

Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

# **Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series**

“This text should be required reading for everyone in contemporary business.” --Peter Woodhull, CEO, Modus21 “The one book that clearly describes and links Big Data concepts to business utility.” --Dr. Christopher Starr, PhD “Simply, this is the best Big Data book on the market!” --Sam Rostam, Cascadian IT Group “...one of the most contemporary approaches I’ve seen to Big Data fundamentals...” --Joshua M. Davis, PhD The Definitive Plain-English Guide to Big Data for Business and Technology Professionals Big Data Fundamentals provides a pragmatic, no-nonsense introduction to Big Data. Best-selling IT author Thomas Erl and his team clearly explain key Big Data concepts, theory and terminology, as well as fundamental technologies and techniques. All coverage is supported with case study examples and numerous simple diagrams. The authors begin by explaining how Big Data can propel an organization forward by solving a spectrum of previously intractable business problems. Next, they demystify key analysis techniques and technologies and show how a Big Data solution environment can be built and integrated to offer competitive advantages. Discovering Big Data’s fundamental concepts and what makes it different from previous forms of

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

data analysis and data science Understanding the business motivations and drivers behind Big Data adoption, from operational improvements through innovation Planning strategic, business-driven Big Data initiatives Addressing considerations such as data management, governance, and security Recognizing the 5 “V” characteristics of datasets in Big Data environments: volume, velocity, variety, veracity, and value Clarifying Big Data’s relationships with OLTP, OLAP, ETL, data warehouses, and data marts Working with Big Data in structured, unstructured, semi-structured, and metadata formats Increasing value by integrating Big Data resources with corporate performance monitoring Understanding how Big Data leverages distributed and parallel processing Using NoSQL and other technologies to meet Big Data’s distinct data processing requirements Leveraging statistical approaches of quantitative and qualitative analysis Applying computational analysis methods, including machine learning 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

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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

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

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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

Web Scraping techniques are getting more popular, since data is as valuable as oil in 21st century. Through this book get some key knowledge about using XPath, regEX; web scraping libraries for R like rvest and RSelenium technologies. Key Features Techniques, tools and frameworks for web scraping with R Scrape data effortlessly from a variety of websites Learn how to selectively choose the data to scrape, and build your dataset Book Description Web scraping is a technique to extract data from websites. It simulates the behavior of a website user to turn the website itself into a web service to retrieve or introduce new data. This book gives you all you need to get started with scraping web pages using R programming. You will learn about the rules of RegEx and XPath, key components for scraping website data. We will show you web scraping techniques, methodologies, and frameworks. With this book's guidance, you will become comfortable with the tools to write and test RegEx and XPath rules. We will focus on examples of dynamic websites for scraping data and how to implement the techniques learned. You will learn how to collect URLs and then create XPath rules for

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

your first web scraping script using rvest library. From the data you collect, you will be able to calculate the statistics and create R plots to visualize them. Finally, you will discover how to use Selenium drivers with R for more sophisticated scraping. You will create AWS instances and use R to connect a PostgreSQL database hosted on AWS. By the end of the book, you will be sufficiently confident to create end-to-end web scraping systems using R. What you will learn Write and create regEX rules Write XPath rules to query your data Learn how web scraping methods work Use rvest to crawl web pages Store data retrieved from the web Learn the key uses of Rselenium to scrape data Who this book is for This book is for R programmers who want to get started quickly with web scraping, as well as data analysts who want to learn scraping using R. Basic knowledge of R is all you need to get started with this book.

The go-to guidebook for deploying Big Data solutions with Hadoop Today's enterprise architects need to understand how the Hadoop frameworks and APIs fit together, and how they can be integrated to deliver real-world solutions. This book is a practical, detailed guide to building and implementing those solutions, with code-level instruction in the popular Wrox tradition. It covers storing data with HDFS and Hbase, processing data with MapReduce, and automating data processing with Oozie. Hadoop security, running Hadoop with Amazon Web Services, best practices, and automating Hadoop processes in real time are also covered in depth. With in-depth code examples in Java and XML and the latest on recent additions to the Hadoop ecosystem, this complete resource also

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

covers the use of APIs, exposing their inner workings and allowing architects and developers to better leverage and customize them. The ultimate guide for developers, designers, and architects who 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.

