

# Books

## Hadoop Books

These books are listed in order of publication, most recent first. The Apache Software Foundation does not endorse any specific book. The links to Amazon are affiliated with the specific author. That said, we also encourage you to support your local bookshops, by buying the book from any local outlet, especially independent ones.

## Books in Print

Here are the books that are currently in print ~~in order of publishing~~, along with the Hadoop version they were written against. One problem anyone writing a book will encounter is that Hadoop is a very fast-moving target, and that things can change fast. Usually this is for the better, when a book says "Hadoop can't" they really mean "the version of Hadoop we worked with couldn't", and that the situation may have improved since then. If you have any query about Hadoop, don't be afraid to ask on the relevant user mailing lists.

```
{{{#!wiki comment/dotted Attention people adding new entries.
# Only reference books about Hadoop and related programs, not random PHP stuff.
# Please include publishing date and version of Hadoop the book is relevant to.
# Please write this in a neutral voice, not "this book will help you", as that implies that the ASF has opinions on the matter. Someone will just edit the claims out.
# Please do not go overboard in exaggerating the outcome of reading a book, "readers of this book will become experts in advanced production-scale Hadoop Algorithms". Such claims will be edited out and not replaced.
# Please don't have tracking URLs. We'll only cut them.
}}}}
```

## Hands-On Big Data Processing with Hadoop 3 (Video)

**Name:** [Hands-On Big Data Processing with Hadoop 3 \(Video\)](#)

**Author:** Sudhanshu Saxena

**Publisher:** Packt

**Date of Publishing:** October 2018

Perform real-time data analytics, stream and batch processing on your application using Hadoop

## Modern Big Data Processing with Hadoop

**Name:** [Modern Big Data Processing with Hadoop](#)

**Author:** V. Naresh Kumar, Prashant Shindgikar

**Publisher:** Packt

**Date of Publishing:** March 2018

A comprehensive guide to design, build and execute effective Big Data strategies using Hadoop

## Deep Learning with Hadoop

**Name:** [Deep Learning with Hadoop](#)

**Author:** Dipayan Dev

**Publisher:** Packt

**Date of Publishing:** February 2017

Build, implement and scale distributed deep learning models for large-scale datasets.

## Hadoop Blueprints

**Name:** [Hadoop Blueprints](#)

**Authors:** Anurag Shrivastava, Tanmay Deshpande

**Publisher:** Packt

**Date of Publishing:** September 2016

Use Hadoop to solve business problems by learning from a rich set of real-life case studies.

## Hadoop: Data Processing and Modelling

**Name:** [Hadoop: Data Processing and Modelling](#)

**Authors:** Garry Turkington, Tanmay Deshpande, Sandeep Karanth

**Publisher:** Packt

**Date of Publishing:** August 2016

Unlock the power of your data with Hadoop 2.X ecosystem and its data warehousing techniques across large data sets.

## Hadoop Explained (Free eBook Download)

**Name:** [Hadoop Explained](#)

**Author:** Aravind Shenoy

**Publisher:** Packt Publishing

Learn how [MapReduce](#) organizes and processes large sets of data and discover the advantages of Hadoop - from scalability to security, see how Hadoop handles huge amounts of data with care

## Hadoop Real-World Solutions Cookbook- Second Edition

**Name:** [Hadoop Real-World Solutions Cookbook- Second Edition](#)

**Author:** Tanmay Deshpande

**Publisher:** Packt Publishing

**Date of Publishing:** March 2016

The book covers recipes that are based on the latest versions of Apache Hadoop 2.X, YARN, Hive, Pig, Sqoop, Flume, Apache Spark, Mahout etc.

## Hadoop Security: Protecting Your Big Data Platform

**Name:** [Hadoop Security: Protecting Your Big Data Platform](#)

**Author:** Ben Spivey, Joey Echeverria

**Publisher:** O'Reilly Media

**Date of Publishing:** June 2015

Covers Hadoop security from a high level, down to how to set up a secure Hadoop cluster and the individual services within it.

## Hadoop and Kerberos: The Madness Beyond the Gate

**Name:** [Hadoop and Kerberos: The Madness Beyond the Gate](#)

**Author:** Steve Loughran

**Date of Publishing:** June, 2015 +

This is an ongoing ebook project attempting to cover the internals of Hadoop + Kerberos. It is targeted at developers and people trying to understand obscure kerberos-related stack traces.

## Apache Oozie Essentials

**Name:** [Apache Oozie Essentials](#)

**Author:** Jagat Jasjit Singh

**Publisher:** Packt Publishing

**Date of Publishing:** December, 2015

This book covers automating data and ML pipelines via Apache Oozie.

