

# Single Node Setup Apache Hadoop

[Getting Started with Impala](#)  
[Hadoop For Dummies](#)  
[Machine Learning with Apache Spark Quick Start Guide](#)  
[Data Analytics with Hadoop](#)  
[Mastering Python for Finance](#)  
[R and Data Mining](#)  
[Hadoop Operations](#)  
[Beginning Apache Hadoop Administration](#)  
[Professional Hadoop](#)  
[Apache Hadoop YARN](#)  
[Information and Communication Technology for Sustainable Development](#)  
[Web 2.0 Fundamentals: With AJAX, Development Tools, and Mobile Platforms](#)  
[Big Data with Hadoop MapReduce](#)  
[Expert Hadoop 2 Administration](#)  
[Large-Scale Graph Processing Using Apache Giraph](#)  
[Big Data Processing With Hadoop](#)  
[Big Data Architect's Handbook](#)  
[Hadoop 2.x Administration Cookbook](#)  
[Mahout in Action](#)  
[Hadoop: The Definitive Guide](#)  
[Mastering MongoDB 6.x](#)  
[Apache Hadoop YARN](#)  
[Innovations in Electronics and Communication Engineering](#)  
[Hadoop 2 Quick-Start Guide](#)  
[Dependability in Sensor, Cloud, and Big Data Systems and Applications](#)  
[Apache Flume: Distributed Log Collection for Hadoop - Second Edition](#)  
[Apache Hadoop 3 Quick Start Guide](#)  
[Optimizing Hadoop for MapReduce](#)  
[Apache Spark in 24 Hours, Sams Teach Yourself](#)  
[Big Data Made Easy](#)  
[Scaling Big Data with Hadoop and Solr - Second Edition](#)  
[Hadoop: The Definitive Guide](#)  
[Information and Communication Technology for Intelligent Systems \(ICTIS 2017\) - Volume 1](#)  
[Practical Hadoop Ecosystem](#)  
[Hadoop Cluster Deployment](#)  
[Beginning Apache Cassandra Development](#)  
[Distributed Computing and Internet Technology](#)  
[Practical Machine Learning](#)  
[Hadoop in Action](#)  
[Big Data Analytics with Hadoop 3](#)

*Single Node Setup Apache Hadoop*

Downloaded from [process.ogleschool.edu](https://process.ogleschool.edu) by guest

## ZION HORTON

[Getting Started with Impala](#) Addison-Wesley Professional

R and Data Mining introduces researchers, post-graduate students, and analysts to data mining using R, a free software environment for statistical computing and graphics. The book provides practical methods for using R in applications from academia to industry to extract knowledge from vast amounts of data. Readers will find this book a valuable guide to the use of R in tasks such as classification and prediction, clustering, outlier detection, association rules, sequence analysis, text mining, social network analysis, sentiment analysis, and more. Data mining techniques are growing in popularity in a broad range of areas, from banking to insurance, retail, telecom, medicine, research, and government. This book focuses on the modeling phase of the data mining process, also addressing data exploration and model evaluation. With three in-depth case studies, a quick reference guide, bibliography, and links to a wealth of online resources, R and Data Mining

is a valuable, practical guide to a powerful method of analysis. Presents an introduction into using R for data mining applications, covering most popular data mining techniques Provides code examples and data so that readers can easily learn the techniques Features case studies in real-world applications to help readers apply the techniques in their work

*Hadoop For Dummies* Packt Publishing Ltd

This book constitutes the refereed proceedings of the 5th International Conference on Dependability in Sensor, Cloud, and Big Data Systems and Applications, DependSys, held in Guangzhou, China, in November 2019. The volume presents 39 full papers, which were carefully reviewed and selected from 112 submissions. The papers are organized in topical sections on dependability and security fundamentals and technologies; dependable and secure systems; dependable and secure applications; dependability and security measures and assessments; explainable artificial intelligence for cyberspace.

