

HBase The Definitive Guide Random Access To Your Planet Size Data

Thank you certainly much for downloading **HBase The Definitive Guide Random Access To Your Planet Size Data** .Maybe you have knowledge that, people have see numerous times for their favorite books in imitation of this HBase The Definitive Guide Random Access To Your Planet Size Data , but end occurring in harmful downloads.

Rather than enjoying a good ebook afterward a cup of coffee in the afternoon, then again they juggled following some harmful virus inside their computer. **HBase The Definitive Guide Random Access To Your Planet Size Data** is manageable in our digital library an online permission to it is set as public therefore you can download it instantly. Our digital library saves in complex countries, allowing you to get the most less latency epoch to download any of our books considering this one. Merely said, the HBase The Definitive Guide Random Access To Your Planet Size Data is universally compatible behind any devices to read.

Hadoop: The Definitive Guide - Tom White 2009-05-29

Hadoop: The Definitive Guide helps you harness the power of your data. Ideal for processing large datasets, the Apache Hadoop framework is an open source implementation of the MapReduce algorithm on which Google built its empire. This comprehensive resource demonstrates how to use Hadoop to build reliable, scalable, distributed systems: programmers will find details for analyzing large datasets, and administrators will learn how to set up and run Hadoop clusters. Complete with case studies that illustrate how Hadoop solves specific problems, this book helps you: Use the Hadoop Distributed File System (HDFS) for storing large datasets, and run distributed computations over those datasets using MapReduce Become familiar with Hadoop's data and I/O building blocks for compression, data integrity, serialization, and persistence Discover common pitfalls and advanced features for writing real-world MapReduce programs Design, build, and administer a dedicated Hadoop cluster, or run Hadoop in the cloud Use Pig, a high-level query language for large-scale data processing Take advantage of

HBase, Hadoop's database for structured and semi-structured data Learn ZooKeeper, a toolkit of coordination primitives for building distributed systems If you have lots of data -- whether it's gigabytes or petabytes -- Hadoop is the perfect solution. Hadoop: The Definitive Guide is the most thorough book available on the subject. "Now you have the opportunity to learn about Hadoop from a master-not only of the technology, but also of common sense and plain talk."-- Doug Cutting, Hadoop Founder, Yahoo!

Advanced Computer Architecture - Dezun Dong 2020-09-04

This book constitutes the refereed proceedings of the 13th Conference on Advanced Computer Architecture, ACA 2020, held in Kunming, China, in August 2020. Due to the COVID-19 pandemic the conference was held online. The 24 revised full papers presented were carefully reviewed and selected from 105 submissions. The papers of this volume are organized in topical sections on: interconnection network, router and network interface architecture; accelerator-based, application-specific and reconfigurable architecture; processor, memory, and storage systems

architecture; model, simulation and evaluation of architecture; new trends of technologies and applications.

Artificial Intelligence and Natural Algorithms - Rijwan Khan 2022-09-23
This book informs the reader about applications of Artificial Intelligence (AI) and nature-inspired algorithms in different situations. Each chapter in this book is written by topic experts on AI, nature-inspired algorithms and data science. The basic concepts relevant to these topics are explained, including evolutionary computing (EC), artificial neural networks (ANN), swarm intelligence (SI), and fuzzy systems (FS). Additionally, the book also covers optimization algorithms for data analysis. The contents include algorithms that can be used in systems designed for plant science research, load balancing, environmental analysis and healthcare. The goal of the book is to equip the reader - students and data analysts - with the information needed to apply basic AI algorithms to resolve actual problems encountered in a professional environment.

Digital Technologies - An overview of concepts, tools and techniques associated with it - Editor IJSMI 2020-09-26

Digital Technologies - An overview of concepts, tools and techniques associated with it Preface Digital technologies transformed the people's day to day life and businesses organizations the way they work. It saved their time, accessibility and cost. It provided new opportunities, new markets, and new products to the organizations. Digital revolution started with the introduction of Computers, Internet, Mobile phones, Social media networks, Cloud computing, Big Data, Internet of Things, 3D Printing, Machine learning, Virtual reality, Natural Language Processing, Block Chain, Artificial Intelligence, Robotics and Quantum Computing. Digital technologies are so dynamic and it becomes difficult to cope up with the pace with which new technologies evolve over time. Currently digital technology is being used in the time of pandemic in many countries to contain the spread through detection, contact tracing and monitoring the people movement using Big Data, Machine learning, mobile applications and Artificial intelligence tools. This book provides an overview of digital technologies and concepts, tools and techniques

associated with it and how organizations and business needs to adapt to these technologies and transform their business operations. Editor IJSMI International Journal of Statistics and Medical Informatics
www.ijsmi.com/book.php <https://www.amazon.com/dp/B08K38DBCQ>
<https://www.amazon.com/dp/B08K3YHW2M>

