

Beginning Apache Pig Springer

Recognizing the pretension ways to get this books **Beginning Apache Pig Springer** is additionally useful. You have remained in right site to start getting this info. acquire the Beginning Apache Pig Springer member that we come up with the money for here and check out the link.

You could purchase guide Beginning Apache Pig Springer or acquire it as soon as feasible. You could speedily download this Beginning Apache Pig Springer after getting deal. So, past you require the ebook swiftly, you can straight get it. Its for that reason enormously easy and therefore fats, isnt it? You have to favor to in this freshen

Processing Big Data with Azure HDInsight -

Vinit Yadav 2017-05-29

Get a jump start on using Azure HDInsight and Hadoop Ecosystem components. As most Hadoop and Big Data projects are written in either Java, Scala, or Python, this book minimizes the effort to learn another language and is written from the perspective of a .NET

developer. Hadoop components are covered, including Hive, Pig, HBase, Storm, and Spark on Azure HDInsight, and code samples are written in .NET only. Processing Big Data with Azure HDInsight covers the fundamentals of big data, how businesses are using it to their advantage, and how Azure HDInsight fits into the big data world. This book introduces Hadoop and big

data concepts and then dives into creating different solutions with HDInsight and the Hadoop Ecosystem. It covers concepts with real-world scenarios and code examples, making sure you get hands-on experience. The best way to utilize this book is to practice while reading. After reading this book you will be familiar with Azure HDInsight and how it can be utilized to build big data solutions, including batch processing, stream analytics, interactive processing, and storing and retrieving data in an efficient manner. What You'll Learn Understand the fundamentals of HDInsight and Hadoop Work with HDInsight cluster Query with Apache Hive and Apache Pig Store and retrieve data with Apache HBase Stream data processing using Apache Storm Work with Apache Spark Who This Book Is For Software developers, technical architects, data scientists/analysts, and Hadoop administrators who want to develop on Microsoft's managed Hadoop offering, HDInsight

Secure Data Science - Bhavani Thuraisingham
2022-04-20

Secure data science, which integrates cyber security and data science, is becoming one of the critical areas in both cyber security and data science. This is because the novel data science techniques being developed have applications in solving such cyber security problems as intrusion detection, malware analysis, and insider threat detection. However, the data science techniques being applied not only for cyber security but also for every application area—including healthcare, finance, manufacturing, and marketing—could be attacked by malware. Furthermore, due to the power of data science, it is now possible to infer highly private and sensitive information from public data, which could result in the violation of individual privacy. This is the first such book that provides a comprehensive overview of integrating both cyber security and data science and discusses both theory and practice in secure

data science. After an overview of security and privacy for big data services as well as cloud computing, this book describes applications of data science for cyber security applications. It also discusses such applications of data science as malware analysis and insider threat detection. Then this book addresses trends in adversarial machine learning and provides solutions to the attacks on the data science techniques. In particular, it discusses some emerging trends in carrying out trustworthy analytics so that the analytics techniques can be secured against malicious attacks. Then it focuses on the privacy threats due to the collection of massive amounts of data and potential solutions. Following a discussion on the integration of services computing, including cloud-based services for secure data science, it looks at applications of secure data science to information sharing and social media. This book is a useful resource for researchers, software developers, educators, and managers who want to understand both the

high level concepts and the technical details on the design and implementation of secure data science-based systems. It can also be used as a reference book for a graduate course in secure data science. Furthermore, this book provides numerous references that would be helpful for the reader to get more details about secure data science.

Big Data Preprocessing - Julián Luengo

2020-03-16

This book offers a comprehensible overview of Big Data Preprocessing, which includes a formal description of each problem. It also focuses on the most relevant proposed solutions. This book illustrates actual implementations of algorithms that helps the reader deal with these problems. This book stresses the gap that exists between big, raw data and the requirements of quality data that businesses are demanding. This is called Smart Data, and to achieve Smart Data the preprocessing is a key step, where the imperfections, integration tasks and other

processes are carried out to eliminate superfluous information. The authors present the concept of Smart Data through data preprocessing in Big Data scenarios and connect it with the emerging paradigms of IoT and edge computing, where the end points generate Smart Data without completely relying on the cloud. Finally, this book provides some novel areas of study that are gathering a deeper attention on the Big Data preprocessing. Specifically, it considers the relation with Deep Learning (as of a technique that also relies in large volumes of data), the difficulty of finding the appropriate selection and concatenation of preprocessing techniques applied and some other open problems. Practitioners and data scientists who work in this field, and want to introduce themselves to preprocessing in large data volume scenarios will want to purchase this book. Researchers that work in this field, who want to know which algorithms are currently implemented to help their investigations, may

also be interested in this book.

