

## Hadoop Security Protecting Your Big Data Platform

Summary Making Sense of NoSQL clearly and concisely explains the concepts, features, benefits, potential, and limitations of NoSQL technologies. Using examples and use cases, illustrations, and plain, jargon-free writing, this guide shows how you can effectively assemble a NoSQL solution to replace or augment the traditional RDBMS you have now. About this Book If you want to understand and perhaps start using the new data storage and analysis technologies that go beyond the SQL database model, this book is for you. Written in plain language suitable for technical managers and developers, and using many examples, use cases, and illustrations, this book explains the concepts, features, benefits, potential, and limitations of NoSQL. Making Sense of NoSQL starts by comparing familiar database concepts to the new NoSQL patterns that augment or replace them. Then, you'll explore case studies on big data, search, reliability, and business agility that apply these new patterns to today's business problems. You'll see how NoSQL systems can leverage the resources of modern cloud computing and multiple-CPU data centers. The final chapters show you how to choose the right NoSQL technologies for your own needs. Managers and developers will welcome this lucid overview of the potential and capabilities of NoSQL technologies. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. What's Inside NoSQL data architecture patterns NoSQL for big data Search, high availability, and security Choosing an architecture About the Authors Dan McCreary and Ann Kelly lead an independent training and consultancy firm focused on NoSQL solutions and are cofounders of the NoSQL Now! Conference. Table of Contents PART 1 INTRODUCTION NoSQL: It's about making intelligent choices NoSQL concepts PART 2 DATABASE PATTERNS Foundational data architecture patterns NoSQL data architecture patterns Native XML databases PART 3 NOSQL SOLUTIONS Using NoSQL to manage big data Finding information with NoSQL search Building high-availability solutions with NoSQL Increasing agility with NoSQL PART 4 ADVANCED TOPICS NoSQL and functional programming Security: protecting data in your NoSQL systems Selecting the right NoSQL solution

Utilize R to uncover hidden patterns in your Big Data About This Book Perform computational analyses on Big Data to generate meaningful results Get a practical knowledge of R programming language while working on Big Data platforms like Hadoop, Spark, H2O and SQL/NoSQL databases, Explore fast, streaming, and scalable data analysis with the most cutting-edge technologies in the market Who This Book Is For This book is intended for Data Analysts, Scientists, Data Engineers, Statisticians, Researchers, who want to integrate R with their current or future Big Data workflows. It is assumed that readers have some experience in data analysis and understanding of data management and algorithmic processing of large quantities of data, however they may lack specific skills related to R. What You Will Learn Learn about current state of Big Data processing using R programming language and its powerful statistical capabilities Deploy Big Data analytics platforms with selected Big Data tools supported by R in a cost-effective and time-saving manner Apply the R language to real-world Big Data problems on a multi-node Hadoop cluster, e.g. electricity consumption across various socio-demographic indicators and bike share scheme usage Explore the compatibility of R with Hadoop, Spark, SQL and NoSQL databases, and H2O platform In Detail Big Data analytics is the process of examining large and complex data sets that often exceed the computational capabilities. R is a leading programming language of data science, consisting of powerful functions to tackle all problems related to Big Data processing. The book will begin with a brief introduction to the Big Data world and its current industry standards. With introduction to the R language and presenting its development, structure, applications in real world, and its shortcomings. Book will progress towards revision of major R functions for data management and transformations. Readers will be introduced to Cloud based Big Data solutions (e.g. Amazon EC2 instances and Amazon RDS, Microsoft Azure and its HDInsight clusters) and also provide guidance on R connectivity with relational and non-relational databases such as MongoDB and HBase etc. It will further expand to include Big Data tools such as Apache Hadoop ecosystem, HDFS and MapReduce frameworks. Also other R compatible tools such as Apache Spark, its machine learning library Spark MLlib, as well as H2O. Style and approach This book will serve as a practical guide to tackling Big Data problems using R programming language and its statistical environment. Each section of the book will present you with concise and easy-to-follow steps on how to process, transform and analyse large data sets.

