Data Warehouse Architecture Best Practices and the Star Schema Approach

White Paper

Author: Mary Welsh – Data Engineer

March 25, 2025

[Overview 1](#_Toc193834558)

[Summary Recommendation 2](#_Toc193834559)

[The Data Warehouse Toolkit Summary 2](#_Toc193834560)

[The Complete Reference: Star Schema Summary 3](#_Toc193834561)

[Databases Illuminated Summary 4](#_Toc193834562)

[Agile Data Warehouse Design Summary 4](#_Toc193834563)

[Astera Articles Summary 5](#_Toc193834564)

[ThoughtSpot Article Summary 6](#_Toc193834565)

[Analytics engineering by Redha Cherif Article Summary 7](#_Toc193834566)

[Medium Article Summary 7](#_Toc193834567)

[Saras Analytics Article Summary 8](#_Toc193834568)

[Geeks for Geeks Article Summary 9](#_Toc193834569)

[Second Brain Articles Summary 9](#_Toc193834570)

[phData Article Summary 11](#_Toc193834571)

[Summary of References 12](#_Toc193834572)

## Overview

This document provides an overview of the importance of good data warehouse architecture, specifically focusing on the Kimball star schema methodology. It will cover key features this architecture can support as well as the benefits and reasoning behind the way it is implemented.

## Summary Recommendation

While there are many other architecture philosophies in the data warehouse world today, it has been shown that the star schema architecture has withstood the tests of time and has proven to be the most effective and popular approach for most business use cases. It can provide support for historical data, maintain data integrity, organize the data for more efficient analysis, improve performance, and much more.

### Key Takeaways

* Star schemas are still relevant in a fast-changing data warehouse landscape. They offer features that can address most business use cases.
* While there are applications for other architectures, there is a reason the Kimball methodology has been widely adopted as a best practice and accepted as an industry standard.
* Data modeling is an integral part of the data warehouse development process. Implementing best practices in architecture can vastly improve the data warehouse from multiple standpoints and for various users and use cases.

## The Data Warehouse Toolkit Summary

Book by Ralph Kimball, Margy Ross

This book is the origin of the Kimball methodology which lays the foundations for the star schema architecture. While it delves deep into the specifics of each of the concepts, it emphasizes the importance and significance of this architecture:

“Dimensional Modeling also has emerged as the leading architecture for building integrated DW/BI systems. When you use the conformed dimensions and conformed facts of a set of dimensional models, you have a practical and predictable framework for incrementally building complex DW/BI systems that are inherently distributed. For all that has changed in our industry, the core dimensional modeling techniques that Ralph Kimball published 17 years ago have withstood the test of time. Concepts such as conformed dimensions, slowly changing dimensions, heterogenous products, factless fact tables, and the enterprise data warehouse bus matrix continue to be discussed in design workshops around the globe.”

## The Complete Reference: Star Schema Summary

Book by Christopher Adamson

Another architectural book highlighting the most popular data warehouse architecture strategies, this book argues that “while these architectures differ in fundamental ways, there is a place for the star schema in each of them.”

It highlights the strategic importance of dimensional design:

“Fundamentally, a dimensional model deals with the measurement of business processes. It describes how a business process is evaluated, and it can be used to frame questions about the process. In this respect, it speaks clearly to the *business users* of the data warehouse.

A dimensional model also has technical implications. Its definition determines the data sources that must be integrated, how information must be cleansed or standardized, and what queries or reports can be built. In this respect, it speaks clearly to the *developers* of the data warehouse.

These business and technical characteristics of the dimensional model make it an ideal focal point for managing data warehouse strategy. A dimensional model can serve as the basis for a *shared* understanding of project priorities, scope, and progress. From a business perspective, it imparts a clear understanding of functional capability; from an IT perspective, is supports a clear understanding of the level of effort required for implementation.”

