
Hadoop Administration Guide

Thank you very much for reading **Hadoop Administration Guide**. Maybe you have knowledge that, people have look numerous times for their favorite novels like this Hadoop Administration Guide, but end up in malicious downloads.

Rather than enjoying a good book with a cup of tea in the afternoon, instead they cope with some harmful bugs inside their laptop.

Hadoop Administration Guide is available in our book collection an online access to it is set as public so you can download it instantly. Our book servers spans in multiple locations, allowing you to get the most less latency time to download any of our books like this one. Merely said, the Hadoop Administration Guide is universally compatible with any devices to read

Downloaded
from
Hadoop Administration srm.nwherald.com
Guide by guest

**ZIMMERMA
N FINLEY**

*Expert
Hadoop 2
Administration*

Apress
"Apache
Hadoop is
helping drive
the Big Data
revolution.
Now, its data
processing

has been
completely
overhauled:
Apache
Hadoop YARN
provides
resource
management

at data center scale and easier ways to create distributed applications that process petabytes of data. And now in Apache Hadoop™ YARN, two Hadoop technical leaders show you how to develop new applications and adapt existing code to fully leverage these revolutionary advances." -- From the Amazon

Hbase Administration Cookbook
Expert Hadoop 2

Administration Managing Spark, YARN, and MapReduce
As part of Packt's cookbook series, each recipe offers a practical, step-by-step solution to common problems found in HBase administration. This book is for HBase administrators, developers, and will even help Hadoop administrators. You are not required to have HBase experience, but are expected to have a basic

understanding of Hadoop and MapReduce.
Hadoop Operations
Packt Publishing Ltd
Let Hadoop For Dummies help harness the power of your data and rein in the information overload Big data has become big business, and companies and organizations of all sizes are struggling to find ways to retrieve valuable information from their massive data sets with becoming overwhelmed.

Enter Hadoop and this easy-to-understand For Dummies guide. Hadoop For Dummies helps readers understand the value of big data, make a business case for using Hadoop, navigate the Hadoop ecosystem, and build and manage Hadoop applications and clusters. Explains the origins of Hadoop, its economic benefits, and its functionality and practical applications. Helps you find

your way around the Hadoop ecosystem, program MapReduce, utilize design patterns, and get your Hadoop cluster up and running quickly and easily. Details how to use Hadoop applications for data mining, web analytics and personalization, large-scale text processing, data science, and problem-solving. Shows you how to improve the value of your Hadoop cluster,

maximize your investment in Hadoop, and avoid common pitfalls when building your Hadoop cluster. From programmers challenged with building and maintaining affordable, scalable 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. [Hadoop 2 Quick-Start Guide](#) "O'Reilly Media, Inc."

If you've been asked to maintain large and complex Hadoop clusters, this book is a must. Demand for operations-specific material has skyrocketed now that Hadoop is becoming the de facto standard for truly large-scale data processing in the data center. Eric Sammer, Principal Solution Architect at Cloudera, shows you the particulars of running Hadoop in production,

from planning, installing, and configuring the system to providing ongoing maintenance. Rather than run through all possible scenarios, this pragmatic operations guide calls out what works, as demonstrated in critical deployments. Get a high-level overview of HDFS and MapReduce: why they exist and how they work Plan a Hadoop deployment, from hardware and OS selection to network

requirements Learn setup and configuration details with a list of critical properties Manage resources by sharing a cluster across multiple groups Get a runbook of the most common cluster maintenance tasks Monitor Hadoop clusters—and learn troubleshooting with the help of real-world war stories Use basic tools and techniques to handle backup and catastrophic

failure
**Cloudera
Administrati
on Handbook**
Packt
Publishing Ltd
If your
organization is
about to enter
the world of
big data, you
not only need
to decide
whether
Apache
Hadoop is the
right platform
to use, but
also which of
its many
components
are best
suited to your
task. This field
guide makes
the exercise
manageable
by breaking
down the
Hadoop
ecosystem
into short,

digestible
sections.
You'll quickly
understand
how Hadoop's
projects,
subprojects,
and related
technologies
work together.
Each chapter
introduces a
different
topic—such as
core
technologies
or data
transfer—and
explains why
certain
components
may or may
not be useful
for particular
needs. When
it comes to
data, Hadoop
is a whole new
ballgame, but
with this
handy
reference,

you'll have a
good grasp of
the playing
field. Topics
include: Core
technologies—
Hadoop
Distributed
File System
(HDFS),
MapReduce,
YARN, and
Spark
Database and
data
management
—Cassandra,
HBase,
MongoDB, and
Hive
Serialization—
Avro, JSON,
and Parquet
Management
and
monitoring—P
uppet, Chef,
Zookeeper,
and Oozie
Analytic
helpers—Pig,
Mahout, and

MLLib Data transfer—Scoop, Flume, distcp, and Storm Security, access control, auditing—Sentry, Kerberos, and Knox Cloud computing and virtualization—Serengeti, Docker, and Whirr

