

Apache Solr 4 Cookbook

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

JSON is becoming the backbone for meaningful data interchange over the internet. This format is now supported by an entire ecosystem of standards, tools, and technologies for building truly elegant, useful, and efficient applications. With this hands-on guide, author and architect Tom Marrs shows you how to build enterprise-class applications and services by leveraging JSON tooling and message/document design. JSON at Work provides application architects and developers with guidelines, best practices, and use cases, along with lots of real-world examples and code samples. You'll start with a comprehensive JSON overview, explore the JSON ecosystem, and then dive into JSON's use in the enterprise. Get acquainted with JSON basics and learn how to model JSON data Learn how to use JSON with Node.js, Ruby on Rails, and Java Structure JSON documents with JSON Schema to design and test APIs Search the contents of JSON documents with JSON Search tools Convert JSON documents to other data formats with JSON Transform tools Compare JSON-based hypermedia formats, including HAL and jsonapi Leverage MongoDB to store and access JSON documents Use Apache Kafka to exchange JSON-based messages between services

Enterprise and web applications require full-fledged, "Google-quality" searchcapabilities, but such features are notoriously difficult to implement and maintain.Hibernate Search builds on the Lucene feature set and offers an easy-to-implemnet interface that integrates seamlessly with Hibernate's leadingdata persistence solution for Java applications. Hibernate Search in Action introduces both the principles of enterprise searchand the implementation details a Java developer will need to use HibernateSearch effectively. This book blends the insights of the Hibernate Search leaddeveloper with the practical techniques required to index and manipulate data,assemble and execute search queries, and create smart filters for better searchresults. Along the way, the reader masters performance-boosting concepts likeusing Hibernate Search in a clustered environment and integrating with thefeatures already in your applications. This book assumes you're a competent Java developer with some experienceusing Hibernate and 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.

Over 60 recipes to help you speed up the development of your Java web applications using the Spring Roo development tool.

Natural Language Processing Recipes

Practical Performant Programming for Humans

Big Data in Practice

Pro Docker

Elasticsearch: The Definitive Guide

Distributed Process Coordination

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 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

Summary Relevant Search demystifies relevance work. Using Elasticsearch, it teaches you how to return engaging search results to your users, helping you understand and leverage the internals of Lucene-based search engines. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Users are accustomed to and expect instant, relevant search results. To achieve this, you must master the search engine. Yet for many developers, relevance ranking is mysterious or confusing. About the Book Relevant Search demystifies the subject and shows you that a search engine is a programmable relevance framework. You'll learn how to apply Elasticsearch or Solr to your business's unique ranking problems. The book demonstrates how to program relevance and how to incorporate secondary data sources, taxonomies, text analytics, and personalization. In practice, a relevance framework requires softer skills as well, such as collaborating with stakeholders to discover the right relevance requirements for your business. By the end, you'll be able to achieve a virtuous cycle of provable, measurable relevance improvements over a search product's lifetime. What's Inside Techniques for debugging relevance? Applying search engine features to real problems? Using the user interface to guide searches? A systematic approach to relevance? A business culture focused on improving search About the Reader For developers trying to build smarter search with Elasticsearch or Solr. About the Authors Doug Turnbull is lead relevance consultant at OpenSource Connections, where he frequently speaks and blogs. John Berrymann is a data engineer at Eventbrite, where he specializes in recommendations and search. Foreword author, Trey Grainger, is a director of engineering at CareerBuilder and author of Solr in Action. Table of Contents The search relevance problem How to use the debug Debugger your first relevance problem Taming tokens Basic multifold search Term-centric search

Shaping the relevance function Providing relevance feedback Designing a relevance-focused search application The relevance-centered enterprise Semantic and personalized search About the Book AngularJS is revolutionizing design patterns and best practices to build stable and high performing apps Key Features Get to grips with the benefits and applicability of using different design patterns in Angular with the help of real-world examples Identify and prevent common problems, programming errors, and anti-patterns Packed with easy-to-follow examples that can be used to create reusable code and extensible designs Book Description This book is an insightful journey through the most valuable design patterns, and it will provide clear guidance on how to use them effectively in Angular. You will explore some of the best ways to work with Angular and how to use it to ensure the stability and performance required in today's web development world. You'll get to know some Angular best practices to improve your productivity and the code base of your application. We will take you on a journey through Angular designs for the real world, using a combination of case studies, design patterns to follow, and anti-patterns to avoid. By the end of the book, you will understand the various features of Angular, and will be able to apply well-known, industry-proven design patterns in your work. What you will learn Understand Angular design patterns and anti-patterns Implement the most useful GoF patterns for Angular Explore some of the most famous navigational patterns for Angular Get to know and implement stability patterns Explore and implement operations patterns Explore the official best practices for Angular Monitor and improve the performance of Angular applications Who this book is for If you want to increase your understanding of Angular and apply it to real-life application development, then this book is for you.

