Data Mart Design for Multidimensional Analysis

20 May 2024

Table of Content

The Objectives

1. To learn about data mart & multidimensional analysis
2. To know about the type & benefit of multidimensional analysis
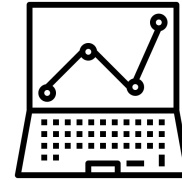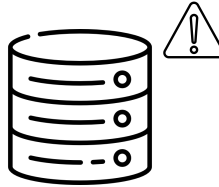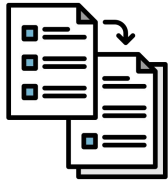3. Student can design & build data mart

Overview

- Banking sector has played a key role in the financial development of a country
- There are many types of historical data
  in multiple heterogeneous databases → **complex process**
- The increasing competition of market changes has demanded bank intelligence for analyzing those enormous data

Challenges

- Data redundancy may exist after the basic data being imported into database
- These mass data are unable to be effectively analyzed and deeply utilized

OLTP vs OLAP
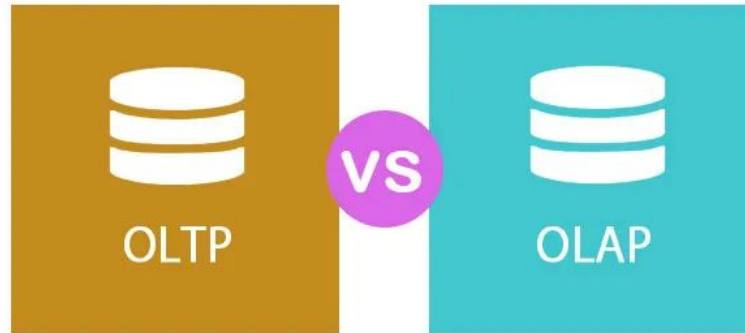
- OLTP & OLAP are **database management systems** for **storing** and **processing** data.
- They require efficient and reliable IT infrastructure to run smoothly.
- You can use them both to query existing data or store new data.
- Both support data-driven decision-making in an organization.

OLTP

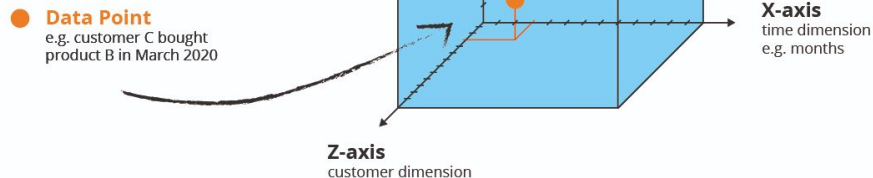**OLTP** or "Online Transactional Processing", refers to systems that facilitates and supports the execution of a large number of **real-time transactions** in a database.

Example: a bank transaction for example, if it fails the whole action must be reversed, if it is successful, it must be recorded and immutable.

