The book was found

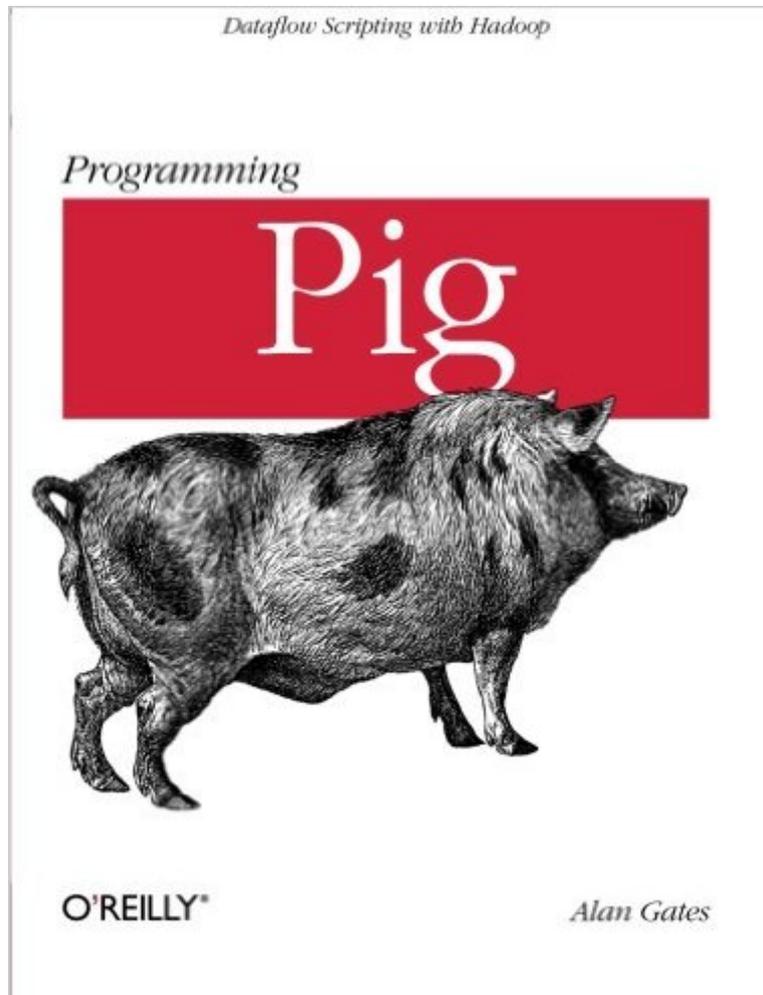# Programming Pig

Dataflow Scripting with Hadoop

Programming

Pig

O'REILLY®

Alan Gates

Synopsis

This guide is an ideal learning tool and reference for Apache Pig, the open source engine for executing parallel data flows on Hadoop. With Pig, you can batch-process data without having to create a full-fledged applicationâ"making it easy for you to experiment with new datasets.Programming Pig introduces new users to Pig, and provides experienced users with comprehensive coverage on key features such as the Pig Latin scripting language, the Grunt shell, and User Defined Functions (UDFs) for extending Pig. If you need to analyze terabytes of data, this book shows you how to do it efficiently with Pig.Delve into Pigâ ™s data model, including scalar and complex data typesWrite Pig Latin scripts to sort, group, join, project, and filter your dataUse Grunt to work with the Hadoop Distributed File System (HDFS)Build complex data processing pipelines with Pigâ ™s macros and modularity featuresEmbed Pig Latin in Python for iterative processing and other advanced tasksCreate your own load and store functions to handle data formats and storage mechanismsGet performance tips for running scripts on Hadoop clusters in less time

Book Information

Paperback: 224 pages

Publisher: O'Reilly Media; 1 edition (October 23, 2011)

Language: English

ISBN-10: 1449302645

ISBN-13: 978-1449302641

Product Dimensions: 7 x 0.8 x 9.2 inches

Shipping Weight: 4 ounces (View shipping rates and policies)