High Performance Python

Solr 1.4 Enterprise Search Server

Alfresco 4 Enterprise Content Management Implementation

SQL Server Query Performance Tuning

Practical Data Integration for the Web

Mastering Hadoop 3

Summary A developer-focused guide to writing applications using Spring Boot. You'll learn how to bypass the tedious configuration steps so that you can concentrate on your application's behavior. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology The Spring Framework simplifies enterprise Java development, but it does require lots of tedious configuration work. Spring Boot radically streamlines spinning up a Spring application. You get automatic configuration and a model with established conventions for build-time and runtime dependencies. You also get a handy command-line interface you can use to write scripts in Groovy. Developers who use Spring Boot often say that they can't imagine going back to hand configuring their applications. About the Book Spring Boot in Action is a developer-focused guide to writing applications using Spring Boot. In it, you'll learn how to bypass configuration steps so you can focus on your application's behavior. Spring expert Craig Walls uses interesting and practical examples to teach you both how to use the default settings effectively and how to override and customize Spring Boot for your unique environment. Along the way, you'll pick up insights from Craig's years of Spring development experience. What's Inside Develop Spring apps more efficiently Minimal to no configuration Runtime metrics with the Actuator Covers Spring Boot 1.3 About the Reader Written for readers familiar with the Spring Framework. About the Author Craig Walls is a software developer, author of the popular book Spring in Action, Fourth Edition, and a frequent speaker at conferences. Table of Contents Bootstrapping Spring Developing your first Spring Boot application Customizing configuration Testing with Spring Boot Getting Groovy with the Spring Boot CLI Applying Grails in Spring Boot Taking a peek inside with the Actuator Deploying Spring Boot applications APPENDIXES Spring Boot developer tools Spring Boot starters Configuration properties Spring Boot dependencies

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.

Queries not running fast enough? Wondering about the in-memory database features in 2014? Tired of phone calls from frustrated users? Grant Fritchey's book SQL Server Query Performance Tuning is the answer to your SQL Server query performance problems. The book is revised to cover the very latest in performance optimization features and techniques, especially including the newly-added, in-memory database features formerly known under the code name Project Hekaton. This book provides the tools you need to approach your queries with performance in mind. SQL Server Query Performance Tuning leads you through understanding the causes of poor performance, how to identify them, and how to fix them. You'll learn to be proactive in establishing performance baselines using tools like Performance Monitor and Extended Events. You'll learn to recognize bottlenecks and defuse them before the phone rings. You'll learn some quick solutions too, but emphasis is on designing for performance and getting it right, and upon heading off trouble before it occurs. Delight your users. Silence that ringing phone. Put the principles and lessons from SQL Server Query Performance Tuning into practice today. Covers the in-memory features from Project Hekaton Helps establish performance baselines and monitor against them Guides in troubleshooting and eliminating of bottlenecks that frustrate users

Implement natural language processing applications with Python using a problem-solution approach. This book has numerous code exercises that will help you to quickly deploy natural language processing techniques, such as text classification, 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.

With applications for Solr and Elasticsearch

A Distributed Real-Time Search and Analytics Engine

JSON at Work

Solr in Action

Server and Configuration Management for Humans

Implement the Gang of Four patterns in your apps with Angular

Apache Solr 4 CookbookPackt Publishing Ltd

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

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.

Apache Solr Search Patterns

Hibernate Search in Action

Java Cookbook

Unlocking Text Data with Machine Learning and Deep Learning using Python

Enhance Your Search with Faceted Navigation, Result Highlighting, Fuzzy Queries, Ranked Scoring, and More

Ansible for DevOps

This book is for Elasticsearch users who want to extend their knowledge and develop new skills. Prior knowledge of the Query DSL and data indexing is expected.

