

БАЗИ ДАНИХ: Лекція №1

НУ “Львівська Політехніка”, кафедра ПЗ

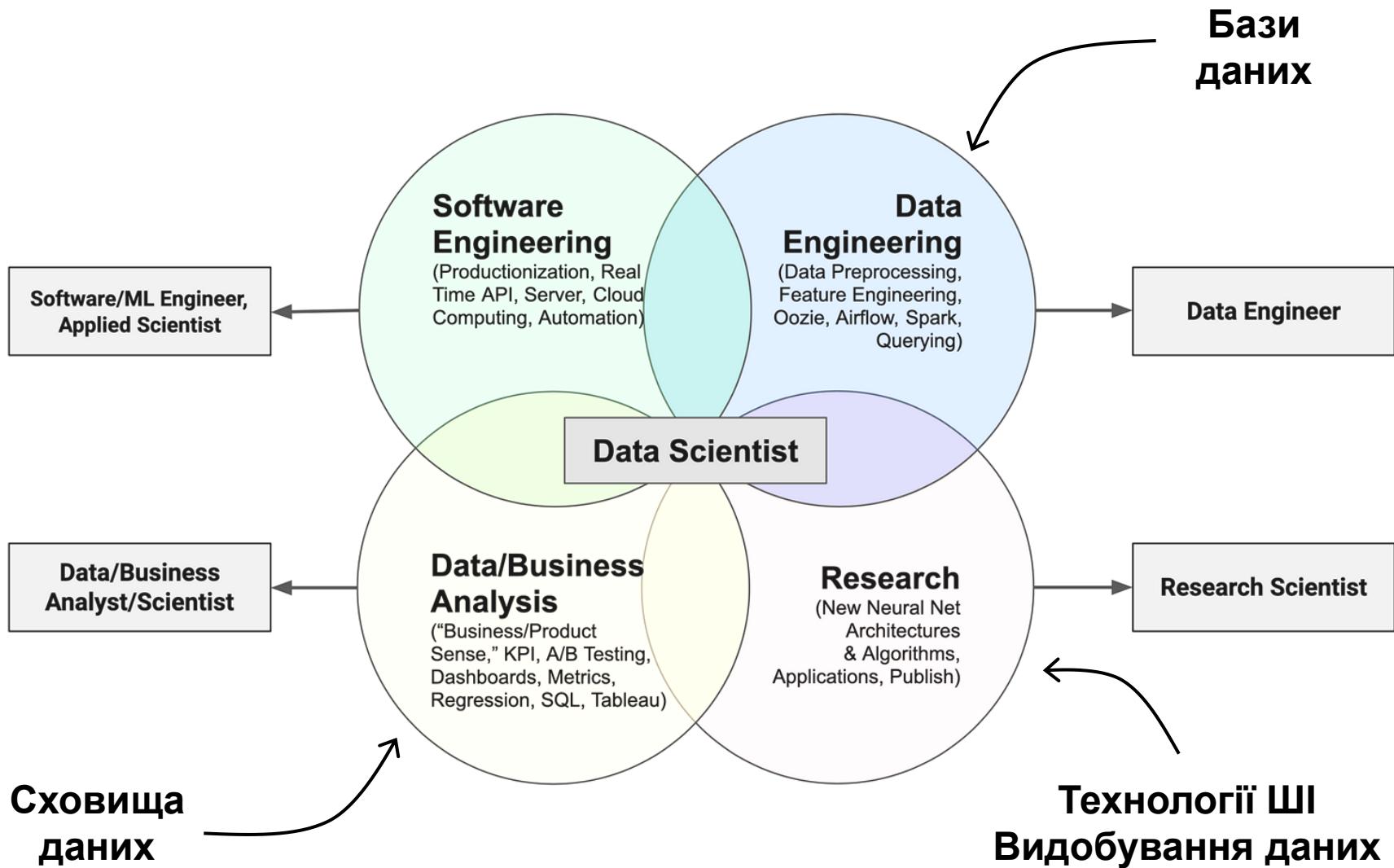
**Системи баз даних.
Поняття, архітектура, моделі даних**

Практичний досвід



- Харківський університет радіоелектроніки
(2001-2013 рр.)
- Національний університет
Львівська політехніка
(з 2023 р.)
- Системи та засоби
штучного інтелекту
- Реляційні бази даних для
накопичення новин
- Софт для управління
корпоративними даними

Інженерія даних



Структура курсу

- 15 лекцій
- 9 лабораторних робіт
 - звіт, результати, висновки, запитання
- 90 годин самостійної роботи
 - практика, виконання завдань
- Письмовий екзамен
- Усна компонента

