

Introduction to Database Systems

January 3, 2020

Contents

1	An Overview of Database Management	13
1.1	Introduction	13
1.2	What Is a Database System?	13
1.3	What Is a Database?	13
1.4	Why Database?	13
1.5	Data Independence	13
1.6	Relational Systems and Others	13
1.7	Summary	13
1.8	Exercises	13
1.9	References and Bibliography	13
2	Database System Architecture	15
2.1	Introduction	16
2.2	The Three Levels of the Architecture	16
2.3	The External Level	16
2.4	The Conceptual Level	16
2.5	The Internal Level	16
2.6	Mappings	16
2.7	The Database Administrator	16
2.8	The Database Management System	16
2.9	Data Communications	16
2.10	Client/Server Architecture	16
2.11	Utilities	16
2.12	Distributed Processing	16
2.13	Summary	16
2.14	Exercises	16
2.15	References and Bibliography	16
3	An Introduction to Relational Databases	17
3.1	Introduction	17
3.2	An Informal Look at the Relational Model	17
3.3	Relations and Relvars	17
3.4	What Relations Mean	17
3.5	Optimization	17

3.6	The Catalog	17
3.7	Base Relvars and Views	17
3.8	Transactions	17
3.9	The Suppliers-and-Parts Database	17
3.10	Summary	17
3.11	Exercises	17
3.12	References and Bibliography	17
4	An Introduction to SQL	19
4.1	Introduction	19
4.2	Overview	19
4.3	The Catalog	19
4.4	Views	19
4.5	Transactions	19
4.6	Embedded SQL	19
4.7	Dynamic SQL and SQL/CU	19
4.8	SQL Is Not Perfect	19
4.9	Summary	19
4.10	Exercises	19
4.11	References and Bibliography	19
5	TYPES	21
5.1	Introduction	21
5.2	Values vs. Variables	21
5.3	Types vs. Representations	21
5.4	Type Definition	21
5.5	Operators	21
5.6	Type Generators	21
5.7	SQL Facilities	21
5.8	Summary	21
5.9	Exercises	21
5.10	References and Bibliography	21
6	Relations	23
6.1	Introduction	23
6.2	Tuples	23
6.3	Relation Types	23
6.4	Relation Values	23
6.5	Relation Variables	23
6.6	SQL Facilities	23
6.7	Summary	23
6.8	Exercises	23
6.9	References and Bibliography	23

7	Relational Algebra	25
7.1	Introduction	25
7.2	Closure Revisited	25
7.3	The Original Algebra: Syntax	25
7.4	The Original Algebra: Semantics	25
7.5	Examples	25
7.6	What Is the Algebra For?	25
7.7	Further Points	25
7.8	Additional Operators	25
7.9	Grouping and Ungrouping	25
7.10	Summary	25
7.11	Exercises	25
7.12	References and Bibliography	25
8	Relational Calculus	27
8.1	Introduction	27
8.2	Tuple Calculus	27
8.3	Examples	27
8.4	Calculus vs. Algebra	27
8.5	Computational Capabilities	27
8.6	SQL Facilities	27
8.7	Domain Calculus	27
8.8	Query-By-Example	27
8.9	Summary	27
8.10	Exercises	27
8.11	References and Bibliography	27
9	Integrity	29
9.1	Introduction	30
9.2	A Closer Look	30
9.3	Predicates and Propositions	30
9.4	Relvar Predicates and Database Predicates	30
9.5	Checking the Constraints	30
9.6	Internal vs. External Predicates	30
9.7	Correctness vs. Consistency	30
9.8	Integrity and Views	30
9.9	A Constraint Classification Scheme	30
9.10	Keys	30
9.11	Triggers (a Digression)	30
9.12	SQL Facilities	30
9.13	Summary	30
9.14	Exercises	30
9.15	References and Bibliography	30