From lambda expressions and JavaFX 8 to new support for network programming and mobile development, Java 8 brings a wealth of changes. This cookbook helps you get up to speed right away with hundreds of hands-on recipes across a broad range of Java topics. You ' ll learn useful techniques for everything from debugging and data structures to GUI development and functional programming. Each recipe includes self-contained code solutions that you can freely use, along with a discussion of how and why they work. If you are familiar with Java basics, this cookbook will bolster your knowledge of the language in general and Java 8 ' s main APIs in particular. Recipes include: Methods for compiling, running, and debugging Manipulating, comparing, and rearranging text Regular expressions for string- and pattern-matching Handling numbers, dates, and times Structuring data with collections, arrays, and other types Object-oriented and functional programming techniques Directory and filesystem operations Working with graphics, audio, and video GUI development, including JavaFX and handlers Network programming on both client and server Database access, using JPA, Hibernate, and JDBC Processing JSON and XML for data storage Multithreading and concurrency

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 drowning in it. Fortunately, there are innovative tools and techniques for managing unstructured information that can throw thesaurus-driven guide to working with text 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 the foundations upon which they are built. Written in a clear and concise style, this book avoids jargon, explaining the subject in terms you can understand. Examples are in plain, natural language processing. Examples are in 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 Lucene committer, 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 very complex processes."—From the Foreword by Liz Uddy, 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 Unlamed text: exploring the next frontier

Apache is far and away the most widely used web server platform in the world. Both free and rock-solid, it runs more than half of the world's web sites, ranging from huge e-commerce operations to corporate intranets and smaller hobby sites, and it continues to maintain its popularity, drawing new users all the time. If you work with Apache on a regular basis, you have plenty of documentation on installing and configuring your server, but where do you go for help with the day-to-day stuff, like adding common modules or fine-tuning your activity logging?The Apache Cookbook is a collection of problems, solutions, and practical examples for webmasters, web administrators, programmers, and everyone else who works with Apache. For every problem addressed in the book, there's a worked-out solution or "recipe"—short, focused pieces of code that you can use immediately. But this book offers more than cut-and-paste code. You also get explanations of how and why the code works, so you can adapt the problem-solving techniques to similar situations.The recipes in the Apache Cookbook range from simple tasks, such as installing the server on Red Hat Linux or Windows, to more complex tasks, such as setting up name-based virtual hosts or securing and managing your proxy server. The two hundred plus recipes in the book cover additional topics such as: Security Aliases, Redirecting, and Rewriting CGI Scripts, the suexec Wrapper, and other dynamic content techniques Error Handling SSL Performance The impressive collection of useful code in this book is a guaranteed timesaver for all Apache users, from novices to advanced practitioners. Instead of poking around mailing lists, online documentation, and other sources, you can rely on the Apache Cookbook for quick solutions to common problems, and then you can spend your time and energy where it matters most.

How to Find, Organize, and Manipulate It

Hadoop MapReduce Cookbook

ZooKeeper

Solr Cookbook - Third Edition

Over 100 recipes for fast, scalable, and reliable search for your enterprise. 4th Edition

How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results

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.

It starts off by discussing Solr and helping you understand how it fits into your architecture, where all databases and document/web crawlers fall short, and Solr shines. The main part of the book is a thorough exploration of nearly every feature that Solr offers. To keep this interesting and realistic, we use a large open source set of metadata about artists, releases, and tracks courtesy of the MusicBrainz.org project. Using this data as a testing ground for Solr, you will learn how to import this data in various ways from CSV to XML to database access.

This book distills 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.

Individual self-contained code recipes. Solve specific problems using individual recipes, or work through the book to develop your capabilities. If you are a big data enthusiast and striving to use Hadoop to solve your problems, this book is for you. Aimed at Java programmers with some knowledge of Hadoop MapReduce, this is also a comprehensive reference for developers and system admins who want to get up to speed using Hadoop.

Data Pipelines with Apache Airflow

Solutions and Examples for Java Developers

Quarkus Cookbook

Taming Text

Elasticsearch Server: Second Edition

Scaling Big Data with Hadoop and Solr - Second Edition

If you are a developer who implements Elasticsearch in your web applications and want to sharpen your understanding of the core elements and applications, this is the book for you. It is assumed that you've got working knowledge of JSON and, if you want to extend Elasticsearch, of Java and related technologies.