The massive datasets required for most modern businesses are too large to safely store and efficiently process on a single server. Hadoop is an open source data processing framework that provides a distributed file system that can manage data stored across clusters of servers and implements the MapReduce data processing model so that users can effectively query and utilize big data. The new Hadoop 2.0 is a stable, enterprise-ready platform supported by a rich ecosystem of tools and related technologies such as Pig, Hive, YARN, Spark, Tez, and many more. Hadoop in Action, Second Edition, provides a comprehensive introduction to Hadoop and shows how to write programs in the MapReduce style. It starts with a few easy examples and then moves quickly to show how Hadoop can be used in more complex data analysis tasks. It covers how YARN, new in Hadoop 2, simplifies and supercharges resource management to make streaming and real-time applications more feasible. Included are best practices and design patterns of MapReduce programming. The book expands on the first edition by enhancing coverage of important Hadoop 2 concepts and systems, and by providing new chapters on data management and data science that reinforce a practical understanding of Hadoop. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. Big Data in a nutshell: It is the ability to retain, process, and understand data like never before. It can mean more data than what you are using today; but it can also mean different kinds of data, a venture into the unstructured world where most of today's data resides. In this book you will learn how cognitive computing systems, like IBM Watson, fit into the Big Data world. Learn about the concept of data-in-motion and InfoSphere Streams, the world's fastest and most flexible platform for streaming data. Capturing, storing, refining, transforming, governing, securing, and analyzing data are important topics also covered in this book.

As more corporations turn to Hadoop to store and process their most valuable data, the risks with a potential breach of those systems increases exponentially. This practical book not only shows security architects how to protect Hadoop data from unauthorized access, but also how to limit the ability of an attacker to corrupt or modify data in the event of a security breach. Security is typically the longest phase of any Hadoop project. In this book, Cloudera professionals Ben Spivey and Joey Echeverria provide a thorough treatment of Hadoop security architecture and implementation to help architects navigate the common challenges of this phase. Example case studies demonstrate how the book's material is immediately applicable to your use cases. Understand the challenges of securing distributed systems, particularly Hadoop Use best practices for preparing Hadoop cluster hardware from a security standpoint Get an overview of the Kerberos network authentication protocol Delve into authorization and accounting principles as they apply to Hadoop Use mechanisms to protect data in a Hadoop cluster, both in transit and at rest Integrate Hadoop data ingest into enterprise-wide security architecture Ensure that security architecture reaches all the way to end-user access

This three-book set constitutes the refereed proceedings of the Second International Conference on Recent Trends in Image Processing and Pattern Recognition (RTIP2R) 2018, held in Solapur, India, in December 2018. The 173 revised full papers presented were carefully reviewed and selected from 374 submissions. The papers are organized in topical sections in the three volumes. Part I: computer vision and pattern recognition; machine learning and applications; and image processing. Part II: healthcare and medical imaging; biometrics and applications. Part III: document image analysis; image analysis in agriculture; and data mining, information retrieval and 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.

The go-to guidebook for deploying Big Data solutions with Hadoop Today's enterprise architects need to understand how the Hadoop frameworks and APIs fit together, and how they can be integrated to deliver real-world solutions. This book is a practical, detailed guide to building and implementing those solutions, with code-level instruction in the popular Wrox tradition. It covers storing data with HDFS and Hbase, processing data with MapReduce, and automating data processing with Oozie. Hadoop security, running Hadoop with Amazon Web Services, best practices, and automating Hadoop processes in real time are also covered in depth. With in-depth code examples in Java and XML and the latest on recent additions to the Hadoop ecosystem, this complete resource also covers the use of APIs, exposing their inner workings and allowing architects and developers to better leverage and customize them. The ultimate guide for developers, designers, and architects who need to build and deploy Hadoop applications Covers storing and processing data with various technologies, automating data processing, Hadoop security, and delivering real-time solutions Includes detailed, real-world examples and code-level guidelines Explains when, why, and how to use these tools effectively Written by a team of Hadoop experts in the programmer-to-programmer Wrox style Professional Hadoop Solutions is the reference enterprise architects and developers need to maximize the power of Hadoop.