## Databases Illuminated Summary

Book by Catherine Ricardo

A much broader introduction to all sorts of database fundamental concepts beyond the scope of just data warehousing, this book has a chapter dedicated to Data Warehouses and Data Mining. In chapter 15 section 3 it states that “unlike an operational database, for which requirements can be specified in advance, a data warehouse must be designed to support *ad hoc* queries and new and unanticipated types of analysis and reporting”

“The data warehouse is...used to support queries for OLAP, to provide information for decision support systems that are used by management for making strategic decisions, and to provide the data for data mining tools that discover new information about patterns in the data.”

Here lies a key feature and benefit of a data warehouse: the ability for end users to query the data to gain greater business insights. This necessarily implies the importance of proper design for organizing the data in such a way that is easily digested and understood by end users.

## Agile Data Warehouse Design Summary

Book by Lawrence Corr, Jim Stagnitto

This book offers an agile approach to the traditional dimensional modeling techniques. It also delves into the specifics of dimensional modeling using the most popular methods. In the introduction it outlines the advantages of dimensional modeling for data warehousing:

“The most obvious advantage of a dimensional model… is its *simplicity*. The small number of tables and joins, coupled with the explicit facts in the center of the diagram, makes it easy to think about how sales can be measured and easy to construct the necessary queries…Limiting the number of tables involved and the length of the join paths…maximizes query performance by leveraging DBMS features such as star-join optimization (which processes multiple joins to a fact table in a single pass).”

“A deeper, less immediately obvious benefit of dimensional models is that they are *process-oriented*. They are not just the result of some aggressive physical data model optimization…to overcome the limitations of database to cope with join intensive BI queries. Instead, the best dimensional models are the result of asking questions to discover which business processes need to be measured, how they should be described in business terms and how they should be measured.”

## Astera Articles Summary

### What is Star Schema? Advantages and Disadvantages

Article by Junaid Baig, March 25, 2024

[https://www.astera.com/knowledge-center/star-schema](https://www.astera.com/knowledge-center/star-schema/)

This article gives a brief overview of the basics of star schema data modeling concepts. These concepts are the fundamental building blocks of the Kimball methodology. It provides both advantages and disadvantages of using star schemas. Advantages include:

* Simplified querying
* Faster performance
* Intuitive analysis
* Robust support

It concludes that the best time to use a star schema is when users have a clear understanding of the data, the data is structured and quantitative, when query performance is the top priority, and when data redundancy will not be an issue.

### What’s the Best Data Warehouse Architecture for Reporting?

Article by Farhan Ahmed Khan, September 3, 2024

[https://www.astera.com/type/blog/best-data-warehouse-architecture-reporting](https://www.astera.com/type/blog/best-data-warehouse-architecture-reporting/)

This article compares 3 popular data warehouse architectures: 3NF/Data Vault, Dimensional Models, and OBT (One Big Table). It provides a brief explanation of each and what they excel in. It mentions reduced data redundancy and easier navigation as advantages of the Dimensional Model approach. It states that “a star schema is the clear winner when talking strictly about consumption for reporting…[as] the queries are less complicated.” It also notes that keeping track of historical data is an important feature of star schemas.

It concludes that “star schema is, well and truly, the most suitable architecture for reporting. While other architectures have their own advantages, a star schema provides the perfect balance between the level of denormalization needed for less complex queries. A star schema also offers structural benefits, such as historical data management and ease of use, making it stand out among the rest of the architectures.”

## ThoughtSpot Article Summary

### Data modeling best practices for data and analytics engineers

Article by Sonny Rivera, March 23, 2023

<https://www.thoughtspot.com/data-trends/data-modeling/data-modeling-best-practices-for-analytics-and-data-engineers>

This article provides tips and tricks for data modeling and architectural best practices. It recommends using a star schema for your analytics models and outlines 5 key benefits:

* Simplicity
* Performance
* Adaptability
* Normalization
* Integration

It also recommends avoiding the OBT (One Big Table) architecture whenever possible due to several challenges such as the following:

* Not scalable
* Duplicate and NULL data, and loss of data granularity
* Problems maintaining data integrity
* Scalability declines even with columnar cloud data platforms as number of columns increases
* Not ideal for complex queries

