

Cloudera Vs Hortonworks Vs Mapr 2017 Cloudera Vs

This Book is published by www.HadoopExam.com (HadoopExam Learning Resources). Where you can find material and training's for preparing for BigData, Cloud Computing, Analytics, Data Science and popular Programming Language. This Book will contain 14 chapters, to cover NiFi concepts and providing 9+ use cases, so that you can understand the various fine grain detail about Apache NiFi. Also, it is recommended that you go through the NiFi Hands On Training provided by HadoopExam. In training we have created concepts as well as practicals by creating simple and complex workflow. While publishing this book there are 19 modules available, which are in-line with this book. As you know, NiFi recently become very popular to solve BigData, IOT (Internet of Things) , IOAT (Internet of Anything's) etc. Having an exclusive skill will certainly give you edge with already lack of BigData resources. To help you HadoopExam.com brings full length Hands on training and this book to understand fundamental concepts of NiFi. We provide many Hands On session for creating simple to complex workflow/dataflow to process the data. As this is a continuously growing and fast paced technology. This technology not only helps in working BigData but also, wherever you need complex and simple DataFlow engine you can use this. NiFi can be integrated with existing technology e.g. Spark, HBase, Cassandra, RDBMS, HDFS and can even be customized as per your requirement. So start learning NiFi with HadoopExam.com premium training and book by getting subscription.

Tackle the real-world complexities of modern machine learning with innovative, cutting-edge, techniques About This Book Fully-coded working examples using a wide range of machine learning libraries and tools, including Python, R, Julia, and Spark Comprehensive practical solutions taking you into the future of machine learning Go a step further and integrate your machine learning projects with Hadoop Who This Book Is For This book has been created for data scientists who want to see machine learning in action and explore its real-world application. With guidance on everything from the fundamentals of machine learning and predictive analytics to the latest innovations set to lead the big data revolution into the future, this is an unmissable resource for anyone dedicated to tackling current big data challenges. Knowledge of programming (Python and R) and mathematics is advisable if you want to get started immediately. What You Will Learn Implement a wide range of algorithms and techniques for tackling complex data Get to grips with some of the most powerful languages in data science, including R, Python, and Julia Harness the capabilities of Spark and Hadoop to manage and process data successfully Apply the appropriate machine learning technique to address real-world problems Get acquainted with Deep learning and find out how neural networks are being used at the cutting-edge of machine learning Explore the future of machine learning and dive deeper into polyglot persistence, semantic data, and more In Detail Finding meaning in increasingly larger and more complex datasets is a growing demand of the modern world. Machine learning and predictive analytics have become the most important approaches to uncover data gold mines. Machine learning uses complex algorithms to make improved predictions of outcomes based on historical patterns and the behaviour of data sets. Machine learning can deliver dynamic insights into trends, patterns, and relationships within data, immensely valuable to business growth and development. This book explores an extensive range of machine learning techniques uncovering hidden tricks and tips for several types of data using practical and real-world examples. While machine learning can be highly theoretical, this book offers a refreshing hands-on approach without losing sight of the underlying principles. Inside, a full exploration of the various algorithms gives you high-quality guidance so you can begin to see just how effective machine learning is at tackling contemporary challenges of big data. This is the only book you need to implement a whole suite of open source tools, frameworks, and languages in machine learning. We will cover the leading data science languages, Python and R, and the underrated but powerful Julia, as well as a range of other big data platforms including Spark, Hadoop, and Mahout. Practical Machine Learning is an essential resource for the modern data scientists who want to get to grips with its real-world application. With this book, you will not only learn the fundamentals of machine learning but dive deep into the complexities of real world data before moving on to using Hadoop and its wider ecosystem of tools to process and manage your structured and unstructured data. You will explore different machine learning techniques for both supervised and unsupervised learning; from decision trees to Naive Bayes classifiers and linear and clustering methods, you will learn strategies for a truly advanced approach to the statistical analysis of data. The book also explores the cutting-edge advancements in machine learning, with worked examples and guidance on deep learning and reinforcement learning, providing you with practical demonstrations and samples that help take the theory—and mystery—out of even the most advanced machine learning methodologies. Style and approach A practical data science tutorial designed to give you an insight into the practical application of machine learning, this book takes you through complex concepts and tasks in an accessible way. Featuring information on a wide range of data science techniques, Practical Machine Learning is a comprehensive data science resource.

If you are a Hadoop programmer who wants to learn about Flume to be able to move datasets into Hadoop in a timely and replicable manner, then this book is ideal for you. No prior knowledge about Apache Flume is necessary, but a basic knowledge of Hadoop and the Hadoop File System (HDFS) is assumed.