OLAP

OLAP Cube
with 3 dimensions

Data Point
e.g. customer C bought
product B in March 2020

Y-axis
product dimension

X-axis
time dimension
e.g. months
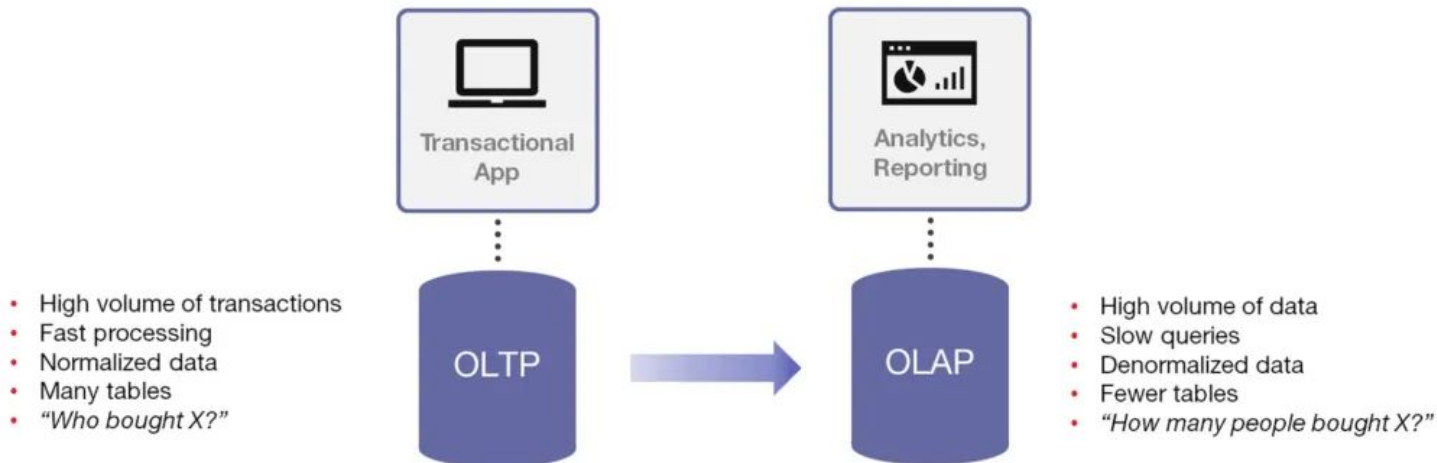
Z-axis
customer dimension

OLAP or "Online Analytical Processing" is **a tool used for the analysis** and treatment of data in real time, widely used to treat a massive amount of information in the various dimensions of a data warehouse.

Example: We are analyzing temporal data and we want to have views by year, day, quarter or semester, this interaction with the user is done by OLAP.

OLAP vs OLTP

OLTP vs OLAP

Transactional App

Analytics, Reporting

- High volume of transactions
- Fast processing
- Normalized data
- Many tables
- *"Who bought X?"*

OLTP → OLAP

- High volume of data
- Slow queries
- Denormalized data
- Fewer tables
- *"How many people bought X?"*
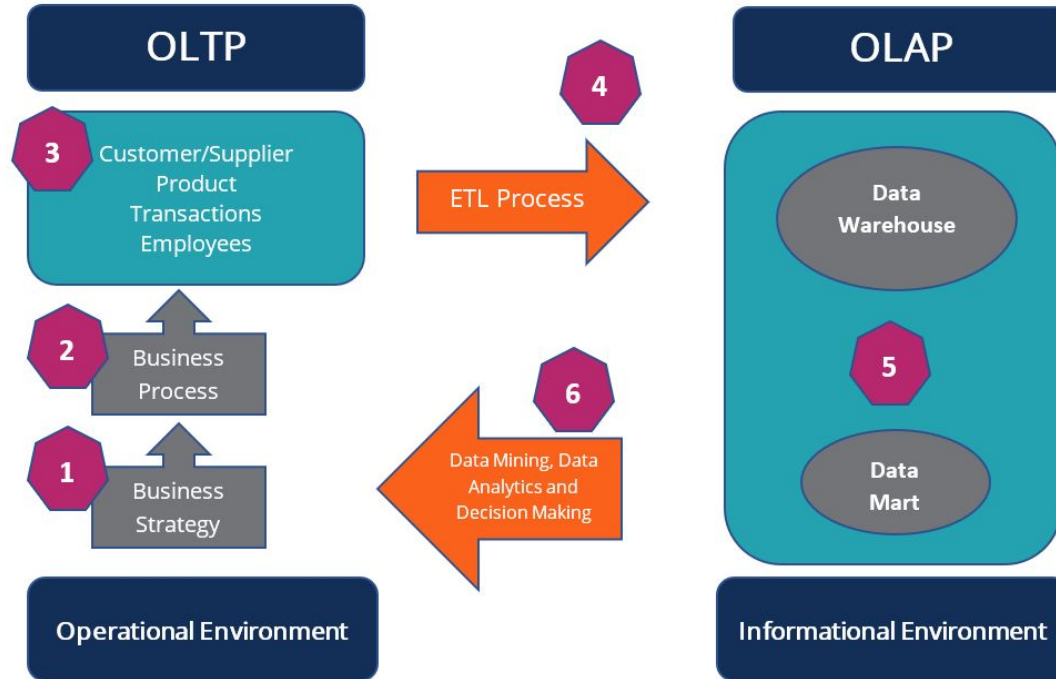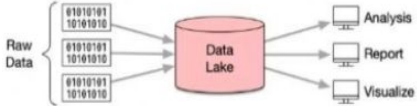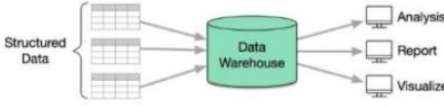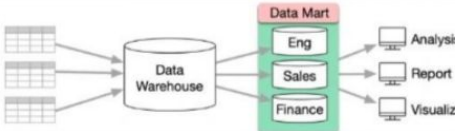
What is Data Mart?

