Apache Solr A Practical Approach To Enterprise Search

When people should go to the ebook stores, search commencement by shop, shelf by shelf, it is truly problematic. This is why we provide the book compilations in this website. It will completely ease you to see guide **Apache Solr A Practical Approach To Enterprise Search** as you such as.

By searching the title, publisher, or authors of guide you truly want, you can discover them rapidly. In the house, workplace, or perhaps in your method can be all best area within net connections. If you point to download and install the Apache Solr A Practical Approach To Enterprise Search , it is very easy then, before currently we extend the associate to purchase and make bargains to download and install Apache Solr A Practical Approach To Enterprise Search consequently simple!

Practical Cassandra - Russell Bradberry 2013
”Eric and Russell were early adopters of Cassandra at SimpleReach. In Practical Cassandra, you benefit from their experience in the trenches administering Cassandra, developing against it, and building one of the first CQL drivers. If you are deploying Cassandra soon, or you inherited a Cassandra cluster to tend, spend some time with the deployment, performance tuning, and maintenance chapters... If you are new to Cassandra, I highly recommend the chapters on data modeling and CQL.” –From the Foreword by Jonathon Ellis, Apache Cassandra Chair Build and Deploy Massively Scalable, Super-fast Data Management Applications with Apache Cassandra Practical Cassandra is the first hands-on developer's guide to building Cassandra systems and applications that deliver breakthrough speed, scalability, reliability, and performance. Fully up to date, it reflects the latest versions of Cassandra–including Cassandra Query Language (CQL), which dramatically lowers the learning curve for Cassandra developers. Pioneering Cassandra developers and Datastax MVPs Russell Bradberry and Eric Lubow walk you through every step of building a real production application that can store enormous amounts of structured, semi-structured, and unstructured data. Drawing on their exceptional expertise, Bradberry and Lubow share practical insights into issues ranging from querying to deployment, management, maintenance, monitoring, and troubleshooting. The authors cover key issues, from architecture to migration, and guide you through crucial decisions about configuration and data modeling. They provide tested sample code, detailed explanations of how Cassandra works ”under the covers,” and new case studies from three cutting-edge users: Ooyala, Hailo, and eBay. Coverage includes Understanding Cassandra's approach, architecture, key concepts, and primary use cases– and why it's so blazingly fast Getting Cassandra up and running on single nodes and large clusters Applying the new design patterns, philosophies, and features that make Cassandra such a powerful data store Leveraging CQL to simplify your transition from SQL-based RDBMSes Deploying and provisioning through the cloud or on bare-metal hardware Choosing the right configuration options for each type of workload Tweaking Cassandra to get maximum performance from your hardware, OS, and JVM Mastering Cassandra's essential tools for maintenance and monitoring Efficiently solving the most common problems with Cassandra deployment, operation, and application development
Gradle Beyond the Basics - Tim Berglund 2013-07-16
If you’re familiar with Gradle’s basics elements—possibly through the author’s previous O’Reilly book, Building and Testing with Gradle—this

more advanced guide provides the recipes, techniques, and syntax to help you master this build automation tool. With clear, concise explanations and lots of ready-to-use code examples, you'll explore four discrete areas of Gradle functionality: file operations, custom Gradle plugins, build lifecycle hooks, and dependency management. Learn how to use Gradle's rich set of APIs and Groovy-based Domain Specific Language to customize build software that actually conforms to your product. By using the techniques in this book, you'll be able to write domain-specific builds that support every other line of code your team creates. Examine Gradle's file API, including copy tasks, pattern matching, content filtering, and the FileCollection interface Understand the process for building and packaging a custom Gradle plug-in Manage build complexity with hook methods and Gradle's rule feature Learn how Gradle handles dependency management natively and through customization Explore Gradle's core plug-ins as well as key examples from the Gradle community

Solr in Action - Timothy Potter 2014-03-25

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 Solr High Performance - Surendra Mohan 2014-03-25

This book is an easy-to-follow guide, full of hands-on, real-world examples. Each topic is explained and demonstrated in a specific and user-friendly flow, from search optimization using Solr to Deployment of Zookeeper applications. This book is ideal for Apache Solr developers and want to learn different techniques to optimize Solr performance with utmost efficiency, along with effectively troubleshooting the problems that usually occur while trying to boost performance. Familiarity with search servers and database querying is expected.

