Data Architecture: A Primer for the Data Scientist

This book is for all data modelers, data architects, and database designers, whether they be novices who want to learn what’s involved in data modeling, or experienced modelers who want to brush up their skills. A novice will not only gain an overview of data modeling, they will also learn how to follow the data modeling process, including the activities required for each step. The experienced practitioner will discover (or rediscover) techniques to ensure that data models accurately reflect business requirements. This book describes rigorous yet easily implemented approaches to:

- Modeling of business information requirements for review by business stakeholders before development of the logical data model
- Normalizing data, based on simple questions rather than the formal definitions which many modelers find intimidating
- Naming and defining concepts and attributes
- Modeling of time-variant data
- Documenting business rules governing both the real world and data
- Data modeling in an Agile project
- Managing data model change in any type of project
- Transforming a business information model to a logical data model against which developers can code
- Implementing the logical data model in a traditional relational DBMS, an SQL:2003-compliant DBMS, an object-relational DBMS, or in XML.

Part 1 describes business information models in-depth, including:

- The importance of modeling business information requirements before embarking on a logical data model
- Business concepts (entity classes)
- Attributes of business concepts
- Attribute classes as an alternative to DBMS data types
- Relationships between business concepts
- Time-variant data
- Generalization and specialization of business concepts
- Naming and defining the components of the business information model
- Business rules governing data, including a distinction between real-world rules and data rules.

Part 2 journeys from requirements to a working data resource, covering:

- Sourcing data requirements
- Developing the business information model
- Communicating it to business stakeholders for review, both as diagrams and verbally
- Managing data model change
- Transforming the business information model into a logical data model of stored data for implementation in a relational or object-relational DBMS
- Attributes
value representation and data constraints (important but often overlooked) · modeling data vault, dimensional and XML data.

Agile Data Warehousing for the Enterprise The need to handle increasingly larger data volumes is one factor driving the adoption of a new class of nonrelational “NoSQL” databases. Advocates of NoSQL databases claim they can be used to build systems that are more performant, scale better, and are easier to program. NoSQL Distilled is a concise but thorough introduction to this rapidly emerging technology. Pramod J. Sadalage and Martin Fowler explain how NoSQL databases work and the ways that they may be a superior alternative to a traditional RDBMS. The authors provide a fast-paced guide to the concepts you need to know in order to evaluate whether NoSQL databases are right for your needs and, if so, which technologies you should explore further. The first part of the book concentrates on core concepts, including schemaless data models, aggregates, new distribution models, the CAP theorem, and map-reduce. In the second part, the authors explore architectural and design issues associated with implementing NoSQL. They also present realistic use cases that demonstrate NoSQL databases at work and feature representative examples using Riak, MongoDB, Cassandra, and Neo4j. In addition, by drawing on Pramod Sadalage's pioneering work, NoSQL Distilled shows how to implement evolutionary design with schema migration: an essential technique for applying NoSQL databases. The book concludes by describing how NoSQL is ushering in a new age of Polyglot Persistence, where multiple data-storage worlds coexist, and architects can choose the technology best optimized for each type of data access.

Business Intelligence and Big Data Understand critical cybersecurity and risk perspectives, insights, and tools for the leaders of complex financial systems and markets. This book offers guidance for decision makers and helps establish a framework for communication between cyber leaders and front-line professionals. Information is provided to help in the analysis of cyber challenges and choosing between risk treatment options. Financial
cybersecurity is a complex, systemic risk challenge that includes technological and operational elements. The interconnectedness of financial systems and markets creates dynamic, high-risk environments where organizational security is greatly impacted by the level of security effectiveness of partners, counterparties, and other external organizations. The result is a high-risk environment with a growing need for cooperation between enterprises that are otherwise direct competitors. There is a new normal of continuous attack pressures that produce unprecedented enterprise threats that must be met with an array of countermeasures. Financial Cybersecurity Risk Management explores a range of cybersecurity topics impacting financial enterprises. This includes the threat and vulnerability landscape confronting the financial sector, risk assessment practices and methodologies, and cybersecurity data analytics. Governance perspectives, including executive and board considerations, are analyzed as are the appropriate control measures and executive risk reporting. What You’ll Learn Analyze the threat and vulnerability landscape confronting the financial sector Implement effective technology risk assessment practices and methodologies Craft strategies to treat observed risks in financial systems Improve the effectiveness of enterprise cybersecurity capabilities Evaluate critical aspects of cybersecurity governance, including executive and board oversight Identify significant cybersecurity operational challenges Consider the impact of the cybersecurity mission across the enterprise Leverage cybersecurity regulatory and industry standards to help manage financial services risks Use cybersecurity scenarios to measure systemic risks in financial systems environments Apply key experiences from actual cybersecurity events to develop more robust cybersecurity architectures Who This Book Is For Decision makers, cyber leaders, and front-line professionals, including: chief risk officers, operational risk officers, chief information security officers, chief security officers, chief information officers, enterprise risk managers, cybersecurity operations directors, technology and cybersecurity risk analysts, cybersecurity architects and engineers, and compliance officers
Non-Invasive Data Governance "There's only one right way to begin an Oracle Designer project: with a detailed standards document from which your whole team can work." "Oracle Designer: A Template for Developing an Enterprise Standards Document brings you: expert techniques for analyzing, developing, and "blueprinting" any enterprise application; a standards template for system development that can be tailored to meet your organization's needs; proven processes for building and improving your own standards documents; system development standards and naming conventions for all major Designer repository objects with detailed explanations of suggested standards diagramming conventions; and white papers on BPM and FHD interactions and Reverse Engineering."--BOOK JACKET.Title Summary field provided by Blackwell North America, Inc. All Rights Reserved

