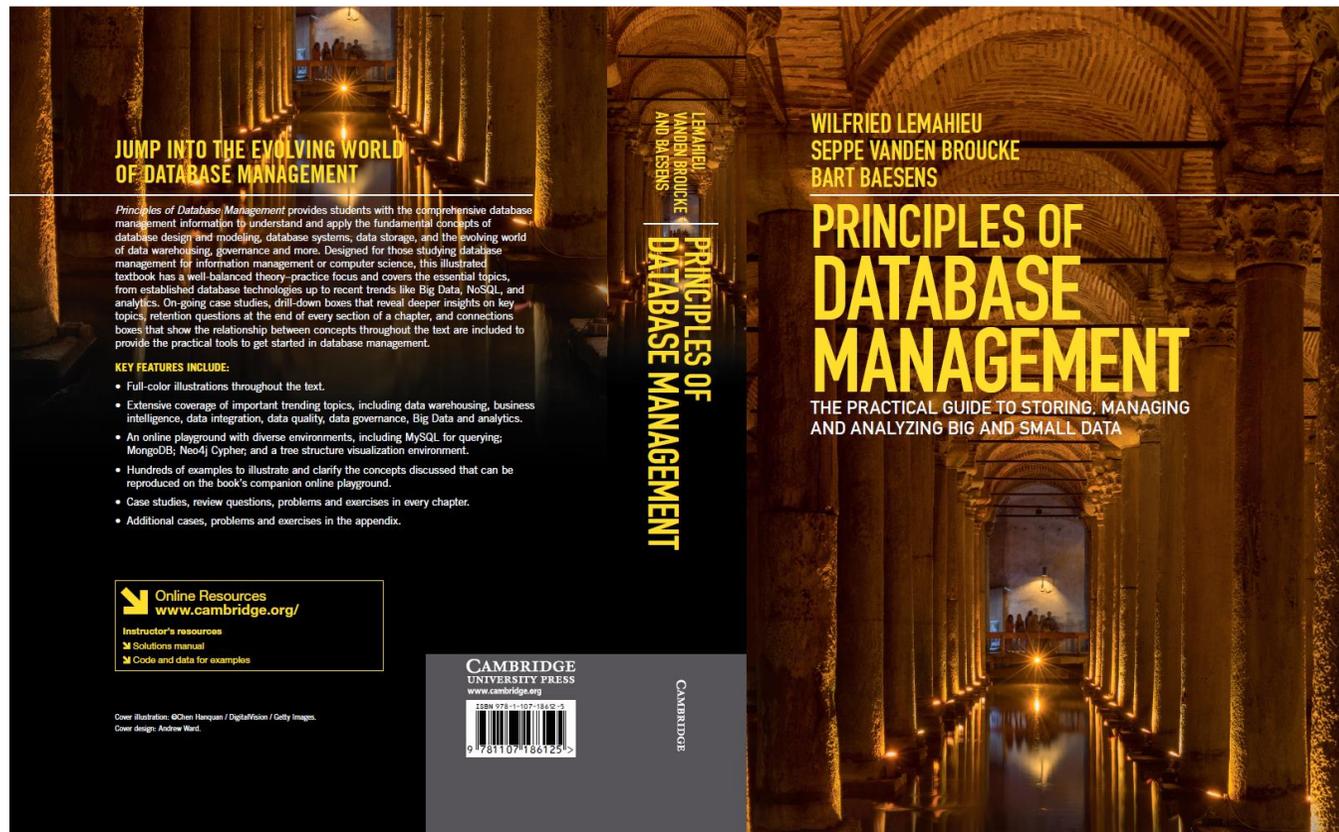


Principles of Database Management

www.pdbmbook.com



Author Team

- Prof. Wilfried Lemahieu
 - professor and dean at KU Leuven (Belgium)
 - more than 30 years database experience
 - Google H-index: 13
 - Wilfried.Lemahieu@kuleuven.be
- Prof. Seppe vanden Broucke
 - professor at KU Leuven (Belgium)
 - more than 20 years database experience
 - sold more than 5,000 books
 - Seppe.vandenBroucke@kuleuven.be
- Prof. Bart Baesens
 - professor at KU Leuven (Belgium) and University of Southampton (UK)
 - more than 250 papers co-authored and 20,000 books sold
 - Google scholar H-index: 55
 - Bart.Baesens@kuleuven.be



