

Hadoop Introduction Core Servlets

The world of medical technologies is undergoing a sea change in the domain of consumer culture. Having a grasp on what appeals to consumers and how consumers are making purchasing decisions is essential to the success of any organization that thrives by offering a product or service. As such, it is vital to examine the consumer-centered aspects of medical technological developments that have a patient-centered focus and allow patients to take part in their own personal health and wellness. *Consumer-Driven Technologies in Healthcare: Breakthroughs in Research and Practice* is a critical source of academic knowledge on the use of smartphones and other technological devices for cancer therapy, fitness and wellness, chronic disease monitoring, and other areas. The tracking of these items using technology has allowed consumers to take control of their own healthcare. Highlighting a range of pertinent topics such as clinical decision support systems, patient engagement, and electronic health records, this publication is an ideal reference source for doctors, nurse practitioners, hospital administrators, medical professionals, IT professionals, academicians, and researchers interested in advancing medical practice through technology.

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.

Learn and implement various techniques related to testing, monitoring and optimization for microservices architecture. Key Features Learn different approaches for testing microservices to design and implement, robust and secure applications Become more efficient while working with microservices Explore Testing and Monitoring tools such as JMeter, Ready API, and AppDynamics Book Description Microservices are the latest "right" way of developing web applications. Microservices architecture has been gaining momentum over the past few years, but once you've started down the microservices path, you need to test and optimize the services. This book focuses on exploring various testing, monitoring, and optimization techniques for microservices. The book starts with the evolution of software architecture style, from monolithic to virtualized, to microservices architecture. Then you will explore methods to deploy microservices and various implementation patterns. With the help of a real-world example, you will understand how external APIs help product developers to focus on core competencies. After that, you will learn testing techniques, such as Unit Testing, Integration Testing, Functional Testing, and Load Testing. Next, you will explore performance testing tools, such as JMeter, and Gatling. Then, we deep dive into monitoring techniques and learn performance benchmarking of the various architectural components. For this, you will explore monitoring tools such as AppDynamics, Dynatrace, AWS CloudWatch, and Nagios. Finally, you will learn to identify, address, and report various performance issues related to microservices. What you will learn Understand the architecture of microservices and how to build services Establish how external APIs help to accelerate the development process Understand testing techniques, such as unit testing, integration testing, end-to-end testing, and UI/functional testing Explore various tools related to the performance testing, monitoring, and optimization of microservices Design strategies for performance testing Identify performance issues and fine-tune performance Who this book is for This book is for developers who are involved with microservices architecture to develop robust and secure applications. Basic knowledge of microservices is essential in order to get the most out of this book.

This book aims at providing fundamental concepts of Python programming. It is a good text book as well as reference book. Here concepts are presented in the form of programs making it quite easy and simple for students to understand. It showcases actual screen shots of the programs from the programming environment to make it more student friendly. Because of the user friendly interface provided in the book a novice learner can also learn Python programming without any difficulty. As Python is open source, programs written in this book can execute on different operating systems like Windows, Linux, and Mac etc. Topics covered in this book Python Language Basics, Python Collections- Compound Data Types, List, Tuple, Dictionary and Set, Setting up Development Environment, Introduction to Anaconda and Installation of Anaconda. All the concepts are explained with practical examples and screenshots of the same are embedded.

Contains the complete reference for all Base SAS procedures. Provides information about what each procedure does and, if relevant, the kind of output that it produces. This title is also available online. SAS Products and Releases: Base SAS: 9.3_M2. Operating Systems: All

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

Our world is being revolutionized by data-driven methods: access to large amounts of data has generated new insights and opened exciting new opportunities in commerce, science, and computing applications. Processing the enormous quantities of data necessary for these advances requires large clusters, making distributed computing paradigms more crucial than ever. MapReduce is a programming model for expressing distributed computations on massive datasets and an execution framework for large-scale data processing on clusters of commodity servers. The programming model provides an easy-to-understand abstraction for designing scalable algorithms, while the execution framework transparently handles many system-level details, ranging from scheduling to synchronization to fault tolerance. This book focuses on MapReduce algorithm design, with an emphasis on text processing algorithms common in natural language processing, information retrieval, and machine learning. We introduce the notion of MapReduce design patterns, which represent general reusable solutions to commonly occurring

problems across a variety of problem domains. This book not only intends to help the reader "think in MapReduce", but also discusses limitations of the programming model as well. This volume is a printed version of a work that appears in the Synthesis Digital Library of Engineering and Computer Science. Synthesis Lectures provide concise, original presentations of important research and development topics, published quickly, in digital and print formats. For more information visit www.morganclaypool.com

