

Buchbesprechungen

Andreas Meier

Online publiziert: 10. Juli 2014
© Springer Fachmedien Wiesbaden 2014

Stefan Edlich, Achim Friedland, Jens Hampe, Benjamin Brauer und Markus Brückner

NoSQL

Einstieg in die Welt nichtrelationaler Web 2.0 Datenbanken

Carl Hanser Verlag, 2011, München, zweite aktualisierte und erweiterte Auflage, 392 S., € 29,90, ISBN 978-3-446-42753-2

Das Fachbuch gibt eine Einführung in die wichtigsten NoSQL-Datenbanken und beschreibt wichtige kommerzielle Produkte im Überblick. Da der Begriff NoSQL noch nicht klar gefasst ist, wird in Kap. 1 eine Definition mit sieben möglichen Kriterien gewagt, wie nicht-relational, Ausrichtung auf verteilte und horizontale Skalierung, open source, Schemafreiheit, einfache Datenreplikation, einfaches API und keine Unterstützung des ACID-Transaktionskonzepts (A = Atomicity, C = Consistency, I = Isolation, D = Durability).

Kapitel 2 beschreibt die wichtigsten Grundlagen für NoSQL-Datenbanken. Neben dem Map/Reduce-Verfahren für nebenläufige Berechnungen auf verteilten Clustern wird das sogenannte CAP-Theorem erläutert (C = Consistency, A = Availability, P = Partition Tolerance). Dieses Theorem von Eric Brewer sagt aus, dass bei verteilten Systemen nicht gleichzeitig alle drei Forderungen der Konsistenz, hohen Verfügbarkeit und Ausfalltoleranz gelten können. Daneben werden weitere Prinzipien wie

A. Meier (✉)
Department of Informatics, University of Fribourg,
Boulevard de Pérolles 90,
1700 Fribourg, Schweiz
E-Mail: andreas.meier@unifr.ch

Consistent Hashing, Multi-version Concurrency Control oder REST (Representational State Transfer) kurz erläutert.

Die nächsten fünf Kapitel widmen sich spezifischen NoSQL-Datenbanken mit entsprechenden Softwareoptionen:

- Wide Column Stores mit den Beispielen HBase, Cassandra und Amazon SimpleDB (Kap. 3)
- Document Stores mit den Vertretern CouchDB und MongoDB (Kap. 4)
- Key/Value-Datenbanken mit Redis, Chordless, Riak und Membase (Kap. 5)
- Graphendatenbanken mit Tinkerpop Graph Processing Stack, AllegroGraph, DEX, HyperGraphDB, InfiniteGraph, InfoGrid, Neo4j, sones GraphDB sowie weiteren graphenbasierten Lösungen (Kap. 6)
- Multimodell-Datenbanken mit OrientDB (Kap. 7)

Kapitel 8 listet weitere NoSQL-Datenbanken und Kap. 9 diskutiert Auswahlkriterien für den Einsatz von NoSQL-Datenbanken. Hier werden Aspekte betreffend Datenanalyse, Transaktionsverwaltung, Performance, Abfragemächtigkeit und Architektur kurz geschildert, neben weiteren Aspekten bezüglich Nutzbarkeit und Sicherheit.

Das Fachbuch ist eines der wenigen deutschsprachigen Bücher, das eine Einführung in die NoSQL-Datenbankwelt systematisch aufarbeitet. Da die Autoren die wichtigsten technologischen Ansätze samt einer Übersicht verfügbarer Software beschreiben, eignet sich das Buch auch als Nachschlagewerk. Definitionsteil (Kap. 1), theoretische Grundlagen (Kap. 2) und die Hinweise zur Auswahl geeigneter Ansätze in Kap. 9 sind noch ausbaufähig. Bei einer Neuauflage wäre zudem wünschenswert, den Inhalt des Kap. 8 über weitere NoSQL-Datenbanken auf die bereits bestehenden Kap. 3 bis 7 aufzuteilen. Fazit: Das Buch führt systematisch in die NoSQL-Technologien ein und schildert konkrete Softwareangebote. Der interessierte Praktiker kommt nicht darum herum, die eine oder andere Lösung mit seinen Use Cases selbst auszutesten.

Eric Redmond, Jim R. Wilson

Seven Databases in Seven Weeks

A Guide to Modern Databases and the NoSQL Movement

The Pragmatic Programmers, 2012, Dallas, Texas, 338 pages, US\$ 35,00, ISBN 978-1-93435-692-0

Das Fachbuch ist beim Verlag The Pragmatic Bookshelf herausgekommen. Der Verlag publiziert Bücher mit der Auflage ‚written by developers for developers‘. Das Buch wurde ins Deutsche übersetzt, leider mit vielen Übersetzungsfehlern. Hier soll deshalb die englische Ausgabe besprochen werden.

Die beiden Autoren fragten sich, wie der Begriff NoSQL zu fassen sei. Zudem: Welche Datenbanksysteme würden darunter fallen, welche nicht? Wie kann man gute Softwarelösungen mit diesen neuen Datenbankansätzen vorantreiben? etc. Dann entschlossen sie sich, sieben Datenbanken vorzustellen: PostgreSQL, Riak, HBase, MongoDB, CouchDB, Neo4j und Redis.

Nach einem kurzen Einstieg wird in Kap. 2 das klassische relationale Datenbanksystem PostgreSQL durchgekämmt. Da es open source ist, kann man gleich loslegen: SQL, Transaktionen, Stored Procedures, Triggers, Textsuche und Auswertungen im mehrdimensionalen Datenwürfel. Am Schluss liefern die Autoren eine Bewertung der Stärken und Schwächen von PostgreSQL.

Kapitel 3 behandelt das System Riak, das für den Key/Value-Ansatz steht. Allerdings ist die Vielfalt für Werte gross: Text, XML, Bilder oder Video Clips. Nun folgen wichtige Ansätze wie REST (representational state transfer), Links und MIME Types, bevor das Map/Reduce-Verfahren allgemein und spezifisch für Riak behandelt wird. Spezielle Verfahren wie Vector Clocks für die Konfliktbehandlung werden ebenso behandelt wie Suchstrategien und Indexverwaltung. Zum Schluss, wie gehabt, werden Stärken und Schwächen des Ansatzes diskutiert.

Die übrigen Systeme wie HBase, MongoDB, CouchDB, Neo4j und Redis werden analog eingeführt. Schritt für Schritt werden spezifische Techniken erläutert, so dass man in sieben Wochen einen vertieften Einblick sowohl in wichtige Techniken wie in Datenbanksysteme erhält. Das Buch ist spannend geschrieben und regt mit kleinen Hausaufgaben an. Fazit: Es richtet sich an Informatik-Interessierte, welche die Methoden und Techniken an konkreten Datenbankumgebungen selber austesten möchten. Es eignet sich als Tutorial für den Übungsbetrieb in Fachhochschulen oder Universitäten. Der Besuch von sieben attraktiven open source Datenbanken, SQL wie NoSQL, lässt wenig Raum für Kritik.