Apache Pulsar in Action - David Kjerrumgaard 2021-12-28

Deliver lightning fast and reliable messaging for your distributed applications with the flexible and resilient Apache Pulsar platform. In Apache Pulsar in Action you will learn how to: Publish from Apache Pulsar into third-party data repositories and platforms Design and develop Apache Pulsar functions Perform interactive SQL queries against data stored in Apache Pulsar Apache Pulsar in Action is a comprehensive and practical guide to building high-traffic applications with Pulsar. You'll learn to use this mature and battle-tested platform to deliver

extreme levels of speed and durability to your messaging. Apache Pulsar committer David Kjerrumgaard teaches you to apply Pulsar's seamless scalability through hands-on case studies, including IOT analytics applications and a microservices app based on Pulsar functions. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Reliable server-to-server messaging is the heart of a distributed application. Apache Pulsar is a flexible real-time messaging platform built to run on Kubernetes and deliver the scalability and resilience required for cloud-based systems. Pulsar supports both streaming and message queuing, and unlike other solutions, it can communicate over multiple protocols including MQTT, AMQP, and Kafka's binary protocol. About the book Apache Pulsar in Action teaches you to build scalable streaming messaging systems using Pulsar. You'll start with a rapid introduction to enterprise messaging and discover the unique benefits of Pulsar. Following crystal-clear explanations and engaging examples, you'll use the Pulsar Functions framework to develop a microservices-based application. Real-world case studies illustrate how to implement the most important messaging design patterns. What's inside Publish from Pulsar into third-party data repositories and platforms Design and develop Apache Pulsar functions Create an event-driven food delivery application About the reader Written for experienced Java developers. No prior knowledge of Pulsar required. About the author David Kjerrumgaard is a committer on the Apache Pulsar project. He currently serves as a Developer Advocate for StreamNative, where he develops Pulsar best practices and solutions. Table of Contents PART 1 GETTING STARTED WITH APACHE PULSAR 1 Introduction to Apache Pulsar 2 Pulsar concepts and architecture 3 Interacting with Pulsar PART 2 APACHE PULSAR DEVELOPMENT ESSENTIALS 4 Pulsar functions 5 Pulsar IO connectors 6 Pulsar security 7 Schema registry PART 3 HANDS-ON APPLICATION DEVELOPMENT WITH APACHE PULSAR 8 Pulsar Functions patterns 9 Resiliency patterns 10 Data access 11 Machine learning in Pulsar 12 Edge analytics

Mastering Apache Solr - Mr. Mathieu Nayrolles 2014-05-15

Topic: In the open source, full-text search community, a leader emerges – Apache Solr. Apache Solr enables you to index and access documents orders of magnitude faster than classical databases and thereby provides a first-class search experience to your end users. Brief Description: Mastering Apache Solr is a practical, hands-on guide containing crisp, relevant, systematically arranged, and progressive chapters. These chapters contain a wealth of information presented in a direct and easy-to-understand manner. This book covers key technical concepts, highlighting Solr's supremacy over classical databases in full-text search, which will help you accelerate your progress in the Solr world. Detailed Description: Mastering Apache Solr starts with an introduction to Apache Solr, its underlying technologies, the main differences between the classical database engines, and gradually moves to more advance topics like boosting performance. In this book, we will look under the hood of a large number of topics and discuss answers to pertinent questions like why denormalize data, how to import classical databases' data inside Apache Solr, how to serve Solr through five different web servers, how to optimize them to serve Solr even faster. An important and major topic covered in this book is Solr's querying mechanism, which will prove to be a strong ally in our journey through this book. We then look at boosting performance and deploying Solr using several servlet servers. Finally, we cover how to communicate with Solr using different programming languages, before deploying it in a cloud-based environment. Who this book is for: Mastering Apache Solr has been written for developers, programmers, and data specialists who want to take a leap towards the future of full-text storage and search and offer a world-class experience to their users. The reader is expected to have a working knowledge of traditional databases, Linux-based operating systems, and XML configuration files. Style and Approach: Mastering Apache Solr is written lucidly and has a dynamically simple approach. From the first page to the last, the book remains practical and focuses on the most important topics used in the world of Apache Solr without neglecting important theoretical fundamentals that help you build a strong foundation. Conclusion: Mastering Apache Solr will empower you to provide a world-

class search experience to your end users through the discovery of the powerful mechanisms presented in this book.

Apache Solr - Dikshant Shahi 2015-12-19

Build an enterprise search engine using Apache Solr: index and search documents; ingest data from varied sources; apply various text processing techniques; utilize different search capabilities; and customize Solr to retrieve the desired results. Apache Solr: A Practical Approach to Enterprise Search explains each essential concept-backed by practical and industry examples--to help you attain expert-level knowledge. The book, which assumes a basic knowledge of Java, starts with an introduction to Solr, followed by steps to setting it up, indexing your first set of documents, and searching them. It then introduces you to information retrieval and its implementation in Apache Solr; this will help you understand your search problem, decide the approach to build an effective solution, and use various metrics to evaluate the results. The book next covers the schema design and techniques to build a text analysis chain for cleansing, normalizing and enriching your documents and addressing different types of search queries. It describes various popular matching techniques which are generally applied to improve the precision and recall of searches. You will learn the end-to-end process of data ingestion from varied sources, metadata extraction, pre-processing and transformation of content, various search components, query parsers and other advanced search capabilities. After covering out-of-the-box features, Solr expert Dikshant Shahi dives into ways you can customize Solr for your business and its specific requirements, along with ways to plug in your own components. Most important, you will learn about implementations for Solr scoring, factors affecting the document score, and tuning the score for the application at hand. The book explains why textual scoring is not sufficient for practical ranking of documents and ways to integrate real-world factors for contributing to the document ranking. You'll see how to influence user experience by providing suggestions and recommendations. You'll also see integration of Solr with important related technologies such as OpenNLP and Tika. Additionally, you will learn about scaling Solr using SolrCloud. This book concludes with coverage of semantic search capabilities, which is crucial for taking the search experience to the next level. By the end of Apache Solr, you will be proficient in designing and developing your search engine.

