

# NoSQL

Relationale Datenbanken sind unflexibel und skalieren schlecht horizontal. Um hier Abhilfe zu schaffen, gibt es NoSQL-Datenbanken mit weniger starrer Struktur und Optimierung für horizontale Skalierung.

## Arten von NoSQL-Datenbanken

- Key-Value: unter einem Schlüssel liegt ein Wert mit unbekannter Struktur
- Dokumentenorientiert: unter einem Schlüssel liegt ein Dokument mit abfragbarer Struktur
- Column Stores: Key-Value in geschachtelt bzw. mehrdimensional
- Graph: Knoten-Objekte und deren Beziehungen

## Map-Reduce

- Idee: Programm kommt zu den Daten und kann auf vielen Maschinen parallel ausgeführt werden
- Es müssen nur zwei Funktionen implementiert werden
  - die Map-Funktion destilliert aus Key-Value-Paaren die gewünschte Sichtweise oder Aufteilung in Form von Key-Value Paaren
  - die Reduce-Funktion fasst alle Werte eines Keys zu einem Ergebnis zusammen. Gegebenenfalls lässt sie sich mehrfach hintereinanderreihen, um sukzessive die Ergebnisse zu sammenzuführen

Beispiel zur Berechnung eines Durchschnittswertes:

```
def map(self, data):  
    yield ('count', (len(data[1]), 1))  
  
def reduce(self, key, values, rereduce):  
    return (sum(value[0] * value[1] for value in values) /  
            sum(value[1] for value in values),  
            sum(value[1] for value in values))
```