

Тема 5.

Основи бізнес-аналітики: бази даних та управління інформацією



План лекції:

1. Структурування даних в традиційному файловому середовищі
2. Підхід бази даних до управління даними
3. Використання баз даних для покращення ефективності бізнесу та прийняття рішень
4. Управління базами даних

Цілі лекції:

- Описати основні концепції організації файлів і проблеми керування ресурсами даних у традиційному файловому середовищі.
- Визначити принципи системи управління базами даних та особливості реляційної бази даних.
- Описати важливі принципи проектування бази даних.
- Оцінити інструменти та технології для надання інформації з баз даних для покращення ефективності бізнесу та прийняття рішень.
- Оцінити роль інформаційної політики, адміністрування даних і забезпечення якості даних в управлінні ресурсами організаційних даних.

Проблемна ситуація:

- **Проблема:** отримання знань про клієнтів і ефективне використання фрагментованих даних про клієнтів.
- **Рішення:** Використовуйте технологію реляційної бази даних для збільшення прибутку та продуктивності.
- Правила доступу до даних і повна база даних клієнтів консолідують дані клієнтів.
- Демонструє роль ІТ у створенні тісної взаємодії з клієнтами та стабілізації інфраструктури.
- Ілюструє роль цифрових технологій у стандартизації того, як дані з різних джерел зберігаються, організовуються та керуються.

1. Структурування даних в традиційному файловому середовищі

Концепції організації файлів

- База даних: група пов'язаних файлів
- Файл: група записів одного типу
- Запис: Група пов'язаних полів
- Поле: група символів у вигляді слів або цифр
 - ❖ Описує сутність (особу, місце, річ, про яку ми зберігаємо інформацію)
 - ❖ Атрибут: кожна характеристика або якість, що описує сутність
 - Наприклад, атрибути **ДАТА** або **ОЦІНКА** належать до сутності **КУРС**

ІЄРАРХІЯ ДАНИХ

Комп'ютерна система організовує дані в ієрархії, яка починається з біта, який представляє або 0, або 1. Біти можуть бути згруповані, щоб утворити байт для представлення одного символу, числа або символу.

Байти можуть бути згруповані для формування поля, а пов'язані поля можна згрупувати для формування запису. Пов'язані записи можна зібрати у файл, а пов'язані файли можна організувати в базу даних.

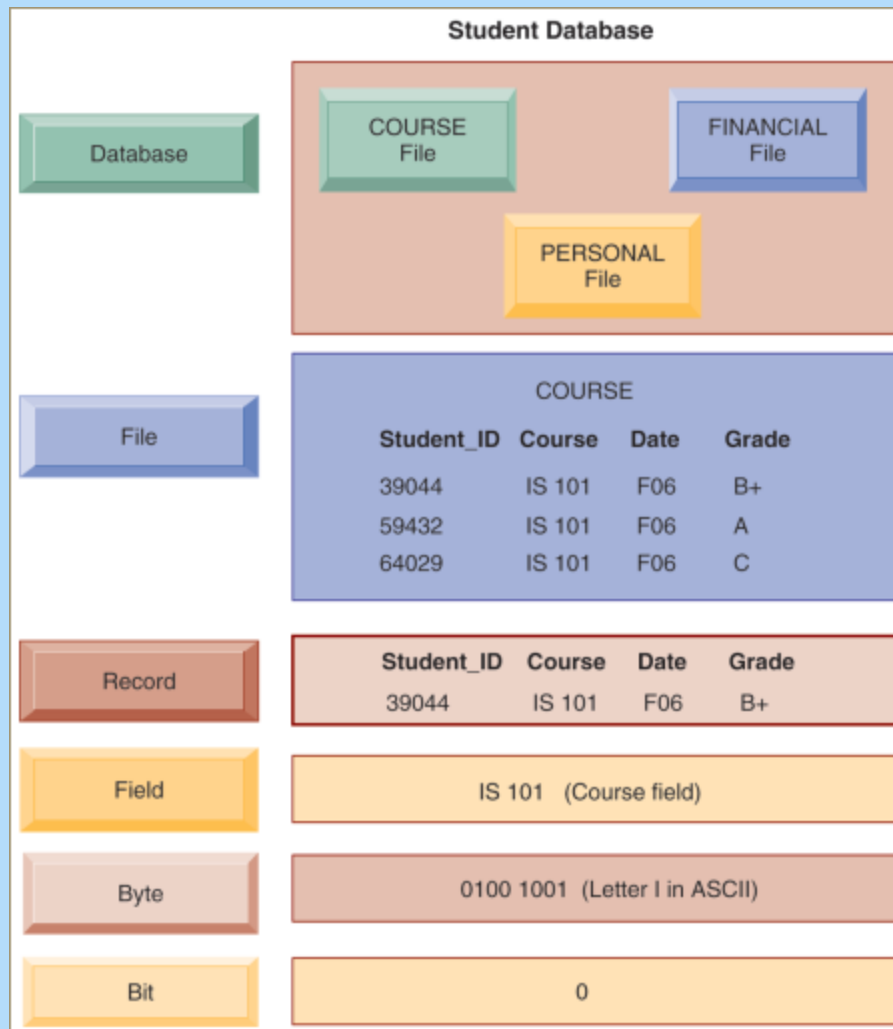


Рис. 5-1

Організація даних у традиційному файловому середовищі

Проблеми з традиційним файловим середовищем (файли зберігаються окремо різними відділами)

- Надмірність даних:
 - ❖ Наявність дублікатів даних у кількох файлах
- Невідповідність даних:
 - ❖ Той самий атрибут має різні значення
- Залежність програми від даних:
 - ❖ Коли зміни в програмі потребують змін у даних, до яких програма має доступ
- Брак гнучкості
- Погана безпека
- Відсутність обміну та доступності даних

Використання традиційного підходу до обробки файлів заохочує кожен функціональний сферу в корпорації розробляти спеціалізовані програми.

