

includes
hands-on
exercises

Database Fundamentals

Ideal for application developers and administrators



Neeraj Sharma

Liviu Perniu

Raul F. Chong

Abhishek Iyer

Chaitali Nandan

Adi-Cristina Mitea

Mallarswami Nonvinkere

Mirela Danubianu

Database Fundamentals

A book for the community by the community

Neeraj Sharma, Liviu Perniu, Raul F. Chong, Abhishek Iyer, Chaitali Nandan,
Adi-Cristina Mitea, Mallarswami Nonvinkere, Mirela Danubianu

FIRST EDITION

First Edition (November 2010)

© Copyright IBM Corporation 2010. All rights reserved.

IBM Canada
8200 Warden Avenue
Markham, ON
L6G 1C7
Canada

This edition covers IBM® DB2® Express-C Version 9.7 for Linux®, UNIX® and Windows®.

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

*IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.*

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

*Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan, Ltd.
3-2-12, Roppongi, Minato-ku, Tokyo 106-8711*

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided "AS IS", without warranty of any kind. IBM shall not be liable for any damages arising out of your use of the sample programs.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "[Copyright and trademark information](http://www.ibm.com/legal/copytrade.shtml)" at www.ibm.com/legal/copytrade.shtml.

Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

Microsoft and Windows are trademarks of Microsoft Corporation in the United States, other countries, or both.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product, or service names may be trademarks or service marks of others.

Table of Contents

Preface	15
Who should read this book?	15
How is this book structured?	15
A book for the community	15
Conventions	15
What's next?	16
About the Authors.....	17
Contributors	19
Acknowledgements	21
Chapter 1 - Databases and information models	23
1.1 What is a database?	23
1.2 What is a database management system?	23
1.2.1 The evolution of database management systems	24
1.3 Introduction to information models and data models	26
1.4 Types of information models	27
1.4.1 Network model	28
1.4.2 Hierarchical model	28
1.4.3 Relational model	29
1.4.4 Entity-Relationship model	30
1.4.5 Object-relational model	31
1.4.6 Other data models	32
1.5 Typical roles and career path for database professionals	32
1.5.1 Data Architect	32
1.5.2 Database Architect	32
1.5.3 Database Administrator (DBA)	33
1.5.4 Application Developer	34
1.6 Summary	34
1.7 Exercises	35
1.8 Review questions	35
Chapter 2 – The relational data model	37
2.1 Relational data model: The big picture	37
2.2 Basic concepts	38
2.2.1 Attributes	38
2.2.2 Domains	39
2.2.3 Tuples	40
2.2.4 Relations	40
2.2.5 Schemas	41
2.2.6 Keys	41
2.3 Relational data model constraints	44
2.3.1 Entity integrity constraint	44
2.3.2 Referential integrity constraint	45
2.3.3 Semantic integrity constraints	46
2.4 Relational algebra	49

2.4.1 Union.....	49
2.4.2 Intersection	49
2.4.3 Difference	50
2.4.4 Cartesian product	51
2.4.5 Selection	52
2.4.6 Projection.....	53
2.4.7 Join	54
2.4.8 Division	56
2.5. Relational calculus	57
2.5.1 Tuple-oriented relational calculus.....	58
2.5.2 Domain-oriented relational calculus	59
2.6 Summary.....	60
2.7 Exercises	60
2.8 Review questions.....	62
Chapter 3 – The conceptual data model.....	65
3.1 Conceptual, logical and physical modeling: The big picture.....	65
3.2 What is a model?	67
3.2.1 Data model	67
3.2.2 Database model.....	67
3.2.3 Conceptual data model concepts	68
3.3 A case study involving a Library Management System - Part 1 of 3	77
3.3.1 Developing the conceptual model	77
3.4 Summary.....	85
3.5 Exercises	85
3.6 Review questions.....	85
Chapter 4 – Relational Database Design	89
4.1 The problem of redundancy.....	89
4.1.1 Insertion Anomalies	90
4.1.2 Deletion Anomalies.....	90
4.1.3 Update Anomalies	90
4.2. Decompositions	91
4.3. Functional Dependencies	92
4.4 Properties of Functional Dependencies.....	94
4.4.1 Armstrong's Axioms.....	94
4.4.2 Computing the closure set of attributes	95
4.4.3 Entailment.....	96
4.5 Normal Forms	96
4.5.1 First Normal Form (1NF).....	96
4.5.2 Second Normal Form (2NF)	98
4.5.3 Third Normal Form (3NF)	99
4.5.4 Boyce-Codd Normal Form (BCNF).....	100
4.6 Properties of Decompositions.....	101
4.6.1 Lossless and Lossy Decompositions.....	102
4.6.2 Dependency-Preserving Decompositions	103
4.7 Minimal Cover	103