# Installing Hadoop 2 6 X On Windows 10

Communication, Networks and Computing
Expert Hadoop 2 Administration
Big Data Analytics
ODROID Magazine
Apache Spark Graph Processing
Learn Hadoop in 24 Hours
Build highly effective analytics solutions to gain valuable insight into your big data
Proceedings of 2011 International Conference on Electronic Engineering, Communication and Management(EECM 2011), held on December 24-25, 2011, Beijing, China
Build your own highly scalable and robust data storage systems that can support a variety of cutting-edge AI applications
Professional NoSQL
Moving Beyond MapReduce and Batch Processing with Apache Hadoop 2
Examples and Case Studies
Scaling Big Data with Hadoop and Solr - Second Edition
BIG DATA AND HADOOP
Apache Spark in 24 Hours, Sams Teach Yourself
IBM Platform Computing Solutions
The Artificial Intelligence Infrastructure Workshop
Euro-Par 2017: Parallel Processing
Hadoop Real-World Solutions Cookbook
Second International Conference, CNC 2020, Gwalior, India, December 29–31, 2020, Revised Selected Papers
Mastering MongoDB 4.x
The First Step towards Hadoop Administration and Management
R and Data Mining
Hadoop 2 Quick-Start Guide
Cloudera Data Platform Private Cloud Base with IBM Spectrum Scale
23rd International Conference on Parallel and Distributed Computing, Santiago de Compostela, Spain, August 28 – September 1, 2017, Proceedings
Learn the Essentials of Big Data Computing in the Apache Hadoop 2 Ecosystem
Integrated Intelligent Computing, Communication and Security
Advances in Information, Communication and Cybersecurity
Cloud Computing: A Hands-On Approach
Pro Apache Hadoop
A Guide for Developers and Administrators
Hadoop Operations
Intelligent Systems and Applications
Python High Performance
Beginning Apache Hadoop Administration
Big Data Analytics with Hadoop 3
Expert techniques to run high-volume and fault-tolerant database solutions using MongoDB 4.x, 2nd Edition

## SYLVIA FORD

CRC Press
A handy reference guide for data analysts and data scientists to help to obtain value from big data analytics using Spark on Hadoop clusters About This Book This book is based on the latest 2.0 version of Apache Spark and 2.7 version of Hadoop integrated with most commonly used tools. Learn all Spark stack components including latest topics such as DataFrames, DataSets, GraphFrames, Structured Streaming, DataFrame based ML Pipelines and SparkR. Integrations with frameworks such as HDFS, YARN and tools such as Jupyter, Zeppelin, NiFi, Mahout, HBase Spark Connector, GraphFrames, H2O and Hivemall. Who This Book Is For Though this book is primarily aimed at data analysts and data scientists, it will also help architects, programmers, and practitioners. Knowledge of either Spark or Hadoop would be beneficial. It is assumed that you have basic programming background in Scala, Python, SQL, or R programming with basic Linux experience. Working experience within big data environments is not mandatory. What You Will Learn Find out and implement the tools and techniques of big data analytics using Spark on Hadoop clusters with wide variety of tools used with Spark and Hadoop Understand all the Hadoop and Spark ecosystem components Get to know all the Spark components: Spark Core, Spark SQL, DataFrames, DataSets, Conventional and Structured Streaming, MLLib, ML Pipelines and Graphx See batch and real-time data analytics using Spark Core, Spark SQL, and Conventional and Structured Streaming Get to grips with data science and machine learning using MLLib, ML Pipelines, H2O, Hivemall, Graphx, SparkR and Hivemall. In Detail Big Data Analytics book aims at providing the fundamentals of Apache Spark and Hadoop. All Spark components – Spark Core, Spark SQL, DataFrames, Data sets, Conventional Streaming, Structured Streaming, MLlib, Graphx and Hadoop core components – HDFS, MapReduce and Yarn are explored in greater depth with implementation examples on Spark + Hadoop clusters. It is moving away from MapReduce to Spark. So, advantages of Spark over MapReduce are explained at great depth to reap benefits of in-memory speeds. DataFrames API, Data Sources API and new Data set API are explained for building Big Data analytical applications. Real-time data analytics using Spark Streaming with Apache Kafka and HBase is covered to help building streaming applications. New Structured streaming concept is explained with an IOT (Internet of Things) use case. Machine learning techniques are covered using MLLib, ML Pipelines and SparkR and Graph Analytics are covered with GraphX and GraphFrames components of Spark. Readers will also get an opportunity to get started with web based notebooks such as Jupyter, Apache Zeppelin and data flow tool Apache NiFi to analyze and visualize data. Style and approach This step-by-step pragmatic guide will make life easy no matter what your level of experience. You will deep dive into Apache Spark on Hadoop clusters through ample exciting real-life examples. Practical tutorial explains data science in simple terms to help programmers and data analysts get started with Data Science

