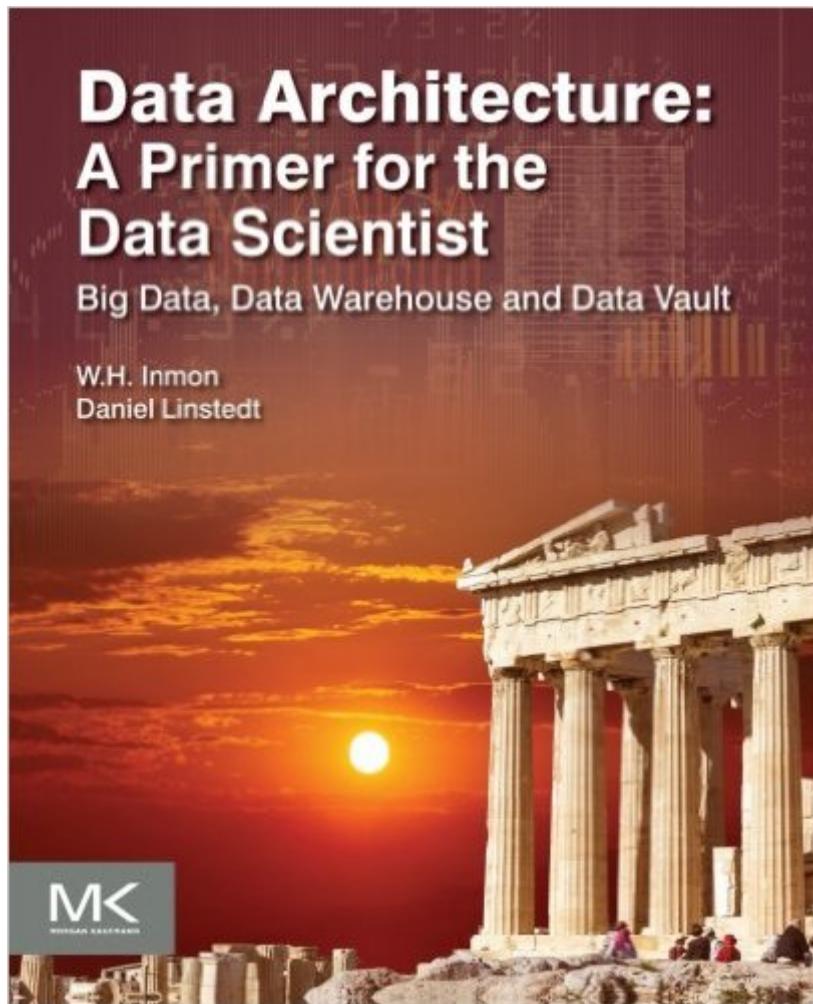


The book was found

# Data Architecture: A Primer For The Data Scientist: Big Data, Data Warehouse And Data Vault



## Synopsis

Today, the world is trying to create and educate data scientists because of the phenomenon of Big Data. And everyone is looking deeply into this technology. But no one is looking at the larger architectural picture of how Big Data needs to fit within the existing systems (data warehousing systems). Taking a look at the larger picture into which Big Data fits gives the data scientist the necessary context for how pieces of the puzzle should fit together. Most references on Big Data look at only one tiny part of a much larger whole. Until data gathered can be put into an existing framework or architecture it can't be used to its full potential. Data Architecture a Primer for the Data Scientist addresses the larger architectural picture of how Big Data fits with the existing information infrastructure, an essential topic for the data scientist. Drawing upon years of practical experience and using numerous examples and an easy to understand framework. W.H. Inmon, and Daniel Linstedt define the importance of data architecture and how it can be used effectively to harness big data within existing systems. You'll be able to:

- Turn textual information into a form that can be analyzed by standard tools.
- Make the connection between analytics and Big Data
- Understand how Big Data fits within an existing systems environment
- Conduct analytics on repetitive and non-repetitive data
- Discusses the value in Big Data that is often overlooked, non-repetitive data, and why there is significant business value in using it
- Shows how to turn textual information into a form that can be analyzed by standard tools
- Explains how Big Data fits within an existing systems environment
- Presents new opportunities that are afforded by the advent of Big Data
- Demystifies the murky waters of repetitive and non-repetitive data in Big Data

## Book Information

Paperback: 378 pages

Publisher: Morgan Kaufmann; 1 edition (December 10, 2014)

Language: English

ISBN-10: 012802044X

ISBN-13: 978-0128020449

Product Dimensions: 7.5 x 0.9 x 9.2 inches

Shipping Weight: 1.8 pounds (View shipping rates and policies)

Average Customer Review: 3.9 out of 5 stars [See all reviews](#) (14 customer reviews)