[Engineering Vibration, Communication and Information Processing](#) - Kanad Ray 2018-10-30
This book discusses the revolution of cycles and rhythms that is expected to take place in different branches of science and engineering in the 21st century, with a focus on communication and information processing. It presents high-quality papers in vibration sciences, rhythms and oscillations, neurosciences, mathematical sciences, and communication. It includes major topics in engineering and structural mechanics, computer sciences, biophysics and biomathematics, as well as other related fields. Offering valuable insights, it also inspires researchers to work in these fields. The papers included in this book were presented at the 1st International Conference on Engineering Vibration, Communication and Information Processing (ICoEVCI-2018), India.
Processing Big Data with Azure HDInsight - Vinit Yadav 2017-06-01

Get a jump start on using Azure HDInsight and Hadoop Ecosystem components. As most Hadoop and Big Data projects are written in either Java, Scala, or Python, this book minimizes the effort to learn another language and is written from the perspective of a .NET developer. Hadoop components are covered, including Hive, Pig, HBase, Storm, and Spark on Azure HDInsight, and code samples are written in .NET only. Processing Big Data with Azure HDInsight covers the fundamentals of big data, how businesses are using it to their advantage, and how Azure HDInsight fits into the big data world. This book introduces Hadoop and big data concepts and then dives into creating different solutions with HDInsight and the Hadoop Ecosystem. It covers concepts with real-world scenarios and code examples, making sure you get hands-on experience. The best way to utilize this book is to practice while reading. After reading this book you will be familiar with Azure HDInsight and how it can be utilized to

build big data solutions, including batch processing, stream analytics, interactive processing, and storing and retrieving data in an efficient manner. What You'll Learn Understand the fundamentals of HDInsight and Hadoop Work with HDInsight cluster Query with Apache Hive and Apache Pig Store and retrieve data with Apache HBase Stream data processing using Apache Storm Work with Apache Spark Who This Book Is For Software developers, technical architects, data scientists/analysts, and Hadoop administrators who want to develop on Microsoft's managed Hadoop offering, HDInsight

Handbook of Cloud Computing - Borko Furht
2010-09-11

Cloud computing has become a significant technology trend. Experts believe cloud computing is currently reshaping information technology and the IT marketplace. The advantages of using cloud computing include cost savings, speed to market, access to greater

computing resources, high availability, and scalability. Handbook of Cloud Computing includes contributions from world experts in the field of cloud computing from academia, research laboratories and private industry. This book presents the systems, tools, and services of the leading providers of cloud computing; including Google, Yahoo, Amazon, IBM, and Microsoft. The basic concepts of cloud computing and cloud computing applications are also introduced. Current and future technologies applied in cloud computing are also discussed. Case studies, examples, and exercises are provided throughout. Handbook of Cloud Computing is intended for advanced-level students and researchers in computer science and electrical engineering as a reference book. This handbook is also beneficial to computer and system infrastructure designers, developers, business managers, entrepreneurs and investors within the cloud computing related industry.

Big Data Processing Using Spark in Cloud -

Mamta Mittal 2018-06-16

The book describes the emergence of big data technologies and the role of Spark in the entire big data stack. It compares Spark and Hadoop and identifies the shortcomings of Hadoop that have been overcome by Spark. The book mainly focuses on the in-depth architecture of Spark and our understanding of Spark RDDs and how RDD complements big data's immutable nature, and solves it with lazy evaluation, cacheable and type inference. It also addresses advanced topics in Spark, starting with the basics of Scala and the core Spark framework, and exploring Spark data frames, machine learning using Mllib, graph analytics using Graph X and real-time processing with Apache Kafka, AWS Kinesis, and Azure Event Hub. It then goes on to investigate Spark using PySpark and R. Focusing on the current big data stack, the book examines the interaction with current big data tools, with Spark being the core processing layer for all types of data. The book is intended for data

engineers and scientists working on massive datasets and big data technologies in the cloud. In addition to industry professionals, it is helpful for aspiring data processing professionals and students working in big data processing and cloud computing environments.

Big Data Analytics with Applications in Insider Threat Detection - Bhavani Thuraisingham
2017-11-22

Today's malware mutates randomly to avoid detection, but reactively adaptive malware is more intelligent, learning and adapting to new computer defenses on the fly. Using the same algorithms that antivirus software uses to detect viruses, reactively adaptive malware deploys those algorithms to outwit antivirus defenses and to go undetected. This book provides details of the tools, the types of malware the tools will detect, implementation of the tools in a cloud computing framework and the applications for insider threat detection.

Assistive Technology Intervention in

Healthcare - Shruti Jain 2021-12-31

Assistive Technology Intervention in Healthcare focuses on various applications of intelligent techniques in biomedical engineering and health informatics. It aims to create awareness about disability reduction and recovery of accidental disability with the help of various rehabilitative systems. Novel technologies in disability treatment, management and assistance, including healthcare devices and their utility from home to hospital, are described. The book deals with simulation, modeling, measurement, control, analysis, information extraction and monitoring of physiological data in clinical medicine and biology. Features Covers the latest evolutionary approaches to solve optimization problems in the biomedical engineering field Explains machine learning-based approaches to improvement in health engineering areas Reviews the IoT, cloud computing and data analytics in healthcare informatics Discusses modeling and simulations in the design of

biomedical equipment Explores monitoring of physiological data This book is aimed at researchers and graduate students in biomedical engineering, clinical engineering and bioinformatics.