Book Outline

Part I. Databases and Database Design

- 1) Fundamental concepts of database management
- 2) Architecture and categorization of DBMSs
- 3) Conceptual data modeling using the (E)ER model and UML class diagram
- 4) Organizational aspects of data management

Part II. Types of Database Systems:

- 5) Legacy databases
- 6) Relational databases: the relational model
- 7) Relational databases: structured query language (SQL)
- 8) Object oriented databases and object persistence
- 9) Extended relational databases
- 10) XML databases
- 11) NoSQL databases

Book Outline (contd.)

Part III. Physical Data Storage, Transaction Management, and Database Access

- 12) Physical file organization and indexing
- 13) Physical database organization
- 14) Basics of transaction management
- 15) Accessing databases and database APIs
- 16) Data distribution and distributed transaction management

Part IV. Data Warehousing, Data Governance and (Big) Data Analytics

- 17) Data warehousing and business intelligence
- 18) Data integration, data quality and data governance
- 19) Big data
- 20) Analytics

Book Outline

Appendices

Appendix A. Cases and questions

Appendix B. Using the online environment

Appendix C. Answer key to review questions

Glossary

Target Audience

	Newcomers	Experienced users	Database users	Managers	Professor (Undergraduate Course)	Professor (Postgraduate course)	Information architect	Database Designer	Database Administrator	Data Scientists
Chapter 1	X		X	X	X		X	X	X	
Chapter 2	X		X	X	X		X	X	X	
Chapter 3	X		X	X	X		X	X	X	
Chapter 4	X		X	X	X		X	X	X	
Chapter 5					X			X	X	
Chapter 6	X				X			X	X	
Chapter 7	X				X			X	X	
Chapter 8	X				X			X	X	
Chapter 9	X				X			X	X	
Chapter 10					X			X	X	
Chapter 11		X			X			X	X	X
Chapter 12						X		X	X	
Chapter 13						X		X	X	
Chapter 14						X			X	
Chapter 15		X				X				
Chapter 16		X				X			X	
Chapter 17		X		X		X			X	X
Chapter 18		X		X		X			X	X
Chapter 19		X		X		X			X	X
Chapter 20		X		X		X				X

Example Courses

- Principles of Database Management
- Database Modeling
- Database Design
- Advanced Database Management
- Data Management
- Data Science

Book in Numbers

- ISBN: 9781107186125
- 3 years of work based on 15+ years of database teaching and research
- 816 pages
- 256 multiple choice questions (and solutions!)
- 150 open questions (and solutions!)
- 20 YouTube lectures (> 20 hours)
- 20 PowerPoint decks (1,400 slides)
- Endorsed by > 20 professors across the globe

Book Selling Points

- Extensive coverage of trending topics: data warehousing, business intelligence, data integration, data quality, data governance, Big Data, and analytics
- Careful balance between theory and practice
- Hundreds of examples to illustrate and clarify concepts discussed
- Case studies, review questions, problems, and exercises in every chapter
- Can be used in traditional and flipped classroom format
- Cross-chapter Sober Scenario

Cross-Chapter Sober Scenario

Sober

1000% Driven by Technology

Sober is a new taxi company deploying self-driving cars to provide cab services. Although it operates its own fleet of self-driving cabs, people can also register their cars as Sober cabs and have them provide taxi services whenever they are not using their cars. For the latter, Sober also wants to keep track of the car owners.

Sober offers two types of taxi services: ride-hailing and ride-sharing. Ride-hailing is a service whereby customers can hail a taxi so they can be picked up and driven to their destination for a time- and distance-based fee. The hailing is an immediate, on-demand service and requests can be made with the Sober App. With just one tap on the screen, a customer can request a cab from anywhere, receive an estimated wait time, and a notification when the car has arrived. Besides the Sober App, users can also hail Sober cabs by hand-waving them as they see them pass, in which case Sober's deep-learning based image recognition system identifies the wave gesture as a cab request. For each use of the ride-hail service, Sober wants to store the time of pick-up and drop-off, the location of pick-up and drop-off, the ride duration, the distance, the number of passengers, the fee, the type of request (via Sober App or hand-waving) and the number and name of the lead customer (the one who pays). The maximum number of passengers for a ride-hail service is six.