Pro Oracle Database 11g RAC on Linux The Data Vault was invented by Dan Linstedt at the U.S. Department of Defense, and the standard has been successfully applied to data warehousing projects at organizations of different sizes, from small to large-size corporations. Due to its simplified design, which is adapted from nature, the Data Vault 2.0 standard helps prevent typical data warehousing failures. "Building a Scalable Data Warehouse" covers everything one needs to know to create a scalable data warehouse end to end, including a presentation of the Data Vault modeling technique, which provides the foundations to create a technical data warehouse layer. The book discusses how to build the data warehouse incrementally using the agile Data Vault 2.0 methodology. In addition, readers will learn how to create the input layer (the stage layer) and the presentation layer (data mart) of the Data Vault 2.0 architecture including implementation best practices. Drawing upon years of practical experience and using numerous examples and an easy to understand framework, Dan Linstedt and Michael Olschimke discuss: How to load each layer using SQL Server Integration Services (SSIS), including automation of the Data Vault loading processes. Important data warehouse technologies and practices. Data Quality Services (DQS) and Master Data Services (MDS) in the context of the Data Vault
Read Book The Official Data Vault Standards
Document Version 1.0 Data Warehouse Architecture.

Provides a complete introduction to data warehousing, applications, and the business context so readers can get-up and running fast. Explains theoretical concepts and provides hands-on instruction on how to build and implement a data warehouse. Demystifies data vault modeling with beginning, intermediate, and advanced techniques. Discusses the advantages of the data vault approach over other techniques, also including the latest updates to Data Vault 2.0 and multiple improvements to Data Vault 1.0.

The Elephant in the Fridge: Guided Steps to Data Vault Success through Building Business-Centered Models Where do ideas that reach policy makers and planners as proposals for Data vault strengthening and reform actually originate? To what extent does management recognize Data vault as a tool to increase the results? How can you negotiate Data vault successfully with a stubborn boss, an irate client, or a deceitful coworker? Who will be responsible for deciding whether Data vault goes ahead or not after the initial investigations? Does Data vault systematically track and analyze outcomes for accountability and quality improvement?

Defining, designing, creating, and implementing a process to solve a challenge or meet an objective is the most valuable role in EVERY group, company, organization and department. Unless you are talking a one-time, single-use project, there should be a process. Whether that process is managed and implemented by humans, AI, or a combination of the two, it needs to be designed by someone with a complex enough perspective to ask the right questions. Someone capable of asking the right questions and step back and say, 'What are we really trying to accomplish here? And is there a different way to look at it?' This Self-Assessment empowers people to do just that - whether their title is entrepreneur, manager, consultant, (Vice-)President, CxO etc - they are the people who rule the future. They are the person who asks the right questions to make Data vault investments work better. This Data vault All-Inclusive Self-Assessment enables You to be that person. All the tools you need to an in-depth Data vault Self-Assessment. Featuring new and updated case-based questions, organized into seven core areas of process design, this
Self-Assessment will help you identify areas in which Data vault improvements can be made. In using the questions you will be better able to: - diagnose Data vault projects, initiatives, organizations, businesses and processes using accepted diagnostic standards and practices - implement evidence-based best practice strategies aligned with overall goals - integrate recent advances in Data vault and process design strategies into practice according to best practice guidelines Using a Self-Assessment tool known as the Data vault Scorecard, you will develop a clear picture of which Data vault areas need attention. Your purchase includes access details to the Data vault self-assessment dashboard download which gives you your dynamically prioritized projects-ready tool and shows your organization exactly what to do next. Your exclusive instant access details can be found in your book.

Data Processing and Modeling with Hadoop Building upon his earlier book that detailed agile data warehousing programming techniques for the Scrum master, Ralph's latest work illustrates the agile interpretations of the remaining software engineering disciplines: Requirements management benefits from streamlined templates that not only define projects quickly, but ensure nothing essential is overlooked. Data engineering receives two new "hyper modeling" techniques, yielding data warehouses that can be easily adapted when requirements change without having to invest in ruinously expensive data-conversion programs. Quality assurance advances with not only a stereoscopic top-down and bottom-up planning method, but also the incorporation of the latest in automated test engines. Use this step-by-step guide to deepen your own application development skills through self-study, show your teammates the world's fastest and most reliable techniques for creating business intelligence systems, or ensure that the IT department working for you is building your next decision support system the right way. Learn how to quickly define scope and architecture before programming starts Includes techniques of process and data engineering that enable iterative and incremental delivery Demonstrates how to plan and execute quality assurance plans and includes a guide to continuous integration and
automated regression testing Presents program management strategies for coordinating multiple agile data mart projects so that over time an enterprise data warehouse emerges Use the provided 120-day road map to establish a robust, agile data warehousing program