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 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

This comprehensive edited volume is the first of its kind, designed to serve as a textbook for long-duration business analytics programs. It can also be used as a guide to the field by practitioners. The book has contributions from experts in top universities and industry. The editors have taken extreme care to ensure continuity across the chapters. The material is organized into three parts: A) Tools, B) Models and C) Applications. In Part A, the tools used by business analysts are described in detail. In Part B, these tools are applied to construct models used to solve business problems. Part C contains detailed applications in various functional areas of business and several case studies. Supporting material can be found in the appendices that develop the pre-requisites for the main text. Every chapter has a business orientation. Typically, each chapter begins with the description of business problems that are transformed into data questions; and methodology is developed to solve these questions. Data analysis is conducted using widely used software, the output and results are clearly explained at each stage of development. These are finally transformed into a business solution. The companion website provides examples, data sets and sample code for each chapter.

This book covers three major parts of Big Data: concepts, theories and applications. Written by world-renowned leaders in Big Data, this book explores the problems, possible solutions and directions for Big Data in research and practice. It also focuses on high level concepts such as definitions of Big Data from different angles; surveys in research and applications; and existing tools, mechanisms, and systems in practice. Each chapter is independent from the other chapters, allowing users to read any chapter directly. After examining the practical side of Big Data, this book presents theoretical perspectives. The theoretical research ranges from Big Data representation, modeling and topology to distribution and dimension reducing. Chapters also investigate the many disciplines that involve Big Data, such as statistics, data mining, machine learning, networking, algorithms, security and differential geometry. The last section of this book introduces Big Data applications from different communities, such as business, engineering and science. Big Data Concepts, Theories and Applications is designed as a reference for researchers and advanced level students in computer science, electrical engineering and mathematics. Practitioners who focus on information systems, big data, data mining, business analysis and other related fields will also find this material valuable.

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 extensions

This book is a step-by-step tutorial filled with practical examples which will show you how to build and manage a Hadoop cluster along with its intricacies. This book is ideal for database administrators, data engineers, and system administrators, and it will act as an invaluable reference if you are planning to use the Hadoop platform in your organization. It is expected that you have basic Linux skills since all the examples in this book use this operating system. It is also useful if you have access to test hardware or virtual machines to be able to follow the examples in the book.

Handbook of Research on Big Data Storage and Visualization Techniques IGI Global

Peramalan (forecasting) adalah mengestimasi atau memperkirakan peristiwa atau situasi yang tidak dapat dikendalikan oleh segala hal yang terkait dengan penawaran, permintaan, produksi, dan penggunaan teknologi dalam sebuah industri atau perusahaan di masa depan berdasarkan keadaan data terdahulu dan data sekarang yang diperlukan untuk menentukan kapan suatu peristiwa akan terjadi, sehingga dapat mengetahui tindakan yang tepat untuk dilakukan perencanaan dalam periode tertentu. Python adalah salah satu bahasa pemrograman yang bersifat open source serta sebagai bahasa yang menggabungkan kemampuan, kapabilitas, dengan sintaksis kode yang sangat jelas, dan dilengkapi dengan fungsionalitas pustaka standar yang besar serta komprehensif. Kami berharap dengan terbitnya buku ini memberikan manfaat yang besar bagi para pembaca baik dari mahasiswa, pengajar maupun masyarakat umum yang membutuhkan referensi dibidang forecasting menggunakan python.

As the age of Big Data emerges, it becomes necessary to take the five dimensions of Big Data- volume, variety, velocity, volatility, and veracity- and focus these dimensions towards one critical emphasis - value. The Encyclopedia of Business Analytics and Optimization confronts the challenges of information retrieval in the age of Big Data by exploring recent advances in the areas of knowledge management, data visualization, interdisciplinary communication, and others. Through its critical approach and practical application, this book will be a must-have reference for any professional, leader, analyst, or manager interested in making the most of the knowledge resources at their disposal.

Unlock the power of your data with Hadoop 2.X ecosystem and its data warehousing techniques across large data sets About This Book Conquer the mountain of data using Hadoop 2.X tools The authors succeed in creating a context for Hadoop and its ecosystem Hands-on examples and recipes giving the bigger picture and helping you to master Hadoop 2.X data processing platforms Overcome the challenging data processing problems using this exhaustive course with Hadoop 2.X Who This Book Is For This course is for Java developers, who know scripting, wanting a career shift to Hadoop - Big Data segment of the IT industry. So if you are a novice in Hadoop or an expert, this book will make you reach the most advanced level in Hadoop 2.X. What You Will Learn Best practices for setup and configuration of Hadoop clusters, tailoring the system to the problem at hand Integration with relational databases, using Hive for SQL queries and Sqoop for data transfer Installing and maintaining Hadoop 2.X cluster and its ecosystem Advanced Data Analysis using the Hive, Pig, and Map Reduce programs Machine learning principles with libraries such as Mahout and Batch and Stream data processing using Apache Spark Understand the changes involved in the process in the move from Hadoop 1.0 to Hadoop 2.0 Dive into YARN and Storm and use YARN to integrate Storm with Hadoop Deploy Hadoop on Amazon Elastic MapReduce and Discover HDFS replacements and learn about HDFS Federation In Detail As Marc Andreessen has said "Data is eating the world," which can be witnessed today being the