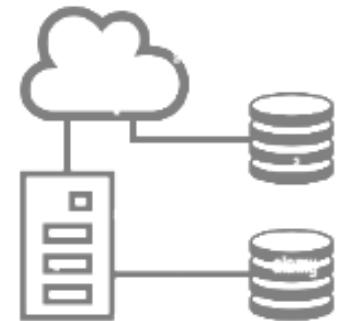
5 семестр

-
- 90 годин курсова робота
 - розробка бази даних
 - інтерфейс користувача
 - завантаження даних із джерел
 - пояснювальна записка

6 семестр

Зміст курсу

- **2 лк, 2 лаб**
 - концепція СУБД, архітектура, типологія, надійність, реляційні моделі
- **4 лк, 3 лаб**
 - мова SQL, цілісність, внесення та вибірка даних, групування даних
- **4 лк, 2 лаб**
 - спеціальні засоби СУБД: транзакції, індекси, тригери, функції, мова PL/SQL
- **2 лк, 1 лаб**
 - підходи до проектування БД, ORM, використання БД у додатках
- **3 лк, 1 лаб**
 - нереляційні БД, структуровані документи, графи, семантичні моделі



Семестрова оцінка

$$|B(T_2, a, b)| \leq 2$$

$$\varphi(5_1 t) \varphi(5_2 t) = \varphi(\sqrt{5_1^2 + 5_2^2} t)$$

$$\sum_{k=1}^{\infty} \int_{b_k v}^{x+b_k v} \left(\int_0^t \Psi_k^*(\tau) d\tau \right) dt - x \int_0^{b_k v} \Psi_k^*(\tau) d\tau - \frac{x^2}{2} B(v) + \int_0^x (x-u) \sum_{k=1}^{\infty} \Psi_k^*(u) du \quad A(v) = \sum_{k=1}^{\infty} b_k \Psi_k^*(b_k v)$$

Поточний контроль		Екзаменаційний контроль		
Лабораторні заняття	Разом	Письмова компонента	Усна компонента	Разом
9 лаб. \times 5	45	5 тест. \times 2 3 теор. \times 10 1 практ. \times 10	5	55

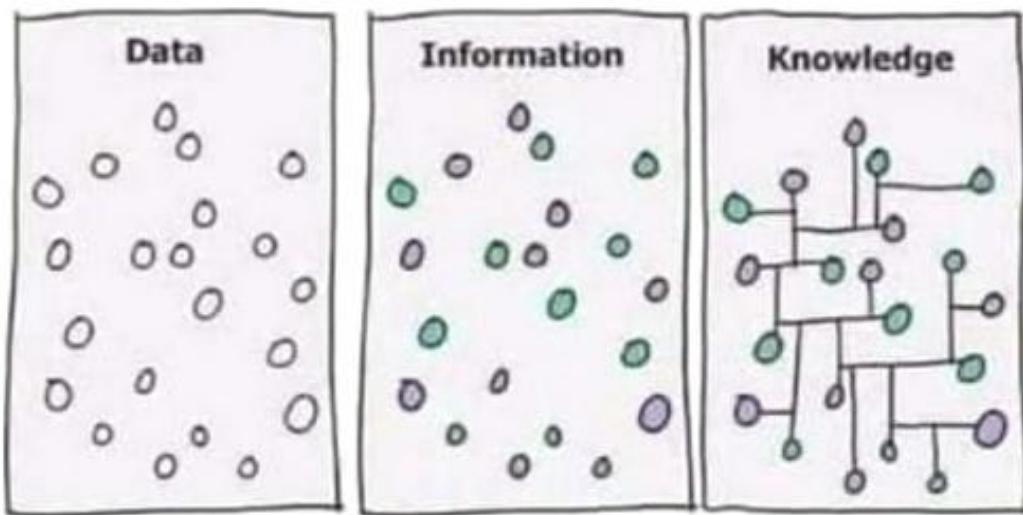
Умова	Формула	Пояснення
ЕК \geq ПК \times 0,5	ПО = ЕК + УК + ПК	Підсумкова оцінка є сумою всіх отриманих балів
ЕК $<$ ПК \times 0,5	ПО = ЕК + УК + ПК \times 0,5	Підсумкова оцінка враховує скореговане значення балів поточного контролю

