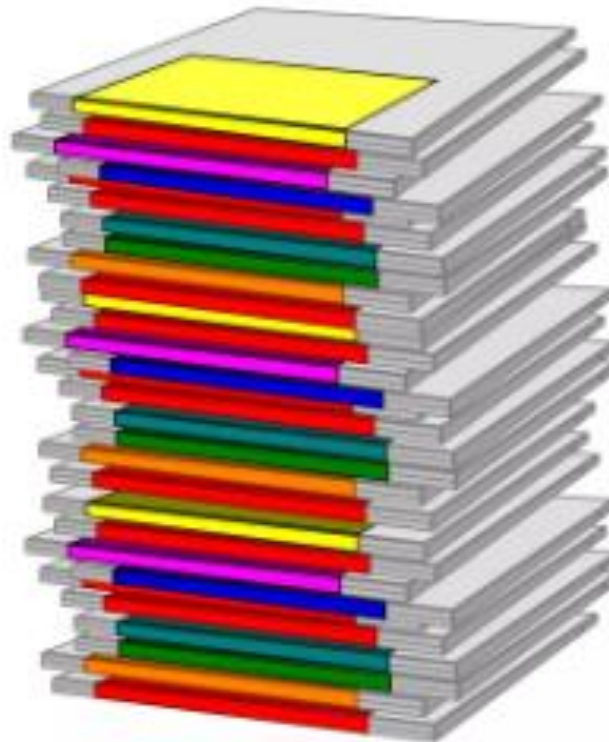


Introduction of Data Warehouse

What is a Data Warehouse?

A single, complete and consistent store of data obtained from a variety of different sources made available to end users in a way that they can understand and use in a business context.