New Horizons for a Data-Driven Economy -

José María Cavanillas 2016-04-04

In this book readers will find technological discussions on the existing and emerging technologies across the different stages of the big data value chain. They will learn about legal aspects of big data, the social impact, and about education needs and requirements. And they will discover the business perspective and how big data technology can be exploited to deliver value within different sectors of the economy. The book is structured in four parts: Part I “The Big Data Opportunity” explores the value potential of big data with a particular focus on the European context. It also describes the legal, business and social dimensions that need to be addressed, and briefly introduces the European

Commission’s BIG project. Part II “The Big Data Value Chain” details the complete big data lifecycle from a technical point of view, ranging from data acquisition, analysis, curation and storage, to data usage and exploitation. Next, Part III “Usage and Exploitation of Big Data” illustrates the value creation possibilities of big data applications in various sectors, including industry, healthcare, finance, energy, media and public services. Finally, Part IV “A Roadmap for Big Data Research” identifies and prioritizes the cross-sectorial requirements for big data research, and outlines the most urgent and challenging technological, economic, political and societal issues for big data in Europe. This compendium summarizes more than two years of work performed by a leading group of major European research centers and industries in the context of the BIG project. It brings together research findings, forecasts and estimates related to this challenging technological context that is becoming the major axis of the new

digitally transformed business environment.
Handbook of Systems Engineering and Risk Management in Control Systems, Communication, Space Technology, Missile, Security and Defense Operations - Anna M.

Doro-on 2022-09-27

This book provides multifaceted components and full practical perspectives of systems engineering and risk management in security and defense operations with a focus on infrastructure and manpower control systems, missile design, space technology, satellites, intercontinental ballistic missiles, and space security. While there are many existing selections of systems engineering and risk management textbooks, there is no existing work that connects systems engineering and risk management concepts to solidify its usability in the entire security and defense actions. With this book Dr. Anna M. Doro-on rectifies the current imbalance. She provides a comprehensive overview of systems engineering and risk

management before moving to deeper practical engineering principles integrated with newly developed concepts and examples based on industry and government methodologies. The chapters also cover related points including design principles for defeating and deactivating improvised explosive devices and land mines and security measures against kinds of threats. The book is designed for systems engineers in practice, political risk professionals, managers, policy makers, engineers in other engineering fields, scientists, decision makers in industry and government and to serve as a reference work in systems engineering and risk management courses with focus on security and defense operations.

Beginning Apache Spark 2 - Hien Luu
2018-08-16

Develop applications for the big data landscape with Spark and Hadoop. This book also explains the role of Spark in developing scalable machine learning and analytics applications with Cloud

technologies. Beginning Apache Spark 2 gives you an introduction to Apache Spark and shows you how to work with it. Along the way, you'll discover resilient distributed datasets (RDDs); use Spark SQL for structured data; and learn stream processing and build real-time applications with Spark Structured Streaming. Furthermore, you'll learn the fundamentals of Spark ML for machine learning and much more. After you read this book, you will have the fundamentals to become proficient in using Apache Spark and know when and how to apply it to your big data applications. What You Will Learn Understand Spark unified data processing platform How to run Spark in Spark Shell or Databricks Use and manipulate RDDs Deal with structured data using Spark SQL through its operations and advanced functions Build real-time applications using Spark Structured Streaming Develop intelligent applications with the Spark Machine Learning library Who This Book Is For Programmers and developers active

in big data, Hadoop, and Java but who are new to the Apache Spark platform.

Deep Learning with Hadoop - Dipayan Dev
2017-02-20

Build, implement and scale distributed deep learning models for large-scale datasets About This Book Get to grips with the deep learning concepts and set up Hadoop to put them to use Implement and parallelize deep learning models on Hadoop's YARN framework A comprehensive tutorial to distributed deep learning with Hadoop Who This Book Is For If you are a data scientist who wants to learn how to perform deep learning on Hadoop, this is the book for you. Knowledge of the basic machine learning concepts and some understanding of Hadoop is required to make the best use of this book. What You Will Learn Explore Deep Learning and various models associated with it Understand the challenges of implementing distributed deep learning with Hadoop and how to overcome it Implement Convolutional Neural Network (CNN)

with deeplearning4j Delve into the implementation of Restricted Boltzmann Machines (RBM) Understand the mathematical explanation for implementing Recurrent Neural Networks (RNN) Get hands on practice of deep learning and their implementation with Hadoop. In Detail This book will teach you how to deploy large-scale dataset in deep neural networks with Hadoop for optimal performance. Starting with understanding what deep learning is, and what the various models associated with deep neural networks are, this book will then show you how to set up the Hadoop environment for deep learning. In this book, you will also learn how to overcome the challenges that you face while implementing distributed deep learning with large-scale unstructured datasets. The book will also show you how you can implement and parallelize the widely used deep learning models such as Deep Belief Networks, Convolutional Neural Networks, Recurrent Neural Networks, Restricted Boltzmann Machines and

autoencoder using the popular deep learning library deeplearning4j. Get in-depth mathematical explanations and visual representations to help you understand the design and implementations of Recurrent Neural network and Denoising AutoEncoders with deeplearning4j. To give you a more practical perspective, the book will also teach you the implementation of large-scale video processing, image processing and natural language processing on Hadoop. By the end of this book, you will know how to deploy various deep neural networks in distributed systems using Hadoop. Style and approach This book takes a comprehensive, step-by-step approach to implement efficient deep learning models on Hadoop. It starts from the basics and builds the readers' knowledge as they strengthen their understanding of the concepts. Practical examples are included in every step of the way to supplement the theory.

