

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

### Definition

Unterschied

Anwendung

Datenherkunft

Wachstum von  
Daten

### Analysemethoden

### Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

R

# Big Data

Isabella Aspodinger, Alexander Pilan

Paris Lodron Universität Salzburg

22. November, 2019

# Inhalt

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Unterschied

Anwendung

Datenherkunft

Wachstum von  
Daten

## Analysemethoden

## Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

R

- 1 Definition
  - Unterschied
  - Anwendung
  - Datenherkunft
  - Wachstum von Daten
- 2 Analysemethoden
- 3 Entwicklungen
  - NoSQL
  - JSON
  - Map Reduce
  - Hadoop
  - R

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

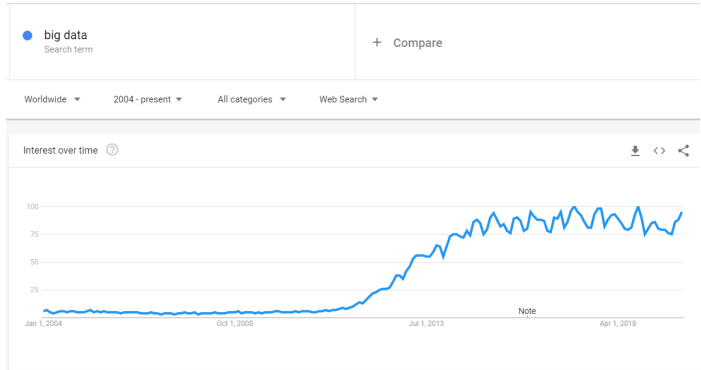
### Definition

Unterschied  
Anwendung  
Datenherkunft  
Wachstum von  
Daten

### Analysemethoden

### Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
R



# Definition

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Unterschied  
Anwendung  
Datenherkunft  
Wachstum von  
Daten  
Analysemethoden  
Entwicklungen  
NoSQL  
JSON  
Map Reduce  
Hadoop  
R

- Volume (Datenvolumen)
- Velocity (Geschwindigkeit der Datenverarbeitung und Veränderungsdynamik)
- Variety (Vielfalt der Datenstrukturen und -klassen)
- Veracity (Echtheit der Daten)
- Value (unternehmerischer Mehrwert)
- Validity (Datenqualität)

# Unterschied

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

### Definition

#### Unterschied

#### Anwendung

#### Datenherkunft

#### Wachstum von Daten

### Analysemethoden

### Entwicklungen

#### NoSQL

#### JSON

#### Map Reduce

#### Hadoop

#### R

## Traditionelle Analytik

- Schrittweise Analyse der kleinen Datenmengen
- Abfassung und Sortierung bevor Bearbeitung
- Daten werden angesammelt, bearbeitet, gespeichert und erst dann analysiert

## Big Data Analytik

- Bearbeitung der ganzen Datenmenge
- Daten werden unverändert bearbeitet
- Analyse und Bearbeitung werden je nach Einlauf durchgeführt

# Anwendung

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Unterschied

## Anwendung

Datenherkunft

Wachstum von  
Daten

## Analysemethoden

## Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

R

- 1 Kundenanalyse
- 2 Risikoanalyse
- 3 Standortbasiertes Targeting
- 4 Kampagneoptimierung
- 5 Produktplatzierungsoptimierung
- 6 Kriminalistik
- 7 Medizin

# Datenherkunft

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Unterschied

Anwendung

**Datenherkunft**

Wachstum von  
Daten

## Analysemethoden

## Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

R

- 1 Aufzeichnungen verschiedener Überwachungssysteme.
- 2 die Nutzung von Kunden- oder Bank- bzw. Bezahlkarten
- 3 die Nutzung eines Smartphones
- 4 Social-Media
- 5 Kraftfahrzeuge
- 6 vernetzte Technik in Häusern
- 7 von Behörden und Unternehmen erhobene und gesammelte Daten.