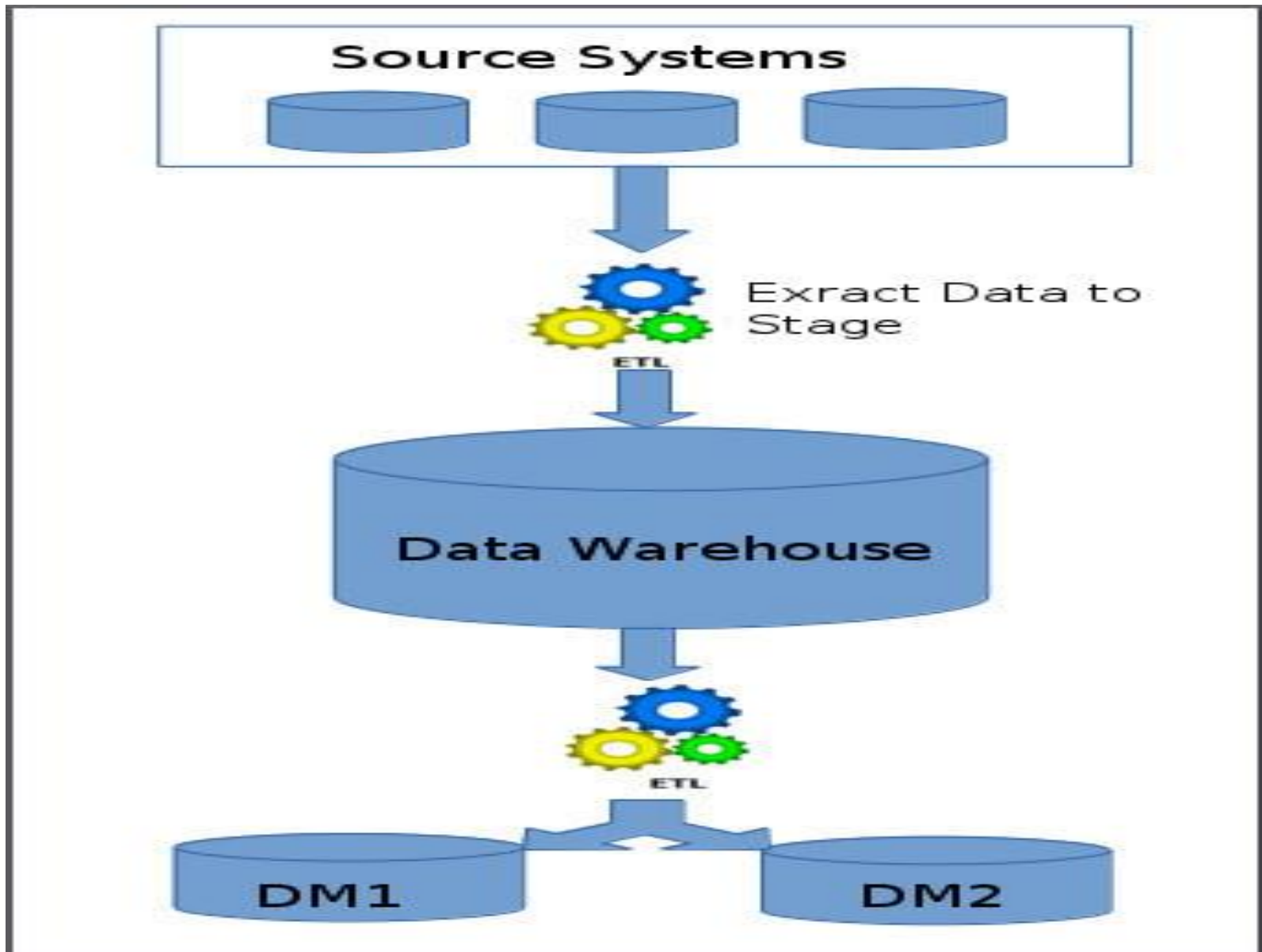


Characterstics

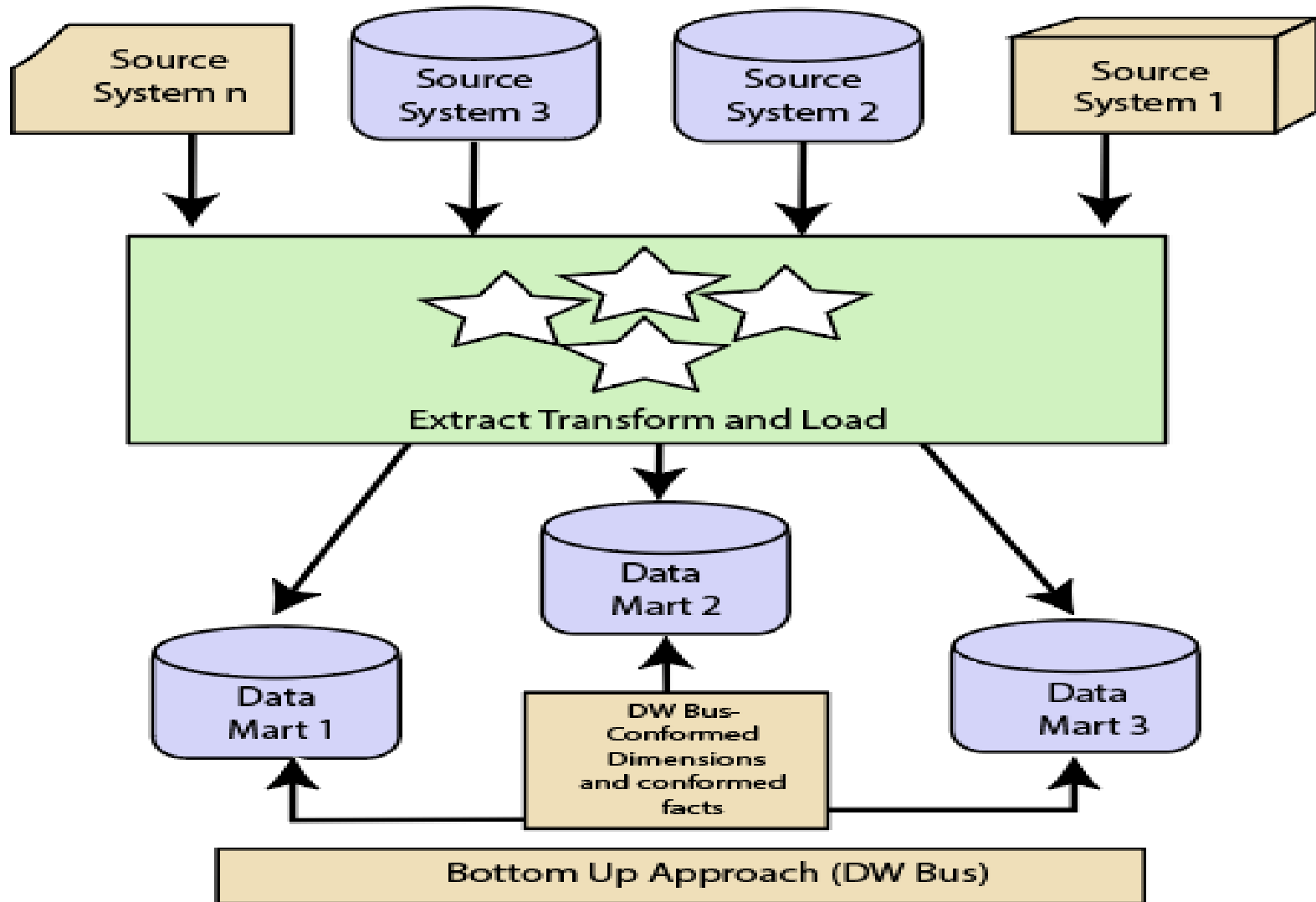
Data warehousing is ...

