Apache Hive Tutorialspoint

Hadoop Operations
Professional Visual Studio 2017
Hadoop: The Definitive Guide
Data Analysis Using Big Data & Hadoop Framework
Proceedings of the International Conference on Soft Computing Systems
Big Data Analytics and Computing for Digital Forensic Investigations
Hadoop For Dummies
Big Data Integration
Frank Kane's Taming Big Data with Apache Spark and Python
Hadoop Practice Guide
Big Data Infrastructure Technologies for Data Analytics
Hadoop Essentials
Parallel and Concurrent Programming in Haskell
Bridging Relational and NoSQL Databases
Introduction to Biomedical Data Science
Hadoop Beginner's Guide
Information Systems Design and Intelligent Applications
Microsoft Azure Essentials - Fundamentals of Azure
Proceedings of 2nd International Conference on Communication, Computing and Networking
Internet of Things and Big Data Technologies for Next Generation Healthcare
Advances in Computational and Bio-Engineering
The Evolution of Business in the Cyber Age
Hands-On Big Data Analytics with PySpark
Apache Hive Cookbook
Hadoop: The Definitive Guide
Expert Hadoop Administration

Handbook of Research on Big Data Storage and Visualization Techniques
From Visual Surveillance to Internet of Things
Python Tutorial 3.11.3
JSON at Work
Big Data Analytics
Programming Hive
Hadoop in Practice
Google BigQuery Analytics
Hadoop in Action
DATA SCIENCE MENGGUNAKAN BAHASA R
Proceedings of International Conference on Recent Advancement on Computer and Communication
Liferay in Action
Learning Spark

CUMMINGS DAKOTA

Hadoop Operations "O'Reilly Media, Inc." 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

Professional Visual Studio 2017

Addison-Wesley Professional

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, programing systems, and computational energy, this publication is geared towards professionals, researchers, and students seeking current research and application topics on the subject.

Hadoop: The Definitive Guide John Wiley & Sons

Relational databases have been predominant for many years and are used throughout various industries. The current system faces challenges related to size and variety of data thus the NoSQL databases emerged. By joining these two database models, there is room for crucial developments in the field of computer science. Bridging Relational and NoSQL Databases is an innovative source of academic content on the convergence process between databases and describes key features of the next database generation. Featuring coverage on a wide variety of topics and perspectives such as BASE approach, CAP theorem, and hybrid and native solutions, this publication is

ideally designed for professionals and researchers interested in the features and collaboration of relational and NoSQL databases.

Data Analysis Using Big Data & Hadoop Framework Springer

Digital forensics has recently gained a notable development and become the most demanding area in today's information security requirement. This book investigates the areas of digital forensics, digital investigation and data analysis procedures as they apply to computer fraud and cybercrime, with the main objective of describing a variety of digital crimes and retrieving potential digital evidence. Big Data Analytics and Computing for Digital Forensic Investigations gives a contemporary view on the problems of information security. It presents the idea that protective mechanisms and software must be integrated along with forensic capabilities into existing forensic software using big data computing tools and techniques. Features Describes trends of digital forensics served for big data and the challenges of evidence acquisition Enables digital forensic investigators and law enforcement agencies to enhance their digital investigation capabilities with the application of data science analytics, algorithms and fusion technique This book is focused on helping professionals as well as researchers to get ready with next-generation security systems to mount the rising challenges of computer fraud and cybercrimes as well as with digital forensic investigations. Dr Suneeta Satpathy has more than ten years of teaching experience in different subjects of the Computer Science and Engineering discipline. She is currently working as an associate professor in the Department of Computer Science and Engineering, College of Bhubaneswar, affiliated with Biju Patnaik University and Technology, Odisha. Her research interests include computer forensics, cybersecurity, data fusion, data mining, big data analysis and decision mining. Dr Sachi Nandan Mohanty is an associate professor in the Department of Computer Science and Engineering at ICFAI Tech, ICFAI Foundation for Higher Education, Hyderabad, India. His research interests include data mining, big data analysis, cognitive science, fuzzy decision-making, brain–computer interface, cognition and computational intelligence.

