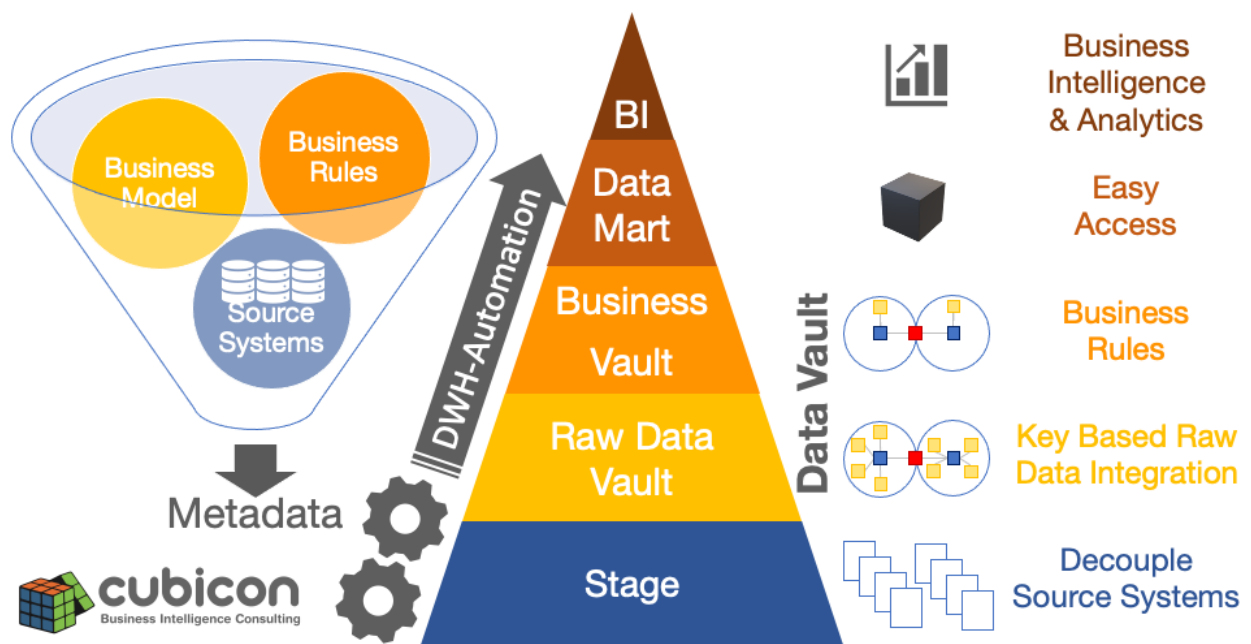


Data Vault Modelling

Aya Mahmoud Mohamed Mahmoud

ITI

Eng. Waleed



Abstract

Data vault modeling is a database modeling method that is designed to provide long-term historical storage of data coming in from multiple operational systems. It is also a method of looking at historical data that deals with issues such as auditing, tracing of data, loading speed and resilience to change as well as emphasizing the need to trace where all the data in the database came from. This means that every row in a data vault must be accompanied by record source and load date attributes, enabling an auditor to trace values back to the source. It was developed by Daniel (Dan) Linstedt in 2000.

Introduction

Data vault modeling makes no distinction between good and bad data ("bad" meaning not conforming to business rules). This is summarized in the statement that a data vault stores "a single version of the facts" (also expressed by Dan Linstedt as "all the data, all of the time") as opposed to the practice in other data warehouse methods of storing "a single version of the truth"[2] where data that does not conform to the definitions is removed or "cleansed".

Methodology

Data Warehouses have served the purpose of providing a source of value-added information for quite some decades. In supporting business users in their day-to-day operations, Data Warehouses integrate multiple sources used throughout the organization, track history for knowledge improvements, auditing and compliancy, and provide the business with an easy-to-access and easy-to-explore insight layer. While Data Warehouses are often modelled in Kimball or Inmon approach, Data Vault is also a vivid possibility. Data Vault is renowned for its stereotypical modelling technique involving Hubs, Links and Satellites. Next to that, it also provides a project approach, architectures, and implementation guidelines.

