1. Record Nr. UNINA9910788111503321

Autore	Karambelkar Hrishikesh Vijay
Titolo	Scaling big data with Hadoop and Solr : understand, design, build, and optimize your big data search engine with Hadoop and Apache Solr / / Hrishikesh Vijay Karambelkar
Pubbl/distr/stampa	Birmingham, England : , : Packt Publishing, , 2015 ©2015
Edizione	[Second edition.]
Descrizione fisica	1 online resource (166 p.)
Collana	Community Experience Distilled
Disciplina	004
Soggetti	Electronic data processing Data mining Big data
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Nota di contenuto	Cover; Copyright; Credits; About the Author; About the Reviewers; www.PacktPub.com; Table of Contents; Preface; Chapter 1: Processing Big Data Using Hadoop and MapReduce; Apache Hadoop's ecosystem; Core components; Understanding Hadoop's ecosystem; Configuring Apache Hadoop; Prerequisites; Setting up ssh without passphrase; Configuring Hadoop; Running Hadoop; Setting up a Hadoop cluster; Common problems and their solutions; Summary; Chapter 2: Understanding Apache Solr; Setting up Apache Solr; Prerequisites for setting up Apache Solr; Running Apache Solr on jetty Running Solr on other J2EE containersHello World with Apache Solr!; Understanding Solr administration; Solr navigation; Common problems and solutions; The Apache Solr architecture; Configuring Solr; Understanding the Solr structure; Defining the Solr schema; Solr fields; Dynamic fields in Solr; Copying the fields; Dealing with field types; Additional metadata configuration; Other important elements of the Solr schema; Configuration files of Apache Solr; Working with solr.xml and Solr core; Instance configuration with solrconfig.xml; Understanding the Solr plugin; Other configuration Loading data in Apache SolrExtracting request handler - Solr Cell; Understanding data import handlers; Interacting with Solr through

SolrJ; Working with rich documents (Apache Tika); Querying for information in Solr; Summary; Chapter 3: Enabling Distributed Search using Apache Solr; Understanding a distributed search; Distributed search patterns; Apache Solr and distributed search; Working with SolrCloud; Why ZooKeeper?; The SolrCloud architecture; Building an enterprise distributed search using SolrCloud; Setting up SolrCloud for development; Setting up SolrCloud for production
Adding a document to SolrCloudCreating shards, collections, and replicas in SolrCloud; Common problems and resolutions; Sharding algorithm and fault tolerance; Document Routing and Sharding; Shard splitting; Load balancing and fault tolerance in SolrCloud; Apache Solr and Big Data - integration with MongoDB; What is NoSQL and how is it related to Big Data?; MongoDB at glance; Installing MongoDB; Creating Solr indexes from MongoDB; Summary; Chapter 4: Big Data Search Using Hadoop and Its Ecosystem; Understanding NoSQL; Working with the Solr HDFS connector; Big data search using Katta
How Katta works?Setting up the Katta cluster; Creating Katta indexes; Using Solr 1045 Patch - map-side indexing; Using Solr 1301 Patch - reduce-side indexing; Distributed search using Apache Blur; Setting up Apache Blur with Hadoop; Apache Solr and Cassandra; Working with Cassandra and Solr; Single node configuration; Integrating with multinode Cassandra; Scaling Solr through Storm; Getting along with Apache Storm; Advanced analytics with Solr; Integrating Solr and R; Summary; Chapter 5: Scaling Search Performance; Understanding the limits; Optimizing search schema
Specifying default search field

Sommario/riassunto

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.