Строки захисту

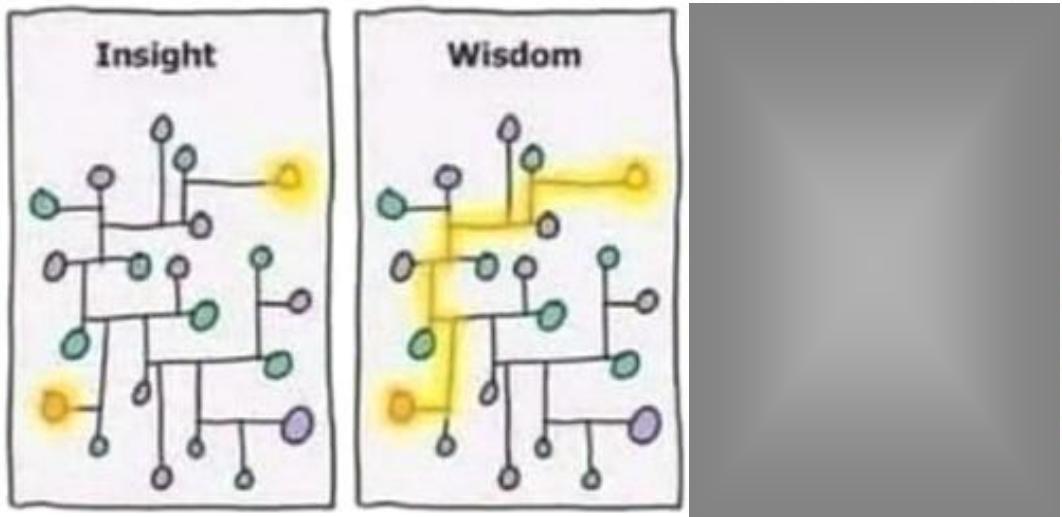
№ з/п	Тема роботи	Повна оцінка	Корегування оцінки		
			Після 4 тижнів	Після 8 тижнів	Після 12 тижнів
1.	Аналіз предметної області та інсталяція СУБД	5	2	1	0
2.	Фізичне проектування бази даних	5	2	1	0
3.	Команди додавання, оновлення та видалення даних	5		3	1
4.	Вирази у запитах на вибірку даних з однієї таблиці	5		3	1
5.	Вибірка даних у запитах для декількох таблиць	5		3	1
6.	Збережені процедури та тригери на модифікацію даних	5			3
7.	Використання індексів при виконанні запитів на вибірку	5			3
8.	Нормалізація та розширення схеми бази даних	5			
9.	Обробка даних у застосунках з використанням СУБД	5			

Дані, інформація, знання

- Зберігання

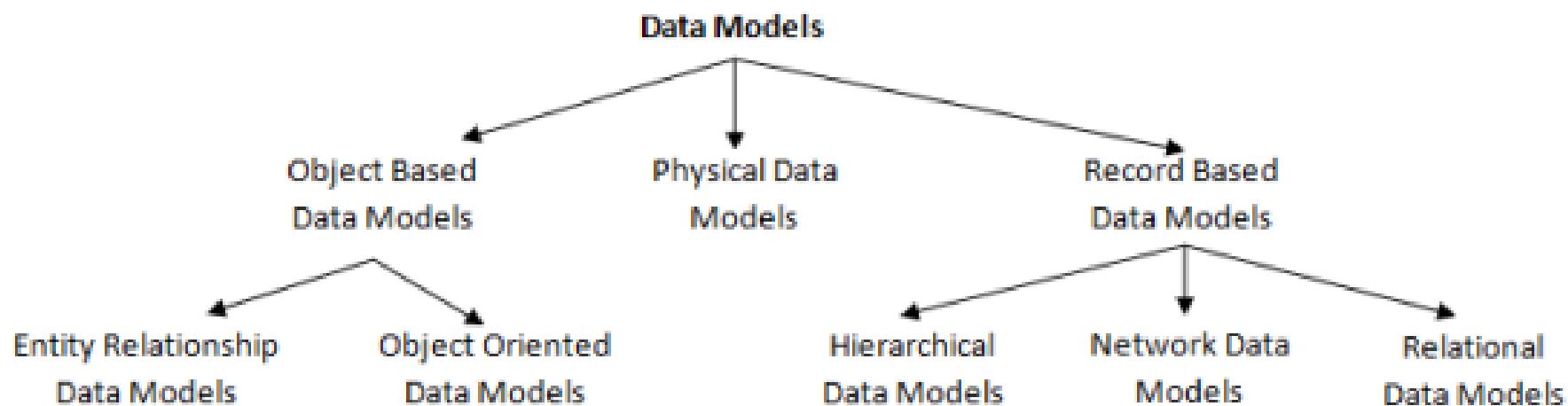


- Інтерпретація



Типологія моделей даних

- **Модель даних** – це представлення об'єктів, подій та існуючих між ними зв'язків. Абстракція, що застосовується до певних даних, в якій акцент робиться на найважливіших аспектах
- **Модель даних** – інтегрований набір понять для опису даних, зв'язків між ними та встановлених обмежень в деякій інформаційній системі



Моделі даних



- Формат документа визначає і розмітку, і дані
- Дескриптори характеризують документ в цілому
- Графи зв'язують мінімальні інформаційні одиниці
- Множинні моделі оперують макрооб'єктами

Предметна область

- **Бізнес-процес** – задачі і заходи, спрямовані на досягнення певної мети в рамках організації. Автоматизація збору та обробки інформації певного бізнес-процесу потребує визначення її структури.
- **Предметна область** – частина реального світу, що розглядається в межах обраного контексту. Визначається межами опису певної діяльності.
- **Сутність і відношення** – ключове явище або об'єкт в предметній області. Сутність має зрозуміли закономірності, особливі властивості, поведінку. Відношення описує існуючий сталий зв'язок між сущностями.
- **Атрибут і параметр** – властивість, яка характеризує об'єкт. Назва та тип виконує функцію метаданих, тобто описує призначення даних та форму зберігання.
- **Дані в базах даних** – це набір конкретних значень відповідних параметрів, які характеризують об'єкт, умову, ситуацію. Мають чітку структуру та визначені обмеження.

