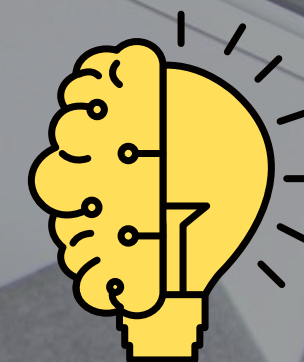


# UNIT 2

## *The Architecture of Data Warehouse and Business Intelligence*

## *Business Intelligence*



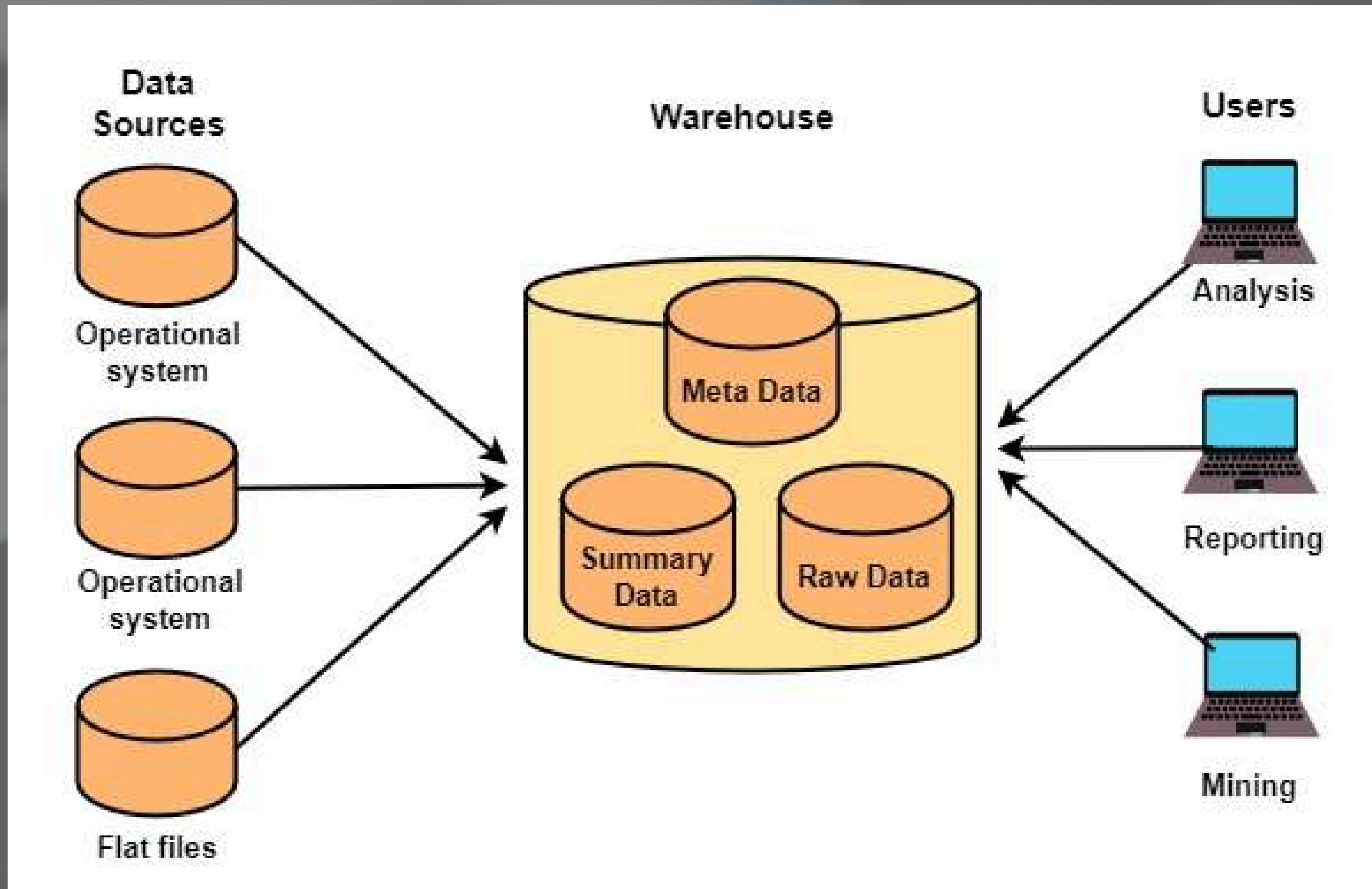
# Data Warehouse-Definition

- A Data Warehouse is a large storage system.
- It is used by businesses to collect, store, and manage data.
- Data is gathered from different sources.
- The data is stored in an organized format.
- It helps in analyzing data effectively.
- Supports better business decision-making



# Data Warehouse-Architecture

- Block Diagram



# Data Warehouse-Architecture

- A Data Warehouse follows a three-tier architecture
  - Helps in organizing and processing data efficiently
  - Improves data analysis and decision making
- a. Bottom Tier → Data Source Layer
  - b. Middle Tier → Storage & Processing Layer
  - c. Top Tier → Presentation Layer



# Data Warehouse-Architecture

## a. Bottom Tier (Data Source Layer)

- Foundation of the Data Warehouse
- Contains databases, files, and external sources
- Raw data is collected from different sources
- Data is extracted, transformed, and loaded using ETL process

## b. Middle Tier (Storage & Processing Layer)

- Stores processed data in the warehouse database
- Uses OLAP technology for fast queries
- Organizes data efficiently
- Supports analysis and business decisions



# ETL Process (Extract, Transform, Load)

## Definition :-

- ETL is a process used in Data Warehousing
- It collects data from different sources
- Cleans and organizes data
- Stores data in structured form for analysis

## Steps of ETL

- Extraction
- Transformation
- Loading



# ETL Process (Extract, Transform, Load)

## Extraction

- First step of ETL
- Raw data is collected from multiple sources
- Sources can be databases, files, cloud apps, spreadsheets
- Goal → gather all required data

## Transformation

- Data is cleaned and formatted
- Errors and duplicates are removed
- Data types may be changed
- Makes data accurate and consistent



