


# Набір даних та їх атрибутів





*Дані у широкому розумінні:* факти, текст, графіки, картинки, звуки, аналогові та цифрові відео-сегменти.

*Джерело даних:* вимірювання, експерименти, арифметичні та логічні операції.

*Дані* – це необроблений матеріал, який надається постачальниками даних та використовується споживачами для формування інформації на основі даних.

## Таблиця “об’єкт - атрибути”

	Атрибути			
Об’єкти	Код клієнта	Вік	Родинний стан	Доход
	1	23	Не одружений	4500
	2	31	Одружений	5000
	3	19	Не одружений	1500
	4	40	Розлучений	6000
	5	55	Одружений	4200


У таблиці горизонтально розташовуються *атрибути об'єкта* або його ознаки. Вертикально у таблиці розташовані *об'єкти*.

*Об'єкт* описується як набір атрибутів.

*Атрибут* – властивість, що характеризує об'єкт.

*Змінна* – властивість або характеристика, спільна для всіх об'єктів, що вивчаються, поява якої може змінюватися від об'єкта до об'єкта.

*Значення* змінної є проявом ознаки.



**Генеральна сукупність** – вся сукупність об'єктів, що вивчаються, яка цікавить дослідника.

**Вибірка** – частина генеральної сукупності, певним способом відібрана з метою дослідження та отримання висновків про властивості та характеристики генеральної сукупності.

**Параметри** – числові характеристики генеральної сукупності.

**Статистики** – числові характеристики вибірки.



**Вимірювання** – процес присвоєння чисел характеристикам об'єктів, що вивчаються, відповідно з певним правилом.

**Шкала** – правило, відповідно до якого об'єктам присвоюються числа.





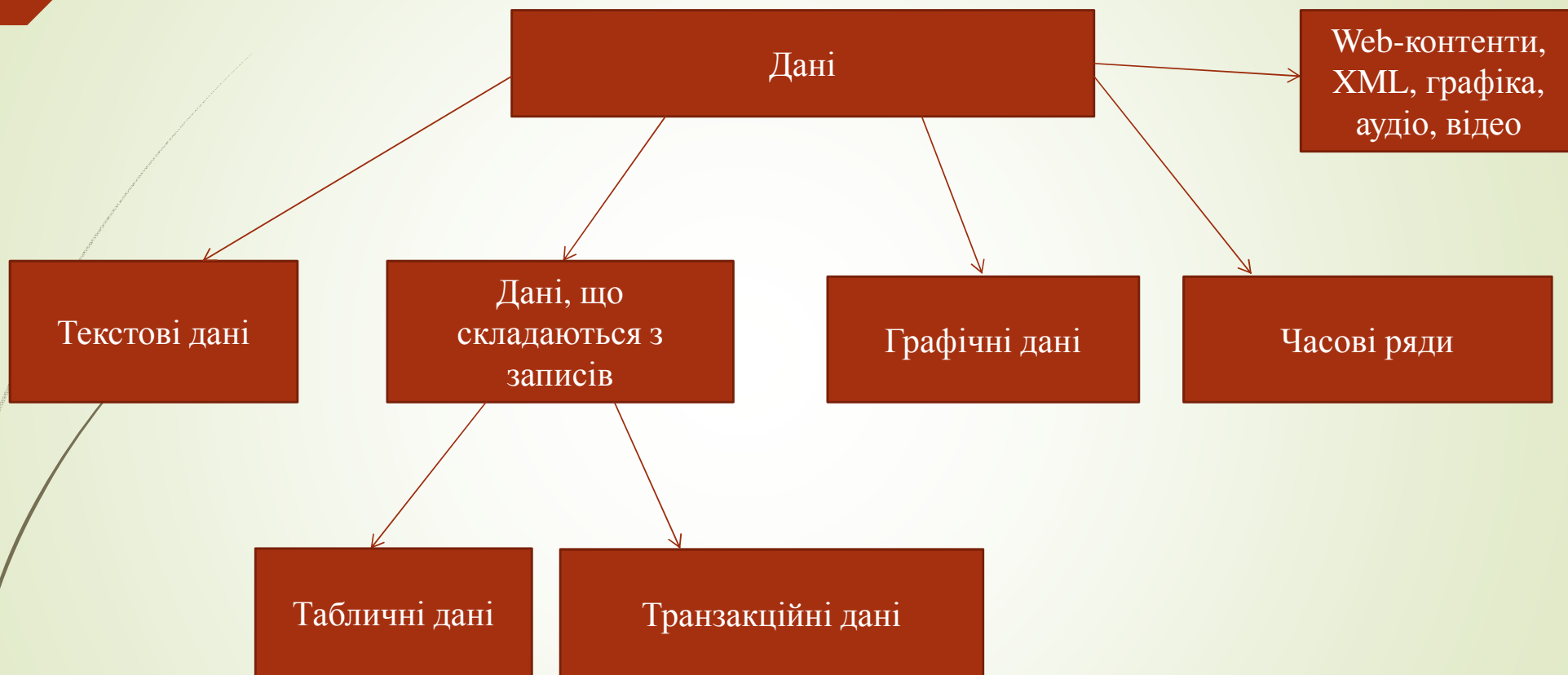
## Шкали:

- номінальна шкала (nominal scale)
- шкала порядку (ordinal scale)
- інтервальна шкала (interval scale)
- відносна шкала (ratio scale)
- дихотомічна шкала (dichotomous scale)



<b>Provides:</b>	<b>Nominal</b>	<b>Ordinal</b>	<b>Interval</b>	<b>Ratio</b>
The "order" of values is known		✓	✓	✓
"Counts," aka "Frequency of Distribution"	✓	✓	✓	✓
Mode	✓	✓	✓	✓
Median		✓	✓	✓
Mean			✓	✓
Can quantify the difference between each value			✓	✓
Can add or subtract values			✓	✓
Can multiple and divide values				✓
Has "true zero"				✓

# Типи наборів даних:



*База даних (Database)* – особливим образом організовані та збережені у електронному вигляді дані.

*Схема даних* – опис логічної структури даних, специфікований мовою опису даних та такий, що обробляється СУБД.

*Схема користувача* – зафіксований для конкретного користувача один варіант порядку полів таблиці.

*СУБД (Database Management System, DBMS)* – оболонка, за допомогою якої при організації структури таблиць та заповнення їх даними утворюється база даних.



## Вимоги до СУБД:

- висока швидкодія;
- простота оновлення даних;
- незалежність даних;
- можливість багатокористувальницького використання даних;
- безпека даних;
- стандартизація побудови та експлуатації бази даних (СУБД);
- адекватність відображення даних відповідної предметної галузі;
- дружлюбний інтерфейс користувача.



## Класифікація видів даних:

- реляційні, багатомірні (куби OLAP)
- змінні, постійні, умовно-постійні
- довідникові, оперативні, архивні
- дані за період, точкові дані
- первинні, вторинні

*Метадані (Metadata) – дані про дані (бізнес-метадані, оперативні метадані).*