**Machine Learning with Apache Spark Quick Start Guide** Apress

A fast paced guide that will help you learn about Apache Hadoop 3 and its ecosystem Key

FeaturesSet up, configure and get started with Hadoop to get useful insights from large data setsWork with the different components of Hadoop such as MapReduce, HDFS and YARN Learn about the new features introduced in Hadoop 3Book 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.

**Data Analytics with Hadoop** Packt Publishing Ltd

Combine advanced analytics including Machine Learning, Deep Learning Neural Networks and Natural Language Processing with modern scalable technologies including Apache Spark to derive actionable insights from Big Data in real-time Key Features Make a hands-on start in the fields of Big Data, Distributed Technologies and Machine Learning Learn how to design, develop and interpret the results of common Machine Learning algorithms Uncover hidden patterns in your data in order to derive real actionable insights and business value Book Description Every person and every organization in the world manages data, whether they realize it or not. Data is used to describe the world around us and can be used for almost any purpose, from analyzing consumer habits to fighting disease and serious organized crime. Ultimately, we manage data in order to derive value from it, and many organizations around the world have traditionally invested in technology to help process their data faster and more efficiently. But we now live in an interconnected world driven by mass data creation and consumption where data is no longer rows and columns restricted to a spreadsheet, but an organic and evolving asset in its own right. With this realization comes major challenges for organizations: how do we manage the sheer size of data being created every second (think not only spreadsheets and databases, but also social media posts, images, videos, music, blogs and so on)? And once we can manage all of this data, how do we derive real value from it? The focus of Machine Learning with Apache Spark is to help us answer these questions in a hands-on manner. We introduce the latest scalable technologies to help us manage and process big data. We then introduce advanced analytical algorithms applied to real-world use cases in order to uncover patterns, derive actionable insights, and learn from this big data. What you will learn Understand how Spark fits in the context of the big data ecosystem Understand how to deploy and configure a local development environment using Apache Spark Understand how to design supervised and unsupervised learning models Build models to perform NLP, deep learning, and cognitive services using Spark ML libraries Design real-time machine learning pipelines in Apache Spark Become familiar with advanced techniques for processing a large volume of data by applying machine learning algorithms Who this book is for This book is aimed at Business Analysts, Data Analysts and Data Scientists who wish to make a hands-on start in order to take advantage of modern Big Data technologies combined with Advanced Analytics.

**Mastering Python for Finance** Packt Publishing Ltd

This book is a step-by-step tutorial filled with practical examples which will show you how to build and manage a Hadoop cluster along with its intricacies. This book is ideal for database administrators, data engineers, and system administrators, and it will act as an invaluable reference if you are planning to use the Hadoop platform in your organization. It is expected that you have basic Linux skills since all the examples in this book use this operating system. It is also useful if you have access to test hardware or virtual machines to be able to follow the examples in the book.

**R and Data Mining** Sams Publishing

Over 100 practical recipes to help you become an expert Hadoop administrator About This Book Become an expert Hadoop administrator and perform tasks to optimize your Hadoop Cluster Import and export data into Hive and use Oozie to manage workflow. Practical recipes will help you plan and secure your Hadoop cluster, and make it highly available Who This Book Is For If you are a system administrator with a basic understanding of Hadoop and you want to get into Hadoop administration, this book is for you. It's also ideal if you are a Hadoop administrator who wants a quick reference guide to all the Hadoop administration-related tasks and solutions to commonly occurring problems What You Will Learn Set up the Hadoop architecture to run a Hadoop cluster smoothly Maintain a Hadoop cluster on HDFS, YARN, and MapReduce Understand high availability with Zookeeper and Journal Node Configure Flume for data ingestion and Oozie to run various

