

# Datenbanken Allgemein

*Jan-Philipp Kolb*

*9 Mai 2017*

## Was sind Datenbanken?

- Eine Datenbank, auch Datenbanksystem (DBS) genannt, ist ein System zur elektronischen Datenverwaltung.
- Die wesentliche Aufgabe eines DBS ist es, große Datenmengen effizient, widerspruchsfrei und dauerhaft zu speichern.

## Wann sollte man R um Datenbanken ergänzen?

- Wenn ein Datensatz in den Arbeitsspeicher passt, gibt es keinen Grund eine Datenbank zu nutzen.
- die Nutzung von Schnittstellen zu Datenbanken macht den Code nur komplizierter und langsamer

Man nutzt die Schnittstelle zu Datenbanken,...

- weil die Daten bereits in einer Datenbank vorgehalten werden
- oder weil der Datensatz nicht in den Arbeitsspeicher passt

## Die drei großen Open-Source Datenbanken

- SQLite, MySQL und PostgreSQL
- für alle drei gibt es Anbindungen in R

```
install.packages("RSQLite")
install.packages("RMySQL")
install.packages("Rpostgresql")
```

- in der Folge liegt der Fokus vor allem auf SQLite und PostgreSQL

## SQLite

- SQLite - Open Source Programmbibliothek mit relationalem Datenbanksystem
- SQLite ist eine schlanke Datenbank und man muss nichts weiter installieren um sie zu nutzen.
- SQLite ist schon in R und Python integriert

## MySQL Datenbank

- Grundlage für dynamische Webauftitte
- am meisten verbreitetes Datenbanksystem
- allerdings schwieriger zu konfigurieren als die anderen SQL Datenbanken SQLite und PostgreSQL

## PostgreSQL

- PostgreSQL ist in den meisten Linux-Distributionen enthalten.
- Bietet eine Schnittstellen zu vielen Programmiersprachen



Figure 1:

## Vergleich zwischen MySQL und PostgreSQL

### Beispiel zu relationalen Datenbanken

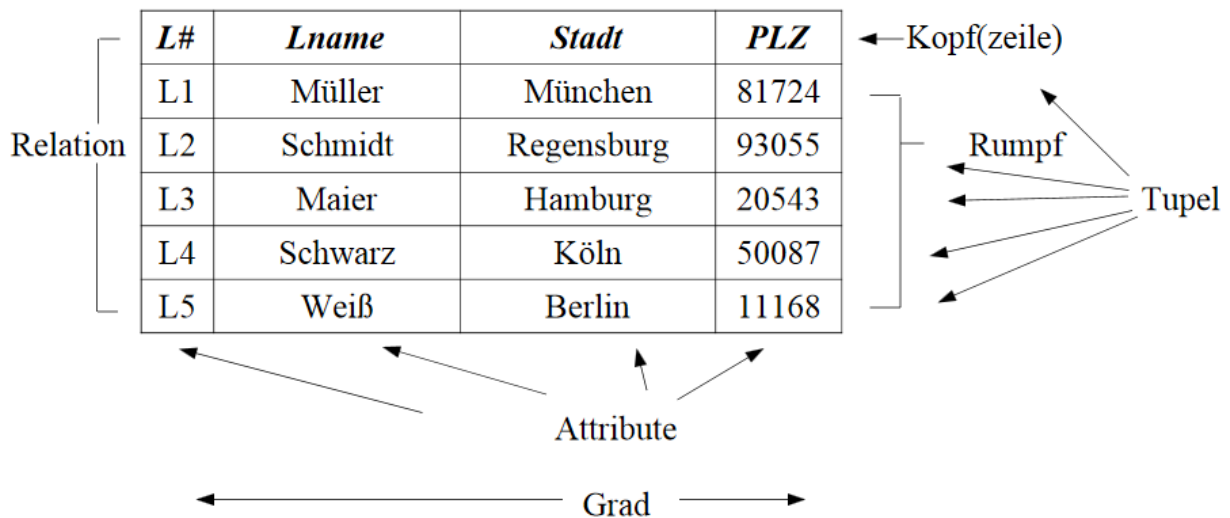


Figure 2:

