Bachelorarbeit

Nils Petersohn

Vergleich und Evaluation zwischen modernen und traditionellen Datenbankkonzepten unter den Gesichtspunkten Skalierung, Abfragemöglichkeit und Konsistenz

Diplom.de

Bibliografische Information der Deutschen Nationalbibliothek:

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

ISBN: 9783842806085

Nils Petersohn

Vergleich und Evaluation zwischen modernen und traditionellen Datenbankkonzepten unter den Gesichtspunkten Skalierung, Abfragemöglichkeit und Konsistenz

Diplom.de

Bachelorarbeit

Nils Petersohn

Vergleich und Evaluation zwischen modernen und traditionellen Datenbankkonzepten unter den Gesichtspunkten Skalierung, Abfragemöglichkeit und Konsistenz

Diplom.de

Nils Petersohn
Vergleich und Evaluation zwischen modernen und traditionellen Datenbankkonzepten unter den Gesichtspunkten Skalierung, Abfragemöglichkeit und Konsistenz

ISBN: 978-3-8428-0608-5
Herstellung: Diplomica® Verlag GmbH, Hamburg, 2010
Zugl. Fachhochschule Brandenburg, Brandenburg, Deutschland, Bachelorarbeit, 2010

Inhaltsverzeichnis

1. Einleitung

Zehntausende Web-Services verwenden relationale Datenbanken, um Daten zu speichern und auszulesen. Im Vergleich zu modernen Konzepten können relationale Datenbanken als wichtigster Stellvertreter für "traditionelle Technologien" bezeichnet werden. Wenn man als Entwickler zu Seiten wie Google.com, Facebook.com, Amazon.com, Digg.com, Ebay.com, Yahoo.com, Twitter.com, oder Dawanda.com surft, wird meist angenommen, dass eine verteilte relationale Datenbank verwendet wird. Die Annahme ist zu 50% richtig, jedoch ist die Datenhaltung meist nicht relational. Diese Großunternehmen verwalten mehrere hundert Gigabytes, bis hin zu 100.000 Gigabyte an Daten, und mussten in den letzten sechs Jahren Lösungen finden, um erfolgreich diese riesigen Datenmengen zu beherrschen. Google erfand vor ca. sieben Jahren ein Verfahren, um Datenmengen im Petabytebereich zu beherrschen. Facebook entwickelte selbst eine Datenbanktechnologie, um die Posteingänge von Benutzern verfügbar zu machen, Twitter.com adaptiert diese Technologie für andere Zwecke (vgl. [Lai10]). Amazon.com entwickelte "Dynamo", um Hochverfügbarkeit für deren weltgrößte E-Commerce Plattform zu schaffen. Diese und andere Eigenentwicklungen entstanden aus der Notwendigkeit heraus, riesige Datenmengen bzw. Datenbanken hoch verfügbar, konsistent und skalierbar zu machen.

1.1. Motivation

Seit den letzten drei Jahren sind alternative "Open-Source-Implementierungen" dieser Entwicklungen entstanden. Die Veröffentlichung der Konzepte und Technologien führten zu einer ganzen Bewegung namens "NoSQL". Sind diese Konzepte vorteilhafter, um eine bessere und für Entwickler einfachere Skalierung, Abfragemöglichkeit und Datenkonsistenz in einem hochverfügbaren Datenbanksystem, zu gewährleisten? Wie werden komplexe Abfragen in modernen und traditionellen verteilten Systemen gemacht und wie werden diese ausgeführt? Speziell stellt sich die Frage, ob das MapReduce Verfahren ein vollständiger Ersatz für SQL ist. Für welche Einsatzzwecke sind beide besonders gut geeignet und für welche weniger?