NoSQL - Ganesh Chandra Deka 2017-05-19

This book discusses the advanced databases for the cloud-based application known as NoSQL. It will explore the recent advancements in NoSQL database technology. Chapters on structured, unstructured and hybrid databases will be included to explore bigdata analytics, bigdata storage and processing. The book is likely to cover a wide range of topics such as cloud computing, social computing, bigdata and advanced databases processing techniques.

Beginning Apache Spark 3 - Hien Luu

2021-10-23

Take a journey toward discovering, learning, and using Apache Spark 3.0. In this book, you will gain expertise on the powerful and efficient distributed data processing engine inside of Apache Spark; its user-friendly, comprehensive, and flexible programming model for processing data in batch and streaming; and the scalable machine learning algorithms and practical utilities to build machine learning applications. Beginning Apache Spark 3 begins by explaining

different ways of interacting with Apache Spark, such as Spark Concepts and Architecture, and Spark Unified Stack. Next, it offers an overview of Spark SQL before moving on to its advanced features. It covers tips and techniques for dealing with performance issues, followed by an overview of the structured streaming processing engine. It concludes with a demonstration of how to develop machine learning applications using Spark MLlib and how to manage the machine learning development lifecycle. This book is packed with practical examples and code snippets to help you master concepts and features immediately after they are covered in each section. After reading this book, you will have the knowledge required to build your own big data pipelines, applications, and machine learning applications. What You Will Learn Master the Spark unified data analytics engine and its various components Work in tandem to provide a scalable, fault tolerant and performant data processing engine Leverage the user-

friendly and flexible programming model to perform simple to complex data analytics using dataframe and Spark SQL Develop machine learning applications using Spark MLlib Manage the machine learning development lifecycle using MLflow Who This Book Is For Data scientists, data engineers and software developers.

Beginning Apache Pig - Balaswamy Vaddeman
2016-12-10

Learn to use Apache Pig to develop lightweight big data applications easily and quickly. This book shows you many optimization techniques and covers every context where Pig is used in big data analytics. Beginning Apache Pig shows you how Pig is easy to learn and requires relatively little time to develop big data applications. The book is divided into four parts: the complete features of Apache Pig; integration with other tools; how to solve complex business problems; and optimization of tools. You'll discover topics such as MapReduce and why it

cannot meet every business need; the features of Pig Latin such as data types for each load, store, joins, groups, and ordering; how Pig workflows can be created; submitting Pig jobs using Hue; and working with Oozie. You'll also see how to extend the framework by writing UDFs and custom load, store, and filter functions. Finally you'll cover different optimization techniques such as gathering statistics about a Pig script, joining strategies, parallelism, and the role of data formats in good performance. What You Will Learn

- Use all the features of Apache Pig
- Integrate Apache Pig with other tools
- Extend Apache Pig
- Optimize Pig Latin code
- Solve different use cases for Pig Latin

Who This Book Is For All levels of IT professionals: architects, big data enthusiasts, engineers, developers, and big data administrators

Network Data Analytics - K. G. Srinivasa
2018-04-26

In order to carry out data analytics, we need powerful and flexible computing software.

However the software available for data analytics is often proprietary and can be expensive. This book reviews Apache tools, which are open source and easy to use. After providing an overview of the background of data analytics, covering the different types of analysis and the basics of using Hadoop as a tool, it focuses on different Hadoop ecosystem tools, like Apache Flume, Apache Spark, Apache Storm, Apache Hive, R, and Python, which can be used for different types of analysis. It then examines the different machine learning techniques that are useful for data analytics, and how to visualize data with different graphs and charts. Presenting data analytics from a practice-oriented viewpoint, the book discusses useful tools and approaches for data analytics, supported by concrete code examples. The book is a valuable reference resource for graduate students and professionals in related fields, and is also of interest to general readers with an understanding of data analytics.

Advances in Artificial Intelligence and Data Engineering - Niranjan N. Chiplunkar
2020-08-13

This book presents selected peer-reviewed papers from the International Conference on Artificial Intelligence and Data Engineering (AIDE 2019). The topics covered are broadly divided into four groups: artificial intelligence, machine vision and robotics, ambient intelligence, and data engineering. The book discusses recent technological advances in the emerging fields of artificial intelligence, machine learning, robotics, virtual reality, augmented reality, bioinformatics, intelligent systems, cognitive systems, computational intelligence, neural networks, evolutionary computation, speech processing, Internet of Things, big data challenges, data mining, information retrieval, and natural language processing. Given its scope, this book can be useful for students, researchers, and professionals interested in the growing applications of artificial intelligence and

data engineering.

Big Data Analytics - V. B. Aggarwal 2017-10-03

This volume comprises the select proceedings of the annual convention of the Computer Society of India. Divided into 10 topical volumes, the proceedings present papers on state-of-the-art research, surveys, and succinct reviews. The volumes cover diverse topics ranging from communications networks to big data analytics, and from system architecture to cyber security. This volume focuses on Big Data Analytics. The contents of this book will be useful to researchers and students alike.

