

DATA WAREHOUSING AND MINING

T.E. D.S, Sem V
Academic Year: 2023-24

Data Warehousing Fundamentals: Characteristics and Architecture

Poonam Pangarkar

Data Warehouse

A Data Warehouse:

- Takes all the data from the operational systems.
- Where necessary, include relevant data from outside
- Integrate all the data from various sources.
- Remove inconsistencies and transform the data.
- Store the data in formats suitable for easy access for decision making.

It involves different functions: data extraction, the function of loading the data, transforming the data, storing the data, and providing user interfaces.

Data Warehouse Architecture

“A Data Warehouse is a **subject oriented, integrated, nonvolatile, and **time variant** collection of data in support of management’s decisions.”**

- Bill Inmon, considered to be the father of Data Warehousing

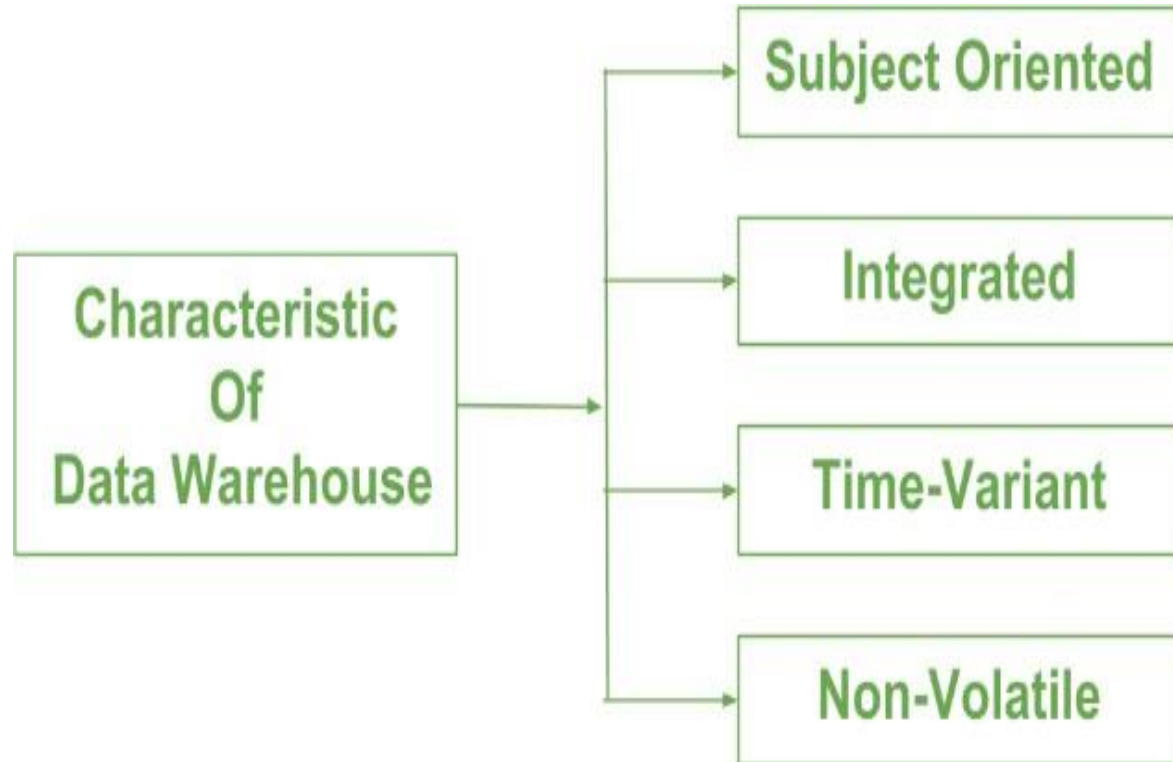
In order to set up this information delivery system, different components or building blocks are needed.

These building blocks are arranged together in the most optimal way to serve the intended purpose.

They are arranged in a suitable architecture

Characteristics of Data Warehouse

- ✓ Subject-Oriented
- ✓ Integrated
- ✓ Time-Variant
- ✓ Non-Volatile



Characteristics of Data Warehouse

Subject Oriented :

