Data Warehouse Structure Decision
Star Schema or Data Vault

Background

Having a data warehouse that drives business intelligence is becoming more relevant everyday as companies collect more data about their customers. Being able to effectively analyze this data can help companies make more informed decisions and stay one step ahead of their competition. Creating a data warehouse also takes stress off your employees and operational systems allowing for more efficient creation of meaningful reports and analytics.

Star Schema is a popular data warehouse structure that transforms operational data into a series of "fact table" structures. The storage of the data focuses on specific business-derived facts that would then be the source of data for the front-end reporting or business analytics tool for data output and visualizations. A data vault structure is another, less common data warehouse structure. Instead of relying on business-derived decisions around "facts" and their associated business rules, a data vault stores all data from the source systems in a historical, structured manner. This can help you realize potential flaws in your business processes while providing the ability to trace all data back to the source systems.

Star Schema or Data Vault?

Prior to finalizing a decision on a data warehouse structure, any requirements for a business intelligence reporting system would need to be completed. It's critical to fully understand what role your data warehouse plays in your business. An analysis of which one would work best given your specific requirements can be established using a Pro vs. Con comparison. Some examples of pros and cons at a high level include:

Data Vault		Star Schema	
Pros	**Cons**	**Pros**	**Cons**
• Source structure changes require minimal-to-no structure changes in DW	• Business rules must be applied within the marts or data pull queries	• Business Rules do not need to be applied during queries	• Stored data has business rules "baked in"
• Business rule changes have zero impact on structure of and loading process into DW	• Data should not be accessed directly from DW – marts must be built	• Data can be accessed directly from DW without building separate marts.	• Source structure changes require large structure changes in DW

Data Flow

Both structures require a load process by which data is brought into the data warehouse. Regardless of the need for real-time, near-real-time, and batched data, the solution can be implemented similarly regardless of the chosen structure.
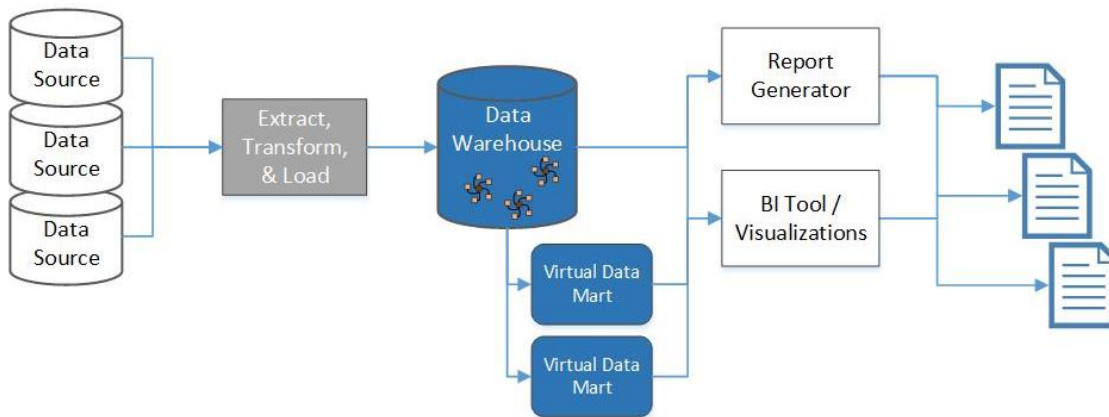

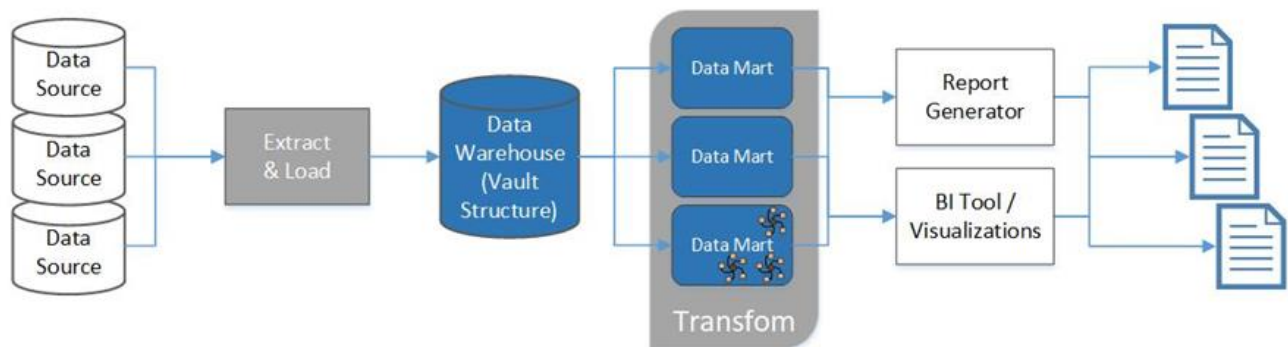Figure 1: Basic Star Schema Data Warehouse Architecture

Figure 2: Basic Data Vault Data Warehouse Architecture

About Data Blueprint

Data Blueprint is a data management consulting firm that puts organizations on the right path to leverage data for competitive advantage and operational efficiency. We unlock business value. It is our belief that strong data management capabilities are the key to sustained long-term value creation. We help clients leverage data as a strategic investment and become data-driven.