This IBM® Platform Computing Solutions Redbooks® publication is the first book to describe each of the available offerings that are part of the IBM portfolio of Cloud, analytics, and High Performance Computing (HPC) solutions for our clients. This IBM Redbooks publication delivers descriptions of the available offerings from IBM Platform Computing that address challenges for our clients in each industry. We include a few implementation and testing scenarios with selected solutions. This publication helps strengthen the position of IBM Platform Computing solutions with a well-defined and documented deployment model within an IBM System x® environment. This deployment model offers clients a planned foundation for dynamic cloud infrastructure, provisioning, large-scale parallel HPC application development, cluster management, and grid applications. This IBM publication is targeted to IT specialists, IT architects, support personnel, and clients. This book is intended for anyone who wants information about how IBM Platform Computing solutions use IBM to provide a wide array of client solutions. 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 Analytics is increasingly an integral part of day-to-day operations at today's leading businesses, and transformation is also occurring through huge growth in mobile and digital channels. Enterprise organizations are attempting to leverage analytics in new ways and transition existing analytics capabilities to respond with more flexibility while making the most efficient use of highly valuable data science skills. The recent growth and adoption of Apache Spark as an analytics framework and platform is very timely and helps meet these challenging demands. The Apache Spark environment on IBM z/OS® and Linux on IBM z Systems™ platforms allows this analytics framework to run on the same enterprise platform as the originating sources of data and transactions that feed it. If most of the data that will be used for Apache Spark analytics, or the most sensitive or quickly changing data is originating on z/OS, then an Apache Spark z/OS based environment will be the optimal choice for performance, security, and governance. This IBM® Redpaper™ publication explores the enterprise analytics market, use of Apache Spark on IBM z Systems™ platforms, integration between Apache Spark and other enterprise data sources, and case studies and examples of what can be achieved with Apache Spark in enterprise environments. It is of interest to data scientists, data engineers, enterprise architects, or anybody looking to better understand how to combine an analytics framework and platform on enterprise systems.

This supplement describes the SAS/ACCESS features and options that are available for SYBASE. When you use this supplement along with SAS/ACCESS Interface to Relational Databases: Reference, you will have a comprehensive guide to SAS/ACCESS software for your database management system (DBMS). This supplement explains how SAS/ACCESS software has been customized to accommodate the particular requirements and capabilities of your DBMS. This title is also available online. This title is designed for users of SAS/ACCESS software who need specific information about SYBASE. Supports releases 9.1 and higher of SAS software.

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.

Let us full stack development with Spring Boot and React JS. DESCRIPTION Designing Application with Spring Boot 2 & React JS is divided into three parts. The first part introduces you to the essentials of the Spring Boot 2.2 Framework and you will learn how to create REST APIs and how to secure REST APIs. Part 2 steps behind the front end application development with React JS and discuss React features and its advantages toward the front end application development. Part 3 expands on that by showing how to deploy backend and frontend application the PaaS platform and also will discuss how to deploy application container technologies such as Docker. KEY FEATURES ? This book has a very specific goal to make developing REST applications easier and focusing on common challenges of the design of the application with best practices. ? This book is providing practical code examples from real-world experiences. ? This book is not only about Spring Boot 2.2 and React JS overview but also has an in-depth discussion about adopted REST Architectural pattern and its constraints to create the REST APIs. ? The book can act as a tool for learning Spring Boot 2.2 and React JS for the first time as well as a guide and reference for those wanting to dig deeper into specific features. ? This book is also providing deeper information about the Spring Security and JWT token-based authentication for your REST applications. ? This does not only provide information about to design an application using Spring Boot and React JS but also providing how to deploy your application to the cloud platform (PaaS). ? Containerization using Docker is another key feature of this book, how to create a Docker image and how to run it. WHAT WILL YOU LEARN ? Exploring Spring Boot 2.2 new features and essential key components such as Starters, Autoconfiguration, CLI, Actuator. ? Develop a REST application using Spring Boot 2.2 and DevTools. ? Exploring Spring Boot Auto Configuration and Customization. ? Creating application profiles based on the

