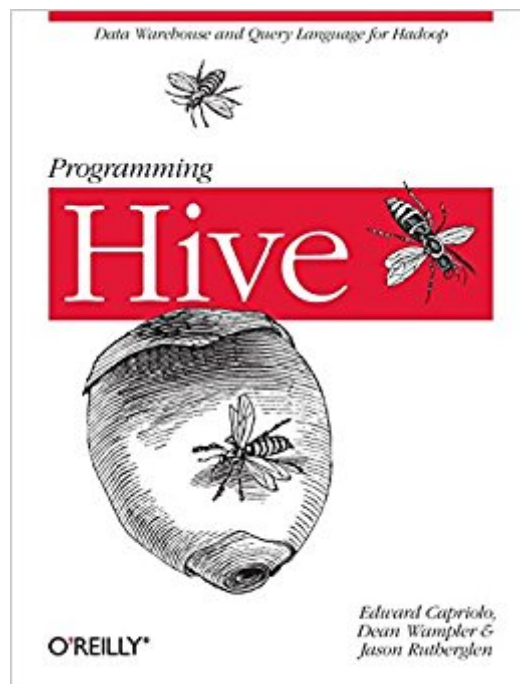The book was found

Programming Hive: Data Warehouse And Query Language For Hadoop

Synopsis

Need to move a relational database application to Hadoop? This comprehensive guide introduces you to Apache Hive, HadoopÃ¢ ¬â„¢s data warehouse infrastructure. YouÃ¢ ¬â„¢ll quickly learn how to use HiveÃ¢ ¬â„¢s SQL dialectÃ¢ ¬â •HiveQLÃ¢ ¬â •to summarize, query, and analyze large datasets stored in HadoopÃ¢ ¬â„¢s distributed filesystem.This example-driven guide shows you how to set up and configure Hive in your environment, provides a detailed overview of Hadoop and MapReduce, and demonstrates how Hive works within the Hadoop ecosystem. YouÃ¢ ¬â„¢ll also find real-world case studies that describe how companies have used Hive to solve unique problems involving petabytes of data.Use Hive to create, alter, and drop databases, tables, views, functions, and indexesCustomize data formats and storage options, from files to external databasesLoad and extract data from tablesÃ¢ ¬â •and use queries, grouping, filtering, joining, and other conventional query methodsGain best practices for creating user defined functions (UDFs)Learn Hive patterns you should use and anti-patterns you should avoidIntegrate Hive with other data processing programsUse storage handlers for NoSQL databases and other datastoresLearn the pros and cons of running Hive on Ã¢ ¬â„¢s Elastic MapReduce

Book Information

Paperback: 350 pages

Publisher: O'Reilly Media; 1 edition (October 6, 2012)

Language: English

ISBN-10: 1449319335

ISBN-13: 978-1449319335

Product Dimensions: 7 x 0.8 x 9.2 inches

Shipping Weight: 1.2 pounds (View shipping rates and policies)

Average Customer Review: 4.4 out of 5 stars   15 customer reviews

Best Sellers Rank: #139,449 in Books (See Top 100 in Books)   #36 inÃ Â Books > Computers & Technology > Databases & Big Data > Data Warehousing   #68 inÃ Â Books > Computers & Technology > Databases & Big Data > SQL   #80 inÃ Â Books > Computers & Technology > Databases & Big Data > Data Modeling & Design

Customer Reviews

Edward Capriolo is currently System Administrator at Media6degrees where he helps design and maintain distributed data storage systems for the internet advertising industry.Edward is a member of the Apache Software Foundation and a committer for the Hadoop-Hive project. He has

experience as a developer as well Linux and network administrator and enjoys the rich world of open source software.Dean Wampler is a Principal Consultant at Think Big Analytics, where he specializes in "Big Data" problems and tools like Hadoop and Machine Learning. Besides Big Data, he specializes in Scala, the JVM ecosystem, JavaScript, Ruby, functional and object-oriented programming, and Agile methods. Dean is a frequent speaker at industry and academic conferences on these topics. He has a Ph.D. in Physics from the University of Washington.Jason Rutherglen is a software architect at Think Big Analytics and specializes in Big Data, Hadoop, search, and security.

