

---

# File Type PDF Apache Solr A Practical Approach To Enterprise Search

---

This is likewise one of the factors by obtaining the soft documents of this **Apache Solr A Practical Approach To Enterprise Search** by online. You might not require more time to spend to go to the book creation as skillfully as search for them. In some cases, you likewise accomplish not discover the message Apache Solr A Practical Approach To Enterprise Search that you are looking for. It will extremely squander the time.

However below, when you visit this web page, it will be fittingly extremely simple to get as skillfully as download guide Apache Solr A Practical Approach To Enterprise Search

It will not take many mature as we tell before. You can reach it while action something else at home and even in your workplace. as a result easy! So, are you question? Just exercise just what we meet the expense of below as capably as review **Apache Solr A Practical Approach To Enterprise Search** what you later to read!

---

## ZCUVS9 - LAILA MIGUEL

---

Building a simple but powerful recommendation system is much easier than you think. Approachable for all levels of expertise, this report explains innovations that make machine learning practical for business production settings—and demonstrates how even a small-scale development team can design an effective large-scale recommendation system. Apache Mahout committers Ted Dunning and Ellen Friedman walk you through a design that relies on careful simplification. You'll learn how to collect the right data, analyze it with an algorithm from the Mahout library, and then easily deploy the recommender using search technology, such as Apache Solr or Elasticsearch. Powerful and effective, this efficient combination does learning offline and delivers rapid response recommendations in real time. Understand the tradeoffs between simple and complex recommenders Collect user data that tracks user actions—rather than their ratings Predict what a user wants based on behavior by others, using Mahout for co-occurrence analysis Use search technology to offer recommendations in real time, complete with item metadata Watch the recommender in action with a music service example Improve your recommender with dithering, multimodal recommendation, and other techniques 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.

Make the most of Angular by leveraging 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 meet 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.

This book is a step-by-step tutorial that will enable you to lever-

age 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.

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

Make the most of Angular by leveraging 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 meet 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.

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.

A fast-paced, example-based guide to learning how to administer, monitor, and optimize Apache Solr. "Administering Solr" is for developers and Solr administrators who have a basic knowledge of Solr and who are looking for ways to keep their Solr server healthy and well maintained. A basic working knowledge of Apache Lucene is recommended, but this is not mandatory.

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.

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 Learn Apply 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 For Data scientists who want to refresh and learn various concepts of natural language processing through coding exercises.

Maximize the power of Magento 2 to improve your e-commerce business About This Book Strategic planning for maximizing your Magento 2 store's operations and sales. Important guidelines and advice for optimizing your Magento 2 store. Packed with screenshots and step-by-step instructions to leverage the hundreds of Magento 2 features. Who This Book Is For This guide is for store owners who use Magento 2 on a daily basis to operate their online stores. Proprietors will use this guide as a reference for ensuring that they are utilizing the full power of Magento 2. Developers and solution providers will find this guide a helpful tool in identifying opportunities to help their clients navigate the complexity of Magento 2. What You Will Learn Strategies for creating multiple stores within a single Magento installation. Use of complex product types such as bundles, downloadables and virtual products. Configuring payments, shipping and taxes. Accommodating multiple languages and currencies. Installing and configuring themes. Managing content and search engine optimization. Creating promotions, up-sells and cross-sells. Leveraging the customer relationship features of Magento 2. Securing and optimizing your installation. In Detail Magento 2 has been completely re-written and re-designed to take the world's most popular open source e-commerce platform to new heights of productivity and usability. With this overhaul comes improved opportunities to leverage the world's most successful open source e-commerce platform for your online retail and wholesale business. The focus of this book is on operational aspects: those actions you, as a proprietor, can take to give your customers a more delightful shopping experience. In this guide, we delve into all aspects of managing a Magento 2 store, from product creation to marketing, customer management to search engine optimization. This book is not only a step-by-step course in setting up your Magento 2 website, it will be a handy reference as continue to improve and refine your on-

line presence. Style and approach A step-by-step guide where it takes the reader from the planning step all the way through launching a new Magento-powered store by building on the initial sample data included in a default Magento 2 installation. Along the way, we explore using discrete action steps how to leverage the full power of Magento 2 for both simple and advanced uses. Screenshots and step-by-step instructions not only teach the new Magento store administrator, but also serve as a handy reference for the more experienced operators.