- **Subject Oriented:** Data that gives information about a particular subject instead of about a company's ongoing operations.
- **Integrated:** Data that is gathered into the data warehouse from a variety of sources and merged into a coherent whole.
- **Time-variant:** All data in the data warehouse is identified with a particular time period.
- **Non-volatile:** Data is stable in a data warehouse. More data is added but data is never removed. This enables management to gain a consistent picture of the business.

Top Down Approach



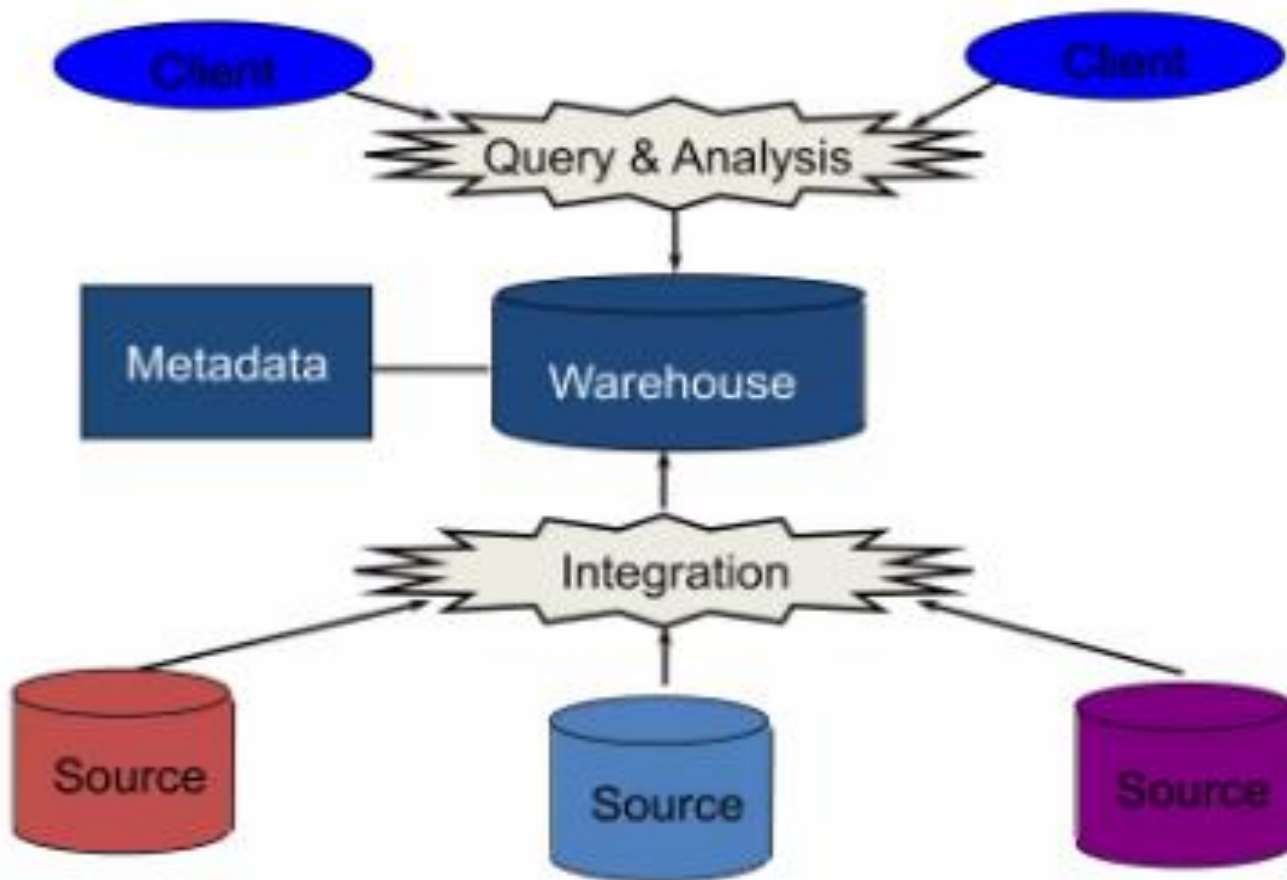
Bottom UP Approach



Bottom Up Design Approach

Architecture

Data Warehouse Architecture



Types of Data Warehouse Architectures

There are mainly three types of Datawarehouse Architectures



The diagram features a central blue circle with the text "Types of Data Warehouse Architectures". To the right of this circle are three horizontal bars, each preceded by a colored sphere. The top bar is teal with a teal sphere and is labeled "Single-Tier Architecture". The middle bar is orange with an orange sphere and is labeled "Two-Tier Architecture". The bottom bar is pink with a pink sphere and is labeled "Three-Tier Architecture".

