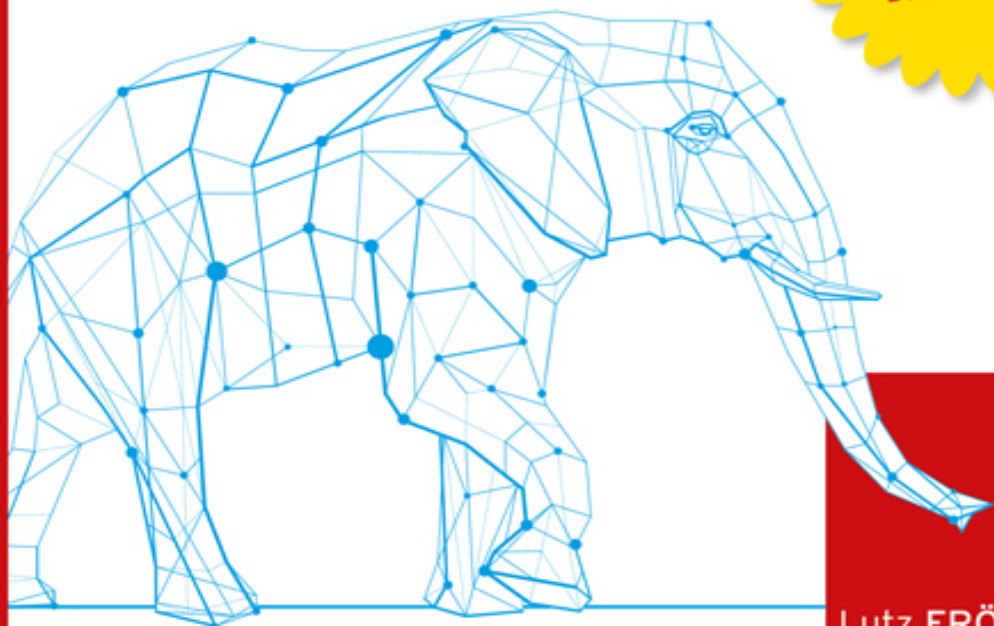


**Update
inside**



Lutz FRÖHLICH

PostgreSQL

**PRAXISBUCH FÜR
ADMINISTRATOREN
UND ENTWICKLER**



Zu Version: PostgreSQL 14

HANSER

HANSER

Lutz Fröhlich

PostgreSQL

Praxisbuch für Administratoren und Entwickler

**Ihr Plus – digitale
Zusatzinhalte!**

Auf unserem Download-Portal
finden Sie zu diesem Titel
kostenloses Zusatzmaterial.

Geben Sie auf [plus.hanser-
fachbuch.de](https://plus.hanser-fachbuch.de) einfach diesen
Code ein:

plus-92Fgh-ci4R7

Der Autor:

Lutz Fröhlich

lutz@lutzfroehlich.de

Alle in diesem Buch enthaltenen Informationen, Verfahren und Darstellungen wurden nach bestem Wissen zusammengestellt und mit Sorgfalt getestet. Dennoch sind Fehler nicht ganz auszuschließen. Aus diesem Grund sind die im vorliegenden Buch enthaltenen Informationen mit keiner Verpflichtung oder Garantie irgendeiner Art verbunden. Autor und Verlag übernehmen infolgedessen keine juristische Verantwortung und werden keine daraus folgende oder sonstige Haftung übernehmen, die auf irgendeine Art aus der Benutzung dieser Informationen – oder Teilen davon – entsteht.

Ebenso übernehmen Autor und Verlag keine Gewähr dafür, dass beschriebene Verfahren usw. frei von Schutzrechten Dritter sind. Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Buch berechtigt deshalb auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Bibliografische Information der Deutschen Nationalbibliothek:

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Dieses Werk ist urheberrechtlich geschützt.

Alle Rechte, auch die der Übersetzung, des Nachdruckes und der Vervielfältigung des Buches, oder Teilen daraus, vorbehalten. Kein Teil des Werkes darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form (Fotokopie, Mikrofilm oder ein anderes Verfahren) – auch nicht für Zwecke der Unterrichtsgestaltung – reproduziert oder unter Verwendung elektronischer Systeme verarbeitet, vervielfältigt oder verbreitet werden.