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

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

predictive modeling Study real-world big data case studies from innovative companies, including Navistar, Cerner, British Telecom, Shopzilla, Thomson Reuters, and Mastercard Who This Book Is For BI and big data warehouse professionals interested in gaining practical and real-world insight into next-generation big data processing and analytics using Apache Kudu, Impala, and Spark; and those who want to learn more about other advanced enterprise topics

Practical Hadoop Ecosystem - Deepak Vohra 2016-09-30

Learn how to use the Apache Hadoop projects, including MapReduce, HDFS, Apache Hive, Apache HBase, Apache Kafka, Apache Mahout, and Apache Solr. From setting up the environment to running sample applications each chapter in this book is a practical tutorial on using an Apache Hadoop ecosystem project. While several books on Apache Hadoop are available, most are based on the main projects, MapReduce and HDFS, and none discusses the other Apache Hadoop ecosystem projects and how they all work together as a cohesive big data development platform. What You Will Learn: Set up the environment in Linux for Hadoop projects using Cloudera Hadoop Distribution CDH 5 Run a MapReduce job Store data with Apache Hive, and Apache HBase Index data in HDFS with Apache Solr Develop a Kafka messaging system Stream Logs to HDFS with Apache Flume Transfer data from MySQL database to Hive, HDFS, and HBase with Sqoop Create a Hive table over Apache Solr Develop a Mahout User Recommender System Who This Book Is For: Apache Hadoop developers. Pre-requisite knowledge of Linux and some knowledge of Hadoop is required.

Apache Spark for the Enterprise: Setting the Business Free - Oliver Draese 2016-02-09

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 SystemsTM 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® RedpaperTM publication explores the enterprise analytics market, use of Apache Spark on IBM z SystemsTM 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.

Data Intensive Computing Applications for Big Data - M. Mittal 2018-01-31

The book ‘Data Intensive Computing Applications for Big Data’ discusses the technical concepts of big data, data intensive computing through machine learning, soft computing and parallel computing paradigms. It brings together researchers to report their latest results or progress in the development of the above mentioned areas. Since there are few books on this specific subject, the editors aim to provide a common platform for researchers working in this area to exhibit their novel findings. The book is intended as a reference work for advanced undergraduates and graduate students, as well as multidisciplinary, interdisciplinary and transdisciplinary research workers and scientists on the subjects of big data and cloud/parallel and distributed computing, and explains didactically many of the core concepts of these approaches for practical applications. It is organized into 24 chapters providing a comprehensive overview of big data analysis using parallel computing and addresses the complete data science workflow in the cloud, as well as dealing with privacy issues and the challenges faced in a data-intensive cloud computing environment. The book explores both

fundamental and high-level concepts, and will serve as a manual for those in the industry, while also helping beginners to understand the basic and advanced aspects of big data and cloud computing.

Hadoop For Dummies - Dirk deRoos 2014-04-14

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, scaleable 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.

Solr Cookbook - Third Edition - Rafal Kuc 2015-01-23

This book is for intermediate Solr Developers who are willing to learn and implement Pro-level practices, techniques, and solutions. This edition will specifically appeal to developers who wish to quickly get to grips with the changes and new features of Apache Solr 5.

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

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

Taming Text - Grant Ingersoll 2012-12-20

Summary Taming Text, winner of the 2013 Jolt Awards for Productivity, is a hands-on, example-driven guide to working with unstructured text in the context of real-world applications. This book explores how to automatically organize text using approaches such as full-text search, proper name recognition, clustering, tagging, information extraction, and summarization. The book guides you through examples illustrating each of these topics, as well as the foundations upon which they are built. About this Book There is so much text in our lives, we are practically drowningin it. Fortunately, there are innovative tools and techniquesfor managing unstructured information that can throw thesmart developer a much-needed lifeline. You'll find them in thisbook. Taming Text is a practical, example-driven guide to working withtext in real applications. This book introduces you to useful techniques like full-text search, proper name recognition,clustering, tagging, information extraction, and summarization.You'll explore real use cases as you systematically absorb thefoundations upon which they are built.Written in a clear and concise style, this book avoids jargon, explainingthe subject in terms you can

understand without a backgroundin statistics or natural language processing. Examples arein Java, but the concepts can be applied in any language. Written for Java developers, the book requires no prior knowledge of GWT. 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. Winner of 2013 Jolt Awards: The Best Books—one of five notable books every serious programmer should read. What's Inside When to use text-taming techniques Important open-source libraries like Solr and Mahout How to build text-processing applications About the Authors Grant Ingersoll is an engineer, speaker, and trainer, a Lucenecommitter, and a cofounder of the Mahout machine-learning project. Thomas Morton is the primary developer of OpenNLP and Maximum Entropy. Drew Farris is a technology consultant, software developer, and contributor to Mahout,Lucene, and Solr. "Takes the mystery out of verycomplex processes."—From the Foreword by Liz Liddy, Dean, iSchool, Syracuse University Table of Contents Getting started taming text Foundations of taming text Searching Fuzzy string matching Identifying people, places, and things Clustering text Classification, categorization, and tagging Building an example question answering system Untamed text: exploring the next frontier