Ride-sharing is another service offered by Sober, which requires more careful planning. It can also be referred to as carpooling and aims at reducing costs, traffic congestion, and the carbon footprint. Because of the planning, both Sober and its customers can negotiate the fee whereby more customers per cab means a lower fee per customer (flexible pricing). To provide an eco-friendly incentive, Sober pledges to plant a tree for each customer who books 20 uses of the Sober ride-sharing service. For each ride-share service, Sober wants to store the time of pick-up and drop-off, the location of pick-up and drop-off, the ride duration, the distance, the number and names of all customers, and the upfront negotiated fee. The maximum number of passengers for a ride-share service is ten.

Due to the novelty of the self-driving car technology, accidents cannot be 100% ruled out. Sober also wants to store information about accident dates, location, and damage amounts per car.

Instructor Support

- Slides
- YouTube videos
- Multiple Choice Quiz Tool
- On-line environment
- Solutions manual

PowerPoint Slides

- 20 PowerPoint slide decks (one per chapter)
- 1,400 slides in total
- available in English, Mandarin and Spanish
- Slides available in ppt format (not pdf!)
 - Easily adjustable to course focus



Lecturer Resources

- > Lecture slides
- > Instructor's manual

Student Resources

- > Appendix on Relational Algebra and Relational Calculus
- > Manual for the Online Playground

> [Back to resources home](#)



Type	Name	Unlocked*	Format	Size
Lecture slides	Chapter_01_Fundamental_Concepts_of_Database_Management.ppt		MS Powerpoint	842KB
Lecture slides	Chapter_02_Architecture_and_Classification_of_DBMS.ppt		MS Powerpoint	1080KB
Lecture slides	Chapter_03_Conceptual_Data_Modeling_using_EER_and_UML.ppt		MS Powerpoint	2357KB
Lecture slides	Chapter_04_Organizational_Aspects_of_Data_Management.ppt		MS Powerpoint	668KB

Lecture Videos

- See [YouTube Book Playlist](#)
- More than 20 hours of free videos
- Currently more than 20,000 views
- Available in (customizable!) MP4 format from book website
- Can be used in **flipped classroom** format
 - Watch lectures upfront
 - Exercises in class

Lecture Videos

The Three Layer Architecture

The diagram illustrates the Three Layer Architecture of a database system. It is divided into three horizontal layers:

- External Layer:** Contains three boxes labeled "View".
- Conceptual/Logical Layer:** Contains four boxes labeled "Data item A", "Data item B", "Data item C", and "Data item D".
- Internal Layer:** Contains three boxes labeled "Files", "Disk", and "Cloud".

Arrows indicate the flow of data and independence:

- Arrows point from each "View" box to "Data item B", "Data item C", and "Data item D".
- Arrows point from "Data item B", "Data item C", and "Data item D" to the "Internal Layer".
- Arrows point from the "Internal Layer" to "Files", "Disk", and "Cloud".

Annotations in red boxes highlight key concepts:

- "External / logical mapping" is associated with the arrows between the External and Conceptual/Logical layers.
- "Logical data independence" is associated with the arrows between the External and Conceptual/Logical layers.
- "Logical / internal mapping" is associated with the arrows between the Conceptual/Logical and Internal layers.
- "Physical data independence" is associated with the arrows between the Conceptual/Logical and Internal layers.

At the bottom of the diagram, a red line is drawn across the text "physical data + logical data independence!".

Video player controls at the bottom show a progress bar at 22:34 / 47:03, a volume icon, and a play button.