Data Warehouse vs Data Lake vs Data Mart

Name	Explanation	Illustration
Data Lake	A vast pool of raw, unstructured data stored in its native format until it's needed for use	
Data Warehouse	A large, structured repository of integrated data from various sources, used for complex querying and historical analysis	
Data Mart	A more focused, department-specific subset of a data warehouse providing quick data retrieval and analysis	
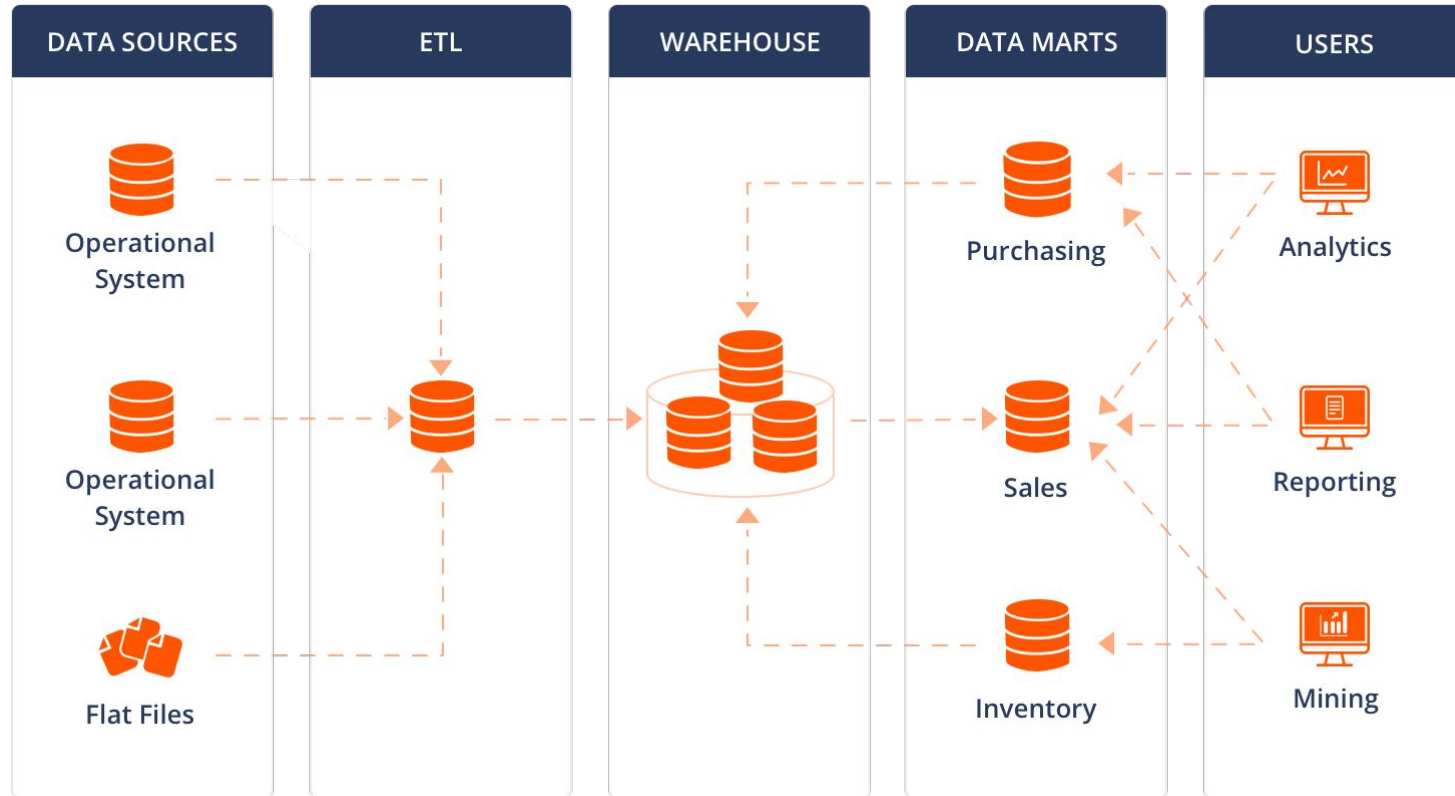
Multidimensional Data Analysis