10 Views	31
10.1 Introduction	31
10.2 What Are Views For?	31
10.3 View Retrievals	31
10.4 View Updates	31
10.5 Snapshots (a Digression)	31
10.6 SQL Facilities	31
10.7 Summary	31
10.8 Exercises	31
10.9 References and Bibliography	31
11 Functional Dependencies	33
11.1 Introduction	33
11.2 Basic Definitions	33
11.3 Trivial and Nontrivial Dependencies	33
11.4 Closure of a Set of Dependencies	33
11.5 Closure of a Set of Attributes	33
11.6 Irreducible Sets of Dependencies	33
11.7 Summary	33
11.8 Exercises	33
11.9 References and Bibliography	33
12 Further Normalization I: INF, 2NF, 3NF, BCNF	35
12.1 Introduction	35
12.2 Nonloss Decomposition and Functional Dependencies	35
12.3 First, Second, and Third Normal Forms	35
12.4 Dependency Preservation	35
12.5 Boyce/Codd Normal Form	35
12.6 A Note on Relation-Valued Attributes	35
12.7 Summary	35
12.8 Exercises	35
12.9 References and Bibliography	35
13 Further Normalization II: Higher Normal Forms	37
13.1 Introduction	37
13.2 Multi-valued Dependencies and Fourth Normal Form	37
13.3 Join Dependencies and Fifth Normal Form	37
13.4 The Normalization Procedure Summarized	37
13.5 A Note on Denormalization	37
13.6 Orthogonal Design (a Digression)	37
13.7 Other Normal Forms	37
13.8 Summary	37
13.9 Exercises	37
13.10References and Bibliography	37

14 Semantic Modeling	39
14.1 Introduction	39
14.2 The Overall Approach	39
14.3 The E/R Model	39
14.4 E/R Diagrams	39
14.5 Database Design with the E/R Model	39
14.6 A Brief Analysis	39
14.7 Summary	39
14.8 Exercises	39
14.9 References and Bibliography	39
15 Recovery	41
15.1 Introduction	41
15.2 Transactions	41
15.3 Transaction Recovery	41
15.4 System Recovery	41
15.5 Media Recovery	41
15.6 Two-Phase Commit	41
15.7 Savepoints (a Digression)	41
15.8 SQL Facilities	41
15.9 Summary	41
15.10 Exercises	41
15.11 References and Bibliography	41
16 Concurrency	43
16.1 Introduction	44
16.2 Three Concurrency Problems	44
16.3 Locking	44
16.4 The Three Concurrency Problems Revisited	44
16.5 Deadlock	44
16.6 Serializability	44
16.7 Recovery Revisited	44
16.8 Isolation Levels	44
16.9 Intent Locking	44
16.10 Dropping ACID	44
16.11 SQL Facilities	44
16.12 Summary	44
16.13 Exercises	44
16.14 References and Bibliography	44
17 Security	45
17.1 Introduction	45
17.2 Discretionary Access Control	45
17.3 Mandatory Access Control	45
17.4 Statistical Databases	45
17.5 Data Encryption	45

17.6 SQL Facilities	45
17.7 Summary	45
17.8 Exercises	45
17.9 References and Bibliography	45
18 Optimization	47
18.1 Introduction	47
18.2 A Motivating Example	47
18.3 An Overview of Query Processing	47
18.4 Expression Transformation	47
18.5 Database Statistics	47
18.6 A Divide-and-Conquer Strategy	47
18.7 Implementing the Relational Operators	47
18.8 Summary	47
18.9 Exercises	47
18.10References and Bibliography	47
19 Missing Information	49
19.1 Introduction	49
19.2 An Overview of the 3 VL Approach	49
19.3 Some Consequences of the Foregoing Scheme	49
19.4 Nulls and Keys	49
19.5 Outer Join (a Digression)	49
19.6 Special Values	49
19.7 SQL Facilities	49
19.8 Summary	49
19.9 Exercises	49
19.10References and Bibliography	49
20 Type Inheritance	51
20.1 Introduction	51
20.2 Type Hierarchies	51
20.3 Polymorphism and Substitutability	51
20.4 Variables and Assignments	51
20.5 Specialization by Constraint	51
20.6 Comparisons	51
20.7 Operators, Versions, and Signatures	51
20.8 Is a Circle an Ellipse?	51
20.9 Specialization by Constraint Revisited	51
20.10SQL Facilities	51
20.11Summary	51
20.12Exercises	51
20.13References and Bibliography	51

21 Distributed Databases	53
21.1 Introduction	53
21.2 Some Preliminaries	53
21.3 The Twelve Objectives	53
21.4 Problems of Distributed Systems	53
21.5 Client/Server Systems	53
21.6 DBMS Independence	53
21.7 SQL Facilities	53
21.8 Summary	53
21.9 Exercises	53
21.10References and Bibliography	53
22 Decision Support	55
22.1 Introduction	55
22.2 Aspects of Decision Support	55
22.3 Database Design for Decision Support	55
22.4 Data Preparation	55
22.5 Data Warehouses and Data Marts	55
22.6 Online Analytical Processing	55
22.7 Data Mining	55
22.8 SQL Facilities	55
22.9 Summary	55
22.10Exercises	55
22.11References and Bibliography	55
23 Temporal Databases	57
23.1 Introduction	57
23.2 What Is the Problem?	57
23.3 Intervals	57
23.4 Packing and Unpacking Relations	57
23.5 Generalizing the Relational Operators	57
23.6 Database Design	57
23.7 Integrity Constraints	57
23.8 Summary	57
23.9 Exercises	57
23.10References and Bibliography	57
24 Logic-Based Databases	59
24.1 Introduction	59
24.2 Overview	59
24.3 Propositional Calculus	59
24.4 Predicate Calculus	59
24.5 A Proof-Theoretic View of Databases	59
24.6 Deductive Database Systems	59
24.7 Recursive Query Processing	59
24.8 Summary	59