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

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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

Apache Spark is a flexible in-memory framework that allows processing of both batch and real-time data. Its unified engine has made it quite popular for big data use cases. This book will help you to quickly get started with Apache Spark 2.0 and write efficient big data applications for a variety of use cases.

An easy-to-follow Apache Hadoop administrator's guide filled with practical screenshots and explanations for each step and configuration. This book is great for administrators interested in setting up and managing a large Hadoop cluster. If you are an administrator, or want to be an administrator, and you are ready to build and maintain a production-level cluster running CDH5, then this book is for you.

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.

Supervised and unsupervised machine learning made easy in Scala with this quick-

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

start guide. Key Features Construct and deploy machine learning systems that learn from your data and give accurate predictions Unleash the power of Spark ML along with popular machine learning algorithms to solve complex tasks in Scala. Solve hands-on problems by combining popular neural network architectures such as LSTM and CNN using Scala with DeepLearning4j library Book Description Scala is a highly scalable integration of object-oriented nature and functional programming concepts that make it easy to build scalable and complex big data applications. This book is a handy guide for machine learning developers and data scientists who want to develop and train effective machine learning models in Scala. The book starts with an introduction to machine learning, while covering deep learning and machine learning basics. It then explains how to use Scala-based ML libraries to solve classification and regression problems using linear regression, generalized linear regression, logistic regression, support vector machine, and Naïve Bayes algorithms. It also covers tree-based ensemble techniques for solving both classification and regression problems. Moving ahead, it covers unsupervised learning techniques, such as dimensionality reduction, clustering, and recommender systems. Finally, it provides a brief overview of deep learning using a real-life example in Scala. What you will learn Get acquainted with JVM-based machine learning libraries for Scala such as Spark ML and Deeplearning4j Learn RDDs, DataFrame, and Spark SQL for analyzing structured and unstructured data Understand supervised and unsupervised learning techniques with best practices

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

and pitfalls Learn classification and regression analysis with linear regression, logistic regression, Naïve Bayes, support vector machine, and tree-based ensemble techniques Learn effective ways of clustering analysis with dimensionality reduction techniques Learn recommender systems with collaborative filtering approach Delve into deep learning and neural network architectures Who this book is for This book is for machine learning developers looking to train machine learning models in Scala without spending too much time and effort. Some fundamental knowledge of Scala programming and some basics of statistics and linear algebra is all you need to get started with this book.

The need to handle increasingly larger data volumes is one factor driving the adoption of a new class of nonrelational “NoSQL” databases. Advocates of NoSQL databases claim they can be used to build systems that are more performant, scale better, and are easier to program. NoSQL Distilled is a concise but thorough introduction to this rapidly emerging technology. Pramod J. Sadalage and Martin Fowler explain how NoSQL databases work and the ways that they may be a superior alternative to a traditional RDBMS. The authors provide a fast-paced guide to the concepts you need to know in order to evaluate whether NoSQL databases are right for your needs and, if so, which technologies you should explore further. The first part of the book concentrates on core concepts, including schemaless data models, aggregates, new distribution models, the CAP theorem, and map-reduce. In the second part, the authors explore architectural

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

and design issues associated with implementing NoSQL. They also present realistic use cases that demonstrate NoSQL databases at work and feature representative examples using Riak, MongoDB, Cassandra, and Neo4j. In addition, by drawing on Pramod Sadalage's pioneering work, NoSQL Distilled shows how to implement evolutionary design with schema migration: an essential technique for applying NoSQL databases. The book concludes by describing how NoSQL is ushering in a new age of Polyglot Persistence, where multiple data-storage worlds coexist, and architects can choose the technology best optimized for each type of data access.