environments. ? Learn to configure backend data using JDBC and Spring Data JPA. ? Learn to configure a DataSource for H2 DB, and also for Maria DB. ? Learn best practices for designing a REST architecture based application. ? Creating a REST application using HATEOAS. ? Consuming REST APIs endpoints with RestTemplate, Traverson, and WebClient. ? Exploring JWT web token for the RESTful APIs and explores how to secure REST APIs using OAuth2 and Spring security. ? Creating TESTING module of the Spring Boot application and Unit & Integration testing. ? Discuss React JS and its components and also discuss React KS features and its advantages and disadvantage. ? Exploring how to create ReactJS components and how to manage ReactJS component lifecycle. ? Taking a quick overview of consuming the REST API using the React application. ? Deploying the application to the Cloud platform (PaaS). ? Containerization and Deploy using Docker containers WHO THIS BOOK IS FOR Designing Application with Spring Boot 2.2 & React JS is for all Java developers who want to learn Spring Boot 2.2 and React JS as in the enterprise application. Therefore, enterprise Java developers will find it particularly useful in the understanding of Spring Boot 2.2 and React JS and how to develop a backend RESTful application using the Spring Boot 2.2 and frontend application using React JS framework. They will most fully appreciate the examples presented in this book. Before reading this book, readers should have basic knowledge of core java, spring, servlet, filter, XML, and JavaScript. TABLE OF CONTENTS Getting Started with Spring Boot 2.2 Customizing Auto-Configuration Configuring Data and CRUD operations Creating REST APIs with Spring Boot 2.2 Securing REST APIs Testing Spring Boot Application Getting Started with React Creating and Styling React Components Consuming the REST API with React JS Deploying and Containerizing Application

What is this book about? Expert One-on-One J2EE Development without EJB shows Javadevelopers and architects how to build robust J2EE applicationswithout having to use Enterprise JavaBeans (EJB). This practical,code-intensive guide provides best practices for using simpler andmore effective methods and tools, including JavaServer pages,servlets, and lightweight frameworks. What does this book cover? The book begins by examining the limits of EJB technology— what it does well and not so well. Then the authors guideyou through alternatives to EJB that you can use to create higherquality applications faster and at lower cost — both agilemethods as well as new classes of tools that have evolved over thepast few years. They then dive into the details, showing solutions based on thelightweight framework they pioneered on SourceForge — one ofthe most innovative open source communities. They demonstrate howto leverage practical techniques and tools, including the popularopen source Spring Framework and Hibernate. This book also guidesyou through productive solutions to core problems, such astransaction management, persistence, remoting, and Web tier design.You will examine how these alternatives affect testing,performance, and scalability, and discover how lightweightarchitectures can slash time and effort on many projects. What will you learn from this book? Here are some details on what you'll find in this book: How to find the simplest and most maintainable architecture foryour application Effective transaction management without EJB How to solve common problems in enterprise software developmentusing AOP and Inversion of Control Web tier design and the place of the Web tier in awell-designed J2EE application Effective data access techniques for J2EE applications withJDBC, Hibernate, and JDO How to leverage open source products to improve productivityand reduce custom coding How to design for optimal performance and scalability

Consumer-Driven Technologies in Healthcare: Breakthroughs in Research and PracticeBreakthroughs in Research and PracticeIGI Global

The go-to guidebook for deploying Big Data solutions withHadoop Today's enterprise architects need to understand how the Hadoopframeworks and APIs fit together, and how they can be integrated todeliver real-world solutions. This book is a practical, detailedguide to building and implementing those solutions, with code-levelinstruction in the popular Wrox tradition. It covers storing datawith HDFS and Hbase, processing data with MapReduce, and automatingdata processing with Oozie. Hadoop security, running Hadoop withAmazon Web Services, best practices, and automating Hadoopprocesses in real time are also covered in depth. With in-depth code examples in Java and XML and the latest onrecent additions to the Hadoop ecosystem, this complete resourcealso covers the use of APIs, exposing their inner workings andallowing architects and developers to better leverage and customizethem. The ultimate guide for developers, designers, and architectswho need to build and deploy Hadoop applications Covers storing and processing data with various technologies,automating data processing, Hadoop security, and deliveringreal-time solutions Includes detailed, real-world examples and code-levelguidelines Explains when, why, and how to use these tools effectively Written by a team of Hadoop experts in theprogrammer-to-programmer Wrox style Professional Hadoop Solutions is the reference enterprisearchitects and developers need to maximize the power of Hadoop.

