Chapter 5

Foundations of Business Intelligence: Databases and Information Management
Student Learning Objectives

- How does a relational database organize data, and how does it differ from an object-oriented database?
- What are the principles of a database management system?
- What are the principal tools and technologies for accessing information from databases to improve business performance and decision making?
Student Learning Objectives

• What is the role of information policy and data administration in the management of organizational data resources?

• Why is data quality assurance so important for a business?
• **Problem:** explosive growth created information management challenges.

• **Solutions:** use MDM to create an enterprise-wide set of data, preventing unnecessary data duplication.
• **Master data management (MDM)** enables companies like R.R. Donnelley to eliminate outdated, incomplete or incorrectly formatted data.

• Demonstrates IT’s role in successful data management.

• Illustrates digital technology’s role in storing and organizing data.
R.R. Donnelley Tries to Master Its Data

- Centralize data management
- Establish data governance rules and procedures
- Standardize data definitions
- Cleanse and reconcile data
- Design master data file
- Implement Purisma data hub

Business Challenges
- Rapid expansion through acquisitions
- Lack of customer knowledge
- Fragmented customer data

People

Information System
- Consolidate customer and vendor data
- Identify customers and sales opportunities

Organization

Business Solutions
- Increase revenue
- Increase productivity

Technology
• **Database:**
  - Collection of related files containing records on people, places, or things.
  - Prior to digital databases, business used file cabinets with paper files.

• **Entity:**
  - Generalized category representing person, place, thing on which we store and maintain information
  - E.g., SUPPLIER, PART

• **Attributes:**
  - Specific characteristics of each entity:
    - SUPPLIER name, address
    - PART description, unit price, supplier
The Database Approach to Data Management

- **Relational database:**
  - Organize data into two-dimensional tables (relations) with columns and rows.
  - One table for each entity:
    - E.g., (CUSTOMER, SUPPLIER, PART, SALES)
  - **Fields** (columns) store data representing an attribute.
  - Rows store data for separate **records**, or tuples.
  - **Key field:** uniquely identifies each record.
  - **Primary key:**
    - One field in each table
    - Cannot be duplicated
    - Provides unique identifier for all information in any row
A relational database organizes data in the form of two-dimensional tables. Illustrated here is a table for the entity SUPPLIER showing how it represents the entity and its attributes. Supplier_Number is the key field.

Figure 5-1
Data for the entity PART have their own separate table. Part_Number is the primary key and Supplier_Number is the foreign key, enabling users to find related information from the SUPPLIER table about the supplier for each part.

**The PART Table**

<table>
<thead>
<tr>
<th>Part_Number</th>
<th>Part_Name</th>
<th>Unit_Price</th>
<th>Supplier_Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>137</td>
<td>Door latch</td>
<td>22.00</td>
<td>8259</td>
</tr>
<tr>
<td>145</td>
<td>Side mirror</td>
<td>12.00</td>
<td>8444</td>
</tr>
<tr>
<td>150</td>
<td>Door molding</td>
<td>6.00</td>
<td>8263</td>
</tr>
<tr>
<td>152</td>
<td>Door lock</td>
<td>31.00</td>
<td>8259</td>
</tr>
<tr>
<td>155</td>
<td>Compressor</td>
<td>54.00</td>
<td>8261</td>
</tr>
<tr>
<td>178</td>
<td>Door handle</td>
<td>10.00</td>
<td>8259</td>
</tr>
</tbody>
</table>
• Establishing relationships
  • Entity-relationship diagram
    • Used to clarify table relationships in a relational database
  • Relational database tables may have:
    • One-to-one relationship
    • One-to-many relationship
    • Many-to-many relationship
      • Requires creating a table (join table, Intersection relation) that links the two tables to join information
This diagram shows the relationship between the entities SUPPLIER and PART.

**Figure 5-3**
The Database Approach to Data Management

• **Normalization**
  - Process of streamlining complex groups of data to:
    - Minimize redundant data elements.
    - Minimize awkward many-to-many relationships.
    - Increase stability and flexibility.

• **Referential integrity rules**
  - Used by relational databases to ensure that relationships between coupled tables remain consistent.
  - E.g., when one table has a foreign key that points to another table, you may not add a record to the table with foreign key unless there is a corresponding record in the linked table.
The shaded areas show which data came from the SUPPLIER, LINE_ITEM, and ORDER tables. The database does not maintain data on Extended Price or Order Total because they can be derived from other data in the tables.

**Figure 5-4**
The final design of the database for suppliers, parts, and orders has four tables. The LINE_ITEM table is a join table that eliminates the many-to-many relationship between ORDER and PART.
Entity-Relationship Diagram for the Database with Four Tables

This diagram shows the relationship between the entities SUPPLIER, ART, LINE_ITEM, and ORDER.

Figure 5-6
DBMS

- Specific type of software for creating, storing, organizing, and accessing data from a database
- Separates the logical and physical views of the data
- **Logical view**: how end users view data
- **Physical view**: how data are actually structured and organized
- **Examples of DBMS**: Microsoft Access, DB2, Oracle Database, Microsoft SQL Server, MySQL
A single human resources database provides many different views of data, depending on the information requirements of the user. Illustrated here are two possible views, one of interest to a benefits specialist and one of interest to a member of the company’s payroll department.
Operations of a Relational DBMS

- **Select:**
  - Creates a subset of all records meeting stated criteria

- **Join:**
  - Combines relational tables to present the server with more information than is available from individual tables

- **Project:**
  - Creates a subset consisting of columns in a table
  - Permits user to create new tables containing only desired information
The Three Basic Operations of a Relational DBMS

The select, project, and join operations enable data from two different tables to be combined and only selected attributes to be displayed.
Capabilities of Database Management Systems

• Data definition capabilities:
  • Specify structure of content of database.

