

ОСНОВНІ ПОНЯТТЯ ТЕОРІЇ БАЗ ДАНИХ

Вступ

Сучасні інформаційні системи, що базуються на концепції інтеграції даних, характеризуються величезними об'ємами збережених даних, складною організацією, необхідністю задовольняти різноманітним вимогам численних користувачів.

Інформаційна система - система, що реалізує автоматизований збір, обробку і маніпулювання даними і включає в себе технічні засоби обробки даних, програмне забезпечення і відповідний персонал.

Мета будь-якої інформаційної системи - обробка даних про об'єкти реального миру. Основою інформаційної системи є база даних. У широкому сенсі слова база даних - це сукупність відомостей про конкретні об'єкти реального миру в якій-небудь предметній області. Під **предметною областю** прийнято розуміти частину реального миру, що підлягає вивченню, для організації управління і автоматизації, наприклад, підприємство, ВНЗ і т.д.

Створюючи базу даних, користувач прагне упорядкувати інформацію за різними ознаками і швидко проводити вибірку з довільним поєднанням ознак. Велике значення при цьому набуває структуризація даних.

Структуризація даних - це введення угод про способи представлення даних.

Неструктурованими називають дані, записані, наприклад, в текстовому файлі.

Нижче приведений приклад неструктурованих і структурованих даних, що містять відомості про студентів (номер особистої справи, прізвище, ім'я, по-батькові і рік народження).

Неструктуровані дані:

Особова справа № 16493. Сергєєв Петро Михайлович, дата народження 1 січня 1976 г.;

О/с № 16593, Петрова Ганна Володимирівна, дата нар. 15 березня 1975 г.;

№ особ. справи 16693, д.нар. 14.04.76, Науменко Андрій Борисович

Легко переконатися, що складно організувати пошук необхідних даних, що зберігаються в неструктурованому вигляді.

Структуровані дані:

№ особової справи	Прізвище	Ім'я	По батькові	Дата народження
16493	Сергєєв	Петро	Михайлович	01.01.76
16593	Петрова	Анна	Володимирівна	15.03.75
16693	Науменко	Андрій	Борисович	14.04.76

Щоб автоматизувати пошук і систематизувати ці дані, необхідно визначити певні угоди про способи представлення даних, тобто дату народження потрібно записувати однаково для кожного студента, вона повинна мати однакову довжину і певне місце серед решти інформації. Ці ж зауваження справедливі і для решти даних (номер особової справи, прізвище, ім'я, по батькові).

Користувачами бази даних можуть бути різні прикладні програми, програмні комплекси, а також фахівці предметної області, які виступають в ролі споживачів або джерел даних і яких називають кінцевими користувачами.

У сучасній технології баз даних передбачається, що створення бази даних, її підтримка і забезпечення доступу користувачів до неї здійснюються централізовано за допомогою спеціального програмного інструментарію - системи управління базами даних.

База даних (БД) - це поійменована сукупність даних, що відображає стан об'єктів і їх відносин в даній предметній області.

Об'єктом називається елемент предметної області, інформація про який зберігається.

Об'єкт може бути реальним (наприклад, людина, виріб або населений пункт) і абстрактним (наприклад, подія, рахунок покупця або курс, що вивчається студентами). Так, в області продажу автомобілів прикладами об'єктів можуть служити МОДЕЛЬ АВТОМОБІЛЯ, КЛІЄНТ і РАХУНОК. На товарному складі - це ПОСТАЧАЛЬНИК, ТОВАР, ВІДПРАВЛЕННЯ і т.д.

Система управління базами даних (СУБД) це комплекс програмних і мовних засобів, призначених для створення, ведення і сумісного застосування баз даних (іноді говорять, актуалізації) багатьма користувачами.

Централізований характер управління даними в базі даних припускає необхідність існування деякої особи (групи осіб), на яку покладаються функції адміністрування даними, що зберігаються в базі.

Класифікація баз даних

За *технологією обробки даних* бази даних діляться на централізовані і розподілені

Централізована база даних зберігається в пам'яті однієї обчислювальної системи. Якщо ця обчислювальна система є компонентом мережі ЕОМ, можливий розподілений доступ до такої бази. Такий спосіб використання баз даних часто застосовують в локальних мережах ПК.