Knowledge Graphs - Mayank Kejriwal 2021-03-30

A rigorous and comprehensive textbook covering the major approaches to knowledge graphs, an active and interdisciplinary area within artificial intelligence. The field of knowledge graphs, which allows us to model, process, and derive insights from complex real-world data, has emerged as an active and interdisciplinary area of artificial intelligence over the last decade, drawing on such fields as natural language processing, data mining, and the semantic web. Current projects involve predicting cyberattacks, recommending products, and even gleaning insights from thousands of papers on COVID-19. This textbook offers rigorous and comprehensive coverage of the field. It focuses systematically on the major approaches, both those that have stood the test of time and the latest deep learning methods.

Big Data: Concepts, Methodologies, Tools, and Applications - Management Association, Information Resources 2016-04-20

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. Big Data: Concepts, Methodologies, Tools, and Applications is a multi-volume compendium of research-based perspectives and solutions within the realm of large-scale and complex data sets. Taking a multidisciplinary approach, this publication presents exhaustive coverage of crucial topics in the field of big data including diverse applications, storage solutions, analysis techniques, and methods for searching and transferring large data sets, in addition to security issues. Emphasizing essential research in the field of data science, this

publication is an ideal reference source for data analysts, IT professionals, researchers, and academics.

Recent Studies on Computational Intelligence - Ashish Khanna
2020-10-16

This book gathers the latest quality research work of Ph.D. students working on the current areas presented in the Doctoral Symposium on Computational Intelligence (DoSCI 2020). The book includes works in the areas of artificial intelligence, deep learning, evolutionary algorithms, swarm intelligence, fuzzy sets and vague sets, rough set theoretic approaches, quantum-inspired computational intelligence, hybrid computational intelligence, machine learning, computer vision, soft computing, distributed computing, parallel and grid computing, cloud computing, high-performance computing, biomedical computing, decision support and decision making. The book is useful for researchers, students, engineers, practitioners and academicians in their advance studies.

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.

HBase: The Definitive Guide - Lars George 2011-08-29

If you're looking for a scalable storage solution to accommodate a virtually endless amount of data, this book shows you how Apache HBase can fulfill your needs. As the open source implementation of Google's BigTable architecture, HBase scales to billions of rows and millions of columns, while ensuring that write and read performance remain constant. Many IT executives are asking pointed questions about HBase. This book provides meaningful answers, whether you're evaluating this non-relational database or planning to put it into practice right away. Discover how tight integration with Hadoop makes scalability with HBase easier Distribute large datasets across an inexpensive cluster of commodity servers Access HBase with native Java clients, or with gateway servers providing REST, Avro, or Thrift APIs Get details on HBase's architecture, including the storage format, write-ahead log, background processes, and more Integrate HBase with Hadoop's MapReduce framework for massively parallelized data processing jobs Learn how to tune clusters, design schemas, copy tables, import bulk data, decommission nodes, and many other tasks

MapReduce Design Patterns - Donald Miner 2012-11-21

Until now, design patterns for the MapReduce framework have been scattered among various research papers, blogs, and books. This handy guide brings together a unique collection of valuable MapReduce patterns that will save you time and effort regardless of the domain, language, or development framework you're using. Each pattern is explained in context, with pitfalls and caveats clearly identified to help you avoid common design mistakes when modeling your big data architecture. This book also provides a complete overview of MapReduce that explains its origins and implementations, and why design patterns are so important. All code examples are written for Hadoop.

Summarization patterns: get a top-level view by summarizing and grouping data Filtering patterns: view data subsets such as records generated from one user Data organization patterns: reorganize data to work with other systems, or to make MapReduce analysis easier Join patterns: analyze different datasets together to discover interesting

relationships Metapatterns: piece together several patterns to solve multi-stage problems, or to perform several analytics in the same job Input and output patterns: customize the way you use Hadoop to load or store data "A clear exposition of MapReduce programs for common data processing patterns—this book is indispensable for anyone using Hadoop." --Tom White, author of Hadoop: The Definitive Guide

Hadoop Operations - Eric Sammer 2012-09-26

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

Programming Hive - Edward Capriolo 2012-09-26

Describes the features and functions of Apache Hive, the data infrastructure for Hadoop.