© 2022 Carl Hanser Verlag München, www.hanser-fachbuch.de

Lektorat: Sylvia Hasselbach

Copy editing: Petra Kienle, Fürstfeldbruck

Umschlagdesign: Marc Müller-Bremer, München, www.rebranding.de

Umschlagrealisation: Max Kostopoulos

Titelmotiv: © [shutterstock.com/yosart](https://www.shutterstock.com/yosart)

Gesamtherstellung: Eberl & Koesel, Altusried-Krugzell
Ausstattung patentrechtlich geschützt. Kösel FD 351, Patent-Nr. 0748702

Print-ISBN 978-3-446-46929-7

E-Book-ISBN 978-3-446-47315-7

E-Pub-ISBN: 978-3-446-47510-6

Update inside.

Mit unserem kostenlosen Update-Service zum Buch erhalten Sie aktuelle Infos zu den neuen Versionen von PostgreSQL.

Und so funktioniert es:

1. Registrieren Sie sich unter:

www.hanser-fachbuch.de/postgresql-update

2. Geben Sie diesen Code ein:

2WHK-SeoTr-bF5x

Der Update-Service läuft bis Mai 2024.

Als registrierter Nutzer werden Sie in diesem Zeitraum persönlich per E-Mail informiert, sobald ein neues Buch-Update zum Download verfügbar ist.

Wenn Sie Fragen haben, wenden Sie sich gerne mit dem Betreff „PostgreSQL“ an:

update-inside@hanser.de

Inhalt

Titelei

Impressum

Inhalt

1 Einführung und Geschichte

1.1 Die Geschichte von PostgreSQL

1.2 Verwendete Versionen

1.3 Konventionen

1.4 Software und Skripte

1.5 Update inside

2 Installation mit Paketen und aus dem Quellcode

2.1 Paketinstallation

2.1.1 Paketinstallation unter Linux

2.1.2 Paketinstallation unter Windows

2.1.3 Paketinstallation unter macOS

2.2 Installation aus dem Quellcode

2.2.1 Installation aus dem Quellcode unter Linux

2.2.2 Installation aus dem Quellcode unter Windows

2.2.3 Installation aus dem Quellcode unter macOS

2.3 Erste Schritte

3 Upgrades und Versionen

3.1 Upgrade mit pg_dumpall

3.2 Upgrade mit pg_upgrade

3.3 Upgrade mit logischer Replikation

3.4 Migration nach Native Partitioning

3.5 Regressionstests

4 Die Architektur von PostgreSQL

4.1 Überblick

4.2 Memory und Prozesse

4.2.1 Hintergrundprozesse

4.2.2 Shared Memory

4.3 VACUUM

4.4 Cluster, Datenbanken und Tabellen

5 Server und Datenbanken administrieren

5.1 Parametereinstellungen

5.1.1 Einstellungen im Betriebssystem

5.1.2 Cluster-Einstellungen

5.1.3 Gebietsschema und Zeichensatz

5.2 Datenbanken verwalten

5.3 Konkurrenz

5.4 Die WAL-Archivierung einschalten

5.5 Wartungsaufgaben

5.5.1 VACUUM

5.5.2 ANALYZE

5.6 Nützliche Skripte und Hinweise

5.6.1 Eine Passwortdatei verwenden

5.6.2 Welche Parameter sind Nicht-Standard?

5.6.3 Eine Session killen

5.6.4 Eine Tabelle nach Excel kopieren

5.6.5 Die Datei .psqlrc

5.6.6 Einen WAL-Switch manuell auslösen

5.6.7 Die PostgreSQL-Server-Logdatei in eine Tabelle laden

5.6.8 Automatisches Rotieren von Logdateien

5.6.9 Nicht verwendete Indexe identifizieren

5.6.10 Microsoft Excel als Datenbank-Client

[5.6.11 Den Inhalt der Kontrolldatei ausgeben](#)

[5.6.12 Platzverbrauch von Tabellen](#)

[5.6.13 Die Anzahl von Verbindungen begrenzen](#)

[5.6.14 Tabellen und Indexe in einen anderen Tablespace legen](#)

[5.6.15 Temporäre Dateien verwalten](#)

[5.6.16 Lang laufende SQL-Anweisungen](#)

[5.7 Beispielschemata](#)

[6 Neue Features](#)

[6.1 JSONB Subscripting](#)

[6.2 B-Tree Bottom-up Deletion](#)

[6.3 Pipeline Queries](#)

[6.4 Große Transaktionen für logische Replikation](#)