Distributed applications demand reliable, high-performance messaging. The Apache Pulsar server-to-server messaging system provides a secure, stable platform without the need for a stream processing engine like Spark. Contributed by Yahoo to the Apache Foundation, Pulsar is mature and battle-tested, handling millions of messages per second for over three years at Yahoo. Apache Pulsar in Action is a comprehensive and practical guide to building high-traffic applications with Pulsar, delivering extreme levels of speed and durability. about the technology Pulsar is a streaming messaging system designed for high performance server-to-server messaging. Built and tested under intense conditions at Yahoo, Pulsar has been proven in production and can handle millions of messages per second. Now free and open-source, Pulsar's unique architecture helps solve some of the challenges of modern development. Pulsar avoids latency in streaming data transmission, making it a powerful tool for IoT Edge analytics. Its unified messaging model improves the performance of microservices architecture, and its tiered storage capabilities allow for larger volumes of data to be handled without fear of data loss. Pulsar's flexible API interface works with Java, C++, Python, and Go, making it easy to incorporate Pulsar into your stack. about the book Apache Pulsar in Action is a hands-on guide to building scalable streaming messaging systems for distributed applications and microservices systems. You'll start with Pulsar's fundamentals, each illustrated by real-world examples, as you get to grips with Pulsar's unique architecture. Pulsar contributor David Kjerrumgaard teaches the skills you need to deploy a Pulsar server, ingest data from third-party systems, and deploy lightweight computing logic with simple functions. You'll learn to employ Pulsar's seamless scalability through relatable case studies, including an IOT analytics application that can be deployed within a resource constrained environment and a microservices application based on Pulsar functions.

At the end of this practical book, you'll be ready to fully take advantage of Pulsar to create high-traffic message-driven applications. what's inside 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 Examples of Pulsar-based microservices that you can download and try yourself about the reader Written for experienced Java developers. No prior knowledge of Pulsar is needed. about the author David Kjerrumgaard is the Director of Solution Architecture at Streamlio, and a contributor to the Apache Pulsar and Apache NiFi projects.

"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

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.

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. 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.

Summary Tika in Action is a hands-on guide to content mining with Apache Tika. The book's many examples and case studies offer real-world experience from domains ranging from search engines to digital asset management and scientific data processing. About the Technology Tika is an Apache toolkit that has built into it everything you and your app need to know about file formats. Using Tika, your applications can discover and extract content from digital documents in almost any format, including exotic ones. About this Book Tika in Action is the ultimate guide to content mining using Apache Tika. You'll learn how to pull usable information from otherwise inaccessible sources, including internet media and file archives. This example-rich book teaches you to build and extend applications based on real-world experience with search engines, digital asset management, and scientific data processing. In addition to architectural overviews, you'll find detailed chapters on features like metadata extraction, automatic language detection, and custom parser development. This book is written for developers who are new to both Scala and Lift and covers just enough Scala to get you started. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book. What's Inside Crack MS Word, PDF, HTML, and ZIP Integrate with search engines, CMS, and other data sources Learn through experimentation Many examples This book requires no previous knowledge of Tika or text mining techniques. It assumes a working knowledge of

Java.

==== Table of Contents PART 1 GETTING STARTED The case for the digital Babel fish Getting started with Tika The information landscape PART 2 TIKA IN DETAIL Document type detection Content extraction Understanding metadata Language detection What's in a file? PART 3 INTEGRATION AND ADVANCED USE The big picture Tika and the Lucene search stack Extending Tika PART 4 CASE STUDIES Powering NASA science data systems Content management with Apache Jackrabbit Curating cancer research data with Tika The classic search engine example 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 relat-

ed 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.

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.

Enhance your Solr indexing experience with advanced techniques and the built-in functionalities available in Apache Solr About This Book Learn about distributed indexing and real-time optimization to change index data on fly Index data from various sources and web crawlers using built-in analyzers and tokenizers This step-by-step guide is packed with real-life examples on indexing data Who This Book Is For This book is for developers who want to increase their experience of indexing in Solr by learning about the various index handlers, analyzers, and methods available in Solr. Beginner level Solr development skills are expected. What You Will Learn Get to know the basic features of Solr indexing and the analyzers/tokenizers available Index XML/JSON data in Solr using the HTTP Post tool and CURL command Work with Data Import Handler to index data from a database Use Apache Tika with Solr to index word documents, PDFs, and much more Utilize Apache Nutch and Solr integration to index crawled data from web pages Update indexes in real-time data feeds Discover techniques to index multi-language and distributed data in Solr Combine the various indexing techniques into a real-life working example of an online shopping web application In Detail Apache Solr is a widely used, open source enterprise search server that delivers powerful indexing and searching features. These features help fetch relevant information from various sources and documentation. Solr also combines with other open source tools such as Apache Tika and Apache Nutch to provide more powerful features. This fast-paced guide starts by helping you set up Solr and get acquainted with its basic building blocks, to give you a better understanding of Solr indexing. You'll quickly move on to indexing text and boosting the indexing time. Next, you'll focus on basic indexing techniques, various index handlers designed to modify documents, and index-

ing a structured data source through Data Import Handler. Moving on, you will learn techniques to perform real-time indexing and atomic updates, as well as more advanced indexing techniques such as de-duplication. Later on, we'll help you set up a cluster of Solr servers that combine fault tolerance and high availability. You will also gain insights into working scenarios of different aspects of Solr and how to use Solr with e-commerce data. By the end of the book, you will be competent and confident working with indexing and will have a good knowledge base to efficiently program elements. Style and approach This fast-paced guide is packed with examples that are written in an easy-to-follow style, and are accompanied by detailed explanation. Working examples are included to help you get better results for your applications.