- In this analysis we use **structured data in the form of a cube** (each side of the cube is a **dimension**)
- **Each table** is seen as a **dimension**, together they form a cube that can have **low or high granularity**, always depending on the requirements of each project.

High granularity   **Low granularity**

Category	
Furniture	742,000
Office Supplies	719,047
Technology	836,154

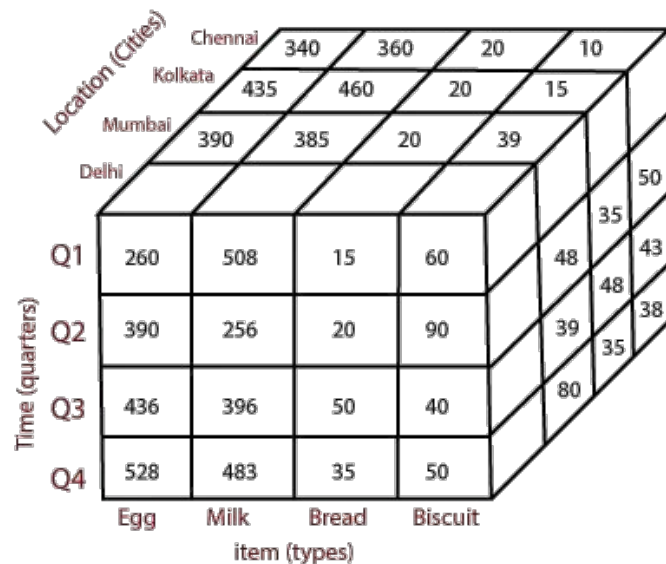| Category	Sub-Catego..	
Furniture	Bookcases	114,880
	Chairs	328,449
	Furnishings	91,705
	Tables	206,966

Category	Sub-Catego..	Manufacturer	
Furniture	Bookcases	Atlantic	18,617
		Bestar	1,898
		Bush	12,707
		DMI	11,047
		Global	3,448
		Hon	5,216
		O'Sullivan	27,308
		Other	1,564
		Riverside	15,611
		Rush	2,977
		SAFCO	1,661
		Sauder	12,826
	Chairs	Bevis	1,775
		DMI	4,620

When we have a **high granularity** of data it means that we have **less details of the data**,
when we have a **low granularity**, we have **more details of the data**.

Dimension vs Fact Tables

Fact Table	Dimension Table
Contains the primary keys of the referenced dimension tables along with some quantitative metrics	Holds the descriptive information for the related fields that are in the fact table's records
Made after dimension table	Created first
Analysis and decision making	Data and process storage
Ex: Customer orders or time-series financial data	It typically represents a physical entity like "customer" or "product."