Types of
Data
Warehouse
Architectures

Single-Tier Architecture

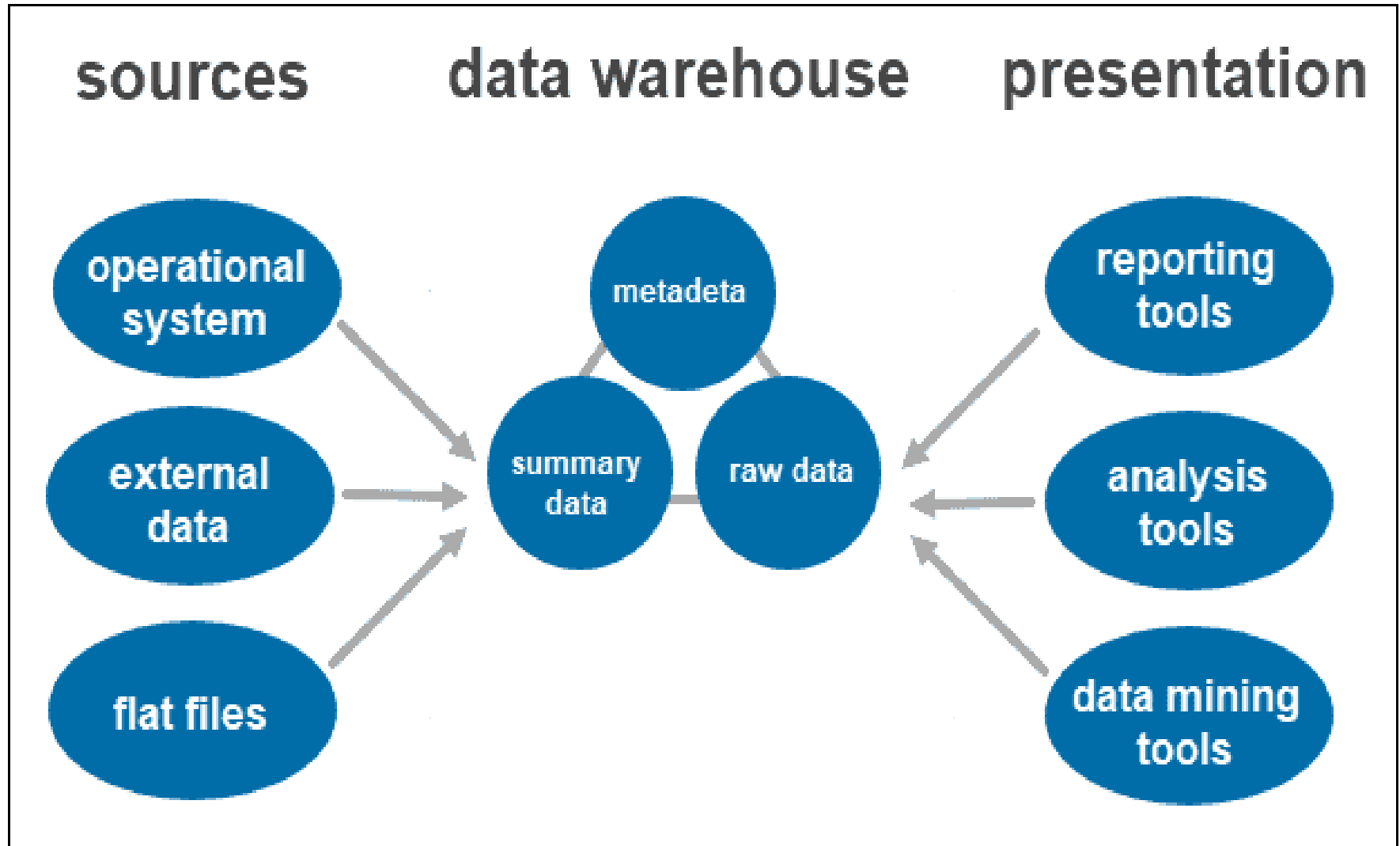
Two-Tier Architecture

Three-Tier Architecture

Single-tier data warehouse architecture

1. The main goal of having such architecture is to remove redundancy by minimizing the amount of data stored.
2. Its primary disadvantage is that it doesn't have a component that separates analytical and transactional processing

