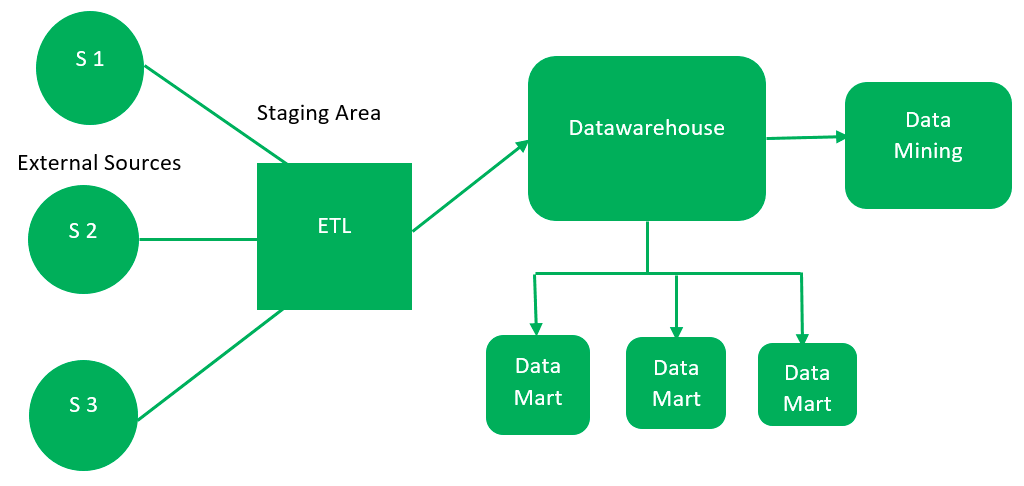
**Design:**

A D**ata-Warehouse** is a heterogeneous collection of data sources organized under a unified schema. There are 2 approaches for constructing a data warehouse:

**What is Top-Down Approach?**

* The initial approach developed by Bill Inmon known as the top-down approach starts with building a single source data warehouse for the whole company.
* It processes external data through the ETL ([Extract, Transform, Load](https://www.geeksforgeeks.org/explain-the-etl-extract-transform-load-process-in-data-engineering/)) process and subsequently stores them in the data warehouse.
* From this warehouse, specialized data marts are developed for specific departments, like finance.
* The strength of this method is that it offers a clear structure for managing data, however, this method can be expensive as well as time-consuming

**Diagram:**



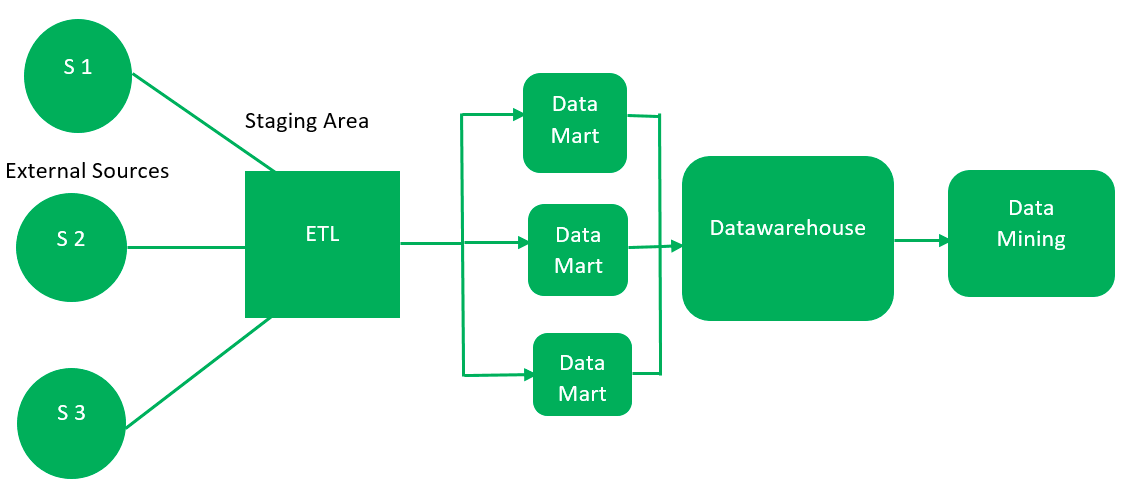
**Components in the Top-Down Approach**

1. **External Sources**: Data is collected from various sources, including structured, semi-structured, and unstructured data types.
2. **Staging Area**: Data from external sources is validated and processed using ETL tools:
   * **Extract**: Data is gathered from external sources.
   * **Transform**: Data is converted into a standardized format.
   * **Load**: Transformed data is stored in the data warehouse.
3. **Data Warehouse**: Acts as a central repository, storing cleansed data in its purest form. Metadata resides here, while detailed data is distributed to data marts.
4. **Data Marts**: Contain subsets of data specific to organizational functions, managed by individual authorities. The number of data marts depends on the organization's functions.
5. **Data Mining**: Involves analyzing the data stored in the warehouse to identify hidden patterns and insights using data mining algorithms.