age of Big Data, businesses are producing data in huge volumes every day and this rise in tide of data need to be organized and analyzed in a more secured way. With proper and effective use of Hadoop, you can build new-improved models, and based on that you will be able to make the right decisions. The first module, Hadoop beginners Guide will walk you through on understanding Hadoop with very detailed instructions and how to go about using it. Commands are explained using sections called “What just happened” for more clarity and understanding. The second module, Hadoop Real World Solutions Cookbook, 2nd edition, is an essential tutorial to effectively implement a big data warehouse in your business, where you get detailed practices on the latest technologies such as YARN and Spark. Big data has become a key basis of competition and the new waves of productivity growth. Hence, once you get familiar with the basics and implement the end-to-end big data use cases, you will start exploring the third module, Mastering Hadoop. So, now the question is if you need to broaden your Hadoop skill set to the next level after you nail the basics and the advance concepts, then this course is indispensable. When you finish this course, you will be able to tackle the real-world scenarios and become a big data expert using the tools and the knowledge based on the various step-by-step tutorials and recipes. Style and approach This course has covered everything right from the basic concepts of Hadoop till you master the advance mechanisms to become a big data expert. The goal here is to help you learn the basic essentials using the step-by-step tutorials and from there moving toward the recipes with various real-world solutions for you. It covers all the important aspects of Hadoop from system designing and configuring Hadoop, machine learning principles with various libraries with chapters illustrated with code fragments and schematic diagrams. This is a compendious course to explore Hadoop from the basics to the most advanced techniques available in Hadoop 2.X.

In this book written for SAP BI, big data, and IT architects, the authors expertly provide clear recommendations for building modern analytics architectures running on SAP HANA technologies. Explore integration with big data frameworks and predictive analytics components. Obtain the tools you need to assess possible architecture scenarios and get guidelines for choosing the best option for your organization. Know your options for on-premise, in the cloud, and hybrid solutions. Readers will be guided through SAP BW/4HANA and SAP HANA native data warehouse scenarios, as well as field-tested integration options with big data platforms. Explore migration options and architecture best practices. Consider organizational and procedural changes resulting from the move to a new, up-to-date analytics architecture that supports your data-driven or data-informed organization. By using practical examples, tips, and screenshots, this book explores: - SAP HANA and SAP BW/4HANA architecture concepts - Predictive Analytics and Big Data component integration - Recommendations for a sustainable, future-proof analytics solutions - Organizational impact and change management Apache® Spark is one of the fastest growing technology in BigData computing world. It supports multiple programming languages like Java, Scala, Python and R. Hence, many existing and new framework started to integrate Spark platform as well in their platform e.g. Hadoop, Cassandra, EMR etc. While creating Spark certification material HadoopExam technical team found that there is no proper material and book is available for the Spark (version 2.x) which covers the concepts as well as use of various features and found difficulty in creating the material. Therefore, they decided to create full length book for Spark (HDPSCD Spark Scala Certification) and outcome of that is this book. In this book technical team try to cover both fundamental concepts of Spark 2.x topics which are part of the certification syllabus as well as add as many exercises as possible and in current version we have around 10 hands on exercises added which you can execute on the Hortonworks sandbox, as this book is focused on the Scala version of the certification, hence all the exercises and their solution provided in the Scala. We have divided the entire book in the 7 chapters, as you move ahead chapter by chapter you would be comfortable with the HDPSCD Spark Scala certification. All the exercises given in this book are written using Scala. However, concepts remain same even if you are using different programming language.

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

Software applications once held on local computers and servers are beginning to shift to the public Internet sphere, and private health information is no exception. The likelihood of placing once restricted and private health records “in the cloud” is increasing. Cloud Computing Applications for Quality Health Care Delivery focuses on cloud technologies that could affect quality in the healthcare field. Leading experts in this area offer their knowledge and contribute to the demystification of healthcare in the Cloud. This publication will prove to be a useful tool for undergraduate and graduate students of healthcare quality and management, healthcare managers, and industry professionals.