Data Modeling for Quality Here you will learn how to develop an attractive, easily readable, conceptual, business-oriented entity/relationship model, using a variation on the UML Class Model notation. This book has two audiences: • Data modelers (both analysts and database designers) who are convinced that UML has nothing to do with them; and • UML experts who don’t realize that architectural data modeling really is different from object modeling (and that the differences are important). David Hay’s objective is to finally bring these two groups together in peace. Here all modelers will receive guidance on how to produce a high quality (that is, readable) entity/relationship model to describe the data architecture of an organization. The notation involved happens to be the one for class models in the Unified Modeling Language, even though UML was originally developed to support object-oriented design. Designers have a different view of the world from those who develop business-oriented conceptual data models, which means that to use UML for architectural modeling requires some adjustments. These adjustments are described in this book. David Hay is the author of Enterprise Model Patterns: Describing the World, a comprehensive model of a generic enterprise. The diagrams were at various levels of abstraction, and they were all rendered in the slightly modified version of UML Class Diagrams presented here. This book is a handbook to describe how to build models such as these. By way of background, an appendix provides a history of the two groups, revealing the sources of their different attitudes towards the system development process. If you are an old-school ER modeler and now find yourself having to come up to speed on UML to get that next job (or keep the current one), this is your guidebook to success. If you are a long time object oriented programmer who has to interact with data modelers, this book is for you too. David has done the hard work of
mapping out how to do a logical entity relationship model using standard (and accepted) UML diagram components. This book shows you step-by-step, with ample examples, how to get from here to there with the least pain possible for all concerned. Kent Graziano Certified Data Vault Master and Oracle ACE Past-President of ODTUG & RMOUG Brilliantly organized: three books hidden in one cohesive work. Notwithstanding the tremendous value provided by cross-training data architects/modelers and object modelers/architects, making each better at what they do, Appendix B presents an absolutely awesome concise, yet detailed, history of modeling objects and data that clearly documents the differences in the approaches over the years and helps bring it all into perspective. This book is packed with useful information. Even the footnotes add clarity and offer interesting and often humorous editorial insight making it a fun read. Whatever viewpoint the reader is coming from this book has something to offer as long as the reader maintains an open mind. Roland Berg Senior Architect Diligent Consulting, Inc. San Antonio, Texas

Data Model Scorecard The data vault methodology presents a unique opportunity to model the enterprise data warehouse using the same automation principles applicable in today's software delivery, continuous integration, continuous delivery and continuous deployment while still maintaining the standards expected for governing a corporation's most valuable asset: data. This book provides at first the landscape of a modern architecture and then as a thorough guide on how to deliver a data model that flexes as the enterprise flexes, the data vault. Whether the data is structured, semi-structured or even unstructured one thing is clear, there is always a model either applied early (schema-on-write) or applied late (schema-on-read). Today's focus on data governance requires that we know what we retain about our customers, the data vault provides that focus by delivering a methodology focused on all aspects about the customer and provides some of the best practices for modern day data compliance. The book will delve into every data vault modelling artefact, its automation with sample code, raw vault, business vault, testing framework, a build
framework, sample data vault models, how to build automation patterns on top of a data vault and even offer an extension of data vault that provides automated timeline correction, not to mention variation of data vault designed to provide audit trails, metadata control and integration with agile delivery tools.

Government Contracts Reporter This book constitutes the refereed proceedings of the 38th International Conference on Conceptual Modeling, ER 2019, held in Salvador, Brazil, in November 2019. The 22 full and 22 short papers presented together with 4 keynotes were carefully reviewed and selected from 142 submissions. This events covers a wide range of topics, covered in the following sessions: conceptual modeling, big data technology I, process modeling and analysis, query approaches, big data technology II, domain specific models I, domain specific models II, decision making, complex systems modeling, model unification, big data technology III, and requirements modeling.

The Business of Data Vault Modeling Data-governance programs focus on authority and accountability for the management of data as a valued organizational asset. Data Governance should not be about command-and-control, yet at times could become invasive or threatening to the work, people and culture of an organization. Non-Invasive Data Governance™ focuses on formalizing existing accountability for the management of data and improving formal communications, protection, and quality efforts through effective stewarding of data resources. Non-Invasive Data Governance will provide you with a complete set of tools to help you deliver a successful data governance program. Learn how: • Steward responsibilities can be identified and recognized, formalized, and engaged according to their existing responsibility rather than being assigned or handed to people as more work. • Governance of information can be applied to existing policies, standard operating procedures, practices, and methodologies, rather than being introduced or emphasized as new processes or methods. • Governance of information can support all data integration, risk management, business intelligence and master data management
activities rather than imposing inconsistent rigor to these initiatives.
• A practical and non-threatening approach can be applied to
governing information and promoting stewardship of data as a
cross-organization asset. • Best practices and key concepts of this
non-threatening approach can be communicated effectively to
leverage strengths and address opportunities to improve.

Corporate Information Factory “Get it done well and get it done
fast” are twin, apparently opposing, demands. Data architects are
increasingly expected to deliver quality data models in challenging
timeframes, and agile developers are increasingly expected to
ensure that their solutions can be easily integrated with the data
assets of the overall organization. If you need to deliver quality
solutions despite exacting schedules, “The Nimble Elephant” will
help by describing proven techniques that leverage the libraries of
published data model patterns to rapidly assemble extensible and
robust designs. The three sections in the book provide guidelines
for applying the lessons to your own situation, so that you can
apply the techniques and patterns immediately to your current
assignments. The first section, Foundations for Data Agility,
addresses some perceived aspects of friction between “data” and
“agile” practitioners. As a starting point for resolving the
differences, pattern levels of granularity are classified, and their
interdependencies exposed. A context of various types of models is
established (e.g. conceptual / logical / physical, and industry /
enterprise / project), and you will learn how to customize patterns
within specific model types. The second section, Steps Towards
Data Agility, shares guidelines on generalizing and specializing,
with cautions on the dangers of going too far. Creativity in using
patterns beyond their intended purpose is encouraged. The short-
term “You Ain’t Gonna Need It” (YAGNI) philosophy of agile
practitioners, and the longer-term strategic perspectives of
architects, are compared and evaluated. Consideration is given to
the potential of enterprise views contributing to project-specific
models. Other topics include industry models, iterative modeling,
creation of patterns when none exist, and patterns for rules-in-data.
The section ends with a perspective on the modeler’s possible role
Read Book The Official Data Vault Standards
Document Version 1 0 Data Warehouse
Architecture

in agile projects, followed by a case study. The final section, A
Bridge to the Land of Object Orientation, provides a pathway for re-
skilling traditional data modelers who want to expand their options
by actively engaging with the ranks of object-oriented developers.
I’m delighted to see that John has put his extensive experience and
broad knowledge of data modeling into print! John’s ability to
simplify the complex, and to share his knowledge and enthusiasm –
and humor – with colleagues, comes through in this very useful and
readable book. I recommend it to anyone working with data. —
Monika Remenyi, Senior Data Architect, Telstra

