

GLOBAL
EDITION



Modern Database Management

THIRTEENTH EDITION

Jeffrey A. Hoffer
V. Ramesh
Heikki Topi



THIRTEENTH EDITION
GLOBAL EDITION

MODERN DATABASE MANAGEMENT

Jeffrey A. Hoffer
University of Dayton

V. Ramesh
Indiana University

Heikki Topi
Bentley University



Harlow, England • London • New York • Boston • San Francisco • Toronto • Sydney • Dubai • Singapore • Hong Kong
Tokyo • Seoul • Taipei • New Delhi • Cape Town • Sao Paulo • Mexico City • Madrid • Amsterdam • Munich • Paris • Milan

Modern Database Management, Global Edition

Table of Contents

Cover

Title Page

Copyright Page

Brief Contents

Contents

Preface

Acknowledgments

Preface

Part I: The Context of Database Management

An Overview of Part I

Chapter 1: The Database Environment and Development Process

Learning Objectives

Data Matter!

Introduction

Basic Concepts and Definitions

Data

Data versus Information

Metadata

Traditional File Processing Systems

File Processing Systems at Pine Valley Furniture Company

Disadvantages of File Processing Systems

Program-Data Dependence

Duplication of Data

Limited Data Sharing

Lengthy Development Times

Excessive Program Maintenance

The Database Approach

Data Models

Entities

Relationships

Relational Databases

Database Management Systems

Advantages of the Database Approach

Table of Contents

Program-Data Independence
Planned Data Redundancy
Improved Data Consistency
Improved Data Sharing
Increased Productivity of Application Development
Enforcement of Standards
Improved Data Quality
Improved Data Accessibility and Responsiveness
Reduced Program Maintenance
Improved Decision Support
Cautions about Database Benefits
Costs and Risks of the Database Approach
New, Specialized Personnel
Installation and Management Cost and Complexity
Conversion Costs
Need for Explicit Backup and Recovery
Organizational Conflict

Integrated Data Management Framework

Components of the Database Environment

The Database Development Process

Systems Development Life Cycle

PlanningEnterprise Modeling
PlanningConceptual Data Modeling
AnalysisConceptual Data Modeling
DesignLogical Database Design
DesignPhysical Database Design and Definition
ImplementationDatabase Implementation
MaintenanceDatabase Maintenance

Alternative Information Systems Development Approaches

Three-Schema Architecture for Database Development

Managing the People Involved in Database Development

Evolution of Database Systems

1960s

1970s

1980s

1990s

2000 and Beyond

The Range of Database Applications

Personal Databases

Departmental Multi-Tiered Client/Server Databases

Enterprise Applications

Enterprise Systems

Data Warehouses

Data Lake

Developing a Database Application for Pine Valley Furniture Company

Table of Contents

Database Evolution at Pine Valley Furniture Company

Project Planning

Analyzing Database Requirements

Designing the Database

Using the Database

Administering the Database

Future of Databases at Pine Valley

Summary

Key Terms

Review Questions

Problems and Exercises

Field Exercises

References

Further Reading

Web Resources

Case: Forondo Artist Management Excellence Inc.

Part II: Database Analysis and Logical Design

An Overview of Part II

Chapter 2: Modeling Data in the Organization

Learning Objectives

Introduction

The E-R Model: An Overview

Sample E-R Diagram

E-R Model Notation

Modeling the Rules of the Organization

Overview of Business Rules

The Business Rules Paradigm

Scope of Business Rules

Good Business Rules

Gathering Business Rules

Data Names and Definitions

Data Names

Data Definitions

Good Data Definitions

Modeling Entities and Attributes

Entities

Entity Type versus Entity Instance

Entity Type versus System Input, Output, or User

Strong versus Weak Entity Types

Naming and Defining Entity Types

Attributes

Table of Contents

- Required versus Optional Attributes
- Simple versus Composite Attributes
- Single-valued versus Multivalued Attributes
- Stored versus Derived Attributes
- Identifier Attribute
- Naming and Defining Attributes

Modeling Relationships

Basic Concepts and Definitions in Relationships

- Attributes on Relationships
- Associative Entities

Degree of a Relationship

- Unary Relationship
- Binary Relationship
- Ternary Relationship

Attributes or Entity?

Cardinality Constraints

- Minimum Cardinality
- Maximum Cardinality

Some Examples of Relationships and Their Cardinalities

- A Ternary Relationship

Modeling Time-Dependent Data

Modeling Multiple Relationships Between Entity Types

Naming and Defining Relationships

E-R Modeling Example: Pine Valley Furniture Company

Database Processing At Pine Valley Furniture

- Showing Product Information
- Showing Product Line Information
- Showing Customer Order Status
- Showing Product Sales

Summary

Key Terms

Review Questions

Problems and Exercises

Field Exercises

References

Further Reading

Web Resources

Case: Forondo Artist Management Excellence Inc.

