



Datenbankmodellierung mit Knowledge Graphen für den Anwendungsfall xyz – Leistungsvergleich mit einer RDBMS und Handlungsempfehlungen

Problemstellung/Ausgangssituation:

NoSQL-Datenbanken und hier speziell graphbasierte Datenbankmodelle (Knowledge Graphen) sind ein hochaktuelles Thema in der (Wirtschafts-)Informatik. Eine strukturierte, tabellenbasierte Datenhaltung in Form eines relationalen Datenbankmanagementsystems (RDBMS) liegt heute den meisten Anwendungen zugrunde, insbesondere den Anwendungen, die extrem große Datenmengen verarbeiten („Big Data“). Dabei werden Beziehungen (Relationen) zeilenweise in vordefinierten Spaltenstrukturen abgelegt, doch beinahe widersprüchlicherweise werden Beziehungen zwischen Elementen (wie bspw. Zeilen) nur über algorithmische Auswertungen erkennbar bzw. erfordern die Hinzunahme weiterer (starrer) relationaler Tabellen. Graphbasierte Datenbanken setzen genau an dieser Stelle an, indem diese (teilweise in einfachsten Strukturen), sämtliche Beziehungen persistieren. So kann das Internet mit all seinen Verflechtungen als gigantischer Knowledge Graph verstanden werden. Erscheint die Graphenmodellierung zunächst vorteilhaft, so stehen dieser anwendungsfallspezifisch auch erhebliche Nachteile gegenüber, die sich bspw. in sehr hohen Laufzeiten bei analytischen Anwendungen aber auch signifikantem Aufwand im Falle von Datenbankreorganisationen manifestieren können.

Masterarbeit:

Im Rahmen einer Masterarbeit soll für einen Anwendungsfall eine graphbasierte Datenbank (bspw. eine open source Datenbank) zugrunde gelegt werden. Der Anwendungsfall kann aus dem eigenen Unternehmen/Kontext stammen oder frei definiert werden. Es bietet sich an, der Anwendung einen Datenbanksatz mit signifikanter Datenmenge zugrunde zu legen, damit eine verlässliche Messung der Leistungsparameter der Implementierung durchgeführt werden kann. Optional kann zudem ein Anwendungsszenario hinzugefügt werden, welches eine Reorganisation des Datenbanksystems erzwingt. Im zweiten Schritt wird der identischen Anwendung ein RDBMS zugrunde gelegt, wobei hierbei ebenfalls von einer open source Implementierung ausgegangen werden darf. Die zuvor gemessenen Leistungsparameter werden ebenfalls bestimmt, und abschließend erfolgt die Erstellung einer empirisch und literaturgestützten Handlungsempfehlung für Praktiker und Entscheider.

Zielgruppe:

Die ausgeschriebene Arbeit adressiert:

- Technisch mittel aber auch stark versierte Studierende
- Betriebswirtschaftlich orientierte Studierende, die Interesse an modernen IT-Konzepten haben und mindestens über gute Grundlagenkenntnisse in der Programmierung verfügen.
- Technisch Interessierte aus dem Beratungs- und Projektmanagementumfeld mit guten Grundlagenkenntnissen in der Programmierung.

Erweitertes Ziel ist es, die Ergebnisse der Arbeit zu veröffentlichen, sowie ggf. die gewonnenen Kenntnisse im Berufsumfeld anzuwenden.