Beginning SQL Queries

From Novice to Professional

Clare Churcher

Apress®

Beginning SQL Queries: From Novice to Professional

ISBN-13 (pbk): 978-1-59059-943-3

ISBN-10 (pbk): 1-59059-943-8

ISBN-13 (electronic): 978-1-4302-0550-0

ISBN-10 (electronic): 1-4302-0550-4

Printed and bound in the United States of America 9 8 7 6 5 4 3 2 1

Lead Editor: Jonathan Gennick
Technical Reviewer: Darl Kuhn
Editorial Board: Clay Andres, Steve Anglin, Ewan Buckingham, Tony Campbell, Gary Cornell, Jonathan Gennick, Matthew Moodie, Joseph Ottinger, Jeffrey Pepper, Frank Pohlmann, Ben Renow-Clarke, Dominic Shakeshaft, Matt Wade, Tom Welsh
Project Manager: Beth Christmas
Copy Editors: Marilyn Smith, Kim Wimpsett
Associate Production Director: Kari Brooks-Copony
Production Editor: Ellie Fountain
Compositor: Susan Glinert
Proofreaders: Linda Seifert, Liz Welch
Indexer: Broccoli Information Management
Artist: April Milne
Cover Designer: Kurt Krames
Manufacturing Director: Tom Debolski

Distributed to the book trade worldwide by Springer-Verlag New York, Inc., 233 Spring Street, 6th Floor, New York, NY 10013. Phone 1-800-SPRINGER, fax 201-348-4505, e-mail orders-ny@springer-sbm.com, or visit http://www.springeronline.com.

For information on translations, please contact Apress directly at 2855 Telegraph Avenue, Suite 600, Berkeley, CA 94705. Phone 510-549-5930, fax 510-549-5939, e-mail info@apress.com, or visit http://www.apress.com.

Apress and friends of ED books may be purchased in bulk for academic, corporate, or promotional use. eBook versions and licenses are also available for most titles. For more information, reference our Special Bulk Sales–eBook Licensing web page at http://www.apress.com/info/bulksales.

The information in this book is distributed on an “as is” basis, without warranty. Although every precaution has been taken in the preparation of this work, neither the author(s) nor Apress shall have any liability to any person or entity with respect to any loss or damage caused or alleged to be caused directly or indirectly by the information contained in this work.

To Mark and Ali

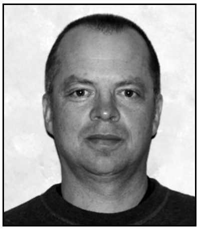
Contents at a Glance

Contents

About the Author

CLARE CHURCHER holds a Ph.D. in physics and has designed several databases for a variety of large and small projects. She is a senior academic in the Applied Computing Group at Lincoln University, where she recently won an Excellence in Teaching Award for her contribution to developing and presenting courses in analysis and design, databases, and programming. She has supervised more than 70 undergraduate projects designing databases for small projects.

About the Technical Reviewer



DARL KUHN is a senior database administrator with Sun Microsystems. Before joining Sun, his work as a consultant ranged from database administration to custom application development. Darl is a coauthor of *RMAN Recipes for Oracle Database 11g* and *Oracle RMAN Pocket Reference*. He is an affiliate professor at Regis University, where he teaches database courses for the department of computer information technology. Darl currently lives near Aguilar, Colorado, with his wife, Heidi, and two daughters, Lisa and Brandi.

Acknowledgments

Expecting your friends and family to put up with you writing a book is a tough call. Writing a second book the following year is really pushing their patience. So I am very indebted to my family and colleagues for putting up with me again. Special thanks to my two “first draft readers.” My husband, Neville Churcher, and my friend and colleague Theresa McLennan both gave me many valuable suggestions on the first attempts at every chapter, and always did it in Clare time (i.e., now!) and with very good grace. Thanks also to all my good friends in the Applied Computing Group at Lincoln University, especially Alan McKinnon for his advice on Chapter 9.

Many thanks to my editor Jonathan Gennick for his insight, expertise, and encouragement, and to both Jonathan and Donna for showing us around their lovely hometown of Munising. Beth Christmas, my Apress project manager, has been a pleasure to work with, and Dahl Kuhn has been a most careful and expert technical reviewer. Thanks to you both.

Introduction

As a query language, SQL is really quite small and should be easy to learn. A few basic ideas and a handful of keywords allow you to tackle a huge range of queries. However, many users often find themselves completely stumped when faced with a particular problem. You may find yourself in that group. It isn't really a great deal of help for someone to say, "This is how I would do it." What you need is a variety of ways to get started on a tricky problem. Once you have made a start on a query, you need to be able to check, amend, and refine your solution until you have what you need.

Two-Pronged Approach

Throughout this book, I approach different types of queries from two directions. The two approaches have their roots in relational algebra and calculus. Don't be alarmed though—I won't be delving into any complex mathematics. However, understanding a question and developing an appropriate SQL query do require logical thinking and precise definitions. The relational algebra and calculus approaches are both useful ways to grasp the logic and precision that are required to get accurate results.

The first approach, which has its roots in relational algebra, looks at *how* tables need to be manipulated in order to retrieve the subset of data you require. I describe the different types of operations that you can perform on tables, including joins, intersections, selections, and so on, and explain how to decide which might help in particular situations. Once you understand what operations are needed, translating them into SQL is relatively straightforward.

The second approach is what I use when I just can't figure out which operations will give me the required results. This approach, based on relational calculus, lets you describe *what* an expected row in your result might be like; that is, what conditions it must obey. By looking at the data, it is surprisingly easy to develop a semiformal description of what a "correct" retrieved row would be like (and, by implication, how you would recognize an "incorrect" row). Because SQL was originally based on relational calculus, translating this semiformal description into a working query is particularly straightforward.

I am always surprised at which approach my students take when confronting a new problem. Some will instantly see the algebra operations that are needed; others will find the calculus approach more obvious. The choice of approach changes from query to query, from person to person, and (I suspect) from day to day. Having more than one way to get started means you are less likely to be completely baffled by a new problem.

Who This Book Is For

This book is for anyone who has a well-designed relational database and needs to extract some information from it. You might have noticed in the previous sentence that the database must be "well designed." I can't overemphasize this point. If your database is badly designed, it will not be able to store accurate and consistent data, so the information your queries retrieve will always be prone to inaccuracies. If you are looking to design a database from scratch, you should read my first book, *Beginning Database Design* (Apress, 2007). The final chapter of this book outlines a few common design problems you are likely to come across and gives some advice about how to mitigate the impact or correct the problem.

For this book, you do not need any theoretical knowledge of relational theory, as I will explain the relevant issues as they come up. The first chapter gives a brief overview of relational database theory, but it will help if you have had some experience working with databases with a few or more tables.

Objective of This Book

In this book, you will be introduced to all the main techniques and keywords needed to create SQL queries. You will learn about joins, intersections, unions, differences, selection of rows, and projection of columns. You will see how to implement these ideas in different ways using simple and nested queries, and you will be introduced to a variety of aggregate functions and summary techniques. You can try out what you learn using the sample data provided through the Apress web page for this book (`http://www.apress.com/book/view/1590599438`). There you will find the Access database used for the examples in the book and some scripts to create the database on a number of other platforms.

Most important of all, you will learn different ways to get started on a troublesome problem. In almost all cases, there are several different ways to express a query. My objective is, for any particular situation, to provide you with a method of attack that matches your psyche and mood (just kidding).