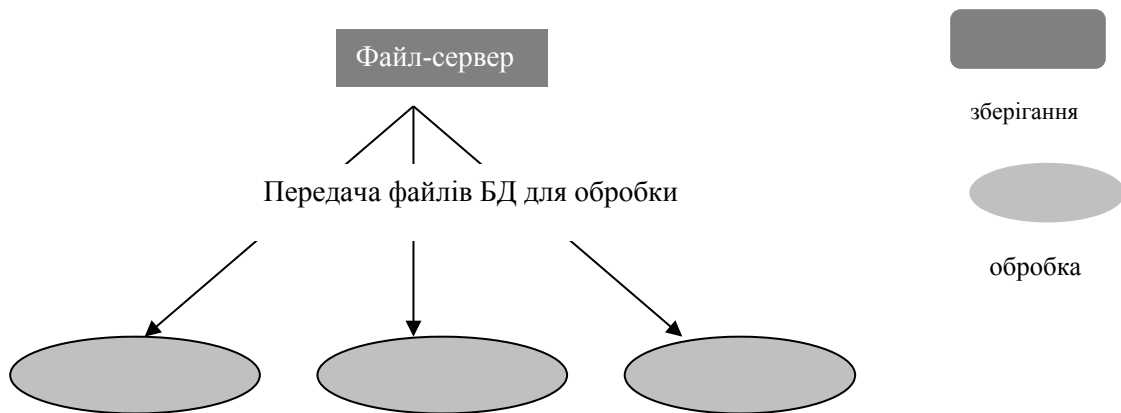
Розподілена база даних складається з декількох частин, що можливо перетинаються або навіть дублюють один одного, і які зберігаються в різних ЕОМ обчислювальної мережі. Робота з такою базою здійснюється за допомогою системи управління розподіленою базою даних (СУРБД).

За *способом доступу до даним* бази даних розділяються на бази даних з локальним доступом і бази даних з видаленим (мережевим) доступом.

Системи централізованих баз даних з мережевим доступом припускають різну **архітектуру** подібних систем:

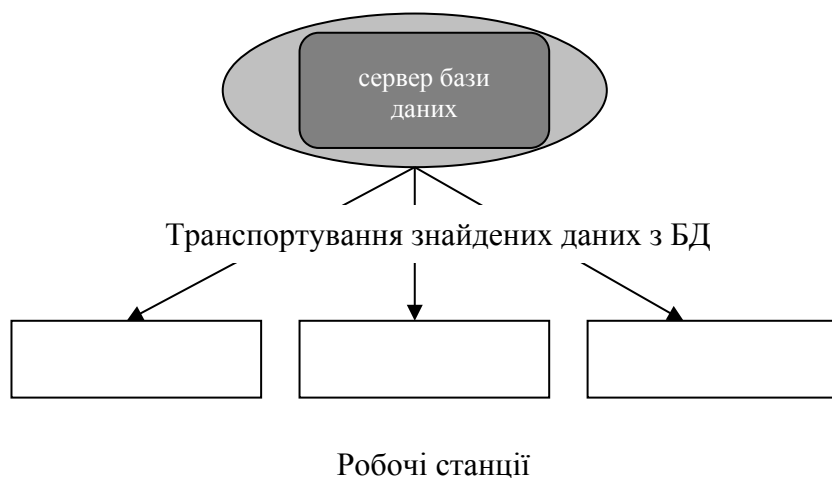
- файл-сервер (товстий клієнт);
- клієнт-сервер (тонкий клієнт).

Файл-сервер. Архітектура систем БД з мережевим доступом припускає виділення однієї з машин мережі як центральної (сервер файлів). На такій машині зберігається спільно використовувана централізована БД. Всі інші машини мережі виконують функції робочих станцій, за допомогою яких підтримується доступ призначеної для користувача системи до централізованої бази даних. Файли бази даних відповідно до призначених для користувача запитів передаються на робочі станції, де в основному і проводиться обробка. При великій інтенсивності доступу до одних і тих же даних продуктивність інформаційної системи падає. Користувачі можуть створювати також на робочих станціях локальні БД, які використовуються ними монопольно. Концепція файл-сервер умовно відображена на мал. 1.