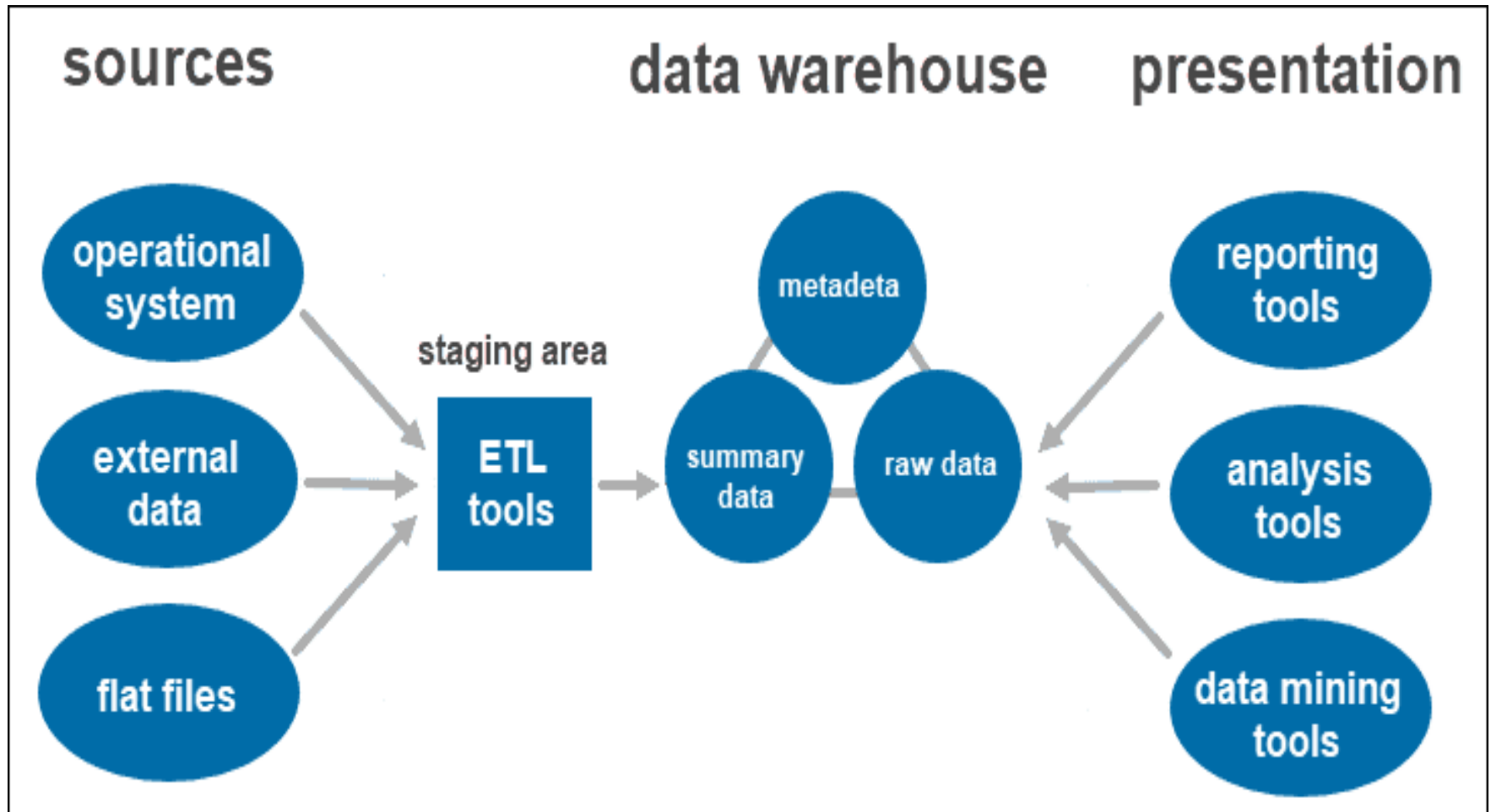
Single-tier-Data-Warehouse



Two-tier data warehouse architecture

1. The two-tier architecture includes a staging area for all data sources, before the data warehouse layer.
2. By adding a staging area between the sources and the storage repository, you ensure all data loaded into the warehouse is cleansed and in the appropriate format.