You've heard the hype about Hadoop: it runs petabyte–scale data mining tasks insanely fast, it runs gigantic tasks on clouds for absurdly cheap, it's been heavily committed to by tech giants like IBM, Yahoo!, and the Apache Project, and it's completely open-source (thus free). But what exactly is it, and more importantly, how do you even get a Hadoop cluster up and running? From Apress, the name you've come to trust for hands–on technical knowledge, Pro Hadoop brings you up to speed on Hadoop. You learn the ins and outs of MapReduce; how to structure a cluster, design, and implement the Hadoop file system; and how to build your first cloud–computing tasks using Hadoop. Learn how to let Hadoop take care of distributing and parallelizing your software—you just focus on the code, Hadoop takes care of the rest. Best of all, you'll learn from a tech professional who's been in the Hadoop scene since day one. Written from the perspective of a principal engineer with down–in–the–trenches knowledge of what to do wrong with Hadoop, you learn how to avoid the common, expensive first errors that everyone makes with creating their own Hadoop system or inheriting someone else's. Skip the novice stage and the expensive, hard–to–fix mistakes...go straight to seasoned pro on the hottest cloud–computing framework with Pro Hadoop. Your productivity will blow your managers away.

Dig deep into the data with a hands-on guide to machine learning with updated examples and more! Machine Learning: Hands-On for Developers and Technical Professionals provides hands-on instruction and fully-coded working examples for the most common machine learning techniques used by developers and technical professionals. The book contains a breakdown of each ML variant, explaining how it works and how it is used within certain industries, allowing readers to incorporate the presented techniques into their own work as they follow along. A core tenant of machine learning is a strong focus on data preparation, and a full exploration of the various types of learning algorithms illustrates how the proper tools can help any developer extract information and insights from existing data. The book includes a full complement of Instructor's Materials to facilitate use in the classroom, making this resource useful for students and as a professional reference. At its core, machine learning is a mathematical, algorithm-based technology that forms

the basis of historical data mining and modern big data science. Scientific analysis of big data requires a working knowledge of machine learning, which forms predictions based on known properties learned from training data. Machine Learning is an accessible, comprehensive guide for the non-mathematician, providing clear guidance that allows readers to: Learn the languages of machine learning including Hadoop, Mahout, and Weka Understand decision trees, Bayesian networks, and artificial neural networks Implement Association Rule, Real Time, and Batch learning Develop a strategic plan for safe, effective, and efficient machine learning By learning to construct a system that can learn from data, readers can increase their utility across industries. Machine learning sits at the core of deep dive data analysis and visualization, which is increasingly in demand as companies discover the goldmine hiding in their existing data. For the tech professional involved in data science, Machine Learning: Hands-On for Developers and Technical Professionals provides the skills and techniques required to dig deeper.

Solve all your Spring 5 problems using complete and real-world code examples. When you start a new project, you'll be able to copy the code and configuration files from this book, and then modify them for your needs. This can save you a great deal of work over creating a project from scratch. The recipes in Spring 5 Recipes cover Spring fundamentals such as Spring IoC container, Spring AOP/ AspectJ, and more. Other recipes include Spring enterprise solutions for topics such as Spring Java EE integration, Spring Integration, Spring Batch, Spring Remoting, messaging, transactions, and working with big data and the cloud using Hadoop and MongoDB. Finally, Spring web recipes cover Spring MVC, other dynamic scripting, integration with the popular Grails Framework (and Groovy), REST/web services, and more. You'll also see recipes on new topics such as Spring Framework 5, reactive Spring, Spring 5 microservices, the functional web framework and much more. This book builds upon the best-selling success of the previous editions and focuses on the latest Spring Framework features for building enterprise Java applications. What You'll Learn Get re-usable code recipes and snippets for core Spring, annotations and other development tools Access Spring MVC for web development Work with Spring REST and microservices for web services development and integration into your enterprise Java applications Use Spring Batch, NoSQL and big data for building and integrating various cloud computing services and resources Integrate Java Enterprise Edition and other Java APIs for use in Spring Use Grails code and much more Who This Book Is For Experienced Java and Spring programmers.

