

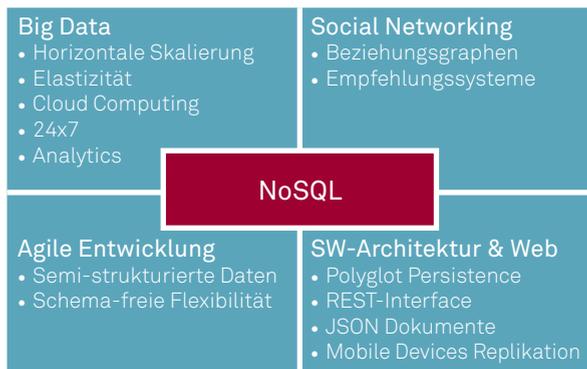
NoSQL

Nicht-relationale Datenbanken und Polyglot Persistence

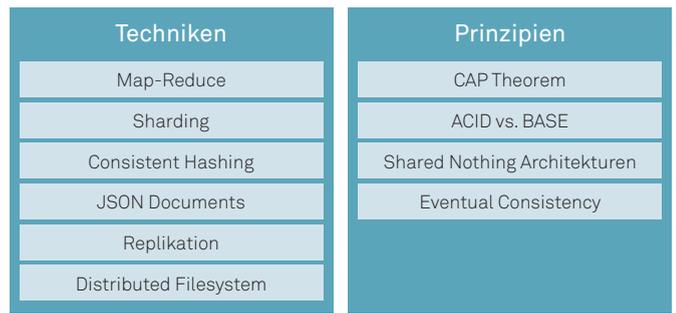
NoSQL („Not only SQL“) ist das Schlagwort für nicht-relationale Datenbanken (DB). Ausgehend von Amazon und Google, die eigene Datenbanken für Massendaten und ihre Cloud-Services erfinden mussten, ist eine große Vielfalt von Produkten entstanden, die unterschiedliche Ansätze verfolgen, um Probleme zu lösen, die mit relationalen Datenbanken (RDBMS) schwer oder gar nicht zu bewältigen sind. Selbst Hersteller von RDBMS wie z.B. Oracle sind auf den Zug aufgesprungen. Mit zunehmendem Einsatz von mehreren spezialisierten Datenbanken wird in Zukunft der Begriff „Polyglot Persistence“ an Bedeutung gewinnen.

Definition

NoSQL-DBs bewältigen die Herausforderungen des Web-Zeitalters mit seiner Vielfalt, Menge und Vernetzung von Daten. Sie verwenden nicht-relationale Datenmodelle. Der Verzicht auf ein striktes Daten-schema macht sie flexibler für Änderungen.



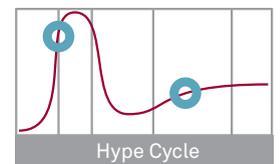
Key-Value Stores und **Wide-Column Stores** sind vor allem für Massendaten und deren verteilte Auswertung gedacht. Sie sind optimiert für unbegrenzte horizontale Skalierung durch Partitionierung auf Commodity-Hardware sowie unterbrechungsfreien Betrieb und hohe Schreib- und Leseperformance. Die **Dokument-Datenbanken** sind ebenfalls hoch skalierbar und Spezialisten für semi-strukturierte Daten und Mischdaten. Sie positionieren sich z. T. speziell für das Mobile Computing. **Graph-Datenbanken** können tief vernetzte Strukturen effizient bewältigen – ein großes Problem für RDBMS. Sie gehören nicht zu den hoch skalierbaren DBs und haben das semantisch reichste Datenmodell.



Die hoch skalierbaren DBs geben Eigenschaften von RDBMS (z.B. Normalisierung, Transaktionen) zugunsten von Geschwindigkeit und Verfügbarkeit auf. Die NoSQL-DBs werden in Zukunft die RDBMS ergänzen. Datenbanken werden für ihr optimales Gebiet ausgewählt und kooperieren, damit innovative Lösungen möglich werden. Diese Strategie wird **Polyglot Persistence** genannt. Die Integration unterschiedlicher Datenspeicher ist in der Enterprise-IT bereits Alltag. Die NoSQL-DBs sind meist offen und lassen sich gut integrieren.

