



**Hochschule
Bonn-Rhein-Sieg**
University of Applied Sciences

Projektarbeit

PostgresSQL - Rekursion auf Basis generischer Stored Procedures

Fachbereich Informatik
Referent: Prof. Dr. Harm Knolle

eingereicht von:
Rolf Kimmelman, Jennifer Wittling, Jan Löffelsender

Sankt Augustin, den 12.11.2018

Zusammenfassung

Problemstellung und Erkenntnisinteresse

In den letzten Jahren haben Graphdatenbanken an Bedeutung gewonnen, da sich mit diesen bestimmte Fragestellungen besonders schnell lösen lassen. Graphdatenbanken haben den Vorteil, dass sich insbesondere Beziehungen zwischen Objekten gut abbilden und sehr performant abfragen lassen. Bei relationalen Datenbanken ist es zur Darstellung von Beziehungen zwischen Objekten erforderlich die verschiedenen Tabellen mittels des JOIN Operators zu verknüpfen. Diese Verknüpfungen können schnell zu einem großen Rechenaufwand und langen Laufzeiten führen. Es soll am Beispiel von Postgres untersucht werden, ob und wie sich Graphen in relationalen Datenbanken abbilden lassen. Weiterhin soll analysiert werden, ob und für welche Problemstellungen es sinnvoller ist Graphen in einer relationalen Datenbank statt einer Graphdatenbank abzubilden. Ist es zukünftig notwendig für die performante Verarbeitung steigender Datenmengen auf neue Technologien, wie Graphdatenbanken zu schwenken oder lassen sich die klassischen relationalen Datenbanken so erweitern, dass diese Problemstellungen ähnlich effizient lösen können.

Aktueller Forschungsstand

NoSQL Datenbanken und insbesondere Graphdatenbanken sind im Gegensatz zu den relationalen Datenbanken flexibler und bei der Lösung bestimmter Probleme weniger rechen- und speicherintensiv. Insbesondere wenn es um die Auflösung von Beziehungen bzw. um die Traversierung über einen Graphen geht, bieten Graphdatenbanken Vorteile gegenüber den herkömmlichen relationalen Datenbanken. In der Praxis wurde jedoch auch die Beobachtung gemacht, dass durch die Verwendung von Stored Procedures die Traversierung über einen Graphen mittels einer relationalen Datenbank ähnlich schnell umgesetzt werden kann, wie mit einer Graphdatenbank.

Zielsetzung

Es soll das Modell als grundlegender technologische Aspekt von Graphdatenbanken kurz erläutert werden. Zielsetzung dieser Arbeit ist es einen Graphen in der relationalen Datenbank Postgres abzubilden und zu vergleichen, wie sich die Traversierung über diesen Graphen effizient umsetzen lässt. Zunächst soll die Umsetzung mittels klassischer SQL Operationen erfolgen. Anschließend sollen die Problemstellungen mittels Stored Procedures, sowie der Rekursion mittels PL/SQL gelöst werden. Die Ergebnisse der verschiedenen Vorgehensweisen sollen miteinander verglichen werden.

Inhaltsverzeichnis

Zusammenfassung	II
1 Graph-Datenbanken - Grundlegende technologische Aspekte	IV
1.1 Einführung	IV
1.2 Modell	IV
1.2.1 Graph	IV
1.2.2 Property Graphen	IV
1.2.3 Hypegraphen	IV
2 Graph-Datenbanken und -Frameworks - Ausgewählte Systeme	VI
2.1 PostgreSQL	VI
2.1.1 Allgemein	VI
2.1.2 Architektur	VI
2.1.3 Datenmodell	VII
2.1.4 Indexe	VII
2.1.5 Anfragemethoden	VII
2.1.6 Konsistenz	VII
3 Graph-Datenbanken im praktischen Einsatz: OLTP	VIII
3.1 PostgreSQL: OLTP	VIII
3.2 Ausgewählte Use Cases	VIII
3.3 Beurteilung	VIII
4 Graph-Datenbanken im praktischen Einsatz: OLAP	IX
4.1 PostgreSQL: OLAP	IX
4.2 Benchmark	IX
4.2.1 Standard SQL	IX
4.2.2 Stored Procedures	IX
4.2.3 PL/SQL-Recursion	IX
4.3 Datenbankzugriffe	IX
4.3.1 Zugriffsart Aggregation	IX
4.3.2 Zugriffsart Traversierung	IX
4.4 Interpretation der Ergebnisse	IX
Literaturverzeichnis	X
Eidesstattliche Erklärung	XII

