

Performanz Evaluation von Graphdatenbanksystemen versus SQL am Beispiel von Kookkurrenzgraphen

Martin Junghanns

`martin.junghanns@studserv.uni-leipzig.de`

Sascha Ludwig

`s.ludwig@studserv.uni-leipzig.de`

Robert Schulze

`robert.schulze@studserv.uni-leipzig.de`

10. Januar 2012

Zusammenfassung

In dieser Arbeit wurden... The research focus (i.e. statement of the problem(s)/research issue(s) addressed); The research methods used (experimental research, case studies, questionnaires, etc.); The results/findings of the research; and The main conclusions and recommendations

Konsolen können Befehle in einer SQL-artigen Syntax an das DBS geschickt und ausgewertet werden.

1 Neo4j

2 OrientDB

2.1 Gebrauch

OrientDB lässt sich einfach über eine Java API steuern. Das Handling von OrientDB wird, nach einer gewissen Einarbeitungszeit, relativ einfach, was auch der guten Javadoc der API geschuldet ist. Neben einer Java-API bietet OrientDB auch die Möglichkeit der Steuerung des Datenbanksystems über eine Konsole. Dies hat den Vorteil, dass man das Datenbanksystem erkunden und kennen lernen kann. In der

2.2 Einschränkungen

Ein großes Bottleneck bei OrientDB ist der Import der Daten aus einer SQL-Datenbank. Auch nach dem befolgen der "Performance Hints" und dem Optimieren diverser JVM Parameter, stieg die Zeit für den Import exponentiell zur Anzahl der einzufügenden Daten an. Dies hat zur Folge, dass große Datenbanken (ab 1M Datensatz) nicht mehr importiert werden konnten. Mittels JProfiler konnte der Persistierungsprozess von OrientDB als HotSpot ausgemacht werden. Ein Kontaktaufnahme mit dem Autor bezüglich dieses Performanzverlustes ist durch uns erfolgt.

2.3 Dokumentation

OrientDB ist gut und ausführlich dokumentiert. Es liegt sowohl eine komplette und ausführliche Javadoc der API vor, als auch Informationsseiten zu grundlegenden und weit-

erführenden Aspekten vor. Auf diesen Webseiten lassen sich Informationen zu fast allen Fragen finden. Allerdings ist das Finden einer Antwort auf eine konkrete Frage teilweise schwierig. Dies ist unter Umständen dem Fakt geschuldet, dass OrientDB sowohl ein Dokumenten-Datenbanksystem, als auch ein Key-Value-Store, als auch ein Graphdatenbanksystem ist. Oft werden in der Dokumentation diese drei Konzepte vermischt, was zu Verwirrungen führen kann. Weiterhin hat das Fehlen eines ausführlichen und kompletten Einstiegbeispiels die Einarbeitung erschwert. Auch fehlen Beispiele für komplexe Anwendungsfälle. Allerdings war der Autor bei speziellen Fragen per E-Mail zu erreichen, wobei man auf eine Antwort lediglich eine angemessene Zeit warten musste.

2.4 Aussicht

OrientDB stellt bezüglich seiner Graphdatenbankenfunktionalität einige interessante Features in Aussicht. OrientDB stellt seit Kurzem eine Implementation der Tinkerpop Blueprints APIs zur Verfügung. Somit steht auch die einfache und leicht lesbare Highlevel-Abfragesprache Gremlin zur Verfügung. Mit dieser Abfragesprache ist es möglich auch komplexe und vielseitige Abfragen verständlich zu formulieren.

3 DEX