Find the right big data solution for your business or organization Big data management is one of the major challenges facing business, industry, and not-for-profit organizations. Data sets such as customer transactions for a mega-retailer, weather patterns monitored by meteorologists, or social network activity can quickly outpace the capacity of traditional data management tools. If you need to develop or manage big data solutions, you'll appreciate how these four experts define, explain, and guide you through this new and often confusing concept. You'll learn what it is, why it matters, and how to choose and implement solutions that work. Effectively managing big data is an issue of growing importance to businesses, not-for-profit organizations, government, and IT professionals Authors are experts in information management, big data, and a variety of solutions Explains big data in detail and discusses how to select and implement a solution, security concerns to consider, data storage and presentation issues, analytics, and much more Provides essential information in a no-nonsense, easy-to-understand style that is empowering Big Data For Dummies cuts through the confusion and helps you take charge of big data solutions for your organization.

The data lake is a daring new approach for harnessing the power of big data technology and providing convenient self-service capabilities. But is it right for your company? This book is based on discussions with practitioners and executives from more than a hundred organizations, ranging from data-driven companies such as Google, LinkedIn, and Facebook, to governments and traditional corporate enterprises. You'll learn what a data lake is, why enterprises need one, and how to build one successfully with the best practices in this book. Alex Gorelik, CTO and founder of Waterline Data, explains why old systems and processes can no longer support data needs in the enterprise. Then, in a collection of essays about data lake implementation, you'll examine data lake initiatives, analytic projects, experiences, and best practices from data experts working in various industries. Get a succinct introduction to data warehousing, big data, and data science Learn various paths enterprises take to build a data lake Explore how to build a self-service model and best practices for providing analysts access to the data Use different methods for architecting your data lake Discover ways to implement a data lake from experts in different industries

This is the eBook of the printed book and may not include any media, website access codes, or print supplements that may come packaged with the bound book. The Comprehensive, Up-to-Date Apache Hadoop Administration Handbook and Reference "Sam Alapati has worked with production Hadoop clusters for six years. His unique depth of experience has enabled him to write the go-to resource for all administrators looking to spec, size, expand, and secure production Hadoop clusters of any size." —Paul Dix, Series Editor In Expert Hadoop® Administration, leading Hadoop administrator Sam R. Alapati brings together authoritative knowledge for creating, configuring, securing, managing, and optimizing production Hadoop clusters in any environment. Drawing on his experience with large-scale Hadoop administration, Alapati integrates action-oriented advice with carefully researched explanations of both problems and solutions. He covers an unmatched range of topics and offers an unparalleled collection of realistic examples. Alapati demystifies complex Hadoop environments, helping you understand exactly what happens behind the scenes when you administer your cluster. You'll gain unprecedented insight as you walk through building clusters from scratch and configuring high availability, performance, security, encryption, and other key attributes. The high-value administration skills you learn here will be indispensable no matter what Hadoop distribution you use or what Hadoop applications you run. Understand Hadoop's architecture from an administrator's standpoint Create simple and fully distributed clusters Run MapReduce and Spark applications in a Hadoop cluster Manage and protect Hadoop data and high availability Work with HDFS commands, file permissions, and storage management Move data, and use YARN to allocate resources and schedule jobs Manage job workflows with Oozie and Hue Secure, monitor, log, and optimize Hadoop Benchmark and troubleshoot Hadoop

This IBM® Redbooks® publication describes how the IBM Big Data Platform provides the integrated capabilities that are required for the adoption of Information Governance in the big data landscape. As organizations embark on new use cases, such as Big Data Exploration, an enhanced 360 view of customers, or Data Warehouse modernization, and absorb ever growing volumes and variety of data with accelerating velocity, the principles and practices of Information Governance become ever more critical to ensure trust in data and help organizations overcome the inherent risks and achieve the wanted value. The introduction of big data changes the information landscape. Data arrives faster than humans can react to it, and issues can quickly escalate into significant events. The variety of data now poses new privacy and security risks. The high volume of information in all places makes it harder to find where these issues, risks, and even useful information to drive new value and revenue are. Information Governance provides an organization with a framework that can align their wanted outcomes with their strategic management principles, the people who can implement those principles, and the architecture and platform that are needed to support the big data use cases. The IBM Big Data Platform, coupled with a framework for Information Governance, provides an approach to build, manage, and gain significant value from the big data landscape.

After a short description of the key concepts of big data the book explores on the secrecy and security threats posed especially by cloud based data storage. It delivers conceptual frameworks and models along with case studies of recent technology.

