**Лекція 2. Основні поняття розподілених баз даних**

**1. Поняття розподіленої баз даних.**

**2. Моделювання розподілених СУБД.**

**3. Розподілена обробка даних у розподілених СУБД.**

**4. Обробка даних за технологією клієнт-сервер.**

**2.1. Розподілена система баз даних**

Розподілена база даних (РБД) чи Dіstrіbuted DаtaBase (DDB), це сукупність розподілених даних представлених у рамках деякої розподіленої реляційної моделі, яка відображує дані і зв'язки між ними.

**Типовий склад РБД**

Розподілені бази даних розробляються в рамках технологічної стратегії реалізація якої і є основною задачею прикладного програміста.

Потреба у проектування розподілених баз даних виникає під час розробки інформаційної системи призначеної для автоматизації функцій підприємства складної структури, підрозділи якого розподілені у просторі та відрізняються за своїм призначенням.

Як правила до такого роду підприємств відносяться великі підприємства, які є ме5режами обслуговування користувачів. Приклади підприємств:

* мережа з виробництва та реалізації меблів
* мережа аптек
* мережа продуктових супермаркетів
* мережа реалізації електронного та обслуговування обладнання

В загальному випадку до складу підприємств складної структури входять наступні підрозділи:

**Типові філії** - сукупність підрозділів з однаковими функціями, які розташовані в різних місцях( наприклад, салони з реалізації меблів у різних міста)

**Спеціалізовані філії** – підрозділі, що входять до мережі, але виконують власний, спеціалізованій набір функцій ( наприклад, загальній склад товарів мережі, виробниче підприємство або транспортне підприємство)

**Центральній офіс –** головне підприємство, яке займається контролем, аналізом та оптимізацією роботи мережі в цілому.

Розробка інформаційної системи для такого підприємства передбачає створення розподіленої бази даних та заснована на наступних методологічних ознаках:

* виділення технологічних даних, які збираються та обробляються на локальних вузлах розподіленої системи(у типових філіях великої установи) та формування локальних баз даних типових філій(бази однакової структури, для кожної типової філії). Локальні бази даних типових філій містіть набір таблиць необхідних для реалізації функцій типової філії. Перелік таблиць формуються на підставі аналізу функцій структурних підрозділів(відділів типової філії
* виділення технологічних даних, які збираються та обробляються на спеціалізованих локальних вузлах розподіленої системи(у спеціалізованих філіях великої установи) та формування спеціалізованих локальних баз даних (для кожної спеціалізованої філії ) Локальні бази даних спеціалізованих філій містіть набір таблиць необхідних для реалізації функцій філії. Перелік таблиць формуються на підставі аналізу функцій структурних підрозділів(відділів спеціалізованої філії
* виділення та формування інтегрованих даних системи, які забезпечують виконання функцій аналізу та контролю роботи мережі та створення "головної бази даних" ( БД центрального офісу великої установи), яка також є частиною розподіленої бази даних і містить інтегровані дані. Перелік таблиць головної формуються на підставі аналізу функцій структурних підрозділів центрального офісу. Особливістю головної БД є формування крім звичайних таблиць, особливих інтегрованих таблиць двох видів:

1. інтегровані показники по підприємству в цілому - інтегратори першого роду(наприклад інтегровані показники з придбання та реалізації товарів по періодам)
2. інтегровані показники по в розрізі філій - інтегратори другого роду(наприклад порівняльні інтегровані показники з придбання та реалізації товарів по філіям )

* виділення сукупності службових технологічних даних. Технологічні дані не несуть основної інформації, вони описують структуру та зміст самої розподіленої бази даних (склад і розташування у мережі). Технологічні дані, будучи частиною сукупності розподілених даних, утворюють службову інструментальну базу метаданих системи. База метаданих містить декілька груп службових таблиць:

1. Основні таблиці для опису розташування локальних баз. В групу входять зокрема таки таблиці як «СЕРВЕР», «БАЗА ДАНИХ», «ТАБЛИЦЯ», «ПОЛЕ»
2. Таблиці для забезпечення підтримки фрагментованих таблиць
3. Таблиці для забезпечення підтримки процедур реплікації
4. Таблиці для забезпечення підтримки розподілених транзакцій

**Принципи розробки РБД**

Розробка розподіленої бази даних потребує застосування засобів реалізації обмінної технології з метою забезпечення синхронності змін, несуперечності даних і функціонування алгоритмів формування головної бази даних.