Proceedings of the International Conference on Soft Computing Systems

Manning Publications

Information Systems Design and Intelligent ApplicationsSpringer

Big Data Analytics and Computing for Digital Forensic Investigations

Springer

Summary Liferay in Action is a comprehensive and authoritative guide to building portals on the Liferay 6 platform. Fully supported and authorized by Liferay, this book guides you smoothly from your first exposure to Liferay through the crucial day-to-day tasks of building and maintaining an enterprise portal that works well within your existing IT infrastructure. About the Technology A portal is a website built around a collection of components that request, display, and share information. Liferay Portal 6, an enterprise-ready development platform, makes it a snap to build portals that integrate with your existing backend systems and provide a rich interactive user experience. Because Liferay uses standard Java and JavaScript, along with

built-in SOAP and JSON support for web services, developers can be productive immediately. And since it's available in both a free, open source version as well as a fully-supported commercial edition, it's an affordable solution for almost any business or organization About the Book Liferay in Action is the official guide to building Liferay portal applications using Java and JavaScript. If you've never used Liferay before, don't worry. This book starts with the basics: setting up your development environment and creating a working portal. Then, it builds on that foundation to help you discover social features, tagging, ratings, and more. You'll also explore the Portlet 2.0 API, and learn to create custom themes and reusable templates. Experienced developers will learn how to use new Liferay APIs to build social and collaborative sites, use the message bus and workflow, implement indexing and search, and more. This book was developed in close collaboration with Liferay engineers, so it answers the right questions, and answers them in depth. No experience with Liferay or the Portlets API is required, but basic knowledge of Java and web technology is assumed. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book. What's Inside Complete coverage of Liferay Portal 6 Covers both the commercial and open source versions Custom portlet development using the Portlet 2.0 spec Liferay's social network API Add functionality with hooks and Ext plugins ========== ========== Table of Contents PART 1 WORKING WITH LIFERAY AND PORTLETS The Liferay difference Getting started with the Liferay development platform PART 2 WRITING APPLICATIONS ON LIFERAY'S PLATFORM A data-driven portlet made easy MVC the Liferay way Designing your site with themes and layout templates Making your site social Enabling user collaboration PART 3 CUSTOMIZING LIFERAY Hooks Extending Liferay effectively A tour of Liferay APIs **Hadoop For Dummies** CRC Press Frank Kane's hands-on Spark training course, based on his bestselling Taming Big Data with Apache Spark and Python video, now available in a book. Understand and analyze large data sets using Spark on a single system or on a cluster. About This Book Understand how Spark can be distributed across computing clusters Develop and run Spark jobs efficiently using Python A hands-on tutorial by Frank Kane with over 15 real-world examples teaching you Big Data processing with Spark Who This Book Is For If you are a data scientist or data analyst who wants to learn Big Data processing using Apache Spark and Python, this book is for you. If you have some programming experience in Python, and want to learn how to process large amounts of data using Apache Spark, Frank Kane's Taming Big Data with Apache Spark and Python will also help you. What You Will Learn Find out how you can identify Big Data problems as Spark problems Install and run Apache Spark on your computer or on a cluster Analyze large data sets across many CPUs using Spark's Resilient Distributed Datasets Implement machine learning on Spark using the MLlib library Process continuous streams of data in real time using the Spark streaming module Perform complex network analysis using Spark's GraphX library Use Amazon's Elastic MapReduce service to run your

Spark jobs on a cluster In Detail Frank Kane's Taming Big Data with Apache Spark and Python is your companion to learning Apache Spark in a hands-on manner. Frank will start you off by teaching you how to set up Spark on a single system or on a cluster, and you'll soon move on to analyzing large data sets using Spark RDD, and developing and running effective Spark jobs quickly using Python. Apache Spark has emerged as the next big thing in the Big Data domain – quickly rising from an ascending technology to an established superstar in just a matter of years. Spark allows you to quickly extract actionable insights from large amounts of data, on a real-time basis, making it an essential tool in many modern businesses. Frank has packed this book with over 15 interactive, fun-filled examples relevant to the real world, and he will empower you to understand the Spark ecosystem and implement production-grade real-time Spark projects with ease. Style and approach Frank Kane's Taming Big Data with Apache Spark and Python is a hands-on tutorial with over 15 real-world examples carefully explained by Frank in a step-by-step manner. The examples vary in complexity, and you can move through them at your own pace.