This book includes state-of-the-art discussions on various issues and aspects of the implementation, testing, validation, and application of big data in the context of healthcare. The concept of big data is

revolutionary, both from a technological and societal well-being standpoint. This book provides a comprehensive reference guide for engineers, scientists, and students studying/involved in the development of big data tools in the areas of healthcare and medicine. It also features a multifaceted and state-of-the-art literature review on healthcare data, its modalities, complexities, and methodologies, along with mathematical formulations. The book is divided into two main sections, the first of which discusses the challenges and opportunities associated with the implementation of big data in the healthcare sector. In turn, the second addresses the mathematical modeling of healthcare problems, as well as current and potential future big data applications and platforms.

Big data is defined as collections of datasets whose volume, velocity or variety is so large that it is difficult to store, manage, process and analyze the data using traditional databases and data processing tools. We have written this textbook to meet this need at colleges and universities, and also for big data service providers.

Explore big data concepts, platforms, analytics, and their applications using the power of Hadoop 3 Key Features Learn Hadoop 3 to build effective big data analytics solutions on-premise and on cloud Integrate Hadoop with other big data tools such as R, Python, Apache Spark, and Apache Flink Exploit big data using Hadoop 3 with real-world examples Book Description Apache Hadoop is the most popular platform for big data processing, and can be combined with a host of other big data tools to build powerful analytics solutions. Big Data Analytics with Hadoop 3 shows you how to do just that, by providing insights into the software as well as its benefits with the help of practical examples. Once you have taken a tour of Hadoop 3's latest features, you will get an overview of HDFS, MapReduce, and YARN, and how they enable faster, more efficient big data processing. You will then move on to learning how to integrate Hadoop with the open source tools, such as Python and R, to analyze and visualize data and perform statistical computing on big data. As you get acquainted with all this, you will explore how to use Hadoop 3 with Apache Spark and Apache Flink for real-time data analytics and stream processing. In addition to this, you will understand how to use Hadoop to build analytics solutions on the cloud and an end-to-end pipeline to perform big data analysis using practical use cases. By the end of this book, you will be well-versed with the analytical capabilities of the Hadoop ecosystem. You will be able to build powerful solutions to perform big data analytics and get insight effortlessly. What you will learn Explore the new features of Hadoop 3 along with HDFS, YARN, and MapReduce Get well-versed with the analytical capabilities of Hadoop ecosystem using practical examples Integrate Hadoop with R and Python for more efficient big data processing Learn to use Hadoop with Apache Spark and Apache Flink for real-time data analytics Set up a Hadoop cluster on AWS cloud Perform big data analytics on AWS using Elastic Map Reduce Who this book is for Big Data Analytics with Hadoop 3 is for you if you are looking to build high-performance analytics solutions for your enterprise or business using Hadoop 3's powerful features, or you're new to big data analytics. A basic understanding of the Java programming language is required.

As more corporations turn to Hadoop to store and process their most valuable data, the risk of a potential breach of those systems increases exponentially. This practical book not only shows Hadoop administrators and security architects how to protect Hadoop data from unauthorized access, it also shows how to limit the ability of an attacker to corrupt or modify data in the event of a security breach. Authors Ben Spivey and Joey Echeverria provide in-depth information about the security features available in Hadoop, and organize them according to common computer security concepts. You'll also get real-world examples that demonstrate how you can apply these concepts to your use cases. Understand the challenges of securing distributed systems, particularly Hadoop Use best practices for preparing Hadoop cluster hardware as securely as possible Get an overview of the Kerberos network authentication protocol Delve into authorization and accounting principles as they apply to Hadoop Learn how to use mechanisms to protect data in a Hadoop cluster, both in transit and at rest Integrate Hadoop data ingest into enterprise-wide security architecture Ensure that security architecture reaches all the way to end-user access

?This two-volume set (CCIS 1075 and CCIS 1076) constitutes the refereed proceedings of the Third International Conference on Advanced Informatics for Computing Research, ICAICR 2019, held in Shimla, India, in June 2019. The 78 revised full papers presented were carefully reviewed and selected from 382 submissions. The papers are organized in topical sections on computing methodologies; hardware; information systems; networks; software and its engineering.

