# Chapter 4:
# Databases and Data Warehouses

# Learning objectives

1. Information resources
2. Database advantages
3. Relational database
4. Master data management
5. Data warehouse
6. Information management

# Information resources

- Structured information
- Unstructured information
- Semi-structured information
- Metadata

# Quality of information

- Accuracy
- Precision
- Completeness
- Consistency
- Timeliness
- Bias
- Duplication

# Managing information

- Record
- Field
- Table

| Department | Object Code | Amount | Category | Description |
|---|---|---|---|---|
| Sales | 4211 | 1888.25 | Computers | Desktop Computers |
| Sales | 4300 | 249.95 | Computer supplies | Image editing software |
| Sales | 4100 | 29.99 | Office supplies | Flash drive |
| Personnel | 4211 | 59.00 | Computers | Stastical software |
| Personnel | 4300 | 14.95 | Computer supplies | Flash drive |
| Personnel | 4211 | 2500.21 | Computers | Laptop Computers |
| Warehouse | 4211 | 59500.00 | Computers | Web server |
| Warehouse | 4211 | 2500.00 | Computers | Printer/copier/scanner/fax |

FIGURE 4-9

# File processing systems

- Redundancy and inconsistency
- Lack of integration
- Inconsistent definitions
- Dependence

# Databases

- Reduced redundancy
- Integrity and accuracy
- Ability to adapt to changes
- Performance and scalability
- Security

# Database architecture

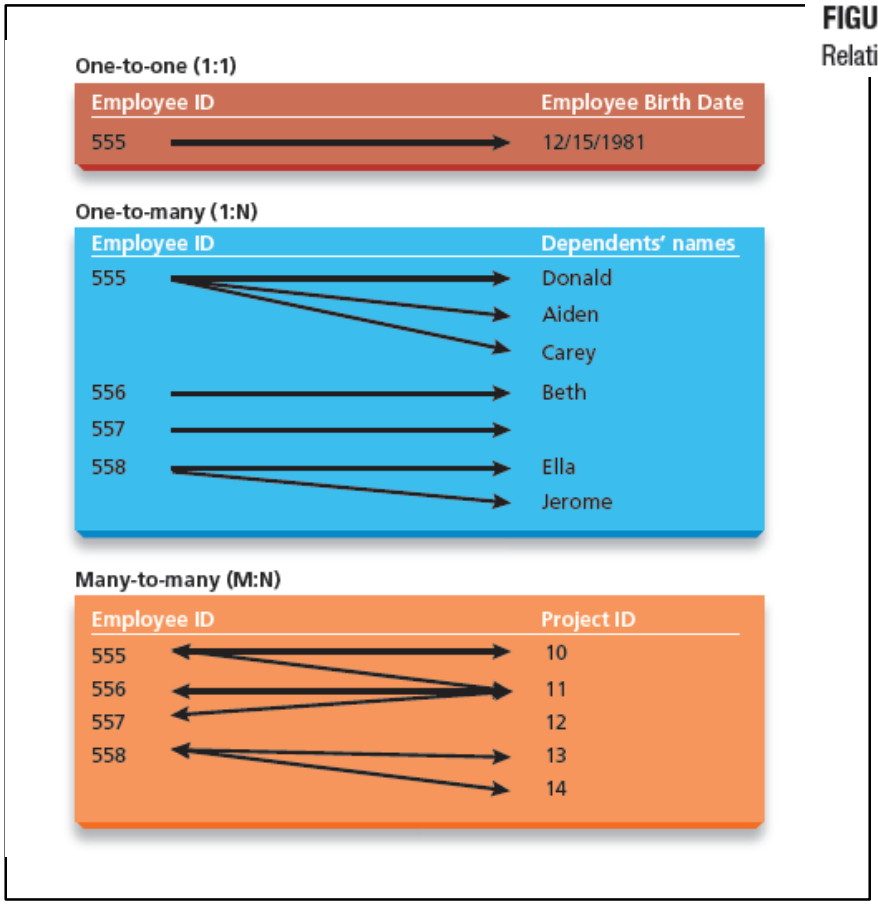



**FIGURE 4-10**
Relationship types.

- One to one (1:1)
- One to many (1:M)
- Many to many (M:M)

# Relational database

- Tables of records
- Link field in one table to field in another table
- Separates data from paths to retrieve data