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

Learn various design patterns and best practices in Spring 5 and use them to solve common design problems. About This Book Explore best practices for designing an application Manage your code easily with Spring's Dependency Injection pattern Understand the benefits that the right design patterns can offer your toolkit Who This Book Is For This book is for developers who would like to use design patterns to address common problems while designing an app using the Spring Framework and Reactive Programming approach. A basic knowledge of the Spring Framework and Java is assumed. What You Will Learn Develop applications using dependency injection patterns Learn best practices to design enterprise applications Explore Aspect-Oriented Programming relating to transactions, security, and caching. Build web applications using traditional Spring MVC patterns Learn to configure Spring using XML, annotations, and Java. Implement caching to improve application performance. Understand concurrency and handle multiple connections inside a web server. Utilizing Reactive Programming Pattern to build Reactive web applications. In Detail Design patterns help speed up the development process by offering well tested and proven solutions to common problems. These patterns coupled with the Spring framework offer tremendous improvements in the development process. The book begins with an overview of Spring Framework 5.0 and design patterns. You will understand the Dependency Injection pattern, which is the main principle behind the decoupling process that Spring performs, thus making it easier to manage your code. You will learn how GoF patterns can be used in Application Design. You will then learn to use Proxy patterns in Aspect Oriented Programming and remoting. Moving on, you will understand the JDBC template patterns and their use in abstracting database access. Then, you will be introduced to MVC patterns to build Reactive web applications. Finally, you will move on to more advanced topics such as Reactive streams and Concurrency. At the end of this book, you will be well equipped to develop efficient enterprise applications using Spring 5 with common design patterns Style and approach The book takes a pragmatic approach, showing various design patterns and best-practice considerations, including the Reactive programming approach with the Spring 5 Framework and ways to solve common development and design problems for enterprise applications.

This book is a step-by-step guide for readers who would like to learn how to build complete enterprise search solutions, with ample real-world examples and case studies. If you are a developer, designer, or architect who would like to build enterprise search solutions for your customers or organization, but have no prior knowledge of Apache Solr/Lucene technologies, this is the book for you.

Looking to study up for the new J2EE 1.5 Sun Certified Web Component Developer (SCWCD) exam? This book will get you way up to speed on the technology you'll know it so

well, in fact, that you can pass the brand new J2EE 1.5 exam. If that's what you want to do, that is. Maybe you don't care about the exam, but need to use servlets and JSPs in your next project. You're working on a deadline. You're over the legal limit for caffeine. You can't waste your time with a book that makes sense only AFTER you're an expert (or worse, one that puts you to sleep). Learn how to write servlets and JSPs, what makes a web container tick (and what ticks it off), how to use JSP's Expression Language (EL for short), and how to write deployment descriptors for your web applications. Master the c:out tag, and get a handle on exactly what's changed since the older J2EE 1.4 exam. You don't just pass the new J2EE 1.5 SCWCD exam, you'll understand this stuff and put it to work immediately. Head First Servlets and JSP doesn't just give you a bunch of facts to memorize; it drives knowledge straight into your brain. You'll interact with servlets and JSPs in ways that help you learn quickly and deeply. And when you're through with the book, you can take a brand-new mock exam, created specifically to simulate the real test-taking experience.

The book is written in such a way that learners without any background in programming are able to follow and understand it entirely. It discusses the concepts of Java in a simple and straightforward language with a clear cut explanation, without beating around the bush. On reading the book, readers are able to write simple programs on their own, as this is the first requirement to become a Java Programmer. The book provides ample solved programs which could be used by the students not only in their examinations but also to remove the fear of programming from their minds. After reading the book, the students gain the confidence to apply for a software development company, face the interview board and come out successful. The book covers sample interview questions which were asked in various interviews. It helps students to prepare for their future careers.

