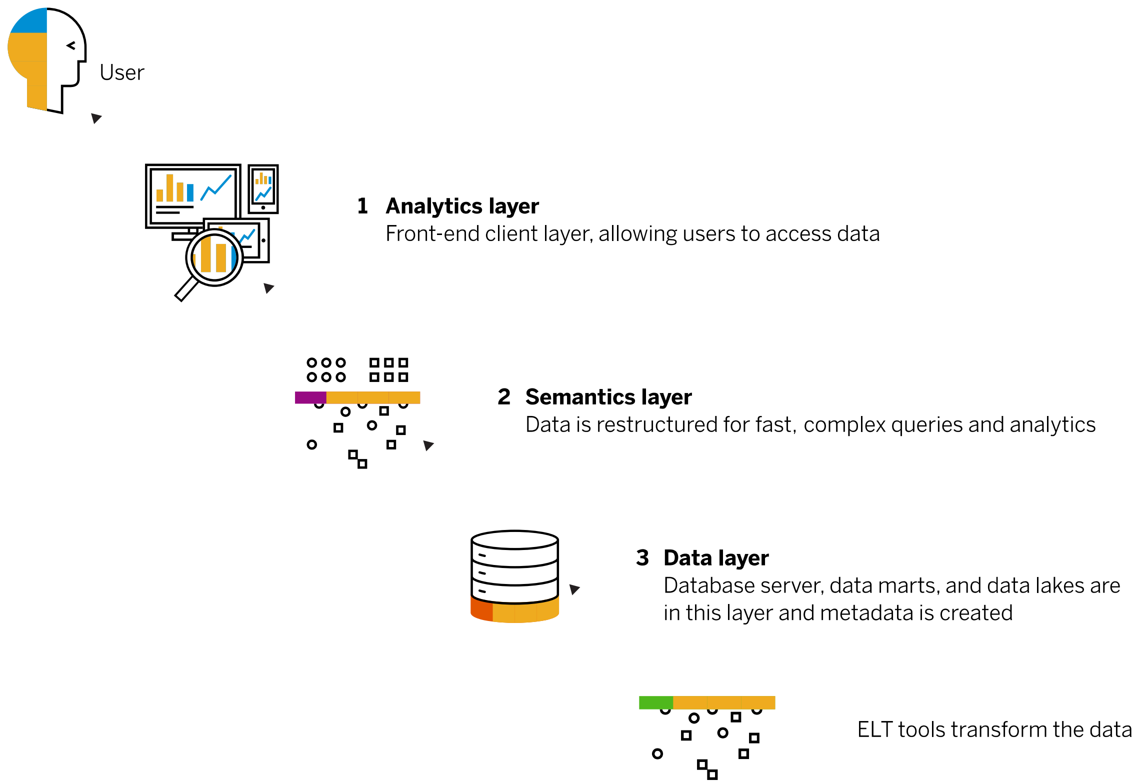

Diagram of data warehouse architecture. A typical data warehouse includes the three separate layers above.

Data layer: Data is extracted from your sources and then transformed and loaded into the bottom tier using ETL tools. The bottom tier consists of your database server, data marts, and data lakes. Metadata is created in this tier – and data integration tools, like data virtualization, are used to seamlessly combine and aggregate data.

Semantics layer: In the middle tier, online analytical processing (OLAP) and online transactional processing (OLTP) servers restructure the data for fast, complex queries and analytics.

Analytics layer: The top tier is the front-end client layer. It holds the data warehouse access tools that let users interact with data, create dashboards and reports, monitor KPIs, mine and analyze data, build apps, and more. This tier often includes a workbench or sandbox area for data exploration and new data model development.

The semantic or business layer that provides natural language phrases and allows everyone to instantly understand data, define relationships between elements in the data model, and enrich data fields with new business information.
Virtual workspaces allow teams to bring data models and connections into one secured and governed place supporting better collaborating with colleagues through one common space and one common data set.
Cloud has further improved decision making by globally empowering employees with a rich set of tools and features to easily perform data analysis tasks. They can connect new apps and data sources without much IT support.