Intelligent Systems - Chiranji Lal Chowdhary 2020-01-06

This volume helps to fill the gap between data analytics, image processing, and soft computing practices. Soft computing methods are used to focus on data analytics and image processing to develop good intelligent systems. To this end, readers of this volume will find quality research that presents the current trends, advanced methods, and hybridized techniques relating to data analytics and intelligent systems.

The book also features case studies related to medical diagnosis with the use of image processing and soft computing algorithms in particular models. Providing extensive coverage of biometric systems, soft computing, image processing, artificial intelligence, and data analytics, the chapter authors discuss the latest research issues, present solutions to research problems, and look at comparative analysis with earlier results. Topics include some of the most important challenges and discoveries in intelligent systems today, such as computer vision concepts and image identification, data analysis and computational paradigms, deep learning techniques, face and speaker recognition systems, and more.

HBase - Lars George 2011-09-05

"HBase: The Definitive Guide" provides the details for evaluating this high-performance, non-relational database, or putting it into practice right away. HBase's adoption rate is beginning to climb, and IT executives are asking pointed questions about this high-capacity database.

Cassandra: The Definitive Guide - Jeff Carpenter 2016-06-29

Imagine what you could do if scalability wasn't a problem. With this hands-on guide, you'll learn how the Cassandra database management system handles hundreds of terabytes of data while remaining highly available across multiple data centers. This expanded second edition—updated for Cassandra 3.0—provides the technical details and practical examples you need to put this database to work in a production environment. Authors Jeff Carpenter and Eben Hewitt demonstrate the advantages of Cassandra's non-relational design, with special attention to data modeling. If you're a developer, DBA, or application architect looking to solve a database scaling issue or future-proof your application, this guide helps you harness Cassandra's speed and flexibility. Understand Cassandra's distributed and decentralized structure Use the Cassandra Query Language (CQL) and cqlsh—the CQL shell Create a working data model and compare it with an equivalent relational model Develop sample applications using client drivers for languages including Java, Python, and Node.js Explore cluster topology and learn how nodes

exchange data Maintain a high level of performance in your cluster
Deploy Cassandra on site, in the Cloud, or with Docker Integrate
Cassandra with Spark, Hadoop, Elasticsearch, Solr, and Lucene

HBase in Action - Amandeep Khurana 2012-11-01

Summary HBase in Action has all the knowledge you need to design, build, and run applications using HBase. First, it introduces you to the fundamentals of distributed systems and large scale data handling. Then, you'll explore real-world applications and code samples with just enough theory to understand the practical techniques. You'll see how to build applications with HBase and take advantage of the MapReduce processing framework. And along the way you'll learn patterns and best practices. About the Technology HBase is a NoSQL storage system designed for fast, random access to large volumes of data. It runs on commodity hardware and scales smoothly from modest datasets to billions of rows and millions of columns. About this Book HBase in Action is an experience-driven guide that shows you how to design, build, and run applications using HBase. First, it introduces you to the fundamentals of handling big data. Then, you'll explore HBase with the help of real applications and code samples and with just enough theory to back up the practical techniques. You'll take advantage of the MapReduce processing framework and benefit from seeing HBase best practices in action. 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 When and how to use HBase Practical examples Design patterns for scalable data systems Deployment, integration, and design Written for developers and architects familiar with data storage and processing. No prior knowledge of HBase, Hadoop, or MapReduce is required. Table of Contents PART 1 HBASE FUNDAMENTALS Introducing HBase Getting started Distributed HBase, HDFS, and MapReduce PART 2 ADVANCED CONCEPTS HBase table design Extending HBase with coprocessors Alternative HBase clients PART 3 EXAMPLE APPLICATIONS HBase by example: OpenTSDB Scaling GIS on HBase PART 4 OPERATIONALIZING HBASE Deploying HBase Operations