John Giles has
written a compelling and engaging book about the importance of
data modeling patterns in the world of agile computing. His book is
clearly and simply written, and it is full of excellent examples drawn
from his extensive experience as a practitioner. You will see the
enthusiasm and passion that John clearly has for his work in data
modeling. And you will see in his book that any interchange with
John will always have its fair share of good humor and wisdom! —
Professor Ron Weber, Dean, Faculty of IT, Monash University

Data Vault Modeling Complete Self-Assessment Guide
You want the rigor of good data architecture at the speed of agile? Then this
is the missing link - your step-by-step guide to Data Vault success.
Success with a Data Vault starts with the business and ends with
the business. Sure, there’s some technical stuff in the middle, and
it is absolutely essential - but it’s not sufficient on its own. This book
will help you shape the business perspective, and weave it into the
more technical aspects of Data Vault modeling. You can read the
foundational books and go on courses, but one massive risk still
remains. Dan Linstedt, the founder of the Data Vault, very clearly
directs those building a Data Vault to base its design on an
“enterprise ontology”. And Hans Hultgren similarly stresses the
importance of the business concepts model. So it’s important. We
get that. But: What on earth is an enterprise ontology/business
concept model, ‘cause I won’t know if I’ve got one if I don’t know
what I’m looking for? If I can’t find one, how do I get my hands on
such a thing? Even if I have one of these wonderful things, how do
I apply it to get the sort of Data Vault that’s recommended? It’s
actually not as hard as some would fear to answer all of these questions, and it’s certainly worth the effort. This book just might save you a world of pain. It’s a supplement to other material on Data Vault modeling, but it’s the vital missing link to finding simplicity for Data Vault success. “Data Warehousing in the context of large healthcare organizations is notoriously difficult – largely in part due to challenges around Health Information Modeling. The Elephant in the Fridge provides clear and rational perspectives that have allowed my team to breakthrough some of our circular debates and ‘too hard basket’ challenges. Success in this space will require technical and health professionals to co-design solutions around a shared information model. This book is written in a way that they can both consume. John’s experience and emphasis on art of Information Modelling is a welcome complement to other works on Data Vault. I wish we had this book a year ago!” Benson Choy, Enterprise Information Architect, eHealth Queensland. “John has a wonderful way of explaining complicated topics in an uncomplicated way. If you’ve heard about taxonomies and are wondering how to apply these to your Data Warehouse, then this is the book for you. Data modelling is about the business, and John explains how this can be achieved by using proven business templates which help to quickly and efficiently define a solid data model. Don’t re-invent the wheel, but start with the model patterns that are already available! This is a highly relevant book, because it helps to match the information delivery to business expectations - by correctly applying taxonomies and (business) model archetypes in a clear and simple way.” Roelant Vos, General Manager - Enterprise Data Management “It has been a delight and privilege working with John on a Data Vault project. I have enjoyed his practical and pragmatic approach and readiness to share his wealth of knowledge. This is an excellent book written based on real life experience. It provides valuable and timely insights into Data Vault modeling and delivery, and can be easily understood by those who are new to the Data Vault. The book contains practical advice and sample patterns to help people get started on a Data Vault project and avoid costly mistakes.” Natalia Bulashenko, Information Solution Architect “This book covers off one of the key
aspects of Data Vault – the Modelling. Far too many DV practitioners come from a technical database background and lack the fundamental Data Modelling skills necessary to identify the “Key Business Concepts” that are core to successful DV projects. Far too many of us start with the low hanging fruit that we have available to us – the source systems. John explains why this is an extremely bad approach with examples, identifying the signs that the project is heading into trouble and then describes what needs to be done to get back on track. This book is NOT about the implementation phase, it’s about setting the corner stone for the whole project. Get this right and you are on the road to a successful project.” Peter Dudley, Consultant, Interactive Innovations

