

Learning PostgreSQL 10

Second Edition

A beginner's guide to building high-performance PostgreSQL database solutions

Salahaldin Juba
Andrey Volkov

Packt>

BIRMINGHAM - MUMBAI

Learning PostgreSQL 10

Second Edition

Copyright © 2017 Packt Publishing

All rights reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without the prior written permission of the publisher, except in the case of brief quotations embedded in critical articles or reviews.

Every effort has been made in the preparation of this book to ensure the accuracy of the information presented. However, the information contained in this book is sold without warranty, either express or implied. Neither the authors, nor Packt Publishing, and its dealers and distributors will be held liable for any damages caused or alleged to be caused directly or indirectly by this book.

Packt Publishing has endeavored to provide trademark information about all of the companies and products mentioned in this book by the appropriate use of capitals. However, Packt Publishing cannot guarantee the accuracy of this information.

First published: November 2015

Second Edition: November 2017

Production reference: 1231117

Published by Packt Publishing Ltd.

Livery Place

35 Livery Street

Birmingham

B3 2PB, UK.

ISBN 978-1-78839-201-3

www.packtpub.com

Credits

Author

Salahaldin Juba
Andrey Volkov

Copy Editors

Tasneem Fatehi
Safis Editing

Reviewers

Dr. Isabel M.D. Rosa
Sheldon E. Strauch

Project Coordinator

Manthan Patel

Commissioning Editor

Amey Varangaonkar

Proofreader

Safis Editing

Acquisition Editor

Varsha Shetty

Indexer

Rekha Nair

Content Development Editor

Aaryaman Singh

Graphics

Tania Dutta

Technical Editor

Dinesh Chaudhary

Production Coordinator

Melwyn Dsa

About the Authors

Salahaldin Juba has over than a decade of experience in the industry and academia, with a focus on database development for large-scale and enterprise applications. He holds a master's degree of science in environmental management with a distinction, and a bachelor's degree of engineering in computer systems. He is also a **Microsoft Certified Solution Developer (MCSD)**.

He has worked mainly with SQL server, PostgreSQL, and Greenplum databases. He has developed applications for scientific communities to process GIS information in a distributed manner, and he has participated in many international projects and standards for image processing during his work in the academic sector.

As a software engineer, he works mainly with defining ETL processes with external parties, defining software solution, promoting SQL best practices, designing OLTP and OLAP application, scouting and evaluating new technologies, and providing training and consultation services.

I would like to express my deepest gratitude to my colleague Andrey Volkov for making this work possible. Also, I would like to thank all the people who provided support, especially the team at Packt for their great support and feedback with proofreading, design, comments, and remarks. I would also like to thank my family for their support despite all of the time that I had to devote to this book over them. Finally, very warm and deep thanks to my late father, Ikrayem Juba, for his utmost support, help, and guidance.

Andrey Volkov has studied information systems in banking. He started his career as a financial analyst in a commercial bank. Using databases as a main tool in his work, he realized that querying the database directly and mastering SQL is much more efficient for ad hoc analysis than using any visual reporting software. He joined the data warehouse team, and after some time, he lead the team by taking the position of the data warehouse architect.

He has worked mainly with Oracle and used to develop logical and physical models of financial and accounting data, implement them in the database, develop ETL processes, and perform analytical tasks. He was also responsible for teaching the users how to use the data warehouse and BI tools. SQL training was also a part of that work.

After 10 years in the financial sector, he changed his field and now works as a senior database developer in a telecommunications company. Here, he works mainly with PostgreSQL databases, being responsible for data modeling and implementing physical data structures, developing stored procedures, integrating databases with other software components, and developing a data warehouse.

Having a lot of experience with both Oracle and PostgreSQL--the leading commercial and one of the most advanced open source RDBMSes--he is able to compare them and recognize and evaluate the key advantages of both of them. Working as a software developer and implementing different kinds of database applications, as well as working as a business analyst and using databases as a tool for analysis, let him learn and understand different database features in different use cases. This experience made him able and willing to work on this book.

I would like to thank my wife and son for their support and for letting me work on weekends and evenings. Big thanks to the editors team for their support, guidance, and organization. And most of all, I would like to thank the main author of the book, Salahaldin Juba, who invited me to work on the book, introduced me to the team, and, in fact, has done most of the work.

