

## Hadoop The Definitive Guide 4th Edition Iteblog

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

"Get ready to unlock the power of your data. With the fourth edition of 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."--

If you're a web designer or app developer interested in sophisticated page styling, improved accessibility, and saving time and effort, this book is for you. This revised edition provides a comprehensive guide to CSS implementation, along with a thorough review of the latest CSS specifications. CSS is a constantly evolving language for describing the presentation of web content on screen, printers, speech synthesizers, screen readers, and chat windows. It is used by all browsers on all screen sizes on all types of IoT devices, including phones, computers, video games, televisions, watches, kiosks, and auto consoles. Authors Eric Meyer and Estelle Weyl show you how to improve user experience, speed development, avoid potential bugs, and add life and depth to your applications through layout, transitions and animations, borders, backgrounds, text properties, and many other tools and techniques. This guide covers: Selectors, specificity, and the cascade Values, units, fonts, and text properties Padding, borders, outlines, and margins Colors, backgrounds, and gradients Floats and positioning tricks Flexible box layout The new Grid layout system 2D and 3D transforms, transitions, and animation Filters, blending, clipping, and masking Media and feature queries

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. Docker containers offer simpler, faster, and more robust methods for developing, distributing, and running software than previously available. With this hands-on guide, you'll learn why containers are so important, what you'll gain by adopting Docker, and how to make it part of your development process. Ideal for developers, operations engineers, and system administrators—especially those keen to embrace a DevOps approach—Using Docker will take you from Docker and container basics to running dozens of containers on a multi-host system with networking and scheduling. The core of the book walks you through the steps needed to develop, test, and deploy a web application with Docker. Get started with Docker by building and deploying a simple web application Use Continuous Deployment techniques to push your application to production multiple times a day Learn various options and techniques for logging and monitoring multiple containers Examine networking and service discovery: how do containers find each other and how do you connect them? Orchestrate and cluster containers to address load-balancing, scaling, failover, and scheduling Secure your system by following the principles of defense-in-depth and least privilege

A comprehensive guide to mastering the most advanced Hadoop 3 concepts Key Features Get to grips with the newly introduced features and capabilities of Hadoop 3 Crunch and process data using MapReduce, YARN, and a host of tools within the Hadoop ecosystem Sharpen your Hadoop skills with real-world case studies and code Book Description Apache Hadoop is one of the most popular big data solutions for distributed storage and for processing large chunks of data. With Hadoop 3, Apache promises to provide a high-performance, more fault-tolerant, and highly efficient big data processing platform, with a focus on improved scalability and increased efficiency. With this guide, you'll understand advanced concepts of the Hadoop ecosystem tool. You'll learn how Hadoop works internally, study advanced concepts of different ecosystem tools, discover solutions to real-world use cases, and understand how to secure your cluster. It will then walk you through HDFS, YARN, MapReduce, and Hadoop 3 concepts. You'll be able to address common challenges like using Kafka efficiently, designing low latency, reliable message delivery Kafka systems, and handling high data volumes. As you advance, you'll discover how to address major challenges when building an enterprise-grade messaging system, and how to use different stream processing systems along with Kafka to fulfil your enterprise goals. By the end of this book, you'll have a complete understanding of how components in the Hadoop ecosystem are effectively integrated to implement a fast and reliable data pipeline, and you'll be equipped to tackle a range of real-world problems in data pipelines. What you will learn Gain an in-depth understanding of distributed computing using Hadoop 3 Develop enterprise-grade applications using Apache Spark, Flink, and more Build scalable and high-performance Hadoop data pipelines with security, monitoring, and data governance Explore batch data processing patterns and how to model data in Hadoop Master best practices for enterprises using, or planning to use, Hadoop 3 as a data platform Understand security aspects of Hadoop, including authorization and authentication Who this book is for If you want to become a big data professional by mastering the advanced concepts of Hadoop, this book is for you. You'll also find this book useful if you're a Hadoop professional looking to strengthen your knowledge of the Hadoop ecosystem. Fundamental knowledge of the Java programming language and basics of Hadoop is necessary to get started with this book.

Imprisoned for life aboard a zeppelin that floats high above a fantastic metropolis, greeting-card writer Harold Winslow pens his memoirs. His only companions are the disembodied voice of Miranda Taligent, the only woman he has ever loved, and the cryogenically frozen body of her father, Prospero, the genius and industrial magnate who drove her insane. As Harold heads toward a last desperate confrontation with Prospero to save Miranda's life, he finds himself an unwitting participant in the creation of the greatest invention of them all: the perpetual motion machine. Beautifully written, stunningly imagined, and wickedly funny, Dexter Palmer's *The Dream of Perpetual Motion* is a heartfelt meditation on the place of love in a world dominated by technology.

