
Hadoop Operations

Hadoop Application Architectures
Hadoop Security
Architecting Modern Data Platforms
Apache Sqoop Cookbook
Hadoop Security
Hadoop Operations
Learning Hadoop 2
Hadoop: Data Processing and Modelling
Big Data and Hadoop
Data Analytics with Hadoop
Proceedings on 25th International Joint Conference on Industrial Engineering and Operations Management - IJCIEOM
Big Data Using Hadoop and Hive
FUTURISTIC TRENDS IN INFORMATION TECHNOLOGY
Professional Hadoop
Hadoop Operations and Cluster Management Cookbook
Architecting Modern Data Platforms
Big Data Analytics: Applications, Hadoop Technologies and Hive
Hadoop Operations
Hadoop: The Definitive Guide
Hadoop Practice Guide
Big Data Analytics Beyond Hadoop
Mastering Hadoop 3
Pro Apache Hadoop
Real-World Hadoop
Hadoop: The Definitive Guide
Professional Hadoop Solutions
Expert Hadoop Administration
Space Operations
Hadoop in Action
Beginning Apache Hadoop Administration
Hadoop For Dummies
Mastering Hadoop
Big Data Forensics - Learning Hadoop Investigations
Hadoop in 24 Hours, Sams Teach Yourself
Applied Big Data Analytics in Operations Management
Programming MapReduce with Scalding
Practical Hive
Mastering Apache Hadoop
Apache Spark in 24 Hours, Sams Teach Yourself
Big Data Analytics with Hadoop 3

Hadoop Operations

Downloaded from archive.imba.com by guest

DUNCAN MATHEWS

Hadoop Application Architectures Sams Publishing

Unleash the Power of Big Data Processing with Apache Hadoop Ecosystem Are you ready to embark on a journey into the world of big data processing and analysis using Apache Hadoop? "Mastering Apache Hadoop" is your comprehensive guide to understanding and harnessing the capabilities of Hadoop for processing and managing massive datasets. Whether you're a data engineer seeking to optimize processing pipelines or a business analyst aiming to extract insights from large data, this book equips you with the knowledge and tools to master the art of Hadoop-based data processing. Key Features: 1. Deep Dive into Hadoop Ecosystem: Immerse yourself in the core components and concepts of the Apache Hadoop ecosystem. Understand the architecture, components, and functionalities that make Hadoop a powerful platform for big data. 2. Installation and Configuration: Master the art of installing and configuring Hadoop on various platforms. Learn about cluster setup, resource management, and configuration settings for optimal performance. 3. Hadoop Distributed File System (HDFS): Uncover the power of HDFS for distributed storage and data management. Explore concepts like replication, fault tolerance, and data placement to ensure data durability. 4. MapReduce and Data Processing: Delve into MapReduce, the core data processing paradigm in Hadoop. Learn how to write MapReduce jobs, optimize performance, and

leverage parallel processing for efficient data analysis. 5. Data Ingestion and ETL: Discover techniques for ingesting and transforming data in Hadoop. Explore tools like Apache Sqoop and Apache Flume for extracting data from various sources and loading it into Hadoop. 6. Data Querying and Analysis: Master querying and analyzing data using Hadoop. Learn about Hive, Pig, and Spark SQL for querying structured and semi-structured data, and uncover insights that drive informed decisions. 7. Data Storage Formats: Explore data storage formats optimized for Hadoop. Learn about Avro, Parquet, and ORC, and understand how to choose the right format for efficient storage and retrieval. 8. Batch and Stream Processing: Uncover strategies for batch and real-time data processing in Hadoop. Learn how to use Apache Spark and Apache Flink to process data in both batch and streaming modes. 9. Data Visualization and Reporting: Discover techniques for visualizing and reporting on Hadoop data. Explore integration with tools like Apache Zeppelin and Tableau to create compelling visualizations. 10. Real-World Applications: Gain insights into real-world use cases of Apache Hadoop across industries. From financial analysis to social media sentiment analysis, explore how organizations are leveraging Hadoop's capabilities for data-driven innovation. Who This Book Is For: "Mastering Apache Hadoop" is an essential resource for data engineers, analysts, and IT professionals who want to excel in big data processing using Hadoop. Whether you're new to Hadoop or seeking advanced techniques, this book will guide you through the intricacies and empower you to harness the full potential of big data technology.