Types of Dimension Tables
(Slowly Changing Dimension/ SCD)

Type 1: Update Changes

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	CA

→

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State
123	ABC	Acme Supply Co	IL

Type 2: Keep Historical

Supplier_Key	Supplier_Code	Supplier_Name	Supplier_State	Start_Date	End_Date
123	ABC	Acme Supply Co	CA	01-Jan-2000	21-Dec-2004
124	ABC	Acme Supply Co	IL	22-Dec-2004	

Type 3: Preserve Limited History

Supplier_Key	Supplier_Code	Supplier_Name	Original_Supplier_State	Effective_Date	Current_Supplier_State
123	ABC	Acme Supply Co	CA	22-Dec-2004	IL

Dimension vs Fact Tables

- Fact tables are **objects to be analyzed**, composed of measures, contexts of each dimension and Foreign Keys, used to link the dimensions to that table.

- Example: In our data warehouse we need to create a sales fact table, for this, we have structured it as follows.

- The dimensions that will compose our suit are:
    - Time dimension
    - Geographic dimension (location)
    - Product dimension
    - Customer dimension, etc.

- The layer will have the Foreign Key for each dimension plus the metrics, such as "total_sales".
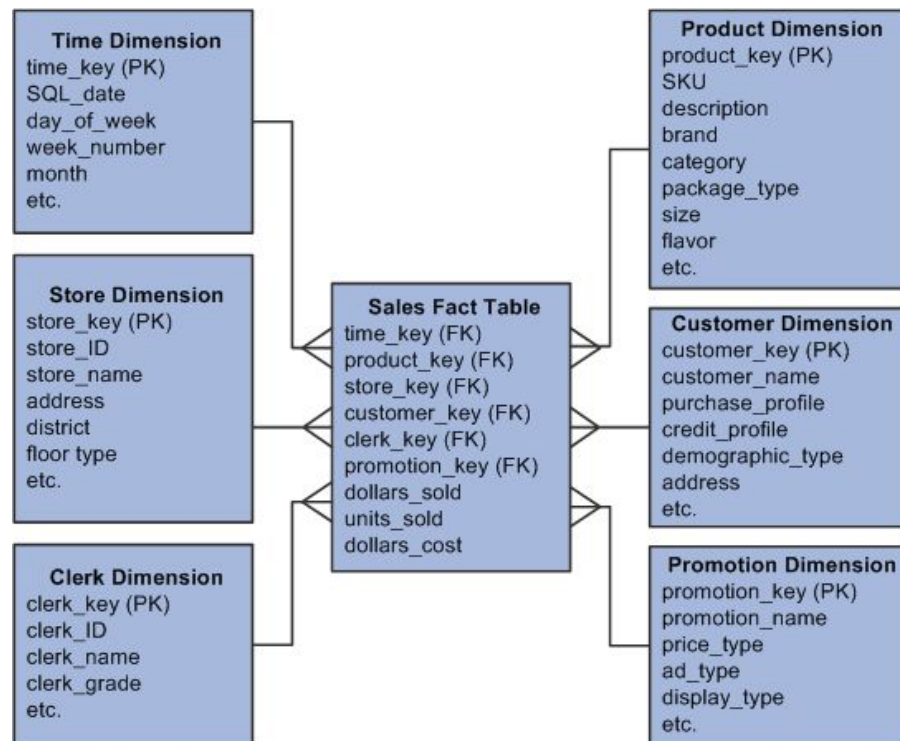
Benefit