Results

In a Data Vault oriented Data Warehouse, the ‘Data Vault layer’ is the beating heart of your information. It houses all data that your organisation has ever recorded and uniformly models it in three components: Hubs, Satellites and Links. The goal of this layer Data Vault layer is to track every single change of attribute in your source systems and provide a consistent ‘base’ on which subsequent objects are built (e.g., Kimball-oriented layer, data/information marts, OLAP cubes).

There are some fundamental rules in Data Vault Modeling that must be followed, or the model itself no longer qualifies to be a Data Vault model. Some of the rules are as follows:

1. Business keys are separated by grain and semantic meaning. That means the keys customer corporation and customer individual must exist or be recorded in two separate hub structures.
2. Relationships, events, and intersections across two or more business keys are placed into link structures.
3. Link structures have no begin or end dates; they are merely an expression of the relationship at the time the data arrived in the warehouse.
4. Satellites are separated by type of data and classification and rate of change. Type of data is typically a single source system.

Discussion

Within the Data Vault, there are layers of data:

1. Source systems, where the data will be created or originate.
2. A staging area that receives the data from the source system, and models it according to its original structure.
3. A core data warehouse containing the raw vault, a layer that allows data to be traced back to the original source system data.
4. A business vault, essentially a semantic layer where business rules are implemented.
5. Data marts, structured to the requirements of the organization. A finance or marketing data mart, for example, would hold relevant data for specific analysis purposes.

Business Vault (BV): It is an extension of a Raw Vault that applies selected business rules, denormalizations, calculations, and other query assistance functions in order to facilitate user access and reporting. Business Vault tables should be refreshed once their dependent Raw Vault tables are refreshed.

Adding Business Vault tables is an option to the Raw Vault. It is not required. As usually it depends on the business and reporting requirements you have to fulfill.

The primary group to use or access the BV is your power users. These are people who understand SQL and relational models well and are not afraid of having many table joins. They may need to do some exploration or data mining but either do not want to wait for the star schemas to be built or are not doing dimensional analysis.

Business Vault also includes some specialty tables that help us in building more efficient queries against the Raw Vault. These are Point-in-Time (PIT) tables and Bridge Tables.

Conclusion

Data Vault modelling can prove an integral part of that environment. With a robust approach designed to maximise the benefits that a Data Vault approach offers, those at the coal face will benefit from vastly improved performance when running analytical models or workflows – enabling organisations to optimise the value of their data at speed.

Resources

1. Linstedt, Dan (December 2010). Super Charge your Data Warehouse. Dan Linstedt. ISBN 978-0-9866757-1-3.
2. Editor. (2022, July 7). The data vault modeling. Lytix. Retrieved January 23, 2023, from <https://lytix.be/the-data-vault-methodology-2/>
3. panelW.H.InmonDanielLinstedtMaryLevins, A. links open overlay, W.H.Inmon, DanielLinstedt, MaryLevins, & AbstractOne of the most important components of the end-state architecture is that of the data vault. The data vault exists to satisfy the need for rock-solid data integrity. Like all other components of the end-state architecture. (2019, May 10). Introduction to data vault modeling. Data Architecture (Second Edition). Retrieved January 23, 2023, from <https://www.sciencedirect.com/science/article/pii/B978012816916200019X>
4. What is Data Vault Modeling and why do we need it? Dataconomy. (2020, August 20). Retrieved January 23, 2023, from <https://dataconomy.com/2020/08/what-is-data-vault-modeling-and-why-do-we-need-it/#:~:text=Within%20the%20Data%20Vault%2C%20there%20are%20layers%20of,layer%20where%20business%20rules%20are%20implemented%3B%20More%20items>
5. Graziano, K. (2016, March 2). The business data vault. Vertabelo Data Modeler. Retrieved January 23, 2023, from <https://www.vertabelo.com/blog/data-vault-series-the-business-data-vault/>