**wichtige Begriffe INSY – NoSQL**

# Big Data

Der Begriff Big Data beschreibt Datenbestände, die aufgrund ihres Umfangs, ihrer Struktur bzw. ihrer Schnelllebigkeit nur begrenzt oder gar nicht mit relationalen Datenbanken verarbeitet werden können.

# 3V

## Volume

Mit Volume ist die schiere Menge an Daten gemeint, die pro Zeiteinheit verarbeitet werden muss.

## Velocity

Velocity beschreibt die Geschwindigkeit, mit der Daten verarbeitet werden müssen.

## Variety

Variety beschreibt die unterschiedlichen Grade der Strukturierung der Daten.

# NoSQL-Systeme

NoSQL Systeme sind Informationssysteme die entwickelt wurden, um den Herausforderungen des Bigdata zu begegnen. NoSQL ist ein Acronym und steht für not only sql.

# Verteilte Datenbank

Ein Verteiltes Datenbanksystem ist ein Informationssystem, das auf mehrere Netzwerkknoten ausgeführt werden kann. Die Daten der Datenbank werden in diesem Fall verteilt über mehrere Knoten hinweg verwaltet.

# Consistency

In verteilten Datenbanken, muss sichergestellt sein, dass nach Abschluss einer Transaktion alle Replikate eines veränderten Datensatzes aktualisiert wurden.

# Availability

Verfügbarkeit beschreibt in erste Linie die Eigenschaft eines Informationssystem bei hoher Systemlast seinen Dienst aufrechterhalten zu können.

# Konsistenzmodell

Das Konsistenzmodell einer Datenbank beschreibt auf welche Weise die Konsistenz der Daten sichergestellt wird.

# Strict Consistency

Alle Anwender sehen immer dieselben Daten. Alle Leseoperationen liefern den Wert, der die letzte abgeschlossene Schreiboperation geschrieben hat.

Strict Consistency wird mit der Hilfe der ACID Transaktionskonzept umgesetzt.

# Eventual Consistency

Anwender sehen nicht immer dieselben Daten. Eine Leseoperation liefert eventuell den zuletzt geschriebenen Wert.

Eventual Consistency wird mit der Hilfe der BASE Transaktionskonzept umgesetzt.

# Transaktion

Als Transaktion bezeichnet man eine Folge von Programmschriften, die als logische Einheit betrachtet werden. Nach fehlerfreier Ausführung der Transaktion wird der Datenbestand in einem konsistenten Zustand hinterlassen.

Daher wird für eine Transaktion insbesondere gefordert, dass sie entweder vollständig und fehlerfrei oder gar nicht ausgeführt wird.

## ACID

Atomicity: Transaktionen sind unteilbar. Wenn eine Aufgabe der Transaktion erledigt ist, müssen auch alle anderen Aufgaben erledigt werden. Bei einer Paypal-Transaktion würde eventuell sonst das Geld verschwinden.

Consistency: Die Daten sind vor und nach der Transaktion logisch konsistent (Normalformen müssen erfüllt sein).

Isolation: Datenänderungen sind außerhalb der Transaktion nicht bekannt. Eine Transaktion darf keine andere beeinflussen.

Durability: Dauerhaftigkeit der Datenänderung nach Abschluss (Beendung der Transaktion).

## BASE

Basic Availability: Das Datenbanksystem ist im Sinne des CAP Theorems grundsätzlich immer verfügbar.

Soft State: Die Daten des Datenbanksystems befinden sich zeitweise in einem inkonsistentem Zustand.

Eventually Consistent: Nach dem Verständnis von BASE wird die Konsistenz der Daten als Zustand betrachtet, der irgendwann erreicht wird.

Es wird in Kauf genommen, dass die Daten eine gewisse Zeit inkonsistent sind.

# Optimistic Locking

Unter Optimistic Locking versteht man eine Sammlung von Verfahren, die die Konsistenz von Daten sicherstellen, ohne Sperren zu verwenden.

## MVCC

Multiversion Concurrency Controll ist ein Verfahren für effiziente, parallele Verarbeitung von Daten. Dabei verzeichtet MVCC dabei in Gänze auf Sperren.

## Consistent Hashing

Consistent Hashing ist ein Verfahren zur Verteilung von Datensätzen in einem verteilten System.

# Map Reduce Algorithmus

Map Reduce ist ein Programmiermodell für nebenläufige Berechnungen von großen Datenmengen auf Computerclustern.

# Skalierbarkeit

Skalierbarkeit bezeichnet die Fähigkeit eines Systems mit wachsender Systemlast seinen Dienst aufrecht halten zu können.

# Modellierungsprinzipien

## Aggregation

In dokumentbasierten Systemen werden in Beziehung stehende Daten in einem einzelnen Dokument verdichtet. Im Gegensatz zum Relationalen Modell werden Geschäftsobjekte nicht in ihre Aspekte aufgeteilt sondern als Einheit gespeichert.

## Relationfree

Vermeiden Sie Beziehungen zwischen Dokumenten.

Ist es für eine Anwendung notwendig Beziehungen zwischen Dokumenten zu definieren, sollte die Modellierung der Daten überdacht werden.

## Duplication

Im hierarchischen Datenmodell können dieselben Daten mehrmals gespeichert werden. Das hierarchische Datenmodell kennt keine Normalformen.

Datenredundanz ist akzeptabel.

## Hierarchical

Datensätze in Dokumentorientierten Systemen sind hierarchisch aufgebaut.

Ein Personendatensatz wird beispielsweise den Namen und andere personenbezogene Daten auf höchster Ebene speichern. Kontaktdaten werden als Subdokument eingefügt.

# Schemadesign

Schemadesign beschreibt den Prozess der Abbildung des fachlichen Modells auf das pyhsikalische Modell eines Informationsystems.

# Schemafreiheit

Schemafreiheit beschreibt die Eigenschaft eines Informationssystems, Datensätze beim Einfügen bzw. Ändern keiner Strukturprüfung zu unterziehen.

# Database

In der Theorie versteht man unter einer Datenbank einen logisch zusammenhängenden Datenbestand. Die Datenbank ist in diesem Sinne ein logischer Namensraum für die Collections eines Geschäftsfalls.

Eine Datenbank dient als logischer Namensraum für die enthaltenen Collections.

# Collection

Collections verwalten Sammlungen gleichartiger Dokumente.

Collections können dabei konzeptionell mit den Tabellen Relationaler Datenbanken verglichen werden.

# View

Views sind Sichten auf Collections. Wie in relationalen Datenbanken erlauben Views lediglich einen lesenden Zugriff auf den Datenbestand.

# Capped Collection

Eine Capped Collection ist eine Collection die strukturell wie ein Ringbuffer aufgebaut ist. Sie wird verwendet um große Mengen von Daten schnell verarbeiten zu können.

# Document

MongoDB speichert Datensätze in Form von Dokumenten.

# Embedded Document

In einem BSON Objekt können andere Objekte gespeichert werden. Eingebettete Objekte werden als Embedded Objects bezeichnet.

# Aufzählungstyp

Ein Aufzählungstyp ist ein Datentyp mit festgelegten Wertemöglichkeiten.

Enums werden verwendet um die logische Struktur und Lesbarkeit von Programmen zu verbessern.