# ETL Process (Extract, Transform, Load)

## Why Transformation is Important

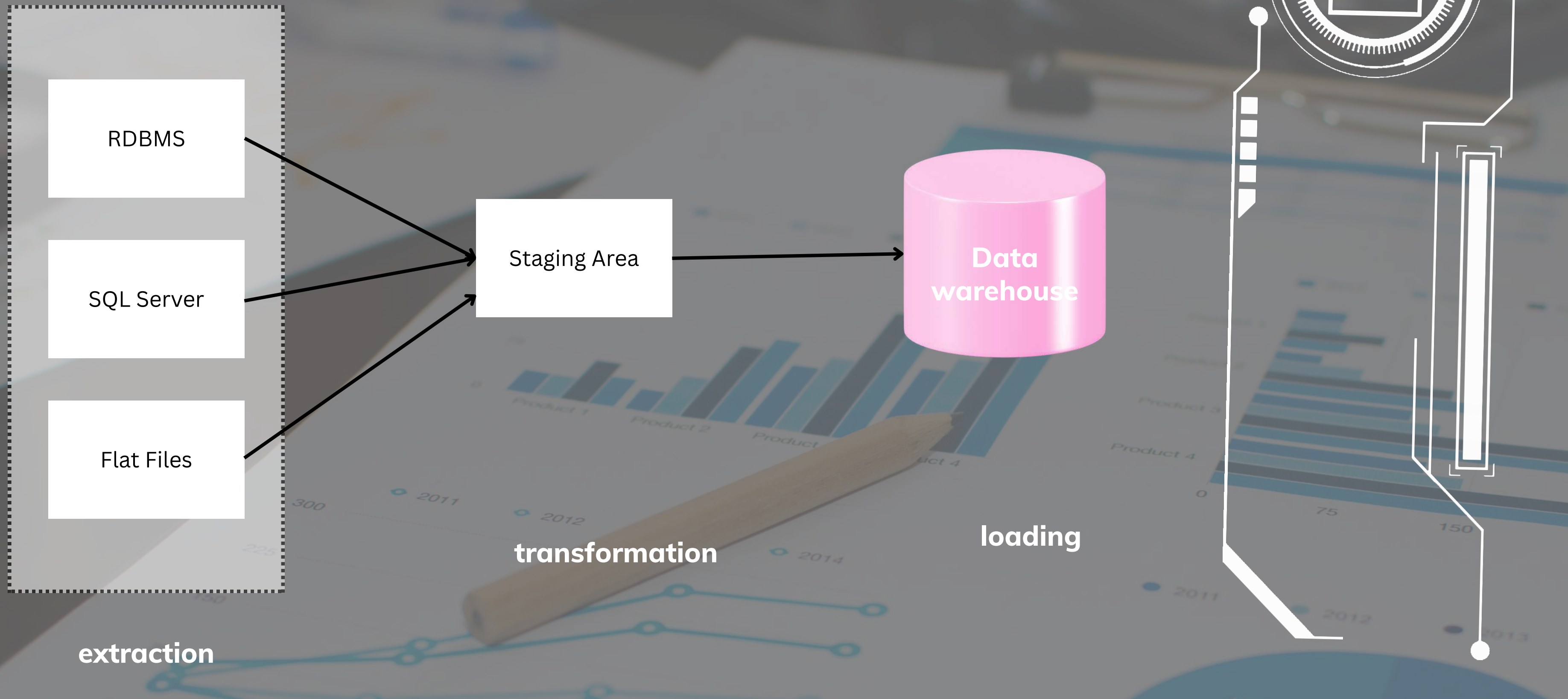
- Improves data quality
- Makes data reliable
- Prepares data for analysis
- Supports better decision making

## Loading

- Final step of ETL
- Processed data is stored in Data Warehouse or Data Lake
- Data can be loaded:
  - In batches (scheduled time)
  - In real time



# ETL Process Diagram :-



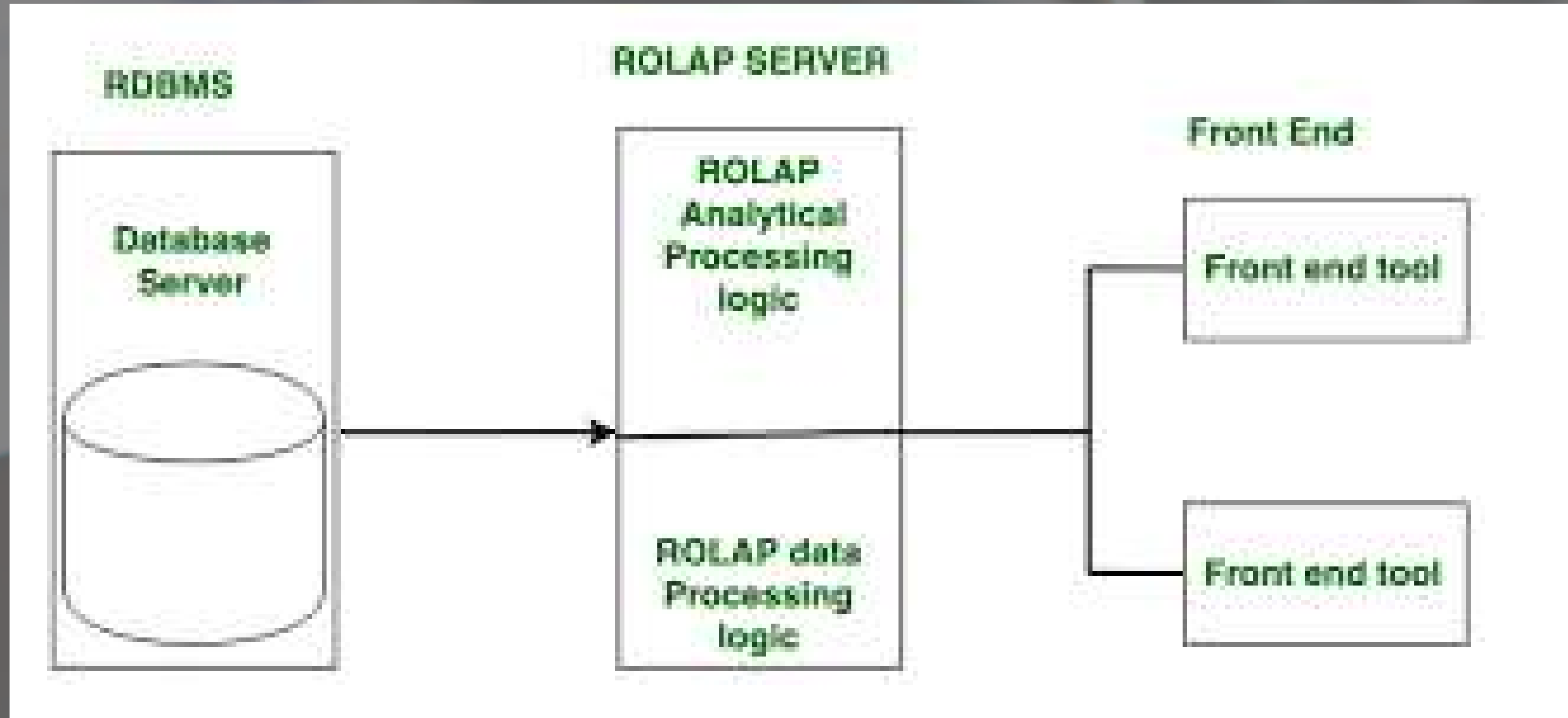
# Difference between DW and BI

Data Warehouse	Business Intelligence (BI)
Stores large amounts of structured data	Analyzes data to generate insights
Acts as a centralized data repository	Acts as a decision-support system
Focuses on data storage and management	Focuses on reporting and visualization
Collects data from multiple sources	Uses stored data for analysis
Backend system	Frontend analytical layer
Handles historical data	Interprets data trends
Built using databases and ETL tools	Uses dashboards, reports, charts
Main goal: store and organize data	Main goal: support decision making