Average Customer Review: 4.0 out of 5 starsÂ Â See all reviewsÂ (16 customer reviews)

Best Sellers Rank: #519,937 in Books (See Top 100 in Books)   #171 inÂ Books > Computers & Technology > Databases & Big Data > Data Warehousing   #255 inÂ Books > Computers & Technology > Databases & Big Data > Data Modeling & Design   #309 inÂ Books > Computers & Technology > Databases & Big Data > Data Mining

Customer Reviews

The book presents an advanced introduction to PIG. Its a book by an insider for insiders and not an introduction to PIG itself. You may end up producing code to run through some data, but may not necessarily gain any understanding.The book reads like a blog. Beyond spell check, it has no editorial oversight whatsover. the content order is adhoc it goes from one topic to another without any apparent continuity. for example, right after a cursory introduction to Map Reduce and PIG, the

discussion goes into an arcane details of commandline and flag settings without any context.the book covers a whole lot of concepts, but the introduction of these concepts itself is weak. For example, projections are introduced as something PIG has in common with SQL.book itself is -2 stars. -1 for 's kindle & oreilly. the publishing quality in this book is horrible. the code fonts smaller than main text, uneven spacing etc. default settings on freely available web publishing softwares produces better content than what and orielly have produced here.

PIG is a powerful programming tool for big data, yet it is simple to write. I got the book to help myself to PIG programming and the book does help with it. If you are new to PIG programming, you will find it useful.I was disappointed that the book only has a cursory reference to piggybank.jar which is a big plus for PIG, and that too at the end with no real examples about it. Also there are tons of PIG examples on the internet which pretty much walk through many scenarios much better than this book .This is a quick reference book, not a PIG Bible

Has some good info especially on how to extend pig functionality, but it leaves a lot to be desired as a primer on how to use pig. Most of the examples are poorly explained. To be honest I learned more about some of the thinking behind pig than I learned about how actually use pig. This book needs a new seriously upgraded edition. It only got three stars because it has no competition.

The company I was working for started using Big Data technologies recently and we were all expected to come up to speed quickly. For PIG, this is the only book I could find. Luckily it's available on Safari, so I didn't have to buy it myself (I didn't feel that the content justified the cost of purchasing this book). You can find a good chunk of the content in this book in online blogs, tutorials, and wikis, but it helped to have this book by my side when we were working on a project since all the relevant information is in a single location. Some of the information in this book is outdated. For instance, it talks about Boolean data type not being supported but the recent versions of PIG do, so be sure to refer to PIG docs from time to time to make sure you have the latest information. PIG is a relatively immature framework - the authors admit this to a certain extent in the book by mentioning that much of the tuning/optimization effort is outsourced to the user (unlike databases which make an effort to optimize queries). This book includes some insights into how to tune for performance (e.g: what types of JOINs to use and when, writing UDFs for performance) which is certainly helpful but the general tone is along the lines "here are somethings to look for but you need to test and find what works best yourself" - in other words, comprehensive examples

illustrating the concepts are missing. Like a lot of other books, some typos exist in this book although it's not too hard to figure out. A basic understanding of hadoop is necessary to use this book but a solid foundation for hadoop is not necessarily required (although it helps a lot). All in all, if you are looking for a single place to refer to for PIG related docs, get this book.

The book is written pretty well, and the examples are clear and easy to follow for the most part. The only lacking aspect in my opinion was a deeper delve into the analytic capabilities for Pig. This is minor, and may actually be a good prompt for a follow-up "Pig Cookbook". Other than that, this is a great reference and has proven very useful.

Good basic primer on pig and its relationship with hadoop. Worth getting though beware that pig like any vital OS project is changing fast.

The book is a very good introduction to Pig written by an insider. It does not assume any previous knowledge of the subject. However, you need some programming experience and familiarity with Hadoop concepts.I don't give it 5 stars only because it is already not quite up-to-date.