**Communication, Networks and Computing** Guru99
As big data becomes more ubiquitous, businesses are wondering how they can best leverage it to gain insight into their most important business questions. Using machine learning (ML) and deep learning (DL) in big data environments can identify historical patterns and build artificial intelligence (AI) models that can help businesses to improve customer experience, add services and offerings, identify new revenue streams or lines of business (LOBs), and optimize business or manufacturing operations. The power of AI for predictive analytics is being harnessed across all industries, so it is important that businesses familiarize themselves with all of the tools and techniques that are available for integration with their data lake environments. In this IBM® Redbooks® publication, we cover the best practices for deploying and integrating some of the best AI solutions on the market, including: IBM Watson Machine Learning Accelerator (see note for product naming) IBM Watson Studio Local IBM Power SystemsTM IBM SpectrumTM Scale IBM Data Science Experience (IBM DSX) IBM Elastic StorageTM Server Hortonworks Data Platform (HDP) Hortonworks DataFlow (HDF) H2O Driverless AI We map out all the integrations that are possible with our different AI solutions and how they can integrate with your existing or new data lake. We also walk you through some of our client use cases and show you how some of the industry leaders are using Hortonworks, IBM PowerAI, and IBM Watson Studio Local to drive decision making. We also advise you on your deployment options, when to use a GPU, and why you should use the IBM Elastic Storage Server (IBM ESS) to improve storage management. Lastly, we describe how to integrate IBM Watson Machine Learning Accelerator and Hortonworks with or without IBM Watson Studio Local, how to access real-time data, and security. Note: IBM Watson Machine Learning Accelerator is the new product name for IBM PowerAI Enterprise. Note: Hortonworks merged with Cloudera in January 2019. The new company is called Cloudera. References to Hortonworks as a business entity in this publication are now referring to the merged company. Product names beginning with Hortonworks continue to be marketed and sold under their original names.

Expert Hadoop 2 Administration Packt Publishing Ltd
Pro Apache Hadoop, Second Edition brings you up to speed on Hadoop – the framework of big data. Revised to cover Hadoop 2.0, the book covers the very latest developments such as YARN (aka MapReduce 2.0), new HDFS high-availability features, and increased scalability in the form of HDFS Federations. All the old content has been revised too, giving the latest on the ins and outs of MapReduce, cluster design, the Hadoop Distributed File System, and more. This book covers everything you need to build your first Hadoop cluster and begin analyzing and deriving value from your business and scientific data. Learn to solve big-data problems the MapReduce way, by breaking a big problem into chunks and creating small-scale solutions that can be flung across thousands upon thousands of nodes to analyze large data volumes in a short amount of wall-clock time. Learn how to let Hadoop take care of distributing and parallelizing your software—you just focus on the code; Hadoop takes care of the rest. Covers all that is new in Hadoop 2.0 Written by a professional involved in Hadoop since day one Takes you quickly to the seasoned pro level on the hottest cloud-computing framework

*Big Data Analytics* Pearson Education
"Apache Hadoop is helping drive the Big Data revolution. Now, its data processing has been completely overhauled: Apache Hadoop YARN provides resource management at data center scale and easier ways to create distributed applications that process petabytes of data. And now in Apache HadoopTM 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