Open Source Intelligence Methods and Tools - Nihad A. Hassan 2018-06-30

Apply Open Source Intelligence (OSINT) techniques, methods, and tools to acquire information from publicly available online sources to support your intelligence analysis. Use the harvested data in different scenarios such as financial, crime, and terrorism investigations as well as performing business competition analysis and acquiring intelligence about individuals and other entities. This book will also improve your skills to acquire information online from both the regular Internet as well as the hidden web through its two sub-layers: the deep web and the dark web. The author includes many OSINT resources that can be used by intelligence agencies as well as by enterprises to monitor trends on a global level, identify risks, and gather competitor intelligence so more effective decisions can be made. You will discover techniques, methods, and tools that are equally used by hackers and penetration testers to gather intelligence about a specific target online. And you will be aware of how OSINT resources can be used in conducting social engineering attacks. Open Source Intelligence Methods and Tools takes a practical approach and lists hundreds of OSINT resources that can be used to gather intelligence from online public sources. The book also covers how to anonymize your digital identity online so you can conduct your searching activities without revealing your identity. What You'll Learn Identify intelligence needs and leverage a broad range of tools and sources to improve data collection, analysis, and decision making in your organization Use OSINT resources to protect individuals and enterprises by discovering data that is online, exposed, and sensitive and hide the data before it is revealed by outside attackers Gather corporate intelligence about business competitors and predict future market directions Conduct advanced searches to gather intelligence from social media sites such as Facebook and Twitter Understand the different layers that make up the Internet and how to search within the invisible web which contains both the deep and the dark webs Who This Book Is For Penetration testers, digital forensics investigators, intelligence services, military, law enforcement, UN agencies, and for-profit/non-profit enterprises

Drupal 7 Media - Liran Tal 2013-01-01

This is a practical, hands-on guide packed with examples that will help you build rich Drupal 7 media web applications.If you are a Drupal site builder and you wish to spice up your web applications with rich media content, then this book is for you. A basic understanding of HTML, JavaScript, and basic PHP module development in Drupal would be helpful, but is not necessary.

Intelligent Systems in Big Data, Semantic Web and Machine Learning - Noreddine Gherabi 2021-05-28

This book describes important methodologies, tools and techniques from the fields of artificial intelligence, basically those which are based on relevant conceptual and formal development. The coverage is wide, ranging from machine learning to the use of data on the Semantic Web,

with many new topics. The contributions are concerned with machine learning, big data, data processing in medicine, similarity processing in ontologies, semantic image analysis, as well as many applications including the use of machine leaning techniques for cloud security, artificial intelligence techniques for detecting COVID-19, the Internet of things, etc. The book is meant to be a very important and useful source of information for researchers and doctoral students in data analysis, Semantic Web, big data, machine learning, computer engineering and related disciplines, as well as for postgraduate students who want to integrate the doctoral cycle.

Introduction to Information Retrieval - Christopher D. Manning 2008-07-07

Class-tested and coherent, this textbook teaches classical and web information retrieval, including web search and the related areas of text classification and text clustering from basic concepts. It gives an up-to-date treatment of all aspects of the design and implementation of systems for gathering, indexing, and searching documents; methods for evaluating systems; and an introduction to the use of machine learning methods on text collections. All the important ideas are explained using examples and figures, making it perfect for introductory courses in information retrieval for advanced undergraduates and graduate students in computer science. Based on feedback from extensive classroom experience, the book has been carefully structured in order to make teaching more natural and effective. Slides and additional exercises (with solutions for lecturers) are also available through the book's supporting website to help course instructors prepare their lectures.

Lucene in Action - Otis Gospodnetic 2010-07-08

When Lucene first hit the scene five years ago, it was nothing short ofamazing. 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" featureinto an indispensable part of most enterprise applications. Lucene now powerssearch in diverse companies including Akamai, Netflix, LinkedIn,Technorati, HotJobs, Epiphany, FedEx, Mayo Clinic, MIT, New ScientistMagazine, and many others. Some things remain the same, though. Lucene still delivers high-performancesearch features in a disarmingly easy-to-use API. Due to its vibrant and diverseopen-source community of developers and users, Lucene is relentlessly improving,with evolutions to APIs, significant new features such as payloads, and ahuge increase (as much as 8x) in indexing speed with Lucene 2.3. And with clear writing, reusable examples, and unmatched advice on bestpractices, Lucene in Action, Second Edition is still the definitive guide todeveloping 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.