Advances in Databases and Information Systems

Treasury, Postal Service, and General Government Appropriations for Fiscal Year 1996: United States Postal Service, Executive Office of the President What amount of memory is required? Who gets access to the data infrastructure, and by what means? How are integrated data products disseminated? Does the solution support monitoring of data collection, cleaning, and integration activities? What are your industry data models and where do you find them? Defining, designing, creating, and implementing a process to solve a challenge or meet an objective is the most valuable role in EVERY group, company, organization and department. Unless you are talking a one-time, single-use project, there should be a process. Whether that process is managed and implemented by humans, AI, or a combination of the two, it needs to be designed by someone with a complex enough perspective to ask the right questions. Someone capable of asking the right questions and step back and say, 'What are we really trying to accomplish here? And is there a different way to look at it?' This Self-Assessment empowers people to do just that - whether their title is entrepreneur, manager, consultant, (Vice-)President, CxO etc - they are the people who rule the future. They are the person who asks the right questions to make Data Vault Modeling
Investments work better. This Data Vault Modeling All-Inclusive Self-Assessment enables You to be that person. All the tools you need to an in-depth Data Vault Modeling Self-Assessment. Featuring 905 new and updated case-based questions, organized into seven core areas of process design, this Self-Assessment will help you identify areas in which Data Vault Modeling improvements can be made. In using the questions you will be better able to: - diagnose Data Vault Modeling projects, initiatives, organizations, businesses and processes using accepted diagnostic standards and practices - implement evidence-based best practice strategies aligned with overall goals - integrate recent advances in Data Vault Modeling and process design strategies into practice according to best practice guidelines Using a Self-Assessment tool known as the Data Vault Modeling Scorecard, you will develop a clear picture of which Data Vault Modeling areas need attention. Your purchase includes access details to the Data Vault Modeling self-assessment dashboard download which gives you your dynamically prioritized projects-ready tool and shows your organization exactly what to do next. You will receive the following contents with New and Updated specific criteria: - The latest quick edition of the book in PDF - The latest complete edition of the book in PDF, which criteria correspond to the criteria in - The Self-Assessment Excel Dashboard - Example pre-filled Self-Assessment Excel Dashboard to get familiar with results generation - In-depth and specific Data Vault Modeling Checklists - Project management checklists and templates to assist with implementation INCLUDES LIFETIME SELF ASSESSMENT UPDATES Every self assessment comes with Lifetime Updates and Lifetime Free Updated Books. Lifetime Updates is an industry-first feature which allows you to receive verified self assessment updates, ensuring you always have the most accurate information at your fingertips.

Data Model Patterns With the aim of simplifying relational database modeling, Database Modeling Step-by-Step presents the standard approach to database normalization and then adds its own approach, which is a more simplistic, intuitive way to building relational database models. Going from basics to contemporary
topics, the book opens with relational data modeling and ends with BigData database modeling following a road map of the evolution in relational modeling and including brief introductions to data warehousing and BigData modeling. A break-down of the elements of a model explains what makes up a relational data model. This is followed by a comparison between standard normalization and a more simplistic intuitive approach to data modeling that a beginner can follow and understand. A brief chapter explains how to use the database programming language SQL (Structured Query Language), which reads from and writes to a relational database. SQL is fundamental to data modeling because it helps in understanding how the model is used. In addition to the relational model, the last three chapters cover important modern world topics including denormalization that leads into data warehouses and BigData database modeling. The book explains how there is not much to logical data modeling in BigData databases because as they are often schema-less, which means that BigData databases do not have schemas embedded into the database itself, they have no metadata and thus not much of a logical data model. Online bonus chapters include a case study that covers relational data modeling and are available at the author’s web site: www.oracletroubleshooter.com/datamodeling.html

Data Architecture: A Primer for the Data Scientist Do You Know If Your Data Warehouse Flexible, Scalable, Secure and Will It Stand The Test Of Time And Avoid Being Part Of The Dreaded "Life Cycle"? The Data Vault took the Data Warehouse world by storm when it was released in 2001. Some of the world's largest and most complex data warehouse situations understood the value it gave especially with the capabilities of unlimited scaling, flexibility and security. Here is what industry leaders say about the Data Vault "The Data Vault is the optimal choice for modeling the EDW in the DW 2.0 framework" - Bill Inmon, The Father of Data Warehousing "The Data Vault is foundationally strong and an exceptionally scalable architecture" - Stephen Brobst, CTO, Teradata "The Data Vault should be considered as a potential standard for RDBMS-based analytic data management by organizations looking to
achieve a high degree of flexibility, performance and openness" - Doug Laney, Deloitte Analytics Institute. "I applaud Dan's contribution to the body of Business Intelligence and Data Warehousing knowledge and recommend this book be read by both data professionals and end users" - Howard Dresner, From the Foreword - Speaker, Author, Leading Research Analyst and Advisor.

You have in your hands the work, experience and testing of 2 decades of building data warehouses. The Data Vault model and methodology has proven itself in hundreds (perhaps thousands) of solutions in Insurance, Crime-Fighting, Defense, Retail, Finance, Banking, Power, Energy, Education, High-Tech and many more. Learn the techniques and implement them and learn how to build your Data Warehouse faster than you have ever done before while designing it to grow and scale no matter what you throw at it. Ready to "Super Charge Your Data Warehouse"?

The Data Model Resource Book, Volume 1

Understand data in a simple way using a data lake. KEY FEATURES

- In-depth practical demonstration of Hadoop/Yarn concepts with numerous examples.
- Includes graphical illustrations and visual explanations for Hadoop commands and parameters.
- Includes details of dimensional modeling and Data Vault modeling.
- Includes details of how to create and define a structure to a data lake.

DESCRIPTION

The book 'Data Processing and Modeling with Hadoop' explains how a distributed system works and its benefits in the big data era in a straightforward and clear manner. After reading the book, you will be able to plan and organize projects involving a massive amount of data. The book describes the standards and technologies that aid in data management and compares them to other technology business standards. The reader receives practical guidance on how to segregate and separate data into zones, as well as how to develop a model that can aid in data evolution. It discusses security and the measures that are utilized to reduce the impact of security. Self-service analytics, Data Lake, Data Vault 2.0, and Data Mesh are discussed in the book. After reading this book, the reader will have a thorough understanding of how to structure a data lake, as well as the ability...
to plan, organize, and carry out the implementation of a data-driven business with full governance and security. WHAT YOU WILL LEARN

● Learn the basics of components to the Hadoop Ecosystem. ● Understand the structure, files, and zones of a Data Lake. ● Learn to implement the security part of the Hadoop Ecosystem. ● Learn to work with the Data Vault 2.0 modeling. ● Learn to develop a strategy to define good governance. ● Learn new tools to work with Data and Big Data

WHO THIS BOOK IS FOR

This book caters to big data developers, technical specialists, consultants, and students who want to build good proficiency in big data. Knowing basic SQL concepts, modeling, and development would be good, although not mandatory.