Reifegrad

Der Reifegrad ist aufgrund der Vielfalt und Herkunft uneinheitlich. Das Interesse an NoSQL ist seit 2010 stark angestiegen. NoSQL fehlt auf keiner einschlägigen Konferenz. Viele Produkte haben eine große Open-Source-Community und kommerziellen Support. Es gibt Hosting-Anbieter, die NoSQL-DBs betreiben. Die Tool-Unterstützung wächst kontinuierlich.





Marktübersicht

Die Übersicht zeigt einige Vertreter der vier NoSQL Kategorien aus dem Abschnitt „Definition“.

Alternativen

NoSQL-DBs sind die Alternative zu RDBMS in Bereichen mit zuvor schwer oder nicht zu bewältigenden Anforderungen. Bei extremen Skalierungsanforderungen gibt es praktisch keine Alternative zu NoSQL. Zu beobachten ist der Trend der hybriden Datenbanken, die eventuell eine nützliche Kombination aus beiden Welten werden. Neu ist die „NewSQL“-Bewegung, deren Vertreter besonders skalierbare RDBMS mit neuartiger Architektur anbieten. Zum Teil wird neueste Hardware wie z.B. Solid State Discs als Grundlage verwendet.

Referenzszenario

Google, Facebook, Twitter, Amazon würden ohne NoSQL nicht existieren. Das CERN (Genf) verwendet NoSQL für den Large Hadron Collider. Nicht jeder hat solche extremen Anforderungen. Dennoch eignen sich NoSQL-Technologien auch für viele alltäglichere Probleme:

- Caching von extrem oft gelesenen Daten
- Empfehlungssysteme, Konfiguratoren
- Indexieren sehr großer Datenmengen
- Sammeln von beliebig großen Zeitreihen, Historien, Log-Daten, Sensordaten, Geodaten und deren Batch-Auswertung
- Online-Verarbeitung tief vernetzter Daten
- Speichern von ungleichmäßigen, gemischten, unbegrenzt wachsenden Daten
- Mobile-Device-Anwendungen mit Offline-Fähigkeit und automatischer Replikation

Business Impact

Mit NoSQL-DBs können innovative Produkte entstehen, da ihre Spezialisierung die Verarbeitung schwieriger Daten überhaupt erst ökonomisch möglich macht. Die Open-Source-Produkte können Vorteile bei den Kosten haben. Die Flexibilität der schema-freien DBs kann die Agile Software-Entwicklung beschleunigen und die Erschließung neuer Märkte erleichtern, deren Anforderungen erst erarbeitet werden müssen. Die Unabhängigkeit von proprietärer Hardware für die Skalierung ist ebenfalls ein ökonomischer Vorteil. Einfachere Architekturmuster für Web-Anwendungen können Zeit und Geld sparen.

Pro	Contra
Horizontale Skalierbarkeit, extreme Lese- und Schreib-Performance (Big Data)	Verzicht auf gewohnte Garantien der RDBMS, wie z. B. Transaktionssicherheit
Flexible Datenmodelle ohne Schema eignen sich für iterative Entwicklung	Design richtet sich z. T. nach vorher zu definierenden Abfragen und nicht nach einem „Modell der Welt“
Stärkere Spezialisierung mit weniger Ballast, der selten benutzt wird	Mehr Verantwortung innerhalb der Anwendungen; schmalerer Einsatzbereich
Keine Produktanschaffungskosten bei Open Source	Weniger verbreitete, unterschiedliche Technologien
Meist sehr leichter Zugang ohne spezielle Treibersoftware	Noch exotische Datenmodellierungsstrategien und Betriebsmodelle

msh systems ag

Robert-Bürkle-Straße 1 | 85737 Ismaning/München
 Telefon: +49 89 96101-0 | Fax: +49 89 96101-1113
 www.msh-systems.com | info@msh-systems.com

Stand: März 2012

<http://www.msh-systems.com/techrefresh>