Hadoop: The Definitive Guide - Tom White 2012-05-10

Ready to unlock the power of your data? With this comprehensive guide, you'll learn how to build and maintain reliable, scalable, distributed systems with Apache Hadoop. This book is ideal for programmers looking to analyze datasets of any size, and for administrators who want to set up and run Hadoop clusters. You'll find illuminating case studies that demonstrate how Hadoop is used to solve specific problems. This third edition covers recent changes to Hadoop, including material on the new MapReduce API, as well as MapReduce 2 and its more flexible execution model (YARN). Store large datasets with the Hadoop Distributed File System (HDFS) Run distributed computations with MapReduce Use Hadoop's data and I/O building blocks for compression, data integrity, serialization (including Avro), and persistence Discover common pitfalls and advanced features for writing real-world MapReduce programs Design, build, and administer a dedicated Hadoop cluster—or run Hadoop in the cloud Load data from relational databases into HDFS, using Sqoop Perform large-scale data processing with the Pig query language Analyze datasets with Hive, Hadoop's data warehousing system Take advantage of HBase for structured and semi-structured data, and ZooKeeper for building distributed systems

Cloud Computing for Science and Engineering - Ian Foster 2017-09-29

A guide to cloud computing for students, scientists, and engineers, with advice and many hands-on examples. The emergence of powerful, always-on cloud utilities has transformed how consumers interact with information technology, enabling video streaming, intelligent personal assistants, and the sharing of content. Businesses, too, have benefited from the cloud, outsourcing much of their information technology to cloud services. Science, however, has not fully exploited the advantages of the cloud. Could scientific discovery be accelerated if mundane chores were automated and outsourced to the cloud? Leading computer scientists Ian Foster and Dennis Gannon argue that it can, and in this book offer a guide to cloud computing for students, scientists, and engineers, with advice and many hands-on examples. The book surveys the technology that underpins the cloud, new approaches to technical

problems enabled by the cloud, and the concepts required to integrate cloud services into scientific work. It covers managing data in the cloud, and how to program these services; computing in the cloud, from deploying single virtual machines or containers to supporting basic interactive science experiments to gathering clusters of machines to do data analytics; using the cloud as a platform for automating analysis procedures, machine learning, and analyzing streaming data; building your own cloud with open source software; and cloud security. The book is accompanied by a website, Cloud4SciEng.org, that provides a variety of supplementary material, including exercises, lecture slides, and other resources helpful to readers and instructors.

Algorithms and Architectures for Parallel Processing - Jaideep Vaidya 2018-12-06

The four-volume set LNCS 11334-11337 constitutes the proceedings of the 18th International Conference on Algorithms and Architectures for Parallel Processing, ICA3PP 2018, held in Guangzhou, China, in November 2018. The 141 full and 50 short papers presented were carefully reviewed and selected from numerous submissions. The papers are organized in topical sections on Distributed and Parallel Computing; High Performance Computing; Big Data and Information Processing; Internet of Things and Cloud Computing; and Security and Privacy in Computing.

Web Engineering - Maria Bielikova 2020-06-09

This book constitutes the proceedings of the 20th International Conference on Web Engineering, ICWE 2020, which was planned to take place in Helsinki, Finland, during June 9-12, 2020. Due to the corona pandemic the conference changed to a virtual format. The total of 24 full and 10 short contributions presented in this volume were carefully reviewed and selected from 78 submissions. The book also contains 4 PhD and 7 demo papers. The papers were organized in topical sections named: User interface technologies; performance of Web technologies; machine learning; testing of Web applications; emotion detection; location-aware applications; sentiment analysis; open data; liquid Web applications; Web-based learning; PhD symposium; demos and posters.