Summary Mondrian in Action teaches business users and developers how to use Mondrian and related tools for strategic business analysis. You'll learn how to design and populate a data warehouse and present the data via a multidimensional model. You'll follow examples showing how to create a Mondrian schema and then expand it to add basic security based on the users' roles. About the Technology Mondrian is an open source, lightning-fast data analysis engine designed to help you explore your business data and perform speed-of-thought analysis.

Mondrian can be integrated into a wide variety of business analysis applications and learning it requires no specialized technical knowledge. About this Book Mondrian in Action teaches you to use Mondrian for strategic business analysis. In it, you'll learn how to organize and present data in a multidimensional manner. You'll follow apt and thoroughly explained examples showing how to create a Mondrian schema and then expand it to add basic security based on users' roles. Developers will discover how to integrate Mondrian using its olap4j Java API and web service calls via XML for Analysis. Written for developers building data analysis solutions. Appropriate for tech-savvy business users and DBAs needing to query and report on data. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. What's Inside Mondrian from the ground up—no experience required A primer on business analytics Using Mondrian with a variety of leading applications Optimizing and restricting business data for fast, secure analysis About the Authors William D. Back is an Enterprise Architect and Director of Pentaho Services. Nicholas Goodman is a Business Intelligence pro who has authored training courses on OLAP and Mondrian. Julian Hyde founded Mondrian and is the project's lead developer. Table of Contents Beyond reporting: business analytics Mondrian: a first look Creating the data mart Multidimensional modeling: making analytics data accessible How schemas grow Securing data Maximizing Mondrian performance Dynamic security Working with Mondrian and Pentaho Developing with Mondrian Advanced analytics

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

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

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.

Get a solid grounding in Apache Oozie, the workflow scheduler system for managing Hadoop jobs. With this hands-on guide, two experienced Hadoop practitioners walk you through the intricacies of this powerful and flexible platform, with numerous examples and real-world use cases. Once you set up your Oozie server, you'll dive into techniques for writing and coordinating workflows, and learn how to write complex data pipelines. Advanced topics show you how to handle shared libraries in Oozie, as well as how to implement and manage Oozie's security capabilities. Install and configure an Oozie server, and get an overview of basic concepts Journey through the world of writing and configuring workflows Learn how the Oozie coordinator schedules and executes workflows based on triggers Understand how Oozie manages data dependencies Use Oozie bundles to package several coordinator apps into a data pipeline Learn about security features and shared library management Implement custom extensions and write your own EL functions and actions Debug workflows and manage Oozie's operational details

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

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

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

As we all know by now, wireless networks offer many advantages over fixed (or wired) networks. Foremost on that list is mobility, since going wireless frees you from the tether of an Ethernet cable at a desk. But that's just the tip of the cable-free iceberg. Wireless networks are also more flexible, faster and easier for you to use, and more affordable to deploy and maintain. The de facto standard for wireless networking is the 802.11 protocol, which includes Wi-Fi (the wireless standard known as 802.11b) and its faster cousin, 802.11g. With easy-to-install 802.11 network hardware available everywhere you turn, the choice seems simple, and many people dive into wireless computing with less thought and planning than they'd give to a wired network. But it's wise to be familiar with both the capabilities and risks associated with the 802.11 protocols. And 802.11 Wireless Networks: The Definitive Guide, 2nd Edition is the perfect place to start. This updated edition covers everything you'll ever need to know about wireless technology. Designed with the system administrator or serious home user in mind, it's a no-nonsense guide for setting up 802.11 on Windows and Linux. Among the wide range of topics covered are discussions on: deployment considerations network monitoring and performance tuning wireless security issues how to use and select access points network monitoring essentials wireless card configuration security issues unique to wireless networks With wireless technology, the advantages to its users are indeed plentiful. Companies no longer have to deal with the hassle and expense of wiring buildings, and households with several computers can avoid fights over who's online. And now, with 802.11 Wireless Networks: The Definitive Guide, 2nd Edition, you can integrate wireless technology into your current infrastructure with the utmost confidence.