Для кожної програми потрібен унікальний файл даних, який, ймовірно, є підмножиною головного файлу.

Ці підмножини головного файлу призводять до надмірності та неузгодженості даних, негнучкості обробки та втрати ресурсів зберігання.

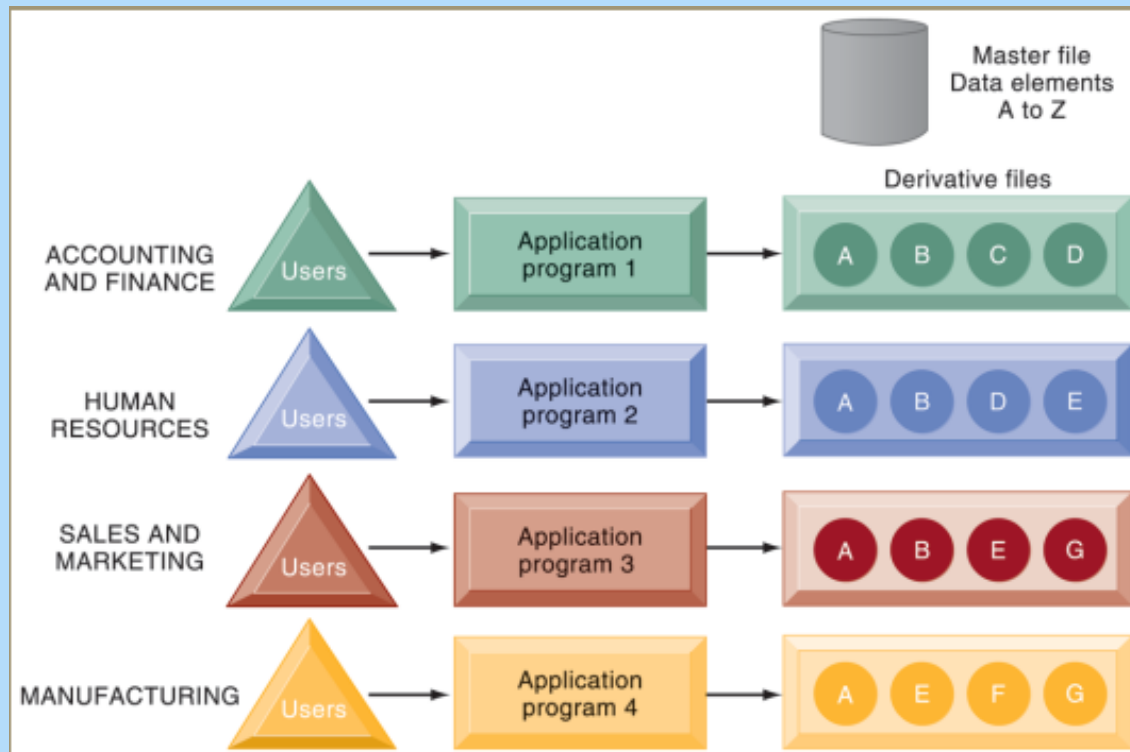


Рис. 5-2

Традиційна обробка файлів

2. Підхід бази даних до управління даними

- База даних
 - ❖ Обслуговує багато програм шляхом централізації даних і контроль надлишкових даних
- Система управління базами даних (СУБД)
 - ❖ Інтерфейси між програмами та фізичними файлами даних
 - ❖ Розділяє логічне та фізичне представлення даних
 - ❖ Вирішує проблеми традиційного файлового середовища
 - Контролює резервування
 - Усуває неузгодженість
 - Роз'єднує програми та дані
 - Дозволяє організації централізовано керувати даними та безпекою даних

Єдина база даних кадрів надає багато різних переглядів даних, залежно від інформаційних потреб користувача. Тут проілюстровано два можливі режими перегляду: один цікавий для фахівця з виплат, а другий — для співробітника відділу оплати праці компанії.