Mastering Apache Solr 7. X - Sandeep Nair 2018-02-19

Accelerate your enterprise search engine and bring relevancy in your search analytics Key Features A practical guide in building expertise with Indexing, Faceting, Clustering and Pagination Master the management and administration of Enterprise Search Applications and services seamlessly Handle multiple data inputs such as JSON, xml, pdf, doc, xls,ppt, csv and much more. Book Description Apache Solr is the only standalone enterprise search server with a REST-like application interface. providing highly scalable, distributed search and index replication for many of the world's largest internet sites. To begin with, you would be introduced to how you perform full text search, multiple filter search, perform dynamic clustering and so on helping you to brush up the basics of Apache Solr. You will also explore the new features and advanced options released in Apache Solr 7.x which will get you numerous performance aspects and making data investigation simpler, easier and powerful. You will learn to build complex queries, extensive filters and how are they compiled in your system to bring relevance in your search tools. You will learn to carry out Solr scoring, elements affecting the document score and how you can optimize or tune the score for the application at hand. You will learn to extract features of documents, writing complex queries in re-ranking the documents. You will also learn advanced options helping you to know what content is

indexed and how the extracted content is indexed. Throughout the book, you would go through complex problems with solutions along with varied approaches to tackle your business needs. By the end of this book, you will gain advanced proficiency to build out-of-box smart search solutions for your enterprise demands. What you will learn Design schema using schema API to access data in the database Advance querying and fine-tuning techniques for better performance Get to grips with indexing using Client API Set up a fault tolerant and highly available server with newer distributed capabilities, SolrCloud Explore Apache Tika to upload data with Solr Cell Understand different data operations that can be done while indexing Master advanced querying through Velocity Search UI, faceting and Query Re-ranking, pagination and spatial search Learn to use JavaScript, Python, SolrJ and Ruby for interacting with Solr Who this book is for The book would rightly appeal to developers, software engineers, data engineers and database architects who are building or seeking to build enterprise-wide effective search engines for business intelligence. Prior experience of Apache Solr or Java programming is must to take the best of this book.

Elasticsearch: The Definitive Guide - Clinton Gormley 2015-01-23
Whether you need full-text search or real-time analytics of structured data—or both—the Elasticsearch distributed search engine is an ideal way to put your data to work. This practical guide not only shows you how to search, analyze, and explore data with Elasticsearch, but also helps you deal with the complexities of human language, geolocation, and relationships. If you're a newcomer to both search and distributed systems, you'll quickly learn how to integrate Elasticsearch into your application. More experienced users will pick up lots of advanced techniques. Throughout the book, you'll follow a problem-based approach to learn why, when, and how to use Elasticsearch features. Understand how Elasticsearch interprets data in your documents Index and query your data to take advantage of search concepts such as relevance and word proximity Handle human language through the effective use of analyzers and queries Summarize and group data to show overall trends, with aggregations and analytics Use geo-points and geo-shapes—Elasticsearch's approaches to geolocation Model your data to take advantage of Elasticsearch's horizontal scalability Learn how to configure and monitor your cluster in production

Scaling Big Data with Hadoop and Solr - Hrishikesh Karambelkar 2013
This book is a step-by-step tutorial that will enable you to leverage the flexible search functionality of Apache Solr together with the Big Data power of Apache Hadoop.Scaling Big Data with Hadoop and Solr provides guidance to developers who wish to build high-speed enterprise search platforms using Hadoop and Solr. This book is primarily aimed at Java programmers who wish to extend the Hadoop platform to make it run as an enterprise search without any prior knowledge of Apache Hadoop and Solr.

Designing Software Architectures - Humberto Cervantes 2016-04-29
Designing Software Architectures will teach you how to design any software architecture in a systematic, predictable, repeatable, and cost-effective way. This book introduces a practical methodology for architecture design that any professional software engineer can use, provides structured methods supported by reusable chunks of design knowledge, and includes rich case studies that demonstrate how to use the methods. Using realistic examples, you'll master the powerful new version of the proven Attribute-Driven Design (ADD) 3.0 method and will learn how to use it to address key drivers, including quality attributes, such as modifiability, usability, and availability, along with functional requirements and architectural concerns. Drawing on their extensive experience, Humberto Cervantes and Rick Kazman guide you through crafting practical designs that support the full software life cycle, from requirements to maintenance and evolution. You'll learn how to successfully integrate design in your organizational context, and how to design systems that will be built with agile methods. Comprehensive coverage includes Understanding what architecture design involves, and where it fits in the full software development life cycle Mastering core design concepts, principles, and processes Understanding how to perform the steps of the ADD method Scaling design and analysis up or

down, including design for pre-sale processes or lightweight architecture reviews Recognizing and optimizing critical relationships between analysis and design Utilizing proven, reusable design primitives and adapting them to specific problems and contexts Solving design problems in new domains, such as cloud, mobile, or big data

Apache Solr 4 Cookbook - Rafał Kuć 2013-01-01

Over 100 practical recipes to make Apache Solr faster, more reliable and return better results.