Apache Spark is one of the fastest growing technology in BigData computing world. It support multiple programming languages like Java, Scala, Python and R. Hence, many existing and new framework started to integrate Spark platform as well in their platform e.g. Hadoop, Cassandra, EMR etc. While creating Spark certification material HadoopExam technical team found that there is no proper material and book is available for the Spark SQL (version 2.x) which covers the concepts as well as use of various features and found difficulty in creating the material. Therefore, they decided to create full length book for Spark SQL and outcome of that is this book. In this book technical team try to cover both fundamental concepts of Spark SQL engine and many exercises approx. 35+ so that most of the programming features can be covered. There are approximately 35 exercises and total 15 chapters which covers the programming aspects of SparkSQL. All the exercises given in this book are written using Scala. However, concepts remain same even if you are using different programming language. This book is good for following audience - Data scientists - Spark Developer - Data Engineer - Data Analytics - Java/Python Developer - Scala Developer

A comprehensive guide to mastering the most advanced Hadoop 3 concepts Key Features Get to grips with the newly introduced features and capabilities of Hadoop 3 Crunch and process data using MapReduce, YARN, and a host of tools within the Hadoop ecosystem Sharpen your Hadoop skills with real-world case studies and code Book Description Apache Hadoop is one of the most popular big data solutions for distributed storage and for processing large chunks of data. With Hadoop 3, Apache promises to provide a high-performance, more fault-tolerant, and highly efficient big data processing platform, with a focus on improved scalability and increased efficiency. With this guide, you’ll understand advanced concepts of the Hadoop ecosystem tool. You’ll learn how Hadoop works internally, study advanced concepts of different ecosystem tools, discover solutions to real-world use cases, and understand how to secure your cluster. It will then walk you through HDFS, YARN, MapReduce, and Hadoop 3 concepts. You’ll be able to address common challenges like using Kafka efficiently, designing low latency, reliable message delivery Kafka systems, and handling high data volumes. As you advance, you’ll discover how to address major challenges when building an enterprise-grade messaging system, and how to use different stream processing systems along with Kafka to fulfil your enterprise goals. By the end of this book, you’ll have a complete understanding of how components in the Hadoop ecosystem are effectively integrated to implement a fast and reliable data pipeline, and you’ll be equipped to tackle a range of

real-world problems in data pipelines. What you will learn Gain an in-depth understanding of distributed computing using Hadoop 3 Develop enterprise-grade applications using Apache Spark, Flink, and more Build scalable and high-performance Hadoop data pipelines with security, monitoring, and data governance Explore batch data processing patterns and how to model data in Hadoop Master best practices for enterprises using, or planning to use, Hadoop 3 as a data platform Understand security aspects of Hadoop, including authorization and authentication Who this book is for If you want to become a big data professional by mastering the advanced concepts of Hadoop, this book is for you. You'll also find this book useful if you're a Hadoop professional looking to strengthen your knowledge of the Hadoop ecosystem. Fundamental knowledge of the Java programming language and basics of Hadoop is necessary to get started with this book.

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

Big Data Application Architecture Pattern Recipes provides an insight into heterogeneous infrastructures, databases, and visualization and analytics tools used for realizing the architectures of big data solutions. Its problem-solution approach helps in selecting the right architecture to solve the problem at hand. In the process of reading through these problems, you will learn harness the power of new big data opportunities which various enterprises use to attain real-time profits. Big Data Application Architecture Pattern Recipes answers one of the most critical questions of this time 'how do you select the best end-to-end architecture to solve your big data problem?'. The book deals with various mission critical problems encountered by solution architects, consultants, and software architects while dealing with the myriad options available for implementing a typical solution, trying to extract insight from huge volumes of data in real-time and across multiple relational and non-relational data types for clients from industries like retail, telecommunication, banking, and insurance. The patterns in this book provide the strong architectural foundation required to launch your next big data application. The architectures for realizing these opportunities are based on relatively less expensive and heterogeneous infrastructures compared to the traditional monolithic and hugely expensive options that exist currently. This book describes and evaluates the benefits of heterogeneity which brings with it multiple options of solving the same problem, evaluation of trade-offs and validation of 'fitness-for-purpose' of the solution.

This book unravels the mystery of Big Data computing and its power to transform business operations. The approach it uses will be helpful to any professional who must present a case for realizing Big Data computing solutions or to those who could be involved in a Big Data computing project. It provides a framework that enables business and technical managers to make optimal decisions necessary for the successful migration to Big Data computing environments and applications within their organizations.

The digital age has presented an exponential growth in the amount of data available to individuals looking to draw conclusions based on given or collected information across industries. Challenges associated with the analysis, security, sharing, storage, and visualization of large and complex data sets continue to plague data scientists and analysts alike as traditional data processing applications struggle to adequately manage big data. The Handbook of Research on Big Data Storage and Visualization Techniques is a critical scholarly resource that explores big data analytics and technologies and their role in developing a broad understanding of issues pertaining to the use of big data in multidisciplinary fields. Featuring coverage on a broad range of topics, such as architecture patterns, programming systems, and computational energy, this publication is geared towards professionals, researchers, and students seeking current research and application topics on the subject.