"Apache Hadoop is helping drive the Big Data revolution. Now, its data processing has been completely overhauled: Apache Hadoop YARN provides resource management at data center scale and easier ways to create distributed applications that process petabytes of data. And now in Apache Hadoop™ YARN, two Hadoop technical leaders show you how to develop new applications and adapt existing code to fully leverage these revolutionary advances." -- From the Amazon

Counsels programmers and administrators for big and small organizations on how to work with large-scale application datasets using Apache Hadoop, discussing its capacity for storing and processing large amounts of data while demonstrating best practices for building reliable and scalable distributed systems. Original.

Every enterprise application creates data, whether it's log messages, metrics, user activity, outgoing messages, or something else. And how to move all of this data becomes nearly as important as the data itself. If you're an application architect, developer, or production engineer new to Apache Kafka, this practical guide shows you how to use this open source streaming platform to handle real-time data feeds. Engineers from Confluent and LinkedIn who are responsible for developing Kafka explain how to deploy production Kafka clusters, write reliable event-driven microservices, and build scalable stream-processing applications with this platform. Through detailed examples, you'll learn Kafka's design principles, reliability guarantees, key APIs, and architecture details, including the replication protocol, the controller, and the storage layer. Understand publish-subscribe messaging and how it fits in the big data ecosystem. Explore Kafka producers and consumers for writing and reading messages Understand Kafka patterns and use-case requirements to ensure reliable data delivery Get best practices for building data pipelines and applications with Kafka Manage Kafka in production, and learn to perform monitoring, tuning, and maintenance tasks Learn the most critical metrics among Kafka's operational measurements Explore how Kafka's stream delivery capabilities make it a perfect source for stream processing systems

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!

Foreword. A transformed scientific method. Earth and environment. Health and wellbeing. Scientific infrastructure. Scholarly communication. Get ready to unlock the power of your data. With the fourth edition of 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. Using Hadoop 2 exclusively, author Tom White presents new chapters on YARN and several Hadoop-related projects such as Parquet, Flume, Crunch, and Spark. You'll learn about recent changes to Hadoop, and explore new case studies on Hadoop's role in healthcare systems and genomics data processing. Learn fundamental components such as MapReduce, HDFS, and YARN Explore MapReduce in depth, including steps for developing applications with it Set up and maintain a Hadoop cluster running HDFS and MapReduce on YARN Learn two data formats: Avro for data serialization and Parquet for nested data Use data ingestion tools such as Flume (for streaming data) and Sqoop (for bulk data transfer) Understand how high-level data processing tools like Pig, Hive, Crunch, and Spark work with Hadoop Learn the HBase distributed database and the ZooKeeper distributed configuration