Observe and process data from **several angle**

Achieve fast query **performance** against business data

Integration with commonly used BI reporting tools

Data Fellowship

Types of Multidimensional Model

Star Schema

Time Dimension

Order ID
Order Date
Year
Quarter
Month

Product Dimension

Product ID
Product Name
Product Category
Unit Price

SALES

Product ID
Order ID
Customer ID
Employer ID
Total
Quantity
Discount

Customer Dimension

Customer ID
Customer Name
Address
City
Zip

Emp Dimension

Emp ID
Emp Name
Title
Department
Region

Snowflake Schema

Time Dimension
OrderID
Order Date
Year
Quarter
Month

Sales
ProductID
OrderID
CustomerID
EmployeeID
Total
Quantity
Discount

Product Dimension
ProductID
ProductName
ProductCategoryID

ProductCategory Dimension
ProductCategoryId
Name
Description
Uniprice

Customer Dimension
CustomerID
CustomerName
Address
CityID

Employee Dimension
EmployeeID
EmployeeName
DepartmentID
Region
territory

City Dimension
CityID
CityName
Zipcode
State
Country

Department Dimension
DepartmentID
Name
Location

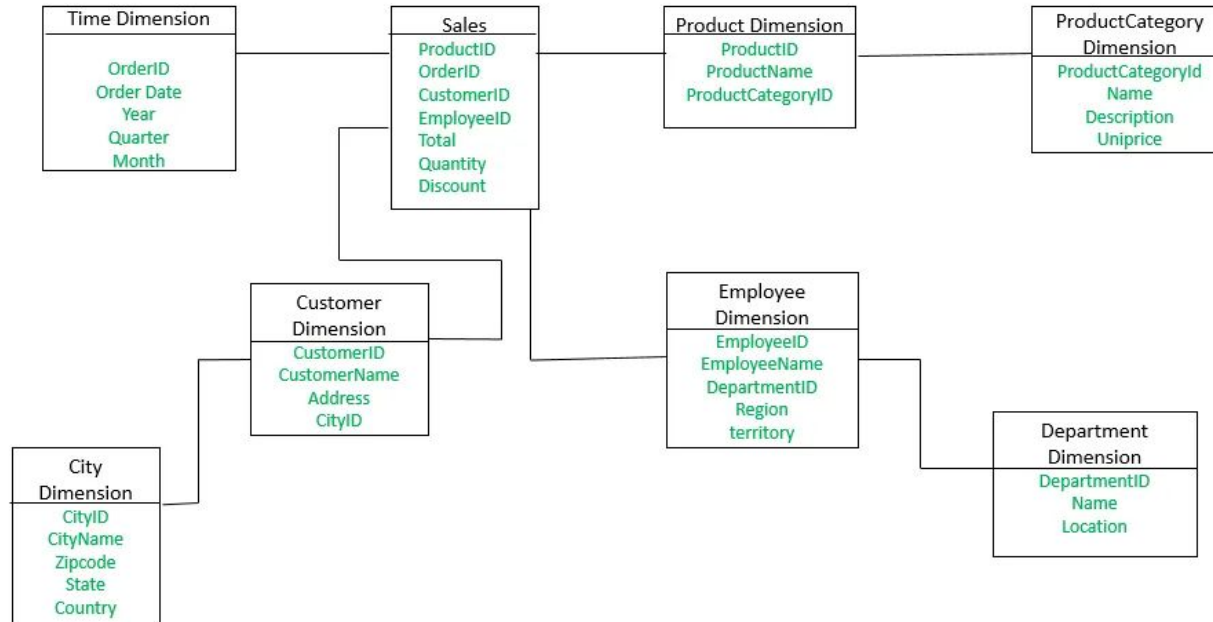**Data Fellowship**

How to Create
a Data Mart

1) Identification of Business Needs

- Understand **the business needs** and **requirements** of the target users utilizing the data mart
  - Outlining the scope of the project
  - Highlighting all risks and limitations
  - This step help choose the type of data mart to use
- **Engage with stakeholders and subject matter expert**s to get detailed information about the specific data elements and metrics needed for analysis
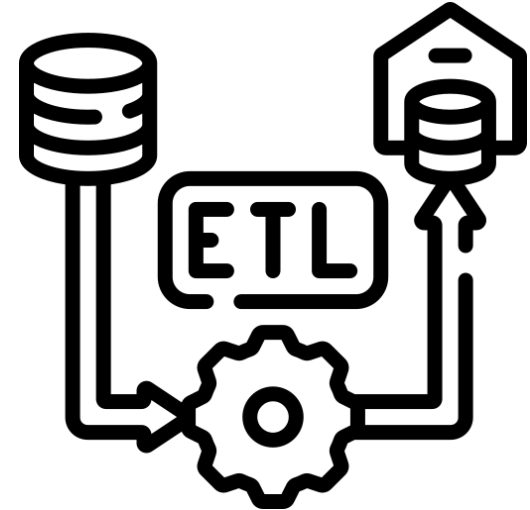
2) Design the Data Mart