Principles of Database Management (book)

Bart Baesens - 1 / 19



-  **Chapter 1 Fundamental Concepts of Database Management**
Bart Baesens
47:04
-  **Chapter 2 Architecture and Classification of DBMS**
Bart Baesens
44:04
-  **Chapter 3 Conceptual Data Modeling using EER and UML**
Bart Baesens
1:18:28
-  **Chapter 4 Organizational Aspects of Data Management**
Bart Baesens
31:50
-  **Chapter 5 Legacy Databases**
Bart Baesens
27:07
-  **Chapter 6 Relational Databases**
Bart Baesens
1:27:25

Chapter 1 Fundamental Concepts of Database Management

2,206 views

18 0 SHARE



Bart Baesens
Published on 24 Feb 2018

SUBSCRIBED 6.1K



Chapter 2 Architecture and Classification of DBMS
Bart Baesens
260 views



Chapter 1: Fundamental Concepts of Database Management
Bart Baesens
47:04

Data model
Definition: a clear description of the data concepts, their relationships, and the constraints that must be satisfied by the data.

On-Line Environment

- Playground includes:
 - MySQL for SQL querying (chapter 7)
 - MongoDB (chapter 11)
 - Neo4j Cypher (chapter 11)
 - Tree structure visualization (chapter 12)
- Playground can be locally installed as a Dockerfile
 - Available at [GitHub](#)
- Installing the playground, see [YouTube](#)
- Working with the playground, see [YouTube](#)

On-Line Environment: SQL

MySQL

SQL Playground

Database: purchase.sql

Write your SQL statement below and press "Run Query" to see the result.

```
1 SELECT PRODNR, SUM(QUANTITY) AS QUANTITY
2 FROM PO_LINE
3 GROUP BY PRODNR
4 HAVING SUM(QUANTITY) > 15
5
```

RUN QUERY ▶

2 row(s) returned

PRODNR (PO_LINE)	QUANTITY ()
0212	23
0668	16

Tables

po_line
product
purchase_order
supplier
supplies

Views

No views defined.

Reset

[Click here to reset the database to its initial state](#) (all your changes will be lost).

On-Line Environment: NoSQL

Neo4j

Using: bookclub

Write your statement below and press "Run" to see the result.

```
1 MATCH (r:Reader)--(:Book)--(:Genre {name:'romance'})
2 RETURN r.name
3
```

Run

Result

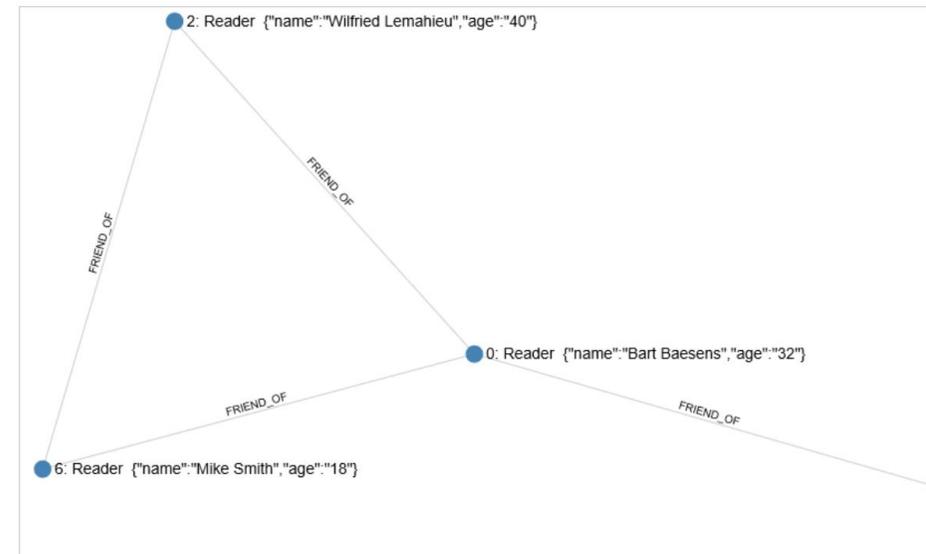
r.name

Elvis Presley

Mike Smith

Anne HatsAway

Result



Multiple Choice Quiz Tool

Test Yourself with the Database Quiz

Want to test yourself? Take a multiple choice quiz based on the multiple choice questions included in the book and check your database knowledge and retention.