ODROID Magazine Hardkernel, Ltd

Need to move a relational database application to Hadoop? This comprehensive guide introduces you to Apache Hive, Hadoop’s data warehouse infrastructure. You’ll quickly learn how to use Hive’s SQL dialect—HiveQL—to summarize, query, and analyze large datasets stored in Hadoop’s distributed filesystem. This example-driven guide shows you how to set up and configure Hive in your environment, provides a detailed overview of Hadoop and MapReduce, and demonstrates how Hive works within the Hadoop ecosystem. You’ll also find real-world case studies that describe how companies have used Hive to solve unique problems involving petabytes of data. Use Hive to create, alter, and drop databases, tables, views, functions, and indexes Customize data formats and storage options, from files to external databases Load and extract data from tables—and use queries, grouping, filtering, joining, and other conventional query methods Gain best practices for creating user defined functions (UDFs) Learn Hive patterns you should use and anti-patterns you should avoid Integrate Hive with other data processing programs Use storage handlers for NoSQL databases and other datastores Learn the pros and cons of running Hive on Amazon’s Elastic MapReduce

*Apache Spark Graph Processing* Springer

This book constitutes the proceedings of the 23rd International Conference on Parallel and Distributed Computing, Euro-Par 2017, held in Santiago de Compostela, Spain, in August/September 2017. The 50 revised full papers presented together with 2 abstract of invited talks and 1 invited paper were carefully reviewed and selected from 176 submissions. The papers are organized in the following topical sections: support tools and environments; performance and power modeling, prediction and evaluation; scheduling and load balancing; high performance architectures and compilers; parallel and distributed data management and analytics; cluster and cloud computing; distributed systems and algorithms; parallel and distributed programming, interfaces and languages; multicore and manycore parallelism; theory and algorithms for parallel computation and networking; prallel numerical methods and applications; and accelerator computing.

Learn Hadoop in 24 Hours Packt Publishing Ltd

Learn how to use Spark to process big data at speed and scale for sharper analytics. Put the principles into practice for faster, slicker big data projects. About This Book A quick way to get started with Spark – and reap the rewards From analytics to engineering your big data architecture, we've got it covered Bring your Scala and Java knowledge – and put it to work on new and exciting problems Who This Book Is For This book is for developers with little to no knowledge of Spark, but with a background in Scala/Java programming. It's recommended that you have experience in dealing and working with big data and a strong interest in data science. What You Will Learn Install and set up Spark in your cluster Prototype distributed applications with Spark's interactive shell Perform data wrangling using the new DataFrame APIs Get to know the different ways to interact with Spark's distributed representation of data (RDDs) Query Spark with a SQL-like query syntax See how Spark works with big data Implement machine learning systems with highly scalable algorithms Use R, the popular statistical language, to work with Spark Apply interesting graph algorithms and graph processing with GraphX In Detail When people want a way to process big data at speed, Spark is invariably the solution. With its ease of development (in comparison to the relative complexity of Hadoop), it's unsurprising that it's becoming popular with data analysts and engineers everywhere. Beginning with the fundamentals, we'll show you how to get set up with Spark with minimum fuss. You'll then get to grips with some simple APIs before investigating machine learning and graph processing – throughout we'll make sure you know exactly how to apply your knowledge. You will also learn how to use the Spark shell, how to load data before finding out how to build and run your own Spark applications. Discover how to manipulate your RDD and get stuck into a range of DataFrame APIs. As if that's not enough, you'll also learn some useful Machine Learning algorithms with the help of Spark MLlib and integrating Spark with R. We'll also make sure you're confident and prepared for graph processing, as you learn more about the GraphX API. Style and approach This book is a basic, step-by-step tutorial that will help you take advantage of all that Spark has to offer.

Build highly effective analytics solutions to gain valuable insight into your big data Apress