Chapter 3: The Enhanced E-R Model

Learning Objectives

Introduction

Table of Contents

Representing Supertypes and Subtypes

Basic Concepts and Notation

An Example of a Supertype/Subtype Relationship

Attribute Inheritance

When to Use Supertype/Subtype Relationships

Representing Specialization and Generalization

Generalization

Specialization

Combining Specialization and Generalization

Specifying Constraints in Supertype/Subtype Relationships

Specifying Completeness Constraints

Total Specialization Rule

Partial Specialization Rule

Specifying Disjointness Constraints

Disjoint Rule

Overlap Rule

Defining Subtype Discriminators

Disjoint Subtypes

Overlapping Subtypes

Defining Supertype/Subtype Hierarchies

An Example of a Supertype/Subtype Hierarchy

Summary of Supertype/Subtype Hierarchies

EER Modeling Example: Pine Valley Furniture Company

Entity Clustering

Packaged Data Models

A Revised Data Modeling Process with Packaged Data Models

Packaged Data Model Examples

Summary

Key Terms

Review Questions

Problems and Exercises

Field Exercises

References

Further Reading

Web Resources

Case: Forondo Artist Management Excellence Inc.

Chapter 4: Logical Database Design and the Relational Model

Learning Objectives

Introduction

The Relational Data Model

Basic Definitions

Table of Contents

Relational Data Structure	
Relational Keys	
Properties of Relations	
Removing Multivalued Attributes from Tables	
Sample Database	
Integrity Constraints	
Domain Constraints	
Entity Integrity	
Referential Integrity	
Creating Relational Tables	
Well-Structured Relations	
Transforming EER Diagrams into Relations	
Step 1: Map Regular Entities	
Composite Attributes	
Multivalued Attributes	
Step 2: Map Weak Entities	
When to Create a Surrogate Key	
Step 3: Map Binary Relationships	
Map Binary One-to-Many Relationships	
Map Binary Many-to-Many Relationships	
Map Binary One-to-One Relationships	
Step 4: Map Associative Entities	
Identifier not Assigned	
Identifier Assigned	
Step 5: Map Unary Relationships	
Unary One-to-Many Relationships	
Unary Many-to-Many Relationships	
Step 6: Map Ternary (and n-ary) Relationships	
Step 7: Map Supertype/Subtype Relationships	
Summary of EER-to-Relational Transformations	
Introduction to Normalization	
Steps in Normalization	
Functional Dependencies and Keys	
Determinants	
Candidate Keys	
Normalization Example: Pine Valley Furniture Company	
Step 0: Represent the View in Tabular Form	
Step 1: Convert to First Normal Form	
Remove Repeating Groups	
Select the Primary Key	
Anomalies in 1NF	
Step 2: Convert to Second Normal Form	
Step 3: Convert to Third Normal Form	
Removing Transitive Dependencies	

Table of Contents

Determinants and Normalization

Step 4: Further Normalization

Merging Relations

An Example

View Integration Problems

Synonyms

Homonyms

Transitive Dependencies

Supertype/Subtype Relationships

A Final Step for Defining Relational Keys

Summary

Key Terms

Review Questions

Problems and Exercises

Field Exercises

References

Further Reading

Web Resources

Case: Forondo Artist Management Excellence Inc.

Part III: Database Implementation and Use

An Overview of Part III

Chapter 5: Introduction to SQL

Learning Objectives

Introduction

Origins of the SQL Standard

The SQL Environment

SQL Data Types

Defining A Database in SQL

Generating SQL Database Definitions

Creating Tables

Creating Data Integrity Controls

Changing Table Definitions

Removing Tables

Inserting, Updating, and Deleting Data

Batch Input

Deleting Database Contents

Updating Database Contents

Internal Schema Definition in RDBMSs

Creating Indexes

Table of Contents

Processing Single Tables

- Clauses of the SELECT Statement
- Using Expressions
- Using Functions
- Using Wildcards
- Using Comparison Operators
- Using Null Values
- Using Boolean Operators
- Using Ranges for Qualification
- Using Distinct Values
- Using IN and NOT IN with Lists
- Sorting Results: The ORDER BY Clause
- Categorizing Results: The GROUP BY Clause
- Qualifying Results by Categories: The HAVING Clause

Summary

Key Terms

Review Questions

Problems and Exercises

Field Exercises

References

Further Reading

Web Resources

Case: Forondo Artist Management Excellence Inc.