## Data Lake Development with Big Data

**Name:** [Data Lake Development with Big Data](#)

**Author:** Pradeep Pasupuleti, Beulah Salome Purra

**Publisher:** Packt Publishing

**Date of Publishing:** November, 2015

This book is for architects and senior managers building a strategy around their current data architecture, helping them identify the need for a Data Lake implementation in an enterprise context.

## Elasticsearch for Hadoop

**Name:** [Elasticsearch for Hadoop](#)

**Author:** Vishal Shukla

**Publisher:** Packt Publishing

**Date of Publishing:** October, 2015

Elasticsearch for Hadoop covers integrating Elasticsearch into Hadoop to visualize and analyze your data.

## YARN Essentials

**Name:** [YARN Essentials](#)

**Authors:** Amol Fasale, Nirmal Kumar

**Publisher:** Packt Publishing

**Date of Publishing:** February, 2015

YARN Essentials is for developers with little knowledge of Hadoop 1.x and want to start afresh with YARN.

## Learning YARN

**Name:** [Learning YARN](#)

**Authors:** Akhil Arora, Shrey Mehrotra

**Publisher:** Packt Publishing

**Date of Publishing:** August, 2015

Learning YARN is intended for those who want to understand what YARN is and how to efficiently use it for the resource management of large clusters.

## Big Data Forensics: Learning Hadoop Investigations

**Name:** [Big Data Forensics: Learning Hadoop Investigations](#)

**Author:** Joe Sremack

**Publisher:** Packt Publishing

**Date of Publishing:** August, 2015

Big Data Forensics: Learning Hadoop Investigations will guide statisticians and forensic analysts with basic knowledge of digital forensics to conduct Hadoop forensic investigations.

## Learning Hadoop 2

**Name:** [Learning Hadoop 2](#)

**Authors:** Garry Turkington, Gabriele Modena

**Publisher:** Packt Publishing

**Date of Publishing:** February, 2015

Learning Hadoop 2 is an introduction guide to building data-processing applications with the wide variety of tools supported by Hadoop 2.

## Hadoop **MapReduce** v2 Cookbook - Second Edition

**Name:** [Hadoop MapReduce v2 Cookbook - Second Edition](#)

**Authors:** Thilina Gunarathne

**Publisher:** Packt Publishing

**Date of Publishing:** February, 2015

Hadoop [MapReduce](#) v2 Cookbook - Second Edition is a beginner's guide to explore the Hadoop [MapReduce](#) v2 ecosystem to gain insights from very large datasets.

## Scaling Big Data with Hadoop and Solr - Second Edition

**Name:** [Scaling Big Data with Hadoop and Solr - Second Edition](#)

**Authors:** Hrishikesh Vijay Karambelkar

**Hadoop Version:** 2.6

**Publisher:** Packt Publishing

**Date of Publishing:** April, 2015

Scaling Big Data with Hadoop and Solr - Second Edition is aimed at developers, designers, and architects who would like to build big data enterprise search solutions for their customers or organizations

## Hadoop for Finance Essentials

**Name:** [Hadoop for Finance Essentials](#)

**Authors:** Rajiv Tiwari

**Publisher:** Packt Publishing

**Date of Publishing:** April, 2015

Hadoop for Finance Essentials is for developers who would like to perform big data analytics with Hadoop for the financial sector.

## Monitoring Hadoop

**Name:** [Monitoring Hadoop](#)

**Authors:** Gurmukh Singh

**Publisher:** Packt Publishing

**Date of Publishing:** April 28, 2015

Monitoring Hadoop is for Hadoop administrators who want to learn how to monitor and diagnose their clusters.

## Hadoop Backup and Recovery Solutions

**Name:** [Hadoop Backup and Recovery Solutions](#)

**Authors:** Gaurav Barot, Chintan Mehta, Amij Patel

**Hadoop Version:** 2.7.x

**Publisher:** Packt Publishing

**Date of Publishing:** July 28, 2015

Hadoop Backup and Recovery Solutions demonstrates the strategies for data recovery from Hadoop backup clusters and troubleshoot problems.

## Hadoop Essentials

**Name:** [Hadoop Essentials](#)

**Authors:** Shiva Achari

**Hadoop Version:** 2.6

**Publisher:** Packt Publishing

**Date of Publishing:** April 29, 2015

Hadoop Essentials explains the key concepts of Hadoop and gives a thorough understanding of the Hadoop ecosystem.

