
Hadoop Mapreduce V2 Cookbook Second Edition

[Hadoop: The Definitive Guide](#)
[Learning Cascading](#)
[Mastering Hadoop 3](#)
[Scaling Big Data with Hadoop and Solr - Second Edition](#)
[Hadoop Application Architectures](#)
[Moving Beyond MapReduce and Batch Processing with Apache Hadoop 2](#)
[Elasticsearch for Hadoop](#)
[Implementing an IBM InfoSphere BigInsights Cluster using Linux on Power](#)
[Hadoop: The Definitive Guide](#)
[Apache Hadoop 3 Quick Start Guide](#)
[Apache Hadoop YARN](#)
[Learn about big data processing and analytics](#)
[Professional Hadoop Solutions](#)
[Big data processing at scale to unlock unique business insights](#)
[Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem](#)
[HDInsight Essentials - Second Edition](#)
[Big Data with Hadoop MapReduce](#)
[Monitoring Hadoop](#)
[Troubleshooting Ubuntu Server](#)
[Big Data Forensics - Learning Hadoop Investigations](#)
[Data Algorithms](#)
[Apache Spark 2.x Cookbook](#)
[19th International Conference on Intelligent Systems Design and Applications \(ISDA 2019\) held December 3-5, 2019](#)
[Hadoop: The Definitive Guide](#)
[Hadoop For Dummies](#)
[Mining of Massive Datasets](#)
[Big Data Analytics with Hadoop 3](#)
[Hadoop in Action](#)
[MapReduce Design Patterns](#)
[Storage and Analysis at Internet Scale](#)
[Data Analysis and Business Modeling with Excel 2013](#)
[Recipes for Scaling Up with Hadoop and Spark](#)
[Information and Communication Technology for Development for Africa](#)
[A Classroom Approach](#)
[Practical Graph Analytics with Apache Giraph](#)
[Hadoop Operations](#)
[Hadoop: Data Processing and Modelling](#)
[Build highly effective analytics solutions to gain valuable insight into your big data](#)
[Programming Hive](#)

Hadoop Mapreduce V2 Cookbook Second Edition

Downloaded from process.ogleschool.edu by guest

ALENA ISAIAH

Hadoop: The Definitive Guide Cambridge University Press

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

Learning Cascading Apress

Describes the features and functions of Apache Hive, the data infrastructure for Hadoop.

Mastering Hadoop 3 "O'Reilly Media, Inc."

"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

Scaling Big Data with Hadoop and Solr - Second Edition 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.

Hadoop Application Architectures "O'Reilly Media, Inc."

This book highlights recent research on intelligent systems and nature-inspired computing. It presents 62 selected papers from the 19th International Conference on Intelligent Systems Design and Applications (ISDA 2019), which was held online. The ISDA is a premier conference in the field of computational intelligence, and the latest installment brought together researchers, engineers and practitioners whose work involves intelligent systems and their applications in industry. Including contributions by authors from 33 countries, the book offers a valuable reference guide for all researchers, students and practitioners in the fields of Computer Science and Engineering.

Moving Beyond MapReduce and Batch Processing with Apache Hadoop 2 John Wiley & Sons

This book is useful for Hadoop administrators who need to learn how to monitor and diagnose their clusters. Also, the book will prove useful for new users of the technology, as the language used is simple and easy to grasp.

Elasticsearch for Hadoop Packt Publishing Ltd

Individual self-contained code recipes. Solve specific problems using individual recipes, or work through the book to develop your capabilities. If you are a big data enthusiast and striving to use Hadoop to solve your problems, this book is for you. Aimed at Java programmers with some knowledge of Hadoop MapReduce, this is also a comprehensive reference for developers and system admins who want to get up to speed using Hadoop.