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.

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 learning 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.

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.

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

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.

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.

Learn everything you need to build highly scalable, robust web applications using Angular release 4 About This Book Apply best

practices and design patterns to achieve higher scalability in your Angular applications Understand the latest features of Angular and create your own components Get acquainted with powerful, advanced techniques in Angular to build professional web applications Who This Book Is For This book is for JavaScript developers with some prior exposure to Angular, at least through basic examples. We assume that you've got working knowledge of HTML, CSS, and JavaScript. What You Will Learn Implement asynchronous programming using Angular Beautify your application with the UI components built to the material design specification Secure your web application from unauthorized users Create complex forms, taking full advantage of 2-way data binding Test your Angular applications using the Jasmine and Protractor frameworks for better efficiency Learn how to integrate Angular with Bootstrap to create compelling web applications Use Angular built-in classes to apply animation in your app In Detail Got some experience of Angular under your belt? Want to learn everything about using advanced features for developing websites? This book is everything you need for the deep understanding of Angular that will set you apart from the developer crowd. Angular has introduced a new way to build applications. Creating complex and rich web applications, with a lighter resource footprint, has never been easier or faster. Angular is now at release 4, with significant changes through previous versions. This book has been written and tested for Angular release 4. Angular is a mature technology, and you'll likely have applications built with earlier versions. This book starts by showing you best practices and approaches to migrating your existing Angular applications so that you can be immediately up-to-date. You will take an in-depth look at components and see how to control the user journey in your applications by implementing routing and navigation. You will learn how to work with asynchronous programming by using Observables. To easily build applications that look great, you will learn all about template syntax and how to beautify applications with Material Design. Mastering forms and data binding will further speed up your application development time. Learning about managing services and animations will help you to progressively enhance your applications. Next you'll use native directives to integrate Bootstrap with Angular. You will see the best ways to test your application with the leading options such as Jasmine and Protractor. At the end of the book, you'll learn how to apply design patterns in Angular, and

see the benefits they will bring to your development. Style and approach This book provides comprehensive coverage of all aspects of development with Angular. You will learn about all the most powerful Angular concepts, with examples and best practices. This book is everything you need for the deep understanding of Angular that will set you apart from the developer crowd.

If you are a Hadoop programmer who wants to learn about Flume to be able to move datasets into Hadoop in a timely and replicable manner, then this book is ideal for you. No prior knowledge about Apache Flume is necessary, but a basic knowledge of Hadoop and the Hadoop File System (HDFS) is assumed.

Unleash the data processing and analytics capability of Apache Spark with the language of choice: Java About This Book Perform big data processing with Spark—without having to learn Scala! Use the Spark Java API to implement efficient enterprise-grade applications for data processing and analytics Go beyond mainstream data processing by adding querying capability, Machine Learning, and graph processing using Spark Who This Book Is For If you are a Java developer interested in learning to use the popular Apache Spark framework, this book is the resource you need to get started. Apache Spark developers who are looking to build enterprise-grade applications in Java will also find this book very useful. What You Will Learn Process data using different file formats such as XML, JSON, CSV, and plain and delimited text, using the Spark core Library. Perform analytics on data from various data sources such as Kafka, and Flume using Spark Streaming Library Learn SQL schema creation and the analysis of structured data using various SQL functions including Windowing functions in the Spark SQL Library Explore Spark Mlib APIs while implementing Machine Learning techniques to solve real-world problems Get to know Spark GraphX so you understand various graph-based analytics that can be performed with Spark In Detail Apache Spark is the buzzword in the big data industry right now, especially with the increasing need for real-time streaming and data processing. While Spark is built on Scala, the Spark Java API exposes all the Spark features available in the Scala version for Java developers. This book will show you how you can implement various functionalities of the Apache Spark framework in Java, without stepping out of your comfort zone. The book starts with an introduction to the Apache Spark 2.x ecosystem, followed by explaining how to install and configure Spark, and refreshes the Java concepts that

will be useful to you when consuming Apache Spark's APIs. You will explore RDD and its associated common Action and Transformation Java APIs, set up a production-like clustered environment, and work with Spark SQL. Moving on, you will perform near-real-time processing with Spark streaming, Machine Learning analytics with Spark MLlib, and graph processing with GraphX, all using various Java packages. By the end of the book, you will have a solid foundation in implementing components in the Spark framework in Java to build fast, real-time applications. Style and approach This practical guide teaches readers the fundamentals of the Apache Spark framework and how to implement components using the Java language. It is a unique blend of theory and practical examples, and is written in a way that will gradually build your knowledge of Apache Spark.