Dive into the world of SQL on Hadoop and get the most out of your Hive data warehouses. This book is your go-to resource for using Hive: authors Scott Shaw, Ankur Gupta, David Kjerrumgaard, and Andreas Francois Vermeulen take you through learning HiveQL, the SQL-like language specific to Hive, to analyze, export, and massage the data stored across your Hadoop environment. From deploying Hive on your hardware or virtual machine and setting up its initial configuration to learning how Hive interacts with Hadoop, MapReduce, Tez and other big data technologies, Practical Hive gives you a detailed treatment of the software. In addition, this book discusses the value of open source software, Hive performance tuning, and how to leverage semi-structured and unstructured data. What You Will Learn Install and configure Hive for new and existing datasets Perform DDL operations Execute efficient DML operations Use tables, partitions, buckets, and user-defined functions Discover performance tuning tips and Hive best practices Who This Book Is For Developers, companies, and professionals who deal with large amounts of data and could use software that can efficiently manage large volumes of input. It is assumed that readers have the ability to work with SQL.

Le développement spectaculaire d'internet, des réseaux sociaux, de la technologie mobile et la multiplication des capteurs provoquent une croissance exponentielle des données à laquelle les entreprises doivent faire face : c'est le phénomène du Big Data. Ses enjeux sont considérables. Au-delà de la simple question technique du stockage, il offre la possibilité de tirer profit du contenu de ces nouvelles sources d'information. Les solutions décisionnelles classiques laissent progressivement place au Business Analytics et aux méthodes prédictives, transformant l'avalanche de données en valeur ajoutée. La technologie est aujourd'hui disponible, les bases de données traditionnelles ont évolué et les solutions dédiées à l'exploitation des données massives, telles que Hadoop, sont désormais opérationnelles. S'appuyant sur différents cas pratiques, Enjeux et usages du Big Data met l'accent sur les méthodes, les techniques et les ressources nécessaires pour permettre aux entreprises d'entrer avec succès dans l'ère de l'information à grande échelle.

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. This book is also meant for Hadoop professionals who want to find solutions to the different challenges they come across in their Hadoop projects.

Apache Hadoop is the technology at the heart of the Big Data revolution, and Hadoop skills are in enormous demand. Now, in just 24 lessons of one hour or less, you can learn all the skills and techniques you'll need to deploy each key component of a Hadoop platform in your local environment or in the cloud, building a fully functional Hadoop cluster and using it with real programs and datasets. Each short, easy lesson builds on all that's come before, helping you master all of Hadoop's essentials, and extend it to meet your unique challenges. Apache Hadoop

in 24 Hours, Sams Teach Yourself covers all this, and much more: Understanding Hadoop and the Hadoop Distributed File System (HDFS) Importing data into Hadoop, and process it there Mastering basic MapReduce Java programming, and using advanced MapReduce API concepts Making the most of Apache Pig and Apache Hive Implementing and administering YARN Taking advantage of the full Hadoop ecosystem Managing Hadoop clusters with Apache Ambari Working with the Hadoop User Environment (HUE) Scaling, securing, and troubleshooting Hadoop environments Integrating Hadoop into the enterprise Deploying Hadoop in the cloud Getting started with Apache Spark Step-by-step instructions walk you through common questions, issues, and tasks; Q-and-As, Quizzes, and Exercises build and test your knowledge; "Did You Know?" tips offer insider advice and shortcuts; and "Watch Out!" alerts help you avoid pitfalls. By the time you're finished, you'll be comfortable using Apache Hadoop to solve a wide spectrum of Big Data problems.

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.

R for Cloud Computing looks at some of the tasks performed by business analysts on the desktop (PC era) and helps the user navigate the wealth of information in R and its 4000 packages as well as transition the same analytics using the cloud. With this information the reader can select both cloud vendors and the sometimes confusing cloud ecosystem as well as the R packages that can help process the analytical tasks with minimum effort, cost and maximum usefulness and customization. The use of Graphical User Interfaces (GUI) and Step by Step screenshot tutorials is emphasized in this book to lessen the famous learning curve in learning R and some of the needless confusion created in cloud computing that hinders its widespread adoption. This will help you kick-start analytics on the cloud including chapters on both cloud computing, R, common tasks performed in analytics including the current focus and scrutiny of Big Data Analytics, setting up and navigating cloud providers. Readers are exposed to a breadth of cloud computing choices and analytics topics without being buried in needless depth. The included references and links allow the reader to pursue business analytics on the cloud easily. It is aimed at practical analytics and is easy to transition from existing analytical set up to the cloud on an open source system based primarily on R. This book is aimed at industry practitioners with basic programming skills and students who want to enter analytics as a profession. Note the scope of the book is neither statistical theory nor graduate level research for statistics, but rather it is for business analytics practitioners. It will also help researchers and academics but at a practical rather than conceptual level. The R statistical software is the fastest growing analytics platform in the world, and is established in both academia and corporations for robustness, reliability and accuracy. The cloud computing paradigm is firmly established as the next generation of computing from microprocessors to desktop PCs to cloud.