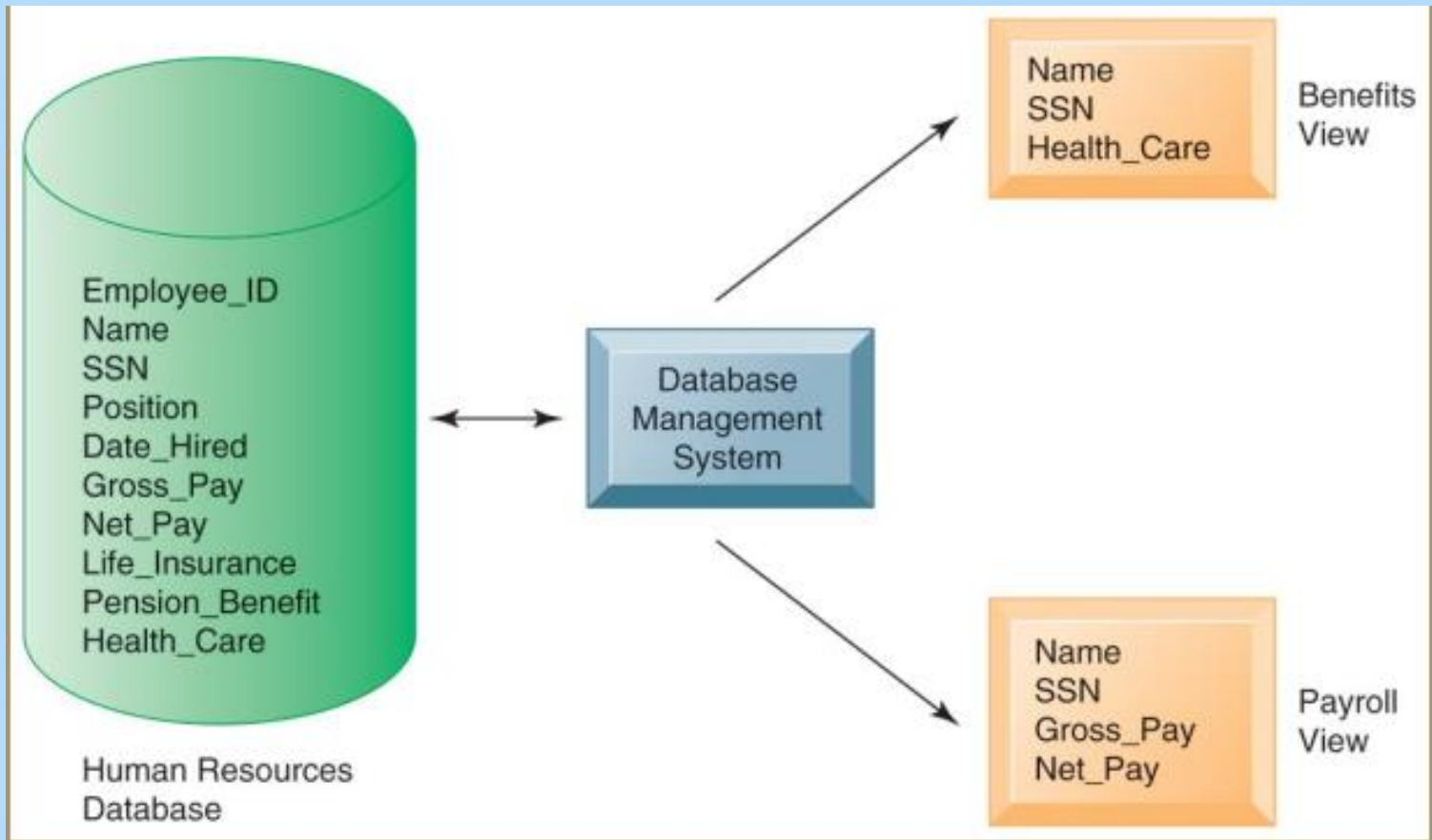


Рис. 5-3

База даних людських ресурсів з кількома програмами

- **Реляційна СУБД**

- ❖ Представлення даних у вигляді двовимірних таблиць, які називаються відношеннями або файли

- ❖ Кожна таблиця містить дані про сутність і атрибути

- **Таблиця: сітка стовпців і рядків**

- ❖ Рядки (кортежі): записи для різних сутностей

- ❖ Поля (стовпці): представляє атрибут для сутності

- ❖ Ключове поле: поле, яке використовується для унікальної ідентифікації кожного запису

- ❖ Первинний ключ: поле в таблиці, яке використовується для ключових полів

- ❖ Зовнішній ключ: первинний ключ, який використовується у другій таблиці як поле пошуку для ідентифікації записів з вихідної таблиці

Реляційна база даних організовує дані у формі двовимірних таблиць. Тут проілюстровано таблиці для сутностей **ПОСТАЧАЛЬНИК** і **ПАРТІЯ**, які показують, як вони представляють кожну сутність та її атрибути. Номер постачальника є первинним ключем для таблиці **SUPPLIER** і зовнішнім ключем для таблиці **PART**.

Columns (Attributes, Fields)					
Supplier_Number	Supplier_Name	Supplier_Street	Supplier_City	Supplier_State	Supplier_Zip
8259	CBM Inc.	74 5 th Avenue	Dayton	OH	45220
8261	B. R. Molds	1277 Gandolly Street	Cleveland	OH	49345
8263	Jackson Composites	8233 Micklin Street	Lexington	KY	56723
8444	Bryant Corporation	4315 Mill Drive	Rochester	NY	11344

Key Field (Primary Key)

Rows (Records, Tuples)

Рис. 5-4

Таблиці реляційної бази даних

Реляційна база даних організовує дані у формі двовимірних таблиць. Тут проілюстровано таблиці для сутностей **ПОСТАЧАЛЬНИК** і **ПАРТІЯ**, які показують, як вони представляють кожну сутність та її атрибути. Номер постачальника є первинним ключем для таблиці **SUPPLIER** і зовнішнім ключем для таблиці **PART**.

PART			
Part_Number	Part_Name	Unit_Price	Supplier_Number
137	Door latch	22.00	8259
145	Side mirror	12.00	8444
150	Door molding	6.00	8263
152	Door lock	31.00	8259
155	Compressor	54.00	8261
178	Door handle	10.00	8259

Primary Key

Foreign Key

Рис. 5-4

Таблиці реляційної бази даних

Операції реляційної СУБД

Три основні операції, які використовуються для створення корисних наборів даних

- ❖ **SELECT**: створює підмножину даних усіх записів, які відповідають зазначеним критеріям
- ❖ **JOIN**: об'єднує реляційні таблиці, щоб надати користувачеві більше інформації, ніж доступно в окремих таблицях
- ❖ **PROJECT**: створює підмножину стовпців у таблиці, створюючи таблиці лише з указаною інформацією

Операції **вибору, об'єднання та проекту** дозволяють об'єднувати дані з двох різних таблиць і відображати лише вибрані атрибути.