**Students**

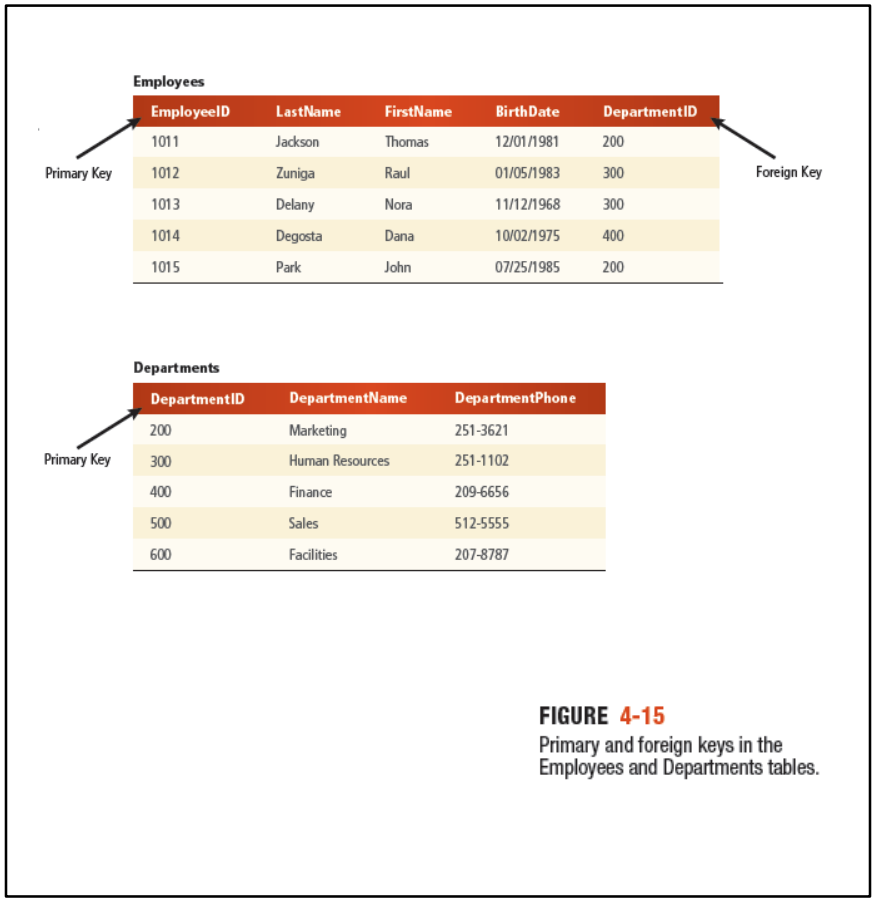
| StudentID | LastName | FirstName | BirthDate |
|---|---|---|---|
| 54001 | Chong | Kevin | 12/01/1987 |
| 65222 | Danelli | Douglas | 01/05/1986 |
| 54555 | Burton | Stephanie | 11/12/1978 |
| 25553 | Washington | Nikia | 10/02/1981 |
| 96887 | Perez | Louis | 07/25/1982 |

**Registrations**

| RegistrationNumber | StudentID | ClassCode | Grade |
|---|---|---|---|
| 10011 | 54001 | 20083BMGT300A | A |
| 10012 | 54001 | 20083HIST450B | C |
| 10013 | 54001 | 20083ECON200F | B |
| 10014 | 54555 | 20083ECON200F | A |
| 10015 | 96887 | 20083HIST410B | I |

**FIGURE 4-14**
Relational database.

# Data model (1:2)



Employees

| EmployeeID | LastName | FirstName | BirthDate | DepartmentID |
|---|---|---|---|---|
| 1011 | Jackson | Thomas | 12/01/1981 | 200 |
| 1012 | Zuniga | Raul | 01/05/1983 | 300 |
| 1013 | Delany | Nora | 11/12/1968 | 300 |
| 1014 | Degosta | Dana | 10/02/1975 | 400 |
| 1015 | Park | John | 07/25/1985 | 200 |

Primary Key

Foreign Key

Departments

| DepartmentID | DepartmentName | DepartmentPhone |
|---|---|---|
| 200 | Marketing | 251-3621 |
| 300 | Human Resources | 251-1102 |
| 400 | Finance | 209-6656 |
| 500 | Sales | 512-5555 |
| 600 | Facilities | 207-8787 |