Chapter 6: Advanced SQL

Learning Objectives

Introduction

Processing Multiple Tables

- Equi-Join
- Natural Join
- Outer Join
- Sample Join Involving Four Tables
- Self-Join
- Subqueries
- Correlated Subqueries
- Using Derived Tables
- Combining Queries
- Conditional Expressions
- More Complicated SQL Queries

Tips for Developing Queries

- Guidelines for Better Query Design

Table of Contents

Using and Defining Views

- Materialized Views

Triggers and Routines

- Triggers

- Routines and Other Programming Extensions

- Example Routine in Oracles PL/SQL

Data Dictionary Facilities

Recent Enhancements and Extensions to SQL

- Analytical and OLAP Functions

- New Temporal Features in SQL

- Other Enhancements

Summary

Key Terms

Review Questions

Problems and Exercises

Field Exercises

References

Further Reading

Web Resources

Case: Forondo Artist Management Excellence Inc.

Chapter 7: Databases in Applications

Learning Objectives

Location, Location, Location!

Introduction

Client/Server Architectures

Databases in Three-Tier Applications

- A Java Web Application

- A Python Web Application

Key Considerations in Three-Tier Applications

- Stored Procedures

- Transactions

- Database Connections

- Key Benefits of Three-Tier Applications

Transaction Integrity

Controlling Concurrent Access

- The Problem of Lost Updates

- Serializability

- Locking Mechanisms

 - Locking Level

Table of Contents

Types of Locks

Deadlock

Managing Deadlock

Versioning

Managing Data Security in an Application Context

Threats to Data Security

Establishing Client/Server Security

Server Security

Network Security

Application Security Issues in Three-Tier Client/Server Environments

Data Privacy

Summary

Key Terms

Review Questions

Problems and Exercises

Field Exercises

References

Further Reading

Web Resources

Case: Forondo Artist Management Excellence Inc.

Chapter 8: Physical Database Design and Database Infrastructure

Learning Objectives

Introduction

The Physical Database Design Process

Who Is Responsible for Physical Database Design?

Physical Database Design as a Basis for Regulatory Compliance

SOX and Databases

IT Change Management

Logical Access to Data

IT Operations

Data Volume and Usage Analysis

Designing Fields

Choosing Data Types

Coding Techniques

Controlling Data Integrity

Handling Missing Data

Denormalizing and Partitioning Data

Denormalization

Opportunities for and Types of Denormalization

Denormalize with Caution

Partitioning

Table of Contents

Designing Physical Database Files

File Organizations

- Heap File Organization
- Sequential File Organizations
- Indexed File Organizations
- Hashed File Organizations

Clustering Files

Designing Controls for Files

Using and Selecting Indexes

- Creating a Unique Key Index
- Creating a Secondary (Nonunique) Key Index
- When to Use Indexes

Designing a Database for Optimal Query Performance

- Parallel Query Processing
- Overriding Automatic Query Optimization

Data Dictionaries and Repositories

- Data Dictionary
- Repositories

Database Software Data Security Features

- Views
- Integrity Controls
- Authorization Rules
- User-Defined Procedures
- Encryption
- Authentication Schemes
 - Passwords
 - Strong Authentication

Database Backup and Recovery

Basic Recovery Facilities

- Backup Facilities
- Journalizing Facilities
- Checkpoint Facility
- Recovery Manager

Recovery and Restart Procedures

- Disk Mirroring
- Restore/Rerun
- Backward Recovery
- Forward Recovery

Types of Database Failure

- Aborted Transactions
- Incorrect Data
- System Failure
- Database Destruction

Disaster Recovery

Table of Contents

Cloud-Based Database Infrastructure

Cloud-Based Models for Providing Data Management Services 407

Benefits and Downsides of Using Cloud-Based Management Services 408

Summary

Key Terms

Review Questions

Problems and Exercises

Field Exercises

References

Further Reading

Web Resources

Case: Forondo Artist Management Excellence Inc.

Part IV: Advanced Database Topics

An Overview of Part IV

Chapter 9: Data Warehousing and Data Integration

Learning Objectives

Introduction

Basic Concepts of Data Warehousing

A Brief History of Data Warehousing

The Need for Data Warehousing

Need for a Company-Wide View

Need to Separate Operational and Informational Systems

Data Warehouse Architectures

Independent Data Mart Data Warehousing Environment

Dependent Data Mart and Operational Data Store Architecture: A Three-Level Approach