Hadoop Security Simon and Schuster

As more corporations turn to Hadoop to store and process their most valuable data, the risk of a potential breach of those systems increases

exponentially. This practical book not only shows Hadoop administrators and security architects how to protect Hadoop data from unauthorized access, it also shows how to limit the ability of an attacker to corrupt or modify data in the event of a security breach. Authors Ben Spivey and Joey Echeverria provide in-depth information about the security features available in Hadoop, and organize them according to common computer security concepts. You'll also get real-world examples that demonstrate how you can apply these concepts to your use cases. Understand the challenges of securing distributed systems, particularly Hadoop Use best practices for preparing Hadoop cluster hardware as securely as possible Get an overview of the Kerberos network authentication protocol Delve into authorization and accounting principles as they apply to Hadoop Learn how to use mechanisms to protect data in a Hadoop cluster, both in transit and at rest Integrate Hadoop data ingest into enterprise-wide security architecture Ensure that security architecture reaches all the way to end-user access

Architecting Modern Data Platforms Springer Nature

This book presents the conference proceedings of the 25th edition of the International Joint Conference on Industrial Engineering and Operations Management. The conference is organized by 6 institutions (from different countries and continents) that gather a large number of members in the field of operational management, industrial engineering and engineering management. This edition of the conference had the title: THE NEXT GENERATION OF PRODUCTION AND SERVICE SYSTEMS in order to emphasis unpredictable and very changeable future. This conference is aimed to enhance connection between academia and industry and to gather researchers and practitioners specializing in operation management, industrial engineering, engineering management and other related disciplines from around the world.

Apache Sqoop Cookbook Sankalp Publication

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

Hadoop Security Springer Nature

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

Hadoop Operations Packt Publishing Ltd

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

Learning Hadoop 2 Packt Pub Limited

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 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.

Hadoop: Data Processing and Modelling "O'Reilly Media, Inc."

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.

Big Data and Hadoop Notion Press

Operations management is a tool by which companies can effectively meet customers' needs using the least amount of resources necessary. With the emergence of sensors and smart metering, big data is becoming an intrinsic part of modern operations management. Applied Big Data Analytics in Operations Management enumerates the challenges and creative solutions and tools to apply when using big data in operations management.

Outlining revolutionary concepts and applications that help businesses predict customer behavior along with applications of artificial neural networks, predictive analytics, and opinion mining on business management, this comprehensive publication is ideal for IT professionals, software engineers, business professionals, managers, and students of management.

Data Analytics with Hadoop Mercury Learning and Information

There's a lot of information about big data technologies, but splicing these technologies into an end-to-end enterprise data platform is a daunting task not widely covered. With this practical book, you'll learn how to build big data infrastructure both on-premises and in the cloud and successfully architect a modern data platform. Ideal for enterprise architects, IT managers, application architects, and data engineers, this book shows you how to overcome the many challenges that emerge during Hadoop projects. You'll explore the vast landscape of tools available in the Hadoop and big data realm in a thorough technical primer before diving into: Infrastructure: Look at all component layers in a modern data platform, from the server to the data center, to establish a solid foundation for data in your enterprise Platform: Understand aspects of deployment, operation, security, high availability, and disaster recovery, along with everything you need to know to integrate your platform with the rest of your enterprise IT Taking Hadoop to the cloud: Learn the important architectural aspects of running a big data platform in the cloud while maintaining enterprise security and high availability

Proceedings on 25th International Joint Conference on Industrial Engineering and Operations Management – IJCIEOM "O'Reilly Media, Inc."