Next-Generation Big Data - Butch Quinto 2018-06-12

Utilize this practical and easy-to-follow guide to modernize traditional enterprise data warehouse and business intelligence environments with next-generation big data technologies. Next-Generation Big Data takes a holistic approach, covering the most important aspects of modern enterprise big data. The book covers not only the main technology stack but also the next-generation tools and applications used for big data warehousing, data warehouse optimization, real-time and batch data ingestion and processing, real-time data visualization, big data governance, data wrangling, big data cloud deployments, and distributed in-memory big data computing. Finally, the book has an extensive and detailed coverage of big data case studies from Navistar, Cerner, British Telecom, Shopzilla, Thomson Reuters, and Mastercard. What You'll Learn Install Apache Kudu, Impala, and Spark to modernize enterprise data warehouse and business intelligence environments, complete with real-world, easy-to-follow examples, and practical advice Integrate HBase, Solr, Oracle, SQL Server, MySQL, Flume, Kafka, HDFS, and Amazon S3 with Apache Kudu, Impala, and Spark Use StreamSets, Talend, Pentaho, and CDAP for real-time and batch data ingestion and processing Utilize Trifacta, Alteryx, and Datameer for data wrangling and interactive data processing Turbocharge Spark with Alluxio, a distributed in-memory storage platform Deploy big data in the cloud using Cloudera Director Perform real-time data visualization and time series analysis using Zoomdata, Apache Kudu, Impala, and Spark Understand enterprise big data topics such as big data governance, metadata management, data lineage, impact analysis, and policy enforcement, and how to use Cloudera Navigator to perform common data governance tasks Implement big data use cases such as big data warehousing, data warehouse optimization, Internet of Things, real-time data ingestion and analytics, complex event processing, and scalable predictive modeling Study real-world big data case studies from innovative companies, including Navistar, Cerner, British Telecom, Shopzilla, Thomson Reuters, and Mastercard Who This Book Is For BI and big data warehouse professionals interested in gaining practical and

real-world insight into next-generation big data processing and analytics using Apache Kudu, Impala, and Spark; and those who want to learn more about other advanced enterprise topics

Advanced Intelligent Systems for Sustainable Development

(AI2SD'2018) - Mostafa Ezziyani 2019-03-06

This book includes the outcomes of the International Conference on Advanced Intelligent Systems for Sustainable Development (AI2SD-2018), held in Tangier, Morocco on July 12-14, 2018. Presenting the latest research in the field of computing sciences and information technology, it discusses new challenges and provides valuable insights into the field, the goal being to stimulate debate, and to promote closer interaction and interdisciplinary collaboration between researchers and practitioners. Though chiefly intended for researchers and practitioners in advanced information technology management and networking, the book will also be of interest to those engaged in emerging fields such as data science and analytics, big data, internet of things, smart networked systems, artificial intelligence, expert systems and cloud computing.

Hadoop in Practice - Alex Holmes 2014-09-29

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

Trino: The Definitive Guide - Matt Fuller 2021-04-14

Perform fast interactive analytics against different data sources using the Trino high-performance distributed SQL query engine. With this practical guide, you'll learn how to conduct analytics on data where it lives, whether it's Hive, Cassandra, a relational database, or a proprietary data store. Analysts, software engineers, and production engineers will learn how to manage, use, and even develop with Trino. Initially developed by Facebook, open source Trino is now used by Netflix, Airbnb, LinkedIn, Twitter, Uber, and many other companies. Matt Fuller, Manfred Moser, and Martin Traverso show you how a single Trino query can combine data from multiple sources to allow for analytics across your entire organization. Get started: Explore Trino's use cases and learn about tools that will help you connect to Trino and query data Go deeper: Learn Trino's internal workings, including how to connect to and query data sources with support for SQL statements, operators, functions, and more

Put Trino in production: Secure Trino, monitor workloads, tune queries, and connect more applications; learn how other organizations apply Trino

Big Data Systems - Jawwad Ahmad Shamsi 2021-05-11

Big Data Systems encompass massive challenges related to data diversity, storage mechanisms, and requirements of massive computational power. Further, capabilities of big data systems also vary with respect to type of problems. For instance, distributed memory systems are not recommended for iterative algorithms. Similarly, variations in big data systems also exist related to consistency and fault tolerance. The purpose of this book is to provide a detailed explanation of big data systems. The book covers various topics including Networking, Security, Privacy, Storage, Computation, Cloud Computing, NoSQL and NewSQL systems, High Performance Computing, and Deep Learning. An illustrative and practical approach has been adopted in which theoretical topics have been aided by well-explained programming and illustrative examples. Key Features: Introduces concepts and evolution of Big Data technology. Illustrates examples for thorough understanding. Contains programming examples for hands on development. Explains a variety of topics including NoSQL Systems, NewSQL systems, Security, Privacy, Networking, Cloud, High Performance Computing, and Deep Learning. Exemplifies widely used big data technologies such as Hadoop and Spark. Includes discussion on case studies and open issues. Provides end of chapter questions for enhanced learning.

Handbook of Research on Big Data Storage and Visualization Techniques - Segall, Richard S. 2018-01-05

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.