workflows Tune the Hadoop cluster for optimal performance Schedule jobs on a Hadoop cluster using the Fair and Capacity scheduler Secure your cluster and troubleshoot it for various common pain points In Detail Hadoop enables the distributed storage and processing of large datasets across clusters of computers. Learning how to administer Hadoop is crucial to exploit its unique features. With this book, you will be able to overcome common problems encountered in Hadoop administration. The book begins with laying the foundation by showing you the steps needed to set up a Hadoop cluster and its various nodes. You will get a better understanding of how to maintain Hadoop cluster, especially on the HDFS layer and using YARN and MapReduce. Further on, you will explore durability and high availability of a Hadoop cluster. You'll get a better understanding of the schedulers in Hadoop and how to configure and use them for your tasks. You will also get hands-on experience with the backup and recovery options and the performance tuning aspects of Hadoop. Finally, you will get a better understanding of troubleshooting, diagnostics, and best practices in Hadoop administration. By the end of this book, you will have a proper understanding of working with Hadoop clusters and will also be able to secure, encrypt it, and configure auditing for your Hadoop clusters. Style and approach This book contains short recipes that will help you run a Hadoop cluster efficiently. The recipes are solutions to real-life problems that administrators encounter while working with a Hadoop cluster

**Hadoop Operations** Springer

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.

**Beginning Apache Hadoop Administration** Packt Publishing Ltd

A comprehensive end-to-end guide that gives hands-on practice in big data and Artificial Intelligence Key Features Learn to build and run a big data application with sample code Explore examples to implement activities that a big data architect performs Use Machine Learning and AI for structured and unstructured data Book Description The big data architects are the "masters" of data, and hold high value in today's market. Handling big data, be it of good or bad quality, is not an easy task. The prime job for any big data architect is to build an end-to-end big data solution that integrates data from different sources and analyzes it to find useful, hidden insights. Big Data Architect's Handbook takes you through developing a complete, end-to-end big data pipeline, which will lay the foundation for you and provide the necessary knowledge required to be an architect in big data. Right from understanding the design considerations to implementing a solid, efficient, and scalable data pipeline, this book walks you through all the essential aspects of big data. It also gives you an overview of how you can leverage the power of various big data tools such as Apache Hadoop and ElasticSearch in order to bring them together and build an efficient big data solution. By the end of this book, you will be able to build your own design system which integrates, maintains, visualizes, and monitors your data. In addition, you will have a smooth design flow in each process, putting insights in action. What you will learn Learn Hadoop Ecosystem and Apache projects Understand, compare NoSQL database and essential software architecture Cloud infrastructure design considerations for big data Explore application scenario of big data tools for daily activities Learn to analyze and visualize results to uncover valuable insights Build and run a big data application with sample code from end to end Apply Machine Learning and AI to perform big data intelligence Practice the daily activities performed by big data architects Who this book is for Big Data Architect's Handbook is for you if you are an aspiring data professional, developer, or IT enthusiast who aims to be an all-round architect in big data. This book is your one-stop solution to enhance your knowledge and carry out easy to complex activities required to become a big data architect.

**Professional Hadoop** "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

**Apache Hadoop YARN** "O'Reilly Media, Inc."

Summary Mahout in Action is a hands-on introduction to machine learning with Apache Mahout. Following real-world examples, the book presents practical use cases and then illustrates how Mahout can be applied to solve them. Includes a free audio- and video-enhanced ebook. About the Technology A computer system that learns and adapts as it collects data can be really powerful. Mahout, Apache's open source machine learning project, captures the core algorithms of recommendation systems, classification, and clustering in ready-to-use, scalable libraries. With Mahout, you can immediately apply to your own projects the machine learning techniques that drive Amazon, Netflix, and others. About this Book This book covers machine learning using Apache Mahout. Based on experience with real-world applications, it introduces practical use cases and illustrates how Mahout can be applied to solve them. It places particular focus on issues of scalability and how to apply these techniques against large data sets using the Apache Hadoop framework. This book is written for developers familiar with Java -- no prior experience with Mahout is assumed. Owners of a Manning eBook purchased anywhere in the world can download a free eBook from manning.com at any time. They can do so multiple times and in any or all formats available (PDF, ePub or Kindle). To do so, customers must register their printed copy on Manning's site by creating a user account and then following instructions printed on the eBook registration insert at the front of the book. What's Inside Use group data to make individual recommendations Find logical clusters within your data Filter and refine with on-the-fly classification Free audio and video extras Table of Contents Meet Apache Mahout PART 1 RECOMMENDATIONS Introducing recommenders Representing recommender data Making recommendations Taking recommenders to production Distributing recommendation computations PART 2 CLUSTERING Introduction to clustering Representing data Clustering algorithms in Mahout Evaluating and improving clustering quality Taking clustering to production Real-world applications of clustering PART 3 CLASSIFICATION Introduction to classification Training a classifier Evaluating and tuning a classifier Deploying a classifier Case study: Shop It To Me