Hadoop has changed the way large data sets are analyzed, stored, transferred, and processed. At such low cost, it provides benefits like supports partial failure, fault tolerance, consistency, scalability, flexible schema, and so on. It also supports cloud computing. More and more number of individuals are looking forward to mastering their Hadoop skills. While initiating with Hadoop, most users are unsure about how to proceed with Hadoop. They are not aware of what are the pre-requisite or data structure they should be familiar with. Or How to make the most efficient use of Hadoop and its ecosystem. To help them with all these queries and other issues this e-book is designed. The book gives insights into many of Hadoop libraries and packages that are not known to many Big data Analysts and Architects. The e-book also tells you about Hadoop MapReduce and HDFS. The example in the e-book is well chosen and demonstrates how to control Hadoop ecosystem through various shell commands. With this book, users will gain expertise in Hadoop technology and its related components. The book leverages you with the best Hadoop content with the lowest price range. After going through this book, you will also acquire knowledge on Hadoop Security required for Hadoop Certifications like CCAH and CCDH. It is a definite guide to Hadoop. Table Of Content Chapter 1: What Is Big Data 1. Examples Of 'Big Data' 2. Categories Of 'Big Data' 3. Characteristics Of 'Big Data' 4. Advantages Of Big Data Processing Chapter 2: Introduction to Hadoop 1. Components of Hadoop 2. Features Of 'Hadoop' 3. Network Topology In Hadoop Chapter 3: Hadoop Installation Chapter 4: HDFS 1. Read Operation 2. Write Operation 3. Access HDFS using JAVA API 4. Access HDFS Using COMMAND-LINE INTERFACE Chapter 5: Mapreduce 1. How MapReduce works 2. How MapReduce Organizes Work? Chapter 6: First Program 1. Understanding MapReducer Code 2. Explanation of SalesMapper Class 3. Explanation of SalesCountryReducer Class 4. Explanation of SalesCountryDriver Class Chapter 7: Counters & Joins In MapReduce 1. Two types of counters 2. MapReduce Join Chapter 8: MapReduce Hadoop Program To Join Data Chapter 9: Flume and Sqoop 1. What is SQOOP in Hadoop? 2. What is FLUME in Hadoop? 3. Some Important features of FLUME Chapter 10: Pig 1. Introduction to PIG 2. Create your First PIG Program 3. PART 1) Pig Installation 4. PART 2) Pig Demo Chapter 11: OOZIE 1. What is OOZIE? 2. How does OOZIE work? 3. Example Workflow Diagram 4. Oozie

workflow application 5. Why use Oozie? 6. FEATURES OF OOZIE
Proceedings of 2011 International Conference on Electronic Engineering, Communication and Management(EECM 2011), held on December 24-25, 2011, Beijing, China "O'Reilly Media, Inc."
A hands-on guide to leveraging NoSQL databases NoSQL databases are an efficient and powerful tool for storing and manipulating vast quantities of data. Most NoSQL databases scale well as data grows. In addition, they are often malleable and flexible enough to accommodate semi-structured and sparse data sets. This comprehensive hands-on guide presents fundamental concepts and practical solutions for getting you ready to use NoSQL databases. Expert author Shashank Tiwari begins with a helpful introduction on the subject of NoSQL, explains its characteristics and typical uses, and looks at where it fits in the application stack. Unique insights help you choose which NoSQL solutions are best for solving your specific data storage needs. Professional NoSQL: Demystifies the concepts that relate to NoSQL databases, including column-family oriented stores, key/value databases, and document databases. Delves into installing and configuring a number of NoSQL products and the Hadoop family of products. Explains ways of storing, accessing, and querying data in NoSQL databases through examples that use MongoDB, HBase, Cassandra, Redis, CouchDB, Google App Engine Datastore and more. Looks at architecture and internals. Provides guidelines for optimal usage, performance tuning, and scalable configurations. Presents a number of tools and utilities relating to NoSQL, distributed platforms, and scalable processing, including Hive, Pig, RRDtool, Nagios, and more.
*Build your own highly scalable and robust data storage systems that can support a variety of cutting-edge AI applications* IBM Redbooks
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
*Professional NoSQL* Sams Publishing
This IBM® Platform Computing Solutions Redbooks® publication is the first book to describe each of the available offerings that are part of the IBM portfolio of Cloud, analytics, and High Performance Computing (HPC) solutions for our clients. This IBM Redbooks publication delivers descriptions of the available offerings from IBM Platform Computing that address challenges for our clients in each industry. We include a few implementation and testing scenarios with selected solutions. This publication helps strengthen the position of IBM Platform Computing solutions with a well-defined and documented deployment model within an IBM System x® environment. This deployment model offers clients a planned foundation for dynamic cloud infrastructure, provisioning, large-scale parallel HPC application development, cluster management, and grid applications. This IBM publication is targeted to IT specialists, IT architects, support personnel, and clients. This book is intended for anyone who wants information about how IBM Platform Computing solutions use IBM to provide a wide array of client solutions.
Moving Beyond MapReduce and Batch Processing with Apache Hadoop 2 Springer Science & Business Media
An expert's guide to build fault tolerant MongoDB application About This Book Master the advanced modeling, querying, and administration techniques in MongoDB and become a MongoDB expert Covers the latest updates and Big Data features frequently used by professional MongoDB developers and administrators If your goal is to become a certified MongoDB professional, this book is your perfect companion Who This Book Is For Mastering MongoDB is a book for database developers, architects, and administrators who want to learn how to use MongoDB more effectively and productively. If you have experience in, and are interested in working with, NoSQL databases to build apps and websites, then this book is for you. What You Will Learn Get hands-on with advanced querying techniques such as indexing, expressions, arrays, and more. Configure, monitor, and maintain highly scalable MongoDB environment like an expert. Master replication and data sharding to optimize read/write performance. Design secure and robust applications based on MongoDB. Administer MongoDB-based applications on-premise or in the cloud Scale MongoDB to achieve your design goals Integrate MongoDB with big data sources to process huge amounts of data In Detail MongoDB has grown to become the de facto NoSQL database with millions of users—from small startups to Fortune 500 companies. Addressing the limitations of SQL schema-based databases, MongoDB pioneered a shift of focus for DevOps and offered sharding and replication maintainable by DevOps teams. The book is based on MongoDB 3.x and covers topics ranging from database querying using the shell, built in drivers, and popular ODM mappers to more advanced topics such as sharding, high availability, and integration with big data sources. You will get an overview of MongoDB and how to play to its strengths, with relevant use cases. After that, you will learn how to query MongoDB effectively and make use of indexes as much as possible. The next part deals with the administration of MongoDB installations on-premise or in the cloud. We deal with database internals in the next section, explaining storage systems and how they can affect performance. The last section of this book deals with replication and MongoDB scaling, along with integration with heterogeneous data sources. By the end this book, you will be equipped with all the required industry skills and knowledge to become a certified MongoDB developer and administrator. Style and approach This book takes a practical, step-by-step approach to explain the concepts of MongoDB. Practical use-cases involving real-world examples are used throughout the book to clearly explain theoretical concepts.
*Examples and Case Studies* Springer
This book highlights the emerging field of intelligent computing and developing smart systems. It includes chapters discussing the outcome of challenging research related to distributed computing, smart machines and their security related research, and also covers next-generation communication techniques and the networking technologies that have the potential to build the future communication infrastructure. Bringing together computing, communications and other aspects of intelligent and