Spark: The Definitive Guide - Bill Chambers 2018-02-08

Learn how to use, deploy, and maintain Apache Spark with this comprehensive guide, written by the creators of the open-source cluster-computing framework. With an emphasis on improvements and new features in Spark 2.0, authors Bill Chambers and Matei Zaharia break down Spark topics into distinct sections, each with unique goals. You'll explore the basic operations and common functions of Spark's structured APIs, as well as Structured Streaming, a new high-level API for building end-to-end streaming applications. Developers and system administrators will learn the fundamentals of monitoring, tuning, and debugging Spark, and explore machine learning techniques and scenarios for employing MLlib, Spark's scalable machine-learning library. Get a gentle overview of big data and Spark Learn about DataFrames, SQL, and Datasets—Spark's core APIs—through worked examples Dive into Spark's low-level APIs, RDDs, and execution of SQL and DataFrames Understand how Spark runs on a cluster Debug, monitor, and tune Spark clusters and applications Learn the power of Structured Streaming, Spark's stream-processing engine Learn how you can apply MLlib to a variety of problems, including classification or recommendation Scaling Big Data with Hadoop and Solr - Second Edition - Hrishikesh Vijay Karambelkar 2015-04-27

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

Smart Technologies, Systems and Applications - Fabián R. Narváez 2020-04-30

This book constitutes refereed proceedings of the First International Conference on Smart Technologies, Systems and Applications, held in Quito, Ecuador, in December 2019. The 27 full papers and 3 short papers presented were carefully reviewed and selected from 90 submissions. The papers of this volume are organized in topical sections on smart technologies; smart systems; smart trends and applications.

Hadoop Application Architectures - Mark Grover 2015-06-30

Get expert guidance on architecting end-to-end data management solutions with Apache Hadoop. While many sources explain how to use various components in the Hadoop ecosystem, this practical book takes you through architectural considerations necessary to tie those components together into a complete tailored application, based on your particular use case. To reinforce those lessons, the book's second section provides detailed examples of architectures used in some of the most commonly found Hadoop applications. Whether you're designing a new Hadoop application, or planning to integrate Hadoop into your existing data infrastructure, Hadoop Application Architectures will skillfully guide you through the process. This book covers: Factors to consider when using Hadoop to store and model data Best practices for moving data in and out of the system Data processing frameworks, including MapReduce, Spark, and Hive Common Hadoop processing patterns, such as removing duplicate records and using windowing analytics Graph, GraphX, and other tools for large graph processing on Hadoop Using workflow orchestration and scheduling tools such as Apache Oozie Near-real-time stream processing with Apache Storm, Apache Spark Streaming, and Apache Flume Architecture examples for clickstream analysis, fraud detection, and data warehousing

Managing Big Data Integration in the Public Sector - Aggarwal, Anil 2015-11-12

The era of rapidly progressing technology we live in generates vast amounts of data; however, the challenge exists in understanding how to aggressively monitor and make sense of this data. Without a better understanding of how to collect and manage such large data sets, it becomes increasingly difficult to successfully utilize them. Managing Big

Data Integration in the Public Sector is a pivotal reference source for the latest scholarly research on the application of big data analytics in government contexts and identifies various strategies in which big data platforms can generate improvements within that sector. Highlighting issues surrounding data management, current models, and real-world applications, this book is ideally designed for professionals, government agencies, researchers, and non-profit organizations interested in the benefits of big data analytics applied in the public sphere.

Learning HBase - Shashwat Shripav 2014-11-25

If you are an administrator or developer who wants to enter the world of Big Data and BigTables and would like to learn about HBase, this is the book for you.

Digital Libraries and Archives - Costantino Grana 2017-10-10

This book constitutes the thoroughly refereed proceedings of the 13th Italian Research Conference on Digital Libraries, IRCDL 2017, held in Modena, Italy, in January 2017. The 15 papers presented were carefully selected from 25 submissions. The papers cover the following topics: bibliometrics and education; multimedia; data management and presentation; cultural heritage; applications.

Distributed Applications and Interoperable Systems - Lydia Y. Chen 2017-06-06

This book constitutes the proceedings of the 17th IFIP International Conference on Distributed Applications and Interoperable Systems, DAIS 2017, held in Neuchâtel, Switzerland, in June 2017. The 11 papers presented together with 4 short papers in this volume were carefully reviewed and selected from 23 submissions. The papers are organized in topical sections on running system efficiently, storing data smartly, roaming in graph, building collaborative services, and making things safe.