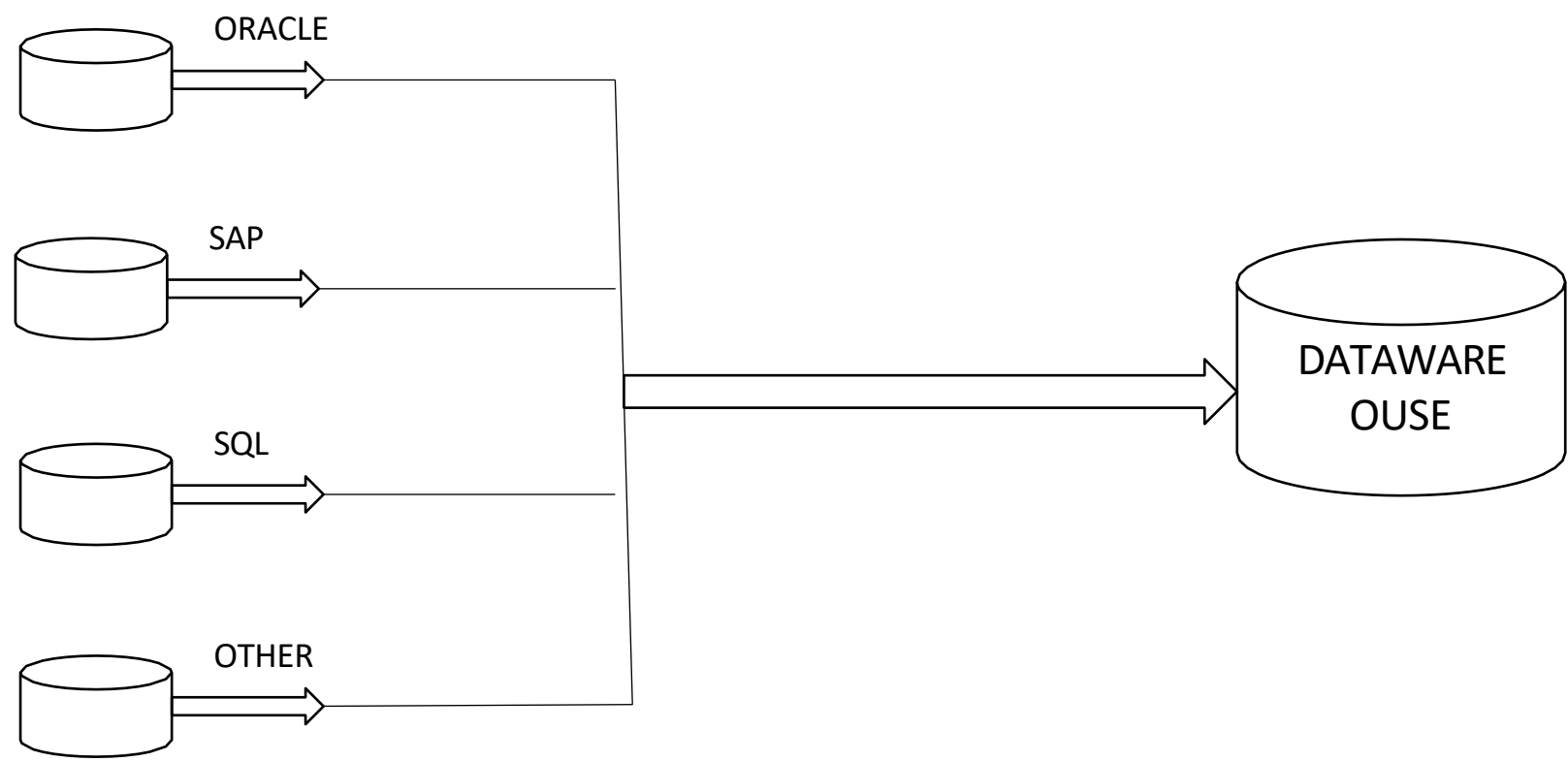
- ✓ In operational systems, data is organized around the applications.
- ✓ In data warehouse the data is stored by the subjects.
- ✓ A data warehouse can be used to analyze a particular subject area. For example, "sales" can be a particular subject.
- ✓ Similarly data can be gathered around subjects such as **Customer, Inventory**
- ✓ It also delivers an easy and precise information around particular theme by eliminating data which is not required to make the decisions.

Characteristics of Data Warehouse

Integrated:

- Data needs to be pulled together from various applications.
- Source data can be in different databases, files, the applications can be on various platforms.
- Before the data from various disparate sources can be usefully stored in a data warehouse, inconsistencies have to be removed.