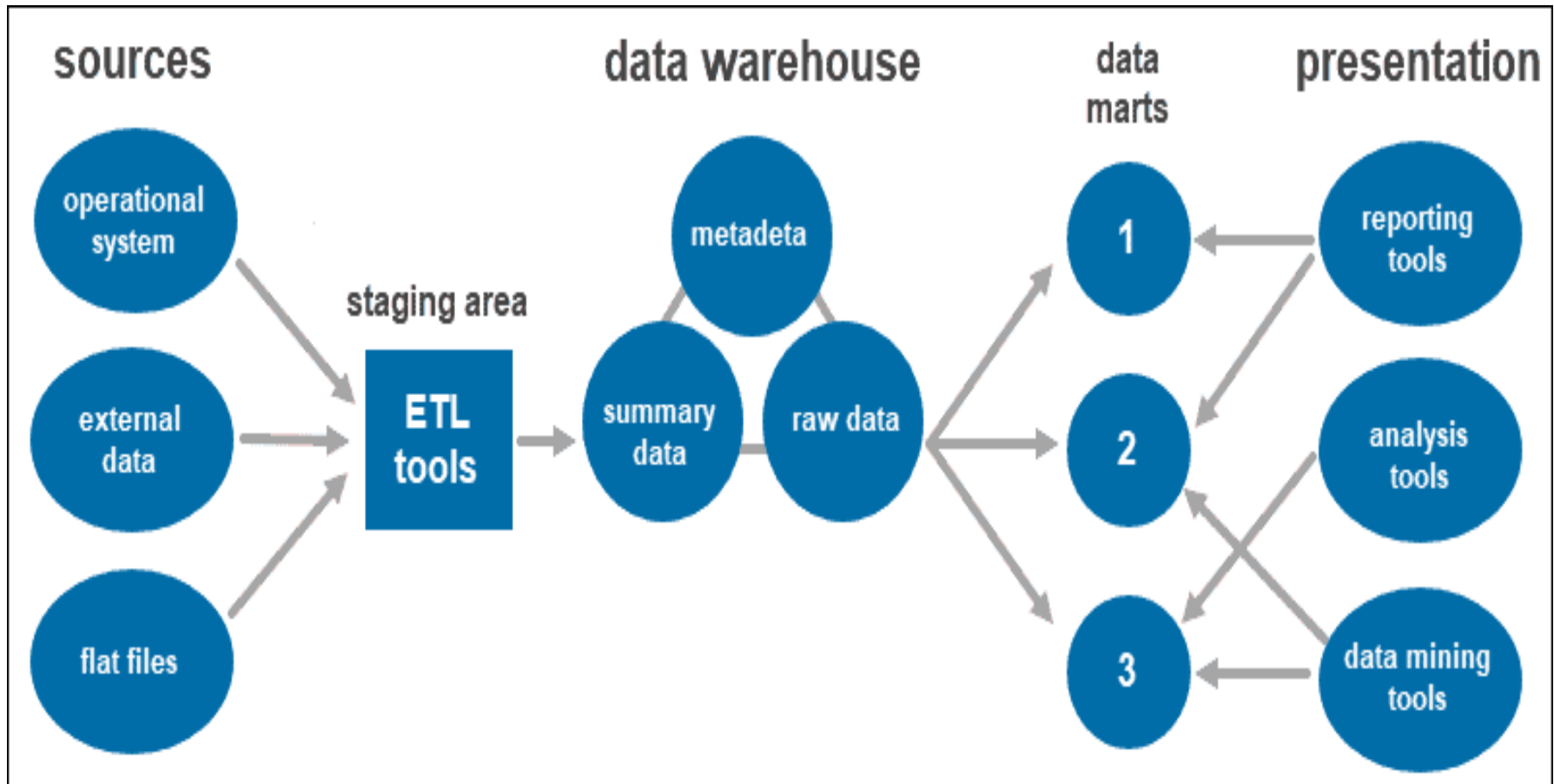
Two-tier-Data-Warehouse



Three-Tier Data Warehouse Architecture

- Data Warehouses usually have a three-level (tier) architecture that includes:
 1. Bottom Tier (Data Warehouse Server)
 2. Middle Tier (OLAP Server)
 3. Top Tier (Front end Tools).
- A **bottom-tier** that consists of the **Data Warehouse server**, which is almost always an RDBMS. It may include several specialized data marts and a metadata repository.
- Data from operational databases and external sources (such as user profile data provided by external consultants) are extracted using application program interfaces called a gateway. A gateway is provided by the underlying DBMS and allows customer programs to generate SQL code to be executed at a server.
- **Examples** of gateways contain **ODBC** (Open Database Connection) and **OLE-DB** (Open-Linking and Embedding for Databases), by **Microsoft**, and **JDBC** (Java Database Connection).

Three-tier-Data-Warehouse



Dimensional Modeling

- **Dimensional modeling** (DM) names a set of techniques and concepts used in data warehouse design.
- Dimensional modeling is one of the methods of data modeling, that help us store the data in such a way that it is relatively easy to retrieve the data from the database.
- Dimensional modeling always uses the concepts of facts (measures), and dimensions (context).