Big Data - Min Chen 2014-05-05

This Springer Brief provides a comprehensive overview of the background and recent developments of big data. The value chain of big data is divided into four phases: data generation, data acquisition, data storage and data analysis. For each phase, the book introduces the general background, discusses technical challenges and reviews the latest advances. Technologies under

discussion include cloud computing, Internet of Things, data centers, Hadoop and more. The authors also explore several representative applications of big data such as enterprise management, online social networks, healthcare and medical applications, collective intelligence and smart grids. This book concludes with a thoughtful discussion of possible research directions and development trends in the field. Big Data: Related Technologies, Challenges and Future Prospects is a concise yet thorough examination of this exciting area. It is designed for researchers and professionals interested in big data or related research. Advanced-level students in computer science and electrical engineering will also find this book useful.

Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 2 - Suresh Chandra Satapathy 2016-05-03

This volume contains 60 papers presented at

ICTIS 2015: International Conference on Information and Communication Technology for Intelligent Systems. The conference was held during 28th and 29th November, 2015, Ahmedabad, India and organized communally by Venus International College of Technology, Association of Computer Machinery, Ahmedabad Chapter and Supported by Computer Society of India Division IV - Communication and Division V - Education and Research. This volume contains papers mainly focused on ICT and its application for Intelligent Computing, Cloud Storage, Data Mining, Image Processing and Software Analysis etc.

Deep Learning with Hadoop - Dipayan Dev
2017-02-28

Implement and scale effective deep learning models on Hadoop About This Book* Get to grips with the deep learning concepts and set up Hadoop to put them to use.* Implement and parallelize deep learning models on Hadoop's YARN framework.* A comprehensive tutorial to

distributed deep learning with Hadoop Who This Book Is For If you are a data scientist who wants to learn how to perform deep learning on Hadoop, this is the book for you. Knowledge of the basic machine learning concepts and some understanding of Hadoop is required to make the best use of this book. What you will learn* Explore Deep Learning and various models associated with it.* Understand the challenges of implementing distributed deep learning with Hadoop and how to overcome it* Implement Convolutional Neural Network (CNN) with deeplearning4j* Delve into the implementation of Restricted Boltzmann Machines (RBM)* Understand the mathematical explanation for implementing Recurrent Neural Networks (RNN)* Get hands on practice of deep learning and their implementation with Hadoop. In Detail Deep Learning involves extracting features and insights from multiple layers of the data. This book will teach you how to deploy the deep learning networks with Hadoop. Starting with

understanding what deep learning is and what the various models associated with deep learning are, this book will then show you how to set up the Hadoop environment for deep learning. In this book, you will also learn how to overcome the challenges that you face while implementing distributed deep learning with Hadoop. The book will also show you how you can implement and parallelize Deep Belief Networks, CNN, RNN, RBM and much more using the popular deep learning library `deeplearning4j`. Get in depth mathematical explanations, visual representations to understand the implementation of Denoising AutoEncoders with `deeplearning4j`. To give you a more practical perspective, the book will also teach you how you can implement image classification, audio processing and natural language processing on Hadoop. By the end of this book, you will know how to deploy deep learning in distributed systems using Hadoop

Resilience in the Digital Age - Fred S. Roberts

2021-02-19

The growth of a global digital economy has enabled rapid communication, instantaneous movement of funds, and availability of vast amounts of information. With this come challenges such as the vulnerability of digitalized sociotechnological systems (STSs) to destructive events (earthquakes, disease events, terrorist attacks). Similar issues arise for disruptions to complex linked natural and social systems (from changing climates, evolving urban environments, etc.). This book explores new approaches to the resilience of sociotechnological and natural-social systems in a digital world of big data, extraordinary computing capacity, and rapidly developing methods of Artificial Intelligence. Most of the book's papers were presented at the Workshop on Big Data and Systems Analysis held at the International Institute for Applied Systems Analysis in Laxenburg, Austria in February, 2020. Their authors are associated with the Task

Group “Advanced mathematical tools for data-driven applied systems analysis” created and sponsored by CODATA in November, 2018. The world-wide COVID-19 pandemic illustrates the vulnerability of our healthcare systems, supply chains, and social infrastructure, and confronts our notions of what makes a system resilient. We have found that use of AI tools can lead to problems when unexpected events occur. On the other hand, the vast amounts of data available from sensors, satellite images, social media, etc. can also be used to make modern systems more resilient. Papers in the book explore disruptions of complex networks and algorithms that minimize departure from a previous state after a disruption; introduce a multigrammatical framework for the technological and resource bases of today’s large-scale industrial systems and the transformations resulting from disruptive events; and explain how robotics can enhance pre-emptive measures or post-disaster responses to increase resiliency. Other papers

explore current directions in data processing and handling and principles of FAIRness in data; how the availability of large amounts of data can aid in the development of resilient STSs and challenges to overcome in doing so. The book also addresses interactions between humans and built environments, focusing on how AI can inform today’s smart and connected buildings and make them resilient, and how AI tools can increase resilience to misinformation and its dissemination.