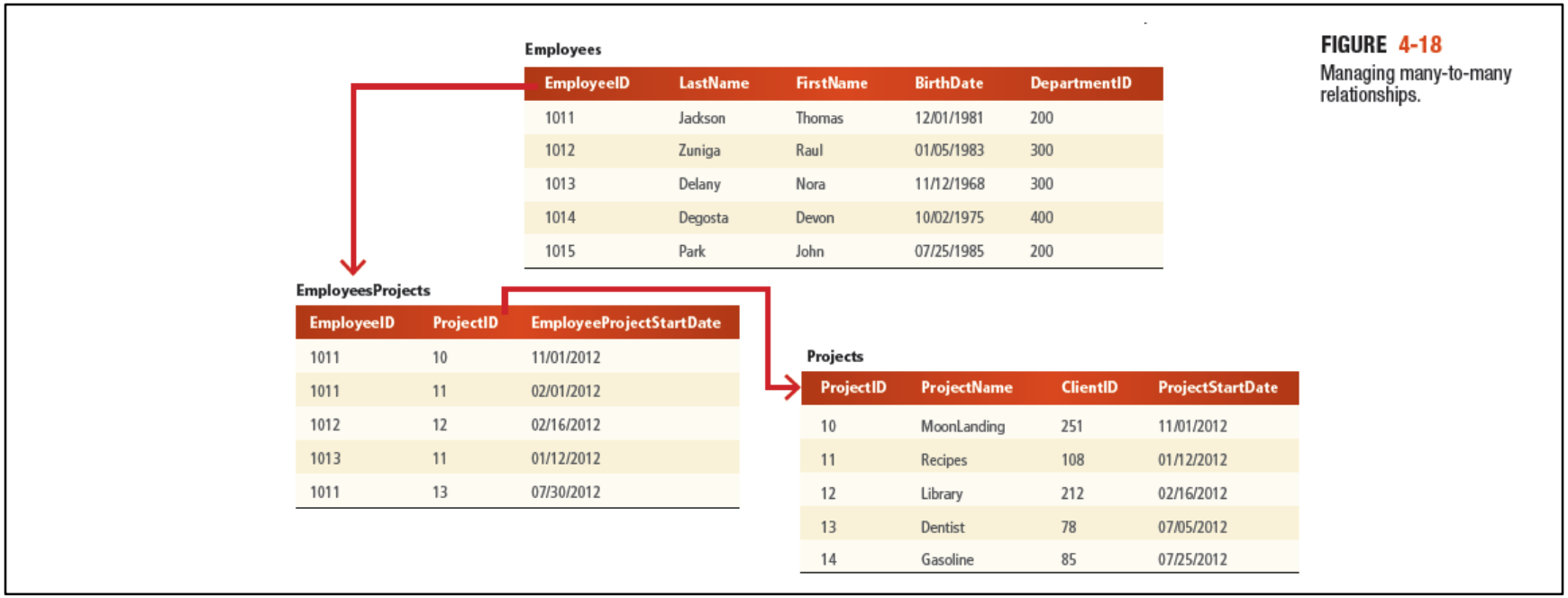
Primary Key

**FIGURE 4-15**
Primary and foreign keys in the Employees and Departments tables.

- Entities and attributes
- Primary key
- Normalization

# Data model (2:2)

- Relationships and foreign keys
- Complex relationships



Employees

| EmployeeID | LastName | FirstName | BirthDate | DepartmentID |
|---|---|---|---|---|
| 1011 | Jackson | Thomas | 12/01/1981 | 200 |
| 1012 | Zuniga | Raul | 01/05/1983 | 300 |
| 1013 | Delany | Nora | 11/12/1968 | 300 |
| 1014 | Degosta | Devon | 10/02/1975 | 400 |
| 1015 | Park | John | 07/25/1985 | 200 |

EmployeesProjects

| EmployeeID | ProjectID | EmployeeProjectStartDate |
|---|---|---|
| 1011 | 10 | 11/01/2012 |
| 1011 | 11 | 02/01/2012 |
| 1012 | 12 | 02/16/2012 |
| 1013 | 11 | 01/12/2012 |
| 1011 | 13 | 07/30/2012 |

Projects

| ProjectID | ProjectName | ClientID | ProjectStartDate |
|---|---|---|---|
| 10 | MoonLanding | 251 | 11/01/2012 |
| 11 | Recipes | 108 | 01/12/2012 |
| 12 | Library | 212 | 02/16/2012 |
| 13 | Dentist | 78 | 07/05/2012 |
| 14 | Gasoline | 85 | 07/25/2012 |

**FIGURE 4-18**
Managing many-to-many relationships.

# Retrieving information

- Structured query language (SQL)
- Interactive voice response (IVR)
- Natural language interfaces

# Managing the database

- Performance tuning and scalability
- Integrity, security, and recovery
- Documentation

