

Анотація навчальної дисципліни вільного вибору здобувача вищої освіти

Дисципліна:	«Сховища та простори даних»
Викладач:	Вороницька Віра Михайлівна, м.т.н., старший викладач
E-mail:	vera.voronitska@gmail.com
Кількість кредитів:	4
Мова викладання:	українська
Вид контролю:	залік
Місце у структурно-логічній схемі:	вивчається в 5 семестрі першого (бакалаврського) рівня вищої освіти за спеціальністю 122 Комп'ютерні науки

Вступ

Концепція сховищ даних.

На початку восьмидесятих років ХХ століття, в період бурхливого розвитку реєструючих інформаційних систем, виникло розуміння обмеженості можливостей їх застосування для аналізу даних і побудови на їх основі систем інтелектуального аналізу даних.

Реєструючі системи створювалися для автоматизації рутинних операцій по веденню бізнесу – виписки рахунків, оформлення договорів, перевірки стану складу тощо. Основними вимогами до таких систем були забезпечення транзакційності змін, що вносилися, і максимізація швидкості їх виконання. Саме ці вимоги визначили вибір реляційних СУБД і моделі представлення даних «суть-зв'язок» в якості основного технічного рішення при побудові реєструючих систем.

Для менеджерів і аналітиків у свою чергу були потрібні системи, які б дозволяли:

- аналізувати інформацію в часовому аспекті;
- формувати довільні запити до системи;
- обробляти великі об'єми даних;
- інтегрувати дані з різних реєструючих систем.

Очевидно, що реєструючі системи не задовольняли жодній з вищезгаданих вимог. У реєструючій системі інформація актуальна тільки на момент звернення до бази даних, в наступний момент часу по тому ж запиту можна отримати абсолютно інший результат.

Інтерфейс реєструючих систем розрахований на проведення жорстко певних операцій і можливості отримання результатів на нерегламентований (ad-hoc) запит сильно обмежені.

Можливість обробки великих масивів даних також мала через налаштування СУБД на виконання коротких транзакцій і неминучого уповільнення роботи решти користувачів.

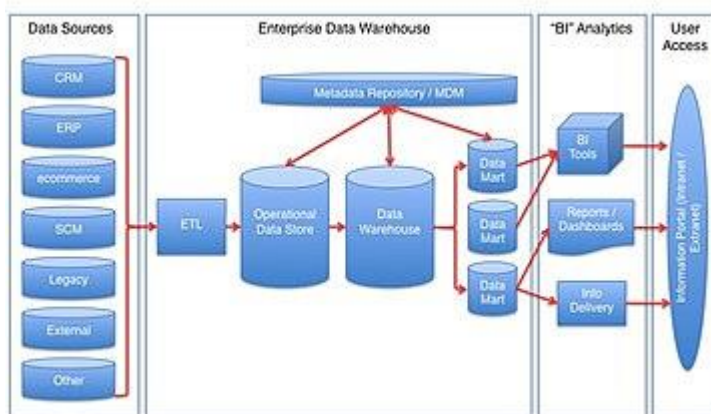
Відповіддю на виниклу потребу стала поява нової технології організації баз даних – технології сховищ даних (Data Warehouse).

За спостереженнями дослідницької компанії Forrester Research, більшість крупних компаній стикаються з наступною проблемою: вони накопичують величезну кількість інформації, яка ніколи не використовується. Практично в будь-якій організації реально функціонує безліч транзакційних систем, орієнтованих на оперативну обробку даних (кожна для конкретного класу задач) і безперервно поповнюють численні бази даних. Окрім цього, часто підприємства володіють величезними об'ємами інформації, що зберігається в так званих успадкованих системах. Всі ці дані розподілені по мережах персональних комп'ютерів, зберігаються на мейнфреймах, робочих станціях і серверах. Таким чином, інформація є, але вона розосереджена, неузгоджена, неструктурована, часто надмірна і не завжди достовірна.

Тому в більшості організацій ці дані до цих пір не можуть бути використані для ухвалення критичних бізнес-рішень. На вирішення цього протиріччя і направлена концепція сховищ даних (Data Warehouse).