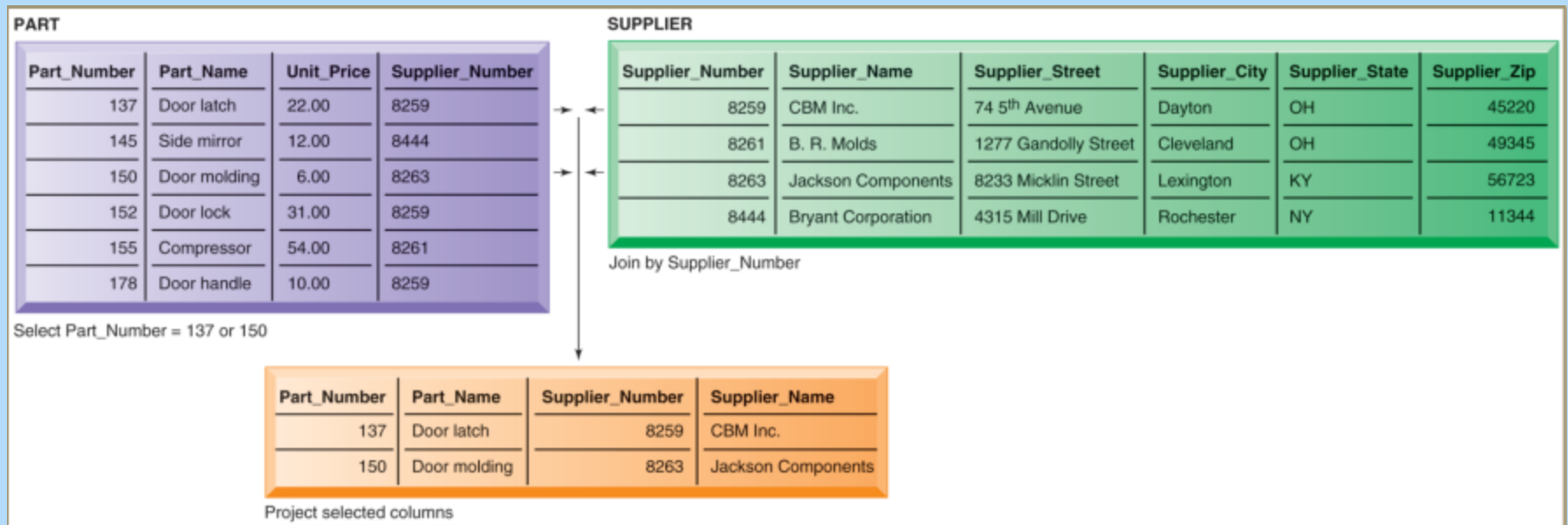


Рис. 5-5

Три основні операції реляційної СУБД

- **Об'єктно-орієнтована СУБД (OODBMS)**

- ❖ Зберігає дані та процедури як об'єкти
- ❖ Об'єктами можуть бути графіка, мультимедіа, Java-аплети
- ❖ Відносно повільно порівняно з реляційною СУБД для обробки великої кількості транзакцій
- ❖ Гібридна об'єктно-реляційна СУБД:
- ❖ забезпечує можливості як OODBMS, так і реляційної СУБД.

- **Бази даних у хмарі**

- ❖ Зазвичай менша функціональність, ніж локальні БД
- ❖ Amazon Web Services, Microsoft SQL Azure

Можливості систем управління базами даних

- **Можливість визначення даних:** визначає структуру вмісту бази даних, яка використовується для створення таблиць і визначення характеристик полів
- **Словник даних:** автоматичний або ручний файл, що зберігає визначення елементів даних та їхні характеристики
- **Мова обробки даних:** використовується для додавання, зміни, видалення, отримання даних із бази даних
 - ❖ Мова структурованих запитів (SQL)
 - ❖ Інструменти користувача Microsoft Access для генерації SQL
- Багато СУБД мають можливості створення звітів для створення відшліфованих звітів (Crystal Reports)

Microsoft Access має елементарну можливість словника даних, який відображає інформацію про розмір, формат та інші характеристики кожного поля в базі даних. Тут відображається інформація, яка зберігається в таблиці SUPPLIER. Маленький значок ключа ліворуч від Supplier_Number вказує на те, що це ключове поле.

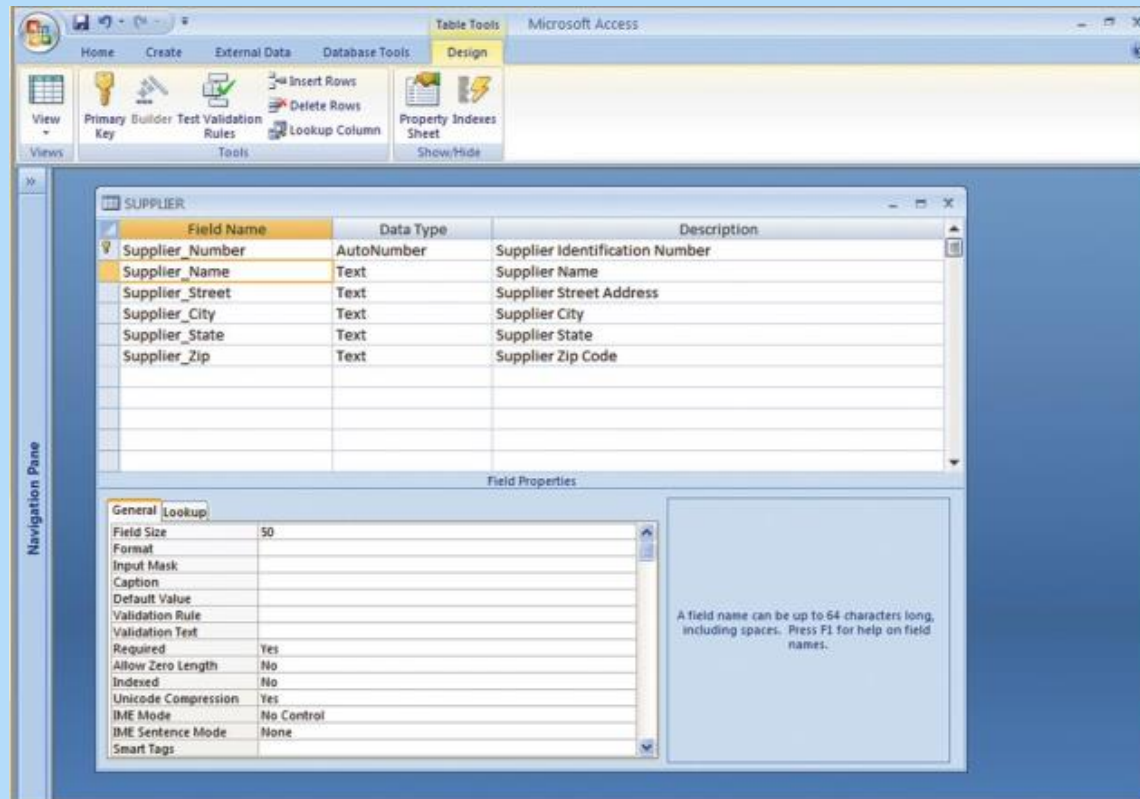


Рис. 5-6

Функції словника даних MICROSOFT ACCESS

На рис. 5-7 проілюстровано оператори SQL для запиту на вибір постачальників для частин 137 або 150. Вони створюють список із такими ж результатами, як на малюнку 5-5.

```
SELECT PART.Part_Number, PART.Part_Name, SUPPLIER.Supplier_Number,  
SUPPLIER.Supplier_Name  
FROM PART, SUPPLIER  
WHERE PART.Supplier_Number = SUPPLIER.Supplier_Number AND  
Part_Number = 137 OR Part_Number = 150;
```

Тут показано, як буде створено запит на малюнку 5-7 за допомогою інструментів створення запитів Microsoft Access. Він показує таблиці, поля та критерії вибору, використані для запиту.