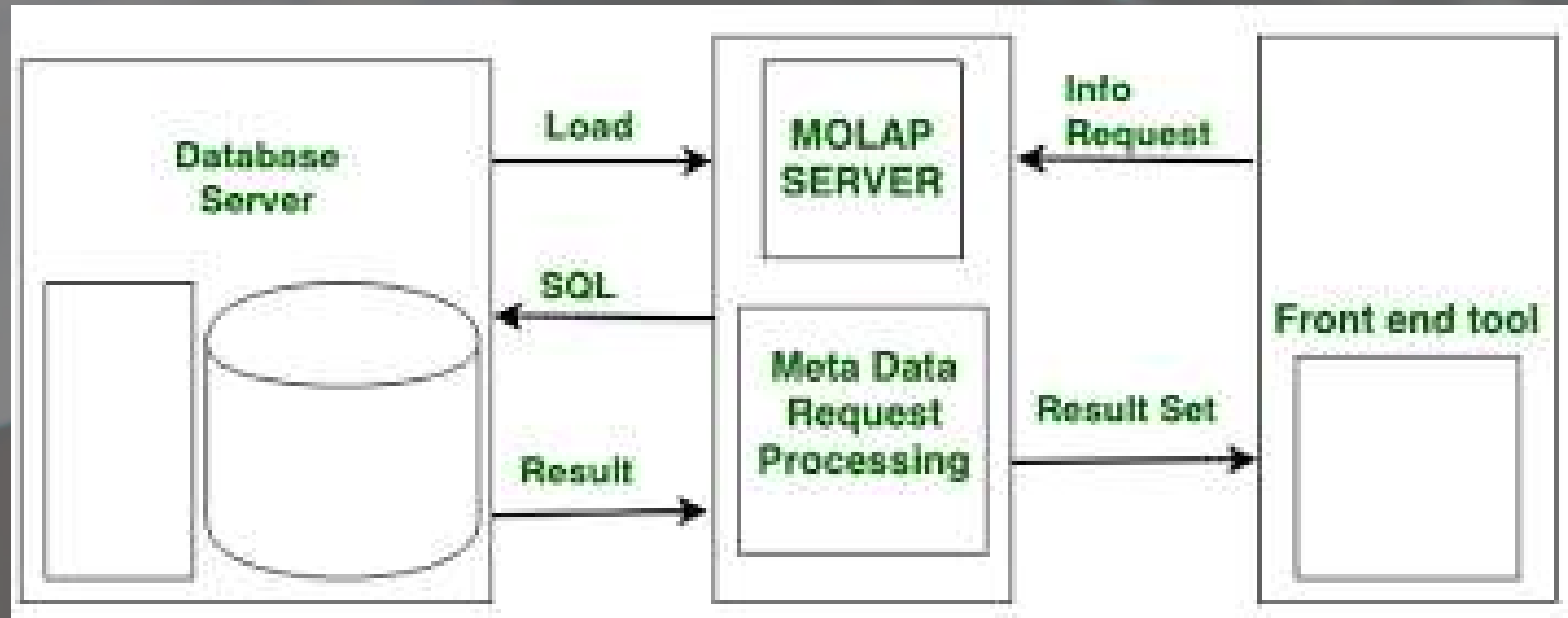
# Difference between rolap and MOLAP

ROLAP (Relational OLAP)	MOLAP (Multidimensional OLAP)
Stores data in relational databases	Stores data in multidimensional cubes
Works directly on relational tables	Uses pre-calculated data cubes
Slower query performance	Faster query performance
Can handle very large data volumes	Best for smaller, summarized data
Requires complex SQL queries	Uses specialized multidimensional queries
Storage efficient	Requires more storage space
Data is calculated at query time	Data is pre-processed
More scalable	Less scalable compared to ROLAP
Easier to update data	Updating cubes is slower

# ROLAP DIAGRAM



# MOLAP DIAGRAM

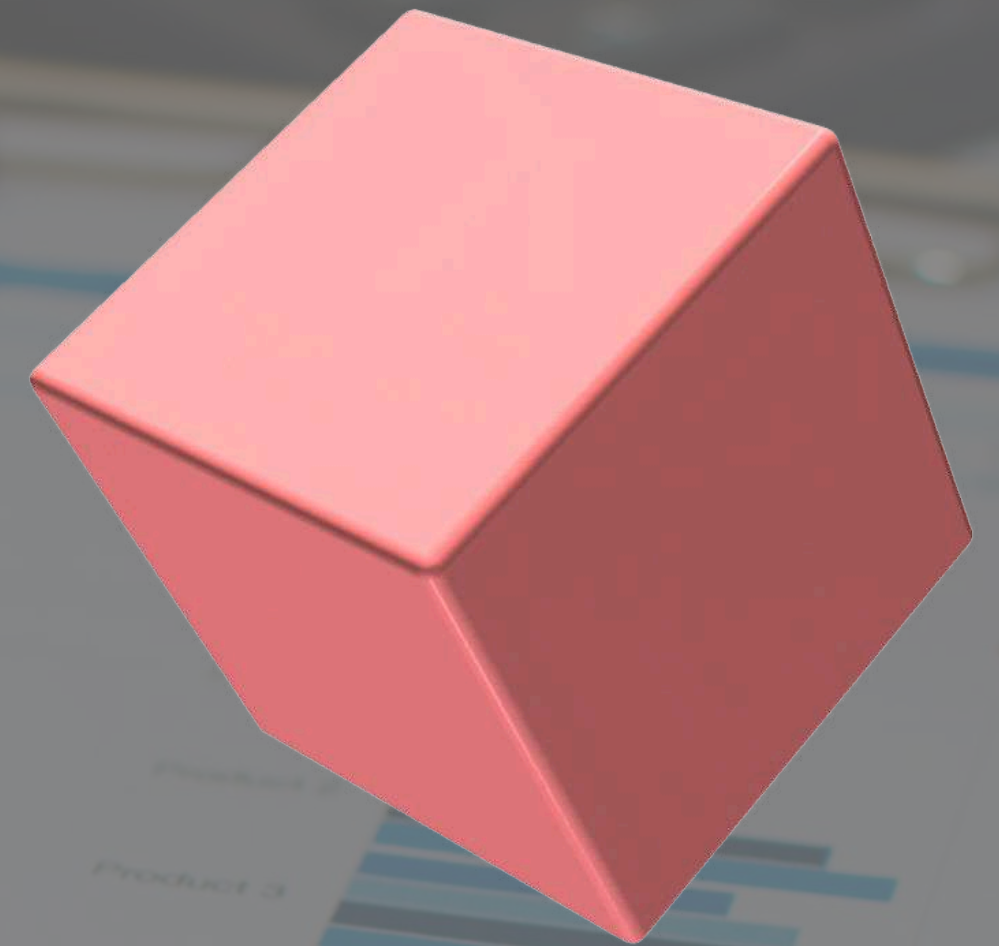


# Difference between olap and OLTP

OLAP (Online Analytical Processing)	OLTP (Online Transaction Processing)
Used for data analysis and reporting	Used for daily transactions
Handles historical data	Handles current, real-time data
Complex queries	Simple queries
Read-intensive system	Write-intensive system
Used by analysts and managers	Used by clerks and end users
Data stored in multidimensional format	Data stored in relational tables
Example: Sales trend analysis	Example: ATM transaction
Slower response time	Fast response time
Focus on decision making	Focus on transaction processing