[6.5 Query-ID für SQL-Anweisungen](#)

[6.6 Sonstige Verbesserungen und Erweiterungen](#)

7 Sicherung und Wiederherstellung

7.1 Online-Sicherung mit Point-in-time-Recovery

7.2 Offline-Sicherung auf Dateisystemebene

7.3 SQL Dump

8 Sicherheit und Überwachung

8.1 Sicherheit

8.1.1 Rollen und Privilegien

8.1.2 Authentifizierung und Zugangskontrolle

8.1.3 Passwortmigration

8.1.4 Rechteverwaltung

8.1.5 Sichere Verbindungen

8.1.6 Out-of-the-box-Sicherheit

8.1.7 Hacker-Attacken abwehren

8.2 Überwachung

8.2.1 Auditing

8.2.2 Monitoring

9 Logical Decoding

9.1 Logical Decoding mit SQL-Funktionen

9.2 Logical Decoding mit Java als Consumer

10 Replikation und Streaming

10.1 Physische Replikation

10.1.1 Vorbereitung und Planung

10.1.2 Konfiguration und Aktivierung

10.1.3 Kaskadenförmige Replikation

10.1.4 Hot Standby

10.1.5 Synchrone Replikation

10.1.6 Die Replikation überwachen

10.1.7 Failover und Switchover

10.2 Logische Replikation

11 Indexe effektiv einsetzen

11.1 B-Tree-Index

11.2 Block-Range-Index (BRIN)

11.3 Hash-Index

11.4 Generalized Inverted Index (GIN)

11.5 Generalized Search Tree-Index (GiST)

11.6 Expression-Index

11.7 Partieller Index

11.8 Individueller Index

11.9 Indexe und Parallelität

12 Textverarbeitung

12.1 Funktionen, Operatoren und Konfigurationen

12.2 Praktische Textverarbeitung

13 Performance Tuning

13.1 Out-of-the-box-Tuning

[13.1.1 Goldene Regeln für neue Server und Datenbanken](#)

[13.1.2 Das Utility pgTune](#)

[13.1.3 Optimierung von Memory-Parametern](#)

[13.2 Performance-Analyse](#)

[13.2.1 Analyse mit dem Statistics Collector](#)

[13.2.2 Der Background Writer](#)

[13.2.3 Analyse mit pgstatspack](#)

[14 Optimierung von SQL-Anweisungen](#)

[14.1 Ausführungsschritte](#)

[14.2 Der SQL-Optimizer](#)

[14.3 Statistiken und Histogramme](#)

[14.4 Zugriffsmethoden](#)

[14.5 Join-Methoden](#)

[14.6 SQL-Optimierung](#)

[14.6.1 Der EXPLAIN-Befehl](#)

[14.6.2 Ausführungspläne verstehen und optimieren](#)

[14.6.3 Parallelisierung von SQL-Ausführungen](#)

[14.6.4 Logging von SQL-Anweisungen](#)

[15 Einsatz großer Datenbanken](#)

[15.1 Skalierung großer Datenbanken](#)

[15.2 Partitionierung von Tabellen](#)

[15.3 Paralleles SQL](#)

[15.4 Materialized Views](#)

[15.5 Indexe für große Tabellen](#)

[16 PostGIS](#)

[16.1 PostGIS und PostgreSQL](#)

[16.2 PostGIS installieren](#)

[16.2.1 Paketorientierte Installation](#)

[16.2.2 Installation aus dem Quellcode](#)