Logical Data Mart and Real-Time Data Warehouse Architecture

Three-Layer Data Architecture

Role of the Enterprise Data Model

Role of Metadata

Some Characteristics of Data Warehouse Data

Status versus Event Data

Transient versus Periodic Data

An Example of Transient and Periodic Data

Transient Data

Periodic Data

Other Data Warehouse Changes

The Derived Data Layer

Characteristics of Derived Data

The Star Schema

Fact Tables and Dimension Tables

Table of Contents

Example Star Schema	
Surrogate Key	
Grain of the Fact Table	
Duration of the Database	
Size of the Fact Table	
Modeling Date and Time	
Variations of the Star Schema	
Multiple Fact Tables	
Factless Fact Tables	
Normalizing Dimension Tables	
Multivalued Dimensions	
Hierarchies	
Slowly Changing Dimensions	
Determining Dimensions and Facts	
Data Integration: An Overview	
General Approaches to Data Integration	
Data Federation	
Data Propagation	
Data Integration for Data Warehousing: The Reconciled Data Layer	
Characteristics of Data after ETL	
The ETL Process	
Mapping and Metadata Management	
Extract	
Cleanse	
Load and Index	
Data Transformation	
Data Transformation Functions	
Record-Level Functions	
Field-Level Functions	
Data Warehouse Administration	
The Future of Data Warehousing: Integration with Other Forms of Data Management and Analytics	
Speed of Processing	
Moving the Data Warehouse into the Cloud	
Dealing with Unstructured Data	
Summary	
Key Terms	
Review Questions	
Problems and Exercises	
Field Exercises	
References	
Further Reading	

Table of Contents

Web Resources

Chapter 10: Big Data Technologies

Learning Objectives

Introduction

Moving Beyond Transactional and Data Warehousing Databases

Big Data

NoSQL

Classification of NoSQL DBMSs

Key-Value Stores

Document Stores

Wide-Column Stores

Graph-Oriented Databases

NoSQL Examples

Redis

MongoDB

Apache Cassandra

Neo4j

A NoSQL Example: MongoDB

Documents

Collections

Relationships

Querying MongoDB

Impact of NoSQL on Database Professionals

Hadoop

Components of Hadoop

The Hadoop Distributed File System (HDFS)

MapReduce

Pig

Hive

HBase

A Practical Introduction to Pig

Loading Data

Transforming Data

A Practical Introduction to Hive

Creating a Table

Loading Data into the Table

Processing the Data

Integrated Analytics and Data Science Platforms

HP HAVEn

Teradata Aster

IBM Big Data Platform

Putting It All Together: Integrated Data Architecture

Summary

Key Terms

Table of Contents

Review Questions

Problems and Exercises

References

Further Reading

Web Resources

Chapter 11: Analytics and Its Implications

Learning Objectives

Introduction

Analytics

Types of Analytics

Use of Descriptive Analytics

SQL OLAP Querying

OLAP Tools

Data Visualization

Business Performance Management and Dashboards

Use of Predictive Analytics

Data Mining Tools

Examples of Predictive Analytics

Use of Prescriptive Analytics

Key User Tools for Analytics

Analytical and OLAP Functions

R 524

Python

Apache Spark

Data Management Infrastructure for Analytics

Impact of Big Data and Analytics

Applications of Big Data and Analytics

Business

E-Government and Politics

Science and Technology

Smart Health and Well-Being

Security and Public Safety

Implications of Big Data Analytics and Decision Making

Personal Privacy versus Collective Benefits

Ownership and Access

Quality and Reuse of Data and Algorithms

Transparency and Validation

Changing Nature of Work

Demands for Workforce Capabilities and Education

Summary

Key Terms

Review Questions

Problems and Exercises

Table of Contents

References

Further Reading

Chapter 12: Data and Database Administration with Focus on Data Quality

Learning Objectives

Introduction

Overview of Data and Database Administration

- Data Administration

- Database Administration

 - Traditional Database Administration

 - Trends in Database Administration

- Evolving Data Administration Roles

The Open Source Movement and Database Management

Data Governance

Managing Data Quality

- Characteristics of Quality Data

 - External Data Sources

 - Redundant Data Storage and Inconsistent Metadata

 - Data Entry Problems

 - Lack of Organizational Commitment

- Data Quality Improvement

 - Get the Business Buy-In

 - Conduct a Data Quality Audit

 - Establish a Data Stewardship Program

 - Improve Data Capture Processes

 - Apply Modern Data Management Principles and Technology

 - Apply TQM Principles and Practices

- Summary of Data Quality

Data Availability

- Costs of Downtime

- Measures to Ensure Availability

 - Hardware Failures

 - Loss or Corruption of Data

 - Human Error

 - Maintenance Downtime

 - Network-Related Problems

Master Data Management

Summary

Key Terms

Review Questions

Problems and Exercises

Field Exercises

References

Table of Contents

Further Reading

Web Resources

Glossary of Acronyms

Glossary of Terms

Index