Dimensional Modeling Steps

Classify the
business
objective

Recognize
Granularity

Classifying
Dimensions,
Attributes

Recognize the
Fact

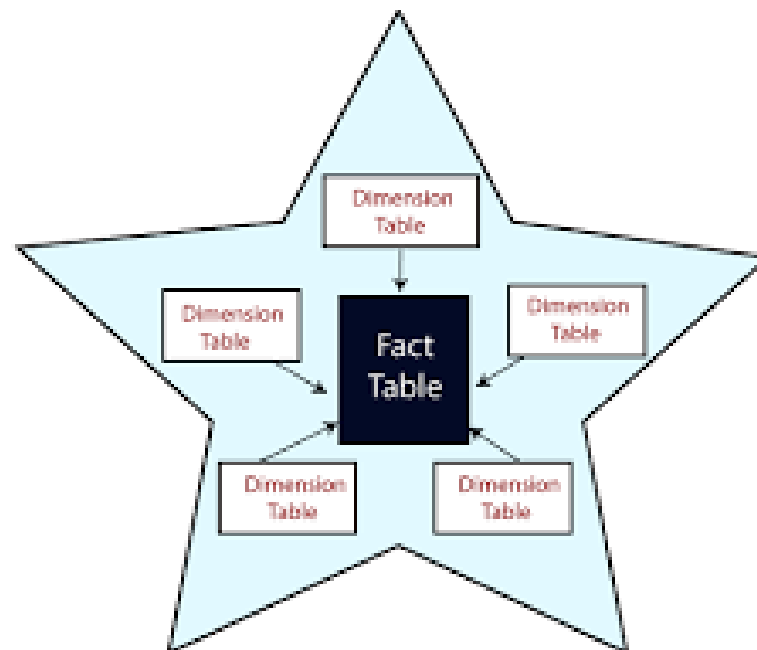
Structure the
Schema

Conceptual Modeling of Data Warehouses

- Star schema
- Snowflake schema

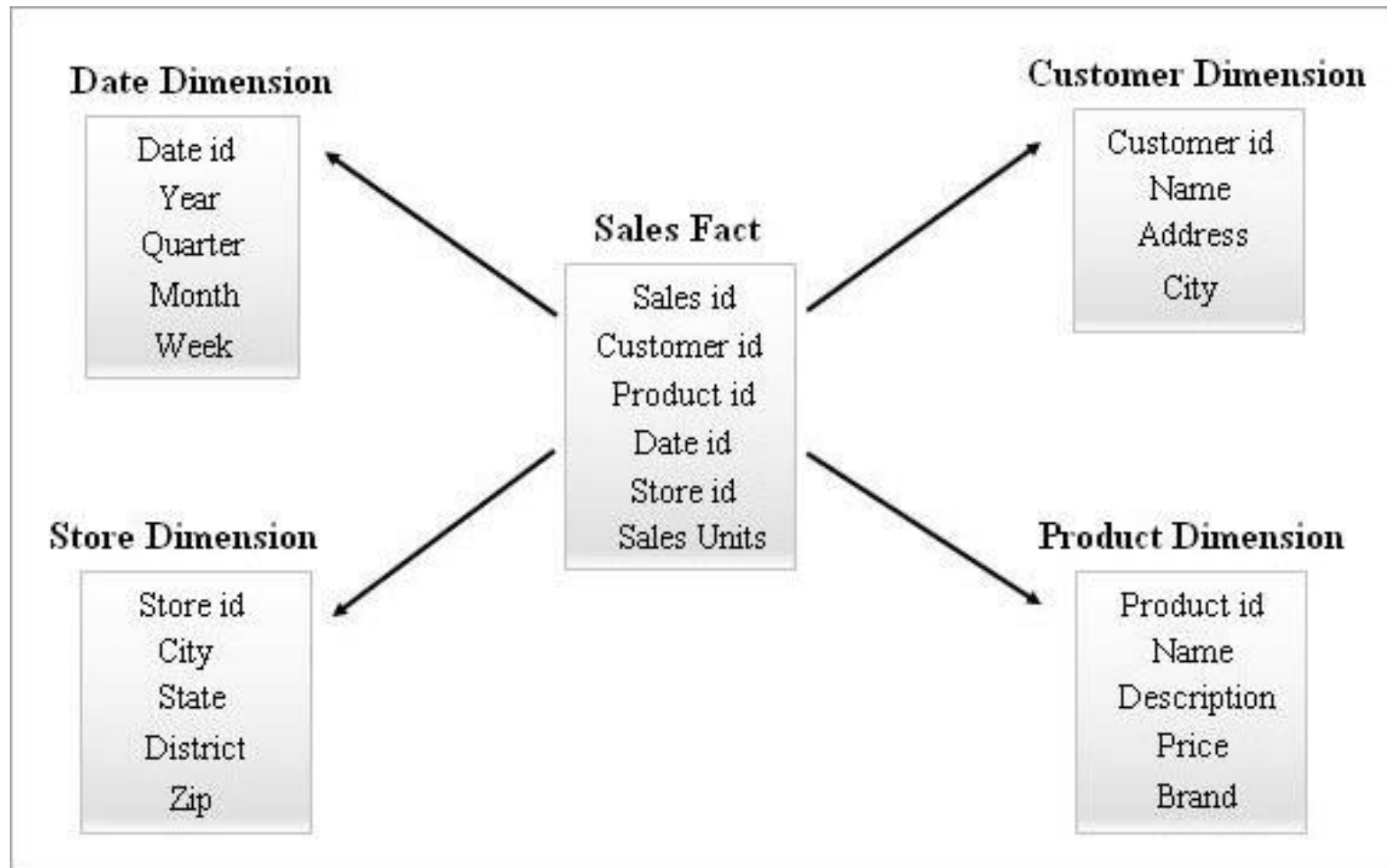
Star Schema:

- Star Schema is a relational database schema for representing multidimensional data. It is the simplest form of data warehouse schema that contains one or more dimensions and fact tables.
- It is called a star schema because the entity-relationship diagram between dimensions and fact tables resembles a star where one fact table is connected to multiple dimensions.
- The center of the star schema consists of a large fact table and it points towards the dimension tables.