Get expert guidance on architecting end-to-end data management solutions with Apache Hadoop. While many sources explain how to use various components in the Hadoop ecosystem, this practical book takes you through architectural considerations necessary to tie those components together into a complete tailored application, based on your particular use case. To reinforce those lessons, the book's second section provides detailed examples of architectures used in some of the most commonly found Hadoop applications. Whether you're designing a new Hadoop application, or planning to integrate Hadoop into your existing data infrastructure, Hadoop Application Architectures will skillfully guide you through the process. This book covers: Factors to consider when using Hadoop to store and model data Best practices for moving data in and out of the system Data processing frameworks, including MapReduce, Spark, and Hive Common Hadoop processing patterns, such as removing duplicate records and using windowing analytics Giraph, GraphX, and other tools for large graph processing on Hadoop Using workflow orchestration and scheduling tools such as Apache Oozie Near-real-time stream processing with Apache Storm, Apache Spark Streaming, and Apache Flume Architecture examples for clickstream analysis, fraud detection, and data warehousing

Big Data Using Hadoop and Hive Addison-Wesley Professional

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 *FUTURISTIC TRENDS IN INFORMATION TECHNOLOGY* "O'Reilly Media, Inc."

Ready to unlock the power of your data? With 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. You'll find illuminating case studies that demonstrate how Hadoop is used to solve specific problems. This third edition covers recent changes to Hadoop, including material on the new MapReduce API, as well as MapReduce 2 and its more flexible execution model (YARN). Store large datasets with the Hadoop Distributed File System (HDFS) Run distributed computations with MapReduce Use Hadoop's data and I/O building blocks for compression, data integrity, serialization (including Avro), 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 Load data from relational databases into HDFS, using Sqoop Perform large-scale data processing with the Pig query language Analyze datasets with Hive, Hadoop's data warehousing system

Take advantage of HBase for structured and semi-structured data, and ZooKeeper for building distributed systems

[Professional Hadoop](#) KHANNA PUBLISHING

The professional's one-stop guide to this open-source, Java-based big data framework Professional Hadoop is the complete reference and resource for experienced developers looking to employ Apache Hadoop in real-world settings. Written by an expert team of certified Hadoop developers, committers, and Summit speakers, this book details every key aspect of Hadoop technology to enable optimal processing of large data sets. Designed expressly for the professional developer, this book skips over the basics of database development to get you acquainted with the framework's processes and capabilities right away. The discussion covers each key Hadoop component individually, culminating in a sample application that brings all of the pieces together to illustrate the cooperation and interplay that make Hadoop a major big data solution. Coverage includes everything from storage and security to computing and user experience, with expert guidance on integrating other software and more. Hadoop is quickly reaching significant market usage, and more and more developers are being called upon to develop big data solutions using the Hadoop framework. This book covers the process from beginning to end, providing a crash course for professionals needing to learn and apply Hadoop quickly. Configure storage, UE, and in-memory computing Integrate Hadoop with other programs including Kafka and Storm Master the fundamentals of Apache Big Top and Ignite Build robust data security with expert tips and advice Hadoop's popularity is largely due to its accessibility. Open-source and written in Java, the framework offers almost no barrier to entry for experienced database developers already familiar with the skills and requirements real-world programming entails. Professional Hadoop gives you the practical information and framework-specific skills you need quickly.

Hadoop Operations and Cluster Management Cookbook Packt Publishing Ltd

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

[Architecting Modern Data Platforms](#) "O'Reilly Media, Inc."

This book is a complete practical approach for Hadoop lovers. It is mainly aimed at beginners who want to have a hands-on experience with Hadoop and its ecosystem. Its simplicity and step-by-step explanation will help students and other readers in the computer science industry to use this book as a reference manual. The book has been divided into various chapters that cover Hadoop installation, Summary on Hadoop core components, General commands in Hadoop with examples, SQOOP-import & export commands with verification steps, Pig Latin Commands, Analysis using Pig Latin, Pig Script examples, HiveQL Queries and expected outputs and HBase with CRUD operations. In short, this book is a guide for programmers and non-programmers to begin their projects in Hadoop. It is also suitable as a reference manual for students and professionals who are new to the Hadoop Ecosystems.