As information systems become ever more pervasive in an increasing number of fields and professions, workers in healthcare and medicine must take into consideration new advances in technologies and infrastructure that will better enable them to treat their patients and serve their communities. Healthcare Administration: Concepts, Methodologies, Tools, and Applications brings together recent research and case studies in the medical field to explore topics such as hospital management, delivery of patient care, and telemedicine, among others. With a focus on some of the most groundbreaking new developments as well as future trends and critical concerns, this three-volume reference source will be a significant tool for medical practitioners, hospital managers, IT administrators, and others actively engaged in the healthcare field.

This book constitutes the refereed proceedings of the 12th Colombian Conference on Computing, CCC 2017, held in Cali, Colombia, in September 2017. The 56 revised full papers presented were carefully reviewed and selected from 186 submissions. The papers are organized in topical sections on information and knowledge management, software engineering and IT architectures, educational informatics, intelligent systems and robotics, human-computer interaction, distributed systems and large-scale architectures, image processing, computer vision and multimedia, security of the information, formal methods, computational logic and theory of computation.

Investors and technology gurus have called big data one of the most important trends to come along in decades. Big Data Bootcamp explains what big data is and how you can use it in your company to become one of tomorrow's market leaders. Along the way, it explains the very latest technologies, companies, and advancements. Big data holds the keys to delivering better customer service, offering more attractive products, and unlocking innovation. That's why, to remain competitive, every organization should become a big data company. It's also why every manager and technology professional should become knowledgeable about big data and how it is transforming not just their own industries but the global economy. And that knowledge is just what this book delivers. It explains components of big data like Hadoop and NoSQL databases; how big data is compiled, queried, and analyzed; how to create a big data application; and the business sectors ripe for big data-inspired products and services like retail, healthcare, finance, and education. Best of all, your guide is David Feinleib, renowned entrepreneur, venture capitalist, and author of Why Startups Fail. Feinleib's Big Data Landscape, a market map featured and explained in the book, is an industry benchmark that has been viewed more than 150,000 times and is used as a reference by VMWare, Dell, Intel, the U.S. Government Accountability Office, and many other organizations. Feinleib also explains:

- Why every businessperson needs to understand the fundamentals of big data or get run over by those who do
- How big data differs from traditional database management systems
- How to create and run a big data project
- The technical details powering the big data revolution

Whether you're a Fortune 500 executive or the proprietor of a restaurant or web design studio, Big Data Bootcamp will explain how you can take full advantage of new technologies to transform your company and your career.

Learn Big Data from the ground up with this complete and up-to-date resource from leaders in the field Big Data: Concepts, Technology, and Architecture delivers a comprehensive treatment of Big Data tools, terminology, and technology perfectly suited to a wide range of business professionals, academic researchers, and students. Beginning with a fulsome overview of what we mean when we say, "Big Data," the book moves on to discuss every stage of the lifecycle of Big Data. You'll learn about the creation of structured, unstructured, and semi-structured data, data storage solutions, traditional database solutions like SQL, data processing, data analytics, machine learning, and data mining. You'll also discover how specific technologies like Apache Hadoop, SQOOP, and Flume work. Big Data also covers the central topic of big data visualization with Tableau, and you'll learn how to create scatter plots, histograms, bar, line, and pie charts with that software. Accessibly organized, Big Data includes illuminating case studies throughout the material, showing you how the included concepts have been applied in real-world settings. Some of those concepts include: The common challenges facing big data technology and technologists, like data heterogeneity and incompleteness, data volume and velocity,

storage limitations, and privacy concerns Relational and non-relational databases, like RDBMS, NoSQL, and NewSQL databases Virtualizing Big Data through encapsulation, partitioning, and isolating, as well as big data server virtualization Apache software, including Hadoop, Cassandra, Avro, Pig, Mahout, Oozie, and Hive The Big Data analytics lifecycle, including business case evaluation, data preparation, extraction, transformation, analysis, and visualization Perfect for data scientists, data engineers, and database managers, Big Data also belongs on the bookshelves of business intelligence analysts who are required to make decisions based on large volumes of information. Executives and managers who lead teams responsible for keeping or understanding large datasets will also benefit from this book.