This book is in Packt's Cookbook series. A Packt Cookbook contains recipes for solutions to the most important problems you face when working with a topic. Inside the Cookbook you will find: A straightforward and easy to follow format, A selection of the most important tasks and problems, Carefully organized instructions for solving the problem efficiently, Clear explanations of what you did, Details for applying the solution to other situations. This is the ideal book for people who have already written a first application with the Play Framework or have just finished reading through the documentation. In other words - anyone who is ready to get to grips with Play! Having a basic knowledge of Java is good, as well as well as some web developer skills - HTML and JavaScript

Your Python code may run correctly, but you need it to run faster. Updated for Python 3, this expanded edition shows you how to locate performance bottlenecks and significantly speed up your code in high-data-volume programs. By exploring the fundamental theory behind design choices, High Performance Python helps you gain a deeper understanding of Python's implementation. How do you take advantage of multicore architectures or clusters? Or build a system that scales up and down without losing reliability? Experienced Python programmers will learn concrete solutions to many issues, along with war stories from companies that use high-performance Python for social media analytics, productionized machine learning, and more. Get a better grasp of NumPy, Cython, and profilers Learn how Python abstracts the underlying computer architecture Use profiling to find bottlenecks in CPU time and memory usage Write efficient programs by choosing appropriate data structures Speed up matrix and vector computations Use tools to compile Python down to machine code Manage multiple I/O and computational operations concurrently Convert multiprocessing code to run on local or remote clusters Deploy code faster using tools like Docker

Discover how data science can help you gain in-depth insight into your business - the easy way! Jobs in data science abound, but few people have the data science skills needed to fill these increasingly important roles. Data Science For Dummies is the perfect starting point for IT professionals and students who want a quick primer on all areas of the expansive data science space. With a focus on business cases, the book explores topics in big data, data science, and data engineering, and how these three areas are combined to produce tremendous value. If you want to pick-up the skills you need to begin a new career or initiate a new project, reading this book will help you understand what technologies, programming languages, and mathematical methods on which to focus. While this book serves as a wildly fantastic guide through the broad, sometimes intimidating field of big data and data science, it is not an instruction manual for hands-on implementation. Here's what to expect: Provides a background in big data and data engineering before moving on to data science and how it's applied to generate value Includes coverage of big data frameworks like Hadoop, MapReduce, Spark, MPP platforms, and NoSQL Explains machine learning and many of its algorithms as well as artificial intelligence and the evolution of the Internet of Things Details data visualization techniques that can be used to showcase, summarize, and communicate the data insights you generate It's a big, big data world out there—let Data Science For Dummies help you harness its power and gain a competitive edge for your organization.

Apache Solr Enterprise Search Server - Third Edition

Apache Cookbook

Apache Solr 4 Cookbook

Ansible

Play Framework Cookbook

Relevant Search

The best-selling author of Big Data is back, this time with a unique and in-depth insight into how specific companies use big data. Big data is on the tip of everyone's tongue. Everyone understands its power and importance, but many fail to grasp the actionable steps and resources required to utilise it effectively. This book fills the knowledge gap by showing how major companies are using big data every day, from an up-close, on-the-ground perspective. From technology, media and retail, to sport teams, government agencies and financial institutions, learn the actual strategies and processes being used to learn about customers, improve manufacturing, spur innovation, improve safety and so much more. Organised for easy dip-in navigation, each chapter follows the same structure to give you the information you need quickly. For each company profiled, learn what data was used, what problem it solved and the processes put it place to make it practical, as well as the technical details, challenges and lessons learned from each unique scenario. Learn how predictive analytics helps Amazon, Target, John Deere and Apple understand their customers Discover how big data is behind the success of Walmart, LinkedIn, Microsoft and more Learn how big data is changing medicine, law enforcement, hospitality, fashion, science and banking Develop your own big data strategy by accessing additional reading materials at the end of each chapter