Building Search Applications - Manu Konchady 2008

Lucene, LingPipe, and Gate are popular open source tools to build powerful search applications. Building Search Applications describes functions from Lucene that include indexing, searching, ranking, and spelling correction to build search engines. With this book you will learn to: Extract tokens from text using custom tokenizers and analyzers from Lucene, LingPipe, and Gate. Construct a search engine index with an optional backend database to manage large document collections. Explore the wide range of Lucene queries to search an index, understand the ranking algorithm for a query, and suggest spelling corrections. Find the names of people, places, and other entities in text using LingPipe and Gate. Categorize documents by topic using classifiers and build groups of self-organized documents using clustering algorithms from LingPipe. Create a Web crawler to scan the Web, Intranet, or desktop using Nutch. Track the sentiment of articles published on the Web with LingPipe.

Liferay Portal Performance Best Practices - Samir Bhatt 2013-06-10

A step-by-step tutorial on implementing Liferay- based portals to learn performance best practices.The book is good for Liferay portal developers and architects who want to learn performance best practices for implementing Liferay- based solutions. It is assumed that you have a working knowledge of the Liferay portal.

Searching the Enterprise - Udo Kruschwitz 2017-07-12

Search has become ubiquitous but that does not mean that search has been solved. Enterprise search, which is broadly speaking the use of information retrieval technology to find information within organisations, is a good example to illustrate this. It is an area that is of huge importance for businesses, yet has attracted relatively little academic interest. Searching the Enterprise explores the main issues involved in enterprise search both from a research as well as a practical point of view. It first plots the landscape of enterprise search and its links to related areas, which allows it to identify key features before surveying the field in more detail. Throughout the monograph, enterprise search is discussed as part of the wider information retrieval research field, and Web search is used as a common reference point as this is likely the search application area that the average reader is most familiar with.

Scaling Apache Solr - Hrishikesh Vijay Karambelkar 2014-07-25

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.

Apache Solr Search Patterns - Jayant Kumar 2015-04-24

This book is for developers who already know how to use Solr and are looking at procuring advanced strategies for improving their search using Solr. This book is also for people who work with analytics to generate graphs and reports using Solr. Moreover, if you are a search architect who is looking forward to scale your search using Solr, this is a must have book for you. It would be helpful if you are familiar with the Java programming language.

Data Lake for Enterprises - Tomcy John 2017-05-31

A practical guide to implementing your enterprise data lake using Lambda Architecture as the base About This Book Build a full-fledged data lake for your organization with popular big data technologies using the Lambda architecture as the base Delve into the big data technologies required to meet modern day business strategies A highly practical guide to implementing enterprise data lakes with lots of examples and real-world use-cases Who This Book Is For Java developers and architects who would like to implement a data lake for their enterprise will find this book useful. If you want to get hands-on experience with the Lambda

Architecture and big data technologies by implementing a practical solution using these technologies, this book will also help you. What You Will Learn Build an enterprise-level data lake using the relevant big data technologies Understand the core of the Lambda architecture and how to apply it in an enterprise Learn the technical details around Sqoop and its functionalities Integrate Kafka with Hadoop components to acquire enterprise data Use flume with streaming technologies for stream-based processing Understand stream- based processing with reference to Apache Spark Streaming Incorporate Hadoop components and know the advantages they provide for enterprise data lakes Build fast, streaming, and high-performance applications using ElasticSearch Make your data ingestion process consistent across various data formats with configurability Process your data to derive intelligence using machine learning algorithms In Detail The term "Data Lake" has recently emerged as a prominent term in the big data industry. Data scientists can make use of it in deriving meaningful insights that can be used by businesses to redefine or transform the way they operate. Lambda architecture is also emerging as one of the very eminent patterns in the big data landscape, as it not only helps to derive useful information from historical data but also correlates real-time data to enable business to take critical decisions. This book tries to bring these two important aspects — data lake and lambda architecture—together. This book is divided into three main sections. The first introduces you to the concept of data lakes, the importance of data lakes in enterprises, and getting you up-to-speed with the Lambda architecture. The second section delves into the principal components of building a data lake using the Lambda architecture. It introduces you to popular big data technologies such as Apache Hadoop, Spark, Sqoop, Flume, and ElasticSearch. The third section is a highly practical demonstration of putting it all together, and shows you how an enterprise data lake can be implemented, along with several real-world use-cases. It also shows you how other peripheral components can be added to the lake to make it more efficient. By the end of this book, you will be able to choose the right big data technologies using the lambda architectural patterns to build your enterprise data lake. Style and approach The book takes a pragmatic approach, showing ways to leverage big data technologies and lambda architecture to build an enterprise-level data lake.