Star Schema

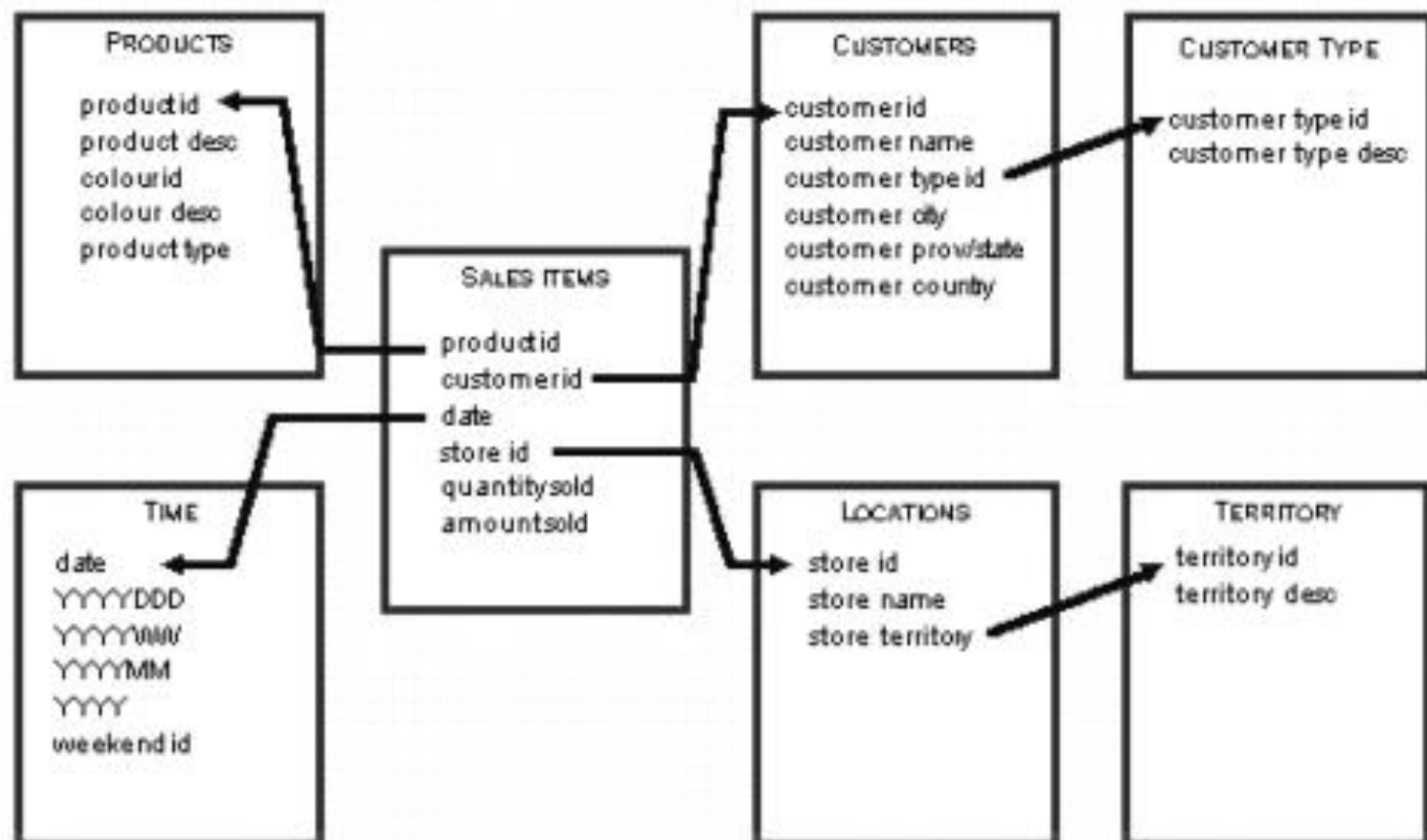
Star Schema Example



Snowflake Schema

- The snowflake schema is a variant of the star schema.
- Here, the centralized fact table is connected to multiple dimensions.
- In the snowflake schema, dimension are present in a normalized form in multiple related tables.
- The snowflake structure materialized when the dimensions of a star schema are detailed and highly structured, having several levels of relationship, and the child tables have multiple parent table.
- The snowflake effect affects only the dimension tables and does not affect the fact tables.

Snowflake Schema



Dimension Table

customer_id	customer_name
1	Customer 1 Name
2	Customer Name 1
3	Customer Name 3

order_id	customer_id	product_id
1	1	3
2	2	1
3	3	2

Fact Table

Dimension Table

order_id	shipment_to_city
1	Detroit
2	New York
3	Las Vegas

Dimension Table

product_id	product_name
1	Football
2	Basketball
3	Golf Ball