Hadoop Operations

Apress
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
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.
Storage and Analysis at Internet Scale
Packt Publishing Ltd
Describes the features and functions of Apache Hive, the data infrastructure for Hadoop.
Hadoop For Dummies
Packt Publishing Ltd
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 <i>Big data processing at scale to unlock unique business insights</i> Apress Perform fast interactive analytics against different data sources using the Trino high-performance distributed SQL query engine. With this practical guide, you'll learn how to conduct analytics on data where it lives, whether it's Hive, Cassandra, a relational database, or a proprietary data store.
---	---	--

Analysts, software engineers, and production engineers will learn how to manage, use, and even develop with Trino. Initially developed by Facebook, open source Trino is now used by Netflix, Airbnb, LinkedIn, Twitter, Uber, and many other companies. Matt Fuller, Manfred Moser, and Martin Traverso show you how a single Trino query can combine data

from multiple sources to allow for analytics across your entire organization. Get started: Explore Trino's use cases and learn about tools that will help you connect to Trino and query data Go deeper: Learn Trino's internal workings, including how to connect to and query data sources with support for SQL statements, operators, functions, and more Put Trino in production: Secure Trino,

monitor workloads, tune queries, and connect more applications; learn how other organizations apply Trino [Learn Azure in a Month of Lunches, Second Edition](#) "O'Reilly Media, Inc." 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—some one 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. "O'Reilly Media, Inc." Learn Azure in a Month of Lunches, Second Edition, is a tutorial on

writing, deploying, and running applications in Azure. In it, you'll work through 21 short lessons that give you real-world experience. Each lesson includes a hands-on lab so you can try out and lock in your new skills. Summary You can be incredibly productive with Azure without mastering every feature, function, and service. Learn Azure in a Month of Lunches, Second

Edition gets you up and running quickly, teaching you the most important concepts and tasks in 21 practical bite-sized lessons. As you explore the examples, exercises, and labs, you'll pick up valuable skills immediately and take your first steps to Azure mastery! This fully revised new edition covers core changes to the Azure UI, new Azure features, Azure containers, and the

upgraded Azure Kubernetes Service. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Microsoft Azure is vast and powerful, offering virtual servers, application templates, and prebuilt services for everything from data storage to AI. To navigate it all, you need a trustworthy guide. In this book, Microsoft

engineer and Azure trainer
Iain Foulds focuses on core skills for creating cloud-based applications. About the book *Learn Azure in a Month of Lunches, Second Edition*, is a tutorial on writing, deploying, and running applications in Azure. In it, you'll work through 21 short lessons that give you real-world experience. Each lesson includes a hands-on lab so you can try out and lock in your new skills. What's inside

Understanding Azure beyond point-and-click
Securing applications and data
Automating your environment
Azure services for machine learning, containers, and more
About the reader This book is for readers who can write and deploy simple web or client/server applications. About the author Iain Foulds is an engineer and senior content developer

with Microsoft.
Table of Contents PART 1 - AZURE CORE SERVICES 1
Before you begin 2
Creating a virtual machine 3
Azure Web Apps 4
Introduction to Azure Storage 5
Azure Networking basics PART 2 - HIGH AVAILABILITY AND SCALE 6
Azure Resource Manager 7
High availability and redundancy 8
Load-balancing applications 9
Applications

that scale 10
 Global
 databases
 with Cosmos
 DB 11
 Managing
 network traffic
 and routing 12
 Monitoring
 and
 troubleshooting
 PART 3 -
 SECURE BY
 DEFAULT 13
 Backup,
 recovery, and
 replication 14
 Data
 encryption 15
 Securing
 information
 with Azure
 Key Vault 16
 Azure Security
 Center and
 updates PART
 4 - THE COOL
 STUFF 17
 Machine
 learning and
 artificial
 intelligence 18

Azure
 Automation 19
 Azure
 containers 20
 Azure and the
 Internet of
 Things 21
 Serverless
 computing
**A Guide for
 Developers
 and
 Administrato
 rs** Packt
 Publishing Ltd
 If your
 organization is
 looking for a
 storage
 solution to
 accommodate
 a virtually
 endless
 amount of
 data, this
 book will show
 you how
 Apache HBase
 can fulfill your
 needs. As the
 open source
 implementatio

n of Google's
 BigTable
 architecture,
 HBase scales
 to billions of
 rows and
 millions of
 columns,
 while ensuring
 that write and
 read
 performance
 remain
 constant.HBas
 e: The
 Definitive
 Guideprovides
 the details you
 require,
 whether you
 simply want to
 evaluate this
 high-
 performance,
 non-relational
 database, or
 put it into
 practice right
 away. HBase's
 adoption rate
 is beginning to
 climb, and

several IT executives are asking pointed questions about this high-capacity database. This is the only book available to give you meaningful answers. Learn how to distribute large datasets across an inexpensive cluster of commodity servers. Develop HBase clients in many programming languages, including Java, Python, and Ruby. Get details on HBase's primary storage

system, HDFS—Hadoop's distributed and replicated filesystem. Learn how HBase's native interface to Hadoop's MapReduce framework enables easy development and execution of batch jobs that can scan entire tables. Discover the integration between HBase and other facets of the Apache Hadoop project. *Hadoop in Practice* "O'Reilly Media, Inc." Learn SAS® administration from the

ground up! Those who are new to SAS platform administration may find themselves full of questions. SAS® Administration from the Ground Up: Running the SAS® 9 Platform in a Metadata Server Environment will save you time, money and frustration. This book walks the reader through setting up and maintaining a SAS platform from scratch. The author

includes tips on best practices and troubleshooting to show you simple ways to streamline your SAS environment and make your work more manageable. Written for both new administrators and seasoned professionals, this book covers: SAS® 9.4 architecture SAS administration tools such as SAS® Management Console, SAS® Environment Manager and SAS®