Зберігання даних є наріжним каменем систем бізнес-аналітики. 87% компаній мають одне або більше функціонуючих сховищ даних, причому більше половини мають декілька подібних сховищ. Фактично, кожні чотири компанії мають двадцять або більше сховищ, які деяким способом логічно інтегровані, 18% цих сховищ має об'єм в терабайт або більше, 30% респондентів чекає, що об'єм їх сховища даних зросте удвічі або навіть у декілька разів за подальші два або три роки. Різноманітність джерел даних, що наповнюють сховища, просто дивує. Природно, кожне сховище використовує транзакційні системи як джерело даних. Проте відомо декілька незвичних джерел даних, такі як Web-сайти електронної комерції (51%), прив'язки до баз даних, розташованих у замовників або бізнес-партнерів (49%), сервіси розширеного доступу до даних, які забезпечують характеристичну інформацію по замовниках (47%), витягання інформації з Web-сайтів (41%), сервіси пошуку/виявлення (34%) і канали новинних даних від крупних новинних агентств (33%). Нарешті, 47% компаній мають одну або більш вітрин даних, і 49% використовують накопичувачі операційних даних.

В основі концепції сховищ даних лежить ідея об'єднання корпоративних даних, розсіяних по системах оперативної обробки даних, історичних архівах і інших зовнішніх джерелах.



Мета концепції сховищ даних – прояснити відмінності в характеристиках даних в операційних і аналітичних системах, визначити вимоги даним, що поміщаються в сховище, визначити загальні принципи і етапи його побудови, основні джерела даних, дати рекомендації по рішенням потенційних проблем, що виникають при їх вивантаженні, очищенні, узгодженні, транспортуванні і завантаженні до цільової бази даних сховища.

Предметом концепції сховищ даних є не аналіз даних, а власне дані, тобто концепція їх підготовки для подальшого аналізу. В той же час концепція сховища даних визначає не просто єдиний логічний погляд на корпоративні дані, а реалізацію єдиного інтегрованого джерела даних.

Передумови для вивчення дисципліни: ОК20 Комп'ютерні мережі, ОК21 Бази даних та інформаційні системи, ВК5 Методи та засоби комп'ютерних інформаційних технологій.

Мета та завдання дисципліни

Метою викладання дисципліни «Сховища та простори даних» є:

- отримання базових знань про системи зберігання даних, особливості сховищ даних та їх призначення;
- формування умінь і навичок проектування сховищ даних та систем бізнес-аналізу;
- знайомство з технологіями інтелектуального аналізу.

Основними **завданнями** вивчення дисципліни «Сховища та простори даних» є:

- вивчення принципів побудови і розробки сховищ даних;
- отримання навичок налаштування сховищ даних;

- проектування і розробка процесу наповнення сховища даних, реалізації запитів до сховищ даних;

Очікувані результати навчання

У результаті освоєння повного курсу навчальної дисципліни «Сховища та простори даних» у здобувачів вищої освіти формуються глибокі, міцні і системні знання, які передбачають вільне володіння понятійним апаратом, розуміння основних задач предмету, його мети та завдання. Студенти повинні:

- знати теорію проектування та розроблення сховищ даних, технології OLAP та Data Mining;
- аналізувати предметну область, визначати потребу використання сховищ даних;
- вміти проектувати сховища даних;
- вміти програмно реалізувати сховища даних засобами наявних СУБД;
- вміти застосовувати технології OLAP та Data Mining.

Програма навчальної дисципліни

Модуль 1. Вступ до дисципліни «Сховища та простори даних». Моделі даних, системи збереження даних. Архітектури сховищ даних.

Тема 1. Архітектури даних. історія розвитку.

Тема 2. Архітектури даних: Бази даних і моделі даних.

Тема 3. Багатовимірні дані. OLAP-технологія, як ключовий компонент СД.

Тема 4. Концепція сховищ даних (СД). OLAP як ключовий компонент СД.

Тема 5. Архітектури сховищ даних.

Тема 6.. Реляційні сховища даних.

Тема 7. Реалізація реляційних сховищ даних.

Модуль 2 .Проектування СД для обраної предметної області.

Тема 8. Віртуальні сховища даних.

Тема 9. Використання сховищ даних. Різні архітектурні рішення СД, реалізація процедур ETL.

Тема 10. Розробка економічної облікової інформаційної системи на основі реляційної бази даних.

Тема 11. Аналіз даних для завантаження в СД.

Тема 12. Розробка моделі СД.

Тема 13. Реалізація СД.

Тема 14. Завантаження даних в СД. Перевірка працездатності.