

BestShift



Machbarkeitsstudie

Abteilung der Informationstechnologie am
Technologischem Gewerbemuseum

September 28, 2015

Contents

Table of Contents	i
1 Introduction	1
2 Datenmanagement & Datenanalyse	3
2.1 Datenmanagement	3
2.1.1 Vergleich NoSQL und relational [1]	3
2.1.2 NoSQL Datenbanken	6
2.1.3 Rationale Datenbanken	6
2.1.4 Data Mining & Data Analysis	7

1 Introduction

To-Do

2 Datenmanagement & Datenanalyse

In dem folgenden Kapitel werden einige Technologien, Libraries Packages, sowie generelle Software gegenübergestellt. Diese werden dann subjektiv bewertet, und die am besten geeigneten werden für das Projekt verwendet.

2.1 Datenmanagement

2.1.1 Vergleich NoSQL und relational [1]

SQL Datenbanken

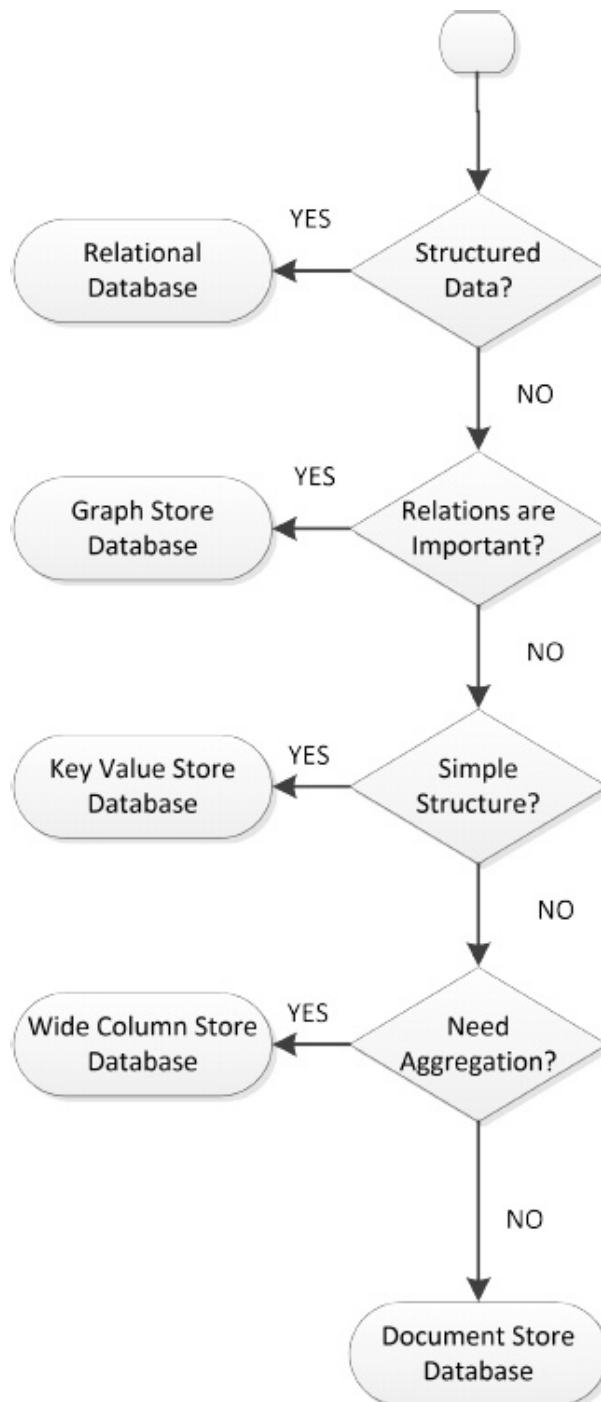
- Typen: Nur eine Art mit kleinen Unterschieden
- Data Storage Model: Individuelle Einträge werden als Reihen (Rows) in Tabellen gespeichert, wobei jede Zelle spezifische Daten über diesen Eintrag beinhaltet.
- Schemas: Struktur und daten type sind im vorhinein festgelegt. Um diese Datenbank Struktur zu ändern muss diese zunächst offline gesetzt werden.
- Skalierbarkeit: Ein einzelner Server muss hierfür mehr Rechenleistung erhalten um größeren Ansprüchen nachzukommen. Es ist prinzipiell möglich SQL Datenbanken auf ein verteiltes System zu erstellen, hierfür werden aber sehr gute Kenntnisse benötigt.
- Data Manipulation: Mittels SELECT, INSERT oder UPDATE
- Konsistenz: Gute konsistenz kann prinzipiell in allen gänglichen DBMS konfiguriert werden.

NoSQL Datenbanken

- Typen: Viele verschiedene, bspw. key-value, document-based, oder graph datenbank.
- Data Storage Model: Hängt vom Typ der Datenbank ab.
- Schemas: Typischerweise dynamisch. Einträge können on-the-fly hinzugefügt werden.
- Skalierbarkeit: Bei Bedarf kann ein Administrator einfach mehrere Cloud Instanzen hinzufügen. Die Datenbank an sich verteilt die Information auf die notwendige Server Anzahl
- Data Manipulation: Durch Objektorientierte APIs
- Konsistenz: Abhängig vom Product

Conclusio

Anhand dem folgenden Graphik wird dann die Entscheidung getroffen, welcher Typ von Datenbank den Anforderungen entspricht



2.1.2 NoSQL Datenbanken

	Cassandra	HBase	MongoDB	Riak	MySQL Cluster	Couchbase
Funktionalität						
Zuverlässigkeit						
Benutzbarkeit						
Efizienz						
Änderbarkeit						
Übertragbarkeit						

2.1.3 Rationale Datenbanken

	Sybase	IBM DB2	Oracle	Microsoft SQL Server	MySQL	PostgreSQL
Funktionalität						
Zuverlässigkeit						
Benutzbarkeit						
Efizienz						
Änderbarkeit						
Übertragbarkeit						

2.1.4 Data Mining & Data Analysis

Durch die Sensoren im Auto sowie durch die zusätzlich durch den CarPC angebrachten wird eine enorme Menge an Daten geliefert. All diese Daten ergeben jedoch erst einen Sinn; wenn wir sie mit geeigneten Verfahren analysieren und auswerten können.

Frameworks

	RapidMiner	WEKA	R-Programming	Orange	KNIME	NLTK
Funktionalität						
Zuverlässigkeit						
Benutzbarkeit						
Efizienz						
Änderbarkeit						
Übertragbarkeit						