smart computing, it contributes to developing a roadmap for future research on intelligent systems.

**Scaling Big Data with Hadoop and Solr - Second Edition** Packt Publishing Ltd

This book is aimed at developers, designers, and architects who would like to build big data enterprise search solutions for their customers or organizations. No prior knowledge of Apache Hadoop and Apache Solr/Lucene technologies is required.

BIG DATA AND HADOOP Notion Press

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.

**Apache Spark in 24 Hours, Sams Teach Yourself** McGraw-Hill Education

Build, process and analyze large-scale graph data effectively with Spark About This Book Find solutions for every stage of data processing from loading and transforming graph data to Improve the scalability of your graphs with a variety of real-world applications with complete Scala code. A concise guide to processing large-scale networks with Apache Spark. Who This Book Is For This book is for data scientists and big data developers who want to learn the processing and analyzing graph datasets at scale. Basic programming experience with Scala is assumed. Basic knowledge of Spark is assumed. What You Will Learn Write, build and deploy Spark applications with the Scala Build Tool. Build and analyze large-scale network datasets Analyze and transform graphs using RDD and graph-specific operations Implement new custom graph operations tailored to specific needs. Develop iterative and efficient graph algorithms using message aggregation and Pregel abstraction Extract subgraphs and use it to discover common clusters Analyze graph data and solve various data science problems using real-world datasets. In Detail Apache Spark is the next standard of open-source cluster-computing engine for processing big data. Many practical computing problems concern large graphs, like the Web graph and various social networks. The scale of these graphs - in some cases billions of vertices, trillions of edges - poses challenges to their efficient processing. Apache Spark GraphX API combines the advantages of both data-parallel and graph-parallel systems by efficiently expressing graph computation within the Spark data-parallel framework. This book will teach the user to do graphical programming in Apache Spark, apart from an explanation of the entire process of graphical data analysis. You will journey through the creation of graphs, its uses, its exploration and analysis and finally will also cover the conversion of graph elements into graph structures. This book begins with an introduction of the Spark system, its libraries and the Scala Build Tool. Using a hands-on approach, this book will quickly teach you how to install and leverage Spark interactively on the command line and in a standalone Scala program. Then, it presents all the methods for building Spark graphs using illustrative network datasets. Next, it will walk you through the process of exploring, visualizing and analyzing different network characteristics. This book will also teach you how to transform raw datasets into a usable form. In addition, you will learn powerful operations that can be used to transform graph elements and graph structures. Furthermore, this book also teaches how to create custom graph operations that are tailored for specific needs with efficiency in mind. The later chapters of this book cover more advanced topics such as clustering graphs, implementing graph-parallel iterative algorithms and learning methods from graph data. Style and approach A step-by-step guide that will walk you through the key ideas and techniques for processing big graph data at scale, with practical examples that will ensure an overall understanding of the concepts of Spark.

