
Java In Data Engineering

Computing with Data
Beginning Java Data Structures and Algorithms
Oracle Database Programming using Java and Web Services
97 Things Every Data Engineer Should Know
Designing Data Structures in Java
Intelligent Data Engineering and Automated Learning--IDEAL 2006
Azure Storage, Streaming, and Batch Analytics
Data Pipelines Pocket Reference
Java Methods for Financial Engineering
Scala for Data Science
Data Teams
Intelligent Data Engineering and Automated Learning
Java Data Objects
Java for Data Scientists Essential Training
Reuse in the Big Data Era
Modern Data Engineering with Apache Spark
Data Science with Java
Java Fundamentals
Java for Engineers and Scientists
Data Engineering with Python
Java Programming for Engineers
Thinking Big
Numerical Methods Using Java
Mastering Java for Data Science
Scala Programming for Big Data Analytics
Fundamentals of Data Engineering
Data Engineering and Management
Computational Life Sciences
Scalable Data Architecture with Java
HBase Essentials
Java Data Mining: Strategy, Standard, and Practice
Java for Data Science
Scala: Guide for Data Science Professionals
Programming Scala
Building Big Data Pipelines with Apache Beam
Think Data Structures
Hadoop For Dummies
Practical Database Programming with Java
Data Structures and Algorithms with Scala

Java In Data Engineering

Downloaded from dev.mabts.edu by guest

KANE ANTON

[Computing with Data](#) Elsevier

Scala will be a valuable tool to have on hand during your data science journey for everything from data cleaning to cutting-edge machine learning About This Book Build data science and data engineering solutions with ease An in-depth look at each stage of the data analysis process — from reading and collecting data to distributed analytics Explore a broad variety of data processing, machine learning, and genetic algorithms through diagrams, mathematical formulations, and source code Who This Book Is For This learning path is perfect for those who are comfortable with Scala programming and now want to enter the field of data science. Some knowledge of statistics is expected. What You Will Learn Transfer and filter tabular data to extract features for machine learning Read, clean, transform, and write data to both SQL and NoSQL databases Create Scala web applications that couple with JavaScript libraries such as D3 to create compelling interactive visualizations Load data from HDFS and HIVE with ease Run streaming and graph analytics in

Spark for exploratory analysis Bundle and scale up Spark jobs by deploying them into a variety of cluster managers Build dynamic workflows for scientific computing Leverage open source libraries to extract patterns from time series Master probabilistic models for sequential data In Detail Scala is especially good for analyzing large sets of data as the scale of the task doesn't have any significant impact on performance. Scala's powerful functional libraries can interact with databases and build scalable frameworks — resulting in the creation of robust data pipelines. The first module introduces you to Scala libraries to ingest, store, manipulate, process, and visualize data. Using real world examples, you will learn how to design scalable architecture to process and model data — starting from simple concurrency constructs and progressing to actor systems and Apache Spark. After this, you will also learn how to build interactive visualizations with web frameworks. Once you have become familiar with all the tasks involved in data science, you will explore data analytics with Scala in the second module. You'll see how Scala can be used to make sense of data through easy to follow recipes. You will learn about Bokeh bindings for exploratory data analysis and quintessential machine learning with algorithms with Spark ML library. You'll get a sufficient understanding of Spark streaming, machine learning for streaming data, and Spark graphX. Armed

with a firm understanding of data analysis, you will be ready to explore the most cutting-edge aspect of data science — machine learning. The final module teaches you the A to Z of machine learning with Scala. You'll explore Scala for dependency injections and implicits, which are used to write machine learning algorithms. You'll also explore machine learning topics such as clustering, dimensionality reduction, Naive Bayes, Regression models, SVMs, neural networks, and more. This learning path combines some of the best that Packt has to offer into one complete, curated package. It includes content from the following Packt products: Scala for Data Science, Pascal Bugnion Scala Data Analysis Cookbook, Arun Manivannan Scala for Machine Learning, Patrick R. Nicolas Style and approach A complete package with all the information necessary to start building useful data engineering and data science solutions straight away. It contains a diverse set of recipes that cover the full spectrum of interesting data analysis tasks and will help you revolutionize your data analysis skills using Scala. [Beginning Java Data Structures and Algorithms](#) O'Reilly Media Examine the techniques and Java tools supporting the growing field of data science About This Book Your entry ticket to the world of data science with the stability and power of Java Explore,

