. Hadoop is one of the most popular Big Data solutions, and forensically investigating a Hadoop cluster requires specialized tools and techniques. With the explosion of Big Data, forensic investigators need to be prepared to analyze the petabytes of data stored in Hadoop clusters. Understanding Hadoop's operational structure and performing forensic analysis with court-accepted tools and best practices will help you conduct a successful investigation. Discover how to perform a complete forensic investigation of large-scale Hadoop clusters using the same tools and techniques

employed by forensic experts. This book begins by taking you through the process of forensic investigation and the pitfalls to avoid. It will walk you through Hadoop's internals and architecture, and you will discover what types of information Hadoop stores and how to access that data. You will learn to identify Big Data evidence using techniques to survey a live system and interview witnesses. After setting up your own Hadoop system, you will collect evidence using techniques such as forensic imaging and application-based extractions. You will analyze Hadoop evidence using advanced tools and techniques to uncover events and statistical information. Finally, data visualization and evidence presentation techniques are covered to help you properly communicate your findings to any audience. Style and approach This book is a complete guide that follows every step of the forensic analysis process in detail. You will be guided through each key topic and step necessary to perform an investigation. Hands-on exercises are presented throughout the book, and technical reference guides and sample documents are included for real-world use.

An Introduction for Data Scientists Packt Publishing Ltd

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. *Cloudera Administration Handbook* Packt Publishing Ltd Do you want to broaden your Hadoop skill set and take your knowledge to the next level? Do you wish to enhance your knowledge of Hadoop to solve challenging data processing problems? Are your Hadoop jobs, Pig scripts, or Hive queries not working as fast as you intend? Are you looking to understand the benefits of upgrading Hadoop? If the answer is yes to any of these, this book is for you. It assumes novice-level familiarity with Hadoop.

Best Sellers - Books :

- [The Inmate: A Gripping Psychological Thriller By Freida Mcfadden](#)
- [Demon Copperhead: A Pulitzer Prize Winner By Barbara Kingsolver](#)
- [I Love You To The Moon And Back](#)
- [The Untethered Soul: The Journey Beyond Yourself By Michael A. Singer](#)
- [Brown Bear, Brown Bear, What Do You See? By Bill Martin Jr.](#)
- [Saved: A War Reporter's Mission To Make It Home](#)
- [How To Win Friends & Influence People \(dale Carnegie Books\)](#)
- [The Five-star Weekend By Elin Hilderbrand](#)
- [Chicka Chicka Boom Boom \(board Book\)](#)
- [Can't Hurt Me: Master Your Mind And Defy The Odds](#)