## Was ist der Unterschied zwischen SQL und NoSQL

- NoSQL steht für not only SQL
- Key-Value-Stores (bspw. CouchDB, MongoDB) und Speicherung unstrukturierter Daten wird durch Schema Evolution ermöglicht
- mit NoSQL lassen sich deutlich größere Datenmengen händeln

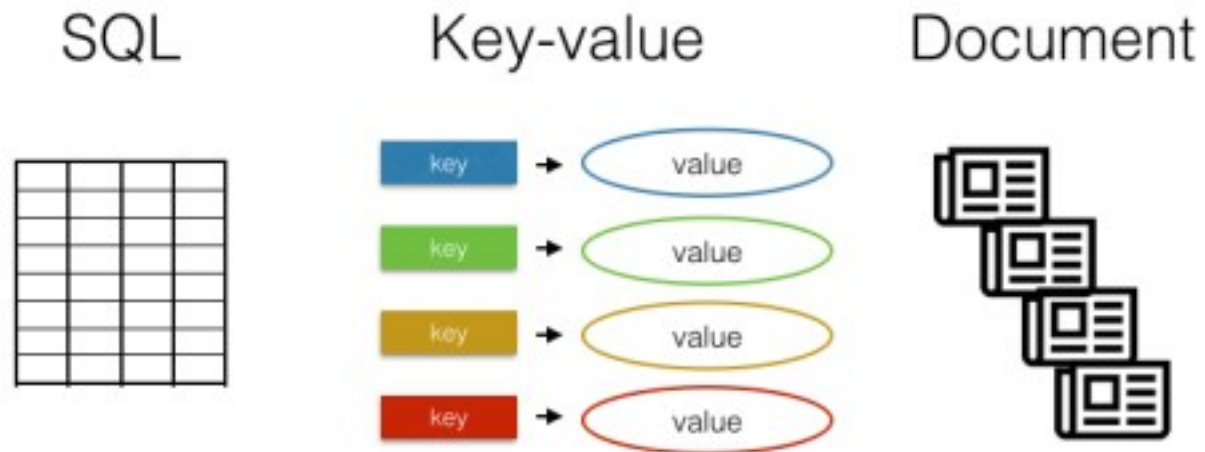


Figure 3:

- Horizontale Skalierbarkeit - wichtig bei Daten wie Video, Audio oder Bild-Dateien
- NoSQL-Bewegung ist nicht proprietär an einen Hersteller gebunden

## MongoDB

- MongoDB - kann sehr schnell und einfach installiert und benutzt werden.
- Schema-freie, dokumentenorientierte NoSQL-Datenbank
- kann Sammlungen von JSON-ähnlichen Dokumenten verwalten

## CouchDB

- dokumentenorientierte Datenbank
- zur Interaktion mit CouchDB kann das Paket `sofa` verwendet werden

## Podcast zu CouchDB

## Quick-R zur Integration von Datenbanken

## SQL lernen...

- ...mit W3School
- ...mit SQLzoo
- ...SQL lernen mit tutorialspoint
- ... mit Beispielen von sql tutorial

## CRE125 CouchDB

Veröffentlicht am **10. Juni 2009**

**Die moderne Key/Value-Datenbank lädt Entwickler zum Entspannen ein**

CouchDB ist ein vergleichsweise neues Projekt im Bereich der Datenbanken, dass durch eine Reihe frischer Ansätze mit den etablierten Konzepten relationaler Datenbanken bricht. CouchDB speichert Daten dokumentenorientiert und ohne feste Struktur und orientiert sich in seinem Kommunikationsmodell am World Wide Web und nutzt dessen Standards wie z.B. HTTP. Im Gespräch mit Tim Pritlove



Figure 4:

## Access to Database Management Systems (DBMS)

### ODBC Interface

The **RODBC** package provides access to databases (including Microsoft Access and Microsoft SQL Server) through an ODBC interface.