About the Reviewers

Dr. Isabel M.D. Rosa has been a Marie Skłodowska-Curie Research Fellow since May 2016 at the German Centre for Integrative Biodiversity Research (iDiv). Born in Lisbon, Portugal, in 1986 she holds a BSc in forestry engineering (2007), an MSc in natural resources management (2009) from the University of Lisbon, and a PhD in computational ecology (2013) from Imperial College London in the United Kingdom. She is the PI of the research project *Using Land Cover Change Models to Address Important Conservation Issues* funded by H2020-MSCA-IF-2015. She also has contributed to two international projects as a team member since 2013, including one European-funded project supported with 1.5 M € budget (Terragenesis, ERC-2011-StG_20101109). She is the author of 15 peer-reviewed publications in scientific journals such as *Nature Ecology and Evolution*, *Current Biology* and *Global Change Biology*, accumulating 252 citations (Google citations, October 2017), H index = 8. During her academic career, she has acquired several skills, such as statistical analysis, programming (R, C++, and Python), working with geographic information systems (ArcGIS and QGIS), and creating databases (PostgreSQL/PostGIS and SQLServer). She also reviewed the book *Learning PostgreSQL*, also by Packt.

Sheldon Strauch is a 23-year veteran of software consulting at companies such as IBM, Sears, Ernst & Young, and Kraft Foods. He has a bachelor's degree in business administration and leverages his technical skills to improve business' self-awareness. His interests include data gathering, management, and mining; maps and mapping; business intelligence; and application of data analysis for continuous improvement. He is currently focused on development of end-to-end data management and mining at Enova International, a financial services company located in Chicago. In his spare time, he enjoys the performing arts, particularly music, and traveling with his wife, Marilyn.

www.PacktPub.com

For support files and downloads related to your book, please visit www.PacktPub.com. Did you know that Packt offers eBook versions of every book published, with PDF and ePub files available? You can upgrade to the eBook version at www.PacktPub.com and as a print book customer, you are entitled to a discount on the eBook copy. Get in touch with us at service@packtpub.com for more details. At www.PacktPub.com, you can also read a collection of free technical articles, sign up for a range of free newsletters and receive exclusive discounts and offers on Packt books and eBooks.



<https://www.packtpub.com/mapt>

Get the most in-demand software skills with Mapt. Mapt gives you full access to all Packt books and video courses, as well as industry-leading tools to help you plan your personal development and advance your career.

Why subscribe?

- Fully searchable across every book published by Packt
- Copy and paste, print, and bookmark content
- On demand and accessible via a web browser

Customer Feedback

Thanks for purchasing this Packt book. At Packt, quality is at the heart of our editorial process. To help us improve, please leave us an honest review on this book's Amazon page at <https://www.amazon.com/dp/1788392019>. If you'd like to join our team of regular reviewers, you can email us at customerreviews@packtpub.com. We award our regular reviewers with free eBooks and videos in exchange for their valuable feedback. Help us be relentless in improving our products!

Table of Contents

Preface	1
Chapter 1: Relational Databases	9
Database management systems	9
A brief history	10
Database categories	10
The NoSQL databases	11
The CAP theorem	11
NoSQL motivation	11
Key-value databases	12
Columnar databases	12
Document databases	12
Graph databases	13
Relational and object relational databases	13
ACID properties	14
The SQL language	14
Relational model concepts	15
Relation	15
Tuple	16
NULL value	16
Attribute	17
Constraint	18
Domain integrity constraint	18
Entity integrity constraint	19
Referential integrity constraints	20
Semantic constraints	21
Relational algebra	22
The select and project operations	23
The rename operation	25
The set theory operations	26
The cartesian product operation	26
Data modeling	28
Data model perspectives	28
The entity-relation model	29
Sample application	30
Entities, attributes, and keys	30
Mapping ER to relations	34
UML class diagrams	34

Summary	35
Chapter 2: PostgreSQL in Action	37
An overview of PostgreSQL	37
PostgreSQL history	38
The advantages of PostgreSQL	38
Business advantages of PostgreSQL	38
PostgreSQL user advantages	39
PostgreSQL applications	40
Success stories	41
Forks	41
PostgreSQL architecture	42
The PostgreSQL community	44
PostgreSQL capabilities	44
Replication	44
Security	46
Extension	46
NoSQL capabilities	48
Foreign data wrappers	49
Performance	51
Installing PostgreSQL	52
Installing PostgreSQL using Advanced Package Tool	53
Client installation	53
Server installation	54
Basic server configuration	56
Installing PostgreSQL on Windows	58
The PostgreSQL clients	60
The psql client	61
psql advanced settings	62
PostgreSQL utility tools	65
Summary	66
Chapter 3: PostgreSQL Basic Building Blocks	69
Database coding	69
Database naming conventions	70
PostgreSQL identifiers	71
Documentation	71
Version control system	72
Database migration tool	72
PostgreSQL objects hierarchy	73
Template databases	73
User databases	74