- Before moving the data into the data warehouse, it has to go through a process of *transformation, consolidation, and integration*.

INTEGRATED



Characteristics of Data Warehouse

Time Variant:

A data warehouse has to contain historical data, not just current values.

Data is stored as snapshots over past and current periods.

Every data structure in the data warehouse contains the time element.

This aspect of the data warehouse is quite significant for both the design and the implementation phases.

Characteristics of Data Warehouse

The **time-variant** nature of the data in a data warehouse

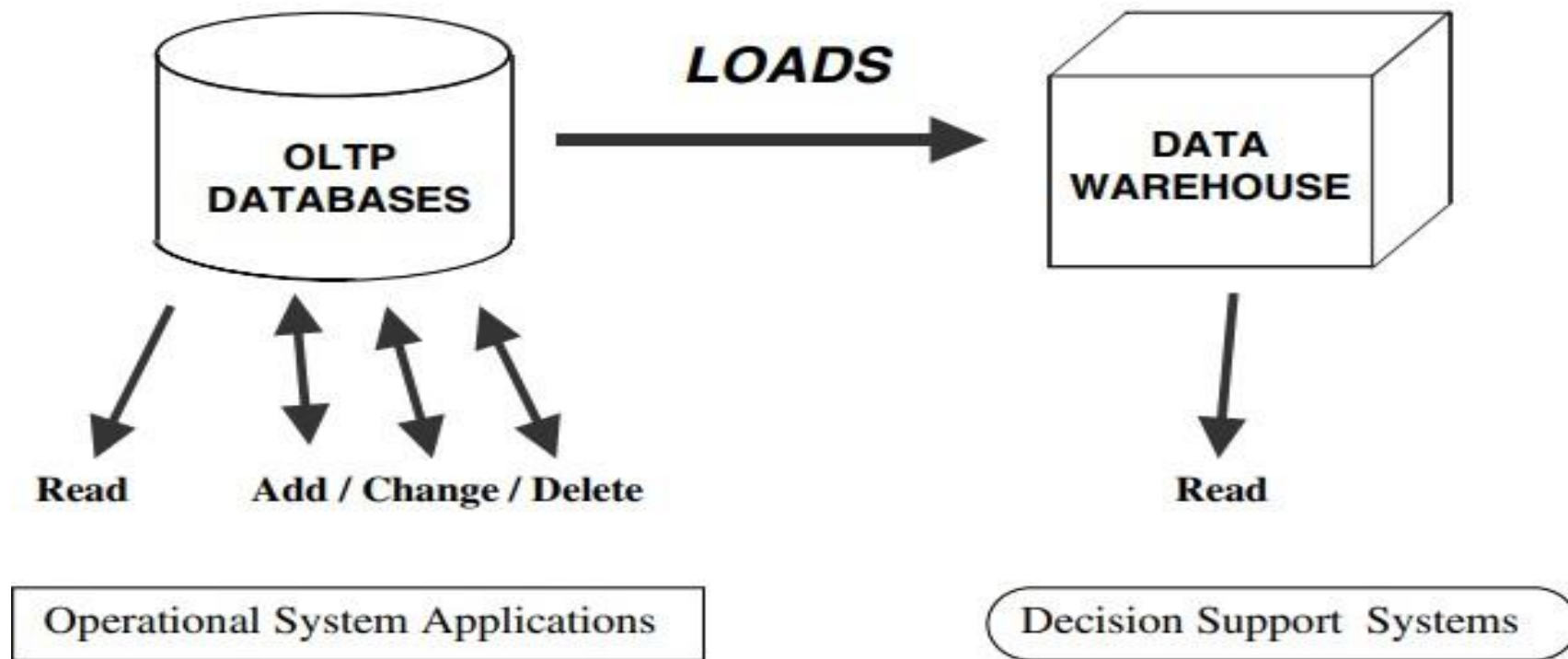
- ✓ Allows for analysis of the past
- ✓ Relates information to the present
- ✓ Enables forecasts for the future