- **Define the data model** and **schema** for the data mart
  - identifying the fact tables (containing quantitative data)
  - Identifying dimension tables (containing descriptive attributes) required for analysis.
- **Choose an appropriate schema design** depending on the analytical needs and data management requirements.
- **Establish relationships** between the fact and dimension tables, defining primary key-foreign key relationships for data integrity.

3) ETL

- Develop the ETL process to **extract source data and transform** it to match the data model of the data mart.Transformation involves:
    - Data cleansing,
    - Validation,
    - Aggregation,
    - Other manipulations to ensure data quality and consistency.
- **Load the processed data into the data mart**, populating the fact and dimension tables.

4) Implementation & Testing

- **Create the data mart** structure in the chosen DBMS
  - Ensuring it can efficiently handle analytical queries
  - Once the data is loaded into the mart, conduct testing to verify data accuracy and performance.
- Perform **user acceptance testing** (UAT) with the target users to validate that the data mart meets their requirements.
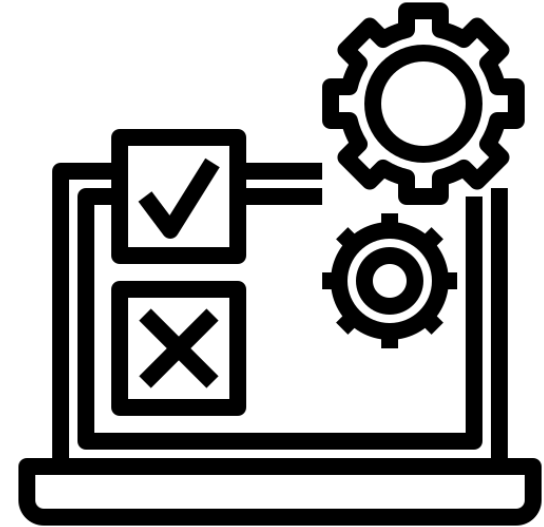
5) Deployment & Maintenance

- **Deploy the data mart to the production environment**, making it accessible to the intended users.
- Once implemented, **monitor performance** and **usage** to ensure it meets the evolving needs of the business users.
- It's also important to **regularly update and maintain the data mart** to accommodate changes in data sources, business requirements, or technology advancements.
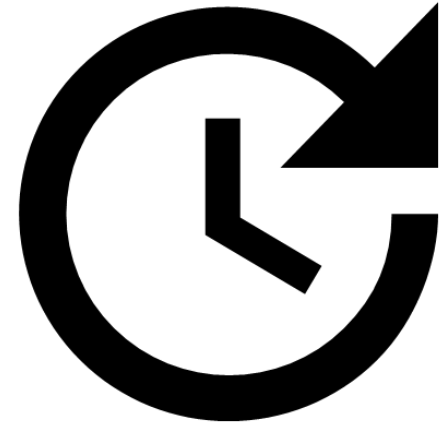
Hands On: Data Mart Creation

Session Summary

- Data marts play a crucial role in enabling organizations to **extract valuable insights** from their vast and diverse datasets
- They enable organizations to **gain more accurate and in-depth insights**
- Lead to **better internal business processes, improved customer experiences, and drive innovation**.
- When implemented effectively, data marts help businesses harness the power of their data, **gain a competitive edge, and make data-driven decisions that drive success and growth**.

References

IF ANY
Font Helvetica or **Lato**
The font size can be adjusted

Thank you!

When we stop learning, we stop growing #neverstoplearning