*IBM Platform Computing Solutions* Addison-Wesley Professional

Gathering the Proceedings of the 2018 Intelligent Systems Conference (IntelliSys 2018), this book offers a remarkable collection of chapters covering a wide range of topics in intelligent systems and computing, and their real-world applications. The Conference attracted a total of 568 submissions from pioneering researchers, scientists, industrial engineers, and students from all around the world. These submissions underwent a double-blind peer review process, after which 194 (including 13 poster papers) were selected to be included in these proceedings. As intelligent systems continue to replace and sometimes outperform human intelligence in decision-making processes, they have made it possible to tackle many problems more effectively. This branching out of computational intelligence in several directions, and the use of intelligent systems in everyday applications, have created the need for such an international conference, which serves as a venue for reporting on cutting-edge innovations and developments. This book collects both theory and application-based chapters on all aspects of artificial intelligence, from classical to intelligent scope. Readers are sure to find the book both interesting and valuable, as it presents state-of-the-art intelligent methods and techniques for solving real-world problems, along with a vision of future research directions.

**The Artificial Intelligence Infrastructure Workshop** Packt Publishing Ltd

Leverage the power of MongoDB 4.x to build and administer fault-tolerant database applications Key Features Master the new features and capabilities of MongoDB 4.x Implement advanced data modeling, querying, and administration techniques in MongoDB Includes rich case-studies and best practices followed by expert MongoDB developers Book Description MongoDB is the best platform for working with non-relational data and is considered to be the smartest tool for organizing data in line with business needs. The recently released MongoDB 4.x supports ACID transactions and makes the technology an asset for enterprises across the IT and fintech sectors. This book provides expertise in advanced and niche areas of managing databases (such as modeling and querying databases) along with various administration techniques in MongoDB, thereby helping you become a successful MongoDB expert. The book helps you understand how the newly added capabilities function with the help of some interesting examples and large datasets. You will dive deeper into niche areas such as high-performance

configurations, optimizing SQL statements, configuring large-scale sharded clusters, and many more. You will also master best practices in overcoming database failover, and master recovery and backup procedures for database security. By the end of the book, you will have gained a practical understanding of administering database applications both on premises and on the cloud; you will also be able to scale database applications across all servers. What you will learn Perform advanced querying techniques such as indexing and expressions Configure, monitor, and maintain a highly scalable MongoDB environment Master replication and data sharding to optimize read/write performance Administer MongoDB-based applications on premises or on the cloud Integrate MongoDB with big data sources to process huge amounts of data Deploy MongoDB on Kubernetes containers Use MongoDB in IoT, mobile, and serverless environments Who this book is for This book is ideal for MongoDB developers and database administrators who wish to become successful MongoDB experts and build scalable and fault-tolerant applications using MongoDB. It will also be useful for database professionals who wish to become certified MongoDB professionals. Some understanding of MongoDB and basic database concepts is required to get the most out of this book.

Euro-Par 2017: Parallel Processing Packt Publishing Ltd

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 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

Hadoop Real-World Solutions Cookbook Springer Nature

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

Related with Installing Hadoop 2 6 X On Windows 10:
• Como Se Hace El Examen De Alcohol Y Drogas : click here