24.9 Exercises	59
24.10References and Bibliography	59
25 Object Databases	61
25.1 Introduction	61
25.2 Objects, Classes, Methods, and Messages	61
25.3 A Closer Look	61
25.4 A Cradie-to-Grave Example	61
25.5 Miscellaneous Issues	61
25.6 Summary	61
25.7 Exercises	61
25.8 References and Bibliography	61
26 Object/Relational Databases	63
26.1 Introduction	63
26.2 The First Great Blunder	63
26.3 The Second Great Blunder	63
26.4 Implementation Issues	63
26.5 Benefits of True Rapprochement	63
26.6 SQL Facilities	63
26.7 Summary	63
26.8 Exercises	63
26.9 References and Bibliography	63
27 The World Wide Web and XML	65
27.1 Introduction	65
27.2 The Web and the Internet	65
27.3 An Overview of XML	65
27.4 XML Data Definition	65
27.5 XML Data Manipulation	65
27.6 XML and Databases	65
27.7 SQL Facilities	65
27.8 Summary	65
27.9 Exercises	65
27.10References and Bibliography	65
28 The TransRelational Model	67
28.1 Introduction	67
28.2 Three Levels of Abstraction	67
28.3 The Basic Idea	67
28.4 Condensed Columns	67
28.5 Merged Columns	67
28.6 Implementing the Relational Operators	67
28.7 Summary	67
28.8 References and Bibliography	67

<i>CONTENTS</i>	11
29 SQL Expressions	69
29.1 Introduction	69
29.2 Table Expressions	69
29.3 Boolean Expressions	69
30 Abbreviations, Acronyms, and Symbols	71
31 Storage Structures and Access Methods (online at http://llzmvwxw.com/cssupport)	73
31.1 Introduction	73
31.2 Database Access: An Overview	73
31.3 Page Sets and Files	73
31.4 Indexing	73
31.5 Hashing	73
31.6 Pointer Chains	73
31.7 Compression Techniques	73
31.8 Summary	73
31.9 Exercises	73
31.10References and Bibliography	73

Chapter 1

An Overview of Database Management

1.1 Introduction

1.2 What Is a Database System?

1.3 What Is a Database?

1.4 Why Database?

1.5 Data Independence

1.6 Relational Systems and Others

1.7 Summary

1.8 Exercises

1.9 References and Bibliography

Chapter 2

Database System Architecture

- 2.1 Introduction
- 2.2 The Three Levels of the Architecture
- 2.3 The External Level
- 2.4 The Conceptual Level
- 2.5 The Internal Level
- 2.6 Mappings
- 2.7 The Database Administrator
- 2.8 The Database Management System
- 2.9 Data Communications
- 2.10 Client/Server Architecture
- 2.11 Utilities
- 2.12 Distributed Processing
- 2.13 Summary
- 2.14 Exercises
- 2.15 References and Bibliography

Chapter 3

An Introduction to Relational Databases

- 3.1 Introduction
- 3.2 An Informal Look at the Relational Model
- 3.3 Relations and Relvars
- 3.4 What Relations Mean
- 3.5 Optimization
- 3.6 The Catalog
- 3.7 Base Relvars and Views
- 3.8 Transactions
- 3.9 The Suppliers-and-Parts Database
- 3.10 Summary
- 3.11 Exercises
- 3.12 References and Bibliography