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

Big data analytics and cloud computing is the fastest growing technologies in current era. This text book serves as a purpose in providing an understanding of big data principles and framework at the beginner's level. The text book covers various essential concepts of big-data analytics and processing tools such as HADOOP and YARN. The Textbook covers an analogical understanding on bridging cloud computing with big-data technologies with essential cloud infrastructure protocol and ecosystem concepts. PART I: Hadoop Distributed File System Basics, Running Example Programs and Benchmarks, Hadoop MapReduce Framework Essential Hadoop Tools, Hadoop YARN Applications, Managing Hadoop with Apache Ambari, Basic Hadoop Administration Procedures PART II: Introduction to Cloud Computing: Origins and Influences, Basic Concepts and Terminology, Goals and Benefits, Risks and Challenges.

Fundamental Concepts and Models: Roles and Boundaries, Cloud Characteristics, Cloud Delivery Models, Cloud Deployment Models. Cloud Computing Technologies:Broadband networks and internet architecture, data center technology, virtualization technology, web technology, multi-tenant technology, service Technology Cloud Infrastructure Mechanisms:Logical Network Perimeter, Virtual Server, Cloud Storage Device, Cloud Usage Monitor, Resource Replication, Ready-made environment

Human factors and usability issues have traditionally played a limited role in security research and secure systems development. Security experts have largely ignored usability issues--both because they often failed to recognize the importance of human factors and because they lacked the expertise to address them. But there is a growing recognition that today's security problems can be solved only by addressing issues of usability and human factors. Increasingly, well-publicized security breaches are attributed to human errors that might have been

prevented through more usable software. Indeed, the world's future cyber-security depends upon the deployment of security technology that can be broadly used by untrained computer users. Still, many people believe there is an inherent tradeoff between computer security and usability. It's true that a computer without passwords is usable, but not very secure. A computer that makes you authenticate every five minutes with a password and a fresh drop of blood might be very secure, but nobody would use it. Clearly, people need computers, and if they can't use one that's secure, they'll use one that isn't. Unfortunately, unsecured systems aren't usable for long, either. They get hacked, compromised, and otherwise rendered useless. There is increasing agreement that we need to design secure systems that people can actually use, but less agreement about how to reach this goal. *Security & Usability* is the first book-length work describing the current state of the art in this emerging field. Edited by security experts Dr. Lorrie Faith Cranor and Dr. Simson Garfinkel, and authored by cutting-edge security and human-computer interaction (HCI) researchers world-wide, this volume is expected to become both a classic reference and an inspiration for future research. *Security & Usability* groups 34 essays into six parts: *Realigning Usability and Security*--with careful attention to user-centered design principles, security and usability can be synergistic. *Authentication Mechanisms*-- techniques for identifying and authenticating computer users. *Secure Systems*--how system software can deliver or destroy a secure user experience. *Privacy and Anonymity Systems*--methods for allowing people to control the release of personal information. *Commercializing Usability: The Vendor Perspective*--specific experiences of security and software vendors (e.g., IBM, Microsoft, Lotus, Firefox, and Zone Labs) in addressing usability. *The Classics*--groundbreaking papers that sparked the field of security and usability. This book is expected to start an avalanche of discussion, new ideas, and further advances in this important field.

An easy-to-follow Apache Hadoop administrator's guide filled with practical screenshots and explanations for each step and configuration. This book is great for administrators interested in setting up and managing a large Hadoop cluster. If you are an administrator, or want to be an administrator, and you are ready to build and maintain a production-level cluster running CDH5, then this book is for you.

Go ahead, be skeptical about big data. The author was—at first. When the term “big data” first came on the scene, bestselling author Tom Davenport (*Competing on Analytics*, *Analytics at Work*) thought it was just another example of technology hype. But his research in the years that followed changed his mind. Now, in clear, conversational language, Davenport explains what big data means—and why everyone in business needs to know about it. *Big Data at Work* covers all the bases: what big data means from a technical, consumer, and management perspective; what its opportunities and costs are; where it can have real business impact; and which aspects of this hot topic have been oversold. This book will help you understand: • Why big data is important to you and your organization • What technology you need to manage it • How big data could change your job, your company, and your industry • How to hire, rent, or develop the kinds of people who make big data work • The key success factors in implementing any big data project • How big data is leading to a new approach to managing analytics With dozens of company examples, including UPS, GE, Amazon, United Healthcare, Citigroup, and many others, this book will help you seize all opportunities—from improving decisions, products, and services to strengthening customer relationships. It will show you how to put big data to work in your own organization so that you too can harness the power of this ever-evolving new resource.