Best Sellers Rank: #264,754 in Books (See Top 100 in Books) #67 in Books > Computers & Technology > Databases & Big Data > Data Warehousing #90 in Books > Computers & Technology > Networking & Cloud Computing > Network Administration > Storage & Retrieval

## Customer Reviews

Putting 'primer' in the title should warn you not to expect too much. Bill Inmon used to deliver more than that. The problem with a primer is that the authors don't have to justify, exemplify or detail anything. Things are presented like this and you have no place to make a choice. It's not even take it or leave it, it's only take it. I mean most of the things look correct if you apply them and you happen to have the chance to have a situation where it fits. If you don't fit, you have no escape. A primer should present only clear simple concepts that are recognized throughout the community and ALL the concepts pertinent to the title. Imagine a data warehouse book where slow changing dimension is not mentioned, nor bitemporality, CWM, metamodel. OLAP is only mentioned in the glossary. Imagine a data architecture book where the words cartesian, constraints, enumeration or domain are not used. Even conceptual model is not used in the standard meaning. Those are cues that all the territory is not covered. I would not recommend this book for a university student, a data professional or a data scientist. Just look at the glossary to convince you. A data model is defined as "an abstraction of data". DW 2.0 is defined as "the second-generation data warehouse architecture". MapReduce is defined as "a language for processing Big Data". A relational model is defined as "a form of data where data is normalized". Even Wikipedia can do better than that. Why putting terms in a glossary in a book if the terms are less precisely defined and/or do not help to contextualize the terms with the subject of the book. It leaves a bad taste for the rest of the book (The semantics may be loose, imprecise with many shortcuts and confusion).

DATA ARCHITECTURE â “ A PRIMER FOR THE DATA SCIENTIST Elsevier Morgan Kaufman This book is not for everyone. If you are looking for a detailed technical book this book is not for you. If you are looking for a rehash of old ideas and concepts that relate to specific subjects such as data warehouse or Big Data then you need to look elsewhere. Instead this book is an architecture book. (If you are a technician that does not understand or appreciate architecture, then you will find this book sort of unintelligible.) And the breadth and scope of this book is beyond anything found in the literature of computer science. First and foremost the book addresses corporate data, in its entirety. All other books address some small aspect of corporate data. But this book is unique in that it addresses the broadest perspective of data. The book handles subjects not found anywhere, such as the fundamental divide between repetitive and non repetitive data. (Hardly any other technical book even recognizes that there is repetitive and non repetitive data.) And the book addresses the

vital topic â€“ how does business value relate to repetitive data and non repetitive data. The book makes the profound point that there is an extreme divide between repetitive data and non repetitive data. This divide is called the â€œGreat Divideâ€. Another aspect of this book that is found nowhere else is that of the explanation of textual disambiguation. It is through textual disambiguation that context of data is discovered. Unlike NLP which is essentially context free, textual disambiguation focusses in on the importance of context in trying to use text for analytical processing.

[Download to continue reading...](#)

Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault  
The Data Warehouse Mentor: Practical Data Warehouse and Business Intelligence Insights  
Building a Scalable Data Warehouse with Data Vault 2.0  
Big Data For Beginners: Understanding SMART Big Data, Data Mining & Data Analytics For improved Business Performance, Life Decisions & More!  
The Data Warehouse Lifecycle Toolkit  
The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling  
Kimball's Data Warehouse Toolkit Classics: 3 Volume Set  
Mastering Data Warehouse Aggregates: Solutions for Star Schema Performance  
Big Data, MapReduce, Hadoop, and Spark with Python: Master Big Data Analytics and Data Wrangling with MapReduce Fundamentals using Hadoop, Spark, and Python  
The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences  
Understanding Cloud, IoT and Big data (Cloud, IoT & Big Data: Basic To AWS SA Professional Book 1)  
Primer of Biostatistics, Seventh Edition (Primer of Biostatistics (Glantz)(Paperback))  
Primer Diario Nana: Mi primer Diario de Susana (Volume 2) (Spanish Edition)  
Primer on the Rheumatic Diseases (Primer on Rheumatic Diseases (Klippel))  
Primer Diario Rosy: mi primer Diario (Volume 1) (Spanish Edition)  
PokÃ©mon Sun and PokÃ©mon Moon: Official Strategy Guide  
Collector's Vault  
Holy Serpent of the Jews: The Rabbis' Secret Plan for Satan to Crush Their Enemies and Vault the Jews to Global Dominion  
Fallout 4 Vault Dweller's Survival Guide: Prima Official Game Guide  
Fallout 4 Vault Dweller's Survival Guide Collector's Edition: Prima Official Game Guide  
ScrewAttack's Video Game Vault: The Best of Nintendo 64

[Dmca](#)