Characteristics of Data Warehouse

Non-volatile Data:

- ✓ The data in the data warehouse is not intended to run the day-to-day business.
- ✓ Data warehouse is not updated every time a single transaction is processed.
- ✓ Data from the operational systems are moved into the data warehouse at specific intervals.
- ✓ Depending on the requirements of the business, these data movements take place twice a day, once a day, once a week, or once in two weeks.
- ✓ Data movements to different data sets may take place at different frequencies
- ✓ The data in a data warehouse is not as volatile as the data in an operational database is
- ✓ The data in a data warehouse is primarily for query and analysis.

Usually the data in the data warehouse is not updated or deleted.



Data warehouse Architecture

Each data warehouse is different, but all are characterized by standard vital components.

It is overall architecture of data communication processing and presentation for the end client.

Three common architectures are:

- Data Warehouse Architecture: Basic
- Data Warehouse Architecture: With Staging Area
- Data Warehouse Architecture: With Staging Area and Data Marts

Data warehouse Architecture

Architecture is the proper arrangement of the components.

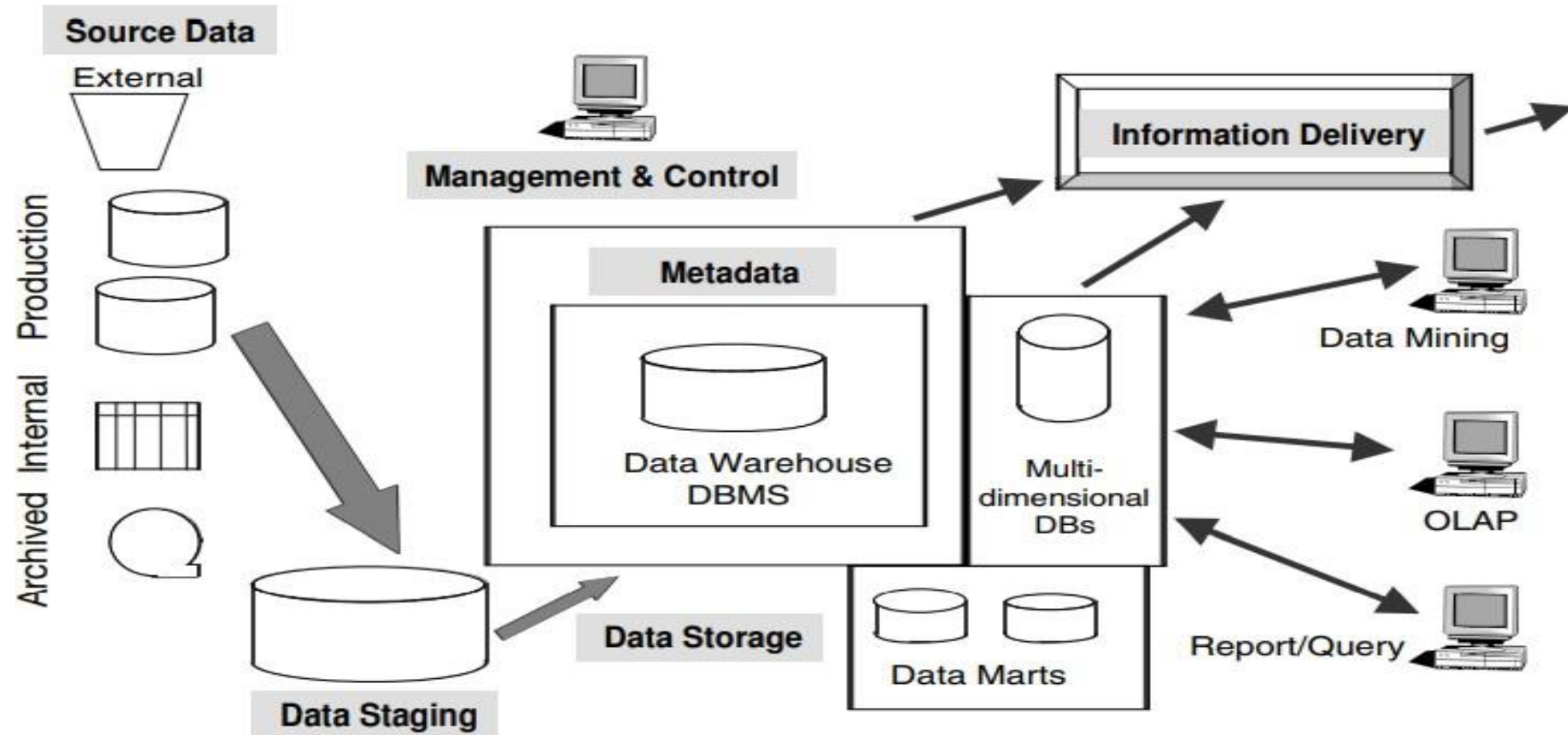
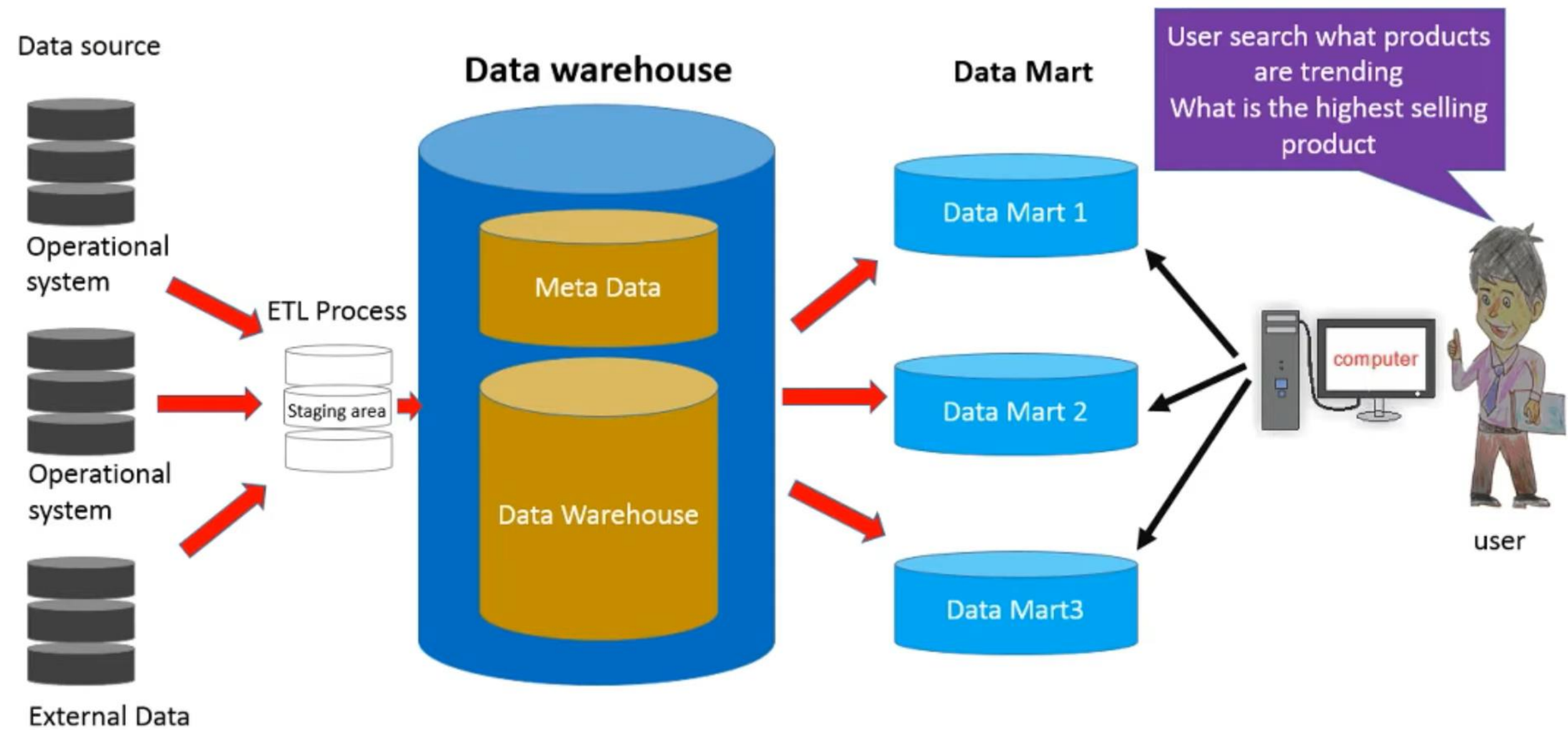
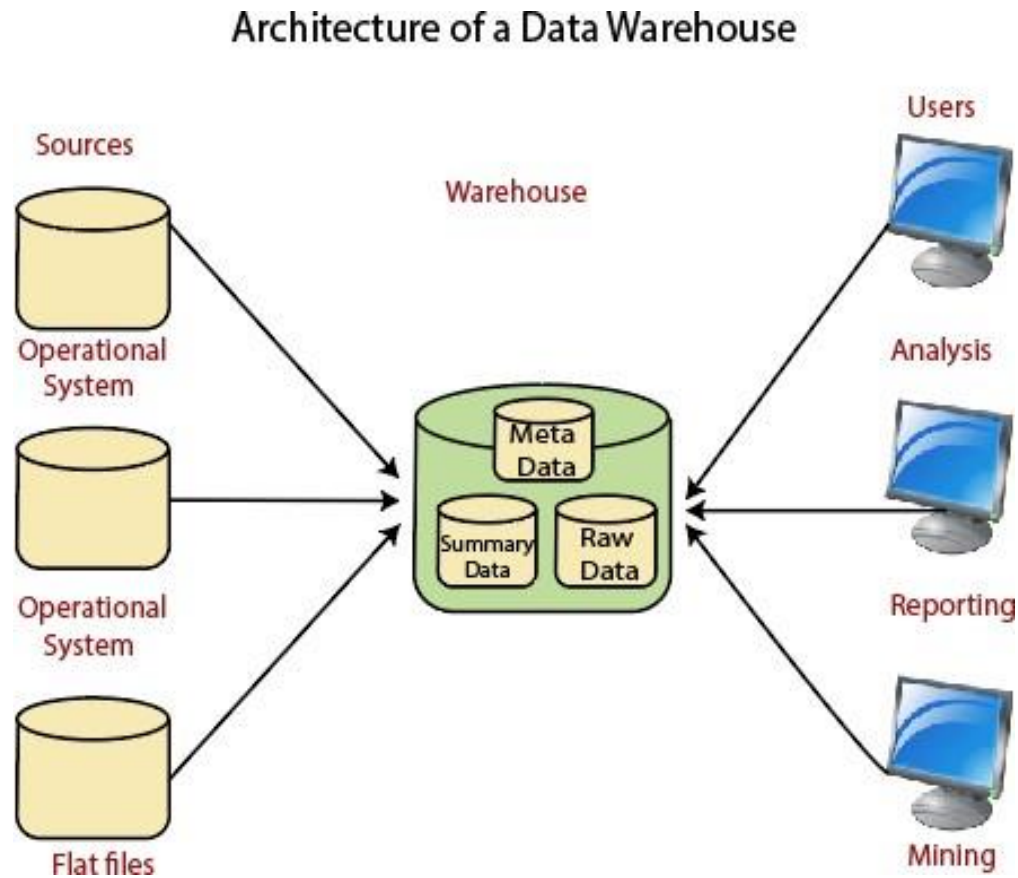