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

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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.

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

The Complete Guide to Data Science with Hadoop—For Technical Professionals, Businesspeople, and Students Demand is soaring for professionals who can solve real data science problems with Hadoop and Spark. Practical Data Science with Hadoop® and Spark is your complete guide to doing just that. Drawing on immense experience with Hadoop and big data, three leading experts bring together everything you need: high-level concepts, deep-dive techniques, real-world use cases, practical applications, and hands-on tutorials. The authors introduce the essentials of data science and the modern Hadoop ecosystem, explaining how Hadoop and Spark have evolved into an effective platform for solving data science problems at scale. In addition to comprehensive application coverage, the authors also provide useful guidance on the important steps of data ingestion, data munging, and visualization. Once the groundwork is in place, the authors focus on specific applications, including machine learning, predictive modeling for sentiment analysis, clustering for document analysis,

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

anomaly detection, and natural language processing (NLP). This guide provides a strong technical foundation for those who want to do practical data science, and also presents business-driven guidance on how to apply Hadoop and Spark to optimize ROI of data science initiatives. Learn What data science is, how it has evolved, and how to plan a data science career How data volume, variety, and velocity shape data science use cases Hadoop and its ecosystem, including HDFS, MapReduce, YARN, and Spark Data importation with Hive and Spark Data quality, preprocessing, preparation, and modeling Visualization: surfacing insights from huge data sets Machine learning: classification, regression, clustering, and anomaly detection Algorithms and Hadoop tools for predictive modeling Cluster analysis and similarity functions Large-scale anomaly detection NLP: applying data science to human language

Get Started Fast with Apache Hadoop® 2, YARN, and Today's Hadoop Ecosystem With Hadoop 2.x and YARN, Hadoop moves beyond MapReduce to become practical for virtually any type of data processing. Hadoop 2.x and the Data Lake concept represent a radical shift away from conventional approaches to data usage and storage. Hadoop 2.x installations offer unmatched scalability and breakthrough extensibility that supports new and existing Big Data analytics processing methods and models. Hadoop® 2 Quick-Start Guide is the first easy, accessible guide to Apache Hadoop 2.x, YARN, and the modern Hadoop ecosystem. Building on his unsurpassed experience teaching Hadoop and Big Data, author Douglas Eadline covers all the

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

basics you need to know to install and use Hadoop 2 on personal computers or servers, and to navigate the powerful technologies that complement it. Eadline concisely introduces and explains every key Hadoop 2 concept, tool, and service, illustrating each with a simple “beginning-to-end” example and identifying trustworthy, up-to-date resources for learning more. This guide is ideal if you want to learn about Hadoop 2 without getting mired in technical details. Douglas Eadline will bring you up to speed quickly, whether you’re a user, admin, devops specialist, programmer, architect, analyst, or data scientist. Coverage Includes Understanding what Hadoop 2 and YARN do, and how they improve on Hadoop 1 with MapReduce Understanding Hadoop-based Data Lakes versus RDBMS Data Warehouses Installing Hadoop 2 and core services on Linux machines, virtualized sandboxes, or clusters Exploring the Hadoop Distributed File System (HDFS) Understanding the essentials of MapReduce and YARN application programming Simplifying programming and data movement with Apache Pig, Hive, Sqoop, Flume, Oozie, and HBase Observing application progress, controlling jobs, and managing workflows Managing Hadoop efficiently with Apache Ambari—including recipes for HDFS to NFSv3 gateway, HDFS snapshots, and YARN configuration Learning basic Hadoop 2 troubleshooting, and installing Apache Hue and Apache Spark