Use Hadoop to solve business problems by learning from a rich set of real-life case studies About This Book Solve real-world business problems using Hadoop and other Big Data technologies Build efficient data lakes in Hadoop, and develop systems for various business cases like improving marketing campaigns, fraud detection, and more Power packed with six case studies to get you going with Hadoop for Business Intelligence Who This Book Is For If you are interested in building efficient business solutions using Hadoop, this is the book for you This book assumes that you have basic knowledge of Hadoop, Java, and any scripting language. What You Will Learn Learn about the evolution of Hadoop as the big data platform Understand the basics of Hadoop architecture Build a 360 degree view of your customer using Sqoop and Hive Build and run classification models on Hadoop using BigML Use Spark and Hadoop to build a fraud detection system Develop a churn detection system using Java and MapReduce Build an IoT-based data collection and visualization system Get to grips with building a Hadoop-based Data Lake for large enterprises Learn about the coexistence of NoSQL and In-Memory databases in the Hadoop ecosystem In Detail If you have a basic understanding of Hadoop and want to put your knowledge to use to build fantastic Big Data solutions for business, then this book is for you. Build six real-life, end-to-end solutions using the tools in the Hadoop ecosystem, and take your knowledge of Hadoop to the next level. Start off by understanding various business problems which can be solved using Hadoop. You will also get acquainted with the common architectural patterns which are used to build Hadoop-based solutions. Build a 360-degree view of the customer by working with different types of data, and build an efficient fraud detection system for a financial institution. You will also develop a system in Hadoop to improve the effectiveness of marketing campaigns. Build a churn detection system for a telecom company, develop an Internet of Things (IoT) system to monitor the environment in a factory, and build a data lake – all making use of the concepts and techniques mentioned in this book. The book covers other technologies and frameworks like Apache Spark, Hive, Sqoop, and more, and how they can be used in conjunction with Hadoop. You will be able to try out the solutions explained in the book and use the knowledge gained to extend them further in your own problem space. Style and approach This is an example-driven book where each chapter covers a single business problem and describes its solution by explaining the structure of a dataset and tools required to process it. Every project is demonstrated with a step-by-step approach, and explained in a very easy-to-understand manner.

The 6th FTRA International Conference on Computer Science and its Applications (CSA-14) will be held in Guam, USA, Dec. 17 - 19, 2014. CSA-14 presents a comprehensive conference focused on the various aspects of advances in engineering systems in computer science, and applications, including ubiquitous computing, U-Health care system, Big Data, UI/UX for human-centric computing, Computing Service, Bioinformatics and Bio-Inspired Computing and will show recent advances on various aspects of computing technology, Ubiquitous Computing Services and its application.

If you are an IT administrator and you want to enter the world of cloud storage using OpenStack Swift, then this book is ideal for you. Basic knowledge of Linux and server technology is beneficial to get the most out of the book.

A practical guide to implementing your enterprise data lake using Lambda Architecture as the base About This Book Build a full-fledged data lake for your organization with popular big data technologies using the Lambda architecture as the base Delve into the big data technologies required to meet modern day business strategies A highly practical guide to implementing enterprise data lakes with lots of examples and real-world use-cases Who This Book Is For Java developers and architects who would like to implement a data lake for their enterprise will find this book useful. If you want to get hands-on experience with the Lambda Architecture and big data technologies by implementing a practical solution using these technologies, this book will also help you. What You Will Learn Build an enterprise-level data lake using the relevant big data technologies Understand the core of the Lambda architecture and how to apply it in an enterprise Learn the technical details around Sqoop and its functionalities Integrate Kafka with Hadoop components to acquire enterprise data Use flume with streaming technologies for stream-based processing Understand stream- based processing with reference to Apache Spark Streaming Incorporate Hadoop components and know the advantages they provide for enterprise data lakes Build fast, streaming, and high-performance applications using ElasticSearch Make your data ingestion process consistent across various data formats with configurability Process your data to derive intelligence using machine learning algorithms In Detail The term "Data Lake" has recently emerged as a prominent term in the big data industry. Data scientists can make use of it in deriving meaningful insights that can be used by businesses to redefine or transform the way they operate. Lambda architecture is also emerging as one of the very eminent patterns in the big data landscape, as it not only helps to derive useful information from historical data but also correlates real-time data to enable business to take critical decisions. This book tries to bring these two important aspects — data lake and lambda architecture—together. This book is divided into three main sections. The first introduces you to the concept of data lakes, the importance of data lakes in enterprises, and getting you up-to-speed with the Lambda architecture. The second section delves into the principal components of building a data lake using the Lambda architecture. It introduces you to popular big data technologies such as Apache Hadoop, Spark, Sqoop, Flume, and ElasticSearch. The third section is a highly practical demonstration of putting it all together, and shows you how an enterprise data lake can be implemented, along with several real-world use-cases. It also shows you how other peripheral components can be added to the lake to make it more efficient. By the end of this book, you will be able to choose the right big data technologies using the lambda architectural patterns to build your enterprise data lake. Style and approach The book takes a pragmatic approach, showing ways to leverage big data technologies and lambda architecture to build an enterprise-level data lake.