Figure 2-6 Data warehouse: building blocks or components.



Data warehouse architecture

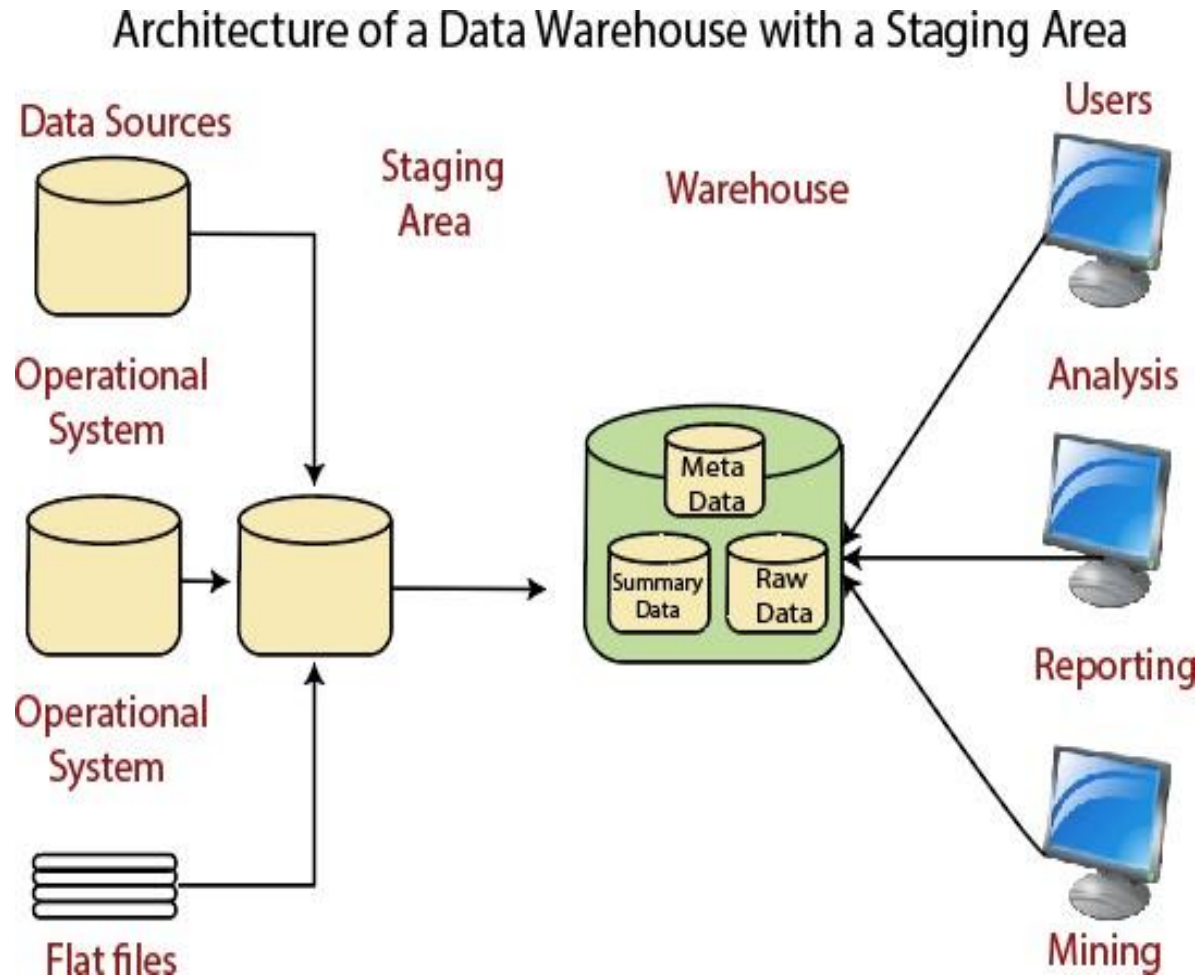
Data Warehouse Architecture: Basic



Meta Data

- A set of data that defines and gives information about other data.
- Meta Data summarizes necessary information about data, which can make finding and work with particular instances of data more accessible.
- For example, author, data build, data changed, and file size are examples of very basic document metadata.
- Metadata is used to direct a query to the most appropriate data source.

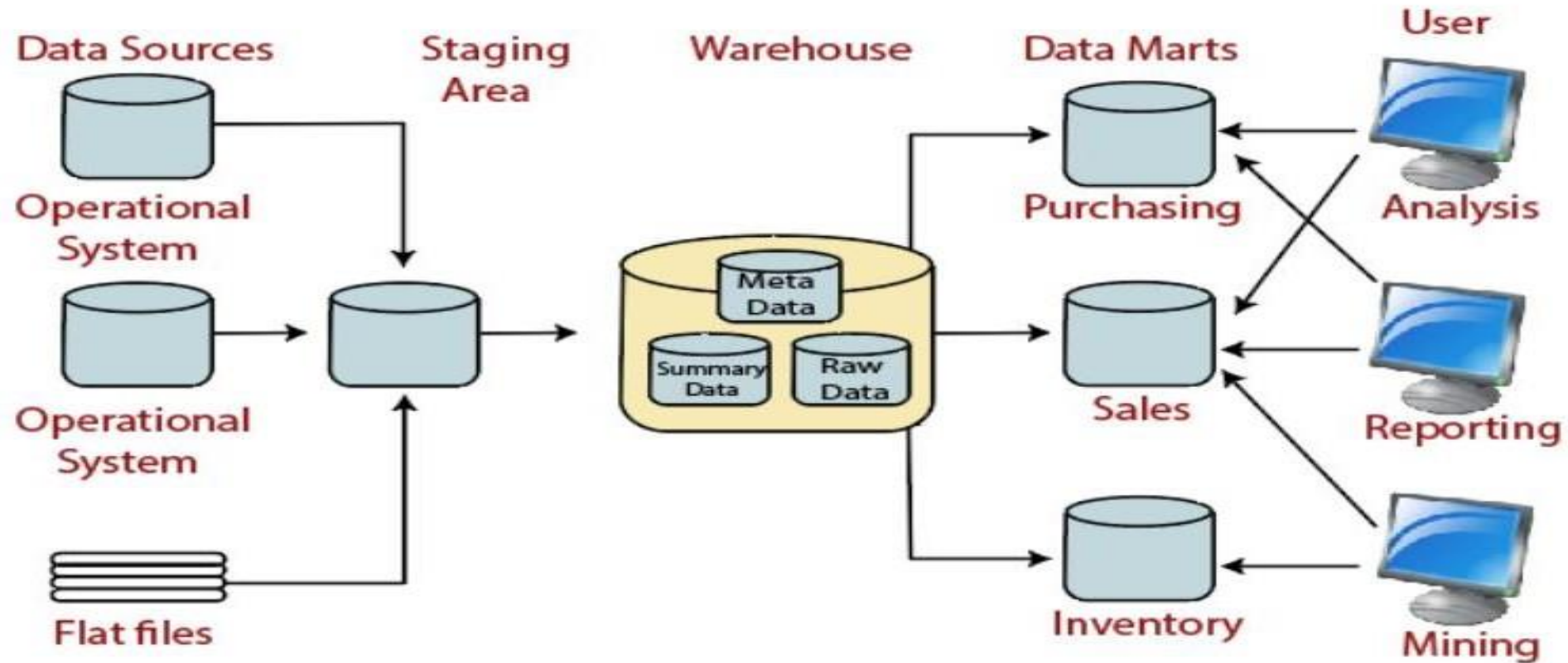
Data Warehouse Architecture: With Staging Area



A staging area simplifies data cleansing and consolidation for operational method coming from multiple source systems, especially for enterprise data warehouses where all relevant data of an enterprise is consolidated.

Data Warehouse Architecture: With Staging Area and Data Marts

Architecture of a Data Warehouse with a Staging Area and Data Marts



Datawarehouse and Datamart

- A data mart is a subset of a Data Warehouse
- The different departments of an organization (like finance, sales, marketing, HR) can have their own datamarts .
- Top down or Bottom up
- The datamart does not contain detailed data of organization
- Datamart are divided into two types
 - 1. Dependent datamart
 - 2. independent datamart