Chapter 4

An Introduction to SQL

4.1 Introduction

4.2 Overview

4.3 The Catalog

4.4 Views

4.5 Transactions

4.6 Embedded SQL

4.7 Dynamic SQL and SQL/CU

4.8 SQL Is Not Perfect

4.9 Summary

4.10 Exercises

4.11 References and Bibliography

Chapter 5

TYPES

5.1 Introduction

5.2 Values vs. Variables

5.3 Types vs . Representations

5.4 Type Definition

5.5 Operators

5.6 Type Generators

5.7 SQL Facilities

5.8 Summary

5.9 Exercises

5.10 References and Bibliography

Chapter 6

Relations

6.1 Introduction

6.2 Tuples

6.3 Relation Types

6.4 Relation Values

6.5 Relation Variables

6.6 SQL Facilities

6.7 Summary

6.8 Exercises

6.9 References and Bibliography

Chapter 7

Relational Algebra

7.1 Introduction

7.2 Closure Revisited

7.3 The Original Algebra: Syntax

7.4 The Original Algebra: Semantics

7.5 Examples

7.6 What Is the Algebra For?

7.7 Further Points

7.8 Additional Operators

7.9 Grouping and Ungrouping

7.10 Summary

7.11 Exercises

7.12 References and Bibliography

Chapter 8

Relational Calculus

8.1 Introduction

8.2 Tuple Calculus

8.3 Examples

8.4 Calculus vs. Algebra

8.5 Computational Capabilities

8.6 SQL Facilities

8.7 Domain Calculus

8.8 Query-By-Example

8.9 Summary

8.10 Exercises

8.11 References and Bibliography

Chapter 9

Integrity

9.1 Introduction

9.2 A Closer Look

9.3 Predicates and Propositions

9.4 Relvar Predicates and Database Predicates

9.5 Checking the Constraints

9.6 Internal vs. External Predicates

9.7 Correctness vs. Consistency

9.8 Integrity and Views

9.9 A Constraint Classification Scheme

9.10 Keys

9.11 Triggers (a Digression)

9.12 SQL Facilities

9.13 Summary

9.14 Exercises

9.15 References and Bibliography

Chapter 10

Views

10.1 Introduction

10.2 What Are Views For?

10.3 View Retrievals

10.4 View Updates

10.5 Snapshots (a Digression)

10.6 SQL Facilities

10.7 Summary

10.8 Exercises

10.9 References and Bibliography

Chapter 11

Functional Dependencies

- 11.1 Introduction
- 11.2 Basic Definitions
- 11.3 Trivial and Nontrivial Dependencies
- 11.4 Closure of a Set of Dependencies
- 11.5 Closure of a Set of Attributes
- 11.6 Irreducible Sets of Dependencies
- 11.7 Summary
- 11.8 Exercises
- 11.9 References and Bibliography

Chapter 12

Further Normalization I: 1NF, 2NF, 3NF, BCNF

- 12.1 Introduction
- 12.2 Nonloss Decomposition and Functional Dependencies
- 12.3 First, Second, and Third Normal Forms
- 12.4 Dependency Preservation
- 12.5 Boyce/Codd Normal Form
- 12.6 A Note on Relation-Valued Attributes
- 12.7 Summary
- 12.8 Exercises
- 12.9 References and Bibliography

Chapter 13

Further Normalization II: Higher Normal Forms

- 13.1 Introduction
- 13.2 Multi-valued Dependencies and Fourth Normal Form
- 13.3 Join Dependencies and Fifth Normal Form
- 13.4 The Normalization Procedure Summarized
- 13.5 A Note on Denormalization
- 13.6 Orthogonal Design (a Digression)
- 13.7 Other Normal Forms
- 13.8 Summary
- 13.9 Exercises
- 13.10 References and Bibliography

Chapter 14

Semantic Modeling

14.1 Introduction

14.2 The Overall Approach

14.3 The E/R Model

14.4 E/R Diagrams

14.5 Database Design with the E/R Model

14.6 A Brief Analysis

14.7 Summary

14.8 Exercises

14.9 References and Bibliography

Chapter 15

Recovery

- 15.1 Introduction
- 15.2 Transactions
- 15.3 Transaction Recovery
- 15.4 System Recovery
- 15.5 Media Recovery
- 15.6 Two-Phase Commit
- 15.7 Savepoints (a Digression)
- 15.8 SQL Facilities
- 15.9 Summary
- 15.10 Exercises
- 15.11 References and Bibliography

Chapter 16

Concurrency

- 16.1 Introduction
- 16.2 Three Concurrency Problems
- 16.3 Locking
- 16.4 The Three Concurrency Problems Revisited
- 16.5 Deadlock
- 16.6 Serializability
- 16.7 Recovery Revisited
- 16.8 Isolation Levels
- 16.9 Intent Locking
- 16.10 Dropping ACID
- 16.11 SQL Facilities
- 16.12 Summary
- 16.13 Exercises
- 16.14 References and Bibliography

Chapter 17

Security

- 17.1 Introduction
- 17.2 Discretionary Access Control
- 17.3 Mandatory Access Control
- 17.4 Statistical Databases
- 17.5 Data Encryption
- 17.6 SQL Facilities
- 17.7 Summary
- 17.8 Exercises
- 17.9 References and Bibliography