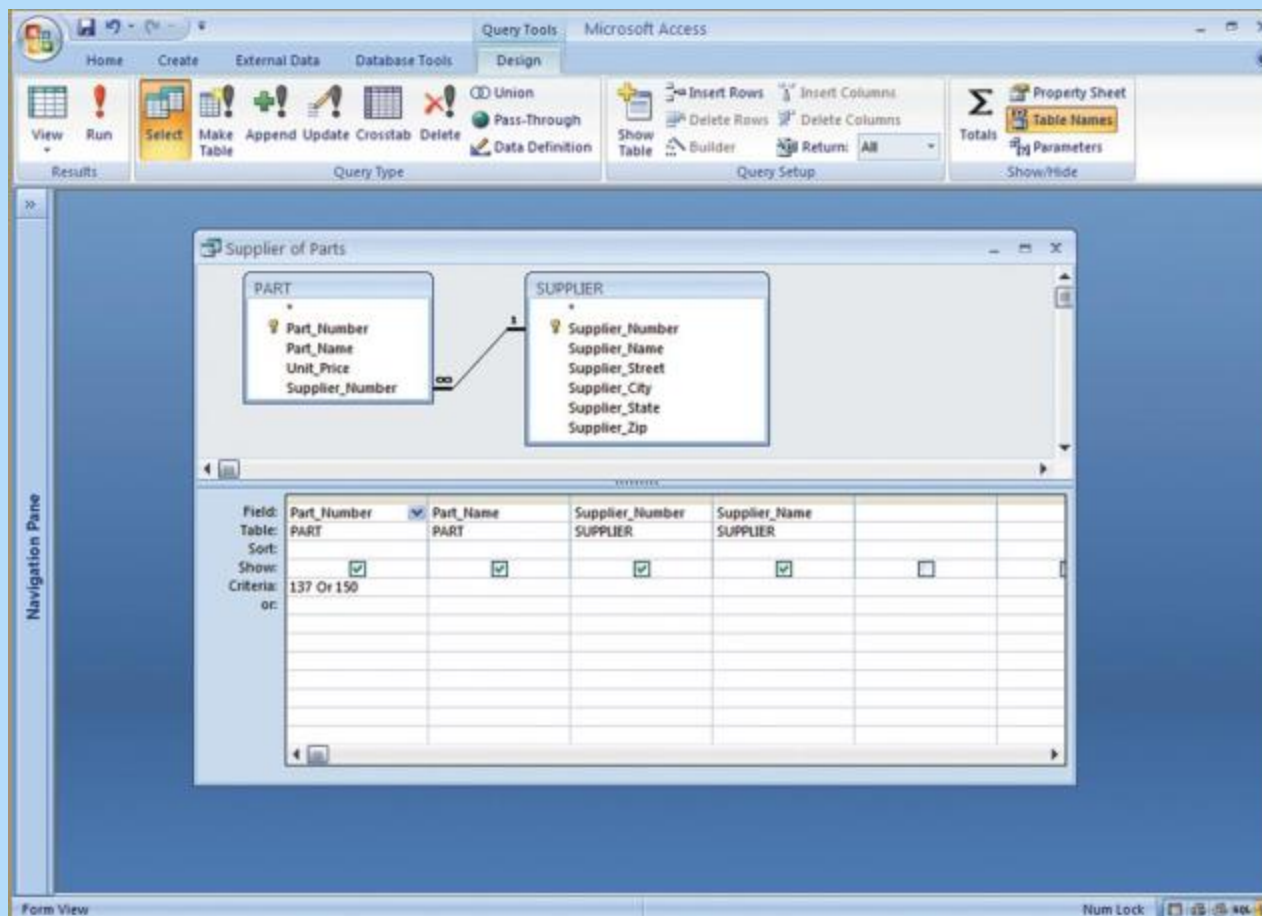


Рис. 5-8

Запит в ACCESS

- ***Проектування баз даних***

- ❖ Концептуальний (логічний) дизайн: абстрактна модель з точки зору бізнесу
- ❖ Фізичний дизайн: як база даних організована на пристроях зберігання даних із прямим доступом

- ***Ідентифікує процес проектування***

- ❖ Зв'язки між елементами даних, надлишковими елементами бази даних
- ❖ Найефективніший спосіб групування елементів даних для задоволення бізнес-вимог, потреб прикладних програм

- ***Нормалізація***

- ❖ Оптимізація складних груп даних для мінімізації надлишкових елементів даних і незручних зв'язків «багато-до-багатьох».

Ненормалізоване відношення містить повторювані групи. Наприклад, для кожного замовлення може бути багато деталей і постачальників. Між Order_Number і Order_Date існує лише однозначна відповідність.