**Bottom-Up Approach:**

* The Bottom-Up Approach, proposed by Ralph Kimball, starts by creating individual data marts focused on specific business functions like marketing or sales.
* These data marts are extracted transformed & loaded first to provide organizations’ ability to generate reports instantly.
* Later, they are integrated into a centralized data warehouse. This approach is more flexible, cost-effective, and suitable for smaller organizations

**Diagram:**



**In the Bottom-Up Approach:**

1. Data is extracted from external sources, similar to the top-down approach.
2. It is processed in the staging area using ETL tools and loaded into data marts instead of directly into a data warehouse.
3. Data marts, addressing specific business areas, are created first and provide immediate reporting capabilities.
4. These data marts are later integrated into a centralized data warehouse.

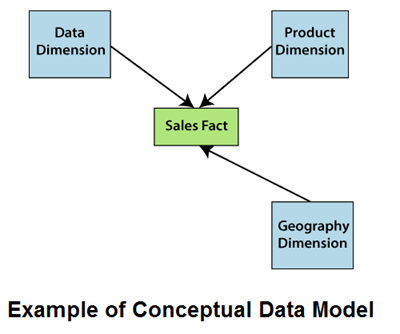
## What is Data Modeling?

* [Data modeling](https://www.geeksforgeeks.org/data-models-in-dbms/) is the process of designing a visual representation of a system or database to establish how data will be stored, accessed, and managed.
* In the context of a data warehouse, data modeling involves defining how different data elements interact and how they are organized for efficient retrieval and analysis.
* The primary goal is to create a blueprint that guides the development of the data warehouse.

**Data Modeling Life Cycle:**

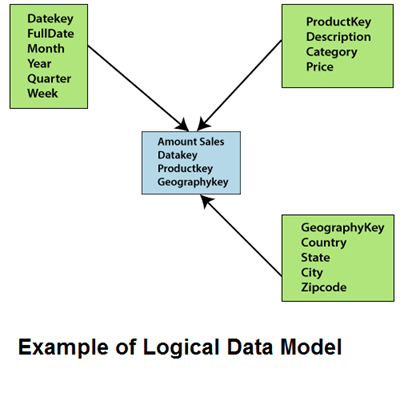
**Conceptual Data Model:** A conceptual data model recognizes the highest-level relationships between the different entities.

* **Components:** Entities, relationships, and attributes.
* **Example:** A conceptual model might define entities like "Customer," "Product," and "Sales" and illustrate the relationships between them.



**Logical Data Model:** Represents the logical structure of the data, including the relationships between entities and the data types for each attribute, without considering physical storage.

* **Components:** Tables, columns, relationships, and constraints.
* **Example:** A logical model might define tables such as "Customer," "Product," and "Sales" with their respective columns like "CustomerID," "ProductID," and "SaleDate."



**Physical Data Model:** Specifies how the data will be physically stored in the database, including indexing, partitioning, and data storage mechanisms.

* **Components:** Tables, indexes, partitions, and storage settings.
* **Example:** A physical model might define storage settings for the "Sales" table, such as partitioning by date to improve query performance.

## Types of Data Warehouse Models:

**Enterprise Warehouse:**

* An Enterprise warehouse collects all of the records about subjects of the entire organization.
* It generally contains detailed information as well as summarized information.
* It required extensive business modeling and may take years to develop and build.

**Data Mart:**

* A **data mart** is a smaller, focused part of a company's data, designed to serve the needs of specific teams or department
* Data Marts is divided into two parts:
* **Independent Data Mart:** Independent data mart is sourced from one or more operational systems.
* **Dependent Data Mart:** Dependent data marts are sourced exactly from enterprise data-warehouses.

**Virtual Data Warehouse**

* A **virtual data warehouse** is a set of views or perceptions over an operational database.
* It doesn't physically store the data but creates virtual views that allow users to query the data as if it were a separate warehouse.