Implementing an IBM InfoSphere BigInsights Cluster using Linux on Power Hadoop MapReduce v2 Cookbook - Second Edition

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

Hadoop: The Definitive Guide Packt Publishing Ltd

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

Apache Hadoop 3 Quick Start Guide IBM Redbooks

Practical Graph Analytics with Apache Giraph helps you build data mining and machine learning applications using the Apache Foundation's Giraph framework for graph processing. This is the same framework as used by Facebook, Google, and other social media analytics operations to derive business value from vast amounts of interconnected data points. Graphs arise in a wealth of data scenarios and describe the connections that are naturally formed in both digital and real worlds. Examples of such connections abound in online social networks such as Facebook and Twitter, among users who rate movies from services like Netflix and Amazon Prime, and are useful even in the context of biological networks for scientific research. Whether in the context of business or science, viewing data as connected adds value by increasing the amount of information available to be drawn from that data and put to use in generating new revenue or scientific opportunities. Apache Giraph offers a simple yet flexible programming model targeted to graph algorithms and designed to scale easily to accommodate massive amounts of data. Originally developed at Yahoo!, Giraph is now a top top-level project at the Apache Foundation, and it enlists contributors from companies such as Facebook, LinkedIn, and Twitter. Practical Graph Analytics with Apache Giraph brings the power of Apache Giraph to you, showing how to harness the power of graph processing for your own data by building sophisticated graph analytics applications using the very same framework that is relied upon by some of the largest players in the industry today.

Apache Hadoop YARN 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, 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.

Learn about big data processing and analytics Simon and Schuster

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

Professional Hadoop Solutions Packt Publishing

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

Big data processing at scale to unlock unique business insights Packt Publishing Ltd

Over 90 hands-on recipes to help you learn and master the intricacies of Apache Hadoop 2.X, YARN, Hive, Pig, Oozie, Flume, Sqoop, Apache Spark, and Mahout About This Book Implement outstanding Machine Learning use cases on your own analytics models and processes. Solutions to common problems when working with the Hadoop ecosystem. Step-by-step implementation of end-to-end big data use cases. Who This Book Is For Readers who have a basic knowledge of big data systems and want to advance their knowledge with hands-on recipes. What You Will Learn Installing and maintaining Hadoop 2.X cluster and its ecosystem. Write advanced Map Reduce programs and understand design patterns. Advanced Data Analysis using the Hive, Pig, and Map Reduce programs. Import and export data from various sources using Sqoop and Flume. Data storage in various file formats such as Text, Sequential, Parquet, ORC, and RC Files. Machine learning principles with libraries such as Mahout Batch and Stream data processing using Apache Spark In Detail Big data is the current requirement. Most organizations produce huge amount of data every day. With the arrival of Hadoop-like tools, it has become easier for everyone to solve big data problems with great efficiency and at minimal cost. Grasping Machine Learning techniques will help you greatly in building predictive models and using this data to make the right decisions for your organization. Hadoop Real World Solutions Cookbook gives readers insights into learning and mastering big data via recipes. The book not only clarifies most big data tools in the market but also provides best practices for using them. The book provides recipes that are based on the latest versions of Apache Hadoop 2.X, YARN, Hive, Pig, Sqoop, Flume, Apache Spark, Mahout and many more such ecosystem tools. This real-world-solution cookbook is packed with handy recipes you can apply to your own everyday issues. Each chapter provides in-depth recipes that can be referenced easily. This book provides detailed practices on the latest technologies such as YARN and Apache Spark. Readers will be able to consider themselves as big data experts on completion of this book. This guide is an invaluable tutorial if you are planning to implement a big data warehouse for your business. Style and approach An easy-to-follow guide that walks you through world of big data. Each tool in the Hadoop ecosystem is explained in detail and the recipes are placed in such a manner that readers can implement them sequentially. Plenty of reference links are provided for advanced reading.

Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem Packt Publishing Ltd

The go-to guidebook for deploying Big Data solutions withHadoop Today's enterprise architects need to understand how the Hadoopframeworks and APIs fit together, and how they can be integrated todeliver real-world solutions. This book is a practical, detailedguide to building and implementing those solutions, with code-levelinstruction in the popular Wrox tradition. It covers storing datawith HDFS and Hbase, processing data with MapReduce, and automatingdata processing with Oozie. Hadoop security, running Hadoop withAmazon Web Services, best practices, and automating Hadoopprocesses in real time are also covered in depth. With in-depth code examples in Java and XML and the latest onrecent additions to the Hadoop ecosystem, this complete resourcealso covers the use of APIs, exposing their inner workings andallowing architects and developers to better leverage and customizethem. The ultimate guide for developers, designers, and architectstwho need to build and deploy Hadoop applications Covers

storing and processing data with various technologies, automating data processing, Hadoop security, and delivering real-time solutions. Includes detailed, real-world examples and code-level guidelines. Explains when, why, and how to use these tools effectively. Written by a team of Hadoop experts in the programmer-to-programmer Wrox style. Professional Hadoop Solutions is the reference enterprise architects and developers need to maximize the power of Hadoop.