This comprehensive new resource provides an introduction to fundamental Attribute Based Access Control (ABAC) models. This book provides valuable information for developing ABAC to improve information sharing within organizations while taking into consideration the planning, design, implementation, and operation. It explains the history and model of ABAC, related standards, verification and assurance, applications, as well as deployment challenges. Readers find authoritative insight into specialized topics including formal ABAC history, ABAC's relationship with other access control models, ABAC model validation and analysis, verification and testing, and deployment frameworks such as XACML. Next Generation Access Model (NGAC) is explained, along with attribute considerations in implementation. The book explores ABAC applications in SOA/workflow domains, ABAC architectures, and includes details on feature sets in commercial and open source products. This insightful resource presents a combination of technical and administrative information for models, standards, and products that will benefit researchers as well as implementers of ABAC systems in the field.

This handbook provides an overarching view of cyber security and digital forensic challenges related to big data and IoT environment, prior to reviewing existing data mining solutions and their potential application in big data context, and existing authentication and access control for IoT devices. An IoT access control scheme and an IoT forensic framework is also presented in this book, and it explains how the IoT forensic framework can be used to guide investigation of a popular cloud storage service. A distributed file system forensic approach is also presented, which is used to guide the investigation of Ceph. Minecraft, a Massively Multiplayer Online Game, and the Hadoop distributed file system environment are also forensically studied and their findings reported in this book. A forensic IoT source camera identification algorithm is introduced, which uses the camera's sensor pattern noise from the captured image. In addition to the IoT access control and forensic frameworks, this handbook covers a cyber defense triage process for nine advanced persistent threat (APT) groups targeting IoT infrastructure, namely: APT1, Molerats, Silent Chollima, Shell Crew, NetTraveler, ProjectSauron, CopyKittens, Volatile Cedar and Transparent Tribe. The characteristics of remote-controlled real-world Trojans using the Cyber Kill Chain are also examined. It introduces a method to leverage different crashes discovered from two fuzzing approaches, which can be used to enhance the effectiveness of fuzzers. Cloud computing is also often associated with IoT and big data (e.g., cloud-enabled IoT systems), and hence a survey of the cloud security literature and a survey of botnet detection approaches are presented in the book. Finally, game security solutions are studied and explained how one may circumvent such solutions. This handbook targets the security, privacy and forensics research community, and big data research community, including policy makers and government agencies, public and private organizations policy makers. Undergraduate and postgraduate students enrolled in cyber security and forensic programs will also find this handbook useful as a reference.

This IBM® Redbooks® publication provides topics to help the technical community take advantage of the resilience, scalability, and performance of the IBM Power Systems™

platform to implement or integrate an IBM Data Engine for Hadoop and Spark solution for analytics solutions to access, manage, and analyze data sets to improve business outcomes. This book documents topics to demonstrate and take advantage of the analytics strengths of the IBM POWER8® platform, the IBM analytics software portfolio, and selected third-party tools to help solve customer's data analytic workload requirements. This book describes how to plan, prepare, install, integrate, manage, and show how to use the IBM Data Engine for Hadoop and Spark solution to run analytic workloads on IBM POWER8. In addition, this publication delivers documentation to complement available IBM analytics solutions to help your data analytic needs. This publication strengthens the position of IBM analytics and big data solutions with a well-defined and documented deployment model within an IBM POWER8 virtualized environment so that customers have a planned foundation for security, scaling, capacity, resilience, and optimization for analytics workloads. This book is targeted at technical professionals (analytics consultants, technical support staff, IT Architects, and IT Specialists) that are responsible for delivering analytics solutions and support on IBM Power Systems.

Data analytics is core to business and decision making. The rapid increase in data volume, velocity and variety offers both opportunities and challenges. While open source solutions to store big data, like Hadoop, offer platforms for exploring value and insight from big data, they were not originally developed with data security and governance in mind. Big Data Management discusses numerous policies, strategies and recipes for managing big data. It addresses data security, privacy, controls and life cycle management offering modern principles and open source architectures for successful governance of big data. The author has collected best practices from the world's leading organizations that have successfully implemented big data platforms. The topics discussed cover the entire data management life cycle, data quality, data stewardship, regulatory considerations, data council, architectural and operational models are presented for successful management of big data. The book is a must-read for data scientists, data engineers and corporate leaders who are implementing big data platforms in their organizations.