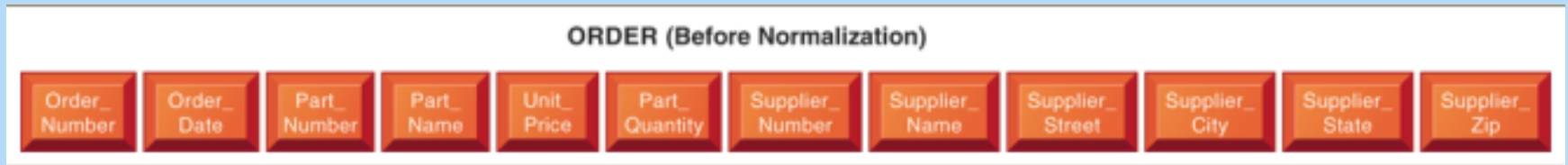


Рис. 5-9 Ненормалізоване відношення для порядку

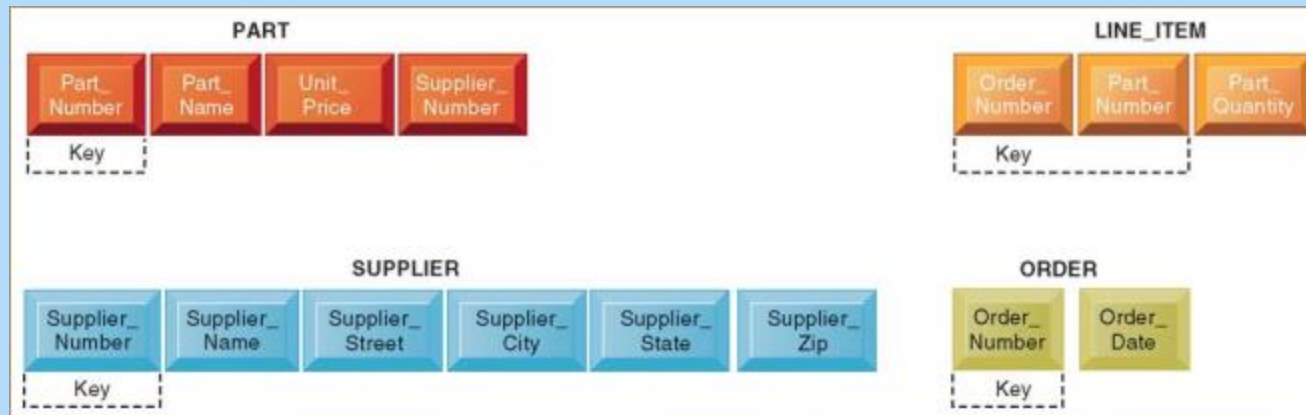


Рис. 5-10 Нормалізовані таблиці, створені на замовлення

❖ Діаграма «сутність-зв'язок»

- Використовується розробниками баз даних для документування моделі даних
- Ілюструє зв'язки між сутностями

❖ Розповсюдження баз даних: зберігання бази даних у кількох місцях

- Розділена: окремі місця зберігають різні частини бази даних
- Реплікація: центральна база даних повністю дублюється в різних місцях

Ця діаграма показує зв'язки між сутностями SUPPLIER, PART, LINE_ITEM і ORDER, які можна використовувати для моделювання бази даних на малюнку 5-10.



Рис. 5-11 Діаграма «сутність-зв'язок»

3. Використання баз даних для покращення ефективності бізнесу та прийняття рішень

❖ «Дуже великі бази даних і системи вимагають спеціальних можливостей, інструментів

- Для аналізу великої кількості даних
- Для доступу до даних з кількох систем

❖ Три ключові прийоми:

- 1. **Data warehousing** - сховище даних
- 2. **Data mining** - видобуток даних
- 3. Інструменти для доступу до внутрішніх баз даних через Інтернет

❖ Data warehouse (сховище даних):

- Зберігає поточні та історичні дані з багатьох основних операційних систем транзакцій
- Консолідує та стандартизує інформацію для використання на підприємстві, але дані не можна змінити
- Система сховища даних забезпечить інструменти запитів, аналізу та звітності

❖ Data marts (вітрини даних):

- Підмножина сховища даних
- Узагальнена або чітко сфокусована частина даних фірми для використання певною групою користувачів
- Зазвичай зосереджується на одній темі чи сфері діяльності

Сховище даних витягує поточні та історичні дані з кількох операційних систем всередині організації. Ці дані поєднуються з даними із зовнішніх джерел і реорганізуються в центральну базу даних, призначену для управлінської звітності та аналізу. Інформаційний довідник надає користувачам інформацію про наявні на складі дані.