TABLE OF CONTENTS

1. Understanding the Current Moment
2. Defining the Zones
3. The Importance of Modeling
4. Massive Parallel Processing
5. Doing ETL/ELT
6. A Little Governance
7. Talking About Security
8. What Are the Next Steps?

Oracle Designer

What strategies can users deploy to develop a successful data warehouse architecture? What are the other data sources that need to be integrated from a reporting perspective (either into the warehouse or directly as a source for reporting)? Who is the audience? What types of data ingestion pipelines do you have, at what frequency? Who is the audience? Defining, designing, creating, and implementing a process to solve a challenge or meet an objective is the most valuable role in EVERY group, company, organization and department. Unless you are talking a one-time, single-use project, there should be a process. Whether that process is managed and implemented by humans, AI, or a combination of the two, it needs to be designed by someone with a complex enough perspective to ask the right questions. Someone capable of asking the right questions and step back and say, 'What are we really trying to accomplish here? And is there a different way to look at it?' This Self-Assessment empowers people to do just that - whether their title is entrepreneur, manager, consultant, (Vice-)President, CxO etc - they are the people who rule the future. They are the person who asks the right questions to make Data Vault Modeling investments work better. This Data
Conceptual Modeling

Financial Cybersecurity Risk Management Author's abstract: The purpose of this Developmental Model is to create an efficient and effective generalized DW model (based on Data Vault 2.0
technology standards) and validate prototype implementation for a Corrections Environment that will provide efficient functional Business Intelligence to the End User. In Database Systems understanding the problem at hand and being able to use the right tool is an essential practice. State Departments of Corrections are often the last agencies to receive funding for technology based on the attitude of a society forced to provide technology based services for prisons. A Correctional System in itself is an ever changing Eco-system and will benefit from the use of a Data Vault System that is able to provide efficient Business Intelligence that is able to change with the environment.

Data Vault: Standard Requirements Agile Data Warehouse Design is a step-by-step guide for capturing data warehousing/business intelligence (DW/BI) requirements and turning them into high performance dimensional models in the most direct way: by modelstorming (data modeling ] brainstorming) with BI stakeholders. This book describes BEAM, an agile approach to dimensional modeling, for improving communication between data warehouse designers, BI stakeholders and the whole DW/BI development team. BEAM provides tools and techniques that will encourage DW/BI designers and developers to move away from their keyboards and entity relationship based tools and model interactively with their colleagues. The result is everyone thinks dimensionally from the outset! Developers understand how to efficiently implement dimensional modeling solutions. Business stakeholders feel ownership of the data warehouse they have created, and can already imagine how they will use it to answer their business questions. Within this book, you will learn: Agile dimensional modeling using Business Event Analysis & Modeling (BEAM ) Modelstorming: data modeling that is quicker, more inclusive, more productive, and frankly more fun! Telling dimensional data stories using the 7Ws (who, what, when, where, how many, why and how) Modeling by example not abstraction; using data story themes, not crow's feet, to describe detail Storyboarding the data warehouse to discover conformed dimensions and plan iterative development Visual modeling:
sketching timelines, charts and grids to model complex process measurement - simply Agile design documentation: enhancing star schemas with BEAM dimensional shorthand notation Solving difficult DW/BI performance and usability problems with proven dimensional design patterns Lawrence Corr is a data warehouse designer and educator. As Principal of DecisionOne Consulting, he helps clients to review and simplify their data warehouse designs, and advises vendors on visual data modeling techniques. He regularly teaches agile dimensional modeling courses worldwide and has taught dimensional DW/BI skills to thousands of students. Jim Stagnitto is a data warehouse and master data management architect specializing in the healthcare, financial services, and information service industries. He is the founder of the data warehousing and data mining consulting firm Llumino.

NoSQL Distilled The world of data warehousing is changing. Big Data & Agile are hot topics. But companies still need to collect, report, and analyze their data. Usually this requires some form of data warehousing or business intelligence system. So how do we do that in the modern IT landscape in a way that allows us to be agile and either deal directly or indirectly with unstructured and semi structured data? The Data Vault System of Business Intelligence provides a method and approach to modeling your enterprise data warehouse (EDW) that is agile, flexible, and scalable. This book will give you a short introduction to Agile Data Engineering for Data Warehousing and Data Vault 2.0. I will explain why you should be trying to become Agile, some of the history and rationale for Data Vault 2.0, and then show you the basics for how to build a data warehouse model using the Data Vault 2.0 standards. In addition, I will cover some details about the Business Data Vault (what it is) and then how to build a virtual Information Mart off your Data Vault and Business Vault using the Data Vault 2.0 architecture. So if you want to start learning about Agile Data Engineering with Data Vault 2.0, this book is for you.

Super Charge Your Data Warehouse The Vault Unleashed leads managers and those new to the Business Intelligence field on a
journey through the landscape. Successful Business Intelligence, or BI, is more than just technology. BI is the melding of people of multiple skills, using a variety of methods and empowering technology to solve business needs. This is best illustrated thru examples. This BI novel uses a fictitious island nation in the South Atlantic faced with threats to its very existence. The story takes the reader thru a connection of the dots from perception of a potential problem thru management execution of solutions. Augmenters must be brought on board. The core team must transfer knowledge to the new augmenters in the background and methodologies necessary to develop insights and make decisions. Most enterprises today are faced with similar challenges in our global economy. Your competition may not be state supported or as aggressive as in this tale. But for today's managers their counterparts are just as real.