**Information and Communication Technology for Sustainable Development** "O'Reilly Media, Inc."

Get Started Fast with Apache Hadoop® 2, YARN, and Today's Hadoop Ecosystem With Hadoop 2.x and YARN, Hadoop moves beyond MapReduce to become practical for virtually any type of data processing. Hadoop 2.x and the Data Lake concept represent a radical shift away from conventional approaches to data usage and storage. Hadoop 2.x installations offer unmatched scalability and breakthrough extensibility that supports new and existing Big Data analytics processing methods and models. Hadoop® 2 Quick-Start Guide is the first easy, accessible guide to Apache Hadoop 2.x, YARN, and the modern Hadoop ecosystem. Building on his unsurpassed experience teaching Hadoop and Big Data, author Douglas Eadline covers all the basics you need to know to install and use Hadoop 2 on personal computers or servers, and to navigate the powerful technologies that complement it. Eadline concisely introduces and explains every key Hadoop 2 concept, tool, and service, illustrating each with a simple "beginning-to-end" example and identifying trustworthy, up-to-date resources for learning more. This guide is ideal if you want to learn about Hadoop 2 without getting mired in technical details. Douglas Eadline will bring you up to speed quickly, whether you're a user, admin, devops specialist, programmer, architect, analyst, or data scientist. Coverage Includes Understanding what Hadoop 2 and YARN do, and how they improve on Hadoop 1 with MapReduce Understanding Hadoop-based Data Lakes versus RDBMS Data Warehouses Installing Hadoop 2 and core services on Linux machines, virtualized sandboxes, or clusters Exploring the Hadoop Distributed File System (HDFS) Understanding the essentials of MapReduce and YARN application programming Simplifying programming and data movement with Apache Pig, Hive, Sqoop, Flume, Oozie, and HBase Observing application progress, controlling jobs, and managing workflows Managing Hadoop efficiently with Apache Ambari—including recipes for HDFS to NFSv3 gateway, HDFS snapshots, and YARN configuration Learning basic Hadoop 2 troubleshooting, and installing Apache Hue and Apache Spark

**Web 2.0 Fundamentals: With AJAX, Development Tools, and Mobile Platforms** Simon and

Schuster

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 with Hadoop MapReduce](#) Springer

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.

*Expert Hadoop 2 Administration* Apress

This volume includes 74 papers presented at ICTIS 2017: Second International Conference on Information and Communication Technology for Intelligent Systems. The conference was held on

25th and 26th March 2017, in Ahmedabad, India and organized jointly by the Associated Chambers of Commerce and Industry of India (ASSOCHAM) Gujarat Chapter, the G R Foundation, the Association of Computer Machinery, Ahmedabad Chapter and supported by the Computer Society of India Division IV - Communication and Division V - Education and Research. The papers featured mainly focus on information and communications technology (ICT) for computation, algorithms and data analytics. The fundamentals of various data analytics and algorithms discussed are useful to researchers in the field.

[Large-Scale Graph Processing Using Apache Giraph](#) Packt Publishing Ltd

This book is aimed at developers, designers, and architects who would like to build big data enterprise search solutions for their customers or organizations. No prior knowledge of Apache Hadoop and Apache Solr/Lucene technologies is required.