Big Data Integration IGI Global

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.

Frank Kane's Taming Big Data with Apache Spark and Python "O'Reilly Media, Inc."

Use PySpark to easily crush messy data at-scale and discover proven techniques to create testable, immutable, and easily parallelizable Spark jobs Key FeaturesWork with large amounts of agile data using distributed datasets and in-memory cachingSource data from all popular data hosting platforms, such as HDFS, Hive, JSON, and S3Employ the easy-to-use PySpark API to deploy big data Analytics for productionBook Description Apache Spark is an open source parallel-processing framework that has been around for quite some time now. One of the many uses of Apache Spark is for data analytics applications across clustered computers. In this book, you will not only learn how to use Spark and the Python API to create high-performance analytics with big data, but also discover techniques for testing, immunizing, and parallelizing Spark jobs. You will learn how to source data from all popular data hosting platforms, including HDFS, Hive, JSON, and S3, and deal with large datasets with PySpark to gain practical big data experience. This book will help you work on prototypes on local machines and subsequently go on to handle messy data

in production and at scale. This book covers installing and setting up PySpark, RDD operations, big data cleaning and wrangling, and aggregating and summarizing data into useful reports. You will also learn how to implement some practical and proven techniques to improve certain aspects of programming and administration in Apache Spark. By the end of the book, you will be able to build big data analytical solutions using the various PySpark offerings and also optimize them effectively. What you will learnGet practical big data experience while working on messy datasetsAnalyze patterns with Spark SQL to improve your business intelligenceUse PySpark's interactive shell to speed up development timeCreate highly concurrent Spark programs by leveraging immutabilityDiscover ways to avoid the most expensive operation in the Spark API: the shuffle operationRe-design your jobs to use reduceByKey instead of groupByCreate robust processing pipelines by testing Apache Spark jobsWho this book is for This book is for developers, data scientists, business analysts, or anyone who needs to reliably analyze large amounts of large-scale, real-world data. Whether you're tasked with creating your company's business intelligence function or creating great data platforms for your machine learning models, or are looking to use code to magnify the impact of your business, this book is for you.

Hadoop Practice Guide John Wiley & Sons
If you have a working knowledge of Haskell, this hands-on book shows you how to use the language’s many APIs and frameworks for writing both parallel and concurrent programs. You’ll learn how parallelism exploits multicore processors to speed up computation-heavy programs, and how concurrency enables you to write programs with threads for multiple interactions. Author Simon Marlow walks you through the process with lots of code examples that you can run, experiment with, and extend. Divided into separate sections on Parallel and Concurrent Haskell, this book also includes exercises to help you become familiar with the concepts presented: Express parallelism in Haskell with the Eval monad and Evaluation Strategies Parallelize ordinary Haskell code with the Par monad Build parallel array-based computations, using the Repa library Use the Accelerate library to run computations directly on the GPU Work with basic interfaces for writing concurrent code Build trees of threads for larger and more complex programs Learn how to build high-speed concurrent network servers Write distributed programs that run on multiple machines in a network

Big Data Infrastructure Technologies for Data Analytics IGI Global
Overview of biomedical data science -- Spreadsheet tools and tips -- Biostatistics primer -- Data visualization -- Introduction to databases -- Big data -- Bioinformatics and precision medicine -- Programming languages for data analysis -- Machine learning -- Artificial intelligence -- Biomedical data science resources -- Appendix A: Glossary -- Appendix B: Using data.world -- Appendix C: Chapter exercises.