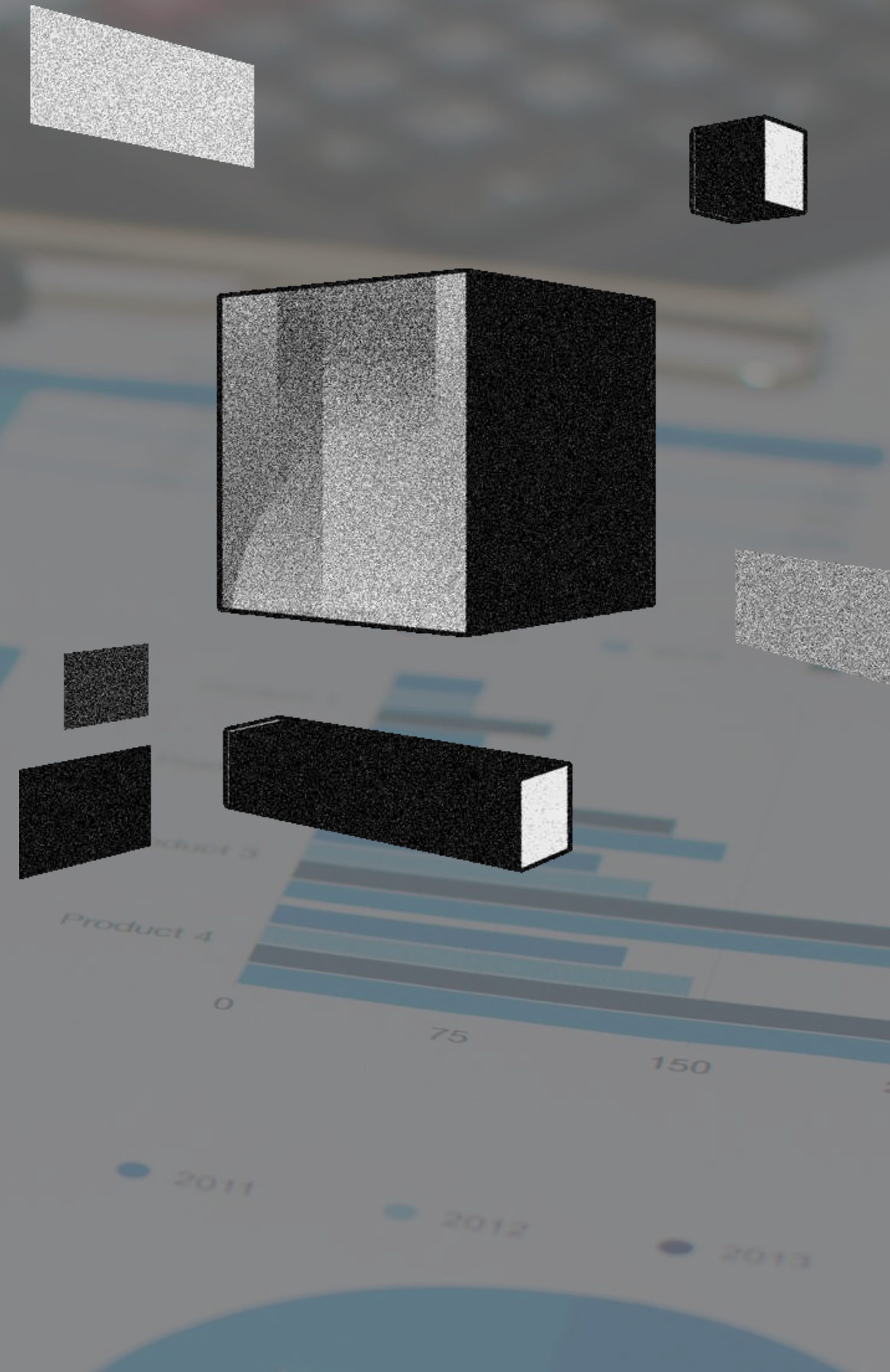
# Cube in Business Intelligence

- A cube in BI is a multidimensional data structure used for quick analysis.
- It arranges data in a way that users can examine information from different perspectives at the same time.
- Data is organized into dimensions (like time, product, region) and measures (like sales or profit).
- It allows fast querying and interactive data exploration.
- Helps simplify complex datasets and makes comparison easier.
- Improves understanding of trends and supports better decision-making.



# Multidimensional Data Cube & OLAP Operations

- A multidimensional data cube is a structured way to store data for quick analysis
- It arranges information into multiple dimensions so users can study data from different viewpoints
- Common dimensions include time, product, region, etc.
- Widely used in analytical systems to discover patterns and insights



# Multidimensional Data Cube & OLAP Operations

## Important OLAP Operations

### 1. Slice

- Selects one layer of data by fixing a single dimension
- Example: Viewing only January sales

### 2. Dice

- Selects a smaller portion of the cube using multiple dimensions
- Example: Electronics sales in January for the USA region



# Multidimensional Data Cube & OLAP Operations

## Drill-Down

- Moves from summary data to detailed data
- Example: Yearly sales → Monthly sales

## Pivot (Rotation)

- Changes the orientation of data for a different perspective
- Example: Switching view from sales by region to sales by product



# Star Schema

## Star Schema

- A database design where one central fact table connects directly to dimension tables
- Structure looks like a star shape
- Fact table stores numeric/measurable data (sales, profit, revenue) and foreign keys
- Dimension tables store descriptive details (time, product, region)

## Advantages

- Faster query performance (fewer joins)