analyse, and visualize your data effectively using easy-to-follow examples Make your Java applications more capable using machine learning Who This Book Is For This book is for Java developers who are comfortable developing applications in Java. Those who now want to enter the world of data science or wish to build intelligent applications will find this book ideal. Aspiring data scientists will also find this book very helpful. What You Will Learn Understand the nature and key concepts used in the field of data science Grasp how data is collected, cleaned, and processed Become comfortable with key data analysis techniques See specialized analysis techniques centered on machine learning Master the effective visualization of your data Work with the Java APIs and techniques used to perform data analysis In Detail Data science is concerned with extracting knowledge and insights from a wide variety of data sources to analyse patterns or predict future behaviour. It draws from a wide array of disciplines including statistics, computer science, mathematics, machine learning, and data mining. In this book, we cover the important data science concepts and how they are supported by Java, as well as the often statistically challenging techniques, to provide you with an understanding of their purpose and application. The book starts with an introduction of data science, followed by the basic data science tasks of data collection, data cleaning, data analysis, and data visualization. This is followed by a discussion of statistical techniques and more advanced topics including machine learning, neural networks, and deep learning. The next section examines the major categories of data analysis including text, visual, and audio data, followed by a discussion of resources that support parallel implementation. The final chapter illustrates an in-depth data science problem and provides a comprehensive, Java-based solution. Due to the nature of the topic, simple examples of techniques are presented early followed by a more detailed treatment later in the book. This permits a more natural introduction to the techniques and concepts presented in the book. Style and approach This book follows a tutorial approach, providing examples of each of the major concepts covered. With a step-by-step instructional style, this book covers various facets of data science and will get you up and running quickly.

Oracle Database Programming using Java and Web Services Dinesh Sachdev

This book describes the principles of model building in financial engineering. It explains those models as designs and working implementations for Java-based applications. The book provides software professionals with an accessible source of numerical methods or ready-to-use code for use in business applications. It is the first book to cover the topic of Java implementations for finance/investment applications and is written specifically to be accessible to software practitioners without prior accountancy/finance training. The book develops a series of packaged classes explained and designed to allow the financial engineer complete flexibility.

97 Things Every Data Engineer Should Know Springer Science & Business Media

In this book, we will be focusing upon following: Apache Hadoop and its components like HDFS and YARN. We will learn about MapReduce framework which is foundation for many big data processing frameworks & technologies. We will walk through Apache Hive, Apache Pig, Apache Flume. Also, detailing Apache Oozie. We will also get an introduction to Apache Sqoop. To get a practical overview, we would implement a case study to analyze Clickstream data and visualize the reports using Jasper iReport Designer tool. Note that this book is written to understand Big Data development. The focus will be minimal on Hadoop Cluster Administration, and/or installing tools & technologies. We will be going through practical exercises rather than keeping it theoretical. It is good to have a basic understanding of programming concepts & any programming language. This book is designed to help developers learn. This book will ensure to keep details simple and practical. Thus, even if you are a novice to IT, by the end of this book you will gain enough knowledge about engineering big data.

Designing Data Structures in Java Springer Science & Business Media

Data Science is booming thanks to R and Python, but Java brings the robustness, convenience, and ability to scale critical to today's data science applications. With this practical book, Java software engineers looking to add data science skills will take a logical journey through the data science pipeline. Author Michael Brzustowicz explains the basic math theory behind each step of the data science process, as well as how to apply these concepts with Java. You'll learn the critical roles that data IO, linear algebra, statistics, data operations, learning and prediction, and Hadoop MapReduce play in the process. Throughout this book, you'll find code examples you can use in your applications. Examine methods for obtaining, cleaning, and arranging data into its purest form Understand the matrix structure that your data should take Learn basic concepts for testing the origin and validity of data Transform your data into stable and usable numerical values Understand

supervised and unsupervised learning algorithms, and methods for evaluating their success Get up and running with MapReduce, using customized components suitable for data science algorithms **Intelligent Data Engineering and Automated Learning--IDEAL 2006** Manning Publications This book constitutes the proceedings of the 18th International Conference on Software and Systems Reuse, ICSR 2019, held in Cincinnati, Ohio, USA in June 2019. The 13 research papers included in this book were carefully reviewed and selected from 32 submissions. In addition, 3 industry innovation papers are included. The papers were organized in topical sections named: software reuse practice; software product line and requirements reuse; reuse and design and evolution; intelligent software reuse; and domain-specific software development.