Until recently, Hadoop deployments existed on hardware owned and run by organizations. Now, of course, you can acquire the computing resources and network connectivity to run Hadoop clusters in the cloud. But there's a lot more to deploying Hadoop to the public cloud than simply renting machines. This hands-on guide shows developers and systems administrators familiar with Hadoop how to install, use, and manage cloud-born clusters efficiently. You'll learn how to architect clusters that work with cloud-provider features—not just to avoid pitfalls, but also to take full advantage of these services. You'll also compare the Amazon, Google, and Microsoft clouds, and learn how to set up clusters in each of them. Learn how Hadoop clusters run in the cloud, the problems they can help you solve, and their potential drawbacks Examine the common concepts of cloud providers, including compute capabilities, networking and security, and storage Build a functional Hadoop cluster on cloud infrastructure, and learn what the major providers require Explore use cases for high availability, relational data with Hive, and complex analytics with Spark Get patterns and practices for running cloud clusters, from designing for price and security to dealing with maintenance

Every day, an increasing amount of our movements, transactions, and choices are becoming digitized and stored up into what has become known as “big data”—revolutionizing the way we do business today. And it's all there for your company to strategically utilize for giant profits! But where to begin? Think Bigger provides a roadmap for organizations looking to develop a profitable big data strategy. Sharing best practices from companies that have implemented a big data strategy including Walmart, InterContinental Hotel Group, Walt Disney, and Shell, this must-have resource for any business not wanting to fall far behind the competition covers the most important big data trends affecting organizations, as well as crucial types of analyses. Big data is changing the way businesses—and even governments—are operated and managed. And now, you too can revolutionize your business by learning how to properly employ the vast amount of digitalized information that is already available to you.

Hadoop Security Protecting Your Big Data Platform"O'Reilly Media, Inc."

With the proliferation of devices connected to the internet and connected to each other, the volume of data collected, stored, and processed is increasing every day, which brings new challenges in terms of information security. As big data expands with the help of public clouds, traditional security solutions tailored to private computing infrastructures and confined to a well-defined security perimeter, such as firewalls and demilitarized zones (DMZs), are no longer effective. New security functions are required to work over the heterogenous composition of diverse hardware, operating systems, and network domains. Security, Privacy, and Forensics Issues in Big Data is an essential research book that examines recent advancements in big data and the impact that these advancements have on information security and privacy measures needed for these networks. Highlighting a range of topics including cryptography, data analytics, and threat detection, this is an excellent reference source for students, software developers and engineers, security analysts, IT consultants, academicians, researchers, and professionals.

Scientific Essay from the year 2018 in the subject Computer Science - IT-Security, grade: 2, , language: English, abstract: In this we discuss security issues for big data Hadoop environment. Big data applications are a great benefit to organization, business and in many small and large scale industries. Security and privacy issues are magnified by velocity, variety and volume of big data. Hadoop projects security as top agenda which in turn represents classified as critical term. With the increasing acceptance of Hadoop, there is increasing trend to create a vast security feature. Therefore a traditional security mechanism, which are tailored to securing a small scale static data are in adequate. The important issues relating to Hadoop are authentication, authorization, editing and encryption within a cluster. In this paper we have highlighted different security aspects of big data Hadoop.

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.

Practical Hadoop Security is an excellent resource for administrators planning a production Hadoop deployment who want to secure their Hadoop clusters. A detailed guide to the security options and configuration within Hadoop itself, author Bhushan Lakhe takes you through a comprehensive study of how to implement defined security within a Hadoop cluster in a hands-on way. You will start with a detailed overview of all the security options available for Hadoop, including popular extensions like Kerberos and OpenSSH, and then delve into a hands-on implementation of user security (with illustrated code samples) with both in-the-box features and with security extensions implemented by leading vendors. No security system is complete without a monitoring and tracing facility, so Practical Hadoop Security next steps you through audit logging and monitoring technologies for Hadoop, as well as ready to use implementation and configuration examples--again with illustrated code samples. The book concludes with the most important aspect of Hadoop security – encryption. Both types of encryptions, for data in transit and data at rest, are discussed at length with leading open source projects that integrate directly with Hadoop at no licensing cost. Practical Hadoop Security: Explains importance of security, auditing and encryption within a Hadoop installation Describes how the leading players have incorporated these features within their Hadoop distributions and provided extensions Demonstrates how to set up and use these features to your benefit and make your Hadoop installation secure without impacting performance or ease of use