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

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

manage and use with traditional tools. The solution: implementing a big data system. As *Big Data Made Easy: A Working Guide to the Complete Hadoop Toolset* shows, Apache Hadoop offers a scalable, fault-tolerant system for storing and processing data in parallel. It has a very rich toolset that allows for storage (Hadoop), configuration (YARN and ZooKeeper), collection (Nutch and Solr), processing (Storm, Pig, and Map Reduce), scheduling (Oozie), moving (Sqoop and Avro), monitoring (Chukwa, Ambari, and Hue), testing (Big Top), and analysis (Hive). The problem is that the Internet offers IT pros wading into big data many versions of the truth and some outright falsehoods born of ignorance. What is needed is a book just like this one: a wide-ranging but easily understood set of instructions to explain where to get Hadoop tools, what they can do, how to install them, how to configure them, how to integrate them, and how to use them successfully. And you need an expert who has worked in this area for a decade—someone just like author and big data expert Mike Frampton. *Big Data Made Easy* approaches the problem of managing massive data sets from a systems perspective, and it explains the roles for each project (like architect and tester, for example) and shows how the Hadoop toolset can be used at each system stage. It explains, in an easily understood manner and through numerous examples, how to use each tool. The book also explains the sliding scale of tools available depending upon data size and when and how to use them. *Big Data Made Easy* shows developers and architects, as well as testers and project managers, how to: Store big data Configure

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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. 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

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

Learn the HBase distributed database and the ZooKeeper distributed configuration service

Plan and Implement Hadoop Virtualization for Maximum Performance, Scalability, and Business Agility Enterprises running Hadoop must absorb rapid changes in big data ecosystems, frameworks, products, and workloads. Virtualized approaches can offer important advantages in speed, flexibility, and elasticity. Now, a world-class team of enterprise virtualization and big data experts guide you through the choices, considerations, and tradeoffs surrounding Hadoop virtualization. The authors help you decide whether to virtualize Hadoop, deploy Hadoop in the cloud, or integrate conventional and virtualized approaches in a blended solution. First, Virtualizing Hadoop reviews big data and Hadoop from the standpoint of the virtualization specialist. The authors demystify MapReduce, YARN, and HDFS and guide you through each stage of Hadoop data management. Next, they turn the tables, introducing big data experts to modern virtualization concepts and best practices. Finally, they bring Hadoop and virtualization together, guiding you through the decisions you'll face in planning, deploying, provisioning, and managing virtualized Hadoop. From security to multitenancy to day-to-day management, you'll find reliable answers for choosing your best Hadoop strategy and executing it. Coverage includes the following:

- Reviewing the frameworks, products, distributions, use cases, and roles associated with Hadoop
- Understanding YARN resource management, HDFS storage,

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

and I/O • Designing data ingestion, movement, and organization for modern enterprise data platforms • Defining SQL engine strategies to meet strict SLAs • Considering security, data isolation, and scheduling for multitenant environments • Deploying Hadoop as a service in the cloud • Reviewing the essential concepts, capabilities, and terminology of virtualization • Applying current best practices, guidelines, and key metrics for Hadoop virtualization • Managing multiple Hadoop frameworks and products as one unified system • Virtualizing master and worker nodes to maximize availability and performance • Installing and configuring Linux for a Hadoop environment

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

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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. A comprehensive guide to rolling out Datadog to monitor infrastructure and applications

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

running in both cloud and datacenter environments Key Features Learn Datadog to proactively monitor your infrastructure and cloud services Use Datadog as a platform for aggregating monitoring efforts in your organization Leverage Datadog's alerting service to implement on-call and site reliability engineering (SRE) processes Book Description Datadog is an essential cloud monitoring and operational analytics tool which enables the monitoring of servers, virtual machines, containers, databases, third-party tools, and application services. IT and DevOps teams can easily leverage Datadog to monitor infrastructure and cloud services, and this book will show you how. The book starts by describing basic monitoring concepts and types of monitoring that are rolled out in a large-scale IT production engineering environment. Moving on, the book covers how standard monitoring features are implemented on the Datadog platform and how they can be rolled out in a real-world production environment. As you advance, you'll discover how Datadog is integrated with popular software components that are used to build cloud platforms. The book also provides details on how to use monitoring standards such as Java Management Extensions (JMX) and StatsD to extend the Datadog platform. Finally, you'll get to grips with monitoring fundamentals, learn how monitoring can be rolled out using Datadog proactively, and find out how to extend and customize the Datadog platform. By the end of this Datadog book, you will have gained the skills needed to monitor your cloud infrastructure and the software applications running on it using Datadog. What you will learn Understand monitoring