Hadoop Essentials Springer Nature
This book has a two-fold mission: to explain and facilitate digital transition in business organizations using information and communications technology and to address the associated growing threat of cyber crime and the challenge of creating

and maintaining effective cyber protection. The book begins with a section on Digital Business Transformation, which includes chapters on tools for integrated marketing communications, human resource workplace digitalization, the integration of the Internet of Things in the workplace, Big Data, and more. The technologies discussed aim to help businesses and entrepreneurs transform themselves to align with today's modern digital climate. The Evolution of Business in the Cyber Age: Digital Transformation, Threats, and Security provides a wealth of information for those involved in the development and management of conducting business online as well as for those responsible for cyber protection and security. Faculty and students, researchers, and industry professionals will find much of value in this volume.

Parallel and Concurrent Programming in Haskell CRC Press

The book provides insights from the 2nd International Conference on Communication, Computing and Networking organized by the Department of Computer Science and Engineering, National Institute of Technical Teachers Training and Research, Chandigarh, India on March 29–30, 2018. The book includes contributions in which researchers, engineers, and academicians as well as industrial professionals from around the globe presented their research findings and development activities in the field of Computing Technologies, Wireless Networks, Information Security, Image Processing and Data Science. The book provides opportunities for the readers to explore the literature, identify gaps in the existing works and propose new ideas for research.

Bridging Relational and NoSQL Databases Springer

JSON is becoming the backbone for meaningful data interchange over the internet. This format is now supported by an entire ecosystem of standards, tools, and technologies for building truly elegant, useful, and efficient applications. With this hands-on guide, author and architect Tom Marrs shows you how to build enterprise-class applications and services by leveraging JSON tooling and message/document design. JSON at Work provides application architects and developers with guidelines, best practices, and use cases, along with lots of real-world examples and code samples. You'll start with a comprehensive JSON overview, explore the JSON ecosystem, and then dive into JSON's use in the enterprise. Get acquainted with JSON basics and learn how to model JSON data Learn how to use JSON with Node.js, Ruby on Rails, and Java Structure JSON documents with JSON Schema to design and test APIs Search the contents of JSON documents with JSON Search tools Convert JSON documents to other data formats with JSON Transform tools Compare JSON-based hypermedia formats, including HAL and jsonapi Leverage MongoDB to store and access JSON documents Use Apache Kafka to exchange JSON-based messages between services

Introduction to Biomedical Data Science "O'Reilly Media, Inc."

The big data era is upon us: data are being generated, analyzed, and used at an unprecedented scale, and data-driven decision making is sweeping through all aspects of society. Since the value of data explodes when it can be linked and fused with other data, addressing the big data integration (BDI) challenge is critical to

realizing the promise of big data. BDI differs from traditional data integration along the dimensions of volume, velocity, variety, and veracity. First, not only can data sources contain a huge volume of data, but also the number of data sources is now in the millions. Second, because of the rate at which newly collected data are made available, many of the data sources are very dynamic, and the number of data sources is also rapidly exploding. Third, data sources are extremely heterogeneous in their structure and content, exhibiting considerable variety even for substantially similar entities. Fourth, the data sources are of widely differing qualities, with significant differences in the coverage, accuracy and timeliness of data provided. This book explores the progress that has been made by the data integration community on the topics of schema alignment, record linkage and data fusion in addressing these novel challenges faced by big data integration. Each of these topics is covered in a systematic way: first starting with a quick tour of the topic in the context of traditional data integration, followed by a detailed, example-driven exposition of recent innovative techniques that have been proposed to address the BDI challenges of volume, velocity, variety, and veracity. Finally, it presents merging topics and opportunities that are specific to BDI, identifying promising directions for the data integration community.

Hadoop Beginner's Guide "O'Reilly Media, Inc."