Big Data Analytics with Java - Rajat Mehta 2017-07-31

Learn the basics of analytics on big data using Java, machine learning and other big data tools About This Book Acquire real-world set of tools for building enterprise level data science applications Surpasses the barrier of other languages in data science and learn create useful object-oriented codes Extensive use of Java compliant big data tools like apache spark, Hadoop, etc. Who This Book Is For This book is for Java developers who are looking to perform data analysis in production environment. Those who wish to implement data analysis in their Big data applications will find this book helpful. What You Will Learn Start from simple analytic tasks on big data Get into more complex tasks with predictive analytics on big data using machine learning Learn real time analytic tasks Understand the concepts with examples and case studies Prepare and refine data for analysis Create charts in order to understand the data See various real-world datasets In Detail This book covers case studies such as sentiment analysis on a tweet dataset, recommendations on a movielens dataset, customer segmentation on an ecommerce dataset, and graph analysis on actual flights dataset. This book is an end-to-end guide to implement analytics on big data with Java. Java is the de facto language for major big data environments, including Hadoop. This book will teach you how to perform analytics on big data with production-friendly Java. This book basically divided into two sections. The first part is an introduction that will help the readers get acquainted with big data environments, whereas the second part will contain a hardcore discussion on all the concepts in analytics on big data. It will take you from data analysis and data visualization to the core concepts and advantages of machine learning, real-life usage of regression and classification using Naive Bayes, a deep discussion on the concepts of clustering,and a review of simple neural networks on big data using deepLearning4j or plain Java Spark code. This book is a must-have book for Java developers who want to start learning big data analytics and

want to use it in the real world. Style and approach The approach of book is to deliver practical learning modules in manageable content. Each chapter is a self-contained unit of a concept in big data analytics. Book will step by step builds the competency in the area of big data analytics. Examples using real world case studies to give ideas of real applications and how to use the techniques mentioned. The examples and case studies will be shown using both theory and code.

Apache Solr PHP Integration - Jayant Kumar 2013-11-25

This book is full of step-by-step example-oriented tutorials which will show readers how to integrate Solr in PHP applications using the available libraries, and boost the inherent search facilities that Solr offers.If you are a developer who knows PHP and is interested in integrating search into your applications, this is the book for you. No advanced knowledge of Solr is required. Very basic knowledge of system commands and the command-line interface on both Linux and Windows is required. You should also be familiar with the concept of Web servers.

Natural Language Processing Recipes - Akshay Kulkarni 2019-01-29

Implement natural language processing applications with Python using a problem-solution approach. This book has numerous coding exercises that will help you to quickly deploy natural language processing techniques, such as text classification, parts of speech identification, topic modeling, text summarization, text generation, entity extraction, and sentiment analysis. Natural Language Processing Recipes starts by offering solutions for cleaning and preprocessing text data and ways to analyze it with advanced algorithms. You'll see practical applications of the semantic as well as syntactic analysis of text, as well as complex natural language processing approaches that involve text normalization, advanced preprocessing, POS tagging, and sentiment analysis. You will also learn various applications of machine learning and deep learning in natural language processing. By using the recipes in this book, you will have a toolbox of solutions to apply to your own projects in the real world, making your development time quicker and more efficient. What You Will LearnApply NLP techniques using Python libraries such as NLTK, TextBlob, spaCy, Stanford CoreNLP, and many more Implement the concepts of information retrieval, text summarization, sentiment analysis, and other advanced natural language processing techniques. Identify machine learning and deep learning techniques for natural language processing and natural language generation problems Who This Book Is ForData scientists who want to refresh and learn various concepts of natural language processing through coding exercises.

Designing Data-Intensive Applications - Martin Kleppmann 2017-03-16

Data is at the center of many challenges in system design today. Difficult issues need to be figured out, such as scalability, consistency, reliability, efficiency, and maintainability. In addition, we have an overwhelming variety of tools, including relational databases, NoSQL datastores, stream or batch processors, and message brokers. What are the right choices for your application? How do you make sense of all these buzzwords? In this practical and comprehensive guide, author Martin Kleppmann helps you navigate this diverse landscape by examining the pros and cons of various technologies for processing and storing data. Software keeps changing, but the fundamental principles remain the same. With this book, software engineers and architects will learn how to apply those ideas in practice, and how to make full use of data in modern applications. Peer under the hood of the systems you already use, and learn how to use and operate them more effectively Make informed decisions by identifying the strengths and weaknesses of different tools Navigate the trade-offs around consistency, scalability, fault tolerance, and complexity Understand the distributed systems research upon which modern databases are built Peek behind the scenes of major online services, and learn from their architectures

Apache Solr - Dikshant Shahi 2015-12-26

Build an enterprise search engine using Apache Solr: index and search documents; ingest data from varied sources; apply various text processing techniques; utilize different search capabilities; and customize Solr to retrieve the desired results. Apache Solr: A Practical Approach to Enterprise Search explains each essential concept-backed by practical and industry examples--to help you attain expert-level knowledge. The book, which assumes a basic knowledge of Java, starts