[Practical Hadoop Migration](#) - Bhushan Lakhe
2016-08-10

Re-architect relational applications to NoSQL, integrate relational database management systems with the Hadoop ecosystem, and transform and migrate relational data to and from Hadoop components. This book covers the best-practice design approaches to re-architecting your relational applications and transforming your relational data to optimize concurrency, security, denormalization, and

performance. Winner of IBM's 2012 Gerstner Award for his implementation of big data and data warehouse initiatives and author of Practical Hadoop Security, author Bhushan Lakhe walks you through the entire transition process. First, he lays out the criteria for deciding what blend of re-architecting, migration, and integration between RDBMS and HDFS best meets your transition objectives. Then he demonstrates how to design your transition model. Lakhe proceeds to cover the selection criteria for ETL tools, the implementation steps for migration with SQOOP- and Flume-based data transfers, and transition optimization techniques for tuning partitions, scheduling aggregations, and redesigning ETL. Finally, he assesses the pros and cons of data lakes and Lambda architecture as integrative solutions and illustrates their implementation with real-world case studies. Hadoop/NoSQL solutions do not offer by default certain relational technology features such as role-based

access control, locking for concurrent updates, and various tools for measuring and enhancing performance. Practical Hadoop Migration shows how to use open-source tools to emulate such relational functionalities in Hadoop ecosystem components. What You'll Learn Decide whether you should migrate your relational applications to big data technologies or integrate them Transition your relational applications to Hadoop/NoSQL platforms in terms of logical design and physical implementation Discover RDBMS-to-HDFS integration, data transformation, and optimization techniques Consider when to use Lambda architecture and data lake solutions Select and implement Hadoop-based components and applications to speed transition, optimize integrated performance, and emulate relational functionalities Who This Book Is For Database developers, database administrators, enterprise architects, Hadoop/NoSQL developers, and IT leaders. Its secondary readership is project and

program managers and advanced students of database and management information systems.

Machine Learning Using R - Karthik Ramasubramanian 2016-12-22

Examine the latest technological advancements in building a scalable machine learning model with Big Data using R. This book shows you how to work with a machine learning algorithm and use it to build a ML model from raw data. All practical demonstrations will be explored in R, a powerful programming language and software environment for statistical computing and graphics. The various packages and methods available in R will be used to explain the topics. For every machine learning algorithm covered in this book, a 3-D approach of theory, case-study and practice will be given. And where appropriate, the mathematics will be explained through visualization in R. All the images are available in color and hi-res as part of the code download. This new paradigm of teaching machine learning will bring about a radical

change in perception for many of those who think this subject is difficult to learn. Though theory sometimes looks difficult, especially when there is heavy mathematics involved, the seamless flow from the theoretical aspects to example-driven learning provided in this book makes it easy for someone to connect the dots..

What You'll Learn Use the model building process flow Apply theoretical aspects of machine learning Review industry-based case studies Understand ML algorithms using R Build machine learning models using Apache Hadoop and Spark Who This Book is For Data scientists, data science professionals and researchers in academia who want to understand the nuances of machine learning approaches/algorithms along with ways to see them in practice using R. The book will also benefit the readers who want to understand the technology behind implementing a scalable machine learning model using Apache Hadoop, Hive, Pig and Spark.

Engineering Vibration, Communication and

Downloaded from ravishingbeasts.com
on by guest

Information Processing - Kanad Ray

2019-11-15

Big Data Applications in Industry 4.0 - P. Kaliraj

2022-02-10

Industry 4.0 is the latest technological innovation in manufacturing with the goal to increase productivity in a flexible and efficient manner. Changing the way in which manufacturers operate, this revolutionary transformation is powered by various technology advances including Big Data analytics, Internet of Things (IoT), Artificial Intelligence (AI), and cloud computing. Big Data analytics has been identified as one of the significant components of Industry 4.0, as it provides valuable insights for smart factory management. Big Data and Industry 4.0 have the potential to reduce resource consumption and optimize processes, thereby playing a key role in achieving sustainable development. Big Data Applications in Industry 4.0 covers the recent advancements

that have emerged in the field of Big Data and its applications. The book introduces the concepts and advanced tools and technologies for representing and processing Big Data. It also covers applications of Big Data in such domains as financial services, education, healthcare, biomedical research, logistics, and warehouse management. Researchers, students, scientists, engineers, and statisticians can turn to this book to learn about concepts, technologies, and applications that solve real-world problems. Features An introduction to data science and the types of data analytics methods accessible today An overview of data integration concepts, methodologies, and solutions A general framework of forecasting principles and applications, as well as basic forecasting models including naïve, moving average, and exponential smoothing models A detailed roadmap of the Big Data evolution and its related technological transformation in computing, along with a brief description of

related terminologies The application of Industry 4.0 and Big Data in the field of education The features, prospects, and significant role of Big Data in the banking industry, as well as various use cases of Big Data in banking, finance services, and insurance Implementing a Data Lake (DL) in the cloud and the significance of a data lake in decision making

Handbook of E-Tourism - Zheng Xiang 2022-09

This handbook provides an authoritative and truly comprehensive overview both of the diverse applications of information and communication technologies (ICTs) within the travel and tourism industry and of e-tourism as a field of scientific inquiry that has grown and matured beyond recognition. Leading experts from around the world describe cutting-edge ideas and developments, present key concepts and theories, and discuss the full range of research methods. The coverage accordingly encompasses everything from big data and analytics to psychology, user behavior, online

marketing, supply chain and operations management, smart business networks, policy and regulatory issues – and much, much more. The goal is to provide an outstanding reference that summarizes and synthesizes current knowledge and establishes the theoretical and methodological foundations for further study of the role of ICTs in travel and tourism. The handbook will meet the needs of researchers and students in various disciplines as well as industry professionals. As with all volumes in Springer’s Major Reference Works program, readers will benefit from access to a continually updated online version.

