

ТЕХНОЛОГІЇ ЗБЕРІГАННЯ ДАНИХ

Лекція 3
Автор Голуб Б.Л.



ПРИЧИНИ ДЕНОРМАЛІЗАЦІЇ ДАНИХ

- Нормалізація даних в СУБД призводить до створення безлічі пов'язаних між собою таблиць. В результаті, виконання складних запитів неминуче призводить до об'єднання багатьох таблиць, що істотно збільшує час відгуку
- Нормалізація робить модель сховища занадто громіздкою, ускладнює її розуміння і погіршує ефективність виконання запиту
- Створення сховища даних передбачає створення ненормалізованої структури даних (допускається надмірність даних і можливість виникнення аномалій при маніпулюванні даними), орієнтованої в першу чергу на високу продуктивність при виконанні аналітичних запитів



Проблематика побудови сховищ даних

- **Інтеграція різнорідних даних.** Дані в СД надходять з різнорідних OLTP-систем, які фізично можуть бути розташовані на різних вузлах мережі. При проектуванні і розробці СД необхідно вирішувати задачу інтеграції різних програмних платформ зберігання.
- **Ефективне зберігання і обробка великих обсягів даних.** Побудова СД припускає накопичення даних за значні періоди часу, що веде до постійного зростання обсягів дискової пам'яті, а також зростанню обсягу оперативної пам'яті, що вимагається для обробки цих даних. При зростанні обсягів даних це зростання не лінійно.
- **Організація багаторівневих довідників метаданих.** Кінцевим користувачам СППР необхідні метадані, що описують структуру, які зберігаються в СД, а також інструменти їх візуалізації.
- **Забезпечення інформаційної безпеки СД.** Зведена інформація про діяльність компанії, як правило, відноситься до комерційної таємниці і підлягає захисту; крім того, в СД можуть міститися персональні дані клієнтів і співробітників, які також необхідно захищати. Для виконання цієї функції має бути розроблена політика безпеки СД і пов'язаної з ним інфраструктури, а також реалізовані передбачені в політиці організаційні та програмно-технічні заходи щодо захисту інформації.



ВІТРИНИ ДАНИХ



ВД - це спрощений варіант **СД**, який містить лише тематично об'єднані дані



ДВА ВИДИ СД

○ *Фізичне* сховище даних

- Дані переносяться з різних оперативних джерел в єдине сховище, до якого й адресуються аналітичні запити
- Як наслідок, великий обсяг надмірних даних та ймовірність отримання неактуальної інформації

○ *Віртуальне* сховище даних

- Аналітичні запити адресуються безпосередньо до оперативних джерел та зберігаються в оперативній пам'яті комп'ютера
- Як наслідок, мінімізація обсягу збережених даних та робота з поточними, актуальними даними

Недоліки віртуального та переваги фізичного СД

- * Необхідність постійної доступності всіх OLTP-джерел
- * Зниження швидкодії OLTP-систем
- * OLTP-системи не орієнтовані на зберігання даних за тривалий період часу, в міру необхідності дані вивантажуються в архівні, тому не завжди є фізична можливість отримання повного набору даних у СД