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 advanced 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 special-

ists 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.

Chapter 7. Case Study : Comparing Twitter Archives; Getting the Data and Distribution of Tweets; Word Frequencies; Comparing Word Usage; Changes in Word Use; Favorites and Retweets; Summary; Chapter 8. Case Study : Mining NASA Metadata; How Data Is Organized at NASA; Wrangling and Tidying the Data; Some Initial Simple Exploration; Word Co-occurrences and Correlations; Networks of Description and Title Words; Networks of Keywords; Calculating tf-idf for the Description Fields; What Is tf-idf for the Description Field Words?; Connecting Description Fields to Keywords; Topic Modeling.

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 process-

ing 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.

A comprehensive guide to using the web application, including such topics as text analysis, faceted search, result grouping, multilingual search, advanced geospatial and data operations, and relevancy tuning.

Big Data: Principles and Paradigms captures the state-of-the-art research on the architectural aspects, technologies, and applications of Big Data. The book identifies potential future directions and technologies that facilitate insight into numerous scientific, business, and consumer applications. To help realize Big Data's

full potential, the book addresses numerous challenges, offering the conceptual and technological solutions for tackling them. These challenges include life-cycle data management, large-scale storage, flexible processing infrastructure, data modeling, scalable machine learning, data analysis algorithms, sampling techniques, and privacy and ethical issues. Covers computational platforms supporting Big Data applications Addresses key principles underlying Big Data computing Examines key developments supporting next generation Big Data platforms Explores the challenges in Big Data computing and ways to overcome them Contains expert contributors from both academia and industry

When Lucene first hit the scene five years ago, it was nothing short of amazing. By using this open-source, highly scalable, super-fast search engine, developers could integrate search into applications quickly and efficiently. A lot has changed since then—search has grown from a "nice-to-have" feature into an indispensable part of most enterprise applications. Lucene now powers search in diverse companies including Akamai, Netflix, LinkedIn, Technorati, HotJobs, Epiphany, FedEx, Mayo Clinic, MIT, New Scientist Magazine, and many others. Some things remain the same, though. Lucene still delivers high-performance search features in a disarmingly easy-to-use API. Due to its vibrant and diverse open-source community of developers and users, Lucene is relentlessly improving, with evolutions to APIs, significant new features such as payloads, and a huge increase (as much as 8x) in indexing speed with Lucene 2.3. And with clear writing, reusable examples, and unmatched advice on best practices, Lucene in Action, Second Edition is still the definitive guide to developing with Lucene. Purchase of the print book comes with an offer of a free PDF, ePub, and Kindle eBook from Manning. Also available is all code from the book.

Work with the intricacies of Angular 2 to build efficient and flexible applications About This Book\* Effectively structure, write, test, and deploy your application\* Harness the full power of Angular 2 by creating your own components\* Query and modify data in various back ends and become proficient with the Promise API Who This Book Is For If you are a JavaScript developer looking for a comprehensive learning path that will take you from the essentials to a high intermediate level, then this is the book for you. We assume that you have worked with Angular before and are new to Angular 2. Working knowledge of HTML, CSS, and JavaScript is ex-

pected. What You Will Learn\* Design and code an effective navigation schema in your web application\* Apply the dependency injection principle when designing Angular 2 components and services\* Secure your web application from unauthorized users\* Quickly create complex forms, taking full advantage of 2-way data binding\* Unit test your code using the Jasmine BDD testing and Protractor frameworks\* Integrate Bootstrap with Angular 2 using native directives\* Utilize Angular 2 built-in classes to apply animation in your app

Detail Are you a JavaScript developer with some experience of Angular JS under your belt? Are you wondering how the experts create more organized, readable, and testable code? Then your wait ends with this book. We'll teach you to build faster, more efficient, and more flexible applications. On this journey, you'll adopt the best practices and see an easier approach to migrating your existing Angular 1 app to Angular 2. You'll work with components, implement routing and navigation, and use asynchronous programming effectively using Observables. You will learn about template syntax and beautify Angular 2 applications with Material Design, then you'll progressively enhance your applications by managing services and animations. Next you'll integrate Bootstrap with Angular 2 using native directives and test your application using Jasmine and Protractor. At the end of the

book, we'll dive deep into design patterns and you'll see how to implement them in your very own application.

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 ex-

tract 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.