16.3 Erste Schritte mit PostGIS

16.4 PostGIS in der Praxis anwenden

17 Applikationen für PostgreSQL entwickeln

17.1 Applikationsdesign

17.2 Entwicklungswerkzeuge

17.3 PostgreSQL Extensions

18 Erweiterungen und Module

18.1 Datentypen

18.2 Funktionen und Sprachen

18.2.1 SQL-Funktionen

18.2.2 Funktionen mit prozeduralen Programmiersprachen

18.2.3 C-Funktionen

18.3 Operatoren

18.4 Das Extension-Netzwerk

18.4.1 Extensions entwickeln

18.4.2 Extensions publizieren

19 PL/pgSQL-Funktionen und Trigger

19.1 PL/pgSQL-Funktionen

19.1.1 Abfragen und Resultsets

19.1.2 Cursor verwenden

19.1.3 DML-Anweisungen

19.1.4 Dynamische SQL-Anweisungen

19.1.5 Fehlerbehandlung

19.2 Trigger

20 Embedded SQL (ECPG)

21 Java-Programmierung

21.1 Eine Entwicklungsumgebung einrichten

21.2 Verarbeitung von Resultsets

21.3 DML-Anweisungen und Transaktionen

21.4 Bindevariablen verwenden

21.5 Java und Stored Functions

21.6 Large Objects

21.7 JDBC-Tracing

22 Die C-Library libpq

22.1 Die Entwicklungsumgebung einrichten

22.2 Programme mit libpq erstellen

23 PHP-Applikationen

23.1 Installation und Konfiguration

23.2 Applikationen mit PHP entwickeln

23.3 Die PDO-API

24 REST API

25 Client-Programmierung mit Perl-DBI

25.1 SELECT-Anweisungen und Resultsets

25.2 DML-Anweisungen

25.3 Bindevariablen verwenden

25.4 Fehlerbehandlung und Tracing

25.5 Nützliche Skripte und Beispiele

25.5.1 Mehrere Server abfragen

25.5.2 Parallele Verbindungen

25.5.3 Large Objects verarbeiten

25.5.4 Asynchrone Abfragen

25.5.5 Datenbanken vergleichen

26 Programmierung mit Python

26.1 Client-Programmierung mit Python

26.2 Server-Programmierung mit PL/Python

27 Data Science und Machine Learning

27.1 Werkzeuge

27.2 Einführung in Data Science

27.3 Ein Beispiel

27.4 Daten verwalten

27.5 Data Cleaning

27.6 Daten analysieren

27.6.1 SQL-Funktionen

27.6.2 Analysen durchführen

27.7 Stimmungslagenanalysen

28 PostgreSQL in die IT-Landschaft einbinden

28.1 Features und Funktionen

28.2 Datensicherheit und Wiederherstellung

28.3 Disaster Recovery

28.4 Überwachung

28.5 Administrierbarkeit

28.6 Verfügbarkeit

28.7 Datensicherheit und Auditing

28.8 Performance und Skalierbarkeit

28.9 Schnittstellen und Kommunikation

28.10 Support

28.11 Fazit

29 Migration von MySQL-Datenbanken

29.1 Unterschiede zwischen MySQL und PostgreSQL

29.2 Eine Migration durchführen

30 Von Oracle nach PostgreSQL migrieren

30.1 Die Migration planen

30.2 Unterschiede zwischen Oracle und PostgreSQL

30.2.1 Unterschiede der Datentypen

[30.2.2 Syntaktische und logische Unterschiede](#)

[30.2.3 Steigerung der Kompatibilität von PostgreSQL](#)

[30.3 Portierung von Oracle PL/SQL](#)

[30.4 Tools zur Unterstützung der Migration](#)

[30.5 Technisches Vorgehen](#)

[30.6 Ein Migrationsbeispiel](#)

[30.6.1 Manuelle Migration](#)

[30.6.2 Migration unter Verwendung von Ora2Pg](#)

[30.6.3 Große Tabellen laden](#)

[31 Replikation zwischen Oracle und PostgreSQL](#)

[31.1 Datenbanklink zwischen Oracle und PostgreSQL](#)

[31.2 Replikation mit Oracle XStream](#)

[32 PostgreSQL in der Cloud](#)

[32.1 Private Cloud](#)

32.2 Public Cloud

32.3 Hybrid Cloud

1 Einführung und Geschichte

Seit dem Erscheinen des Buchs „PostgreSQL 10 – Praxisbuch für Administratoren und Entwickler“ sind fast vier Jahre vergangen. In dieser Zeit hat sich bezüglich der Weiterentwicklung und des Einsatzes von PostgreSQL-Datenbanken viel getan. Ein guter Grund, die neue Version mit ihren neuen Features und Verbesserungen zu präsentieren.

So wie bei anderen Herstellern von Datenbanksystemen ist man dazu übergegangen, anstelle von großen Major-Release-Paketen neue Features kontinuierlicher zur Verfügung zu stellen. Das hat zur Folge, dass die Versionen schneller hochgezählt werden als in der Vergangenheit. Das Buch bezieht sich auf die Version 14. Es ist keine einfache Überarbeitung des Vorgängers, sondern wurde inhaltlich neu strukturiert und erweitert. Das Buch präsentiert sich weiterhin im bekannten Stil des Autors mit vielen Beispielen und Skripten.