# Star Schema

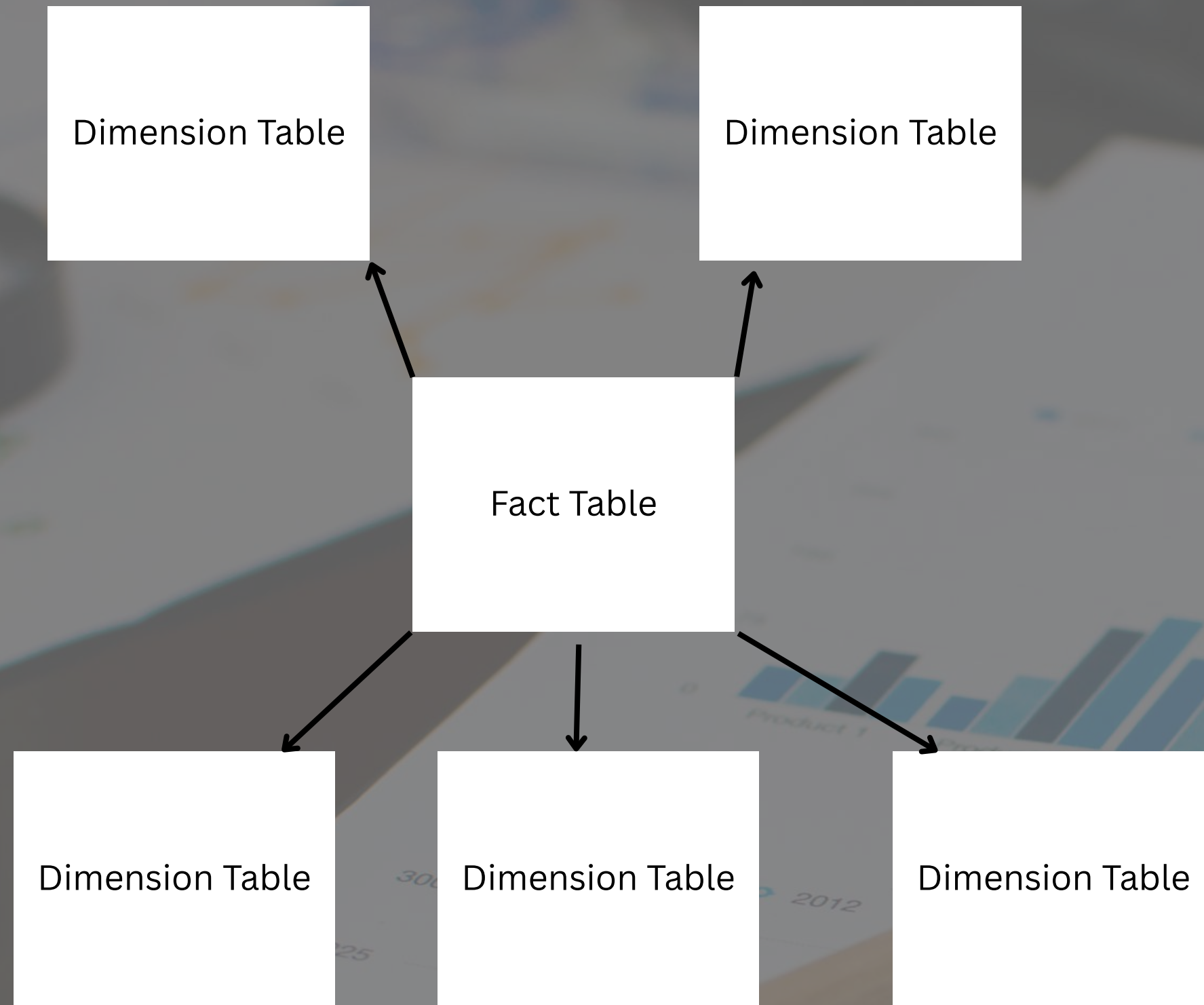
- Simple structure and easy to understand
- Best suited for OLAP and reporting

## Disadvantages

- Data redundancy (values repeated in dimension tables)
- Uses more storage due to denormalized data



# Star Schema - Diagram

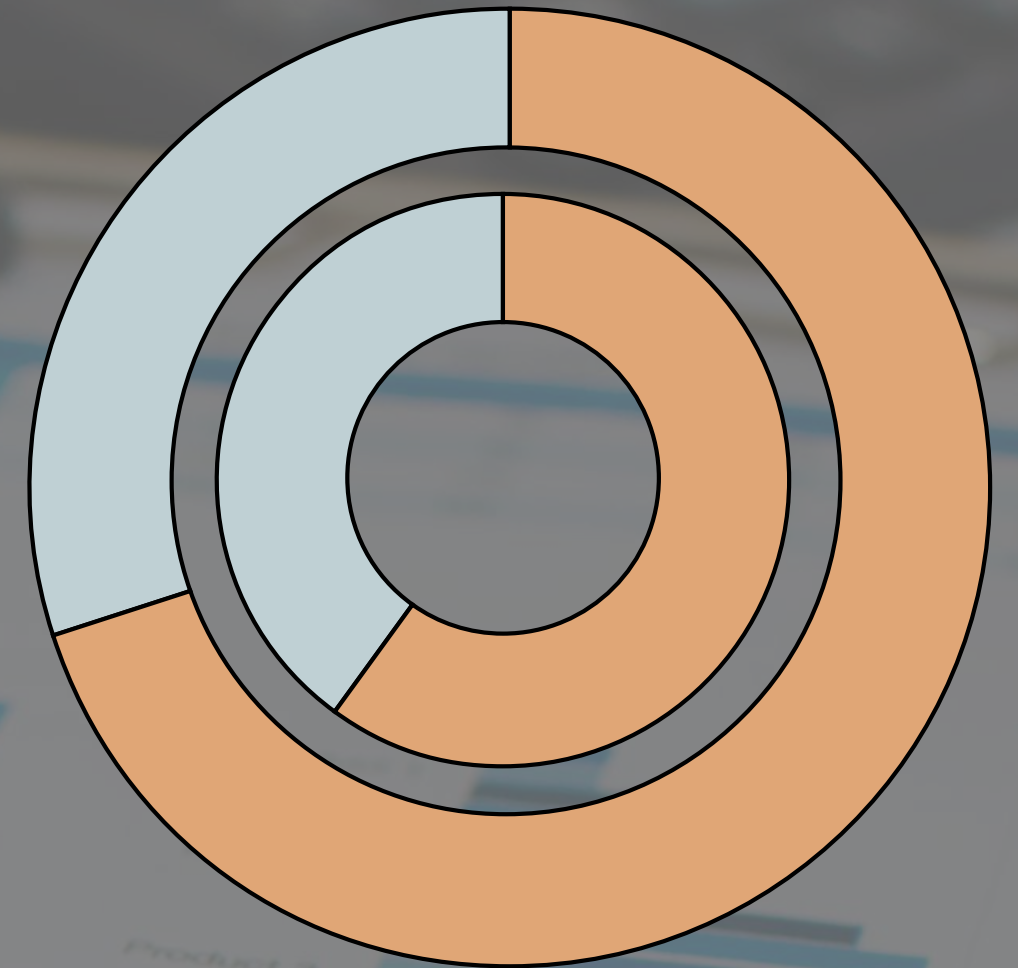


# Snowflake Schema

- An advanced version of Star Schema
- Dimension tables are normalized into multiple related tables
- Structure looks like a snowflake pattern
- Used in data warehouse design

## Components

- Fact Table → Stores measurable data (sales, quantity, profit)
- Normalized Dimension Tables → Data is split into smaller related tables instead of one large table