# Wachstum von Daten

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Unterschied

Anwendung

Datenherkunft

## Wachstum von Daten

## Analysemethoden

## Entwicklungen

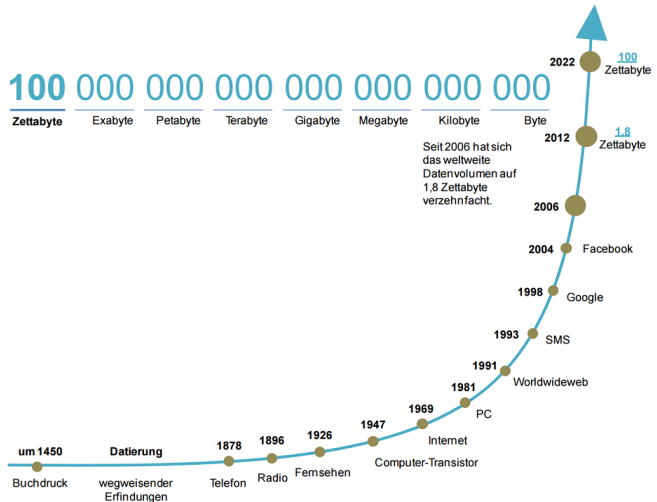
NoSQL

JSON

Map Reduce

Hadoop

R





# Analysemethoden

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Unterschied  
Anwendung  
Datenherkunft  
Wachstum von  
Daten

## Analysemethoden

## Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
R

- Repräsentative Stichprobe
- Kopplungsanalyse
- Predictive Analytics
- Data Mining

# Entwicklungen

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Unterschied

Anwendung

Datenherkunft

Wachstum von  
Daten

## Analysemethoden

## Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

R

- NOSQL (Not Only SQL)
- JSON
- Map Reduce
- Hadoop
- Spark
- R

# NoSQL

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Unterschied

Anwendung

Datenherkunft

Wachstum von  
Daten

## Analysemethoden

## Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

R

- Objektdatenbanken
- Grid- und Cloud-Datenbanken
- XML-Datenbanken
- Andere nicht-relationale Datenbanken

# NoSQL

## Kriterien

### Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

### Definition

Unterschied

Anwendung

Datenherkunft

Wachstum von  
Daten

### Analysemethoden

### Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

R

- Nichtrelationales Datenmodell
- Schemafrei (oder nur schwache Restriktionen)
- Bieten einfache API
- Verteilte Architektur, optimiert für einfache Replikation und horizontale
- Skalierung
- Kein ACID-Konsistenzmodell
- Open Source

# NoSQL

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

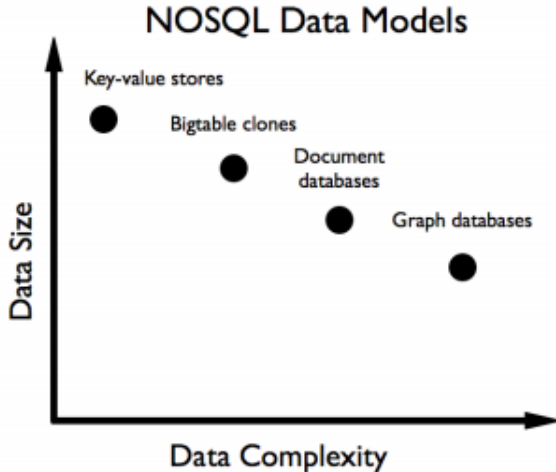
## Definition

Unterschied  
Anwendung  
Datenherkunft  
Wachstum von  
Daten

## Analysemethoden

## Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
R



# JavaScript Object Notation

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

JSON ist ein kompaktes Datenformat in einer einfach lesbaren Textform zum Zweck des Datenaustauschs zwischen Anwendungen.

```
{
  "Herausgeber": "Xema",
  "Nummer": "1234-5678-9012-3456",
  "Deckung": 2e+6,
  "Waehrung": "EURO",
  "Inhaber":
    {
      "Name": "Mustermann",
      "Vorname": "Max",
      "maennlich": true,
      "Hobbys": ["Reiten", "Golfen", "Lesen"],
      "Alter": 42,
      "Kinder": [],
      "Partner": null
    }
}
```