HD Insight Essentials - Second Edition Packt Publishing Ltd

Integrate Elasticsearch into Hadoop to effectively visualize and analyze your data. About This Book Build production-ready analytics applications by integrating the Hadoop ecosystem with Elasticsearch. Learn complex Elasticsearch queries and develop real-time monitoring Kibana dashboards to visualize your data. Use Elasticsearch and Kibana to search data in Hadoop easily with this comprehensive, step-by-step guide. Who This Book Is For This book is targeted at Java developers with basic knowledge on Hadoop. No prior Elasticsearch experience is expected. What You Will Learn Set up the Elasticsearch-Hadoop environment. Import HDFS data into Elasticsearch with MapReduce jobs. Perform full-text search and aggregations efficiently using Elasticsearch. Visualize data and create interactive dashboards using Kibana. Check and detect anomalies in streaming data using Storm and Elasticsearch. Inject and classify real-time streaming data into Elasticsearch. Get production-ready for Elasticsearch-Hadoop based projects. Integrate with Hadoop eco-system such as Pig, Storm, Hive, and Spark. In Detail The Hadoop ecosystem is a de-facto standard for processing tera-bytes and peta-bytes of data. Lucene-enabled Elasticsearch is becoming an industry standard for its full-text search and aggregation capabilities. Elasticsearch-Hadoop serves as a perfect tool to bridge the worlds of Elasticsearch and Hadoop ecosystem to get best out of both the worlds. Powered with Kibana, this stack makes it a cakewalk to get surprising insights out of your massive amount of Hadoop ecosystem in a flash. In this book, you'll learn to use Elasticsearch, Kibana and Elasticsearch-Hadoop effectively to analyze and understand your HDFS and streaming data. You begin with an in-depth understanding of the Hadoop, Elasticsearch, Marvel, and Kibana setup. Right after this, you will learn to successfully import Hadoop data into Elasticsearch by writing MapReduce job in a real-world example. This is then followed by a comprehensive look at Elasticsearch essentials, such as full-text search analysis, queries, filters and aggregations; after which you gain an understanding of creating various visualizations and interactive dashboard using Kibana. Classifying your real-world streaming data and identifying trends in it using Storm and Elasticsearch are some of the other topics that we'll cover. You will also gain an insight about key concepts of Elasticsearch and Elasticsearch-hadoop in distributed mode, advanced configurations along with some common configuration presets you may need for your production deployments. You will have "Go production checklist" and high-level view for cluster administration for post-production. Towards the end, you will learn to integrate Elasticsearch with other Hadoop eco-system tools, such as Pig, Hive and Spark. Style and approach A concise yet comprehensive approach has been adopted with real-time examples to help you grasp the concepts easily.

Big Data with Hadoop MapReduce "O'Reilly Media, Inc."

Unlock the power of your data with Hadoop 2.X ecosystem and its data warehousing techniques across large data sets. About This Book Conquer the mountain of data using Hadoop 2.X tools. The authors succeed in creating a context for Hadoop and its ecosystem. Hands-on examples and recipes giving the bigger picture and helping you to master Hadoop 2.X data processing platforms. Overcome the challenging data processing problems using this exhaustive course with Hadoop 2.X. Who This Book Is For This course is for Java developers, who know scripting, wanting a career shift to Hadoop - Big Data segment of the IT industry. So if you are a novice in Hadoop or an expert, this book will make you reach the most advanced level in Hadoop 2.X. What You Will Learn Best practices for setup and configuration of Hadoop clusters, tailoring the system to the problem at hand. Integration with relational databases, using Hive for SQL queries and Sqoop for data transfer. Installing and maintaining Hadoop 2.X cluster and its ecosystem. Advanced Data Analysis using the Hive, Pig, and Map Reduce programs. Machine learning principles with libraries such as Mahout and Batch and Stream data processing using Apache Spark. Understand the changes involved in the process in the move from Hadoop 1.0 to Hadoop 2.0. Dive into YARN and Storm and use YARN to integrate Storm with Hadoop. Deploy Hadoop on Amazon Elastic MapReduce and Discover HDFS replacements and learn about HDFS Federation. In Detail As Marc Andreessen has said "Data is eating the world," which can be witnessed today being the age of Big Data, businesses are producing data in huge volumes every day and this rise in tide of data need to be organized and analyzed in a more secured way. With proper and effective use of Hadoop, you can build new-improved models, and based on that you will be able to make the right decisions. The first module, Hadoop beginners Guide will walk you through on understanding Hadoop with very detailed instructions and how to go about using it. Commands are explained using sections called "What just happened" for more clarity and understanding. The second module, Hadoop Real World