**Big Data Processing With Hadoop** Jones & Bartlett Learning

The book proposes new technologies and discusses future solutions for design infrastructure for ICT. The book contains high quality submissions presented at Second International Conference on Information and Communication Technology for Sustainable Development (ICT4SD - 2016) held at Goa, India during 1 - 2 July, 2016. The conference stimulates the cutting-edge research discussions among many academic pioneering researchers, scientists, industrial engineers, and students from all around the world. The topics covered in this book also focus on innovative issues at international level by bringing together the experts from different countries.

[Big Data Architect's Handbook](#) John Wiley & Sons

Designed for a broad spectrum of people with technically diverse backgrounds, this book covers the most recent developments in Web 2.0 programming topics and applications, including up-to-date material on cloud computing, Google AppEngine, Social Networks, Comet, HTML5, semantic technology, and a chapter on the future of the Web. This book prepares readers for more advanced technical topics in Web 2.0. The accompanying CD-ROM and companion website provide code samples from the book and appendices with an extensive set of links (over 1,000) for supplemental material and links for the Twitter and Facebook pages. (Please note, eBook version does not include CD-ROM).

[Hadoop 2.x Administration Cookbook](#) Addison-Wesley Professional

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

*Mahout in Action* Springer

Many corporations are finding that the size of their data sets are outgrowing the capability of their systems to store and process them. The data is becoming too big to manage and use with traditional tools. The solution: implementing a big data system. As *Big Data Made Easy: A Working Guide to the Complete Hadoop Toolset* shows, Apache Hadoop offers a scalable, fault-tolerant system for storing and processing data in parallel. It has a very rich toolset that allows for storage (Hadoop), configuration (YARN and ZooKeeper), collection (Nutch and Solr), processing (Storm, Pig, and Map Reduce), scheduling (Oozie), moving (Sqoop and Avro), monitoring (Chukwa, Ambari, and Hue), testing (Big Top), and analysis (Hive). The problem is that the Internet offers IT pros wading into big data many versions of the truth and some outright falsehoods born of ignorance. What is needed is a book just like this one: a wide-ranging but easily understood set of instructions to explain where to get Hadoop tools, what they can do, how to install them, how to configure them, how to integrate them, and how to use them successfully. And you need an expert who has worked in this area for a decade—someone just like author and big data expert Mike Frampton. *Big Data Made Easy* approaches the problem of managing massive data sets from a systems perspective, and it explains the roles for each project (like architect and tester, for example) and shows how the Hadoop toolset can be used at each system stage. It explains, in an easily understood manner and through numerous examples, how to use each tool. The book also explains the sliding scale of tools available depending upon data size and when and how to use them. *Big Data Made Easy* shows developers and architects, as well as testers and project managers, how to: Store big data Configure big data Process big data Schedule processes Move data among SQL and NoSQL systems Monitor data Perform big data analytics Report on big data processes and projects Test big data systems *Big Data Made Easy* also explains the best part, which is that this toolset is free. Anyone can download it and—with the help of this book—start to use it within a day. With the skills this book will teach you under your belt, you will add value to your company or client immediately, not to mention your career.

*Hadoop: The Definitive Guide* IGI Global

Bigdata is one of the most demanding markets in the IT sector. If you are an administrator or a 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.

Best Sellers - Books :

- [America's Cultural Revolution: How The Radical Left Conquered Everything By Christopher F. Rufo](#)
- [Never Never: A Romantic Suspense Novel Of Love And Fate](#)
- [The Very Hungry Caterpillar By Eric Carle](#)
- [The Inmate: A Gripping Psychological Thriller By Freida Mcfadden](#)
- [How To Catch A Leprechaun](#)
- [Ugly Love: A Novel](#)
- [Feel-good Productivity: How To Do More Of What Matters To You By Ali Abdaal](#)
- [November 9: A Novel](#)
- [Things We Never Got Over \(knockemout\) By Lucy Score](#)
- [Remarkably Bright Creatures: A Read With Jenna Pick By Shelby Van Pelt](#)