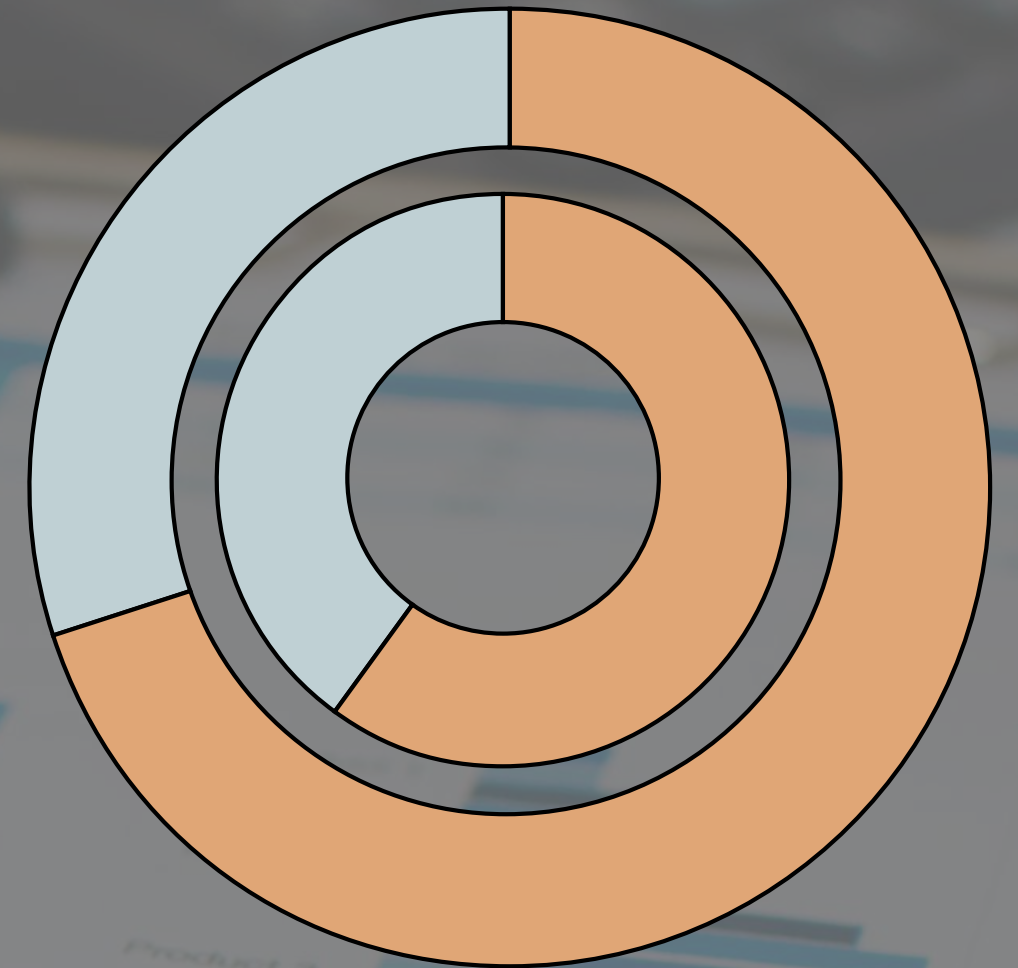
# Snowflake Schema

## Advantages

- Less data redundancy (no repeated data)
- Saves storage space
- Better data consistency and integrity
- Good for large data warehouses

## Star vs Snowflake — Which is Better?

- Choice depends on business needs and system design
- Star Schema is better when:
- Fast query performance is needed
- Simple design is preferred
- Used for BI reports and dashboards

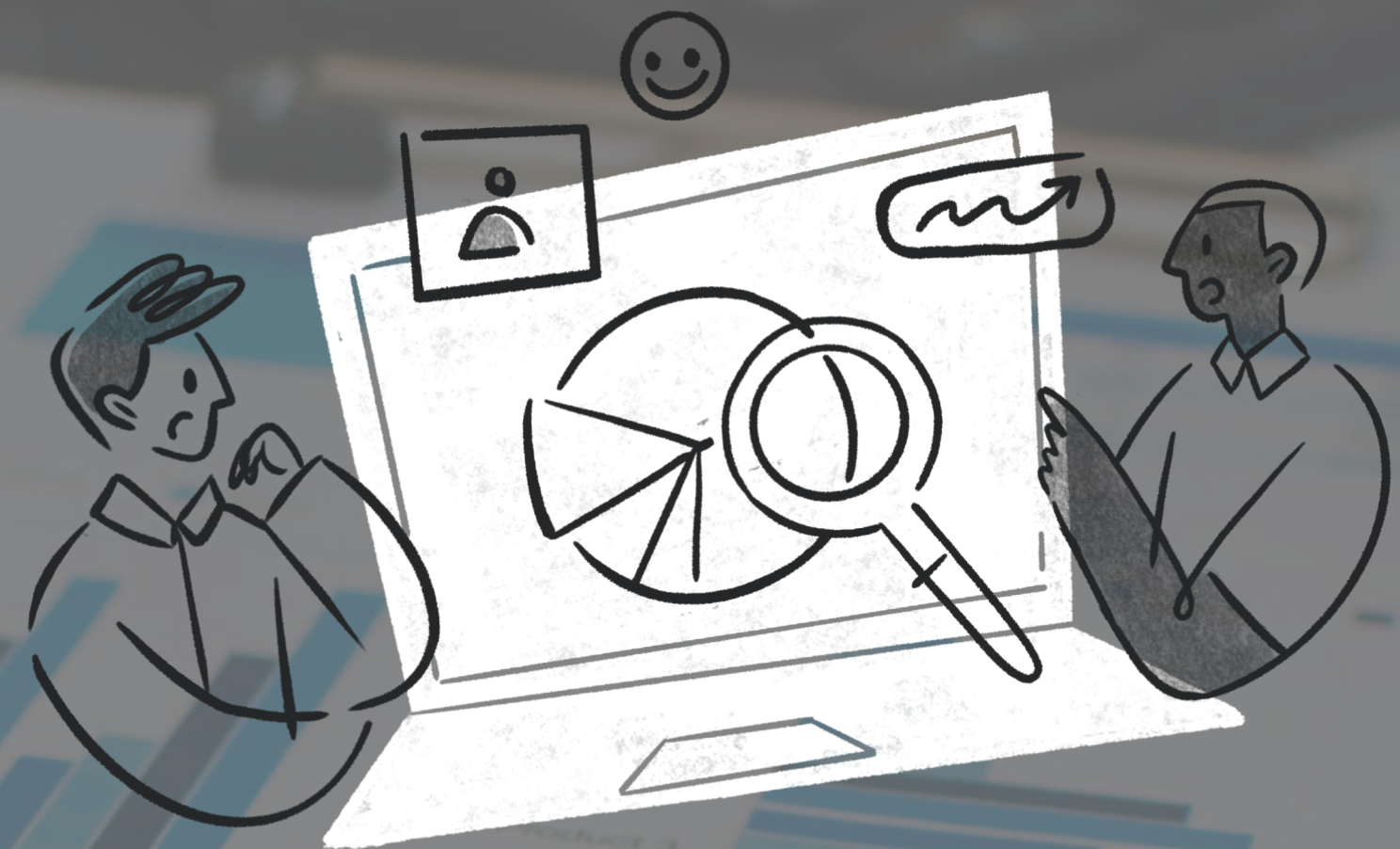


# Fact Constellation Schema

- A complex data warehouse design
- Contains multiple fact tables
- Fact tables share common dimension tables
- Also called Galaxy Schema