У більшості випадків для реалізації розподіленої бази даних використовується реляційна модель представлення даних,яка забезпечує однорідність представлення даних і зв'язків між ними у вигляді концептуальної схеми бази даних.

Реалізація реляційної моделі для складного підприємства приводить до створення розподіленої бази даних(системи реляційних розподілених баз даних).

РБД, яка створюється заново, є однорідною. Разом з цим, нерідко вона створюється як сукупність групи локальних баз даних, що уже функціонують у ряді систем. У цьому випадку виникає неоднорідна РБД. Обидва типи баз поміщаються в Систему Керування Розподіленою Базою Даних (СКРБД).

У загальному випадку локальні бази даних, що складають розподілену базу даних, не обов'язково повинні бути однорідними (тобто вестися однією СКБД) чи оброблятися в середовищі однієї і тієї ж операційної системи і на комп'ютерах того самого типу.

Наприклад, одна база даних може бути базою Oracle на комп'ютері SUN з операційною системою SUN OS(UNІX), друга база даних може вестися СКБД DB2 на мейнфреймі ІBM 3090 з операційною системою MVS, а для ведення третьої бази може використовуватися СКБД SQL/DS також на мейнфреймі ІBM, але з операційною системою VM. Обов'язкова тільки одна умова - ЛБД зв'язані між собою логічно, але фізично розміщені на декількох машинах (вузлах), що входять в одну комп'ютерну мережу.

При цьому мережна підтримка РБД характеризується наступними особливостями:

* кожен вузол має власні бази даних;
* вузли працюють погоджено, тобто користувач може одержати доступ до даних на будь-якому вузлі мережі так, ніби-то вони знаходяться на його власному вузлі;
* функціонування РБД відбувається незалежно від типів використовуваних у системах пристроїв;
* забезпечується просторова прозорість, що дає можливість користувачу не знати, де розташовані компоненти бази;
* підтримка повної функціональності, тобто можливість виконання всіх операцій, що можливі в базі даних, яка знаходиться в одній системі;
* гарантія цілісності даних, яку забезпечують функції спостереження за даними, виправлення помилок.

Найважливіші функціональні характеристики розподіленої бази даних такі:

* автономність вузлів розподіленої бази даних яка означає, що ведення кожної бази даних може відбуватися незалежно від інших;
* обробка розподілених запитів - таких запитів (SQL-команда), у ході виконання яких відбувається доступ до об'єктів (таблиць чи представлень) різних баз даних;
* виконання розподілених транзакцій. При виконанні розподілених транзакцій здійснюється погоджене керування усіма залученими базами даних. При цьому часто використовується технологія двофазної передачі інформації для виконання розподілених транзакцій.

Суть РБД полягає в організації доступу користувачів до великих об’ємів інформації. Це дозволяє розмістити дані так, що останні, з одного боку, знаходяться в пунктах їх найбільшого опиту, а з іншого боку, за допомогою транзакцій забезпечується доступ до будь-яких даних, не залежно від того, де вони знаходяться.

Сучасні комп'ютерні технології при проектуванні розподілених баз даних розглядаються в різних варіантах архітектури "клієнт-сервер" (дворівнева, трирівнева), що орієнтується на об’єктно-орієнтований підхід. У цій архітектурі будь-який об'єкт, що використовує ресурси іншого об'єкта, визначається як клієнт, а об'єкт, що поставляє ресурси називається сервером відповідно до встановленого протоколу обміну.

**2.2. Моделювання розподілених СКБД**

Існують різні моделі для побудови схем розподіленої бази даних. Завдяки схемі користувач розглядає РБД як єдину БД.

Різні автори пропонують різні підходи до побудови набору схем, які розроблюються при проектуванні баз даних. Однією з найбільш розповсюджених схем є схема Арсеньєва-Яковлєва (Інтеграція розподілених баз даних, "ЛАНЬ", 2001 рік)

Концептуальне проектування

Концептуальна схема (модель)

Інформаційне проектування

Інфологічна схема (модель)

Даталогічне подання

Реляційна схема обраної СКБД

Схема Арсеньєва-Яковлєва застосовується для проектування всіх складових розподіленої базі даних(БД типових філій, БД спец філій, головної БД та бази метаданих)

**2.2.1. Концептуальне моделювання**

Складається з наступних операцій:

* визначення джерел знань про предмет (закон евристики, існуючі рішення, документи);
* вибір системи класифікації об'єктів;
* виділення абстракції об'єктів і процесів, тобто виділення класів (типізація, абстрагування);
* побудова структури базових класів;
* побудова схеми взаємодії.