1 Graph-Datenbanken - Grundlegende technologische Aspekte

1.1 Einführung

1.2 Modell

1.2.1 Graph

Ein Graph ist mathematisch folgendermaßen definiert:

Definition. *Ein Graph(graph) $G = (V, E, \gamma)$ ist ein Tripel bestehend aus:*

- V , einer nicht leeren Menge von Knoten(vertices)
- E , einer Menge von Kanten (edges) und
- γ , einer Inzidenzabbildung (incidence relation), mit
 $\gamma : E \longrightarrow \{X | X \subseteq V, 1 \leq |X| \leq 2\}$

Zwei Knoten $a, b \in V$ heißen adjazent(adjacent) genau dann wenn $\exists e \in E : \gamma(e) = \{a, b\}$.

Ein Knoten $a \in V$ und eine Kante $e \in E$ heißen inzident (incident) genau dann wenn $a \in \gamma(e)$.¹

1.2.2 Property Graphen

Property Graphen zeichnen sich durch ihre den Kanten und Knoten zugewiesenen Eigenschaften auf.

¹Vgl. [Bec18, Seite 21]

1.2.3 Hypegraphen

Ein Hypergraph ist folgendermaßen definiert:

Definition. *Let $X = \{x_1, x_2, \dots, x_n\}$ be a finite set, and let $E = \{e_1, e_2, \dots, e_m\}$ be a family of subsets of X such that*

$$e_i \neq \emptyset (i = 1, 2, \dots, m) \cup_{i=1}^m e_i = X.$$

The pair $H = (X, E)$ is called a hypergraph with vertex set X and hyperedge set E . The elements x_1, x_2, \dots, x_n of X are vertices of hypergraph H , and the sets e_1, e_2, \dots, e_m are hyperedges of hypergraph H .²

²Vgl. [ZSHZ18, Seite 2]

2 Graph-Datenbanken und -Frameworks - Ausgewählte Systeme

2.1 PostgreSQL

2.1.1 Allgemein

- Kategorie / Modell

PostgreSQL ist ein Relationales Datenbank System [Groa]

- Version

Aktuelle Version: 11

- Historie

PostgreSQL ist aus dem POSTGRES projekt der University of California at Berkeley entstanden, welches unter der Leitung von Professor Michael Stonebraker im Jahre 1986 began. SQL Interpreter seit 1994, das System hieß zu diesem Zeitpunkt Postgres95. 1996 wurde es in PostgrSQL umbenannt.[Groa]

- Hersteller

- Lizenz

Open-Source

2.1.2 Architektur

- Programmiersprache des Systems
- Systemkomponenten, Systemarchitektur

client / server Architektur [Groa]

- Betriebsart

Cluster - Menge an Datenbanken, die von PostgreSQL-Server verwaltet werden [Fro18]

- Protokoll der Schnittstelle

TCP/IP

- API

2.1.3 Datenmodell

- Standardsprache: SQL
- Objektbegriff, Konzepte: Relationale Datenbank - Abbildung in Tabellen
- Datentypen:

Sehr viele unterstützte Datentypen, es lassen sich aber auch eigene Datentypen mittels des CREATE TYPE befehls erstellen.[Grob]