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

fundamentals, including metrics, monitors, alerts, and thresholds Implement core monitoring requirements using Datadog features Explore Datadog's integration with cloud platforms and tools Extend Datadog using custom scripting and standards such as JMX and StatsD Discover how proactive monitoring can be rolled out using various Datadog features Understand how Datadog can be used to monitor microservices in both Docker and Kubernetes environments Get to grips with advanced Datadog features such as APM and Security Monitoring Who this book is for This book is for DevOps engineers, site reliability engineers (SREs), IT Production engineers, software developers and architects, cloud engineers, system administrators, and anyone looking to monitor and visualize their infrastructure and applications with Datadog. Basic working knowledge of cloud and infrastructure is useful. Working experience of Linux distribution and some scripting knowledge is required to fully take advantage of the material provided in the book.

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.

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

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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. 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

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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 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

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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

Hadoop in Action teaches readers how to use Hadoop and write MapReduce programs. The intended readers are programmers, architects, and project managers who have to process large amounts of data offline. Hadoop in Action will lead the reader from obtaining a copy of Hadoop to setting it up in a cluster and writing data analytic programs. The book begins by making the basic idea of Hadoop and MapReduce easier to grasp by applying the default Hadoop installation to a few easy-to-follow tasks, such as analyzing changes in word frequency across a body of documents. The book continues through the basic concepts of MapReduce applications developed using Hadoop, including a close look at framework components, use of Hadoop for a variety of data analysis tasks, and numerous examples of Hadoop in action. Hadoop in Action will explain how to use Hadoop and present design patterns and practices of programming MapReduce. MapReduce is a complex idea both conceptually and in its implementation, and Hadoop users are challenged to learn all the knobs and levers for running Hadoop. This book takes you beyond the mechanics of running Hadoop, teaching you to write meaningful programs in a MapReduce framework.

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

This book assumes the reader will have a basic familiarity with Java, as most code examples will be written in Java. Familiarity with basic statistical concepts (e.g. histogram, correlation) will help the reader appreciate the more advanced data processing examples. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book.

Summary Hadoop in Practice, Second Edition provides over 100 tested, instantly useful techniques that will help you conquer big data, using Hadoop. This revised new edition covers changes and new features in the Hadoop core architecture, including MapReduce 2. Brand new chapters cover YARN and integrating Kafka, Impala, and Spark SQL with Hadoop. You'll also get new and updated techniques for Flume, Sqoop, and Mahout, all of which have seen major new versions recently. In short, this is the most practical, up-to-date coverage of Hadoop available anywhere. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Book It's always a good time to upgrade your Hadoop skills! Hadoop in Practice, Second Edition provides a collection of 104 tested, instantly useful techniques for analyzing real-time streams, moving data securely, machine learning, managing large-scale clusters, and taming big data using Hadoop. This completely revised

# Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

edition covers changes and new features in Hadoop core, including MapReduce 2 and YARN. You'll pick up hands-on best practices for integrating Spark, Kafka, and Impala with Hadoop, and get new and updated techniques for the latest versions of Flume, Sqoop, and Mahout. In short, this is the most practical, up-to-date coverage of Hadoop available. Readers need to know a programming language like Java and have basic familiarity with Hadoop. What's Inside