## Definition

Unterschied

Anwendung

Datenherkunft

Wachstum von  
Daten

## Analysemethoden

## Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

R

# Map Reduce

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

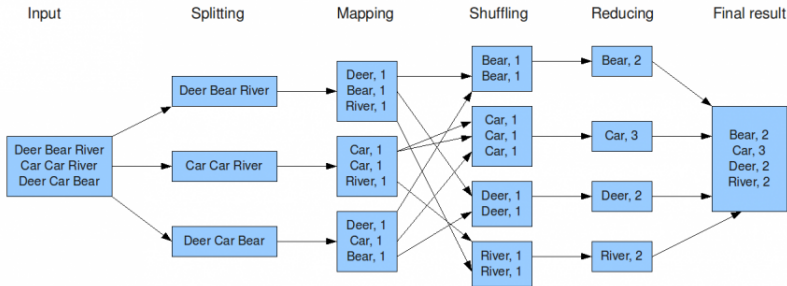
Unterschied  
Anwendung  
Datenherkunft  
Wachstum von  
Daten

## Analysemethoden

## Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
R

The overall MapReduce word count process



# Map Reduce

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Wortzahl

### Definition

Unterschied  
Anwendung  
Datenherkunft  
Wachstum von  
Daten

### Analysemethoden

### Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
R

```
1 function map(String name, String documentPart):  
2     for each word w in documentPart:  
3         emit (w, 1)  
4  
5 function reduce(String word, List<Int>  
6     partialCounts):  
7     sum = 0  
8     for each pc in partialCounts:  
9         sum += pc  
10        emit (word, sum)
```



# Hadoop

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Unterschied

Anwendung

Datenherkunft

Wachstum von

Daten

## Analysemethoden

## Entwicklungen

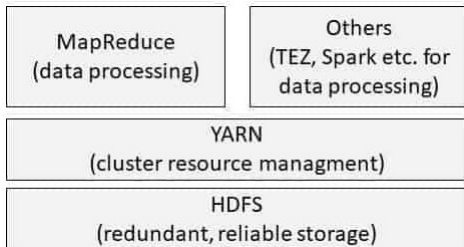
NoSQL

JSON

Map Reduce

Hadoop

R



- HDFS (Hadoop Distributed File System)
- YARN (Yet Another Resource Negotiator)
- Map Reduce

# HDFS

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

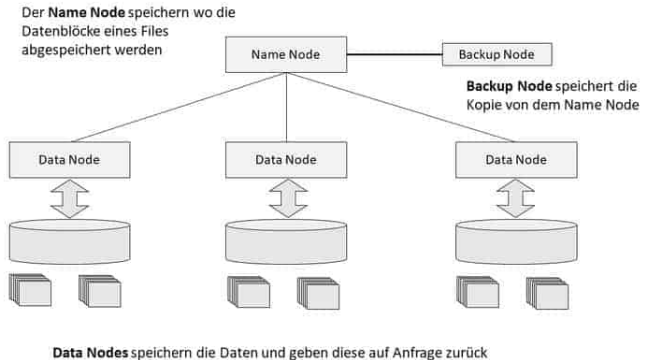
## Definition

Unterschied  
Anwendung  
Datenherkunft  
Wachstum von  
Daten

## Analysemethoden

## Entwicklungen

NoSQL  
JSON  
Map Reduce  
**Hadoop**  
R



## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Unterschied  
Anwendung  
Datenherkunft  
Wachstum von  
Daten

## Analysemethoden

## Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
R

## Paradigmen:

- funktional
- dynamisch
- objektorientiert

## R Beispiel Code

```
1 | Gewicht <- c(60, 72, 57, 90, 95, 72)
2 | Groesse <- c(1.75, 1.80, 1.65, 1.90, 1.74, 1.91)
3 | BMI <- Gewicht / Groesse^2
4 | sum(Gewicht)
5 | length(Gewicht)
6 | sum(Gewicht) / length(Gewicht)
7 | table(Gewicht)
```