service

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

Data in all domains is getting bigger. How can you work with it efficiently? Recently updated for Spark 1.3, this book introduces Apache Spark, the open source cluster computing system that makes data analytics fast to write and fast to run. With Spark, you can tackle big datasets quickly through simple APIs in Python, Java, and Scala. This edition includes new information on Spark SQL, Spark Streaming, setup, and Maven coordinates. Written by the developers of Spark, this book will have data scientists and engineers up and running in no time. You'll learn how to express parallel jobs with just a few lines of code, and cover applications from simple batch jobs to stream processing and machine learning. Quickly dive into Spark capabilities such as distributed datasets, in-memory caching, and the interactive shell Leverage Spark's powerful built-in libraries, including Spark SQL, Spark Streaming, and MLlib Use one programming paradigm instead of mixing and matching tools like Hive, Hadoop, Mahout, and Storm Learn how to deploy interactive, batch, and streaming applications Connect to data sources including HDFS, Hive, JSON, and S3 Master advanced topics like data partitioning and shared variables What would happen if you optimized a data store for the operations application developers actually use? You'd arrive at MongoDB, the reliable document-oriented database. With this concise guide, you'll learn how to build elegant database applications with MongoDB and PHP. Written by the Chief Solutions Architect at 10gen—the company that develops and supports this open source database—this book takes you through MongoDB basics such as queries, read-write operations, and administration, and then dives into MapReduce, sharding, and other advanced topics. Get out of the relational database rut, and take advantage of a high-performing system optimized for operations and scale. Learn step-by-step the tools you need to build PHP applications with MongoDB Perform Create, Read, Update, and Delete (CRUD) operations, and learn how to perform queries to retrieve data Administer your database, and access and manipulate data with the MongoDB Shell Use functions to work with sets, arrays, and multiple documents to perform synchronous, asynchronous, and atomic operations Discover PHP's community tools and libraries, and why they're valuable Work with regular expressions, aggregation, MapReduce, replication, and sharding A beginner's guide to storing, managing, and analyzing data with the updated features of Elastic 7.0 Key Features Gain access to new features and updates introduced in Elastic Stack 7.0 Grasp the fundamentals of Elastic Stack including Elasticsearch, Logstash, and Kibana Explore useful tips for using Elastic Cloud and deploying Elastic Stack in production environments Book Description The Elastic Stack is a powerful combination of tools for techniques such as distributed search, analytics, logging, and visualization of data. Elastic Stack 7.0 encompasses new features and capabilities that will enable you to find unique insights into analytics using these techniques. This book will give you a fundamental understanding of what the stack is all about, and help you use it efficiently to build powerful real-time data processing applications. The first few sections of the book will help you understand how to set up the stack by installing tools, and exploring their basic configurations. You'll then get up to speed with using Elasticsearch for distributed searching and analytics, Logstash for logging, and Kibana for data visualization. As you work through the book, you will discover the technique of creating custom plugins using Kibana and Beats. This is followed by coverage of the Elastic X-Pack, a useful extension for effective security and monitoring. You'll also find helpful tips on how to use Elastic Cloud and deploy Elastic Stack in production environments. By the end of this book, you'll be well versed with the fundamental Elastic Stack functionalities and the role of each component in the stack to solve different data processing problems. What you will learn Install and configure an Elasticsearch architecture Solve the full-text search problem with Elasticsearch Discover powerful analytics capabilities through aggregations using Elasticsearch Build a data pipeline to transfer data from a variety of sources into Elasticsearch for analysis Create interactive dashboards for effective storytelling with your data using Kibana Learn how to secure, monitor and use Elastic Stack's alerting and reporting capabilities Take applications to an on-premise or cloud-based production environment with Elastic Stack Who this book is for This book is for entry-level data professionals, software engineers, e-commerce developers, and full-stack developers who want to learn about Elastic Stack and how the real-time processing and search engine works for business analytics and enterprise search applications. Previous experience with Elastic Stack is not required, however knowledge of data warehousing and database concepts will be helpful.

This book presents a collection of papers from the 10th ISAAC Congress 2015, held in Macau, China. The papers, prepared by respected international experts, address recent results in Mathematics, with a special focus on Analysis. By structuring the content according to the various mathematical topics, the volume offers specialists and non-specialists alike an excellent source of information on the state-of-the-art in Mathematical Analysis and its interdisciplinary applications.

Learn how to write, tune, and port SQL queries and other statements for a Big Data environment, using Impala—the massively parallel processing SQL query engine for Apache Hadoop. The best practices in this practical guide help you design database schemas that not only interoperate with other Hadoop components, and are convenient for administrators to manage and monitor, but also accommodate future expansion in data size and evolution of software capabilities. Written by John Russell, documentation lead for the Cloudera Impala project, this book gets you working with the most recent Impala releases quickly. Ideal for database developers and business analysts, the latest revision covers analytics functions, complex types, incremental statistics, subqueries, and submission to the Apache incubator. Getting Started with Impala includes advice from Cloudera's development team, as well as insights from its consulting engagements with customers. Learn how Impala integrates with a wide range of Hadoop components Attain high performance and scalability for huge data sets on production clusters Explore common developer tasks, such as porting code to Impala and optimizing performance Use tutorials for working with billion-row tables, date- and time-based values, and other techniques Learn how to transition from rigid schemas to a flexible model that evolves as needs change Take a deep dive into joins and the roles of statistics

Hadoop in Action teaches readers how to use Hadoop and write MapReduce programs. The intended readers are programmers, architects, and project managers who have to process large amounts of data offline. Hadoop in Action will lead the reader from obtaining a copy of Hadoop to setting it up in a cluster and writing data analytic programs. The book begins by making the basic