## Analytics engineering by Redha Cherif Article Summary

### Demystifying Data Modeling: Which Approach is Right for Your Business?

Article by Redha Cherif, December 15, 2024

<https://redhacherif.substack.com/p/demystifying-data-modeling-which>

This article focuses on the 3 most popular data modeling approaches: Kimball dimensional modeling, One Big Table, and the Inmon approach. It delves into the basics of each and highlights the differences as well as pros and cons. The overall article key takeaway recommends using the Kimball dimensional modeling with star schema alongside One Big Tables. It specifically says that the two architectures can be used in combination, with the underlying architecture using the traditional star schema and the OBT built on top of that. This will leverage advantages of both approaches. It argues that OBTs can be the first tables which can be checked for analytics use cases and should answer 10 to 20% of use cases, with the rest falling on the traditional star schema model.

The article also advocates for the Kimball approach instead of Inmon because Inmon “is more complex to implement and requires more time before being able to deliver business-driven analytics solutions. Moreover, Kimball’s modeling is optimized for querying and analysis compared to Inmon’s which require further aggregation. Moreover, in real-world, using a business-focused approach (Kimball) is very often a better solution than using an enterprise-focused (Inmon) one.”

## Medium Article Summary

### In-Depth Guide to Star Schema in Data Warehouse Modeling: Concepts, Design Principles, and Advantages

Article by Nilimesh Halder, PhD, October 19, 2023

<https://medium.com/analysts-corner/in-depth-guide-to-star-schema-in-data-warehouse-modeling-concepts-design-principles-and-f540f3c4744b>

This article provides an introduction to star schema dimensional modeling and its concepts and components. It says that “the star schema is a widely used data warehouse modeling technique that offers simplicity, efficiency, and improved query performance in business intelligence (BI) and analytical applications.” It delves into the various features and highlights several advantages such as:

* Simplicity
* Query Performance
* Scalability
* Flexible Data Analysis
* Efficient Data Storage
* Easier Maintenance

It concludes that the star schema “is a powerful data warehouse modeling technique that offers simplicity, efficiency, and improved query performance for business intelligence and analytical applications…The star schema enables flexible data analysis, efficient data storage, and easier maintenance, ultimately supporting data-driven decision-making in various industries and applications.”

## Saras Analytics Article Summary

### 11 Best Practices for Data Modeling

Article by Saras Analytics team, March 21, 2025

<https://www.sarasanalytics.com/blog/data-modeling-best-practices>

This article defines what data modeling is and provides several best practices and reasons both are important. It states that “Data Modeling provides business and technical stakeholders with a clear, visual representation of complicated data ideas, to their benefit” and notes that “Data modeling is, at its heart, a paradigm of data comprehension prior to analysis or action...its significance and need will increase exponentially as diverse domains increase their adoption rates.”

## Geeks for Geeks Article Summary

### Data Modeling Techniques For Data Warehouse

Article by Geeks for Geeks team, February 14, 2025

[https://www.geeksforgeeks.org/data-modeling-techniques-for-data-warehouse](https://www.geeksforgeeks.org/data-modeling-techniques-for-data-warehouse/)

This article introduces the types of data models such as conceptual, representational/logical, and physical. It lists some of the popular modeling approaches including the star schema but does not recommend one over the others. Instead, it notes that the star schema is the simplest and most commonly used schema; its advantages including simplicity and ease of understanding and efficiency for querying large datasets.

It underscores the importance of data modeling in data warehouses for several key reasons:

* Improved Data Quality
* Efficient Data Retrieval
* Scalability
* Reduced Redundancy

## Second Brain Articles Summary

### Data Modeling – The Unsung Hero of Data Engineering: An Introduction to Data Modeling (Part 1)

Article by Simon Späti, April 3, 2023

<https://www.ssp.sh/blog/data-modeling-for-data-engineering-introduction>

This blog post is another high-level introduction to data modeling. On its importance it notes: “In an ever faster-growing and changing data landscape, having a clear data structure and architecture in place is essential. The best way is to have a good data model and strategy.” It goes on to highlight a very common problem many orgs have: “Most data projects initially ignore data modeling because of time constraints, fast-paced development, or sometimes lack of knowledge. Instead of investing time into sketching out how the data should flow from its source, the organization’s requirements, to how the end user wants to analyze its data, projects get built quickly and with a “dirty” MVP. Data modeling goes hand in hand with clear goals. It integrates technical and data requirements as well as business and performance requirements. With these, aligned within the company, it’s much easier to understand the business from a data perspective and be successful in the long run.”