Мета концептуальної побудови - побудова фундаментальної семантики в представленій області.

Нотації опису різні, наприклад у мові UML концептуальна схема може бути представлена у видгляді діаграм класів без баз деталізації по атрибутах і методах.

**2.2.2. Інфологічне моделювання**

Виконується на основі концептуальної моделі як її уточнення і деталізація.

Основні операції:

* формування базових атрибутів класів, які виходять з типізації;
* формування додаткових атрибутів;
* визначення сукупного спільного використання атрибутів, тобто породження прототипів таблиць;
* визначення зв'язків логічної цілісності і зв'язків навігації;
* побудова інформаційно - логічної моделі.

Нотації різні.

У мові UML - повна діаграма статичних класів.

**2.2.3. Даталогічне моделювання**

Даталогічний аналіз і проектування це остаточне проектування реляційної схеми даних з врахуванням середовища обраної СКБД.

Проектування виконується на основі інфологічної схеми і полягає у виконанні таких пунктів:

* виділення таблиць (R-відношень);
* визначення фізичних форматів атрибутів на основі типів СКБД;
* визначення складу індексованих полів;
* визначення зв'язків навігації і логічної цілісності;
* установка бізнес-правил;
* проектування представлень;
* проведення нормалізації реляційних схем (усунення надлишковості, багатозначності).

Результат даталогічного проектування - реляційна схема бази даних.

**2.3. Розподілена обробка даних - Dіsdtrіbuted Data Processіng (DDP)**

Виконання операцій з базами даних на одній машині і додатків на інший і є розподілена обробка даних. Зв'язок між додатком і процесами виконання операцій з базами даних реалізується як апаратурою, так і програмним забезпеченням. Сутність DDP полягає у тому, що користувач одержує можливість працювати з мережними службами і прикладними процесами, розташованими в декількох взаємозв'язаних абонентських системах. При цьому можливі кілька видів робіт, які він може виконувати:

* віддалений запит, наприклад, команда, що дозволяє посилати одиночну заявку на виконання обробки даних;
* віддалена транзакція, що здійснює направлення групи запитів прикладному процесу;
* розподілена транзакція, що дає можливість використання декількох серверів і прикладних процесів, які виконуються у групі абонентських систем.

**2.4. Обробка даних за технологією клієнт-сервер**

**2.4.1 Дволанкова архітектура «клієнт-сервер баз даних»**

Для використання РБД необхідно наступне апаратне і програмне забезпечення:

- комп'ютер-сервер бази даних;

- комп'ютер(и)-клієнти;

- комунікаційна мережа;

- мережне програмне забезпечення;

- прикладне програмне забезпечення.

Технологія клієнт-сервер є реалізацією розподіленої обробки даних. У системі архітектури клієнт-сервер обробка даних розділена між комп'ютером-клієнтом і комп'ютером-сервером, зв'язок між якими відбувається по мережі. Цей поділ процесів обробки даних засновано на групуванні функцій. Як правило, комп'ютер-сервер баз даних виділяється для виконання операцій з базами даних, а комп'ютер-клієнт виконує прикладні програми. На рисунку 1 показана проста система архітектури клієнт-сервер, до складу якої входять комп'ютер, що діє як сервер, і інші комп'ютер, що діють як його клієнти. Кожна машина виконує різні функції і має свої власні ресурси.

Застосування- реалізація логіки

Застосування- реалізація логіки

Сервер БД - виконання операції з базою даних

Комп'ютер-сервер

ІBM - сумісний ПК

ІBM – сумісний ПК

Комп’ютери - клієнти

мережа

**2.4.2 Триланкова архітектура «клієнт-сервер застосувань сервер БД»**

З метою підвищення ефективності обробки розподіленої інформації застосовуються три ланкова технологія **«клієнт - сервер застосувань - сервер БД»** клієнт-сервер є реалізацією розподіленої обробки даних. У системі архітектури клієнт-сервер обробка даних розділена між комп'ютером-клієнтом, комп'ютером-сервером застосувань, і комп'ютером-сервером, зв'язок між якими відбувається по мережі. Цей поділ процесів обробки даних засновано на групуванні функцій. Як правило, комп'ютер-сервер баз даних виділяється для виконання операцій з базами даних, а комп'ютер-сервер застосувань виконує загальні типові функції потрібні всім клієнтам, а комп'ютер – клієнт реалізує інтерфейс користувача та здійснює звернення до комп’ютера з сервером застосувань . На рисунку 2 показана проста система архітектури **«клієнт - сервер застосувань - сервер БД»** клієнт-сервер, до складу якої входять комп'ютер, що діє як сервер, комп'ютер, що діє як сервер застосувань та і інші комп'ютері, що діють клієнти. Кожна машина виконує різні функції і має свої власні ресурси.