Introduction to federated data warehouse

Nowadays, corporate usually has a set of heterogeneous system landscape that contains transaction systems and business intelligence tools which provide analytical capabilities for each individual department needs.

Each department views a business model from its own perspective. For example, a product in Sales can be defined as a material in Manufacturing and equipment in Service Management. In order to integrate those heterogeneous systems that aim to provide analytic capabilities across the different functions and departments, the federated data warehouse was invented.

A federated data warehouse is a practical approach to achieving the "single version of the truth" across the organization. The federated data warehouse is used to integrate key business measures and dimensions. The foundations of the federated data warehouse are the common business model and common staging area.

The architecture of federated data warehouse

Regional federation possible in federated data warehouse

The big organization has various regions that provide businesses to customers globally. Different regional data warehouses were built for each region to meet the specific business needs. A global data warehouse also was built to provide analytical capabilities to the executive at the global level.

The difference between the regional and global data warehouse systems is the nature of data resided at each system level. In the regional federated data warehouse architecture picture below, there are two data flows between regional and global data warehouses:

•Upward federation – only fact data are moved from regional data warehouse to global data warehouse. The aggregation of data can take place at a global data warehouse after data integrated or during data movement.
•Downward federation – in the downward federation, the reference flows from the global to the regional level. This ensures the consistency and integrity of data across the organization. Transactional data from corporate operational systems such as ERP and CRM are sourced at the global level and then extracted, transformed, and loaded into a respective regional data warehouse.

What is Dimensional Modeling?

Dimensional modeling represents data with a cube operation, making more suitable logical data representation with OLAP data management. The perception of Dimensional Modeling was developed by **Ralph Kimball** and is consist of **"fact"** and **"dimension"** tables.

In dimensional modeling, the transaction record is divided into either **"facts,"** which are frequently numerical transaction data, or **"dimensions,"** which are the reference information that gives context to the facts. For example, a sale transaction can be damage into facts such as the number of products ordered and the price paid for the products, and into dimensions such as order date, user name, product number, order ship-to, and bill-to locations, and salesman responsible for receiving the order.

Objectives of Dimensional Modeling

The purposes of dimensional modeling are:

1.To produce database architecture that is easy for end-clients to understand and write queries.

2.To maximize the efficiency of queries. It achieves these goals by minimizing the number of tables and relationships between them.

Advantages of Dimensional Modeling

Following are the benefits of dimensional modeling are:

Dimensional modeling is simple: Dimensional modeling methods make it possible for warehouse designers to create database schemas that business customers can easily hold and comprehend. There is no need for vast training on how to read diagrams, and there is no complicated relationship between different data elements.

Dimensional modeling promotes data quality: The star schema enable warehouse administrators to enforce referential integrity checks on the data warehouse. Since the fact information key is a concatenation of the essentials of its associated dimensions, a factual record is actively loaded if the corresponding dimensions records are duly described and also exist in the database.

By enforcing foreign key constraints as a form of referential integrity check, data warehouse DBAs add a line of defense against corrupted warehouses data.

Performance optimization is possible through aggregates: As the size of the data warehouse increases, performance optimization develops into a pressing concern. Customers who have to wait for hours to get a response to a query will quickly become discouraged with the warehouses. Aggregates are one of the easiest methods by which query performance can be optimized.

Disadvantages of Dimensional Modeling

1. To maintain the integrity of fact and dimensions, loading the data warehouses with a record from various operational systems is complicated.

2. It is severe to modify the data warehouse operation if the organization adopting the dimensional technique changes the method in which it does business.

Elements of Dimensional Modeling

Fact

It is a collection of associated data items, consisting of measures and context data. It typically represents business items or business transactions.

Dimensions

It is a collection of data which describe one business dimension. Dimensions decide the contextual background for the facts, and they are the framework over which OLAP is performed.

Measure

It is a numeric attribute of a fact, representing the performance or behavior of the business relative to the dimensions.