Thoroughly updated for Hadoop 2 How to write YARN applications Integrate real-time technologies like Storm, Impala, and Spark Predictive analytics using Mahout and RR Readers need to know a programming language like Java and have basic familiarity with Hadoop. About the Author Alex Holmes works on tough big-data problems. He is a software engineer, author, speaker, and blogger specializing in large-scale Hadoop projects. Table of Contents

PART 1 BACKGROUND AND FUNDAMENTALS Hadoop in a heartbeat Introduction to YARN

PART 2 DATA LOGISTICS Data serialization—working with text and beyond Organizing and optimizing data in HDFS Moving data into and out of Hadoop

PART 3 BIG DATA PATTERNS Applying MapReduce patterns to big data Utilizing data structures and algorithms at scale Tuning, debugging, and testing

PART 4 BEYOND MAPREDUCE SQL on Hadoop Writing a YARN application

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

Today, billions of devices are Internet-connected, IoT standards and protocols are stabilizing, and technical professionals must increasingly solve real problems with IoT technologies. Now, five leading Cisco IoT experts present the first comprehensive, practical reference for making IoT work. IoT Fundamentals brings together knowledge previously available only in white papers, standards documents, and other hard-to-find sources—or nowhere at all. The authors begin with a high-level overview of IoT and introduce key concepts needed to successfully design IoT solutions. Next, they walk through each key technology, protocol, and technical building block that combine into complete IoT solutions. Building on these essentials, they present several detailed use cases, including manufacturing, energy, utilities, smart+connected cities, transportation, mining, and public safety. Whatever your role or existing infrastructure, you'll gain deep insight what IoT applications can do, and what it takes to deliver them. Fully covers the principles and components of next-generation wireless networks built with Cisco IOT solutions such as IEEE 802.11 (Wi-Fi), IEEE 802.15.4-2015 (Mesh), and LoRaWAN Brings together real-world tips, insights, and best practices for designing and implementing next-generation wireless networks Presents start-to-finish configuration examples for common deployment scenarios Reflects the extensive first-hand experience of Cisco experts

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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

Hadoop 2 Quick-Start Guide Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem Addison-Wesley Professional

Build efficient, high-performance & scalable systems to process large volumes of data with Apache Ignite Key Features Understand Apache Ignite's in-memory technology Create High-Performance app components with Ignite Build a real-time data streaming and complex event processing system Book Description Apache Ignite is a distributed in-memory platform designed to scale and process large volume of data. It can be integrated with microservices as well as monolithic systems, and can be used as a scalable, highly available and performant deployment platform for microservices. This book will teach you to use Apache Ignite for building a high-performance, scalable, highly available

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

system architecture with data integrity. The book takes you through the basics of Apache Ignite and in-memory technologies. You will learn about installation and clustering Ignite nodes, caching topologies, and various caching strategies, such as cache aside, read and write through, and write behind. Next, you will delve into detailed aspects of Ignite's data grid: web session clustering and querying data. You will learn how to process large volumes of data using compute grid and Ignite's map-reduce and executor service. You will learn about the memory architecture of Apache Ignite and monitoring memory and caches. You will use Ignite for complex event processing, event streaming, and the time-series predictions of opportunities and threats. Additionally, you will go through off-heap and on-heap caching, swapping, and native and Spring framework integration with Apache Ignite. By the end of this book, you will be confident with all the features of Apache Ignite 2.x that can be used to build a high-performance system architecture. What you will learn Use Apache Ignite's data grid and implement web session clustering Gain high performance and linear scalability with in-memory distributed data processing Create a microservice on top of Apache Ignite that can scale and perform Perform ACID-compliant CRUD operations on an Ignite cache Retrieve data from Apache Ignite's data grid using SQL, Scan and Lucene Text query Explore complex event processing concepts

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

and event streaming Integrate your Ignite app with the Spring framework Who this book is for The book is for Big Data professionals who want to learn the essentials of Apache Ignite. Prior experience in Java is necessary.