2.1.4 Indexe

2.1.5 Abfragemethoden

2.1.6 Konsistenz

3 Graph-Datenbanken im praktischen Einsatz: OLTP

3.1 PostgreSQL: OLTP

3.2 Ausgewählte Use Cases

3.3 Beurteilung

4 Graph-Datenbanken im praktischen Einsatz: OLAP

4.1 PostgreSQL: OLAP

4.2 Benchmark

4.2.1 Standard SQL

4.2.2 Stored Procedures

4.2.3 PL/SQL-Recursion

4.3 Datenbankzugriffe

4.3.1 Zugriffsart Aggregation

4.3.2 Zugriffsart Traversierung

4.4 Interpretation der Ergebnisse

Literaturverzeichnis

- [Ang12] ANGLES, Renzo: A comparison of current graph database models. In: *Data Engineering Workshops (ICDEW), 2012 IEEE 28th International Conference on IEEE*, 2012, S. 171–177
- [APPDSLP13] ANGLES, Renzo ; PRAT-PÉREZ, Arnau ; DOMINGUEZ-SAL, David ; LARRIBA-PEY, Josep-Lluís: Benchmarking database systems for social network applications. In: *First International Workshop on Graph Data Management Experiences and Systems* ACM, 2013, S. 15
- [AU95] AHO, Alfred V. ; ULLMAN, Jeffrey D.: *Foundations of computer science*. USA : Computer Science Press, 1995 <http://infolab.stanford.edu/~ullman/focs.html>
- [Bec18] BECKER, Peter: *Graphentheorie*. <http://www2.inf.h-brs.de/~pbecke2m/graphentheorie/einfuehrung.pdf>. Version: 2018
- [Eis03] EISENTRAUT, Peter: *PostgreSQL Das Offizielle Handbuch*. mitp-Verlag GmbH/Bonn, 2003
- [Fro18] FROELICH, Lutz: *PostgreSQL 10*. Carl Hanser Verlag München, 2018
- [Groa] GROUP, The PostgreSQL Global D.: *PostgreSQL 11.1 Documentation*. <https://www.postgresql.org/files/documentation/pdf/11/postgresql-11-A4.pdf>
- [Grob] GROUP, The PostgreSQL Global D.: *PostgreSQL 8.4 Documentation*. <https://www.postgresql.org/docs/8.4/datatype.html>
- [Gru17] GRUCIA, Jelena: *PostgreSQL and GraphQL*. <https://blog.cloudboost.io/postgresql-and-graphql-2da30c6cde26>. Version: 2017
- [KHA⁺16] KUCUK, Ahmet ; HAMDI, Shah M. ; AYDIN, Berkay ; SCHUH, Michael A. ; ANGRYK, Rafal A.: Pg-Trajectory: A PostgreSQL/PostGIS based data model for spatiotemporal trajectories. In: *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom)(BDCloud-SocialCom-SustainCom)* IEEE, 2016, S. 81–88

- [Kud15] KUDRASS, Thomas: *Taschenbuc Datenbanken*. Fachbuchverlag Leipzig im Carl Hanser Verlag, 2015
- [Red12] REDMOND, Eric: *Sieben Wochen, sieben Datenbanken*. O'Reilly Verlag, 2012
- [Sas18] SASAKI, Bryce M.: *Graph Databases for Beginners: The Basics of Data Modeling*. <https://neo4j.com/blog/data-modeling-basics/>. Version: 2018
- [ZSHZ18] ZHANG, Hongliang ; SONG, Lingyang ; HAN, Zhu ; ZHANG, Yingjun: *Hypergraph Theory in Wireless Communication Networks*. Springer, 2018

Eidesstattliche Erklärung

Ich versichere an Eides Statt durch meine eigenhändige Unterschrift, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe angefertigt habe. Alle Stellen, die wörtlich oder dem Sinn nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche kenntlich gemacht. Ich versichere außerdem, dass ich keine andere als die angegebene Literatur verwendet habe. Diese Versicherung bezieht sich auch auf alle in der Arbeit enthaltenen Zeichnungen, Skizzen, bildlichen Darstellungen und dergleichen.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

*Sankt Augustin,
den 27. November
2018*

Ort, Datum

Rolf Kimmelman, Jennifer
Wittling, Jan Löffelsender