UML and Data Modeling 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

Modeling the Agile Data Warehouse with Data Vault The nature of an information system; Naming; Relationships; Attributes; Types and categories and sets; Models; The record model; The other three popular models; The modelling of relationships; Elementary concepts; Philosophy.

Database Modeling Step by Step This is the digital version of the printed book (Copyright ©1996). Learning the basics of a modeling technique is not the same as learning how to use and apply it. To develop a data model of an organization is to gain insights into its nature that do not come easily. Indeed, analysts are often expected to understand subtleties of an organization's structure that may have evaded people who have worked there for years. Here's help for those analysts who have learned the basics of data modeling (or "entity/relationship modeling") but who need to obtain the insights required to prepare a good model of a real business. Structures common to many types of business are analyzed in areas such as accounting, material requirements planning, process manufacturing, contracts, laboratories, and documents. In each chapter, high-level data models are drawn from the following business areas: The Enterprise and Its World The Things of the Enterprise Procedures and Activities Contracts Accounting The Laboratory Material Requirements Planning Process Manufacturing Documents Lower-Level Conventions

Data and Reality Data Modeling for Agile Data Warehouse using Data Vault Modeling Approach. Includes Enterprise Data Warehouse Architecture. This is a complete guide to the data vault
data modeling approach. The book also includes business and program considerations for the agile data warehousing and business intelligence program. There are over 200 diagrams and figures concerning modeling, core business concepts, architecture, business alignment, semantics, and modeling comparisons with 3NF and Dimensional modeling.

Agile Data Warehouse Design The "father of data warehousing" incorporates the latest technologies into his blueprint for integrated decision support systems. Today's corporate IT and data warehouse managers are required to make a small army of technologies work together to ensure fast and accurate information for business managers. Bill Inmon created the Corporate Information Factory to solve the needs of these managers. Since the First Edition, the design of the factory has grown and changed dramatically. This Second Edition, revised and expanded by 40% with five new chapters, incorporates these changes. This step-by-step guide will enable readers to connect their legacy systems with the data warehouse and deal with a host of new and changing technologies, including Web access mechanisms, e-commerce systems, ERP (Enterprise Resource Planning) systems. The book also looks closely at exploration and data mining servers for analyzing customer behavior and departmental data marts for finance, sales, and marketing.

An Introduction to Agile Data Engineering Using Data Vault 2.0

This book constitutes the thoroughly refereed proceedings of the 20th East European Conference on Advances in Databases and Information Systems, ADBIS 2016, held in Prague, Czech Republic, in August 2016. The 21 full papers presented together with two keynote papers and one keynote abstract were carefully selected and reviewed from 85 submissions. The papers are organized in topical sections such as data quality, mining, analysis and clustering; model-driven engineering, conceptual modeling; data warehouse and multidimensional modeling, recommender systems; spatial and temporal data processing; distributed and parallel data processing; internet of things and sensor networks.
Practical Data Science A quick and reliable way to build proven databases for core business functions Industry experts raved about The Data Model Resource Book when it was first published in March 1997 because it provided a simple, cost-effective way to design databases for core business functions. Len Silverston has now revised and updated the hugely successful 1st Edition, while adding a companion volume to take care of more specific requirements of different businesses. This updated volume provides a common set of data models for specific core functions shared by most businesses like human resources management, accounting, and project management. These models are standardized and are easily replicated by developers looking for ways to make corporate database development more efficient and cost effective. This guide is the perfect complement to The Data Model Resource CD-ROM, which is sold separately and provides the powerful design templates discussed in the book in a ready-to-use electronic format. A free demonstration CD-ROM is available with each copy of the print book to allow you to try before you buy the full CD-ROM.

The Vault Unleashed


The Data Vault Guru Over the past 5 years, the concept of big data has matured, data science has grown exponentially, and data architecture has become a standard part of organizational decision-making. Throughout all this change, the basic principles that shape the architecture of data have remained the same. There remains a need for people to take a look at the "bigger picture" and to understand where their data fit into the grand scheme of things. Data Architecture: A Primer for the Data Scientist, Second Edition addresses the larger architectural picture of how big data fits within the existing information infrastructure or data warehousing systems. This is an essential topic not only for data scientists, analysts, and managers but also for researchers and engineers who increasingly need to deal with large and complex sets of data.
Until data are gathered and can be placed into an existing framework or architecture, they cannot be used to their full potential. Drawing upon years of practical experience and using numerous examples and case studies from across various industries, the authors seek to explain this larger picture into which big data fits, giving data scientists the necessary context for how pieces of the puzzle should fit together. New case studies include expanded coverage of textual management and analytics New chapters on visualization and big data Discussion of new visualizations of the end-state architecture

Data Vault Modeling A Complete Guide - 2020 Edition Data models are the main medium used to communicate data requirements from business to IT, and within IT from analysts, modelers, and architects, to database designers and developers. Therefore it’s essential to get the data model right. But how do you determine right? That’s where the Data Model Scorecard comes in. The Data Model Scorecard is a data model quality scoring tool containing ten categories aimed at improving the quality of your organization’s data models. Many of my consulting assignments are dedicated to applying the Data Model Scorecard to my client’s data models – I will show you how to apply the Scorecard in this book. This book, written for people who build, use, or review data models, contains the Data Model Scorecard template and an explanation along with many examples of each of the ten Scorecard categories. There are three sections: In Section I, Data Modeling and the Need for Validation, receive a short data modeling primer in Chapter 1, understand why it is important to get the data model right in Chapter 2, and learn about the Data Model Scorecard in Chapter 3. In Section II, Data Model Scorecard Categories, we will explain each of the ten categories of the Data Model Scorecard. There are ten chapters in this section, each chapter dedicated to a specific Scorecard category: Chapter 4: Correctness · Chapter 5: Completeness · Chapter 6: Scheme · Chapter 7: Structure · Chapter 8: Abstraction · Chapter 9: Standards · Chapter 10: Readability · Chapter 11: Definitions · Chapter 12: Consistency · Chapter 13: Data In Section III,
Validating Data Models, we will prepare for the model review (Chapter 14), cover tips to help during the model review (Chapter 15), and then review a data model based upon an actual project (Chapter 16).