Which chapters do you want to include?

Make sure to select at least one chapter.

[Select all](#), [Deselect all](#)

- | | |
|---|---|
| <input type="checkbox"/> Chapter 1: Fundamental Concepts of Database Management | <input type="checkbox"/> Chapter 11: NoSQL Databases |
| <input type="checkbox"/> Chapter 2: Architecture and Categorization of DBMSs | <input type="checkbox"/> Chapter 12: Physical File Organization and Indexing |
| <input type="checkbox"/> Chapter 3: Conceptual Data Modeling | <input type="checkbox"/> Chapter 13: Physical Database Organization |
| <input type="checkbox"/> Chapter 4: Organizational Aspects of Data Management | <input type="checkbox"/> Chapter 14: Basics of Transaction Management |
| <input type="checkbox"/> Chapter 5: Legacy Databases | <input type="checkbox"/> Chapter 15: Accessing Databases and Database APIs |
| <input type="checkbox"/> Chapter 6: Relational Databases: The Relational Model | <input type="checkbox"/> Chapter 16: Data Distribution and Distributed Transaction Management |
| <input type="checkbox"/> Chapter 7: Relational Databases: Structured Query Language (SQL) | <input type="checkbox"/> Chapter 17: Data Warehousing and Business Intelligence |
| <input type="checkbox"/> Chapter 8: Object Oriented Databases and Object Persistence | <input type="checkbox"/> Chapter 18: Data Integration, Data Quality and Data Governance |
| <input type="checkbox"/> Chapter 9: Extended Relational Databases | <input type="checkbox"/> Chapter 19: Big Data |
| <input type="checkbox"/> Chapter 10: XML Databases | <input type="checkbox"/> Chapter 20: Analytics |

Care for an extra challenge?

- Shuffle the answer options of correct/incorrect style questions across selected chapters

Let's go!

Multiple Choice Quiz Tool

Test Yourself with the Database Quiz

Using Cypher, how do you get a list of all movies Wilfried Lemahieu has liked, when he has given at least four stars?

- MATCH (b:User)-[:LIKES]-(m:Movie)
WHERE b.name = "Wilfried Lemahieu"
AND l.stars >= 4
RETURN m
- SELECT (b:User)--m:Movie
WHERE b.name = "Wilfried Lemahieu"
AND m.starts >= 4
- MATCH (b:User)-[:LIKES]-(m:Movie)
WHERE b.name = "Wilfried Lemahieu"
AND m.stars >=4
RETURN m
- MATCH (b:User)--(m:Movie)
WHERE b.name = "Wilfried Lemahieu"
AND l.stars >= 4
RETURN m

[Try another question](#) So far, you've answered 0 question(s) out of 2 correctly.

Sorry, that is **incorrect**. The **correct** answer is shown in blue.

Check up on the following chapter(s) to check your knowledge: 11.

Solutions Manual

- Solutions to all multiple choice and open questions
- Can be used for assignments, coursework and exam questions
- 180 pages
- Additional exercises and solutions will be made available on www.pdbmbook.com

Example Endorsements

- “This accessible, authoritative book **introduces the reader the most important fundamental concepts** of data management, while providing a practical view of recent advances. Both are essential for data professionals today.” – Foster Provost, Professor of Data Science & Information Systems, New York University, Stern School of Business
- “This guide to big and small data management addresses both fundamental principles and practical deployment. It reviews a range of databases and their relevance for analytics. The **book is useful to practitioners because it contains many case studies, links to open-source software, and a very useful abstraction of analytics** that will help them better choose solutions. It is important to academics because it promotes database principles which are key to successful and sustainable data science.” – Sihem Amer-Yahia, Laboratoire d’Informatique de Grenoble; Editor-in-Chief, The VLDB journal (International Journal on Very Large DataBases)
- “As we are entering a new technological era of intelligent machines powered by data-driven algorithms, understanding fundamental concepts of data management and their most current practical applications has become more important than ever. **This book is a timely guide for anyone interested in getting up to speed with the state of the art in database systems**, big data technologies, and data science. It is full of insightful examples and case studies with direct industrial relevance.” – Nesime Tatbul, Intel Labs and MIT