Azure Storage, Streaming, and Batch Analytics Apress

Let Hadoop For Dummies help harness the power of your data and rein in the information overload Big data has become big business, and companies and organizations of all sizes are struggling to find ways to retrieve valuable information from their massive data sets with becoming overwhelmed. Enter Hadoop and this easy-to-understand For Dummies guide. Hadoop For Dummies helps readers understand the value of big data, make a business case for using Hadoop, navigate the Hadoop ecosystem, and build and manage Hadoop applications and clusters. Explains the origins of Hadoop, its economic benefits, and its functionality and practical applications Helps you find your way around the Hadoop ecosystem, program MapReduce, utilize design patterns, and get your Hadoop cluster up and running quickly and easily Details how to use Hadoop applications for data mining, web analytics and personalization, large-scale text processing, data science, and problem-solving Shows you how to improve the value of your Hadoop cluster, maximize your investment in Hadoop, and avoid common pitfalls when building your Hadoop cluster From programmers challenged with building and maintaining affordable, scaleable data systems to administrators who must deal with huge volumes of information effectively and efficiently, this how-to has something to help you with Hadoop.

Data Pipelines Pocket Reference Packt Publishing Ltd

This book introduces basic computing skills designed for industry professionals without a strong computer science background. Written in an easily accessible manner, and accompanied by a user-friendly website, it serves as a self-study guide to survey data science and data engineering for those who aspire to start a computing career, or expand on their current roles, in areas such as applied statistics, big data, machine learning, data mining, and informatics. The authors draw from their combined experience working at software and social network companies, on big data products at several major online retailers, as well as their experience building big data systems for an AI startup. Spanning from the basic inner workings of a computer to advanced data manipulation techniques, this book opens doors for readers to quickly explore and enhance their computing knowledge. Computing with Data comprises a wide range of computational topics essential for data scientists, analysts, and engineers, providing them with the necessary tools to be successful in any role that involves computing with data. The introduction is self-contained, and chapters progress from basic hardware concepts to operating systems, programming languages, graphing and processing data, testing and programming tools, big data frameworks, and cloud computing. The book is fashioned with several audiences in mind. Readers without a strong educational background in CS--or those who need a refresher--will find the chapters on hardware, operating systems, and programming languages particularly useful. Readers with a strong educational background in CS, but without significant industry background, will find the following chapters especially beneficial: learning R, testing, programming, visualizing and processing data in Python and R, system design for big data, data stores, and software craftsmanship.

Java Methods for Financial Engineering John Wiley & Sons

This book constitutes the refereed proceedings of the 7th International Conference on Intelligent Data Engineering and Automated Learning, IDEAL 2006. The 170 revised full papers presented were carefully selected from 557 submissions. The papers are organized in topical sections on learning and information processing, data mining, retrieval and management, bioinformatics and bio-inspired models, agents and hybrid systems, financial engineering, as well as a special session on nature-inspired data technologies.

Scala for Data Science Springer

Leverage Java in your data science career. Learn how to use Java for two components of data science?data engineering and data analysis.

Data Teams Packt Publishing Ltd

Orchestrate data architecting solutions using Java and related technologies to evaluate,

