
Online Library HBase The Definitive Guide Random Access To Your Planet Size Data

This is likewise one of the factors by obtaining the soft documents of this **HBase The Definitive Guide Random Access To Your Planet Size Data** by online. You might not require more mature to spend to go to the books instigation as without difficulty as search for them. In some cases, you likewise complete not discover the statement HBase The Definitive Guide Random Access To Your Planet Size Data that you are looking for. It will no question squander the time.

However below, afterward you visit this web page, it will be fittingly agreed easy to get as competently as download guide HBase The Definitive Guide Random Access To Your Planet Size Data

It will not agree to many period as we notify before. You can realize it though behave something else at home and even in your workplace. correspondingly easy! So, are you question? Just exercise just what we pay for below as capably as evaluation **HBase The Definitive Guide Random Access To Your Planet Size Data** what you similar to to read!

QY4324 - DEVIN LAILA

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.

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.

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

This book constitutes the refereed proceedings of the First International Symposium on Benchmarking, Measuring, and Optimization, Bench 2018, held in Seattle, WA, USA, in December 2018. The 20 full papers presented were carefully reviewed and selected from 51 submissions. The papers are organized in topical sections named: AI Benchmarking; Cloud; Big Data; Modelling and Prediction; and Algorithm and Implementations.

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.

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.

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

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

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: 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!

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

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.

This book constitutes the proceedings of the 14th IFIP WG 10.3 International Conference on Network and Parallel Computing, NPC 2017, held in Hefei, China, in October 2017. The 9 full papers and 10 short papers presented in this book were carefully reviewed and selected from 88 submissions. The papers cover traditional areas of network and parallel computing including parallel applications, distributed algorithms, software environments, and distributed tools.

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 Giraph, 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

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).

With the proliferation of Software-as-a-Service (SaaS) offerings, it is becoming increasingly important for individual SaaS providers to operate their services at a low cost. This book investigates SaaS from the perspective of the provider and shows how operational costs can be reduced by using “multi tenancy,” a technique for consolidating a large number of customers onto a small number of servers. Specifically, the book addresses multi tenancy on the database level, focusing on in-memory column databases, which are the backbone of many important new enterprise applications. For efficiently implementing multi tenancy in a farm of databases, two fundamental challenges must be addressed, (i) workload modeling and (ii) data placement. The first involves estimating the (shared) resource consumption for multi tenancy on a single in-memory database server. The second consists in assigning tenants to servers in a way that minimizes the number of required servers (and thus costs) based on the assumed workload model. This step also entails replicating tenants for performance and high availability. This book presents novel solutions to both problems.

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.

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

If you’re looking for a scalable storage solution to accommodate a virtually endless amount of data, this updated edition shows you how Apache HBase can meet your needs. Modeled after Google’s BigTable architecture, HBase scales to billions of rows and millions of columns, while ensuring that write and read performance remain constant. Fully revised for HBase 1.0, this second edition brings you up to speed on the new HBase client API, as well as security features and new case studies that demonstrate HBase use in the real world. Whether you just started to evaluate this non-relational database, or plan to put it into practice right away, this book has your back. Launch into basic, advanced, and administrative features of HBase’s new client-facing API Use new classes to integrate HBase with Hadoop’s MapReduce framework Explore HBase’s architecture, including the storage format, write-ahead log, and background processes Dive into advanced usage, such extended client and server options Learn cluster sizing, tuning, and monitoring best practices Design schemas, copy tables, import bulk data, decommission nodes, and other tasks Go deeper into HBase security, including Kerberos and encryption at rest

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.

Computational Intelligence for Medical Internet of Things (MIoT) Applications: Machine Intelligence Applications for IoT in Healthcare explores machine intelligence techniques necessary for effective MIoT research and practice, taking a practical approach for practitioners and students entering the field. This book investigates advanced concepts and applications in the MIoT field, guiding readers through emerging developments and future trends. A wide range of international authors guide readers through advanced concepts, including deep learning, neural network, and big data analytic approaches for the classification, indexing, retrieval, analysis, and inferencing of healthcare data. Presents the state-of-the-art in machine intelligence and related technologies and methodologies for IoT in healthcare Discusses emerging developments and trends in machine intelligence for business and decision-making strategy in healthcare Features new models, practical solutions, prototypes, frameworks and technological advances related to machine intelligence for MIoT applications

Let Hadoop For Dummies help harness the power of your data and rein in the information overload Big data has become big business, and companies and organizations of all sizes are struggling to find ways to retrieve valuable information from their massive data sets with becoming overwhelmed. Enter Hadoop and this easy-to-understand For Dummies guide. Hadoop For Dummies helps readers understand the value of big data, make a business case for using Hadoop, navigate the Hadoop ecosystem, and build and manage Hadoop applications and clusters. Explains the origins of Hadoop, its economic benefits, and its functionality and practical applications Helps you find your way around the Hadoop ecosystem, program MapReduce, utilize design patterns, and get your Hadoop cluster up and running quickly and easily Details how to use Hadoop applications for data mining, web analytics and personalization, large-scale text processing, data science, and problem-solving Shows you how to improve the value of your Hadoop cluster, maximize your investment in Hadoop, and avoid common pitfalls when building your Hadoop cluster From programmers challenged with building and

maintaining affordable, scalable data systems to administrators who must deal with huge volumes of information effectively and efficiently, this how-to has something to help you with Hadoop.

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

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: 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.

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.

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

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.

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.

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.

“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"--Provided by publisher.

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

Perform fast interactive analytics against different data sources using the Presto 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 Presto. Initially developed by Facebook, open source Presto 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 Presto query can combine data from multiple sources to allow for analytics across your entire organization. Get started: Explore Presto's use cases and learn about tools that will help you connect to Presto and query data Go deeper: Learn Presto's internal workings, including how to connect to and query data sources with support for SQL statements, operators, functions, and more Put Presto in production: Secure Presto, monitor workloads, tune queries, and connect more applications; learn how other organizations apply Presto

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

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.

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, how-

ever, has not fully exploited the advantages of the cloud. Could scientific discovery be accelerated if mundane chores were automated and out-sourced 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.

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.

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

Principles of Big Graph: In-depth Insight, Volume 128 in the Advances in Computer series, highlights new advances in the field with this new volume presenting interesting chapters on a variety of topics, including CESDAM: Centered subgraph data matrix for large graph representation, Bivariate, cluster and suitability analysis of NoSQL Solutions for big graph applications, An empirical investigation on Big Graph using deep learning, Analyzing correlation between quality and accuracy of graph clustering, geneBF: Filtering protein-coded gene graph data using bloom filter, Processing large graphs with an alternative representation, MapReduce based convolutional graph neural networks: A comprehensive review. Fast exact triangle counting in large graphs using SIMD acceleration, A comprehensive investigation on attack graphs, Qubit representation of a binary tree and its operations in quantum computation, Modified ML-KNN: Role of similarity measures and nearest neighbor configuration in multi label text classification on big social network graph data, Big graph based online learning through social networks, Community detection in large-scale real-world networks, Power rank: An interactive web page ranking algorithm, GA based energy efficient modelling of a wireless sensor network, The major challenges of big graph and their solutions: A review, and An investigation on socio-cyber crime graph. Provides an update on the issues and challenges faced by current researchers Updates on future research agendas Includes advanced topics for intensive research for researchers

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.