I had joined Nextag a few months before this book got released and my solo-artist project was to convert all MySQL aggregations to Hadoop. I was thinking MR Java, HBase - and then I started reading this book. Start to end it took me less than a week and I decided that I want to give Hive a shot. It was possibly Hive 0.8 in late 2012 if my memory serves me right. I joined the Hive user group and also had the good fortune of communicating with two of the authors Capriolo and Wampler ad seeking clarifications. I don't want to go into the details of how well written this book is. Thye developers who read this review will get it when that reading this book and actively participating in the Hive community helped me a great deal to execute a project all the way from start to production - all by myself...If the authors want to publish a V2 , I would love to write a case study :-)

Very help and easy to use!

Good book - I used it to study for a presentation I needed to give our group at work. It's needing an update with Impala/Stinger/Presto now available to improve performance.

Well written, good book. However how fast Hive/hadoop is moving, I think this book will be out of data (already is in parts) soon. I'm assuming they'll come out with updated v2 of the book in the future.

Similar to my review on the Hadoop book, if, you are planning on taking the Hadoop Developer certification, this book is a good reference to read before taking the exam.

It's a great resource, and remains on my desk; however it already feels outdated (bought in October

2012.) There are many commands and functions that aren't in this book for whatever reason.Still, a great start - the floodgates should be open to Hive reference material!

this give good information with some good example.good for basic.but it's little old.example might not work. but should give good start for starter

Good overview and provided a good working knowledge of the queries for Hadoop.

Programming Hive: Data Warehouse and Query Language for Hadoop Warehouse Management: A Complete Guide to Improving Efficiency and Minimizing Costs in the Modern Warehouse Data Analytics and Python Programming: 2 Bundle Manuscript: Beginners Guide to Learn Data Analytics, Predictive Analytics and Data Science with Python Programming Big Data For Business: Your Comprehensive Guide to Understand Data Science, Data Analytics and Data Mining to Boost More Growth and Improve Business - Data Analytics Book, Series 2 Data Analytics: What Every Business Must Know About Big Data And Data Science (Data Analytics for Business, Predictive Analysis, Big Data Book 1) Data Analytics: Applicable Data Analysis to Advance Any Business Using the Power of Data Driven Analytics (Big Data Analytics, Data Science, Business Intelligence Book 6) Python Programming: Python Programming for Beginners, Python Programming for Intermediates, Python Programming for Advanced C++: The Ultimate Crash Course to Learning the Basics of C++ (C programming, C++ in easy steps, C++ programming, Start coding today) (CSS,C Programming, ... Programming,PHP, Coding, Java Book 1) Analytics: Data Science, Data Analysis and Predictive Analytics for Business (Algorithms, Business Intelligence, Statistical Analysis, Decision Analysis, Business Analytics, Data Mining, Big Data) The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling SQL for Beginners: Learn the Structured Query Language for the Most Popular Databases including Microsoft SQL Server, MySQL, MariaDB, PostgreSQL, and Oracle Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale C++ and Python Programming: 2 Manuscript Bundle: Introductory Beginners Guide to Learn C++ Programming and Python Programming C++ and Python Programming 2 Bundle Manuscript. Introductory Beginners Guide to Learn C++ Programming and Python Programming Python Programming: The Complete Step By Step Guide to Master Python Programming and Start Coding Today! (Computer Programming Book 4) Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data Data Analytics For Beginners: Your Ultimate Guide To Learn and Master Data Analysis. Get Your Business Intelligence Right Ã¢â ¬â œ Accelerate Growth and Close More Sales

(Data Analytics Book Series) Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking Analytics: Business Intelligence, Algorithms and Statistical Analysis (Predictive Analytics, Data Visualization, Data Analytics, Business Analytics, Decision Analysis, Big Data, Statistical Analysis) Discovering Knowledge in Data: An Introduction to Data Mining (Wiley Series on Methods and Applications in Data Mining)