recommend and present the most suitable solution to leadership and clients Key FeaturesLearn how to adapt to the ever-evolving data architecture technology landscapeUnderstand how to choose the best suited technology, platform, and architecture to realize effective business valueImplement effective data security and governance principlesBook Description Java architectural patterns and tools help architects to build reliable, scalable, and secure data engineering solutions that collect, manipulate, and publish data. This book will help you make the most of the architecting data solutions available with clear and actionable advice from an expert. You'll start with an overview of data architecture, exploring responsibilities of a Java data architect, and learning about various data formats, data storage, databases, and data application platforms as well as how to choose them. Next, you'll understand how to architect a batch and real-time data processing pipeline. You'll also get to grips with the various Java data processing patterns, before progressing to data security and governance. The later chapters will show you how to publish Data as a Service and how you can architect it. Finally, you'll focus on how to evaluate and recommend an architecture by developing performance benchmarks, estimations, and various decision metrics. By the end of this book, you'll be able to successfully orchestrate data architecture solutions using Java and related technologies as well as to evaluate and present the most suitable solution to your clients. What you will learnAnalyze and use the best data architecture patterns for problemsUnderstand when and how to choose Java tools for a data architectureBuild batch and real-time data engineering solutions using JavaDiscover how to apply security and governance to a solutionMeasure performance, publish benchmarks, and optimize solutionsEvaluate, choose, and present the best architectural alternativesUnderstand how to publish Data as a Service using GraphQL and a REST APIWho this book is for Data architects, aspiring data architects, Java developers and anyone who wants to develop or optimize scalable data architecture solutions using Java will find this book useful. A basic understanding of data architecture and Java programming is required to get the best from this book.

Intelligent Data Engineering and Automated Learning "O'Reilly Media, Inc."

Leverage the power of Scala with different tools to build scalable, robust data science applications About This Book A complete guide for scalable data science solutions, from data ingestion to data visualization Deploy horizontally scalable data processing pipelines and take advantage of web frameworks to build engaging visualizations Build functional, type-safe routines to interact with relational and NoSQL databases with the help of tutorials and examples provided Who This Book Is For If you are a Scala developer or data scientist, or if you want to enter the field of data science, then this book will give you all the tools you need to implement data science solutions. What You Will Learn Transform and filter tabular data to extract features for machine learning Implement your own algorithms or take advantage of MLLib's extensive suite of models to build distributed machine learning pipelines Read, transform, and write data to both SQL and NoSQL databases in a functional manner Write robust routines to query web APIs Read data from web APIs such as the GitHub or Twitter API Use Scala to interact with MongoDB, which offers high performance and helps to store large data sets with uncertain query requirements Create Scala web applications that couple with JavaScript libraries such as D3 to create compelling interactive visualizations Deploy scalable parallel applications using Apache Spark, loading data from HDFS or Hive In Detail Scala is a multi-paradigm programming language (it supports both object-oriented and functional programming) and scripting language used to build applications for the JVM. Languages such as R, Python, Java, and so on are mostly used for data science. It is particularly good at analyzing large sets of data without any significant impact on performance and thus Scala is being adopted by many developers and data scientists. Data scientists might be aware that building applications that are truly scalable is hard. Scala, with its powerful functional libraries for interacting with databases and building scalable frameworks will give you the tools to construct robust data pipelines. This book will introduce you to the libraries for ingesting, storing, manipulating, processing, and visualizing data in Scala. Packed with real-world examples and interesting data sets, this book will teach you to ingest data from flat files and web APIs and store it in a SQL or NoSQL database. It will show you how to design scalable architectures to process and modelling your data, starting from simple concurrency constructs such as parallel collections and futures, through to actor systems and Apache Spark. As well as Scala's emphasis on functional structures and immutability, you will learn how to use the right parallel construct for the job at hand, minimizing development time without compromising scalability. Finally, you will learn how to build beautiful interactive visualizations using web frameworks. This book gives tutorials on some of the most common Scala libraries for data science, allowing you to quickly get up to speed with building

data science and data engineering solutions. Style and approach A tutorial with complete examples, this book will give you the tools to start building useful data engineering and data science solutions straightaway
Springer

This is a definitive guide to JDO API. It provides a thorough introduction to JDO (Java Data Objects), starting with a simple application that demonstrates many of JDO's capabilities. It shows the reader how to make classes persistent, how to configure JDO at runtime, how to make queries and more.