Für die Themenwahl wurde auch die allgemeine Entwicklung im IT-Sektor berücksichtigt. Für die Installation gibt es nun auch eine Anleitung für macOS. Macbook-Besitzer können nun direkt einsteigen und wahlweise aus Paketen oder dem Quellcode installieren.

In [Kapitel 4](#) finden Sie wiederum eine ausführliche Einführung in die Architektur, um das Verständnis für interne Prozesse und Abläufe zu fördern. Das Thema „Sicherheit und Überwachung“ wurde um interne Details zur Authentifizierung und Hacker-Abwehr erweitert.

PostgreSQL tummelt sich zunehmend im Bereich der großen Datenbanken. Die Themen „Performance- und SQL-Optimierung“ sowie der Umgang mit großen Datenbanken bilden deshalb einen Schwerpunkt. Dazu wurde das [Kapitel 11 „Indexe effizient einsetzen“](#) neu aufgenommen. Es zeigt die Überlegenheit von PostgreSQL in diesem Umfeld gegenüber manch anderem Datenbanksystem. Für den Einsatz von großen Datenbanken zeigen die Möglichkeiten der Parallelisierung, wie PostgreSQL-Datenbanken und SQL-Anweisungen gegen rapides Wachstum skalieren können.

Ein weiterer Schwerpunkt ist der Bereich „Data Science und maschinelles Lernen“. Es werden die Möglichkeiten und Vorteile für den Einsatz von PostgreSQL als Datenbasis, aber auch der vorhandenen Features für Analyse und Auswertungen von Daten dargestellt. Hierbei spielt die Textverarbeitung, der ein eigenes Kapitel gewidmet wurde, eine wichtige Rolle. Die Datenbankprogrammierung mit Python, client- und serverseitig, wurde ebenfalls neu aufgenommen und ergänzt die Darstellung und Beispiele der anderen Programmiersprachen. Das Buch wird damit Administratoren und Entwicklern gleichermaßen gerecht.

Am Thema „Cloud“ führt längst kein Weg mehr vorbei. Die Cloud-Angebote haben sich vervielfacht. Aber auch die Einsatzmöglichkeiten und Features haben sich erweitert. Aktuelle Trends wie die Hybrid Cloud machen auch vor PostgreSQL nicht halt.

Das Buch ist als Einstieg und Nachschlagewerk für IT-Profis geschrieben und setzt Basiskenntnisse von relationalen Datenbanken voraus. Auf eine Erläuterung von gängigen Begriffen wird deshalb bewusst verzichtet, auch um den Umfang des Buchs überschaubar zu halten. Dennoch finden Sie viele Beispiele und Praxistipps, die auch Einsteigern die Möglichkeit bieten, sich in das Produkt einzuarbeiten.

PostgreSQL hat in den vergangenen Jahren an Verbreitung und Popularität erheblich zugenommen. Dazu haben in erheblichem Maß die permanente Erweiterung mit neuen Features und die Anpassung an die Belange der Anwender beigetragen.

PostgreSQL ist der lebende Beweis, dass Open-Source-Software nicht nur mit kommerziellen Produkten mithalten kann, sondern in einigen Bereichen sogar überlegen ist. Der kommerzielle Druck steht nicht im Vordergrund und lässt die Entwickler-Community frei arbeiten und Innovationen umsetzen.

Neben einem robusten Transaktionskern sowie einer hohen Zuverlässigkeit bietet PostgreSQL viele Features eines modernen Datenbankbetriebssystems und kann problemlos in eine vorhandene IT-Infrastruktur integriert werden. Durch den hohen Kompatibilitätsgrad zu Oracle ist der Migrationsaufwand überschaubar und ein Mischbetrieb gut umzusetzen.

PostgreSQL kann auf allen populären Plattformen wie Linux, macOS, Solaris oder Windows eingesetzt werden. Obwohl es sich um ein Open-Source-Produkt handelt, kann kommerzieller Support zu einem vernünftigen Preis hinzugekauft werden. Einem professionellen Einsatz steht damit nichts im Wege.

Freuen Sie sich auf einen PostgreSQL-Server 14 mit spannenden neuen Features!