This book is a detailed, practical, hands-on guide packed with real-life scenarios and examples which will show you how to implement an ElasticSearch search engine on your own websites. If you are a web developer or a user who wants to learn more about ElasticSearch, then this is the book for you. You do not need to know anything about ElasticSearch, Java, or Apache Lucene in order to use this book, though basic knowledge about databases and queries is required. Summary Deep Learning for Search teaches you how to improve the effectiveness of your search by implementing neural network-based techniques. By the time you're finished with the book, you'll be ready to build amazing search engines that deliver the results your users need and that get better as time goes on! Foreword by Chris Mattmann. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the Technology Deep learning handles the toughest search challenges, including imprecise search terms, badly indexed data, and retrieving images with minimal metadata. And with modern tools like DL4J and TensorFlow, you can apply powerful DL techniques without a deep background in data science or natural language processing (NLP). This book will show you how. About the Book Deep Learning for Search teaches you to improve your search results with neural networks. You'll review how DL relates to search basics like indexing and ranking. Then, you'll walk through in-depth examples to upgrade your search with DL techniques using Apache Lucene and DeepLearning4J. As the book progresses, you'll explore advanced topics like searching through images, translating user queries, and designing search engines that improve as they learn! What's Inside Accurate and relevant rankings Searching across languages Content-based image search Search with recommendations About the Reader For developers comfortable with Java or a similar language and search basics. No experience with deep learning or NLP needed. About the Author Tommaso Teofili is a software engineer with a passion for open source and machine learning. As a member of the Apache Software Foundation, he contributes to a number of open source projects, ranging from topics like information retrieval (such as Lucene and Solr) to natural language processing and machine translation (including OpenNLP, Joshua, and UIMA). He currently works at Adobe, developing search and indexing infrastructure components, and researching the areas of natural language processing, information retrieval, and deep learning. He has presented search and machine learning talks at conferences including BerlinBuzzwords, International Conference on Computational Science, ApacheCon, EclipseCon, and others. You can find him on Twitter at @tteofili. Table of Contents PART 1 - SEARCH MEETS DEEP LEARNING Neural search Generating synonyms PART 2 - THROWING NEURAL NETS AT A SEARCH ENGINE From plain retrieval to text generation More-sensitive query suggestions Ranking search results with word embeddings Document embeddings for rankings and recommendations PART 3 - ONE STEP BEYOND Searching across languages Content-based image search A peek at performance

Ansible is a simple, but powerful, server and configuration management tool. Learn to use Ansible effectively, whether you manage one server—or thousands.

Data Science For Dummies

Natural Language Processing with Java and LingPipe Cookbook

CMIS and Apache Chemistry in Action

Deep Learning for Search

Elasticsearch Server

Elasticsearch 7.0 Cookbook

Search, analyze, and manage data effectively with Elasticsearch 7 Key FeaturesExtend Elasticsearch functionalities and learn how to deploy on Elastic CloudDeploy and manage simple Elasticsearch nodes as well as complex cluster topologiesExplore the capabilities of Elasticsearch 7 with easy-to-follow recipesBook Description Elasticsearch is a Lucene-based distributed search server that allows users to index and search unstructured content with petabytes of data. With this book, you'll be guided through comprehensive recipes on what's new in Elasticsearch 7, and see how to create and run complex queries and analytics. Packed with recipes on performing index mapping, aggregation, and scripting using Elasticsearch, this fourth edition of Elasticsearch Cookbook will get you acquainted with numerous solutions and quick techniques for performing both every day and uncommon tasks such as deploying Elasticsearch nodes, integrating other tools to Elasticsearch, and creating different visualizations. You will install Kibana to monitor a cluster and also extend it using a variety of plugins. Finally, you will integrate your Java, Scala, Python, and big data applications such as Apache Spark and Pig with Elasticsearch, and create efficient data applications powered by enhanced functionalities and custom plugins. By the end of this book, you will have gained in-depth knowledge of implementing Elasticsearch architecture, and you'll be able to manage, search, and store data efficiently and effectively using Elasticsearch. What you will learnCreate an efficient architecture with ElasticsearchOptimize search results by executing analytics aggregationsBuild complex queries by managing indices and documentsMonitor the performance of your cluster and nodesDesign advanced mapping to take full control of index stepsIntegrate Elasticsearch in Java, Scala, Python, and big data applicationsInstall Kibana to monitor clusters and extend it for pluginsWho this book is for If you ' re a software engineer, big data infrastructure engineer, or Elasticsearch developer, you'll find this book useful. This Elasticsearch book will also help data professionals working in the e-commerce and FMCG industry who use Elastic for metrics evaluation and search analytics to get deeper insights for better business decisions. Prior experience with Elasticsearch will help you get the most out of this book.