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

Information Systems Design and Intelligent Applications Microsoft Press

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

Microsoft Azure Essentials - Fundamentals of Azure Springer

Data Science merupakan suatu bidang ilmu yang (relatif) baru di Indonesia yang tujuan utamanya adalah "memahami dan menganalisis segala fenomena yang berkaitan dengan data yang bertipe terstruktur, semi terstruktur, serta tidak terstruktur", yang datang dari berbagai sumber data/datasource yangsangat beragam (meskipun saat ini sumber data yang terutama berasal dari Big Data dan/atau loT/Internet of Things) menggunakan pengetahuan (yang terutama) Matematika, Statistika, Ilmu Informasi (Information Science), serta Ilmu Komputer (terutama Kecerdasan Buatan/AI-Artificial Intelligence). McKinsey & Company pada tahun 2018 melakukan survei tentang berapa sesungguhnya kebutuhan Data Scientist di Amerika Serikat pada tahun itu dan hasilnya menunjukkan bahwa dibutuhkan sekitar 140.000-190.000 Data Scientist (peningkatan sekitar 6,5 kali lipat dibandingkan tahun sebelumnya). Dalam hal ini, kita juga bisa memperkirakan bahwa kebutuhan Data Scientist di Indonesia dalam beberapa tahun mendatang juga akan sangat tinggi (terutama jika kita mengingat salah satu kebijakan ekonomi pemerintah Indonesia saat ini yang pada tahun- tahun mendatang akan lebih fokus pada berbagai bentuk perekonomian digital seperti (contohnya) perdagangan barang dan jasa melalui jaringan Internet [e-Commerce]). Dengan demikian, buku yang berjudul Data Science Menggunakan Bahasa R: Analisis Data, Visualisasi, serta Pemodelan ini diharapkan akan memberi pemahaman tentang beberapa metode Data Science yang paling sering digunakan oleh seorang Data Scientist

untuk melakukan berbagai analisis data (data analytics) kepada kalangan akademisi para mahasiswa di jurusan-jurusan yang terkait dengan Ilmu Komputer dan/atau Informatika, dan pada para praktisi di seluruh Indonesia.

Proceedings of 2nd International Conference on Communication, Computing and Networking Packt Publishing Ltd

This comprehensive book focuses on better big-data security for healthcare organizations. Following an extensive introduction to the Internet of Things (IoT) in healthcare including challenging topics and scenarios, it offers an in-depth analysis of medical body area networks with the 5th generation of IoT communication technology along with its nanotechnology. It also describes a novel strategic framework and computationally intelligent model to measure possible security vulnerabilities in the context of e-health. Moreover, the book addresses healthcare systems that handle large volumes of data driven by patients' records and health/personal information, including big-data-based knowledge management systems to support clinical decisions. Several of the issues faced in storing/processing big data are presented along with the available tools, technologies and algorithms to deal with those problems as well as a case study in healthcare analytics. Addressing trust, privacy, and security issues as well as the IoT and big-data challenges, the book highlights the advances in the field to guide engineers developing different IoT devices and evaluating the performance of different IoT techniques. Additionally, it explores the impact of such technologies on public, private, community, and hybrid scenarios in healthcare. This book offers professionals, scientists and engineers the latest technologies, techniques, and strategies for IoT and big data.

Internet of Things and Big Data Technologies for Next Generation Healthcare CRC Press

Big data analytics emerged as a revolution in the field of information technology. It is the ability of the organization to stay agile which gives it a competitive edge over its competitors. Data harvesting and data analytics enable the organization identify new opportunities which in turn results in efficient operations, leads to smarter business moves and higher business turnovers. All these issues are addressed by big data analytics and its initiatives. Chapter 4 focuses on architecture of Pig, Apache Pig execution modes, Pig data types and operators. Apache Pig Latin data model is based on nested relations. The chapter provides description of different components of Pig Latin data model. The lab session includes installing Pig over Hadoop and exploring different Pig Latin operators.Chapter 5 deals with common services provides by zookeeper, architecture and components of zookeeper and zookeeper operation modes. The salient feature of the chapter is exploration of leader election algorithm and security of ZNodes through access control list. The chapter concludes with the hands-on lab sessions on installation of zookeeper and exposure to zookeeper command-line interface. Chapter 6 discusses different types of No SQL databases, transformation rules from one data model to another and performs in-depth analysis of HBase data model. The features which are difficult to comprehend such as data compaction, data locality, HBase read and write operations are simplified with easy to understand figures

and explanation. As a part of hands-on lab sessions, installation of HBase over Hadoop and exercises based on HBase general commands, DDL commands and DML commands are dealt with.