Pragmatic Enterprise Architecture is a practical hands-on instruction manual for enterprise architects. This book prepares you to better engage IT, management, and business users by equipping you with the

tools and knowledge you need to address the most common enterprise architecture challenges. You will come away with a pragmatic understanding of and approach to enterprise architecture and actionable ideas to transform your enterprise. Experienced enterprise architect James V. Luisi generously shares life cycle architectures, transaction path analysis frameworks, and more so you can save time, energy, and resources on your next big project. As an enterprise architect, you must have relatable frameworks and excellent communication skills to do your job. You must actively engage and support a large enterprise involving a hundred architectural disciplines with a modest number of subject matter experts across business, information systems, control systems, and operations architecture. They must achieve their mission using the influence of ideas and business benefits expressed in simple terms so that any audience can understand what to do and why. Pragmatic Enterprise Architecture gives you the tools to accomplish your goals in less time with fewer resources. Expand your Enterprise Architecture skills so you can do more in less time with less money with the priceless tips presented Understand the cost of creating new Enterprise Architecture disciplines and contrast those costs to letting them go unmanaged Includes 10 life cycle architectures so that you can properly assess the ROI of performing activities such as outsourcing, insourcing, restructuring, mergers and acquisitions, and more Complete appendix of eight transaction path analysis frameworks provide DBA guidelines for proper physical database design

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. 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.

Learn all you need to know about seven key innovations disrupting business analytics today. These innovations—the open source business model, cloud analytics, the Hadoop ecosystem, Spark and in-memory analytics, streaming analytics, Deep Learning, and self-service analytics—are radically changing how businesses use data for competitive advantage. Taken together, they are disrupting the business analytics value chain, creating new opportunities. Enterprises who seize the opportunity will thrive and prosper, while others struggle and decline: disrupt or be disrupted. Disruptive Business Analytics provides strategies to profit from disruption. It shows you how to organize for insight, build and provision an open source stack, how to practice lean data warehousing, and how to assimilate disruptive innovations into an organization. Through a short history of business analytics and a detailed survey of products and services, analytics authority Thomas W. Dinsmore provides a practical explanation of the most compelling innovations available today. What You'll Learn Discover how the open source business model works and how to make it work for you See how cloud computing completely changes the economics of analytics Harness the power of Hadoop and its ecosystem Find out why Apache Spark is everywhere Discover the potential of streaming and real-time analytics Learn what Deep Learning can do and why it matters See how self-service analytics can change the way organizations do business Who This Book Is For Corporate actors at all levels of responsibility for analytics: analysts, CIOs, CTOs, strategic decision makers, managers, systems architects, technical marketers, product developers, IT personnel, and consultants.

Tap the power of Big Data with Microsoft technologies Big Data is here, and Microsoft's new Big Data platform is a valuable tool to help your company get the very most out of it. This timely book shows you how to use HDInsight along with HortonWorks Data Platform for Windows to store, manage, analyze, and share Big Data throughout the enterprise. Focusing primarily on Microsoft and HortonWorks technologies but also covering open source tools, Microsoft Big Data Solutions explains best practices, covers on-premises and cloud-based solutions, and features valuable case studies. Best of all, it helps you integrate these new solutions with technologies you already know, such as SQL Server and Hadoop. Walks you through how to integrate Big Data solutions in your company using Microsoft's HDInsight Server, HortonWorks Data Platform for Windows, and open source tools Explores both on-premises and cloud-based solutions Shows how to store, manage, analyze, and share Big Data through the enterprise Covers topics such as Microsoft's approach to Big Data, installing and configuring HortonWorks Data Platform for Windows, integrating Big Data with SQL Server, visualizing data with Microsoft and HortonWorks BI tools, and more Helps you build and execute a Big Data plan Includes contributions from the Microsoft and HortonWorks Big Data product teams If you need a detailed roadmap for designing and implementing a fully deployed Big Data solution, you'll want Microsoft Big Data Solutions.

This book briefly covers internationally contributed chapters with artificial intelligence and applied mathematics-oriented background-details. Nowadays, the world is under attack of intelligent systems covering all fields to make them practical and meaningful for humans. In this sense, this edited book provides the most recent research on use of engineering capabilities for developing intelligent systems. The chapters are a collection from the works presented at the 2nd International Conference on Artificial Intelligence and Applied Mathematics in Engineering held within 09-10-11 October 2020 at the Antalya, Manavgat (Turkey). The target audience of the book covers scientists, experts, M.Sc. and Ph.D. students, post-docs, and anyone interested in intelligent systems and their usage in different problem domains. The book is suitable to be used as a reference work in the courses associated with artificial intelligence and applied mathematics.

[Copyright: 1dd0c6d593b49a7f6c662bb1c5221e54](#)