Reasons for creating a datamart

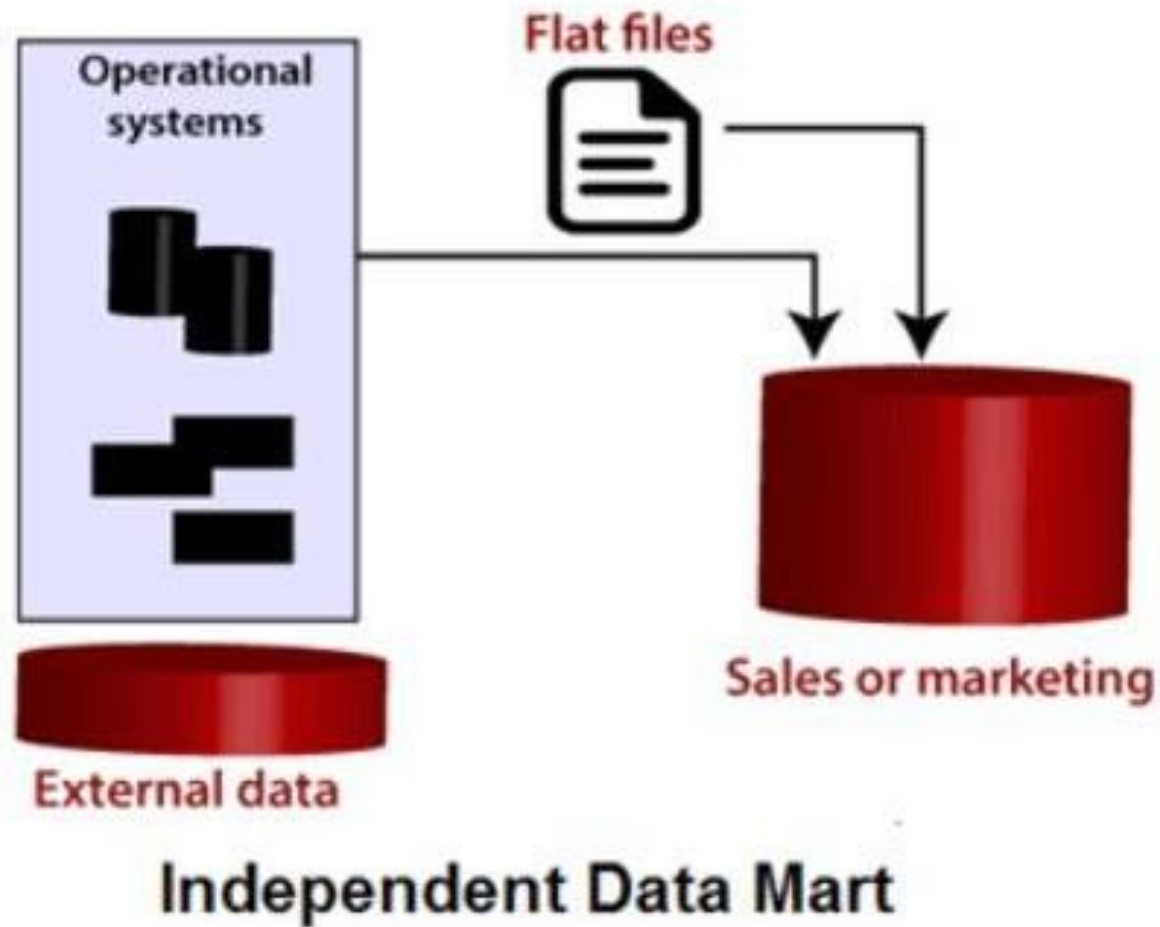
- Creates collective data by a group of users
- Easy access to frequently needed data
- Ease of creation
- Improves end-user response time
- Lower cost than implementing a complete data warehouses
- Potential clients are more clearly defined than in a comprehensive data warehouse
- It contains only essential business data and is less cluttered.

Dependent Datamart



Dependent Data Mart

Independent Datamart



DEPENDENT DATA MARTS

A logical subset or a physical subset of a larger data warehouse

Get data from a central data warehouse that has already been created

Obtain data by aggregating, restricting and summarizing the data of the data warehouse

Depend on the central data warehouse

Less secure

Control over data is less

INDEPENDENT DATA MARTS

A type of data mart that draws data from different sources without using the central data warehouse

Get data directly from operational sources and/or external sources of data

Get data from multiple transaction systems in one subject area or department to support specific business needs

Standalone data marts and do not depend on the central data warehouse

More secure

Have more control over their data

Data warehouse design strategies

1. Bottom up design :

- In this approach, data marts are first created to provide reporting and analytical capabilities for specific business processes or departments, data marts provide atomic data or summarizes if needed.
- These data marts are then integrated together to create enterprise DW.
- The most important management task in this approach is to ensure that the dimensions among the data marts are consistent.

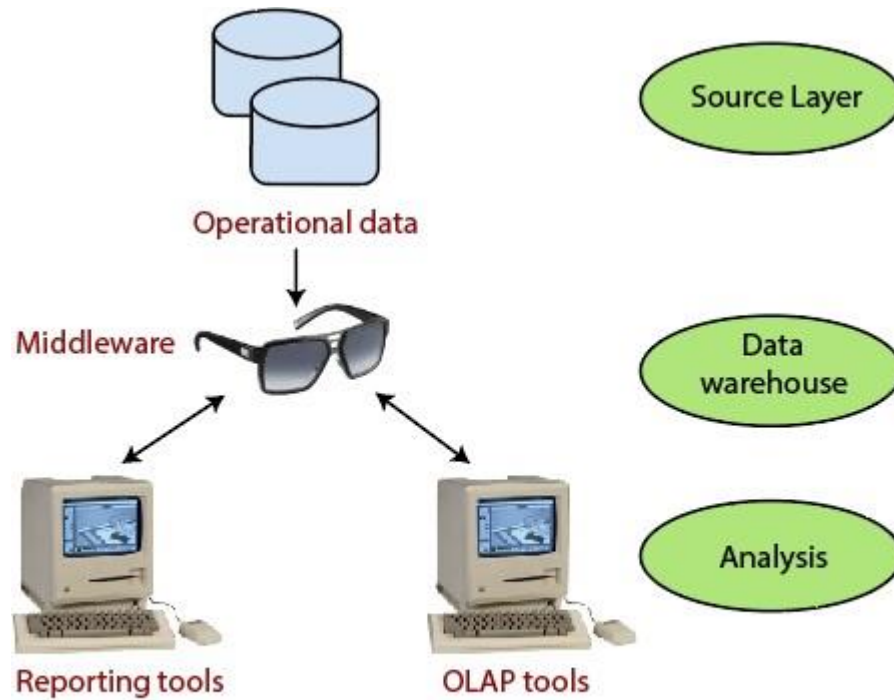
2. Top down design :