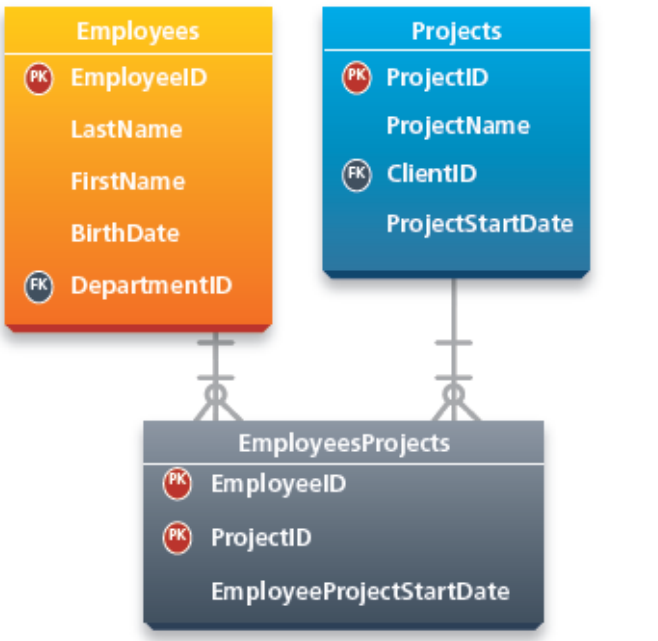



**FIGURE 4-21**
Sample database schema.

# Multiple databases

- Integration challenges
- Shadow systems
- Master data management
- Data stewards

# Data warehouses

- Building data warehouses
- Extract, transform, and load (ETL)
- Data mining

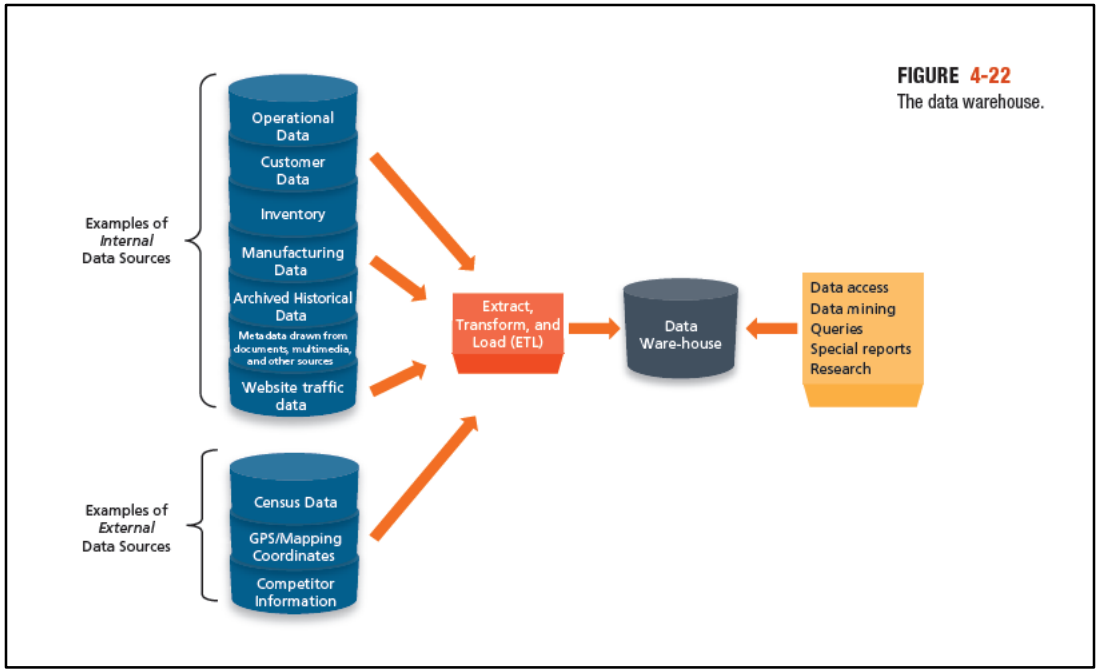



**FIGURE 4-22**
The data warehouse.

# Human element

- Ownership issues
- Databases without boundaries
- Stakeholders

# Summary

1. Information resources
2. Database advantages
3. Relational database
4. Master data management
5. Data warehouse
6. Information management

# UK police case

- Video surveillance
- Automatic plate number recognition
- Database
- Queries and data mining
- Privacy

**FIGURE 4-29**
Capturing license plate numbers for law enforcement.

*Source:* Ann Cantelow/Shutterstock.

# Colgate Palmolive case

- $15 billion sales, 70 countries
- Consistency in products and data
- Colgate Business Planning (CBP)—profit, loss and ROI by product, region, and retailer
- Reinvested $100 million in most profitable promotions, goal $300 million