If you're like most R users, you have deep knowledge and love for statistics. But as your organization continues to collect huge amounts of data, adding tools such as Apache Spark makes a lot of sense. With this practical book, data scientists and professionals working with large-scale data applications will learn how to use Spark from R to tackle big data and big compute problems. Authors Javier Luraschi, Kevin Kuo, and Edgar Ruiz show you how to use R with Spark to solve different data analysis problems. This book covers relevant data science topics, cluster computing, and issues that should interest even the most advanced users. Analyze, explore, transform, and visualize data in Apache Spark with R Create statistical models to extract information and predict outcomes; automate the process in production-ready workflows Perform analysis and modeling across many machines using distributed computing techniques Use large-scale data from multiple sources and different formats with ease from within Spark Learn about alternative modeling frameworks for graph processing, geospatial analysis, and genomics at scale Dive into advanced topics including custom transformations, real-time data processing, and creating custom Spark

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series extensions

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.

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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.

The book contains the latest trend in IT industry 'BigData and Hadoop'. It explains how big is 'Big Data' and why everybody is trying to implement this into their IT project. It includes research work on various topics, theoretical and practical approach, each component of the architecture is described along with current industry trends. Big Data and Hadoop have taken together are a new skill as per the industry standards. Readers will get a compact book along with the industry experience and would be a reference to help readers. KEY FEATURES

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

Overview Of Big Data, Basics of Hadoop, Hadoop Distributed File System, HBase, MapReduce, HIVE: The Dataware House Of Hadoop, PIG: The Higher Level Programming Environment, SQOOP: Importing Data From Heterogeneous Sources, Flume, Ozzie, Zookeeper & Big Data Stream Mining, Chapter-wise Questions & Previous Years Questions

This IBM® Redbooks® publication provides topics to help the technical community take advantage of the resilience, scalability, and performance of the IBM Power Systems™ platform to implement or integrate an IBM Data Engine for Hadoop and Spark solution for analytics solutions to access, manage, and analyze data sets to improve business outcomes. This book documents topics to demonstrate and take advantage of the analytics strengths of the IBM POWER8® platform, the IBM analytics software portfolio, and selected third-party tools to help solve customer's data analytic workload requirements. This book describes how to plan, prepare, install, integrate, manage, and show how to use the IBM Data Engine for Hadoop and Spark solution to run analytic workloads on IBM POWER8. In addition, this publication delivers documentation to complement available IBM analytics solutions to help your data analytic needs. This publication strengthens the position of IBM analytics and big data solutions with a well-defined and documented deployment model within an IBM POWER8

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

virtualized environment so that customers have a planned foundation for security, scaling, capacity, resilience, and optimization for analytics workloads. This book is targeted at technical professionals (analytics consultants, technical support staff, IT Architects, and IT Specialists) that are responsible for delivering analytics solutions and support on IBM Power Systems.

Get started fast with Apache Hadoop 2, with the first easy, accessible guide to this revolutionary Big Data technology. Building on his unsurpassed experience teaching Hadoop and Big Data, Dr. Douglas Eadline covers all the basics you need to know to install and use Hadoop 2 on both personal computers and servers, and navigate the entire Apache Hadoop ecosystem. Eadline demystifies Hadoop 2, explains the problems it solves, shows how it relates to Big Data, and demonstrates both administrators and users work with it. He explains the central role of MapReduce in Hadoop 1, and how (and why) YARN and Hadoop 2 move beyond MapReduce. You'll find essential information on: Planning and performing Hadoop 2 installations -- including decisions about hardware, software, clustering, and HDFS Using the Hadoop Distributed File System (HDFS) and working around its tradeoffs Running and benchmarking Hadoop 2 programs Working with MapReduce -- including basic programming examples Using higher-level tools, including Pig and Hive Getting started with Apache

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

Hadoop YARN frameworks Administering Hadoop 2 with Ambari, radmin, and automated scripts From its Getting Started checklist/flowchart to its roadmap of additional resources, Hadoop 2 Quick-Start Guide is your perfect Hadoop 2 starting point -- and your fastest way to start mastering Big Data.