Essentials of Business Analytics -

Bhimasankaram Pochiraju 2019-07-10

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.

Practical Hadoop Ecosystem - Deepak Vohra
2016-09-30

Learn how to use the Apache Hadoop projects, including MapReduce, HDFS, Apache Hive, Apache HBase, Apache Kafka, Apache Mahout, and Apache Solr. From setting up the environment to running sample applications each chapter in this book is a practical tutorial on using an Apache Hadoop ecosystem project. While several books on Apache Hadoop are available, most are based on the main projects, MapReduce and HDFS, and none discusses the other Apache Hadoop ecosystem projects and how they all work together as a cohesive big data development platform. What You Will Learn: Set up the environment in Linux for Hadoop projects using Cloudera Hadoop Distribution CDH 5 Run a MapReduce job Store data with Apache Hive, and Apache HBase Index data in HDFS with Apache Solr Develop a Kafka messaging system Stream Logs to HDFS with

Apache Flume Transfer data from MySQL database to Hive, HDFS, and HBase with Sqoop Create a Hive table over Apache Solr Develop a Mahout User Recommender System Who This Book Is For: Apache Hadoop developers. Pre-requisite knowledge of Linux and some knowledge of Hadoop is required.

Applications of Security, Mobile, Analytic, and Cloud (SMAC) Technologies for Effective Information Processing and Management - Karthikeyan, P. 2018-06-29

From cloud computing to big data to mobile technologies, there is a vast supply of information being mined and collected. With an abundant amount of information being accessed, stored, and saved, basic controls are needed to protect and prevent security incidents as well as ensure business continuity. Applications of Security, Mobile, Analytic, and Cloud (SMAC) Technologies for Effective Information Processing and Management is a vital resource that discusses various research findings and

innovations in the areas of big data analytics, mobile communication and mobile applications, distributed systems, and information security. With a focus on big data, the internet of things (IoT), mobile technologies, cloud computing, and information security, this book proves a vital resource for computer engineers, IT specialists, software developers, researchers, and graduate-level students seeking current research on SMAC technologies and information security management systems.

Designing Data-Intensive Applications - Martin Kleppmann 2017-03-16

Data is at the center of many challenges in system design today. Difficult issues need to be figured out, such as scalability, consistency, reliability, efficiency, and maintainability. In addition, we have an overwhelming variety of tools, including relational databases, NoSQL datastores, stream or batch processors, and message brokers. What are the right choices for your application? How do you make sense of all

these buzzwords? In this practical and comprehensive guide, author Martin Kleppmann helps you navigate this diverse landscape by examining the pros and cons of various technologies for processing and storing data. Software keeps changing, but the fundamental principles remain the same. With this book, software engineers and architects will learn how to apply those ideas in practice, and how to make full use of data in modern applications. Peer under the hood of the systems you already use, and learn how to use and operate them more effectively Make informed decisions by identifying the strengths and weaknesses of different tools Navigate the trade-offs around consistency, scalability, fault tolerance, and complexity Understand the distributed systems research upon which modern databases are built Peek behind the scenes of major online services, and learn from their architectures

Computational Methods and Data Engineering - Vijendra Singh 2020-11-04

This book gathers selected high-quality research papers from the International Conference on Computational Methods and Data Engineering (ICMDE 2020), held at SRM University, Sonipat, Delhi-NCR, India. Focusing on cutting-edge technologies and the most dynamic areas of computational intelligence and data engineering, the respective contributions address topics including collective intelligence, intelligent transportation systems, fuzzy systems, data privacy and security, data mining, data warehousing, big data analytics, cloud computing, natural language processing, swarm intelligence, and speech processing.

NEO 2015 - Oliver Schütze 2016-09-15

This volume comprises a selection of works presented at the Numerical and Evolutionary Optimization (NEO) workshop held in September 2015 in Tijuana, Mexico. The development of powerful search and optimization techniques is of great importance in today's world that requires researchers and practitioners to tackle

a growing number of challenging real-world problems. In particular, there are two well-established and widely known fields that are commonly applied in this area: (i) traditional numerical optimization techniques and (ii) comparatively recent bio-inspired heuristics. Both paradigms have their unique strengths and weaknesses, allowing them to solve some challenging problems while still failing in others. The goal of the NEO workshop series is to bring together people from these and related fields to discuss, compare and merge their complimentary perspectives in order to develop fast and reliable hybrid methods that maximize the strengths and minimize the weaknesses of the underlying paradigms. Through this effort, we believe that the NEO can promote the development of new techniques that are applicable to a broader class of problems. Moreover, NEO fosters the understanding and adequate treatment of real-world problems particularly in emerging fields that affect us all

such as health care, smart cities, big data, among many others. The extended papers the NEO 2015 that comprise this book make a contribution to this goal.