It seems to me that all the information in this book is freely available online, and much more easily searchable online. I was disappointed by its brevity and lack of fresh helpful information.

Download to continue reading...

Programming #8:C Programming Success in a Day & Android Programming In a Day! (C Programming, C++programming, C++ programming language, Android , Android Programming, Android Games) Programming #57: C++ Programming Professional Made Easy & Android Programming in a Day (C++ Programming, C++ Language, C++for beginners, C++, Programming ... Programming, Android, C, C Programming) Programming #45: Python Programming Professional Made Easy & Android Programming In a Day! (Python Programming, Python Language, Python for beginners, ... Programming Languages, Android Programming) Guinea Pigs Owner Handbook: The Complete Beginner's Guide to Guinea Pig Care and Facts (How to Care for Guinea Pigs, Guinea Pig Facts Book 1) Guinea Pig Care Secrets: Kids Guide to a Happy Guinea Pig (Kids Pet Care & Guides Book 3) Guinea Pig Pets: Train Your Guinea Pig The Easy Way!: The 7 Day Guide Programming: Computer Programming for Beginners: Learn the Basics of Java, SQL & C++ - 3. Edition (Coding, C Programming, Java Programming, SQL Programming, JavaScript, Python, PHP) Raspberry Pi 2: Raspberry Pi 2 Programming Made Easy (Raspberry Pi, Android Programming,

Programming, Linux, Unix, C Programming, C+ Programming) Android: Programming in a Day! The Power Guide for Beginners In Android App Programming (Android, Android Programming, App Development, Android App Development, ... App Programming, Rails, Ruby Programming) DOS: Programming Success in a Day: Beginners guide to fast, easy and efficient learning of DOS programming (DOS, ADA, Programming, DOS Programming, ADA ... LINUX, RPG, ADA Programming, Android, JAVA) ASP.NET: Programming success in a day: Beginners guide to fast, easy and efficient learning of ASP.NET programming (ASP.NET, ASP.NET Programming, ASP.NET ... ADA, Web Programming, Programming) C#: Programming Success in a Day: Beginners guide to fast, easy and efficient learning of C# programming (C#, C# Programming, C++ Programming, C++, C, C Programming, C# Language, C# Guide, C# Coding) FORTRAN Programming success in a day:Beginners guide to fast, easy and efficient learning of FORTRAN programming (Fortran, Css, C++, C, C programming, ... Programming, MYSQL, SQL Programming) Prolog Programming; Success in a Day: Beginners Guide to Fast, Easy and Efficient Learning of Prolog Programming (Prolog, Prolog Programming, Prolog Logic, ... Programming, Programming Code, Java) R Programming: Learn R Programming In A DAY! - The Ultimate Crash Course to Learning the Basics of R Programming Language In No Time (R, R Programming, ... Course, R Programming Development Book 1) Parallel Programming: Success in a Day: Beginners' Guide to Fast, Easy, and Efficient Learning of Parallel Programming (Parallel Programming, Programming, ... C++ Programming, Multiprocessor, MPI) MYSQL Programming Professional Made Easy 2nd Edition: Expert MYSQL Programming Language Success in a Day for any Computer User! (MYSQL, Android programming, ... JavaScript, Programming, Computer Software) Programming Raspberry Pi 3: Getting Started With Python (Programming Raspberry Pi 3, Raspberry Pi 3 User Guide, Python Programming, Raspberry Pi 3 with Python Programming) VBScript: Programming Success in a Day: Beginner's Guide to Fast, Easy and Efficient Learning of VBScript Programming (VBScript, ADA, ASP.NET, C#, ADA ... ASP.NET Programming, Programming, C++, C) Apps: Mobile App Trends in 2015 (iOS, Xcode Programming, App Development, iOS App Development, App Programming, Swift, Without Coding) ((Android, Android ... App Programming, Rails, Ruby Programming))