Терміни баз даних

- **Метадані** – супровід даних, може бути частиною моделі. Виконує формальний опис набору або елемента даних.
- **Схема** даних – опис структури відповідно до обраної інформаційної моделі. Визначає типи, обмеження, зв'язки.
- **Стан** або вміст – сукупність інформації, що зберігається в певний момент часу. Зміна даних переводить систему в інший стан.
- **Бізнес-логіка** або логіка домену – правила обробки даних з урахуванням особливостей предметної області.
- **Цілісність** – коректність та несуперечливість даних, що зберігаються. Механізми підтримки запобігають порушенню цілісності при введені або зміні даних.
- **Транзакція** – послідовність операцій, які виконуються користувачем з метою доступу або зміни вмісту бази даних. Логічно пов'язані операції щодо об'єкта, вдале виконання яких або одномоментно переведе його або в інший цілісний стан, або залишіть у попередньому.

Визначення бази даних

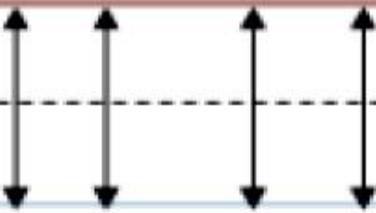
База даних – це сукупність логічно зв'язаних даних (і опис цих даних), яка відображає стан об'єктів та їх зв'язків в певній предметній області, динамічно змінюється у процесі свого функціонування і є доступна багатьом користувачам [Дейт]

Christopher Date: Introduction to Database Systems [1975]

База даних – це єдине велике сховище даних, яке один раз визначається, а потім функціонує у багатокористувачькому режимі [Ульман]

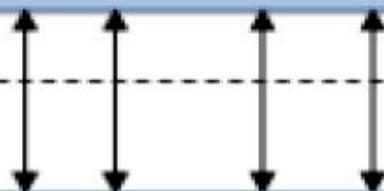
Edgar Codd: A Relational Model of Data for Large Shared Data Banks [1970]

External Level



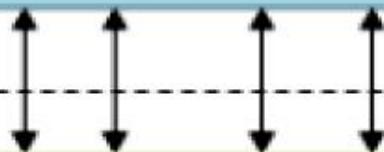
User Interface

Logical / Conceptual Level



Logical Record Interface

Internal Level



Physical Record Interface

Physical Level



OS

Трирівнева архітектура

Presentation layer

Рівень представлення інформації

- забезпечує інтерфейс з користувачем (людиною або іншою програмною системою)
- отримання інформації від користувача відбувається за допомогою різних уніфікованих засобів
- видача результатів реалізується генерацією звітів

Business Logic layer

Рівень бізнес-логіки

- визначає функціональність і працевздатність системи
- обробка даних відповідно до предметної області
- блоки програмного коду використовуються багаторазово для створення складних програм

Data Access layer

Рівень зберігання даних

- забезпечує надійне фізичне зберігання, додавання, модифікацію і вибірку даних
- на цьому рівні контролюється цілісність і узгодженість даних, відбувається виконання транзакцій

Відображення та незалежність даних

- Розділення на рівні дає змогу вирішити проблему **логічної та фізичної** незалежності даних
- Зміни на одному з рівнів не повинні спричиняти зміни на інших, змінюється лише відповідне **відображення**
- Дає можливість змінювати логічну модель бази даних без зміни фізичної структури
- Технологія **зберігання даних** та їхня фізична структура незалежні від коду прикладних програм

Бази даних в архітектурі ПЗ

Тип архітектури	Файл-сервер	Клієнт-сервер (бізнес-логіка на клієнти)	Клієнт-сервер (бізнес-логіка на сервері)	<i>N-рівнева архітектура (Multitier Architecture)</i>
Рівень				
Представлення інформації	Клієнт	Клієнт	Клієнт	Клієнт
Бізнес-логіки	Клієнт	Клієнт	Сервер БД	Сервер прикладних програм (кластер)
Зберігання даних	Файл-сервер (або клієнт)	Сервер БД	Сервер БД	Сервер БД
Реалізація	Усі три рівні утворюють єдиний програмний модуль	Інтерфейс користувача і бізнес логіка утворюють єдиний модуль. Дані зберігаються на сервері	Усі бізнес-логіка реалізована у вигляді збережених процедур, які виконуються на сервері БД	Усі рівні реалізовані на різних машинах

