Undergraduate Topics in Computer Science

Undergraduate Topics in Computer Science (UTiCS) delivers high-quality instructional content for undergraduates studying in all areas of computing and information science. From core foundational and theoretical material to final-year topics and applications, UTiCS books take a fresh, concise, and modern approach and are ideal for self-study or for a one- or two-semester course. The texts are all authored by established experts in their fields, reviewed by an international advisory board, and contain numerous examples and problems. Many include fully worked solutions.

For further volumes:
www.springer.com/series/7592

Peter Lake · Paul Crowther

Concise Guide to Databases

A Practical Introduction

Foreword by Professor Richard Hill

Springer

Peter Lake
Sheffield Hallam University
Sheffield, UK

Paul Crowther
Sheffield Hallam University
Sheffield, UK

ISSN 1863-7310 ISSN 2197-1781 (electronic)
Undergraduate Topics in Computer Science
ISBN 978-1-4471-5600-0 ISBN 978-1-4471-5601-7 (eBook)
DOI 10.1007/978-1-4471-5601-7
Springer London Heidelberg New York Dordrecht

Library of Congress Control Number: 2013955488

Dedicated to our mate Andy McEwan

Paul and Peter

Foreword

From tablets of stone through to libraries of parchments; from paper-based files to the electronic era, there is not one aspect of modern business that has avoided the need to collect, collate, organize and report upon data. The proliferation of databases and database technologies within modern times, has now been further secured by the use of the Internet to enable database integration on a massive scale.

In amongst the innovation, the basic concepts remain. The need to organize—a topic that Codd reminded us could be best done by relational models—is now being challenged, as processor power and storage space become cheap and utility-like with the advent of Cloud Computing infrastructure. But a glance at the past does much to inform future thinking, and this book serves to prepare the foundations of a mature approach to using database technologies in the 21st Century.

In many cases, both established and emerging database technologies are readily available and free to use. As such they may appear free to implement, which fuels rapid adoption of technology that may not have been proven sufficiently, without the formal governance that other business norms might impose. This creates an exciting, risky, domain where commercial models can make or lose money depending upon how they embrace and realize the potential of the technology. Conversely, there is also an opportunity to solve problems when things go awry—and the accelerated innovation that we now witness, presents more opportunities and pitfalls, if we do not possess the requisite understanding of how databases should serve our needs.

Proficiency in the field of databases is a combination of technical understanding, conceptual knowledge and business acumen. All of these traits are underpinned by education, and the need for professionals to continually update their knowledge. Since professionals not only face the challenge of when to introduce a technology, but also when *not* to adopt, it is important to understand the impact of failure as well as success. This book takes readers through the essential basics, before charting a path towards technical skill acquisition in the real-life context of business.

Professor Richard Hill
Head of Subject, Computing and Mathematics
University of Derby, Derby, UK
June 2013

About Richard Hill:

Richard Hill, PhD, is Professor of Intelligent Systems and Head of Department in the School of Computing and Mathematics, at the University of Derby, UK. Professor Hill has published widely in the areas of multi agent systems, computational intelligence, intelligent cloud computing and emerging technologies for distributed systems, and has organised a number of international conferences. Latterly, Professor Hill has edited and co-authored several book collections and textbooks, including ‘Guide to Cloud Computing: Principles and Practice’, published by Springer UK.

Preface

Overview and Goals

Databases are not new and there are many text books available which cover various database types, especially relational. What is changing, however, is that Relational Database Management Systems (RDBMS) are no longer the only database solution. In an era where Big Data is the current buzzword and Data Scientists are tomorrow's big earners, it is important to take a wider view of database technology.

Key objectives for this book include:

- Present an understanding of the key technologies involved in Database Systems in general and place those technologies in an historic context
- Explore the potential use of a variety of database types in a business environment
- Point out areas for further research in a fast moving domain
- Equip readers with an understanding of the important aspects of a database professional's job
- Provide some hands-on experience to further assist in the understanding of the technologies involved

Organisation and Features

This book is organised into three parts:

- Part I introduces database concepts and places them in both a historic and business context;
- Part II provides insights into some of the major database types around today and also provides some hands-on tutorials in the areas concerned;
- Part III is devoted to issues and challenges which face Database Professionals.