The blog goes into details of the history of data modeling and the rise and fall of its perception of importance throughout time. It also outlines the high-level key concepts.

### Data Modeling – The Unsung Hero of Data Engineering: An Introduction to Data Modeling (Part 2)

Article by Simon Späti, May 3, 2023

<https://www.ssp.sh/blog/data-modeling-for-data-engineering-approaches-techniques>

This blog post is the second part in the 3-part series on data modeling. This part provides an overview of data modeling approaches, techniques, and common problems with data modeling. It highlights the importance of dimensional modeling today: “is dimensional modeling and all its associated concepts still relevant today? The answer is a resounding yes—perhaps even more so than before. As you’ve read in the preceding chapters, dimensional modeling aims to achieve a focus on business value. In today’s rapidly evolving world, this crucial aspect is sometimes overlooked.

By incorporating a robust dimensional model at the core of every data project, data engineers are compelled to consider critical questions related to granularity, entities, metrics, and more. Addressing these essential aspects upfront and working towards them is invaluable for achieving business goals and driving project success.”

The blog also notes that problems commonly arise in the last “mile” of the data analytics cycle and that “addressing common problems in data modeling and ensuring tight integration into the overall data architecture is essential for success.”

A diagram of a person's problem

AI-generated content may be incorrect.

## phData Article Summary

### What is Data Modeling and How Do I Choose the Right One?

Article by Nitin Mule, August, 24, 2021

<https://www.phdata.io/blog/how-to-model-and-choose-the-right-data-model>

This article is another overview of data modeling, its importance, and highlights 2 popular methodologies: data vault and star schema. It notes that “an effective data model provides a solid foundation for any Data Warehouse to sustain growing data volumes and adjust easily to addition or deletion of data entities.”

Of the star schema, this article notes the following advantages:

* Simpler Queries
* Simplified Business Reporting Logic
* Can Feed OLAP Cubes

The article recommends using the star schema (specifically in Snowflake) for building smaller, focused data marts for one or a few departments. It mentions the use of conforming dimensions to eliminate data silos between data marts which can further enforce consistency in business definitions across departments.

## Summary of References

* The Data Warehouse Toolkit – Ralph Kimball, Margy Ross
* The Complete Reference: Star Schema – Christopher Adamson
* Databases Illuminated – Catherine Ricardo
* Agile Data Warehouse Design – Lawrence Corr, Jim Stagnitto
* [https://www.astera.com/knowledge-center/star-schema](https://www.astera.com/knowledge-center/star-schema/)
* [https://www.astera.com/type/blog/best-data-warehouse-architecture-reporting](https://www.astera.com/type/blog/best-data-warehouse-architecture-reporting/)
* <https://www.thoughtspot.com/data-trends/data-modeling/data-modeling-best-practices-for-analytics-and-data-engineers>
* <https://redhacherif.substack.com/p/demystifying-data-modeling-which>
* <https://medium.com/analysts-corner/in-depth-guide-to-star-schema-in-data-warehouse-modeling-concepts-design-principles-and-f540f3c4744b>
* <https://www.sarasanalytics.com/blog/data-modeling-best-practices>
* [https://www.geeksforgeeks.org/data-modeling-techniques-for-data-warehouse](https://www.geeksforgeeks.org/data-modeling-techniques-for-data-warehouse/)
* <https://www.ssp.sh/blog/data-modeling-for-data-engineering-introduction>
* <https://www.ssp.sh/blog/data-modeling-for-data-engineering-approaches-techniques>
* <https://www.phdata.io/blog/how-to-model-and-choose-the-right-data-model>