СУБД як інформаційна система

СУБД – це програмне забезпечення підтримки інтегрованої сукупності даних, призначене для створення, зберігання, ведення і використання бази даних багатьма користувачами

Обробка даних:

- Специфікація
- Модифікація
- Отримання
- Адміністрування

Архітектура системи:

- Серверні
- Хмарні
- Вбудовані
- Операційні

Типові функції СУБД

- Визначення структури даних, її ініціалізація і проведення початкового завантаження даних
- Доступність для користувачів опису моделі у формі спеціальної бази метаданих
- Підтримка цілісності даних, яка контролює, щоб дані та запити на зміни відповідали встановленим правилам
- Авторизація та контроль доступу, який надає операції із даними лише санкціонованим користувачам
- Підтримка паралельного виконання транзакцій, що гарантує коректне оновлення бази даних при одночасних змінах
- Відновлення бази даних на випадок пошкодження
- Допоміжні адміністративні утиліти: імпорт та експорт, засоби моніторингу, статистичного аналізу, оптимізації фізичних структур зберігання
- Підтримка єдиної інтегрованої командної мови, що містить необхідні засоби для взаємодії користувача із БД

Набір вимог ACID

■ **Atomicity**

— Атомарність

- транзакція розглядається як одна цілісна операція взаємодії з набором даних, що зберігаються у системі
- при будь-якій помилці відхиляються всі зміни, зроблені з початку цієї транзакції

■ **Consistency**

— Узгодженість

- система має перебувати в узгодженному, несуперечливому стані до початку виконання транзакції і по її завершенню

■ **Isolation**

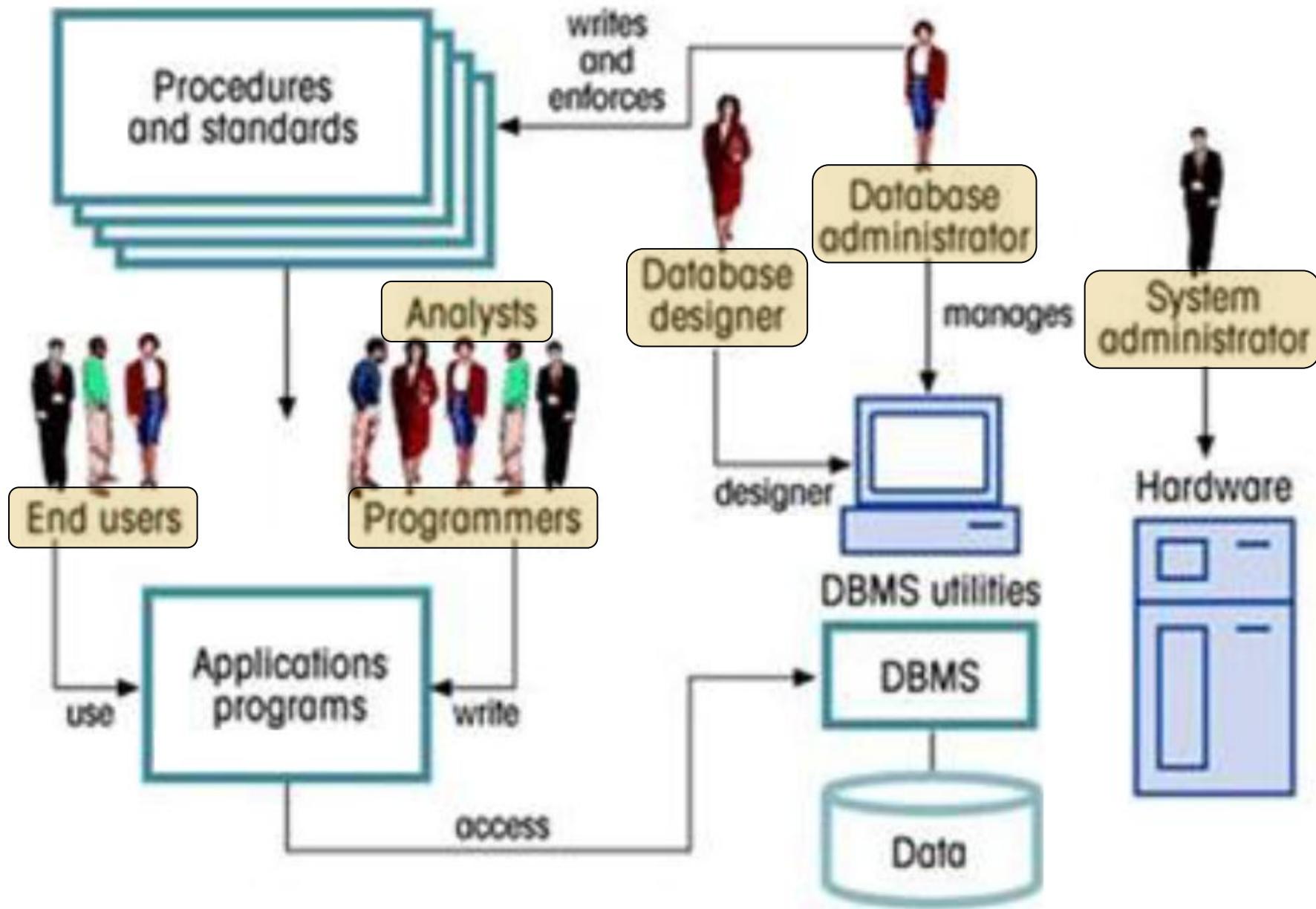
— Ізольованість

- зміни не будуть видимі іншим користувачам за межами транзакції до її вдалого завершення
- проміжний стан даних транзакції має тимчасовий характер

