1. Record Nr. UNINA9911006788503321

Autore	Rochester Eric
Titolo	Clojure data analysis cookbook / / Eric Rochester
Pubbl/distr/stampa	Birmingham, UK, : Packt Pub., c2013
ISBN	1-68015-416-8 1-299-44085-1 1-78216-265-8
Edizione	[1st edition]
Descrizione fisica	1 online resource (342 p.)
Disciplina	005.133
Soggetti	Database searching Clojure (Computer program language)
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: Importing Data for Analysis; Introduction; Creating a new project; Reading CSV data into Incanter datasets; Reading JSON data into Incanter datasets; Reading data from Excel with Incanter; Reading data from JDBC databases; Reading XML data into Incanter datasets; Scraping data from tables in web pages; Scraping textual data from web pages; Reading RDF data; Reading RDF data with SPARQL; Aggregating data from different formats; Chapter 2: Cleaning and Validating Data IntroductionCleaning data with regular expressions; Maintaining consistency with synonym maps; Identifying and removing duplicate data; Normalizing numbers; Rescaling values; Normalizing dates and times; Lazily processing very large data sets; Sampling from very large data sets; Fixing spelling errors; Parsing custom data formats; Validating data with Valip; Chapter 3: Managing Complexity with Concurrent Programming; Introduction; Managing program complexity with STM; Managing program complexity with agents; Getting better performance with commute; Combining agents and STM Maintaining consistency with ensureIntroducing safe side effects into the STM; Maintaining data consistency with validators; Tracking processing with watchers; Debugging concurrent programs with

watchers; Recovering from errors in agents; Managing input with sized queues; Chapter 4: Improving Performance with Parallel Programming; Introduction; Parallelizing processing with pmap; Parallelizing processing with Incanter; Partitioning Monte Carlo simulations for better pmap performance; Finding the optimal partition size with simulated annealing; Parallelizing with reducers
Generating online summary statistics with reducersHarnessing your GPU with OpenCL and Calx; Using type hints; Benchmarking with Criterium; Chapter 5: Distributed Data Processing with Cascalog; Introduction; Distributed processing with Cascalog and Hadoop; Querying data with Cascalog; Distributing data with Apache HDFS; Parsing CSV files with Cascalog; Complex queries with Cascalog; Aggregating data with Cascalog; Defining new Cascalog operators; Composing Cascalog queries; Handling errors in Cascalog workflows; Transforming data with Cascalog
Executing Cascalog queries in the Cloud with PalletChapter 6: Working with Incanter Datasets; Introduction; Loading Incanter's sample datasets; Loading Clojure data structures into datasets; Viewing datasets interactively with view; Converting datasets to matrices; Using infix formulas in Incanter; Selecting columns with ; Selecting rows with ; Filtering datasets with where; Grouping data with group-by; Saving datasets to CSV and JSON; Projecting from multiple datasets with join; Chapter 7: Preparing for and Performing Statistical Data Analysis with Incanter; Introduction
Generating summary statistics with rollup

Sommario/riassunto

Full of practical tips, the ""Clojure Data Analysis Cookbook"" will help you fully utilize your data through a series of step-by-step, real world recipes covering every aspect of data analysis.Prior experience with Clojure and data analysis techniques and workflows will be beneficial, but not essential.