Target Audiences

This book has been written specifically to support the following audiences:

Advanced undergraduate students and *postgraduate students* should find the combination of theoretical and practical examples database usage of interest. We imagine this text would be of particular relevance for modern Computer Science,

Software Engineering, and Information Technology courses. However, any course that makes reference to databases, and in particular to the latest developments in computing will find this text book of use. As such, *University Instructors* may adopt the book as a core text.

Especially in Part II, this book adopts a learning-by-doing approach, with the extensive worked examples explaining how to use the variety of databases available to address today's business needs. Practising *Database Professionals*, and *Application Developers* will also be able to use this book to review the current state of the database domain.

Suggested Uses

A Concise Guide to Databases can be used as a solid introduction to the concept of databases. The book is suitable as both a comprehensive introduction to databases, as well as a reference text as the reader develops their skills and abilities through practical application of the ideas. For *University Instructors*, we suggest the following programme of study for a twelve-week semester format:

- Weeks 1–3: Part I
- Weeks 4–8: Part II
- Weeks 9–12: Part III
- Week 12: Assessment

Review Questions

Each chapter concludes with a set of review questions that make specific reference to the content presented in the chapter, plus an additional set of further questions that will require further research. The review questions are designed in such a way that the reader will be able to tackle them based on the chapter contents. They are followed by discussion questions, that often require research, extended reading of other material or discussion and collaboration. These can be used as classroom discussion topics by tutors or used as the basis of summative assignments.

Hands-on Exercises

The technology chapters include extended hands-on exercises. Readers will then progressively engage in more complex activities, building skills and knowledge along the way. Such an approach ensures that a solid foundation is built before more advanced topics are undertaken. Some of the material here is Open Source, whilst some examples are Oracle specific, but even these latter can be applied to other SQL databases.

Chapter Summary

A brief summary of each of the twelve chapters is as follows:

Chapter 1: Data is the lifeblood of all business systems and we place the use of data in its historical context and review some of the key concepts in handling data.

Chapter 2: Provides an examination of the way that data has been handled throughout history, using databases of a variety of types.

Chapter 3: Considers how we actually store data. Turning information into a series of 1s and 0s is at the heart of every current database system and so an understanding of issues like physical storage and distribution are important concepts to understand.

Chapter 4: The *de facto* standard database solution is, without doubt, the relational database. In this chapter we look at how RDBMS works and provide worked examples.

Chapter 5: The NoSQL movement is still relatively new. Databases which store data without schemas and which do not necessarily provide transactional security may seem like a bad idea to experienced relational database practitioners, but these tools do certainly have their place in today's data rich society. We review the area in general and then look at specific examples of a Column-based and a Document-based database, with hands-on tutorials for each.

Chapter 6: Look at many leading database vendors' web sites and you will see that we are in the Big Data era. We explore what this actually means and, using a tutorial, review one of the key concepts in this era—that of MapReduce.

Chapter 7: Object databases were once thought of as the next important design for databases. When used by developers using Object programming they can seem very appealing still. There are half-way house solutions also available—Oracle, for example, has an Object-Relational option. We explore this area with more tutorial material.

Chapter 8: Reading data from disk is far slower than reading from RAM. Computing technologies now exist that can allow databases to run entirely in memory, making for very rapid data processing. These databases may well become the norm as RAM becomes cheaper and hard disk technology becomes less able to improve in performance.

Chapter 9: Once you have designed your database, especially when supporting a web- or cloud-based solution, you need to be sure that it can grow if the business that the application supports is successful. Scalability is about ensuring that you can cope with many concurrent users, or huge amounts of data, or both.

Chapter 10: Once your system is built, you need to be able to have it available for use permanently (or as close to permanently as can be achieved within the financial resources at your disposal). We review key concepts such as back-up, recovery, and disaster recovery.

Chapter 11: For a DBA the dreaded phone call is "my report is running very slowly". For a start, what is mean by slowly? What is the user used to? Then there is the problem of how you establish where the problem is—is it hardware related? Or Network related? At the Server or Client end? The solution may be indexes, or

partitions: we review a variety of performance related techniques. We include some tutorial material which explores some performance management tools.

Chapter 12: Data is one of an organisation's most important assets. It needs to be protected from people wanting to either take it, or bring the system down. We look at physical and software-related weaknesses and review approaches to making our databases secure.

Sheffield, UK

Peter Lake
Paul Crowther

Contents

Part II Database Types