■ **Durability**

— Довговічність

- незалежно від проблем після відновлення працездатності системи результати завершених транзакцій зберігаються
- якщо отримано повідомлення про успішне завершення транзакції, то користувач може бути впевнений, що всі внесені дані збережені у повному обсязі



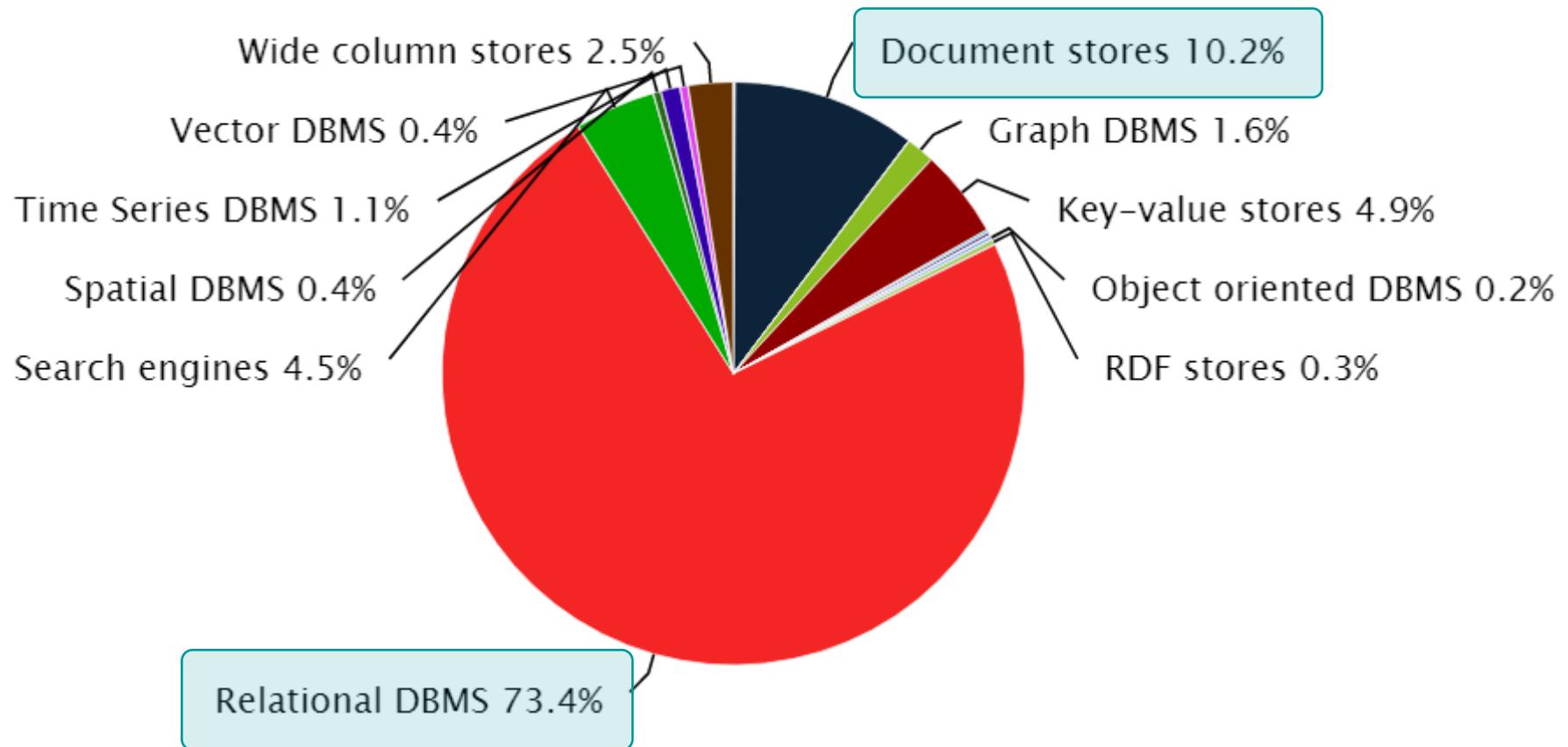
Розподіл обов'язків

- **Адміністратор даних**
 - відповідає за управління даними, планування життєвого циклу даних, розробку та ведення стандартів, бізнес-правил та процедур обробки
- **Адміністратор бази даних**
 - відповідає за фізичну реалізацію бази даних, проектування та втілення проекту, безпеку та цілісність даних, супровід системи і забезпечення продуктивності
- **Розробник логічної бази даних**
 - займається ідентифікацією даних, виявленням сутностей та атрибутів, зв'язками, встановлює обмеження на дані, що зберігаються
- **Розробник фізичної бази даних**
 - згідно логічній моделі виконує фізичну реалізацію: побудову таблиць та обмежень цілісності; вибір структур збереження та методів доступу до даних; проектуванням необхідних мір захисту інформації
- **Прикладні програмісти**
 - після створення бази даних працюють над створенням програм, які надають користувачам необхідні їм функціональні можливості на основі специфікацій, наданих системними аналітиками
- **Користувачі або клієнти**
 - база даних проєктується, створюється та ведеться для того, щоб обслуговування інформаційні потреби кінцевих користувачів

Вибір типу СУБД

- Складність предметної області
- Первісна структурованість сутностей
- Обсяги даних для зберігання
- Змінюваність даних та схеми
- Архітектура прикладного ПЗ
- Плани на взаємодію з БД
- Плани на супровід системи
- Техніко-економічні аспекти