idea of Hadoop and MapReduce easier to grasp by applying the default Hadoop installation to a few easy-to-follow tasks, such as analyzing changes in word frequency across a body of documents. The book continues through the basic concepts of MapReduce applications developed using Hadoop, including a close look at framework components, use of Hadoop for a variety of data analysis tasks, and numerous examples of Hadoop in action. Hadoop in Action will explain how to use Hadoop and present design patterns and practices of programming MapReduce. MapReduce is a complex idea both conceptually and in its implementation, and Hadoop users are challenged to learn all the knobs and levers for running Hadoop. This book takes you beyond the mechanics of running Hadoop, teaching you to write meaningful programs in a MapReduce framework. This book assumes the reader will have a basic familiarity with Java, as most code examples will be written in Java. Familiarity with basic statistical concepts (e.g. histogram, correlation) will help the reader appreciate the more advanced data processing examples. 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 could you do with data if scalability wasn't a problem? With this hands-on guide, you'll learn how Apache Cassandra handles hundreds of terabytes of data while remaining highly available across multiple data centers -- capabilities that have attracted Facebook, Twitter, and other data-intensive companies. Cassandra: The Definitive Guide provides the technical details and practical examples you need to assess this database management system and put it to work in a production environment. Author Eben Hewitt demonstrates the advantages of Cassandra's nonrelational design, and pays special attention to data modeling. If you're a developer, DBA, application architect, or manager looking to solve a database scaling issue or future-proof your application, this guide shows you how to harness Cassandra's speed and flexibility. Understand the tenets of Cassandra's column-oriented structure Learn how to write, update, and read Cassandra data Discover how to add or remove nodes from the cluster as your application requires Examine a working application that translates from a relational model to Cassandra's data model Use examples for writing clients in Java, Python, and C# Use the JMX interface to monitor a cluster's usage, memory patterns, and more Tune memory settings, data storage, and caching for better performance

Plugins will soon be a thing of the past. The Definitive Guide to HTML5 Video is the first authoritative book on HTML5 video, the new web standard that allows browsers to support audio and video elements natively. This makes it very easy for web developers to publish audio and video, integrating both within the general presentation of web pages. For example, media elements can be styled using CSS (style sheets), integrated into SVG (scalable vector graphics), and manipulated in a Canvas. The book offers techniques for providing accessibility to media elements, enabling consistent handling of alternative representations of media resources. The Definitive Guide to HTML5 Video dives deep into the markup that is introduced for media element support in browsers. You'll explore the default user interface offered through browsers for media elements, as well as the JavaScript API provided to control their behavior. You'll also learn how to account for H.264 and Ogg Theora codecs.

This pocket guide presents the most crucial information about SQL in a compact and easily accessible format, covering the four commonly used SQL variants--Oracle, IBM DB2, Microsoft SQL Server, and MySQL. Topics include: Data manipulation statements (SELECT, DELETE, INSERT, UPDATE, MERGE) and transaction control statements (START TRANSACTION, SAVEPOINT, COMMIT, ROLLBACK). Common SQL functions (date, numeric, math, trigonometric, string, conversion, aggregate) Such topics as literals, NULLs, CASE expressions, datatype conversion, regular expressions, grouping and summarizing data, joining tables, and writing queries (hierarchical, recursive, union, flashback) and subqueries. Instead of presenting complex and confusing syntax diagrams, the book teaches by example, showing the SQL statements and options that readers are most like to use. All example data is available on the O'Reilly web site. "If you need fast, accurate SQL information, with examples for multiple database engines, be sure to check out this book."--Chris Kempster, Senior DBA and author of SQL Server 2000 for the Oracle DBA, [www.chriskempster.com](http://www.chriskempster.com)

"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

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

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 This book is useful for Hadoop administrators who need to learn how to monitor and diagnose their clusters. Also, the book will prove useful for new users of the technology, as the language used is simple and easy to grasp.

Get up to speed on the latest Ethernet capabilities for building and maintaining networks for everything from homes and offices to data centers and server machine rooms. This thoroughly revised, comprehensive guide covers a wide range of Ethernet technologies, from basic operation to network management, based on the authors' many years of field experience. When should you upgrade to higher speed Ethernet? How do you use switches to build larger networks? How do you troubleshoot the system? This book provides the answers. If you're looking to build a scalable network with Ethernet to satisfy greater bandwidth and market requirements, this book is indeed the definitive guide. Examine the most widely used media systems, as well as advanced 40 and 100 gigabit Ethernet Learn about Ethernet's four basic elements and the IEEE standards Explore full-duplex Ethernet, Power over Ethernet, and Energy Efficient Ethernet Understand structured cabling systems and the components you need to build your Ethernet system Use Ethernet switches to expand and improve network design Delve into Ethernet performance, from specific channels to the entire network Get troubleshooting techniques for problems common to twisted-pair and fiber optic systems

[Copyright: 0a873e3ce807b119aef4ed7d73af5ea7](https://www.it-ebooks.info/book/1000000/0a873e3ce807b119aef4ed7d73af5ea7)