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

Summary Solr in Action is a comprehensive guide to implementing scalable search using Apache Solr. This clearly written book walks you through well-documented examples ranging from basic keyword searching to scaling a system for billions of documents and queries. It will give you a deep understanding of how to implement core Solr capabilities. About the Book Whether you're handling big (or small) data, managing documents, or building a website, it is important to be able to quickly search through your content and discover meaning in it. Apache Solr is your tool: a ready-to-deploy, Lucene-based, open source, full-text search engine. Solr can scale across many servers to enable real-time queries and data analytics across billions of documents. Solr in Action teaches you to implement scalable search using Apache Solr. This easy-to-read guide balances conceptual discussions with practical examples to show you how to implement all of Solr's core capabilities. You'll master topics like text analysis, faceted search, hit highlighting, result grouping, query suggestions, multilingual search, advanced geospatial and data operations, and relevancy tuning. This book assumes basic knowledge of Java and standard database technology. No prior knowledge of Solr or Lucene is required. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. What's Inside How to scale Solr for big data Rich real-world examples Solr as a NoSQL data store Advanced multilingual, data, and relevancy tricks Coverage of versions through Solr 4.7 About the Authors Trey Grainger is a director of engineering at CareerBuilder. Timothy Potter is a senior member of the engineering team at LucidWorks. The authors work on the scalability and reliability of Solr, as well as on recommendation engine and big data analytics technologies. Table of Contents PART 1 MEET SOLR Introduction to Solr Getting to know Solr Key Solr concepts Configuring Solr Indexing Text analysis PART 2 CORE SOLR CAPABILITIES Performing queries and handling results Faceted search Hit highlighting Query suggestions Result grouping/field collapsing Taking Solr to production PART 3 TAKING SOLR TO THE NEXT LEVEL SolrCloud Multilingual search Complex query operations Mastering relevancy

"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

A tutorial introducing Java basics covers programming principles, integrating applets with Web applications, and using threads, arrays, and sockets.

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 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 When Lucene first hit the scene five years ago, it was nothing short of amazing. By using this open-source, highly scalable, super-fast search engine, developers could integrate search into

applications quickly and efficiently. A lot has changed since then—search has grown from a "nice-to-have" feature into an indispensable part of most enterprise applications. Lucene now powers search in diverse companies including Akamai, Netflix, LinkedIn, Technorati, HotJobs, Epiphany, FedEx, Mayo Clinic, MIT, New Scientist Magazine, and many others. Some things remain the same, though. Lucene still delivers high-performance search features in a disarmingly easy-to-use API. Due to its vibrant and diverse open-source community of developers and users, Lucene is relentlessly improving, with evolutions to APIs, significant new features such as payloads, and a huge increase (as much as 8x) in indexing speed with Lucene 2.3. And with clear writing, reusable examples, and unmatched advice on best practices, *Lucene in Action, Second Edition* is still the definitive guide to developing with Lucene. 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.

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

Learn advanced analytical techniques and leverage existing tool kits to make your analytic applications more powerful, precise, and efficient. This book provides the right combination of architecture, design, and implementation information to create analytical systems that go beyond the basics of classification, clustering, and recommendation. *Pro Hadoop Data Analytics* emphasizes best practices to ensure coherent, efficient development. A complete example system will be developed using standard third-party components that consist of the tool kits, libraries, visualization and reporting code, as well as support glue to provide a working and extensible end-to-end system. The book also highlights the importance of end-to-end, flexible, configurable, high-performance data pipeline systems with analytical components as well as appropriate visualization results. You'll discover the importance of mix-and-match or hybrid systems, using different analytical components in one application. This hybrid approach will be prominent in the examples. What You'll Learn Build big data analytic systems with the Hadoop ecosystem Use libraries, tool kits, and algorithms to make development easier and more effective Apply metrics to measure performance and efficiency of components and systems Connect to standard relational databases, noSQL data sources, and more Follow case studies with example components to create your own systems Who This Book Is For Software engineers, architects, and data scientists with an interest in the design and implementation of big data analytical systems using Hadoop, the Hadoop ecosystem, and other associated technologies.

Servlet and JavaServer Pages (JSP) are the underlying technologies for developing web applications in Java. They are essential for any programmer to master in order to effectively use frameworks such as JavaServer Faces, Struts 2 or Spring MVC. Covering Servlet 3.1 and JSP 2.3, this book explains the important programming concepts and design models in Java web development as well as related technologies and new features in the latest versions of Servlet and JSP. With comprehensive coverage and a lot of examples, this book is a guide to building real-world applications.

