

1. Record Nr.	UNINA9910815187903321
Autore	Handiekar Sachin
Titolo	Apache Solr for indexing data : enhance your Solr indexing experience with advanced techniques and the built-in functionalities available in Apache Solr / / Sachin Handiekar, Anshul Johri
Pubbl/distr/stampa	Birmingham : , : Packt Publishing, , 2015
ISBN	1-78355-324-3
Edizione	[1st edition]
Descrizione fisica	1 online resource (160 p.)
Collana	Community experience distilled
Soggetti	Electronic information resource searching Indexing - Computer programs
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Nota di contenuto	Cover; Copyright; Credits; About the Authors; About the Reviewers; www.PacktPub.com; Table of Contents; Preface; Chapter 1: Getting Started; Overview and installation of Solr; Installing Solr in OS X (Mac); Running Solr; Installing Solr in Windows; Installing Solr on Linux; The Solr architecture and directory structure; Solr directory structure; Cores in Solr (Multicore Solr); Summary; Chapter 2: Understanding Analyzers, Tokenizers, and Filters; Introducing analyzers; Analysis phases; Tokenizers; Standard tokenizer; Keyword tokenizer; Lowercase tokenizer; N-gram tokenizer; Filters Lowercase filterSynonym filter; Porter stem filter; Running your analyzer; Summary; Chapter 3: Indexing Data; Indexing data in Solr; Introducing field types; Defining fields; Defining an unique key; Copy fields and dynamic fields; Building our musicCatalogue example; Using the Solr Admin UI; Facet searching; Summary; Chapter 4: Index Data - Basic Technique and Using Index Handlers; Inserting data into Solr; Configuring UpdateRequestHandler; Indexing documents using XML; Adding and updating documents; Deleting a document; Indexing documents using JSON; Adding a single document Adding multiple JSON documentsSequential JSON update commands; Indexing updates using CSV; Summary; Chapter 5: Index Data Using Structured Data Source Using DIH; Indexing data from MySQL; Configuring datasource; DIH commands; Indexing data using XPath;

Summary; Chapter 6: Indexing Data Using Apache Tika; Introducing Apache Tika; Configuring Apache Tika in Solr; Indexing PDF and Word documents; Summary; Chapter 7: Apache Nutch; Introducing Apache Nutch; Installing Apache Nutch; Configuring Solr with Nutch; Summary; Chapter 8: Commits, Real-Time Index Optimizations, and Atomic Updates
Understanding soft commit, optimize, and hard commitUsing atomic updates in Solr; Using RealTime Get; Summary; Chapter 9: Advanced Topics - Multilanguage, Deduplication, and Others; Multilanguage indexing; Removing duplicate documents (deduplication); Content streaming; UIMA integration with Solr; Summary; Chapter 10: Distributed Indexing; Setting up SolrCloud; The collections API; Updating configuration files; Distributed indexing and searching; Summary; Chapter 11: Case Study of Using Solr in E-Commerce; Creating an AutoSuggest feature; Facet navigation; Search filtering and sorting
Relevancy boostingSummary; Index

Sommario/riassunto

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 indexing 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 knowledg...