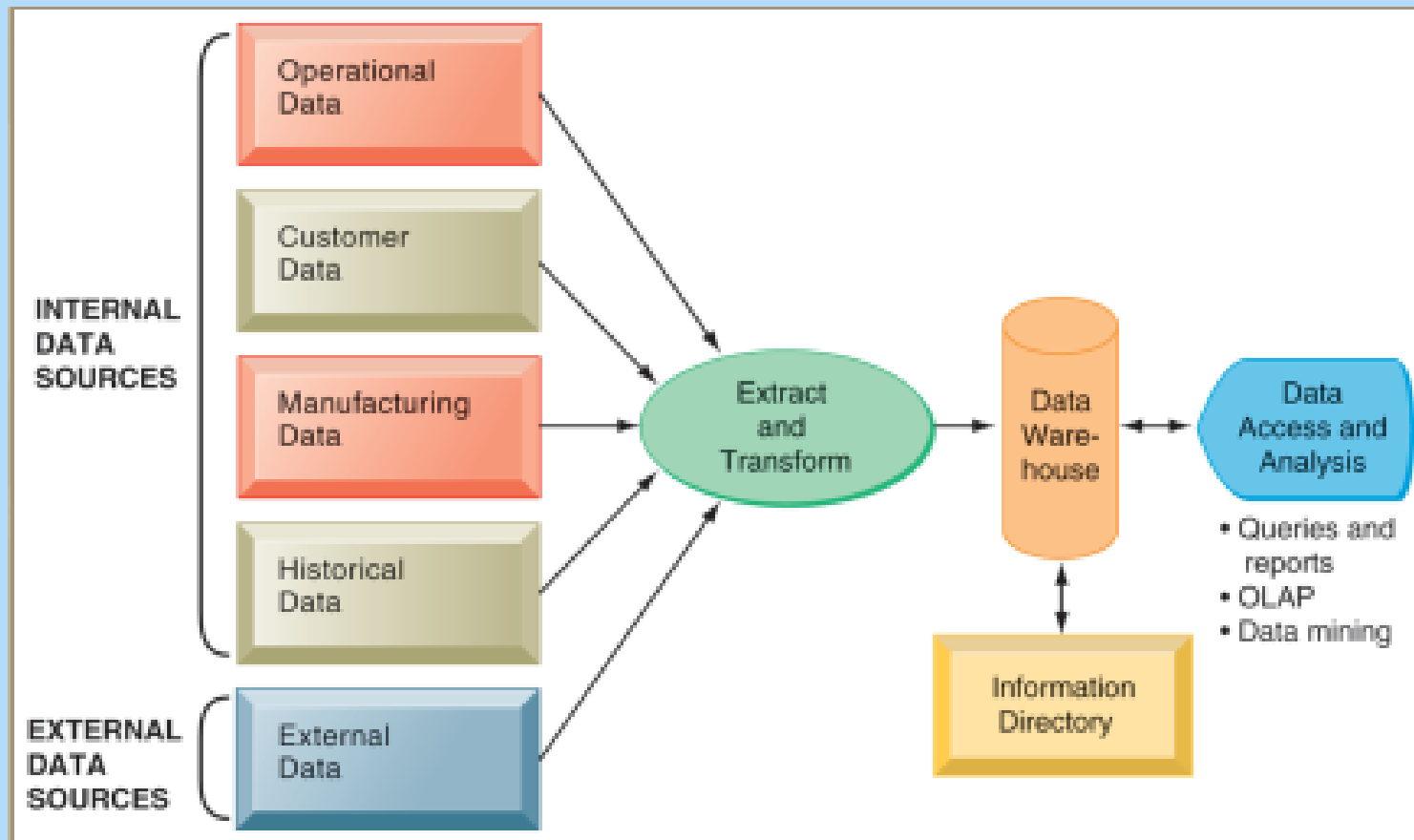


Рис. 5-12

Компоненти сховища даних

Бізнес-аналітика:

- ❖ Інструменти для консолідації, аналізу та надання доступу до величезних обсягів даних, щоб допомогти користувачам приймати кращі бізнес-рішення
- ❖ Наприклад, Harrah's Entertainment аналізує клієнтів, щоб розробити профілі азартних ігор і визначити найприбутковіших клієнтів
- ❖ Основні інструменти включають:
 - Програмне забезпечення для запитів до бази даних і звітності
 - Онлайн аналітична обробка (OLAP)
 - Data mining (видобуток даних)

Онлайн аналітична обробка (OLAP)

- ❖ Підтримує багатовимірний аналіз даних
 - Перегляд даних за допомогою кількох вимірів
 - Кожен аспект інформації (продукт, ціна, вартість, регіон, період часу) має різний вимір
 - Наприклад, скільки пральних машин продали на сході в червні в порівнянні з іншими регіонами?
- ❖ OLAP дозволяє швидко отримувати відповіді в режимі онлайн на спеціальні запити

На рисунку представлено продукт і регіон. Якщо ви повернете куб на 90 градусів, грань, яка буде відображатися, — це продукт у порівнянні з фактичними та прогнозованими продажами. Якщо ви знову повернете куб на 90 градусів, ви побачите регіон у порівнянні з фактичними та прогнозованими продажами. Можливі інші види.

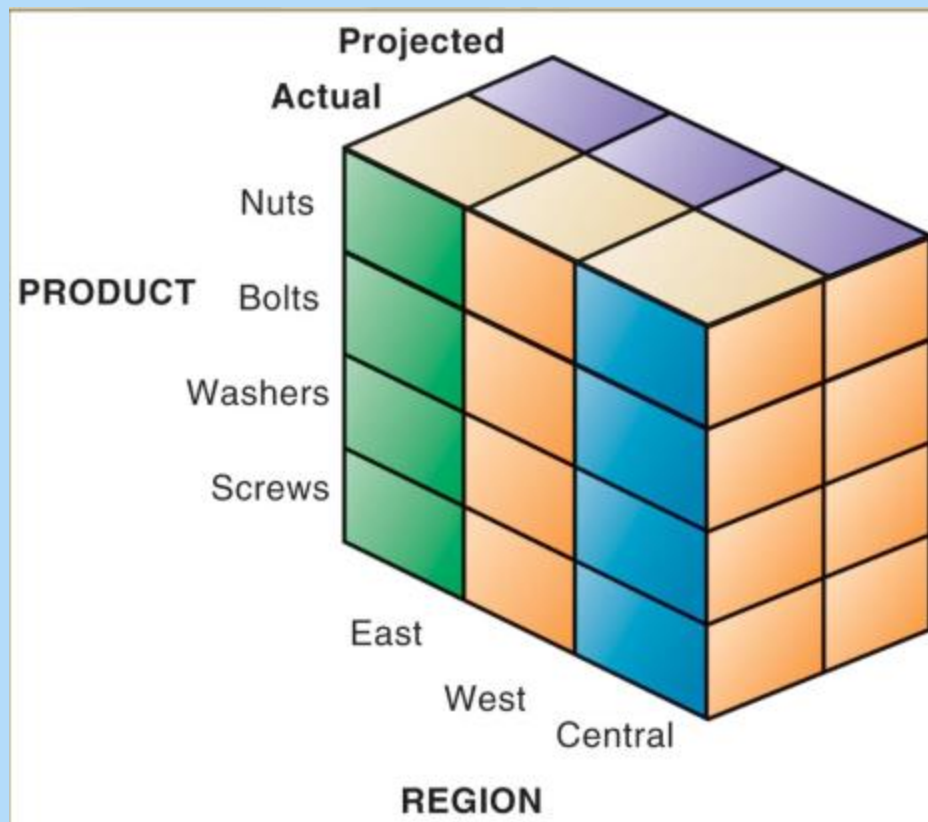


Рис. 5-13

Багатовимірна модель даних

Видобування даних (Data mining)