Figure 5:

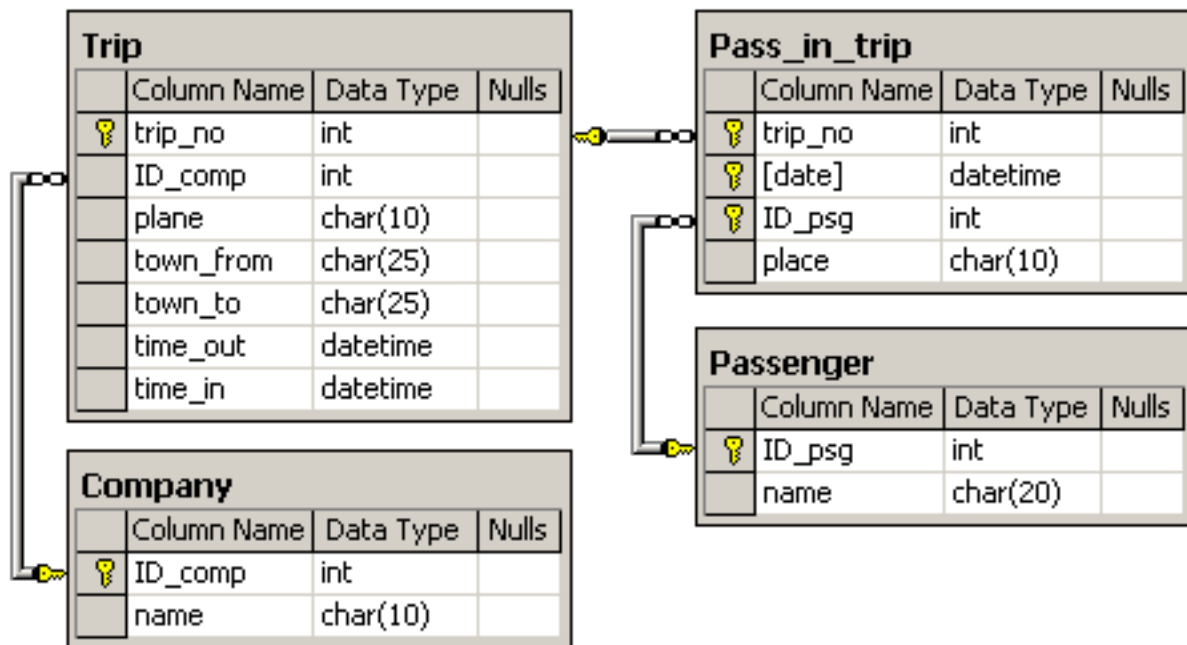


Figure 6:

## How to use SQL Union in Set Operations

In this puzzle, we're going to learn how to write a SQL UNION without using Set Operations. Set operations allow us to compare rows from two or more tables to arrive at a result. For several classes of problems, it is much easier to use a set operation, than join data. But can we [...]

[Continue reading](#)

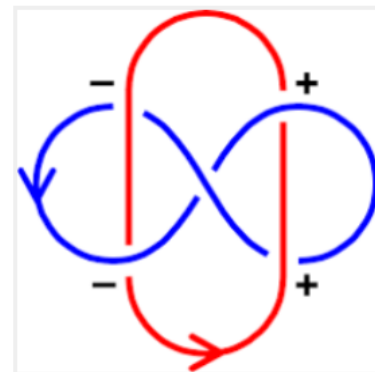


Figure 7:

## Weitergehendes Lernen

### Weitere Ressourcen

- 10 einfache Schritte um SQL zu verstehen
- Abfrage planen
- Video um NoSQL zu verstehen