In this fast-paced book on the Docker open standards platform for developing, packaging and running portable distributed applications, Deepak Vorhadiscusses how to build, ship and run applications on any platform such as a PC, the cloud, data center or a virtual machine. He describes how to install and create Docker images, and the advantages of Docker containers.The remainder of the book is devoted to discussing using Docker with important software solutions. He begins by discussing using Docker with a traditional RBDBMS using Oracle and MySQL. Next he moves on to NoSQL with chapter on MongoDB Cassandra, and Couchbase. Then he addresses the use of Docker in the Hadoop ecosystem with complete chapters on utilizing not only Hadoop, but Hive, HBase, Sqoop, Kafka, Solr and Spark. What You Will Learn How to install a Docker image How to create a Docker container How to run an Application in a Docker Container Use Docker with Apache Hadoop Ecosystem Use Docker with NoSQL Databases Use Docker with RDBMS Who This Book Is ForApache Hadoop Developers. Database developers. NoSQL Developers.

Optimized for Kubernetes, Quarkus is designed to help you create Java applications that are cloud first, container native, and serverless capable. With this cookbook, author Alex Soto Bueno and Jason Porter from Red Hat provide detailed solutions for installing, interacting with, and using Quarkus in the development and production of microservices. The recipes in this book midlevel to senior developers familiar with Java enterprise application development how to get started with Quarkus quickly. You ' ll become familiar with how Quarkus works within the wider Java ecosystem and discover ways to adapt this framework to your particular needs. You ' ll learn how to shorten the development cycle by enabling the dev mode Connect to and communicate with Kafka Develop with the reactive programming model Easily add fault tolerance to your services Build your application as a Kubernetes-ready container Ease development with OpenAPI and test a native Quarkus application

Building distributed applications is difficult enough without having to coordinate the actions that make them work. This practical guide shows how Apache ZooKeeper helps you manage distributed systems, so you can focus mainly on application logic. Even with ZooKeeper, implementing coordination tasks is not trivial, but this book provides good practices to give you a head start, and points out caveats that developers and administrators alike need to watch for along the way. In three separate sections, ZooKeeper contributors Flavio Junqueira and Benjamin Reed introduce the principles of distributed systems, provide ZooKeeper programming techniques, and include the information you need to administer this service. Learn how ZooKeeper solves common coordination tasks Explore the ZooKeeper API ' s Java and C implementations and how they differ Use methods to track and react to ZooKeeper state changes Handle failures of the network, application processes, and ZooKeeper itself Learn about ZooKeeper ' s trickier aspects dealing with concurrency, ordering, and configuration Use the Curator high-level interface for connection management Become familiar with ZooKeeper internals and administration tools

Spring Roo 1.1 Cookbook
Spring Boot in Action
Big data processing at scale to unlock unique business insights
Angular Design Patterns
Mastering Elasticsearch - Second Edition
ElasticSearch Cookbook - Second Edition

ElasticSearch is an open source search server built on Apache Lucene. It was built to provide a scalable search solution with built-in support for near real-time search and multi-tenancy. Jumping into the world of ElasticSearch by setting up your own custom cluster, this book will show you how to create a fast, scalable, and flexible search solution. By learning the ins-and-outs of data indexing and analysis, "ElasticSearch Server" will start you on your journey to mastering the powerful capabilities of ElasticSearch. With practical chapters covering how to search data, extend your search, and go deep into cluster administration and search analysis, this book is perfect for those new and experienced with search servers. In "ElasticSearch Server" you will learn how to revolutionize your website or application with faster, more accurate, and flexible search functionality. Starting with chapters on setting up your own ElasticSearch cluster and searching and extending your search parameters you will quickly be able to create a fast, scalable, and completely custom search solution. Building on your knowledge further you will learn about ElasticSearch's query API and become confident using powerful filtering and faceting capabilities. You will develop practical knowledge on how to make use of ElasticSearch's near real-time capabilities and support for multi-tenancy. Your journey then concludes with chapters that help you monitor and tune your ElasticSearch cluster as well as advanced topics such as shard allocation, gateway configuration, and the discovery module.