## Hadoop in Practice, Second Edition

**Name:** [Hadoop in Practice, Second Edition](#)

**Author:** Alex Holmes

**Hadoop Version:** 2.x

**Publisher:** Manning

**Date of Publishing:** Fall 2014.

**Sample Chapters:** [Chapter 2: Introduction to YARN](#), [Chapter 9: SQL on Hadoop](#)

The second edition of Hadoop in Practice includes over 100 Hadoop techniques. This edition covers Hadoop 2 (YARN and [MapReduce](#) 2) and updates include new techniques that show how to integrate Kafka, Impala, and Spark SQL with Hadoop.

## Optimizing Hadoop for MapReduce

**Name:** [Optimizing Hadoop for MapReduce](#)

**Author:** Khaled Tannir

**Publisher:** Packt Publishing

**Date of Publishing:** February 21, 2014

**Sample Chapter:** [Chapter 3: Detecting System Bottlenecks](#)

Optimizing Hadoop for MapReduce book is an example-based tutorial that deals with Optimizing Hadoop for MapReduce job performance.

## Scaling Big Data with Hadoop and Solr

**Name:** [Scaling Big Data with Hadoop and Solr](#)

**Author:** Hrishikesh Karambelkar

**Publisher:** Packt Publishing

**Date of Publishing:** August 26, 2013

**Sample Chapter:** [Chapter 2: Understanding Solr](#)

Scaling Big Data with Hadoop and Solr is a step-by-step guide to building a search engine while scaling data. Starting with the basics of Apache Hadoop and Solr, this book then dives into advanced topics of optimizing search with some real-world use cases and sample Java code.

## Hadoop Operations and Cluster Management Cookbook

**Name:** [Hadoop Operations and Cluster Management Cookbook](#)

**Author:** Shumin Guo

**Hadoop Version:** 2.x

**Publisher:** Packt Publishing

**Date of Publishing:** July 24, 2013

**Sample Chapter:** [Chapter 3: Configuring a Hadoop Cluster](#)

Hadoop Operations and Cluster Management Cookbook is a guide for designing and managing a Hadoop cluster.

## Hadoop Beginner's Guide

**Name:** [Hadoop Beginner's Guide](#)

**Author:** Garry Turkington

**Hadoop Version:** 1.0.x

**Publisher:** Packt Publishing

**Date of Publishing:** February 22, 2013

**Sample Chapter:** [Chapter 4: Developing MapReduce Programs](#)

Written for complete beginners to Hadoop, covers how to install and run Hadoop on a local Ubuntu host or create an on-demand Hadoop cluster on Amazon Web Services (EC2), before getting to grips with MapReduce.

## Hadoop Real World Solutions Cookbook

**Name:** [Hadoop Real World Solutions Cookbook](#)

**Author:** Jonathan Owens, Brian Femiano, Jon Lentz

**Hadoop Version:** CDH3

**Publisher:** Packt Publishing

**Date of Publishing:** February 7, 2013

**Sample Chapter:** [Chapter 6: Big Data Analysis](#)

Collection of real world code analytics and design patterns using various tools from the Hadoop community. Each recipe walks the reader through the implementation, or in some cases debugging and configuration tuning. The book covers various tools including MapReduce, Hive, Pig, MRUnit, serialization using Avro/Thrift/ProtoBufs, Giraph, Accumulo and several others.

## Hadoop MapReduce Cookbook

**Name:** [Hadoop MapReduce Cookbook](#)

**Author:** Srinath Perera, Thilina Gunarathne

**Hadoop Version:** 1.0.x

**Publisher:** Packt Publishing

**Date of Publishing:** January 25, 2013

**Sample Chapter:** [Chapter 6: Analytics](#)

Hadoop MapReduce Cookbook is a guide to processing large and complex data sets using Hadoop [MapReduce](#).

## Hadoop Operations

**Name:** [Hadoop Operations](#)

**Author:** Eric Sammers

**Hadoop Version:** 1.x, CDH3.x

**Publisher:** O'Reilly Press

**Date of Publishing:** September 2012.

A guide to running large-scale Hadoop clusters, written by someone who has practical experience in such deployments.

## Hadoop in Practice

**Name:** [Hadoop in Practice](#)

**Author:** Alex Holmes

**Hadoop Version:** 1.0

**Publisher:** Manning

**Date of Publishing:** Fall 2012.

**Sample Chapter:** [Chapter 1](#)

## Hadoop: The Definitive Guide, 3rd Edition

**Name:** [Hadoop: The Definitive Guide, 3rd Edition](#)

**Author:** Tom White

**Hadoop Version:** 1.x

**Publisher:** O'Reilly

**Date of Publishing:** May 2012

**Sample Chapter:** [Sample Chapter](#)

## Hadoop in Action

**Name:** [Hadoop in Action](#)

**Author:** Chuck Lam

**Hadoop Version:** 0.19-0.20

**Publisher:** Manning

**Date of Publishing:** December, 2010

**Sample Chapter:** [Chapter 1](#)

Hadoop in Action introduces the subject and shows how to write programs in the [MapReduce](#) style. It starts with a few easy examples and then moves quickly to show Hadoop use in more complex data analysis tasks. Included are best practices and design patterns of [MapReduce](#) programming.