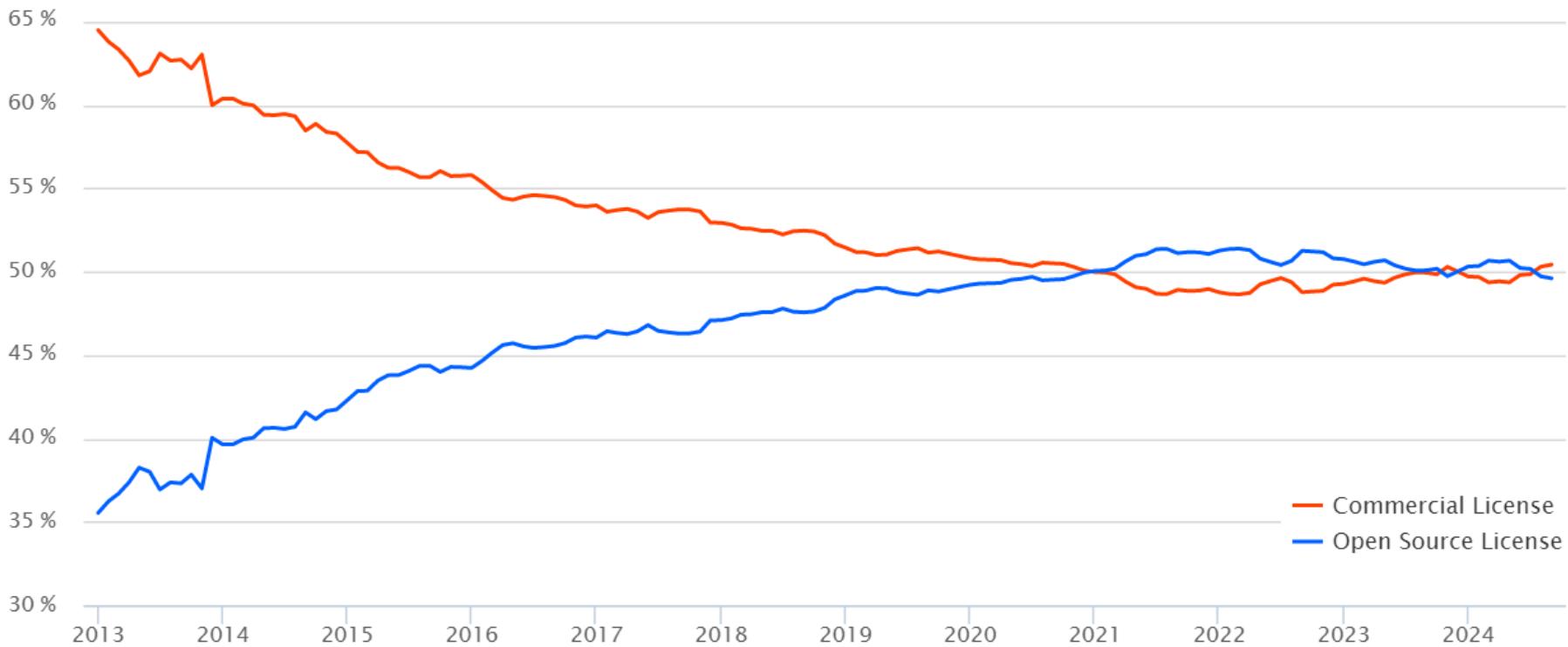
Популярність типів СУБД



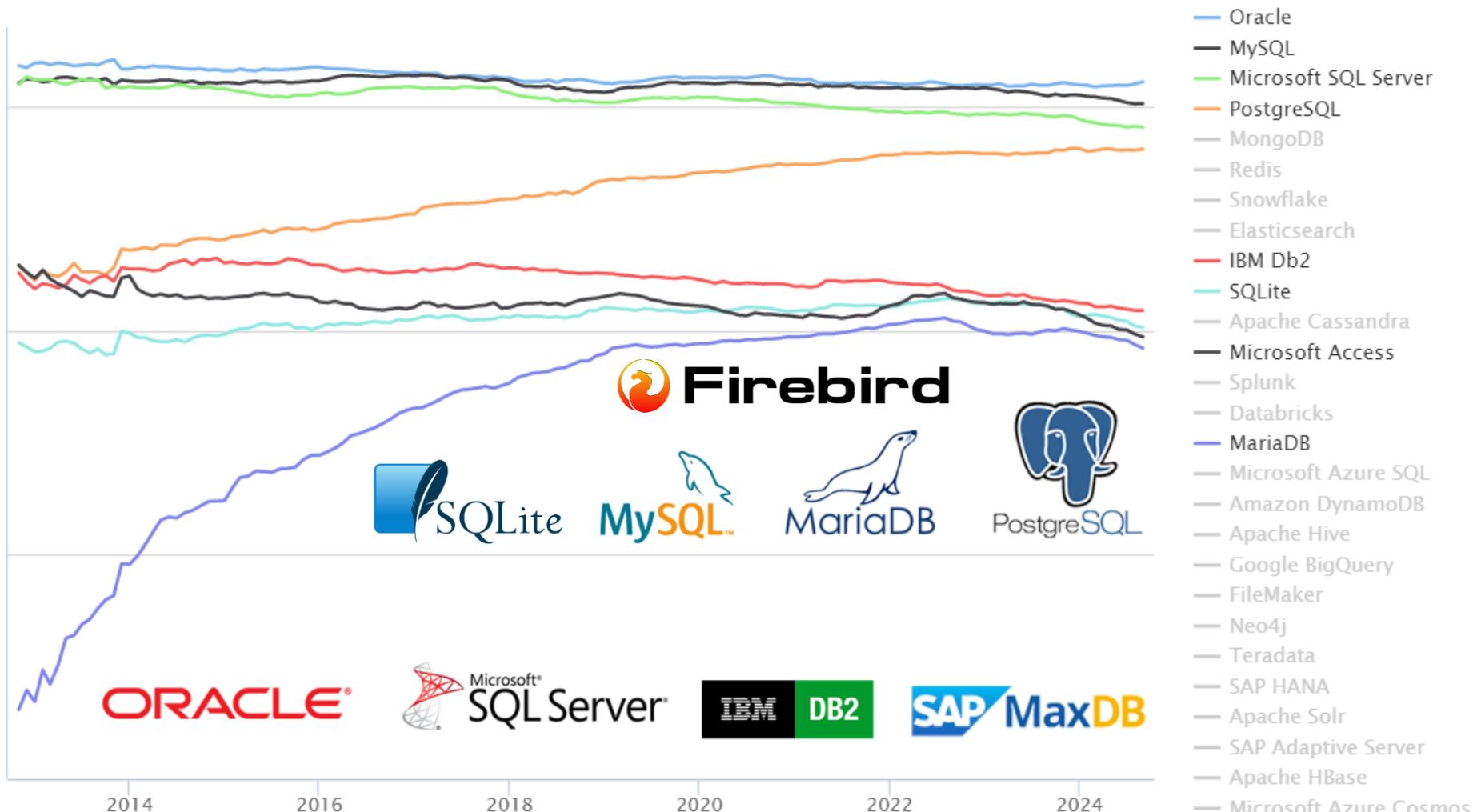
Комерційні та відкриті системи

Rank	System	Score	Overall Rank
1.	Oracle	1287	1.
2.	Microsoft SQL Server	808	3.
3.	Snowflake	134	7.
4.	IBM Db2	123	9.
5.	Microsoft Access	94	12.

Rank	System	Score	Overall Rank
1.	MySQL	1029	2.
2.	PostgreSQL	644	4.
3.	MongoDB	410	5.
4.	Redis	149	6.
5.	Elasticsearch	129	8.



Тренди реляційних баз даних



https://db-engines.com/en/ranking_trend

Дякую за увагу!



DATABASE SYSTEMS

The Complete Book

Second Edition



**Database Modeling
& Design:
Logical Design**

Fourth Edition

Hector Garcia-Molina

Jeffrey D. Ullman

Jennifer Widom

*Department of Computer Science
Stanford University*

**TOBY TEOREY
SAM LIGHTSTONE
TOM NADEAU**

- <https://www.postgresql.org/docs/16/>
- <https://dev.mysql.com/doc/refman/8.4/en/>
- <https://learn.microsoft.com/en-us/sql/t-sql/>
- <https://docs.oracle.com/en/database/oracle/oracle-database/23/sqlrf/>