Database Systems for Advanced Applications. DASFAA 2022

International Workshops - Uday Kiran Rage 2022-08-16

This volume constitutes the papers of several workshops which were held in conjunction with the 27th International Conference on Database Systems for Advanced Applications, DASFAA 2022, held as virtual event

in April 2022. The 30 revised full papers presented in this book were carefully reviewed and selected from 65 submissions. DASFAA 2022 presents the following five workshops: · First workshop on Pattern mining and Machine learning in Big complex Databases (PMBD 2021) · 6th International Workshop on Graph Data Management and Analysis (GDMA 2022) · First International Workshop on Blockchain Technologies (IWBT2022) · 8th International Workshop on Big Data Management and Service (BDMS 2022) · First workshop on Managing Air Quality Through Data Science · 7th International Workshop on Big Data Quality Management (BDQM 2022).

Mathematics of Big Data - Jeremy Kepner 2018-07-17

The first book to present the common mathematical foundations of big data analysis across a range of applications and technologies. Today, the volume, velocity, and variety of data are increasing rapidly across a range of fields, including Internet search, healthcare, finance, social media, wireless devices, and cybersecurity. Indeed, these data are growing at a rate beyond our capacity to analyze them. The tools—including spreadsheets, databases, matrices, and graphs—developed to address this challenge all reflect the need to store and operate on data as whole sets rather than as individual elements. This book presents the common mathematical foundations of these data sets that apply across many applications and technologies. Associative arrays unify and simplify data, allowing readers to look past the differences among the various tools and leverage their mathematical similarities in order to solve the hardest big data challenges. The book first introduces the concept of the associative array in practical terms, presents the associative array manipulation system D4M (Dynamic Distributed Dimensional Data Model), and describes the application of associative arrays to graph analysis and machine learning. It provides a mathematically rigorous definition of associative arrays and describes the properties of associative arrays that arise from this definition. Finally, the book shows how concepts of linearity can be extended to encompass associative arrays. Mathematics of Big Data can be used as a textbook or reference by engineers, scientists, mathematicians, computer

scientists, and software engineers who analyze big data.

Architecting Modern Data Platforms - Jan Kunigk 2018-12-05

There's a lot of information about big data technologies, but splicing these technologies into an end-to-end enterprise data platform is a daunting task not widely covered. With this practical book, you'll learn how to build big data infrastructure both on-premises and in the cloud and successfully architect a modern data platform. Ideal for enterprise architects, IT managers, application architects, and data engineers, this book shows you how to overcome the many challenges that emerge during Hadoop projects. You'll explore the vast landscape of tools available in the Hadoop and big data realm in a thorough technical primer before diving into: Infrastructure: Look at all component layers in a modern data platform, from the server to the data center, to establish a solid foundation for data in your enterprise Platform: Understand aspects of deployment, operation, security, high availability, and disaster recovery, along with everything you need to know to integrate your platform with the rest of your enterprise IT Taking Hadoop to the cloud: Learn the important architectural aspects of running a big data platform in the cloud while maintaining enterprise security and high availability

Cloud Computing in Remote Sensing - Lizhe Wang 2019-07-11

This book provides the users with quick and easy data acquisition, processing, storage and product generation services. It describes the entire life cycle of remote sensing data and builds an entire high performance remote sensing data processing system framework. It also develops a series of remote sensing data management and processing standards. Features: Covers remote sensing cloud computing Covers remote sensing data integration across distributed data centers Covers cloud storage based remote sensing data share service Covers high performance remote sensing data processing Covers distributed remote sensing products analysis

Big Scientific Data Management - Jianhui Li 2019-08-06

This book constitutes the refereed proceedings of the First International Conference on Big Scientific Data Management, BigSDM 2018, held in Beijing, Greece, in November/December 2018. The 24 full papers

presented together with 7 short papers were carefully reviewed and selected from 86 submissions. The topics involved application cases in the big scientific data management, paradigms for enhancing scientific discovery through big data, data management challenges posed by big scientific data, machine learning methods to facilitate scientific discovery, science platforms and storage systems for large scale scientific applications, data cleansing and quality assurance of science

data, and data policies.

Programming Pig - Alan Gates 2011-10-06

This guide is an ideal learning tool and reference for Apache Pig, the programming language that helps programmers describe and run large data projects on Hadoop. With Pig, they can analyze data without having to create a full-fledged application--making it easy for them to experiment with new data sets.