



**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

Exposé

Überbegriff Inhalt: Was soll im Exposé geleistet werden?

- Problemstellung „Welches theoretische, praktische Problem ist Ausgangspunkt der Arbeit?“
 - Es soll untersucht werden wie sich Graphen in der relationalen Datenbank Postgresql abbilden lassen
 - Ist es besser eine Relationale Datenbank als Graphdatenbank zu verwenden oder eine "richtige" Graphdatenbank zu verwenden.
 - Welche Problemstellungen lassen sich durch die Modellierung von Daten in Form von Graphen lösen
 - Optimierungsmöglichkeiten? Handgestrikte SQLs vs. Stored Procedure
- Forschungsstand Eigene Vorkenntnisse im Exposé darlegen: Wie ist der aktuelle Forschungsstand zum Thema der Arbeit?
 - In einer relationalen Datenbank wird beim Auflösen von Beziehungen (z.B. Vater-Kind Beziehung oder wer kennt wen Beziehung) ein JOIN verwendet, Bei einem JOIN wird das kartesische Produkt verwendet, welches schon nach relativ wenigen Eingabeelementen viel Rechenleistung beansprucht.
- Forschungsstand Eigene Vorkenntnisse im Exposé darlegen: Wie ist der aktuelle Forschungsstand zum Thema der Arbeit?
 - In der Praxis wurde folgende Beobachtung gemacht: Das Traversieren über einen Graph mit Hilfe von Stored Procedures auf einer relationalen Datenbank ist vermutlich ähnlich schnell, wie das Traversieren mit Hilfe einer klassischen Graphdatenbank.
- Wissenslücke/Erkenntnisinteresse Warum will man sich mit diesem Problem beschäftigen? Welche Wissenslücke kann man dadurch schließen?
 - Praktische Anwendungsgebiete wie Empfehlung engines o.Ä.
 - Vorteile von Relationalen Datenbanken.
 - Auswirkungen auf bestehende Anwendungen. Was sind die Vorteile? Wie wären die Auswirkungen auf ein Datenbankschema?
- Fragestellung Welche Frage (Forschungsfrage) soll in der Arbeit beantwortet werden?
 - Warum sind Graphdatenbanken so schnell?

- Ziel/Hypothese Im Exposé darlegen, welches Ziel erreicht werden soll: Was soll bewiesen oder widerlegt werden? Kurz: „Was will ich wissen? (Fragestellung), wozu will ich das wissen? (Ziel)“
 - Wie performant lässt sich die Traversierung mit Hilfe von Stored Procedures bzw SQL-Erweiterungen abbilden? Kann man hier Abschätzungen zur Komplexität machen?
 - Welche Besonderheiten bietet Postgresql?
 - Schneiden die relationale Datenbanken im Vergleich zu Graphdatenbanken in Hinsicht auf Performance beim Traversieren von Graphen ähnlich gut ab, wenn man unkonventionelle Methoden verwendet (Traversieren mit Hilfe von rekursiven Stored Procedures).
- Theoriebezug Welche Theorien sollen als Basis für die Bearbeitung der Fragestellung dienen?
 - Graphentheorie
 - Relationale Algebra
 - Komplexitätstheorie
 - Rekursion - Besonderheit Rekursion in SQL

Beispielgliederung

1. Graph-Datenbanken - Grundlegende technologische Aspekte
 - 1.1. Einführung
 - 1.2. Modell (Graph, Property Graphen, Hypergraphen)
2. Graph-Datenbanken und -Frameworks - PostgreSQL
 - 2.1. Allgemein
 - 2.2. Architektur
 - 2.3. Datenmodell
 - 2.4. Indexe
 - 2.5. Anfragemethoden
 - 2.6. Konsistenz
3. Graph-Datenbanken im praktischen Einsatz: OLTP
 - 3.1. Ausgewählte Use Cases
 - 3.2. Weitere Zugriffstechniken
 - 3.3. Vergleich mit relationalen Datenbanksystemen
 - 3.4. Beurteilung
4. Graph-Datenbanken im praktischen Einsatz: OLAP
 - 4.1. Benchmark
 - 4.2. Datenbankzugriffe
 - 4.3. Zugriffsart Aggregation
 - 4.4. Zugriffsart Traversierung
 - 4.5. 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*. Computer Science Press, 1995 <http://infolab.stanford.edu/~ullman/focs.html>
- [Eis03] EISENTRAUT, Peter: *PostgreSQL Das Offizielle Handbuch*. mitp-Verlag GmbH/Bonn, 2003
- [Fro18] FROELICH, Lutz: *PostgreSQL*. Carl Hanser Verlag München, 2018
- [Gru17] GRUCIA, Jelena: *PostgreSQL and GraphQL*. <https://blog.cloudboost.io/postgresql-and-graphql-2da30c6cde26>. Version: 2017
- [KHA⁺16] KUCUK, Ahmet ; HAMDİ, 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