Example Endorsements

- “It is a pleasure to study this new book on database systems. The book offers a fantastically **fresh approach** to database teaching. The **mix of theoretical and practical contents** is almost perfect, the content is **up-to-date** and covers the recent ones, the examples are nice, and the database testbed provides an excellent way of understanding the concepts. Coupled with the authors ‘expertise, this book is an important addition to the database field.” – Arnab Bhattacharya, Indian Institute of Technology, Kanpur
- “Data science success stories and big data applications are only possible because of advances in database technology. This book provides both a **broad and deep** introduction to databases. It covers the different types of database systems (from relational to noSQL) and manages to bridge the gap between data modeling and the underlying basic principles. The book is highly recommended for anyone that wants to understand how modern information systems deal with ever-growing volumes of data.” – Wil van der Aalst, RWTH Aachen University

Media Coverage

Data Integration Vs. Data Quality: Friends or Foes?

Data Integration Vs. Data Quality: Friends or Foes?



By [Bart Baesens](#), [Seppe vanden Broucke](#), [Wilfried Lemahieu](#)
Advisor
Posted May 1, 2018 in [Data Analytics & Digital Technologies](#)

[This article is based on the authors' upcoming book, [Principles of Database Management: The Practical Guide to Storing, Managing and Analyzing Big and Small Data.](#)]

To SQL or not To SQL: that is the question!

◀ Previous post Next post ▶

Like 41 Share 41 G+ Share 18

Tags: [Databases](#), [NoSQL](#), [Relational Databases](#), [Scalability](#), [SQL](#)

This article looks at the emergence of the NoSQL movement and compares it to a traditional relational database.

DATA WITH INSIGHT LEADS YOU TO DECISIONS WITH CLARITY.

NYU STERN MS in Business Analytics

Data with Insight Leads you to Decisions with Clarity
NYU MS in Business Analytics

By [W. Lemahieu](#), [S. vanden Broucke](#), [B. Baesens](#), [KU Leuven](#)

This article is based upon our upcoming book [Principles of Database Management: The Practical Guide to Storing, Managing and Analyzing Big and Small Data](#), [www.pdbmbook.com](#) See also our corresponding YouTube channel with free video lectures.

Relational database systems (RDBMS) pay a lot of attention to data consistency and compliance with a formal database schema. New data or modifications to



Big Data Experts: Top Job Titles

Job overviews for information architect, database designer, data owner, data steward, database administrator and data scientist. Depending upon the size of the database and the company, multiple profiles may be merged into one job description.

Email Article Print Article Comment on this article

WEBINAR: On-Demand **How to Help Your Business Become an AI Early Adopter** [Watch →](#)

Posted May 1, 2018
By [Guest Author](#)

This article is based on an upcoming book, [Principles of Database Management: The Practical Guide to Storing, Managing and Analyzing Big and Small Data.](#)

[Submit Feedback ▶](#)
[More Articles ▶](#)

Data management entails the proper management of data as well as the corresponding data definitions or metadata. It aims at ensuring that (meta-) data is of good quality and thus a key resource for effective and efficient managerial decision making. Data quality (DQ) is often defined as 'fitness for use,' which

implies the relative nature of the concept.

Data that is of acceptable quality in one decision context may be perceived to be of poor quality in

A Database Perspective on Data Security



May 11, 2018

[Baesens](#)

Although security is often related to privacy, they are not synonymous. Data security can be defined as the set of policies and techniques to ensure the