## Hadoop: The Definitive Guide, 2nd Edition

**Name:** [Hadoop: The Definitive Guide, 2nd Edition](#)

**Author:** Tom White

**Hadoop Version:** 0.20-0.21

**Publisher:** O'Reilly

**Date of Publishing:** September 2010

## Pro Hadoop

**Name:** [Pro Hadoop](#)

**Author:** Jason Venner

**Hadoop Version:** 0.20

**Publisher:** Apress

**Date of Publishing:** June 22, 2009

Jason says "This book is a step by step guide to writing, running and debugging Map/Reduce jobs using Hadoop, and to installing and managing Hadoop Clusters. It is ideal for training new Map/Reduce users and Cluster administrators and for polishing existing Hadoop skills."

## Hadoop: The Definitive Guide

**Name:** [Hadoop: The Definitive Guide](#)

**Author:** Tom White

**Hadoop Version:** 0.20

**Publisher:** O'Reilly

**Date of Publishing:** June 19, 2009

## Forthcoming Books

### Hadoop in Action, Second Edition

**Name:** [Hadoop in Action, Second Edition](#)

**Author:** Chuck P. Lam, Mark W. Davis

**Hadoop Version:** 2.x

**Publisher:** Manning

**Date of Publishing (est.):** October 2015

---

---

## Hadoop Videos

---

### Hands-On Big Data Analysis with Hadoop 3 (Video)

**Name:** [Hands-On Big Data Analysis with Hadoop 3 \(Video\)](#)

**Author:** Tomasz Lelek

**Publisher:** Packt

**Date of Publishing:** August 2018

Perform real-time data analytics with Hadoop

---

### Hands-On Beginner's Guide on Big Data and Hadoop 3 (Video)

**Name:** [Hands-On Beginner's Guide on Big Data and Hadoop 3 \(Video\)](#)

**Author:** Milind Jagre

**Publisher:** Packt

**Date of Publishing:** July 2018

Effectively store, manage, and analyze large Datasets with HDFS, SQOOP, YARN, and [MapReduce](#)

---

### Hadoop Administration and Cluster Management (Video)

**Name:** [Hadoop Administration and Cluster Management \(Video\)](#)

**Author:** Gurmukh Singh

**Publisher:** Packt

**Date of Publishing:** May 2018

Planning, deploying, managing, monitoring and performance-tuning your Hadoop cluster with Apache Hadoop

---

### Solving 10 Hadoop'able Problems (Video)

**Name:** [Solving 10 Hadoop'able Problems \(Video\)](#)

**Author:** Tomasz Lelek

**Publisher:** Packt

**Date of Publishing:** February 2018

Need solutions to your big data problems? Here are 10 real-world projects demonstrating problems solved using Hadoop

---

### Learn By Example: Hadoop, [MapReduce](#) for Big Data problems (Video)

**Name:** [Learn By Example: Hadoop, MapReduce for Big Data problems \(Video\)](#)

**Author:** Loonycorn

**Publisher:** Packt

**Date of Publishing:** Jan 2018

A hands-on workout in Hadoop, [MapReduce](#) and the art of thinking "parallel"

---

### The Ultimate Hands-on Hadoop (Video)

**Name:** [The Ultimate Hands-on Hadoop \(Video\)](#)

**Author:** Frank Kane



**Publisher:** Packt

**Date of Publishing:** June 2017

Design distributed systems that manage Big Data using Hadoop and related technologies.

---

## Getting Started with Hadoop 2.x (Video)

**Name:** [Getting Started with Hadoop 2.x \(Video\)](#)

**Author:** A K M Zahiduzzaman

**Publisher:** Packt

**Date of Publishing:** April 30, 2017

Build a strong foundation by exploring Hadoop ecosystem with real-world examples.

---

## Taming Big Data with [MapReduce](#) and Hadoop - Hands On! (Video)

**Name:** [Taming Big Data with MapReduce and Hadoop - Hands On! \(Video\)](#)

**Author:** Frank Kane

**Publisher:** Packt

**Date of Publishing:** September 12, 2016

Master the art of processing Big Data using Hadoop and [MapReduce](#) with the help of real-world examples.

---

Hadoop in Action introduces the subject and shows how to write programs in the [MapReduce](#) style. It starts with a few easy examples and then moves quickly to show Hadoop use in more complex data analysis tasks. Included are best practices and design patterns of [MapReduce](#) programming.