



Introduction to Data and Database



Additional Rules? Nope Just to make the class better!

1. There is no stupid questions!
2. Make sure to mute your mic so it won't disturb others
3. **You must enjoy the class!** If not, immediately raise your issue by click raise hand or comments in the chat



One thing we should know before we start ...



“Without big data, you are blind and deaf and in the middle of a freeway”

- Geoffrey Moore

Table of Content

What will We Learn Today?

1. What is Data?
2. What is Database?
3. What is SQL and DBMS?
4. What is data types?
5. Data "Lifecycle"
6. Google Collabs environment
7. DBeaver Installment





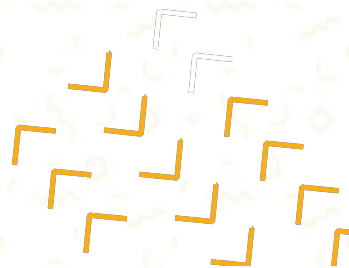


DATA

noun /'deɪ.tə, dæ.tə/

Information, especially **facts** or **numbers**, collected to be **examined** and considered and used to help **decision-making**, or information in an electronic form that can be stored and used by a computer.

Format: **text, spreadsheet, image, audio, video, etc.**





Categorical Data

- Nominal (classification): **gender, cities, name**
- Ordinal (weighted classification): **education, ranking**

Numerical Data

- Discret (count): **total kids**
- Continuous (measurement)
 - Interval: **temperature**
 - Ratio: **height, weight**



Knowledge

Given **meaning** change information into knowledge. It's already contextual, synthesized, and capable to give learning.

Raw Data

Can be retrieved from anywhere.
We know nothing yet!



Wisdom

1 Highest level is when we add **insight** to knowledge that enables understanding, integration, and actionable.

Information

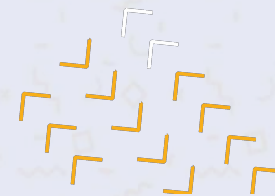
3 Once given some **context**, it becomes more useful, organised, and structured.





Pick the right definition!

- **It's 1 PM**
- **1 PM is lunch time**
- **1**
- **His family's favorite for lunch is Indonesian food**
- **Data**
- **Information**
- **Knowledge**
- **Wisdom**





DATABASE





DATABASE

an **organized collection** of data **stored** and **accessed** electronically.

For example, a company database may include tables for products, employees, and financial records.

Advantages:

1. **Reduced data redundancy**
2. **Greater data integrity and ensures consistency of data**
3. **Improved data access to users through use of host and query languages**
4. **Provides greater security and privacy of data**
5. **Shows you the big picture**
6. **Robust backup and recovery**



Some Tech Definitions

- **DBMS (Database Management System)**
 - software designed to store, retrieve, define, and manage data in a database
- **SQL (Structured Query Language)**
 - a domain-specific language used in programming and designed for managing data held in a relational database management system
- **DDL (Data Definition Language)**
 - SQL command to define the schema
 - CREATE, DROP, ALTER, TRUNCATE, COMMENT, RENAME
- **DML (Data Manipulation Language)**
 - SQL command to query informations, fill data, change data, update data, or delete data from database
 - SELECT, INSERT, UPDATE, DELETE



DATABASE



Structured



Unstructured



DATABASE (Structured)

- **PRO**
 - Easier implementation for Machine Learning
 - Easily used by business users/end users
 - Compatible with lot of tools
- **CONS**
 - Limited to relational database only
- **TOOLS**
 - MySQL, MS SQL, PostgreSQL, SQLite, etc



DATABASE (Unstructured)

- **PRO**
 - Keep original format as it is
 - Faster processing time
 - Save more storage while saved in data lake
- **CONS**
 - Need extra technical skills
 - Need specific tools
- **TOOLS**
 - MongoDB, DynamoDB, Hadoop, etc



DATABASE - Data Types

Numeric

int
bigint
smallint
float
decimal
real
bit

Date/time

date
time
datetime
timestamp
year

Character/String

char
varchar
text

Unicode Character

nchar
nvarchar
ntext

Binary

binary
varbinary

Miscellaneous

clob
blob
xml
json



DATABASE - Data Models

Relationship Cardinality

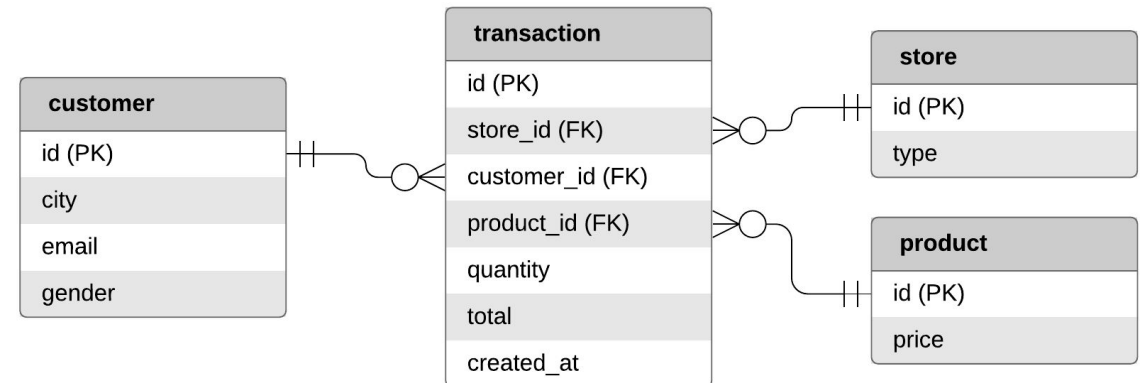
- one to one (1:1)
- one to many (1:M)
- many to many (M:N)

Participant Constraint

- Mandatory
 - At least there's one entity that associated with entity A
- Optional
 - Allowed to have none entity that associated with entity A

Entity types

- Strong: Can stand alone
- Weak: Need others
- Associative: Created by other entities





DATA LIFECYCLE





DATA LIFECYCLE

also called the **information life cycle**, refers to the entire period of time that data exists in your system from first capture onwards.

PS: 8 would go to 1 again

1

Generation

2

Collection

3

Processing

4

Storage

5

Management

6

Analysis

7

Visualization

8

Interpretation



GOOGLE COLABS ENVIRONMENT



DBEAVER INSTALLMENT



**Thank
YOU**