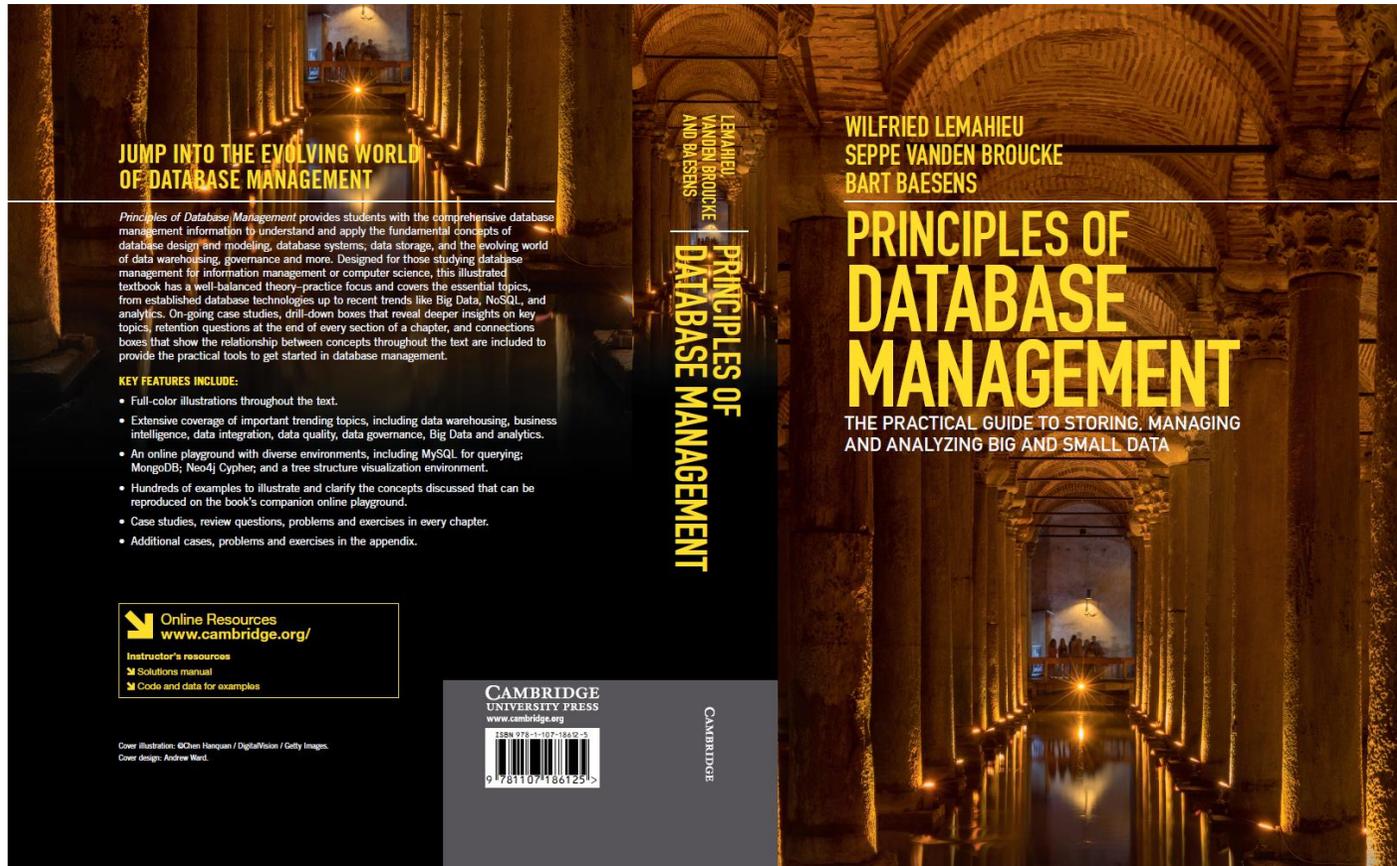
confidentiality, availability, and integrity of data at all times. Data privacy refers to the fact that the parties accessing and using the data do so only in ways that comply with the agreed-upon purposes of data use in their roles. These purposes can be expressed as part of a company's policy, but are also subject to legislation. In this way, several aspects of security can be considered as necessary instruments to guarantee data privacy.

More concretely, data security pertains to the following concerns:

Guaranteeing Data Availability: Ensures that the data is accessible to all authorized users and applications, even in the occurrence of partial system malfunctions. Many techniques exist to safeguard data by means of backup and/or replication. Examples are tape backup, hard disk backup, electronic vaulting, replication, and mirroring.

Authentication and Access Control: Refers to the tools and formats to express which users and applications

More Information?



www.pdbmbook.com