**Комп'ютер-сервер БД**

**Комп'ютер-сервер застосувань**

Сервер БД – реалізація операцій з базою даних базі даних

Сервер БД – застосування для виконання типових функцій клієнта( операции с БД)

Застосування «клієнт – реалізація інтерфейсу користувача

**Комп'ютер-клієнт**

**Комп'ютери-клієнти**

Основна функція комп'ютера-клієнта у дволанковій архітектурі полягає у виконанні клієнтського застосування. Яке реалізує інтерфейс користувача і всю бізнес логіку та здійсненні зв'язку із сервером баз даних, коли цього вимагає додаток.

Основна функція комп'ютера-клієнта у триланковій архітектурі полягає у виконанні клієнтського застосування. яке реалізує інтерфейс користувача і виконує з’єднання з сервером застосувань та здійснює звернення до сервера застосувань для виконання типових функцій

Зазвичай клієнти - це повільні і недорогі машини. У залежності від вимог, пропонованих додатками, можуть знадобитися більш потужні комп'ютери-клієнти.

**Комп'ютери-сервери застосувань**

Основна функція комп'ютера-сервера застосувань полягає у виконанні застосування призначеного для очікування звернень від клієнтів, виконання загальних функцій та в разі потреби здійсненні зв'язку із сервером баз даних, коли цього вимагає виконуваня функція

**Комп'ютер-сервер бази даних**

Як випливає вже із самого терміну, головна функція комп'ютера-сервера полягає в обслуговуванні потреб клієнта. Термін "Сервер" використовується для позначення двох різних груп функцій: файл-сервер і сервер баз даних. Файли-сервери не призначені для виконання операцій з базами даних, їх основна функція - поділ файлів між декількома користувачами, тобто забезпечення одночасного доступу багатьох користувачів до файлів на комп'ютері файлі-сервері. Прикладом файлу-сервера є операційна система NetWare компанії Novell.

Комп'ютер із сервером баз даних (наприклад Oracle) виділяється тільки для виконання операцій з базами даних.

Основні функції сервера баз даних.:

- зв'язок із клієнтом, аналіз і виконання запиту до бази даних, включаючи повернення клієнту результату запиту (набору рядків з бази даних);

- керування одночасним доступом до бази даних багатьох користувачів, пере направлення запитів до інших серверів мережі;

- виконання реляційних операцій, керування словником-довідником даних і забезпечення захисту інформації.

Одна з важливих вимог до сервера - це те, що операційна система, у середовищі якої розміщений сервер баз даних, повинна бути багатозадачною (і, бажано, але не обов'язково, загального користування). Багато різновидів систем UNІX, MVS, VM та інші операційні системи є і багатозадачними, і загального користування.

При виборі комп'ютера-сервера бази даних варто взяти до уваги, потужність процесора, ємність оперативної пам'яті, обсяг дисків, швидкість роботи з дисками, число одночасно працюючих користувачів і вартість.

**2.4.4 Принципи взаємодії між клієнтськими та серверними частинами**

Доступ до бази даних від прикладної програми чи користувача відбувається шляхом звертання до клієнтської частини системи. Як основний інтерфейс між клієнтською і серверною частинами виступає мова баз даних SQL.

Ця мова являє собою стандарт інтерфейсу СКБД у відкритих системах. Збірна назва SQL-сервер відноситься до всіх серверів баз даних, заснованих на SQL. Сервери баз даних, інтерфейс яких заснований винятково мовою SQL, мають свої переваги і свої недоліки. Очевидна перевага - стандартність інтерфейсу. Клієнтські частини будь-якої SQL-орієнтованої СКБД могли б працювати з будь-яким SQL-сервером (хоча поки це не зовсім так). Недолік досить очевидний. При такому високому рівні інтерфейсу між клієнтською і серверною частинами системи на стороні клієнта працює занадто мало програм СКБД. Це нормально, якщо на стороні клієнта використовується малопотужна робоча станція. Але якщо клієнтський комп'ютер має достатню потужність, то часто виникає бажання покласти на нього більше функцій керування базами даних, розвантаживши сервер, що є вузьким місцем усієї системи.

Одним з перспективних напрямків СКБД є гнучке конфігурування системи, при якому розподіл функцій між клієнтською і користувацькою частинами СКБД визначається при установці системи.