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—Puppet, 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

Massive, disruptive change is coming to IT as software as a service (SaaS), SOA, mashups, Web 2.0, and cloud computing truly come of age. Now, one of the world's leading IT innovators explains what it all means—coherently, thoroughly, and authoritatively. Writing for IT executives, architects, and developers alike, world-renowned expert David S. Linthicum explains why the days of managing IT organizations as private fortresses will rapidly disappear as IT inevitably becomes a global community. He demonstrates how to run IT when critical elements of customer, product, and business data and processes extend far beyond the firewall—and how to use all that information to deliver real-time answers about everything from an individual customer's credit to the location of a specific cargo container. *Cloud Computing and SOA Convergence in Your Enterprise* offers a clear-eyed assessment of the challenges associated with this new world—and offers a step-by-step program for getting there with maximum return on investment and minimum risk. Using multiple examples, Linthicum Reviews the powerful cost, value, and risk-related drivers behind the move to cloud computing—and explains why the shift will accelerate Explains the technical underpinnings, supporting technologies, and best-practice methods you'll need

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

to make the transition Helps you objectively assess the promise of cloud computing and SOA for your organization, quantify value, and make the business case Walks you through evaluating your existing IT infrastructure and finding your most cost-effective, safest path to the “cloud” Shows how to choose the right candidate data, services, and processes for your cloud computing initiatives Guides you through building disruptive infrastructure and next-generation process platforms Helps you bring effective, high-value governance to the clouds If you’re ready to begin driving real competitive advantage from cloud computing, this book is the start-to-finish roadmap you need to make it happen.

Learn how to quickly generate business intelligence, insights and create interactive dashboards for digital storytelling through various data sources with Redash Key Features Learn the best use of visualizations to build powerful interactive dashboards Create and share visualizations and data in your organization Work with different complexities of data from different data sources Book Description Data exploration and visualization is vital to Business Intelligence, the backbone of almost every enterprise or organization. Redash is a querying and visualization tool developed to simplify how marketing and business development departments are exposed to data. If you want to learn to create interactive dashboards with Redash, explore different visualizations, and

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

share the insights with your peers, then this is the ideal book for you. The book starts with essential Business Intelligence concepts that are at the heart of data visualizations. You will learn how to find your way round Redash and its rich array of data visualization options for building interactive dashboards. You will learn how to create data storytelling and share these with peers. You will see how to connect to different data sources to process complex data, and then visualize this data to reveal valuable insights. By the end of this book, you will be confident with the Redash dashboarding tool to provide insight and communicate data storytelling. What you will learn

- Install Redash and troubleshoot installation errors
- Manage user roles and permissions
- Fetch data from various data sources
- Visualize and present data with Redash
- Create active alerts based on your data
- Understand Redash administration and customization
- Export, share and recount stories with Redash visualizations
- Interact programmatically with Redash through the Redash API

Who this book is for This book is intended for Data Analysts, BI professionals and Data Developers, but can be useful to anyone who has a basic knowledge of SQL and a creative mind. Familiarity with basic BI concepts will be helpful, but no knowledge of Redash is required.

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

## Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data Analytics Series

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 MongoDB has grown to become the de facto NoSQL database with millions of users, from small start-ups to Fortune 500 companies. It can solve problems that are considered difficult, if not impossible, for aging RDBMS technologies. Written for version 4 of MongoDB, this book is the easiest way to get started with

Read Free Hadoop 2 Quick Start Guide Learn The Essentials Of Big Data  
Computing In The Apache Hadoop 2 Ecosystem Addison Wesley Data  
Analytics Series

MongoDB.

[Copyright: 822c70a81bc18d945f26f4d612a8fecc](#)