“This text should be required reading for everyone in contemporary business.” --Peter Woodhull, CEO, Modus21 “The one book that clearly describes and links Big Data concepts to business utility.” --Dr.

Christopher Starr, PhD “Simply, this is the best Big Data book on the market!” --Sam Rostam, Cascadian IT Group “...one of the most contemporary approaches I've seen to Big Data fundamentals...”

--Joshua M. Davis, PhD The Definitive Plain-English Guide to Big Data for Business and Technology Professionals Big Data Fundamentals provides a pragmatic, no-nonsense introduction to Big Data. Best-selling IT author Thomas Erl and his team clearly explain key Big Data concepts, theory and terminology, as well as fundamental technologies and techniques. All coverage is supported with case study examples and numerous simple diagrams. The authors begin by explaining how Big Data can propel an organization forward by solving a spectrum of previously intractable business problems. Next, they demystify key analysis techniques and technologies and show how a Big Data solution environment can be built and integrated to offer competitive advantages. Discovering Big Data's fundamental concepts and what makes it different from previous forms of data analysis and data science Understanding the business motivations and drivers behind Big Data adoption, from operational improvements through innovation Planning strategic, business-driven Big Data initiatives Addressing considerations such as data management, governance, and security Recognizing the 5 “V” characteristics of datasets in Big Data environments: volume, velocity, variety, veracity, and value Clarifying Big Data's relationships with OLTP, OLAP, ETL, data warehouses, and data marts Working with Big Data in structured, unstructured, semi-structured, and metadata formats Increasing value by integrating Big Data resources with corporate performance monitoring Understanding how Big Data leverages distributed and parallel processing Using NoSQL and other technologies to meet Big Data's distinct data processing requirements Leveraging statistical approaches of quantitative and qualitative analysis Applying computational analysis methods, including machine learning

The book contains the latest trend in IT industry 'BigData and Hadoop'. It explains how big is 'Big Data' and why everybody is trying to implement this into their IT project. It includes research work on various topics, theoretical and practical approach, each component of the architecture is described along with current industry trends. Big Data and Hadoop have taken together are a new skill as per the industry standards. Readers will get a compact book along with the industry experience and would be a reference to help readers. KEY FEATURES Overview Of Big Data, Basics of Hadoop, Hadoop Distributed File System, HBase, MapReduce, HIVE: The Dataware House Of Hadoop, PIG: The Higher Level Programming Environment, SQOOP: Importing Data From Heterogeneous Sources, Flume, Ozzie, Zookeeper & Big Data Stream Mining, Chapter-wise Questions & Previous Years Questions

This book constitutes the proceedings of the 8th International Conference on Big Data Analytics, BDA 2020, which took place during December 15-18, 2020, in Sonapat, India. The 11 full and 3 short papers included in this volume were carefully reviewed and selected from 48 submissions; the book also contains 4 invited and 3 tutorial papers. The contributions were organized in topical sections named as follows: data science systems; data science architectures; big data analytics in healthcare; information interchange of Web data resources; and business analytics.

In recent years, technological advances have led to significant developments within a variety of business applications. In particular, data-driven research provides ample opportunity for enterprise growth, if utilized efficiently. Privacy and Security Policies in Big Data is a pivotal reference source for the latest research on innovative concepts on the management of security and privacy analytics within big data. Featuring extensive coverage on relevant areas such as kinetic knowledge, cognitive analytics, and parallel computing, this publication is an ideal resource for professionals, researchers, academicians, advanced-level students, and technology developers in the field of big data.

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

[Copyright: 93c419caea1faf3e0321a30d0a59b521](https://www.pdfdrive.com/hadoop-security-protecting-your-big-data-platform-p123456789.html)