Building a Scalable Data Warehouse with Data Vault 2.0 Learn how to build a data science technology stack and perform good data science with repeatable methods. You will learn how to turn data lakes into business assets. The data science technology stack demonstrated in Practical Data Science is built from components in general use in the industry. Data scientist Andreas Vermeul en demonstrates in detail how to build and provision a technology stack to yield repeatable results. He shows you how to apply practical methods to extract actionable business knowledge from data lakes consisting of data from a polyglot of data types and dimensions. What You'll Learn Become fluent in the essential concepts and terminology of data science and data engineering Build and use a technology stack that meets industry criteria Master the methods for retrieving actionable business knowledge Coordinate the handling of polyglot data types in a data lake for repeatable results Who This Book Is For Data scientists and data engineers who are required to convert data from a data lake into actionable knowledge for their business, and students who aspire to be data scientists and data engineers

ECCWS 2018 17th European Conference on Cyber Warfare and Security V2 This book constitutes revised tutorial lectures of the 7th European Business Intelligence and Big Data Summer School, eBISS 2017, held in Bruxelles, Belgium, in July 2017. The tutorials were given by renowned experts and covered advanced aspects of business intelligence and big data. This summer school, presented by leading researchers in the field, represented an opportunity for postgraduate students to equip themselves with the theoretical, practical, and collaboration skills necessary for developing challenging business intelligence applications.

The Nimble Elephant
An Efficient Data Vault-based Data Warehouse Model Implementation for a Corrections Environment How will the Data vault modeling team and the organization measure complete success of Data vault modeling? Why should we adopt a Data vault modeling framework? What's the best design framework for Data vault modeling organization now that, in a post industrial-age if the top-down, command and control model is no longer relevant? Has the Data vault modeling work been fairly and/or equitably divided and delegated among team members who are qualified and capable to perform the work? Has everyone contributed? How to deal with Data vault modeling Changes? Defining, designing, creating, and implementing a process to solve a challenge or meet an objective is the most valuable role in EVERY group, company, organization and department. Unless you are talking a one-time, single-use project, there should be a process. Whether that process is managed and implemented by humans, AI, or a combination of the two, it needs to be designed by someone with a complex enough perspective to ask the right questions. Someone capable of asking the right questions and step back and say, 'What are we really trying to accomplish here? And is there a different way to look at it?' This Self-Assessment empowers people to do just that - whether their title is entrepreneur, manager, consultant, (Vice-)President, CxO etc - they are the people who rule the future. They are the person who asks the right questions to make Data vault modeling investments work better. This Data vault modeling All-Inclusive Self-Assessment enables You to be that person. All the tools you need to an in-depth Data vault modeling Self-Assessment. Featuring 701 new and updated case-based questions, organized into seven core areas of process design, this Self-Assessment will help you identify areas in which Data vault modeling improvements can be made. In using the questions you will be better able to: - diagnose Data vault modeling projects, initiatives, organizations, businesses and processes using accepted diagnostic standards and practices - implement evidence-based best practice strategies aligned with overall goals - integrate recent advances in Data vault modeling and process design strategies into practice according to best practice guidelines Using
a Self-Assessment tool known as the Data vault modeling Scorecard, you will develop a clear picture of which Data vault modeling areas need attention. Your purchase includes access details to the Data vault modeling self-assessment dashboard download which gives you your dynamically prioritized projects-ready tool and shows your organization exactly what to do next. Your exclusive instant access details can be found in your book.

Treasury, Postal Service, and General Government Appropriations for Fiscal Year 1996 Pro Oracle Database 11g RAC on Linux provides full-life-cycle guidance on implementing Oracle Real Application Clusters in a Linux environment. Real Application Clusters, commonly abbreviated as RAC, is Oracle’s industry-leading architecture for scalable and fault-tolerant databases. RAC allows you to scale up and down by simply adding and subtracting inexpensive Linux servers. Redundancy provided by those multiple, inexpensive servers is the basis for the failover and other fault-tolerance features that RAC provides. Written by authors well-known for their talent with RAC, Pro Oracle Database 11g RAC on Linux gives you a rock-solid and technically flawless foundation on which to build your RAC-management skills. Authors Julian Dyke and Steve Shaw share their hard-won experience in building RAC clusters, showing you how to build for success using the very latest Oracle technologies, such as Automatic Storage Management (ASM) and Oracle Clusterware. You’ll learn to troubleshoot performance and other problems. You’ll even learn how to correctly deploy RAC in a virtual-machine environment based upon Oracle VM, which is the only virtualization solution supported by Oracle Corporation. RAC is a complex and powerful technology. It demands expertise in its deployment. You can’t just “wing it” in creating a RAC solution. Julian and Steve have earned the right to term themselves expert—in Pro Oracle Database 11g RAC on Linux, they offer a rigorous and technically-correct treatment of RAC that helps you build a solid foundation of expertise and achieve success. Rigorous and technically accurate content Complete coverage of RAC, from planning to implementation to rollout to ongoing maintenance and troubleshooting Up-to-date with
the very latest RAC features