with an introduction to Solr, followed by steps to setting it up, indexing your first set of documents, and searching them. It then introduces you to information retrieval and its implementation in Apache Solr; this will help you understand your search problem, decide the approach to build an effective solution, and use various metrics to evaluate the results. The book next covers the schema design and techniques to build a text analysis chain for cleansing, normalizing and enriching your documents and addressing different types of search queries. It describes various popular matching techniques which are generally applied to improve the precision and recall of searches. You will learn the end-to-end process of data ingestion from varied sources, metadata extraction, pre-processing and transformation of content, various search components, query parsers and other advanced search capabilities. After covering out-of-the-box features, Solr expert Dikshant Shahi dives into ways you can customize Solr for your business and its specific requirements, along with ways to plug in your own components. Most important, you will learn about implementations for Solr scoring, factors affecting the document score, and tuning the score for the application at hand. The book explains why textual scoring is not sufficient for practical ranking of documents and ways to integrate real-world factors for contributing to the document ranking. You'll see how to influence user experience by providing suggestions and recommendations. You'll also see integration of Solr with important related technologies such as OpenNLP and Tika. Additionally, you will learn about scaling Solr using SolrCloud. This book concludes with coverage of semantic search capabilities, which is crucial for taking the search experience to the next level. By the end of Apache Solr, you will be proficient in designing and developing your search engine.

Building Modular Cloud Apps with OSGi - Paul Bakker 2013-09-09
If you’re an experienced Java developer in the enterprise, this practical, hands-on book shows you how to use OSGi to design, develop, and deploy modular cloud applications. You’ll quickly learn how to use OSGi, through concise code examples and a set of best practices derived from the authors’ experiences with real-world projects. Through the course of this book, you’ll learn to develop modern web applications with tools and techniques such as RESTful Web Services, NoSQL, provisioning, elasticity, Auto Scaling, hotfixes, and automatic failover. Code samples are available from GitHub. Work with dynamic OSGi services to create modular applications Explore the basics of OSGi bundles and modular application design Learn advanced topics, including semantic versioning, integration testing, and configuring components Understand OSGi pitfalls, anti-patterns, and features you should avoid Create a modular architecture for cloud-based web applications Discover how maintainability, extensibility, scalability, and testability are affected by modular design Get a look at various options for creating web applications with a modular approach Interact with persistent storage services, including relational databases and NoSQL Examine alternatives for deploying modular applications to the cloud

Mastering Apache Solr 7.x - Sandeep Nair 2018-02-22
Accelerate your enterprise search engine and bring relevancy in your search analytics Key Features A practical guide in building expertise with Indexing, Faceting, Clustering and Pagination Master the management and administration of Enterprise Search Applications and services seamlessly Handle multiple data inputs such as JSON, xml, pdf, doc, xls,ppt, csv and much more. Book Description Apache Solr is the only standalone enterprise search server with a REST-like application interface. providing highly scalable, distributed search and index replication for many of the world's largest internet sites. To begin with, you would be introduced to how you perform full text search, multiple filter search, perform dynamic clustering and so on helping you to brush up the basics of Apache Solr. You will also explore the new features and advanced options released in Apache Solr 7.x which will get you numerous performance aspects and making data investigation simpler, easier and powerful. You will learn to build complex queries, extensive filters and how are they compiled in your system to bring relevance in your search tools. You will learn to carry out Solr scoring, elements affecting the document score and how you can optimize or tune the score for the application at hand. You will learn to extract features of

documents, writing complex queries in re-ranking the documents. You will also learn advanced options helping you to know what content is indexed and how the extracted content is indexed. Throughout the book, you would go through complex problems with solutions along with varied approaches to tackle your business needs. By the end of this book, you will gain advanced proficiency to build out-of-box smart search solutions for your enterprise demands. What you will learn Design schema using schema API to access data in the database Advance querying and fine-tuning techniques for better performance Get to grips with indexing using Client API Set up a fault tolerant and highly available server with newer distributed capabilities, SolrCloud Explore Apache Tika to upload data with Solr Cell Understand different data operations that can be done while indexing Master advanced querying through Velocity Search UI, faceting and Query Re-ranking, pagination and spatial search Learn to use JavaScript, Python, SolrJ and Ruby for interacting with Solr Who this book is for The book would rightly appeal to developers, software engineers, data engineers and database architects who are building or seeking to build enterprise-wide effective search engines for business intelligence. Prior experience of Apache Solr or Java programming is must to take the best of this book.

Apache Solr Enterprise Search Server - Third Edition - David Smiley 2015-05-26

This book is for developers who want to learn how to get the most out of Solr in their applications, whether you are new to the field, have used Solr but don't know everything, or simply want a good reference. It would be helpful to have some familiarity with basic programming concepts, but no prior experience is required.

Alfresco 4 Enterprise Content Management Implementation - Munwar Shariff 2013-01-01

This book distils the hands-on approach of the training courses into a concise, practical book. The emphasis is on getting up and running fast and discovering the scope and power of Alfresco 4 incrementally through practical examples. Though this book is not a developer guide, various examples in the book will help developers to extend Alfresco functionality and to integrate Alfresco with external systems.This book is designed for experienced users, business owners, or system administrators who want to install and use Alfresco in their teams or businesses. Because Alfresco is free, many teams can install and experiment with its ECM features without any upfront cost, often without management approval. You need to have a degree of technical confidence, but you do not require specialist system admin or developer skills to get a basic system up and running.Though this book is not a developer guide, various examples in the book will help you to extend Alfresco functionality and to integrate Alfresco with external systems.