Мал.1 Схема обробки інформації в БД за принципом файл-сервер

Клієнт-сервер. У цій концепції розуміється, що крім зберігання централізованої бази даних центральна машина (сервер бази даних) повинна/може забезпечувати виконання основного обсягу по обробці даних. Запит на дані, що подається клієнтом (робочою станцією), породжує пошук і підготовку даних на сервері. Витягнуті дані (але не файли) транспортуються по мережі від сервера до клієнта. Специфікою архітектури клієнт-сервер є використання мови запитів SQL. Концепція клієнт-сервер умовно зображена на мал. 2.



Мал.2 Схема обробки інформації в БД за принципом клієнт-сервер

Поняття бази даних тісно пов'язане з такими поняттями структурних елементів, як поле (атрибут, домен), запис (кортеж), відношення (таблиця) (мал.3).



Мал.3. Основні структурні елементи БД

Структурні елементи бази даних

Поле (атрибут) - елементарна одиниця логічної організації даних, яка відповідає неподільній одиниці інформації (домену). Для опису поля використовуються наступні характеристики:

найменування, наприклад, Прізвище, Ім'я, По-батькові, Дата народження;

тип, наприклад, символьний, числовий, календарний;

довжина, наприклад, 15 байт, причому визначатиметься максимально можливою кількістю символів;

точність для числових даних, наприклад два десяткові знаки для відображення дробової частини числа

Запис (кортеж) - сукупність логічно зв'язаних полів.

Екземпляр запису - окрема реалізація запису, що містить конкретні значення її полів.

Відношення (таблиця) - сукупність екземплярів записів однієї структури.

Опис логічної структури запису файлу містить послідовність розташування полів записи і їх основні характеристики, як це показано на мал. 4 і 5.

Ім'я таблиці					
Поле (атрибут)		Ознака ключа	Формат поля		
Ім'я (найменування)	Повне найменування		Тип	Довжина	Точність (для чисел)
ім'я 1					
ім'я N					

Мал. 4. Опис логічної структури запису (кортежу) таблиці

У структурі запису можуть бути вказані поля, значення яких є ключами: первинними (ПК) і вторинними (ВК).

Первинний ключ (ПК) - це одне або декілька полів, що однозначно ідентифікують запис. Якщо первинний ключ складається з одного поля, він називається простим, якщо з декількох полів - складеним ключом.

Вторинний ключ (ВК) - це одне або декілька полів, які виконують роль пошукових ознак або ознак групування. На відміну від первинного, значення вторинного ключа може повторюватися в декількох записах файлу, тобто він не є унікальним. Якщо за значенням первинного ключа може бути знайдений один єдиний екземпляр запису, то за вторинним - декілька.

Ім'я таблиці: СТУДЕНТ					
Поле		Ознака ключа	Формат поля		
Позначення	Повне найменування		Тип	Довжина	Точність
Номер Npp	№ особової справи	•	Симв	5	
Прізвище Р	Прізвище студента		Симв	15	
Ім'я І	Ім'я студента		Симв	10	
По-батькові В	По-батькові студента		Симв	15	
Дата BD	Дата народження		Дата	8	

Мал.5. Опис логічної структури запису файлу (таблиці) СТУДЕНТ

Моделі даних

Ядром будь-якої бази даних є модель даних. Модель даних представляє собою множину структур даних, обмежень цілісності та операцій маніпулювання даними. За допомогою моделі даних подаються об'єкти предметної області та взаємозв'язки між ними.

Цілісність БД - властивість бази даних, яка означає, що у БД міститься повна, без протиріч і адекватно відображаюча сутність предметної області інформація. Підтримка цілісності БД містить у собі перевірку цілісності даних, відношень та обмежень у БД, і її відновлення у випадку появи протиріч. Цілісність БД описується за допомогою обмежень цілісності у вигляді умов, яким мають задовольняти дані, що зберігаються у БД.

Модель даних - сукупність структур даних і операцій по їх обробці.

СУБД базується на використанні ієрархічної, мережевої, реляційної або пост-реляційної (нереляційної) моделі даних, на комбінації цих моделей або на деякій їх підмножині.

Будемо спочатку розглядати 3 основні типи моделей даних: ієрархічну, мережеву і реляційну, після чого розглянемо приклад не реляційної бази даних MongoDB.