[Java Data Objects](#) Springer Nature

While teaching Java programming at Minnesota State University, the authors noticed that engineering students were enrolling in Java programming courses in order to obtain basic programming skills, but there were no Java books suitable for courses intended for engineers. They realized the need for a comprehensive Java programming tutorial that offers basic programming skills that can be applied in the field of engineering. With this in mind, the authors developed Java Programming for Engineers in order to meet the needs of both engineers and engineering students. The text uses the personal computer as a development platform and assumes no prior programming experience or knowledge. The only skills expected of the reader are basic keyboarding and user-level familiarity with the PC. Topics covered range from mathematical expressions to linear systems to engineering graphics. Chapters on problem solving skills and the designing of engineering applications walk readers through real word problems they might encounter. Divided into two parts, Part 1 is a description of the Java language, of the fundamentals of object orientation, input and output operations, and error handling. Part 2 is about Java programming for engineers. It starts with computer number systems, fixed- and variable-precision numeric data, mathematical programming in Java as could be of interest to engineers, and concludes with an overview of Java Graphics.

[Java for Data Scientists Essential Training](#) "O'Reilly Media, Inc."

Whether you are a software developer, systems architect, data analyst, or business analyst, if you want to take advantage of data mining in the development of advanced analytic applications, Java Data Mining, JDM, the new standard now implemented in core DBMS and data mining/analysis software, is a key solution component. This book is the essential guide to the usage of the JDM standard interface, written by contributors to the JDM standard. Data mining introduction - an overview of data mining and the problems it can address across industries; JDM's place in strategic solutions to data mining-related problems JDM essentials - concepts, design approach and design issues, with detailed code examples in Java; a Web Services interface to enable JDM functionality in an SOA environment; and illustration of JDM XML Schema for JDM objects JDM in practice - the use of JDM from vendor implementations and approaches to customer applications, integration, and usage; impact of data mining on IT infrastructure; a how-to guide for building applications that use the JDM API Free, downloadable KJDM source code referenced in the book available here

[Reuse in the Big Data Era](#) Packt Publishing Ltd

"Designing Data Structures in Java" provides a solid foundation for anyone seeking to understand the how and the why of programming data structures. Intended for the reader with an introductory Java background, this book aims to meet the needs of students enrolled in a typical "Data Structures and Algorithms with Java" (CS2) course. Starting with a description of the software development process, the book takes a problem-solving approach to programming, and shows how data structures form the building blocks of well-designed and cleanly-implemented programs.

Topics include: Problem solving, Abstraction, Java objects and references, Arrays, Abstract Data Types, Ordered lists, Sorting, Algorithm evaluation, Binary searches, Stacks, Queues, Linked Lists, Double-ended lists, Recursion, Doubly-linked lists, Binary Search Trees, Traversals, Heaps, and more. Mr. Brouillette's 25+ years of experience as a software engineer and educator allow him to bring a unique and refreshing perspective to the topic of data structures which is rigorous, accessible and practical. Material is presented in a 'top down' approach, beginning with explanations of why different data structures are used, continuing with clearly illustrated concepts of how the structures work, and ending with clear, neat Java code examples. Succinct graphics provide visual representations of the ideas, and verbal explanations supplement the documented code. Each chapter ends with a Chapter Checklist summary page which distills and highlights the most important ideas from the chapter. The book is intended as a step by step explanation and exploration of the how and why of using Data Structures in modern computer program development. Even though the Java language is used in the explanation and implementation of the various structures, the concepts are applicable to other languages which the reader may encounter in the future. The topics included have been sequenced to build upon each other, always with the perspective of the beginning programming student in mind. There are discussions of software engineering concepts and goals, and motivations for learning different data structures. This text brings the beginning Java student from novice programmer to the next level of programming maturity.

[Modern Data Engineering with Apache Spark](#) Packt Publishing Ltd

Implement, run, operate, and test data processing pipelines using Apache Beam Key FeaturesUnderstand how to improve usability and productivity when implementing Beam pipelinesLearn how to use stateful processing to implement complex use cases using Apache BeamImplement, test, and run Apache Beam pipelines with the help of expert tips and techniquesBook Description Apache Beam is an open source unified programming model for implementing and executing data processing pipelines, including Extract, Transform, and Load (ETL), batch, and stream processing. This book will help you to confidently build data processing pipelines with Apache Beam. You'll start with an overview of Apache Beam and understand how to use it to implement basic pipelines. You'll also learn how to test and run the pipelines efficiently. As you progress, you'll explore how to structure your code for reusability and also use various Domain Specific Languages (DSLs). Later chapters will show you how to use schemas and query your data using (streaming) SQL. Finally, you'll understand advanced Apache Beam concepts, such as implementing your own I/O connectors. By the end of this book, you'll have gained a deep understanding of the Apache Beam model and be able to apply it to solve problems. What you will learnUnderstand the core concepts and architecture of Apache BeamImplement stateless and stateful data processing pipelinesUse state and timers for processing real-time event processingStructure your code for reusabilityUse streaming SQL to process real-time data for increasing productivity and data accessibilityRun a pipeline using a portable runner and implement data processing using the Apache Beam Python SDKImplement Apache Beam I/O connectors using the Splittable DoFn APIWho this book is for This book is for data engineers, data scientists, and data analysts who want to learn how Apache Beam works. Intermediate-level knowledge of the Java programming language is assumed.