## Purpose

- Used when a system handles multiple business processes
- Allows analysis across different departments
- Provides more flexibility than Star or Snowflake schema



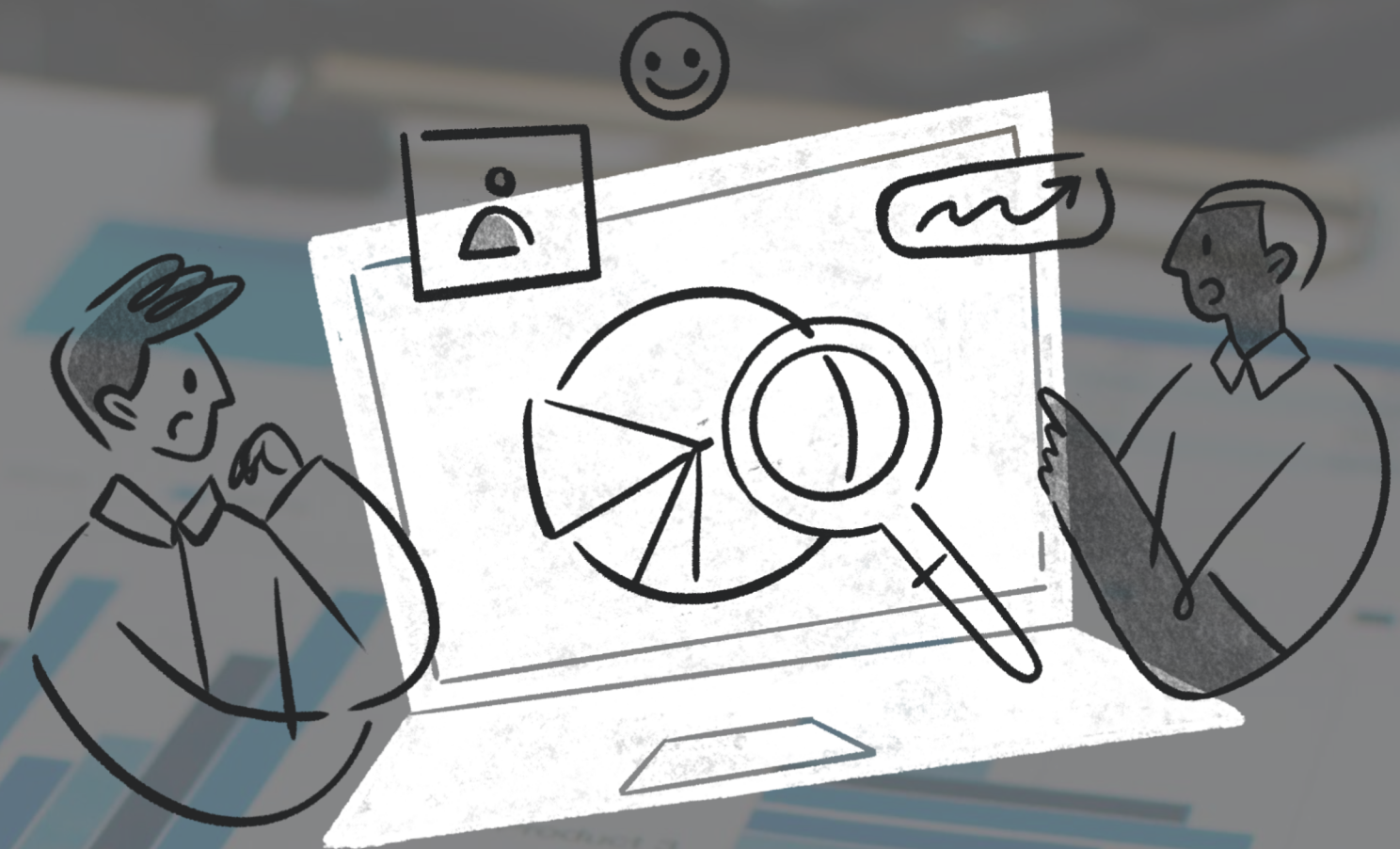
# Fact Constellation Schema

## Structure

- Each fact table represents a different process
- All fact tables are connected through shared dimensions
- Common dimensions may include Time, Product, Location

## Example Scenario

- A retail company wants to analyze both sales and shipping data together.