1.1 Die Geschichte von PostgreSQL

PostgreSQL geht zurück auf das POSTGRES-Projekt, das an der University of California at Berkeley in den 1980er-Jahren angesiedelt war. Die erste vorzeigbare Version erschien im Jahre 1987 als Postgres-Version 1. Als Reaktion auf die ersten Kritiken wurde das noch heute in PostgreSQL vorhandene Rule-System entwickelt. Version 3 erschien im Jahre 1991 mit einer Weiterentwicklung der Abfrageeinheit. 1993 beendete die University of California das Projekt mit der Version 4.2, um die rasant wachsenden Supportanforderungen nicht mehr tragen zu müssen.

Nach Hinzufügen eines SQL-Abfrageinterpreters im Jahr 1995 wurde die Software unter dem Begriff Postgres95 ins Web gestellt, mit dem Quellcode des originalen Berkeley-Postgres. Das Produkt war zu dieser Zeit komplett in ANSI C geschrieben. Durch Verbesserungen in den Bereichen Wartbarkeit und Performance lief es schließlich bis zu 50 % schneller als das originale Berkeley-Postgres.

Die Entscheidung, die Jahreszahl aus dem Produktnamen zu entfernen, fiel im Jahre 1996. Damit wurde Postgres95 zu PostgreSQL und es begann die ständige Weiterentwicklung von PostgreSQL als Open-Source-Produkt. Obwohl es über viele Jahre ein Schattendasein im Licht der großen kommerziellen Datenbanken, aber auch der durch den Internet-Boom schnell verbreiteten Open-Source-Datenbank MySQL führte, erfolgte seine konsequente Weiterentwicklung durch die Community.

Heute präsentiert sich PostgreSQL als ausgereift und stabil und erfüllt alle Anforderungen an ein modernes relationales

Datenbanksystem. Für viele überraschend: Die Performance ist vergleichbar mit so manchem kommerziellen Produkt.

1.2 Verwendete Versionen

Das Buch bezieht sich auf die während der Manuskripterstellung vorliegende aktuelle Version 14. Schauen Sie regelmäßig nach weiteren Veröffentlichungen, insbesondere für neuere Features und Versionen, auf der Webseite des Verlags und der Autoren-Webseite vorbei. Alles rund um die PostgreSQL-Community finden Sie auf der Webseite <http://www.postgresql.org>.

1.3 Konventionen

Begriffe in spitzen Klammern bezeichnen eine zu ersetzende Variable (so ist zum Beispiel der Ausdruck `<VERSION>` in der Regel durch die aktuelle Version 14.1 zu ersetzen). Die meisten Darstellungen beziehen sich gleichermaßen auf UNIX- und Windows-Betriebssysteme. Die Darstellung der Umgebungsvariablen erfolgt im Wesentlichen im UNIX-Format, das heißt z. B. `$BIN` statt `%BIN%` für Windows. Sie können das Format einfach nach Windows übertragen. Das Gleiche gilt für das Trennzeichen der Pfade: „/“ unter Unix sowie „\“ unter Windows.

1.4 Software und Skripte

Sie können die aktuelle Version von PostgreSQL aus dem Internet herunterladen und installieren. Es wird die Installation aus dem Quellcode empfohlen, um alle Beispiele nachvollziehen zu können. Ideal ist es, wenn Sie auf einem Linux-, macOS- oder Windows-Betriebssystem arbeiten.

Alle nummerierten Listings im Buch können als Datei von der Webseite des Verlags sowie von der Autoren-Webseite <https://www.lutzfroehlich.de> heruntergeladen werden:

Geben Sie auf

<https://plus.hanser-fachbuch.de>

diesen Code ein:

```
plus-92Fgh-ci4R7
```

1.5 Update inside

PostgreSQL stellt jährlich neue Major-Release-Versionen mit neuen Features und Erweiterungen zur Verfügung. Davon sind alle Bereiche vom Kernsystem bis hin zur Applikationsentwicklung betroffen.

Damit Sie möglichst lange mit diesem Buch arbeiten können, haben Sie die Möglichkeit, sich für den kostenlosen Update inside-Service zu registrieren: Geben Sie unter

www.hanser-fachbuch.de/postgresql-update

diesen Code ein:

```
2WHK-SeoTr-bF5x
```

Dann erhalten Sie bis Mai 2024 Aktualisierungen in Form zusätzlicher Kapitel als PDF. Darin stelle ich Ihnen wichtige Neuerungen vor und gehe auf Änderungen ein, die die Inhalte dieses Buches betreffen.

Putbus und Oasis del Sol, im April 2022

Lutz Fröhlich

lutz@lutzfroehlich.de