[Data Science with Java](#) Packt Publishing Ltd

Presenting Java details on a need to know basis, this concise introduction to Java develops and executes practical examples useful for solving problems in an engineering environment. KEY

TOPICS Focused on the proper way to write reliable Java programs, it outlines excellent software development practices; uses a top-down design technique to break the program up into logical portions; and demonstrates object re-usability by building later examples on the classes and methods created in earlier examples. The book also includes a number of supporting classes to make it easier for engineers to format numbers in Java, to display data in linear or logarithmic plots, to work with complex numbers, to manipulate arrays, and to perform such common engineering calculations as FFTs, convolutions, correlations, etc. Engineering and Science professionals.

[Java Fundamentals](#) CRC Press

Take advantage of today's sky-high demand for data engineers. With this in-depth book, current and aspiring engineers will learn powerful real-world best practices for managing data big and small. Contributors from notable companies including Twitter, Google, Stitch Fix, Microsoft, Capital One, and LinkedIn share their experiences and lessons learned for overcoming a variety of specific and often nagging challenges. Edited by Tobias Macey, host of the popular Data Engineering Podcast, this book presents 97 concise and useful tips for cleaning, prepping, wrangling, storing, processing, and ingesting data. Data engineers, data architects, data team managers, data scientists, machine learning engineers, and software engineers will greatly benefit from the wisdom and experience of their peers. Topics include: The Importance of Data Lineage - Julien Le Dem Data Security for Data Engineers - Katharine Jarmul The Two Types of Data Engineering and Data Engineers - Jesse Anderson Six Dimensions for Picking an Analytical Data Warehouse - Gleb Mezhanskiy The End of ETL as We Know It - Paul Singman Building a Career as a Data Engineer - Vijay Kiran Modern Metadata for the Modern Data Stack - Prukalpa Sankar Your Data Tests Failed! Now What? - Sam Bail

[Java for Engineers and Scientists](#) Springer

Gain the key language concepts and programming techniques of Scala in the context of big data analytics and Apache Spark. The book begins by introducing you to Scala and establishes a firm contextual understanding of why you should learn this language, how it stands in comparison to Java, and how Scala is related to Apache Spark for big data analytics. Next, you'll set up the Scala environment ready for examining your first Scala programs. This is followed by sections on Scala fundamentals including mutable/immutable variables, the type hierarchy system, control flow expressions and code blocks. The author discusses functions at length and highlights a number of associated concepts such as functional programming and anonymous functions. The book then delves deeper into Scala's powerful collections system because many of Apache Spark's APIs bear a strong resemblance to Scala collections. Along the way you'll see the development life cycle of a Scala program. This involves compiling and building programs using the industry-standard Scala Build Tool (SBT). You'll cover guidelines related to dependency management using SBT as this is critical for building large Apache Spark applications. Scala Programming for Big Data Analytics concludes by demonstrating how you can make use of the concepts to write programs that run on the Apache Spark framework. These programs will provide distributed and parallel computing, which is critical for big data analytics. What You Will LearnSee the fundamentals of Scala as a general-purpose programming languageUnderstand functional programming and object-oriented programming constructs in ScalaUse Scala collections and functions Develop, package and run Apache Spark applications for big data analyticsWho This Book Is For Data scientists, data analysts and data engineers who intend to use Apache Spark for large-scale analytics. /div

Related with Java In Data Engineering:

[© Java In Data Engineering Space Personal Financial Guide](#)

[© Java In Data Engineering Spanish Present Tense Practice](#)

[© Java In Data Engineering Sparknotes For How To Read Literature Like A Professor](#)