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 *An Airflow bible. Useful for all kinds of users, from novice to expert." - Rambabu Posa, Sai Aashika Consultancy Data Pipelines with Apache Airflow teaches you how to build and maintain effective data pipelines. A successful pipeline moves data efficiently, minimizing pauses and blockages between tasks, keeping every process along the way operational. Apache Airflow provides a single customizable environment for building and managing data pipelines, eliminating the need for a hodgepodge collection of tools, snowflake code, and homegrown processes. Using real-world scenarios and examples, Data Pipelines with Apache Airflow teaches you how to simplify and automate data pipelines, reduce operational overhead, and smoothly integrate all the technologies in your stack. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. About the technology Data pipelines manage the flow of data from initial collection through consolidation, cleaning, analysis, visualization, and more. Apache Airflow provides a single platform you can use to design, implement, monitor, and maintain your pipelines. Its easy-to-use UI, plug-and-play options, and flexible Python scripting make Airflow perfect for any data management task. About the book Data Pipelines with Apache Airflow teaches you how to build and maintain effective data pipelines. You'll explore the most common usage patterns, including aggregating multiple data sources, connecting to and from data lakes, and cloud deployment. Part reference and part tutorial, this practical guide covers every aspect of the directed acyclic graphs (DAGs) that power Airflow, and how to customize them for your pipeline's needs. What's inside Build, test, and deploy Airflow pipelines as DAGs Automate moving and transforming data Analyze historical datasets using backfilling Develop custom components Set up Airflow in production environments About the reader For DevOps, data engineers, machine learning engineers, and sysadmins with intermediate Python skills. About the author Bas Harenslak and Julian de Ruyter are data engineers with extensive experience using Airflow to develop pipelines for major companies. Bas is also an Airflow committer. Table of Contents PART 1 - GETTING STARTED 1 Meet Apache Airflow 2 Anatomy of an Airflow DAG 3 Scheduling in Airflow 4 Templating tasks using the Airflow context 5 Defining dependencies between tasks PART 2 - BEYOND THE BASICS 6 Triggering workflows 7 Communicating with external systems 8 Building custom components 9 Testing 10 Running tasks in containers PART 3 - AIRFLOW IN PRACTICE 11 Best practices 12 Operating Airflow in production 13 Securing Airflow 14 Project: Finding the fastest way to get around NYC PART 4 - IN THE CLOUDS 15 Airflow in the clouds 16 Airflow on AWS 17 Airflow on Azure 18 Airflow in GCP

Summary CMIS and Apache Chemistry in Action is a comprehensive guide to the CMIS standard and related ECM concepts, written by the authors of the standard. In it, you'll tackle hands-on examples for building applications on CMIS repositories from both the client and the server sides. You'll learn how to create new content-centric applications that install and run in any CMIS-compliant repository. About The Technology Content Management Interoperability Services (CMIS) is an OASIS standard for accessing content management systems. It specifies a vendor- and language-neutral way to interact with any compliant content repository. Apache Chemistry provides complete reference implementations of the CMIS standard with robust APIs for developers writing tools, applications, and servers. About This Book CMIS and Apache Chemistry in Action is a comprehensive guide to the CMIS standard and related ECM concepts. In it, you'll find clear teaching and instantly useful examples for building content-centric client and server-side applications that run against any CMIS-compliant repository. In fact, using the CMIS Workbench and the InMemory Repository from Apache Chemistry, you'll have running code talking to a real CMIS server by the end of chapter 1. This book requires some familiarity with content management systems and a standard programming language like Java or C#. No exposure to CMIS or Apache Chemistry is assumed. Purchase of the print book includes a free eBook in PDF, Kindle, and ePub formats from Manning Publications. What's inside The only CMIS book endorsed by OASIS Complete coverage of the CMIS 1.0 and 1.1 specifications Cookbook-style tutorials and real-world examples About the Authors Florian Müller, Jay Brown, and Jeff Potts are among the original authors, contributors, and leaders of Apache Chemistry and the OASIS CMIS specification. They continue to shape CMIS implementations at Al fresco, IBM, and SAP. Table of Contents PART 1 UNDERSTANDING CMIS Introducing CMIS Exploring the CMIS domain model Creating, updating, and deleting objects with CMIS CMIS metadata: types and properties Query PART 2 HANDS-ON CMIS CLIENT DEVELOPMENT Meet your new project: The Blend The Blend: read and query functionality The Blend: create, update, and delete functionality Using other client libraries Building mobile apps with CMIS PART 3 ADVANCED TOPICS CMIS bindings Security and control Performance Building a CMIS server