Chapter 18

Optimization

18.1 Introduction

18.2 A Motivating Example

18.3 An Overview of Query Processing

18.4 Expression Transformation

18.5 Database Statistics

18.6 A Divide-and-Conquer Strategy

18.7 Implementing the Relational Operators

18.8 Summary

18.9 Exercises

18.10 References and Bibliography

Chapter 19

Missing Information

19.1 Introduction

19.2 An Overview of the 3 VL Approach

19.3 Some Consequences of the Foregoing Scheme

19.4 Nulls and Keys

19.5 Outer Join (a Digression)

19.6 Special Values

19.7 SQL Facilities

19.8 Summary

19.9 Exercises

19.10 References and Bibliography

Chapter 20

Type Inheritance

20.1 Introduction

20.2 Type Hierarchies

20.3 Polymorphism and Substitutability

20.4 Variables and Assignments

20.5 Specialization by Constraint

20.6 Comparisons

20.7 Operators, Versions, and Signatures

20.8 Is a Circle an Ellipse?

20.9 Specialization by Constraint Revisited

20.10 SQL Facilities

20.11 Summary

20.12 Exercises

20.13 References and Bibliography

Chapter 21

Distributed Databases

- 21.1 Introduction
- 21.2 Some Preliminaries
- 21.3 The Twelve Objectives
- 21.4 Problems of Distributed Systems
- 21.5 Client/Server Systems
- 21.6 DBMS Independence
- 21.7 SQL Facilities
- 21.8 Summary
- 21.9 Exercises
- 21.10 References and Bibliography

Chapter 22

Decision Support

22.1 Introduction

22.2 Aspects of Decision Support

22.3 Database Design for Decision Support

22.4 Data Preparation

22.5 Data Warehouses and Data Marts

22.6 Online Analytical Processing

22.7 Data Mining

22.8 SQL Facilities

22.9 Summary

22.10 Exercises

22.11 References and Bibliography

Chapter 23

Temporal Databases

23.1 Introduction

23.2 What Is the Problem?

23.3 Intervals

23.4 Packing and Unpacking Relations

23.5 Generalizing the Relational Operators

23.6 Database Design

23.7 Integrity Constraints

23.8 Summary

23.9 Exercises

23.10 References and Bibliography

Chapter 24

Logic-Based Databases

24.1 Introduction

24.2 Overview

24.3 Propositional Calculus

24.4 Predicate Calculus

24.5 A Proof-Theoretic View of Databases

24.6 Deductive Database Systems

24.7 Recursive Query Processing

24.8 Summary

24.9 Exercises

24.10 References and Bibliography

Chapter 25

Object Databases

25.1 Introduction

25.2 Objects, Classes, Methods, and Messages

25.3 A Closer Look

25.4 A Cradle-to-Grave Example

25.5 Miscellaneous Issues

25.6 Summary

25.7 Exercises

25.8 References and Bibliography

Chapter 26

Object/Relational Databases

- 26.1 Introduction
- 26.2 The First Great Blunder
- 26.3 The Second Great Blunder
- 26.4 Implementation Issues
- 26.5 Benefits of True Rapprochement
- 26.6 SQL Facilities
- 26.7 Summary
- 26.8 Exercises
- 26.9 References and Bibliography

Chapter 27

The World Wide Web and XML

- 27.1 Introduction
- 27.2 The Web and the Internet
- 27.3 An Overview of XML
- 27.4 XML Data Definition
- 27.5 XML Data Manipulation
- 27.6 XML and Databases
- 27.7 SQL Facilities
- 27.8 Summary
- 27.9 Exercises
- 27.10 References and Bibliography

Chapter 28

The TransRelational Model

28.1 Introduction

28.2 Three Levels of Abstraction

28.3 The Basic Idea

28.4 Condensed Columns

28.5 Merged Columns

28.6 Implementing the Relational Operators

28.7 Summary

28.8 References and Bibliography

Chapter 29

SQL Expressions

29.1 Introduction

29.2 Table Expressions

29.3 Boolean Expressions

Chapter 30

Abbreviations, Acronyms, and Symbols

Chapter 31

Storage Structures and Access Methods (online at <http://llzmvwxw.com/cssupport>)

- 31.1 Introduction
- 31.2 Database Access: An Overview
- 31.3 Page Sets and Files
- 31.4 Indexing
- 31.5 Hashing
- 31.6 Pointer Chains
- 31.7 Compression Techniques
- 31.8 Summary
- 31.9 Exercises
- 31.10 References and Bibliography