- ❖ Більш орієнтоване на пошук, ніж OLAP
- ❖ Знаходить приховані шаблони, зв'язки у великих базах даних і виводить правила для передбачення майбутньої поведінки
- ❖ Наприклад, пошук закономірностей у даних клієнта для один-на-один маркетингові кампанії або для виявлення прибуткових клієнтів.
- ❖ Типи інформації, яку можна отримати шляхом аналізу даних
 - Асоціації
 - Послідовності
 - Класифікація
 - Кластеризація
 - Прогнозування

❖ Прогнозний аналіз

- Використовує методи аналізу даних, історичні дані та припущення щодо майбутніх умов для прогнозування результатів подій
- Наприклад, ймовірність того, що клієнт відповість на пропозицію

❖ Видобування тексту

- Витягує ключові елементи з великих неструктурованих наборів даних (наприклад, збережених електронних листів)

Веб-майнінг (Web mining)

- ❖ Виявлення та аналіз корисних шаблонів та інформації з WWW
 - Наприклад, щоб зрозуміти поведінку клієнтів, оцінити ефективність Веб-сайт тощо.
- ❖ Майнінг веб-контенту
 - Знання, отримані з вмісту веб-сторінок
- ❖ Майнінг веб-структури
 - Наприклад, посилання на веб-сторінку та з неї
- ❖ Майнінг веб-використання
 - Дані взаємодії користувача, записані веб-сервером

Бази даних і Інтернет

- ❖ Багато компаній використовують Інтернет, щоб зробити деякі внутрішні бази даних доступними для клієнтів або партнерів
- ❖ Типова конфігурація включає:
 - Веб-сервер
 - Сценарії сервера програм/проміжного програмного забезпечення/CGI
 - Сервер бази даних (хостинг DBM)
- ❖ Переваги використання Web для доступу до бази даних:
 - Простота використання програмного забезпечення браузера
 - Веб-інтерфейс вимагає незначних змін у базі даних або взагалі не вимагає їх
 - Недорого додати веб-інтерфейс до системи

Користувачі отримують доступ до внутрішньої бази даних організації через Інтернет за допомогою своїх настільних ПК і програмного забезпечення веб-браузера.

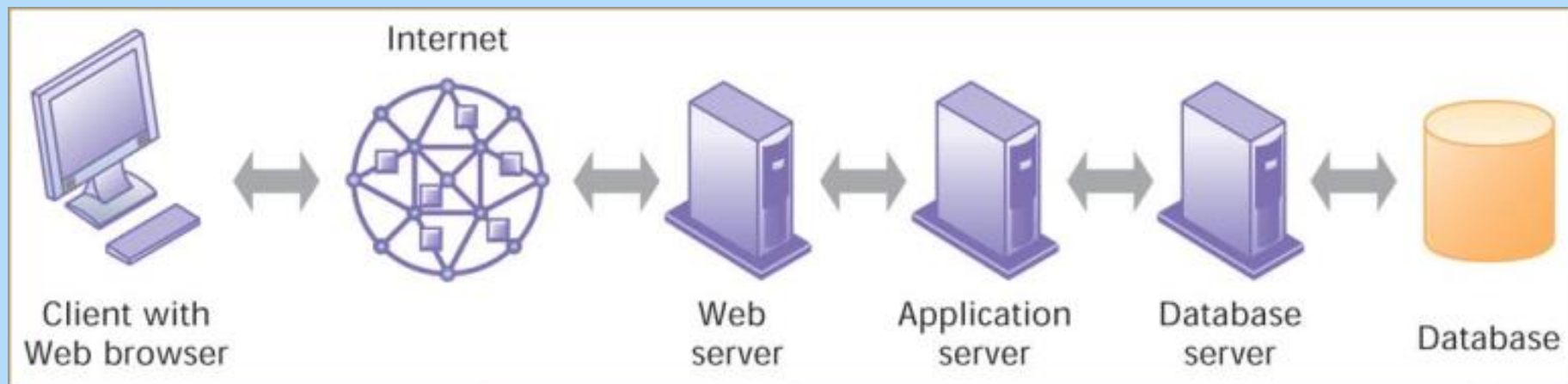


Рис. 5-14

Зв'язок внутрішніх баз даних з мережею

4. Управління базами даних

Формування інформаційної політики

- ❖ Правила фірми, процедури, ролі для спільного використання, управління, стандартизація даних
- ❖ Адміністрування даних:
 - Функція фірми, відповідальна за конкретні політики та процедури керувати даними
- ❖ Керування даними:
 - Політики та процеси для керування доступністю, зручністю використання, цілісністю, і безпека корпоративних даних, особливо в тому, що стосується державних постанов
- ❖ Адміністрування бази даних:
 - Визначення, організація, впровадження, підтримка бази даних; виконується групою проектування та керування базами даних

Забезпечення якості даних

- ❖ Більше 25% критичних даних у базах даних компаній зі списку Fortune 1000 є неточними або неповними
- ❖ Більшість проблем із якістю даних виникають через неправильне введення
- ❖ Перш ніж створити нову базу даних, потрібно:
 - Виявляти та виправляти помилкові дані
 - Встановіть кращі процедури для редагування даних після запуску бази даних

❖ Аудит якості даних:

- Структурований огляд точності та рівня повноти даних в інформаційній системі
 - ✓ • Вибірки опитування з файлів даних, або
 - ✓ • Опитуйте кінцевих користувачів щодо сприйняття якості

❖ Очищення даних

- - Програмне забезпечення для виявлення та виправлення даних, які є неправильними, неповними, неправильно відформатованими або зайвими
- - Забезпечує узгодженість між різними наборами даних з окремих інформаційних систем

Література по курсу:

Laudon K.C. , Laudon J.P. (2020). Management Information Systems: Managing the Digital Firm. 16 ed. Published by Pearson.

Дякую за увагу!