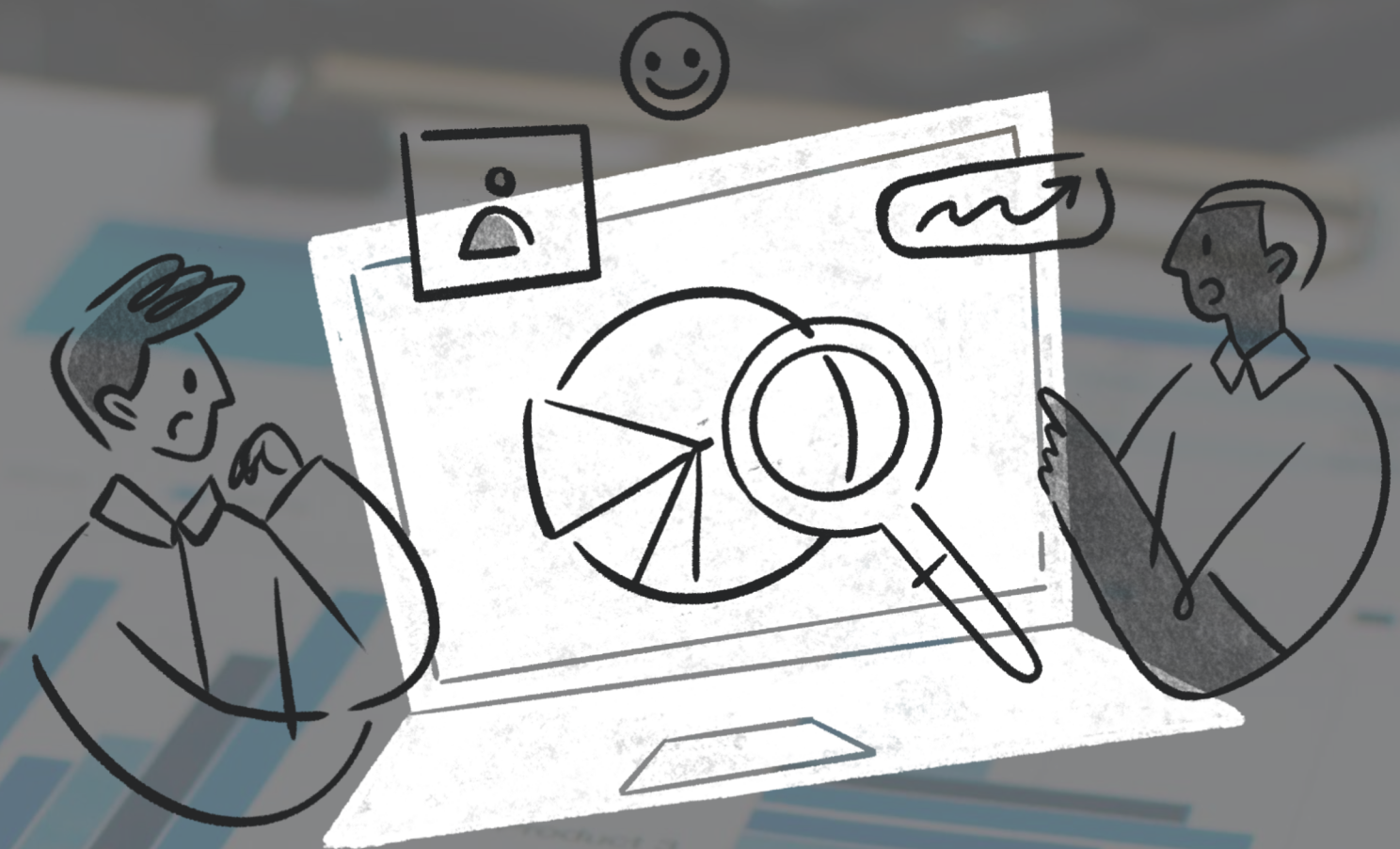
# Fact Constellation Schema

## Sales Fact Table

- Stores product sales information
- Example fields:
  - Sales ID
  - Product ID
  - Customer ID
  - Date
  - Sales Amount

## Shipping Fact Table

- Stores shipment details
- Example fields:



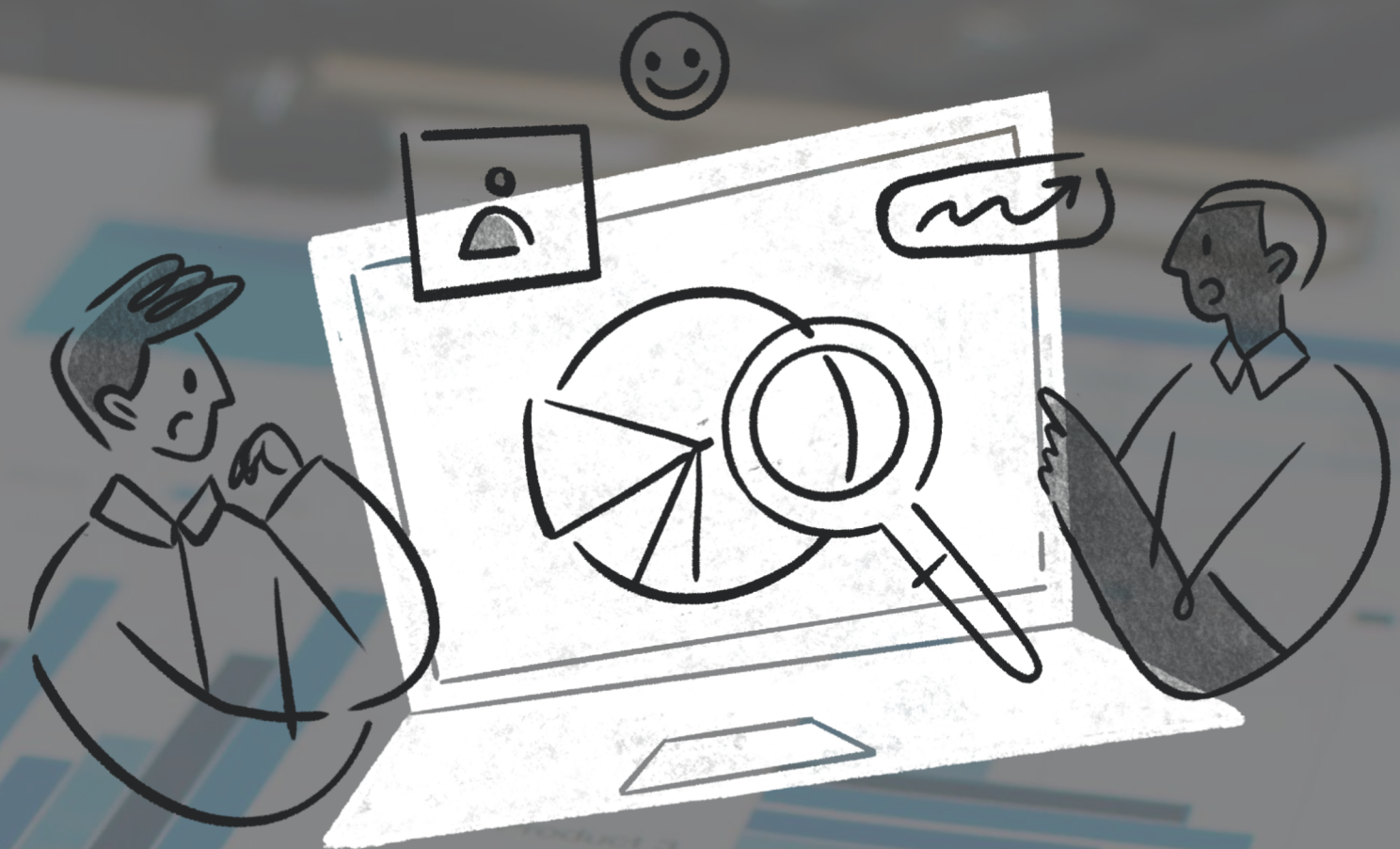
# Fact Constellation Schema

- Shipment ID
- Product ID
- Date
- Shipping Cost
- Delivery Time

## Shared Dimensions

- Product Dimension
- Time Dimension
- Customer Dimension

Shared dimensions allow combined analysis of multiple processes.



# BI Tools

## Business Intelligence (BI) Tools

- Software that helps organizations collect, process, analyze, and visualize data
- Used to support better business decisions
- Converts raw data into meaningful insights

## Popular BI Tools

- Microsoft Power BI
- Tableau
- Looker
- Domo
- Qlik Sense



# BI Tools

## Power BI

- One of the most widely used BI tools
- Connects to many data sources
- Creates interactive dashboards
- Generates real-time reports
- Integrates well with other Microsoft services
- Used for business analytics and visualization



## Tableau

- Powerful data visualization platform
- Creates interactive and shareable dashboards

# BI Tools

- Drag-and-drop interface (easy for beginners)
- Works with large datasets
- Connects with databases, cloud services, and spreadsheets
- Used for reporting, analysis, and forecasting

## Benefits of BI Tools

- Faster decision making
- Clear data visualization
- Easy trend identification
- Improved business performance





SHARE

subscribe



Thank you