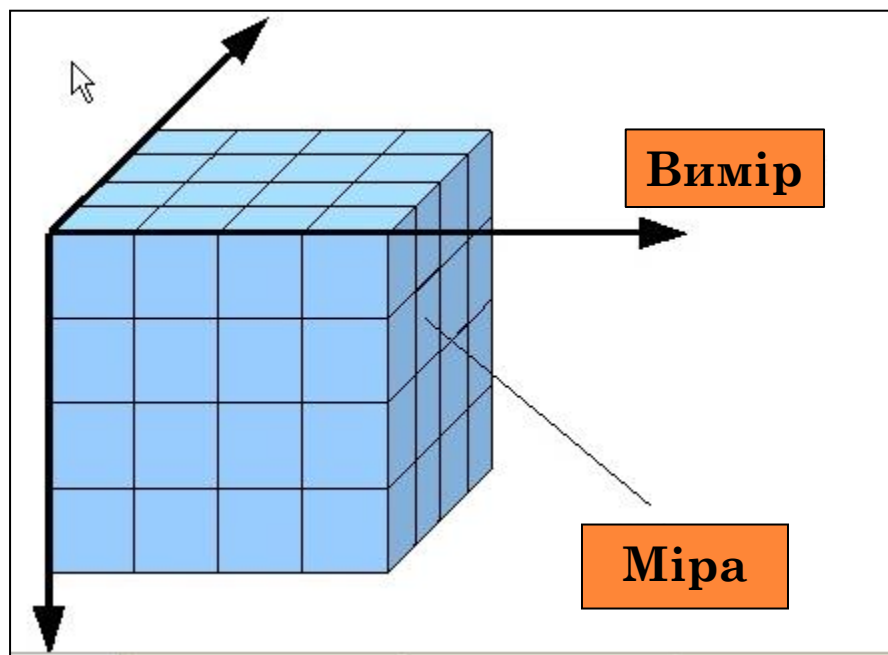


БАГАТОМІРНА МОДЕЛЬ ДАНИХ

- *Вимір* (Dimensions) – це послідовність значень одного з параметрів, які аналізуються
- Множинність вимірів передбачає представлення даних у вигляді *багатомірної моделі*
- *Багатомірне концептуальне представлення* – це множинна перспектива, що складається з декількох незалежних вимірів
- Кожний вимір може бути представлений у вигляді *ієрархічної структури*
- На перетині осей вимірів розташовуються дані, кількісні характеристики фактів, що аналізуються – *міри* (Measures)



ЗОБРАЖЕННЯ БАГАТОМІРНОЇ МОДЕЛІ



ПРАВИЛА КОДДА ДЛЯ OLAP

- Концептуальне багатомірне подання. OLAP-модель повинна бути багатомірною у своїй основі.
- Прозорість. Користувач здатний одержати всі необхідні дані з OLAP-машини, навіть не підозрюючи, звідки вони беруться. Прозорість повинна досягатися й при взаємодії аналітичного інструмента з гомогенним і гетерогенним середовищами БД.
- Доступність. OLAP-модель повинна надавати свою власну логічну схему для доступу до гетерогенного середовища БД і виконувати відповідні перетворення для надання даних користувачеві.
- Постійна продуктивність при розробці звітів. Продуктивність формування звітів не повинна істотно падати з ростом кількості вимірів і розмірів бази даних.
- Клієнт-серверна архітектура. Потрібно, щоб продукт був не тільки клієнт-серверним, але й щоб серверний компонент був би досить інтелектуальним для того, щоб різні клієнти могли підключатися з мінімумом зусиль і програмування.
- Загальна багатомірність. Всі виміри повинні бути рівноправні, кожний вимір має бути еквівалентним й у структурі, і в операційних можливостях.



ПРАВИЛА КОДДА

- Динамічне керування розрідженими матрицями. OLAP системи мають автоматично набудовувати свою фізичну схему залежно від типу моделі, обсягів даних і розрідженості бази даних.
- Багато користувальницька підтримка. OLAP-інструмент повинен надавати можливості спільного доступу (запиту й доповнення), цілісності й безпеці.
- Необмежені перехресні операції. Всі види операцій повинні бути дозволені для будь-яких вимірів.
- Інтуїтивна маніпуляція даними. Маніпулювання даними здійснювалося за допомогою прямих дій над осередками в режимі перегляду без використання меню й множинних операцій.
- Гнучкі можливості одержання звітів. Виміри повинні бути розміщені у звіті так, як це потрібно користувачеві.
- Необмежена розмірність і число рівнів агрегації. Дослідження про можливе число необхідних вимірів, що вимагаються в аналітичній моделі, показало, що одночасно може використатися до 19 вимірів. Звідси випливає настійна рекомендація, щоб аналітичний інструмент був здатний одночасно надати як мінімум 15 вимірів, а переважно 20. Більше того, кожне із загальних вимірів не повинне бути обмежене по числу обумовлених користувачем-аналітиком рівнів агрегації й шляхів консолідації.