Intelligent Tools for Building a Scientific Information Platform - Robert Bembenik
2013-01-12

This book is a selection of results obtained within two years of research performed under SYNAT - a nation-wide scientific project aiming at creating an infrastructure for scientific content storage and sharing for academia, education and open knowledge society in Poland. The selection refers to the research in artificial intelligence, knowledge discovery and data mining, information retrieval and natural language processing, addressing the problems of implementing intelligent tools for building a scientific information platform. This book is a continuation and extension of the ideas presented in "Intelligent Tools for Building a Scientific Information Platform" published as

volume 390 in the same series in 2012. It is based on the SYNAT 2012 Workshop held in Warsaw. The papers included in this volume present an overview and insight into information retrieval, repository systems, text processing, ontology-based systems, text mining, multimedia data processing and advanced software engineering.

Springer Handbook of Geographic Information - Wolfgang Kresse 2022-06-24

This handbook provides an exhaustive, one-stop reference and a state-of-the-art description of geographic information and its use. This new, substantially updated edition presents a complete and rigorous overview of the fundamentals, methods and applications of the multidisciplinary field of geographic information systems. Designed to be a useful and readable desk reference book, but also prepared in various electronic formats, this title allows fast yet comprehensive review and easy retrieval of essential reliable key information. The Springer

Handbook of Geographic Information is divided into three parts. Part A, Basics and Computer Science, provides an overview on the fundamentals, including descriptions of databases and encoding of geographic information. It also covers the underlying mathematical and statistics methods and modeling. A new chapter exemplifies the emerging use and analysis of big data in a geographic context. Part B offers rigorous descriptions of gathering, processing and coding of geographic information in a standardized way to allow interoperable use in a variety of systems; from traditional methods such as geodesy and surveying to state-of-the-art remote sensing and photogrammetry; from cartography to geospatial web services. Discussions on geosemantic interoperability and security of open distributed geospatial information systems complete the comprehensive coverage. The final part describes a wide array of applications in science, industry and society at large, such as

agriculture, defense, transportation, energy and utilities, health and human services. The part is enhanced by new chapters on smart cities and building information modeling, as well as a complete overview of the currently available open-source geographic information systems. Using standardized international terminology, in accordance with ISO/TC 211 and INSPIRE, this handbook facilitates collaboration between different disciplines and is a must have for practitioners and new comers in industry and academia.

Knowledge Graphs and Big Data Processing -
Valentina Janev 2020-01-01

This open access book is part of the LAMBDA Project (Learning, Applying, Multiplying Big Data Analytics), funded by the European Union, GA No. 809965. Data Analytics involves applying algorithmic processes to derive insights.

Nowadays it is used in many industries to allow organizations and companies to make better decisions as well as to verify or disprove existing

theories or models. The term data analytics is often used interchangeably with intelligence, statistics, reasoning, data mining, knowledge discovery, and others. The goal of this book is to introduce some of the definitions, methods, tools, frameworks, and solutions for big data processing, starting from the process of information extraction and knowledge representation, via knowledge processing and analytics to visualization, sense-making, and practical applications. Each chapter in this book addresses some pertinent aspect of the data processing chain, with a specific focus on understanding Enterprise Knowledge Graphs, Semantic Big Data Architectures, and Smart Data Analytics solutions. This book is addressed to graduate students from technical disciplines, to professional audiences following continuous education short courses, and to researchers from diverse areas following self-study courses. Basic skills in computer science, mathematics, and statistics are required.

Recent Innovations in Computing - Pradeep Kumar Singh 2021-01-12

This book features selected papers presented at the 3rd International Conference on Recent Innovations in Computing (ICRIC 2020), held on 20-21 March 2020 at the Central University of Jammu, India, and organized by the university's Department of Computer Science & Information Technology. It includes the latest research in the areas of software engineering, cloud computing, computer networks and Internet technologies, artificial intelligence, information security, database and distributed computing, and digital India.

Computing and Network Sustainability - H.R . Vishwakarma 2017-07-05

The book is compilation of technical papers presented at International Research Symposium on Computing and Network Sustainability (IRSCNS 2016) held in Goa, India on 1st and 2nd July 2016. The areas covered in the book are sustainable computing and security, sustainable

systems and technologies, sustainable methodologies and applications, sustainable networks applications and solutions, user-centered services and systems and mobile data management. The novel and recent technologies presented in the book are going to be helpful for researchers and industries in their advanced works.

Knowledge Discovery in Big Data from Astronomy and Earth Observation - Petr Skoda 2020-04-23

Knowledge Discovery in Big Data from Astronomy and Earth Observation: Astrogeoinformatics bridges the gap between astronomy and geoscience in the context of applications, techniques and key principles of big data. Machine learning and parallel computing are increasingly becoming cross-disciplinary as the phenomena of Big Data is becoming common place. This book provides insight into the common workflows and data science tools used for big data in astronomy and

geoscience. After establishing similarity in data gathering, pre-processing and handling, the data science aspects are illustrated in the context of both fields. Software, hardware and algorithms of big data are addressed. Finally, the book offers insight into the emerging science which combines data and expertise from both fields in studying the effect of cosmos on the earth and

its inhabitants. Addresses both astronomy and geosciences in parallel, from a big data perspective Includes introductory information, key principles, applications and the latest techniques Well-supported by computing and information science-oriented chapters to introduce the necessary knowledge in these fields