- In this approach, the DW is first created using a normalised enterprise data model.
- Atomic data i.e. data at the lowest possible level, are stored in DW, dimensional data marts are created from DW for specific business processors or departments are created from DW.
- This approach generates highly consistent dimensional views of data across data marts since all the data marts are created from the central re position. it is also proven to be robust against the business changes.

Data warehouse vs Datamart

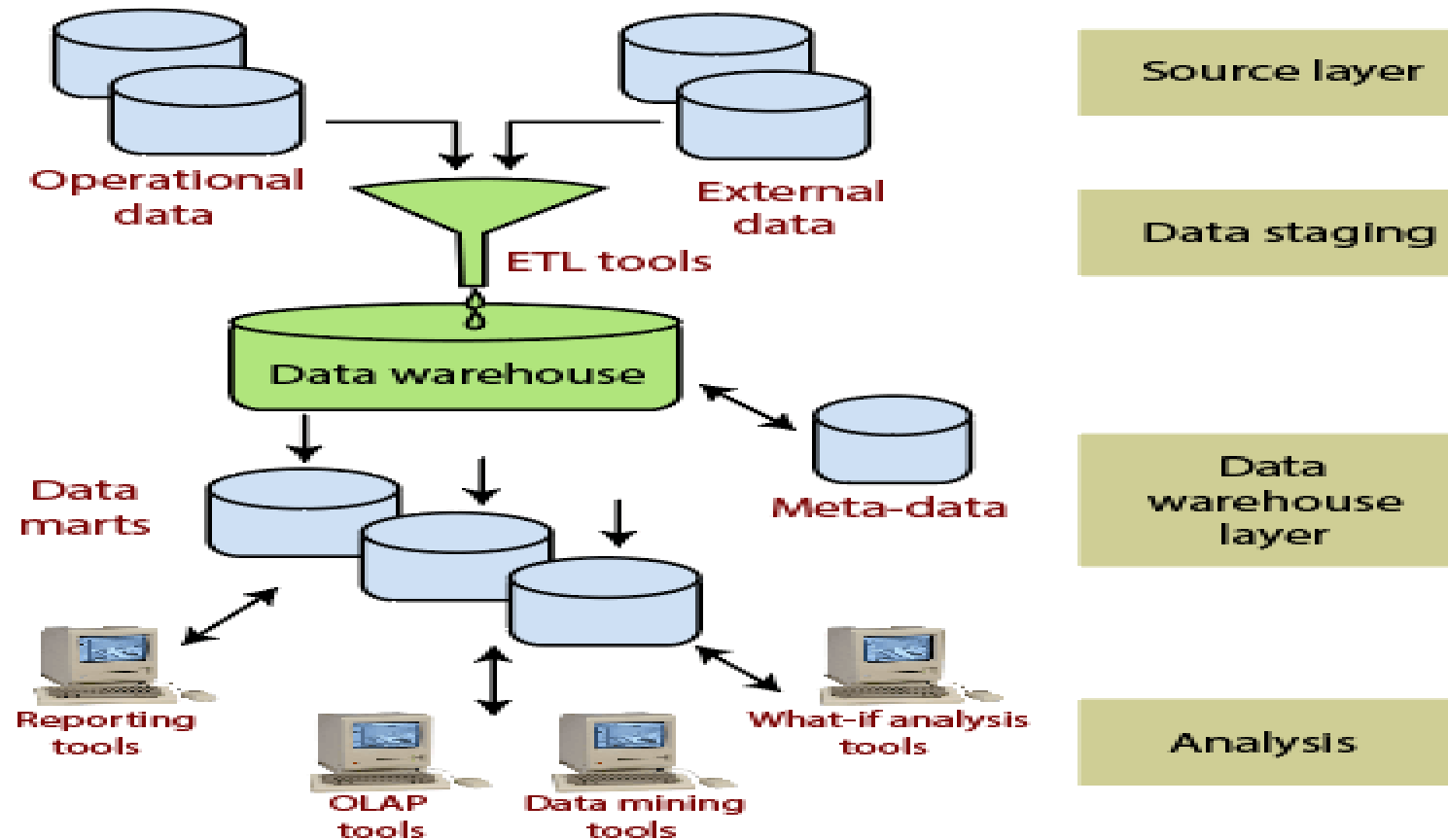
Attributes	Datawarehouse	Datamarts
View	It gives overall enterprise view	It gives departmental view
Time	DW takes more time to implement	DM takes less time to implement
Size	DW size is very large (>100TB)	DM size is less (>10TB)
Response	Response from DW is slow as compared to DM	Response from DM is fast as compared to DW
Union	DW is an union of all departments	DM is a single business process(a piece if DW)

Single Tier Architecture



Single-Tier Data Warehouse Architecture

Two Tier Architecture



Two-Tier Data Warehouse Architecture