• Data dictionary:
  • Automated or manual file storing definitions of data elements and their characteristics.

• Querying and reporting:
  • Data manipulation language
    • Structured query language (SQL)
    • Microsoft Access query-building tools
  • Report generation, e.g., Crystal Reports
Microsoft Access has a rudimentary data dictionary capability that displays information about the size, format, and other characteristics of each field in a database. Displayed here is the information maintained in the SUPPLIER table. The small key icon to the left of Supplier_Number indicates that it is a key field.

Figure 5-9
Example of an SQL Query

```
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;
```

Illustrated here are the SQL statements for a query to select suppliers for parts 137 or 150. They produce a list with the same results as Figure 5-8.

Figure 5-10
Illustrated here is how the query in Figure 5-10 would be constructed using Microsoft Access query-building tools. It shows the tables, fields, and selection criteria used for the query.

Figure 5-11
Object-Oriented DBMS (OODBMS)

- Stores data and procedures that act on those data as objects to be retrieved and shared
- Better suited for storing graphic objects, drawings, video, than DBMS designed for structuring data only
- Used to manage multimedia components or Java applets in Web applications
- Relatively slow compared to relational DBMS
- Object-relational DBMS: provide capabilities of both types
• Databases provide information to help the company run the business more efficiently, and help managers and employees make better decisions.

• Tools for analyzing, accessing vast quantities of data:
  • Data warehousing
  • Multidimensional data analysis
  • Data mining
  • Utilizing Web interfaces to databases
Data Warehouses

- **Data warehouse:**
  - Database that stores current and historical data that may be of interest to decision makers
  - Consolidates and standardizes data from many systems, operational and transactional databases
  - Data can be accessed but not altered

- **Data mart:**
  - Subset of data warehouses that is highly focused and isolated for a specific population of users
Components of a Data Warehouse

The data warehouse extracts current and historical data from multiple operational systems inside the organization. These data are combined with data from external sources and reorganized into a central database designed for management reporting and analysis. The information directory provides users with information about the data available in the warehouse.

Figure 5-12
Business Intelligence, Multidimensional Data Analysis, and Data Mining

- **Business intelligence**: tools for consolidating, analyzing, and providing access to large amounts of data to improve decision making
  - Software for database reporting and querying
  - Tools for multidimensional data analysis (online analytical processing)
  - Data mining
- E.g., Harrah’s Entertainment gathers and analyzes customer data to create gambling profile and identify most profitable customers
A series of analytical tools works with data stored in databases to find patterns and insights for helping managers and employees make better decisions to improve organizational performance.

Figure 5-13
Online Analytical Processing (OLAP)

- Supports multidimensional data analysis, enabling users to view the same data in different ways using multiple dimensions
  - Each aspect of information—product, pricing, cost, region, or time period—represents a different dimension
  - E.g., comparing sales in East in June versus May and July
- Enables users to obtain online answers to ad hoc questions such as these in a fairly rapid amount of time
The view that is showing is product versus region. If you rotate the cube 90 degrees, the face that will show is product versus actual and projected sales. If you rotate the cube 90 degrees again, you will see region versus actual and projected sales. Other views are possible.

Figure 5-14
Data Mining

- Finds hidden patterns and relationships in large databases and infers rules from them to predict future behavior
- Types of information obtainable from data mining
  - Associations: occurrences linked to single event
  - Sequences: events linked over time
  - Classifications: patterns describing a group an item belongs to
  - Clusters: discovering as yet unclassified groupings
  - Forecasting: uses series of values to forecast future values
Data Mining

• **One popular use of data mining:** analyzing patterns in customer data for one-to-one marketing campaigns or for identifying profitable customers

• **Predictive analysis:**
  - Uses data mining techniques, historical data, and assumptions about future conditions to predict outcomes of events, such as the probability a customer will respond to an offer or purchase a specific product

• **Data mining versus privacy concerns**
  - Used to create detailed data image about each individual
• **Text Mining**
  - Unstructured data (mostly text files) accounts for 80 percent of an organization’s useful information.
  - Text mining allows businesses to extract key elements from, discover patterns in, and summarize large unstructured data sets.

• **Web Mining**
  - Discovery and analysis of useful patterns and information from the Web
  - Content mining, structure mining, usage mining
Databases and the Web

- Firms use the Web to make information from their internal databases available to customers and partners.

- Middleware and other software make this possible
  - Web server
  - Application servers or CGI
  - Database server

- Web interfaces provide familiarity to users and savings over redesigning and rebuilding legacy systems.
Users access an organization’s internal database through the Web using their desktop PCs and Web browser software.

Figure 5-15
Establishing an Information Policy

- **Information policy**
  - States organization’s rules for organizing, managing, storing, sharing information

- **Data administration**
  - Responsible for specific policies and procedures through which data can be managed as a resource

- **Database administration**
  - Database design and management group responsible for defining and organizing the structure and content of the database, and maintaining the database.
Ensuring Data Quality

- **Poor data quality**: major obstacle to successful customer relationship management
- **Data quality problems**: caused by
  - Redundant and inconsistent data produced by multiple systems
  - Data input errors
- **Data quality audit**: structured survey of the accuracy and completeness of data
- **Data cleansing**: detects and corrects incorrect, incomplete, improperly formatted, and redundant data