Big Data Analytics: Applications, Hadoop Technologies and Hive IGI Global

Related with Hadoop Operations:

- Telpas Practice Test 2022 : [click here](#)

Perform forensic investigations on Hadoop clusters with cutting-edge tools and techniques About This Book Identify, collect, and analyze Hadoop evidence forensically Learn about Hadoop's internals and Big Data file storage concepts A step-by-step guide to help you perform forensic analysis using freely available tools Who This Book Is For This book is meant for statisticians and forensic analysts with basic knowledge of digital forensics. They do not need to know Big Data Forensics. If you are an IT professional, law enforcement professional, legal professional, or a student interested in Big Data and forensics, this book is the perfect hands-on guide for learning how to conduct Hadoop forensic investigations. Each topic and step in the forensic process is described in accessible language. What You Will Learn Understand Hadoop internals and file storage Collect and analyze Hadoop forensic evidence Perform complex forensic analysis for fraud and other investigations Use state-of-the-art forensic tools Conduct interviews to identify Hadoop evidence Create compelling presentations of your forensic findings Understand how Big Data clusters operate Apply advanced forensic techniques in an investigation, including file carving, statistical analysis, and more In Detail Big Data forensics is an important type of digital investigation that involves the identification, collection, and analysis of large-scale Big Data systems. Hadoop is one of the most popular Big Data solutions, and forensically investigating a Hadoop cluster requires specialized tools and techniques. With the explosion of Big Data, forensic investigators need to be prepared to analyze the petabytes of data stored in Hadoop clusters. Understanding Hadoop's operational structure and performing forensic analysis with court-accepted tools and best practices will help you conduct a successful investigation. Discover how to perform a complete forensic investigation of large-scale Hadoop clusters using the same tools and techniques employed by forensic experts. This book begins by taking you through the process of forensic investigation and the pitfalls to avoid. It will walk you through Hadoop's internals and architecture, and you will discover what types of information Hadoop stores and how to access that data. You will learn to identify Big Data evidence using techniques to survey a live system and interview witnesses. After setting up your own Hadoop system, you will collect evidence using techniques such as forensic imaging and application-based extractions. You will analyze Hadoop evidence using advanced tools and techniques to uncover events and statistical information. Finally, data visualization and evidence presentation techniques are covered to help you properly communicate your findings to any audience. Style and approach This book is a complete guide that follows every step of the forensic analysis process in detail. You will be guided through each key topic and step necessary to perform an investigation. Hands-on exercises are presented throughout the book, and technical reference guides and sample documents are included for real-world use.

[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, 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.

[Hadoop: The Definitive Guide](#) "O'Reilly Media, Inc."

Dr.P.Pushpa, Lecturer, School of Software Engineering, East China University of Technology, Nanchang, Jiangxi, China. Dr.V.Thamilarasi, Assistant Professor, Department of Computer Science, Sri Sarada College for Women(Autonomous), Salem, Tamil Nadu, India. Dr. S. Lakshmi Prabha, Associate Professor, Department of Computer Science, Seethalakshmi Ramaswami College, Tiruchirappalli, Tamil Nadu, India. Mrs.Sudha Nagarajan, Assistant Professor, Department of Computer Science, Excel College for Commerce and Science, Komarapalayam, Namakkal, Tamil Nadu, India.

[Hadoop Practice Guide](#) "O'Reilly Media, Inc."

If you are a system or application developer interested in learning how to solve practical problems using the Hadoop framework, then this book is ideal for you. You are expected to be familiar with the Unix/Linux command-line interface and have some experience with the Java programming language. Familiarity with Hadoop would be a plus.