Deployment Manager Users, groups, and roles Metadata library administration Security Also included is a master administration checklist, with helpful resources provided for each task. **Scaling Big Data with Hadoop and Solr** "O'Reilly Media, Inc." 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. [Apache Sqoop Cookbook](#) Morgan & Claypool Publishers This book is a step-by-step tutorial that will enable you to leverage the flexible search functionality of Apache Solr together with the Big Data

power of Apache Hadoop. Scaling Big Data with Hadoop and Solr provides guidance to developers who wish to build high-speed enterprise search platforms using Hadoop and Solr. This book is primarily aimed at Java programmers who wish to extend the Hadoop platform to make it run as an enterprise search without any prior knowledge of Apache Hadoop and

Solr. [Hadoop: The Definitive Guide](#) John Wiley & Sons 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 Security

Manning
Publications
Data is
arriving faster
than you can
process it and
the overall
volumes keep
growing at a
rate that
keeps you
awake at
night. Hadoop
can help you
tame the data
beast.
Effective use
of Hadoop
however
requires a
mixture of
programming,
design, and
system
administration
skills. "Hadoop
Beginner's
Guide"

removes the mystery from Hadoop, presenting Hadoop and related technologies with a focus on building working systems and getting the job done, using cloud services to do so when it makes sense. From basic concepts and initial setup through developing applications and keeping the system running as the data grows, the book gives the understanding needed to effectively use Hadoop to

solve real world problems. Starting with the basics of installing and configuring Hadoop, the book explains how to develop applications, maintain the system, and how to use additional products to integrate with other systems. While learning different ways to develop applications to run on Hadoop the book also covers tools such as Hive, Sqoop, and Flume that show how Hadoop can be integrated

with relational databases and log collection. In addition to examples on Hadoop clusters on Ubuntu uses of cloud services such as Amazon, EC2 and Elastic MapReduce are covered. *Running the SAS9 Platform in a Metadata Server Environment* "O'Reilly Media, Inc." 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. [Programming Hive](#) Packt Pub Limited Learn to design, build, and manage your infrastructure

<p>on the most popular of all the Cloud platforms—Amazon Web Services About This Book Learn how to leverage various Amazon Web Services (AWS) components and services to build a secure, reliable, and robust environment to host your applications on Deep dive into the core AWS service offerings with hands-on tutorials, real-world use case scenarios, and</p>	<p>best practices A self-paced, systematic, and step-by-step guide to learning and implementing AWS in your own environment Who This Book Is For This book is for those who want to learn and leverage AWS. Although no prior experience with AWS is required, it is recommended that you have some hands-on experience of Linux, Web Services, and basic networking What You Will Learn A brief</p>	<p>introduction to Cloud Computing and AWS accompanied by steps to sign up for your first AWS account Create and manage users, groups, and permissions using AWS Identity and Access Management services Get started with deploying and accessing EC2 instances, working with EBS Volumes and Snapshots Customize and create your very own Amazon Machine Image Design and deploy</p>
--	--	---

your instances on a highly secured, network isolated environment using Amazon VPC Effectively monitor your AWS environment using specialized alarms, custom monitoring metrics, and much more Explore the various benefits of Database-as-a-Service offerings and leverage them using Amazon RDS and Amazon DynamoDB Take an in-depth look at

what's new with AWS, including EC2 Container Service and Elastic File System In Detail AWS is at the forefront of Cloud Computing today. Many businesses are moving away from traditional datacenters and toward AWS because of its reliability, vast service offerings, lower costs, and high rate of innovation. Because of its versatility and flexible design, AWS can be used to

accomplish a variety of simple and complicated tasks such as hosting multitier websites, running large scale parallel processing, content delivery, petabyte storage and archival, and lots more. Whether you are a seasoned sysadmin or a rookie, this book will provide you with all the necessary skills to design, deploy, and manage your applications on the AWS

cloud platform. The book guides you through the core AWS services such as IAM, EC2, VPC, RDS, and S3 using a simple real world application hosting example that you can relate to. Each chapter is designed to provide you with the most information possible about a particular AWS service coupled with easy to follow hands-on steps, best practices, tips, and recommendations. By the

end of the book, you will be able to create a highly secure, fault tolerant, and scalable environment for your applications to run on. Style and approach This in-depth and insightful guide is filled with easy-to-follow examples, real-world use cases, best practices, and recommendations that will help you design and leverage AWS. Modern Big Data Processing with Hadoop Packt Publishing Ltd

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.