Best Sellers - Books :

- [The Courage To Be Free: Florida's Blueprint For America's Revival](#)
- [The Seven Husbands Of Evelyn Hugo: A Novel By Taylor Jenkins Reid](#)
- [Little Blue Truck's Valentine](#)
- [The Silent Patient By Alex Michaelides](#)
- [Stone Maidens By Lloyd Devereux Richards](#)
- [I Love You Like No Otter: A Funny And Sweet Board Book For Babies And Toddlers \(punderland\) By Rose Rossner](#)
- [Spare](#)
- [The Very Hungry Caterpillar](#)
- [If He Had Been With Me](#)
- [Our Class Is A Family \(our Class Is A Family & Our School Is A Family\)](#)

Solutions Cookbook, 2nd edition, is an essential tutorial to effectively implement a big data warehouse in your business, where you get detailed practices on the latest technologies such as YARN and Spark. Big data has become a key basis of competition and the new waves of productivity growth. Hence, once you get familiar with the basics and implement the end-to-end big data use cases, you will start exploring the third module, Mastering Hadoop. So, now the question is if you need to broaden your Hadoop skill set to the next level after you nail the basics and the advance concepts, then this course is indispensable. When you finish this course, you will be able to tackle the real-world scenarios and become a big data expert using the tools and the knowledge based on the various step-by-step tutorials and recipes. Style and approach This course has covered everything right from the basic concepts of Hadoop till you master the advance mechanisms to become a big data expert. The goal here is to help you learn the basic essentials using the step-by-step tutorials and from there moving toward the recipes with various real-world solutions for you. It covers all the important aspects of Hadoop from system designing and configuring Hadoop, machine learning principles with various libraries with chapters illustrated with code fragments and schematic diagrams. This is a compendious course to explore Hadoop from the basics to the most advanced techniques available in Hadoop 2.X.

Monitoring Hadoop Packt Publishing Ltd

This IBM® Redbooks® publication demonstrates and documents how to implement and manage an IBM PowerLinux™ cluster for big data focusing on hardware management, operating systems provisioning, application provisioning, cluster readiness check, hardware, operating system, IBM InfoSphere® BigInsights™, IBM Platform Symphony®, IBM Spectrum™ Scale (formerly IBM GPFSTM), applications monitoring, and performance tuning. This publication shows that IBM PowerLinux clustering solutions (hardware and software) deliver significant value to clients that need cost-effective, highly scalable, and robust solutions for big data and analytics workloads. This book documents and addresses topics on how to use IBM Platform Cluster Manager to manage PowerLinux BigData data clusters through IBM InfoSphere BigInsights, Spectrum Scale, and Platform Symphony. This book documents how to set up and manage a big data cluster on PowerLinux servers to customize application and programming solutions, and to tune applications to use IBM hardware architectures. This document uses the architectural technologies and the software solutions that are available from IBM to help solve challenging technical and business problems. This book is targeted at technical professionals (consultants, technical support staff, IT Architects, and IT Specialists) that are responsible for delivering cost-effective Linux on IBM Power Systems™ solutions that help uncover insights among client's data so they can act to optimize business results, product development, and scientific discoveries.

[Troubleshooting Ubuntu Server](#) "O'Reilly Media, Inc."

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.

Big Data Forensics - Learning Hadoop Investigations Packt Publishing Ltd

Now in its second edition, this book focuses on practical algorithms for mining data from even the largest datasets.