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

Learn how to build scalable, resilient, and effective applications in Java that suit your software requirements. Key Features Explore advanced technologies that Java 11 delivers such as web programming and parallel computing Discover modern programming paradigms such as microservices, cloud computing and enterprise structures Build highly responsive applications with this practical introduction to Reactive programming Book Description Java is one of the most commonly used software languages by programmers and developers. In this book, you'll learn the new features of Java 11 quickly and experience a simple and powerful approach to software development. You'll see how to use the Java runtime tools, understand the Java environment, and create a simple namesorting Java application. Further on, you'll learn about advanced technologies that Java delivers, such as web programming and parallel computing, and will develop a mastermind game. Moving on, we provide more simple examples, to build a foundation before diving into some complex data structure problems that will solidify your Java 11 skills. With a special focus on the features of new projects: Project Valhalla, Project Panama, Project Amber, and Project Loom, this book will help you get employed as a top-notch Java developer. By the end of the book, you'll have a firm foundation to continue your journey toward becoming a professional Java developer. What you will learn Compile, package, and run a program using a build management tool Get to know the principles of

test-driven development Separate the wiring of multiple modules from application logic Use Java annotations for configuration Master the scripting API built into the Java language Understand static versus dynamic implementation of code Who this book is for This book is for anyone who wants to learn the Java programming language. No programming experience required. If you have prior experience, it will help you through the book more easily.

A catalog of solutions to commonly occurring design problems, presenting 23 patterns that allow designers to create flexible and reusable designs for object-oriented software. Describes the circumstances in which each pattern is applicable, and discusses the consequences and trade-offs of using the pattern within a larger design. Patterns are compiled from real systems, and include code for implementation in object-oriented programming languages like C++ and Smalltalk. Includes a bibliography. Annotation copyright by Book News, Inc., Portland, OR

Enterprise Integration Patterns provides an invaluable catalog of sixty-five patterns, with real-world solutions that demonstrate the formidable of messaging and help you to design effective messaging solutions for your enterprise. The authors also include examples covering a variety of different integration technologies, such as JMS, MSMQ, TIBCO ActiveEnterprise, Microsoft BizTalk, SOAP, and XSL. A case study describing a bond trading system illustrates the patterns in practice, and the book offers a look at emerging standards, as well as insights into what the future of enterprise integration might hold. This book provides a consistent vocabulary and visual notation framework to describe large-scale integration solutions across many technologies. It also explores in detail the advantages and limitations of asynchronous messaging architectures. The authors present practical advice on designing code that connects an application to a messaging system, and provide extensive information to help you determine when to send a message, how to route it to the proper destination, and how to monitor the health of a messaging system. If you want to know how to manage, monitor, and maintain a messaging system once it is in use, get this book.

Apache Spark is a fast, scalable, and flexible open source distributed processing engine for big data systems and is one of the most active open source big data projects to date. In just 24 lessons of one hour or less, Sams Teach Yourself Apache Spark in 24 Hours helps you build practical Big Data solutions that leverage Spark's amazing speed, scalability, simplicity, and versatility. This book's straightforward, step-by-step approach shows you how to deploy, program, optimize, manage, integrate, and extend Spark—now, and for years to come. You'll discover how to create powerful solutions encompassing cloud computing, real-time stream processing, machine learning, and more. Every lesson builds on what you've already learned, giving you a rock-solid foundation for real-world success. Whether you are a data analyst, data engineer, data scientist, or data steward, learning Spark will help you to advance your career or embark on a new career in the booming area of Big Data. Learn how to

- Discover what Apache Spark does and how it fits into the Big Data landscape
- Deploy and run Spark locally or in the cloud
- Interact with Spark from the shell
- Make the most of the Spark Cluster Architecture
- Develop Spark applications with Scala and functional Python
- Program with the Spark API, including transformations and actions
- Apply practical data engineering/analysis approaches designed for Spark
- Use Resilient Distributed Datasets (RDDs) for caching, persistence, and output
- Optimize Spark solution performance
- Use Spark with SQL (via Spark SQL) and with NoSQL (via Cassandra)
- Leverage cutting-edge functional programming techniques
- Extend Spark with streaming, R, and Sparkling Water
- Start building Spark-based machine learning and graph-processing applications
- Explore advanced messaging technologies, including Kafka
- Preview and prepare for Spark's next generation of innovations

Instructions walk you through common questions, issues, and tasks; Q-and-As, Quizzes, and Exercises build and test your knowledge; "Did You Know?" tips offer insider advice and shortcuts; and "Watch Out!" alerts help you avoid pitfalls. By the time you're finished, you'll be comfortable using Apache Spark to solve a wide spectrum of Big Data problems.

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

[Copyright: 5fd679d0c7d992175a901271bf65a010](#)