

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

### Definition

Anwendung  
Datenherkunft  
Wachstum von  
Daten  
Unterschied

### Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
Spark

# Big Data

Isabella Aspodinger, Alexander Pilan

November 20, 2019

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Anwendung  
Datenherkunft  
Wachstum von  
Daten  
Unterschied

## Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
Spark

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

# Definition

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Anwendung  
Datenherkunft  
Wachstum von  
Daten  
Unterschied

## Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
Spark

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

# Anwendung

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

### Definition

#### Anwendung

Datenherkunft

Wachstum von  
Daten

Unterschied

### Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

Spark

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

# Datenherkunft

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Anwendung

**Datenherkunft**

Wachstum von  
Daten

Unterschied

## Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

Spark

- 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

Anwendung

Datenherkunft

Wachstum von  
Daten

Unterschied

## Entwicklungen

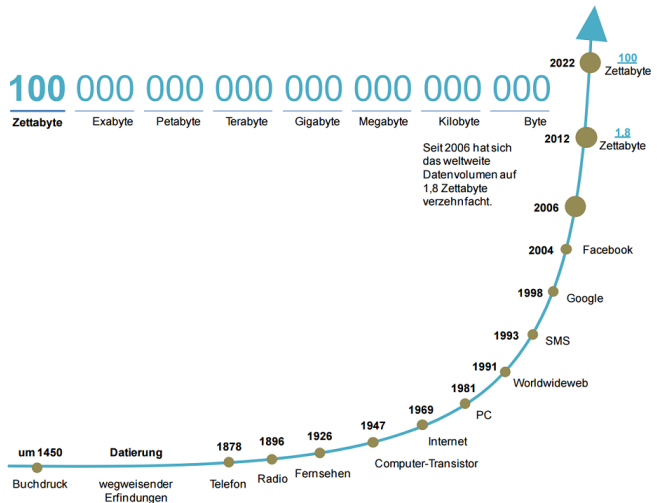
NoSQL

JSON

Map Reduce

Hadoop

Spark



# Unterschied

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Anwendung

Datenherkunft

Wachstum von  
Daten

Unterschied

## Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

Spark

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

# Entwicklungen

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

### Definition

Anwendung  
Datenherkunft  
Wachstum von  
Daten  
Unterschied

### Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
Spark

Klassische relationale Datenbanksysteme sowie Statistik- und Visualisierungsprogramme sind oft nicht in der Lage, derart große Datenmengen zu verarbeiten. Für Big Data kommen daher neue Arten von Datenspeicher- und Analyse-Systemen zum Einsatz, die parallel auf bis zu Hunderten oder Tausenden von Prozessoren bzw. Servern arbeiten. Dabei gibt es u. a. folgende Herausforderungen:



## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

### Definition

Anwendung  
Datenherkunft  
Wachstum von  
Daten  
Unterschied

### Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
Spark

- Verarbeitung vieler Datensätze
- Verarbeitung vieler Spalten innerhalb eines Datensatzes
- Schneller Import großer Datenmengen
- Sofortige Abfrage importierter Daten (Realtime Processing)
- Kurze Antwortzeiten (Latenz und Verarbeitungsdauer) auch bei komplexen Abfragen
- Möglichkeit zur Verarbeitung vieler gleichzeitiger Abfragen (Concurrent Queries)
- Analyse verschiedenartiger Informationstypen (Zahlen, Texte, Bilder, ...)

# NoSQL

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Anwendung  
Datenherkunft  
Wachstum von  
Daten  
Unterschied

## Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
Spark

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

# NoSQL

## Kriterien

### Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

### Definition

Anwendung

Datenherkunft

Wachstum von  
Daten

Unterschied

### Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

Spark

- 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

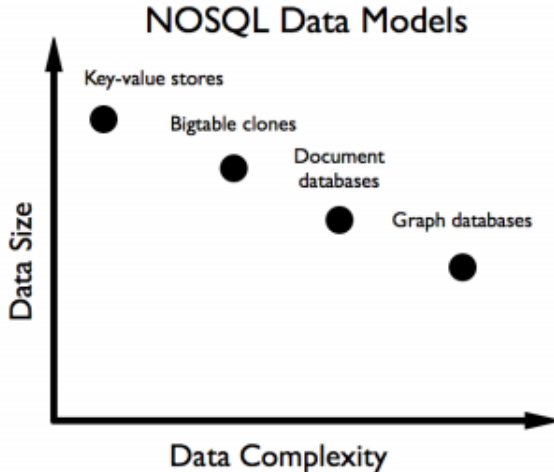
Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Anwendung  
Datenherkunft  
Wachstum von  
Daten  
Unterschied

## Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
Spark



# JSON

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Anwendung  
Datenherkunft  
Wachstum von  
Daten  
Unterschied

## Entwicklungen

NoSQL  
**JSON**  
Map Reduce  
Hadoop  
Spark

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

# Map Reduce

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

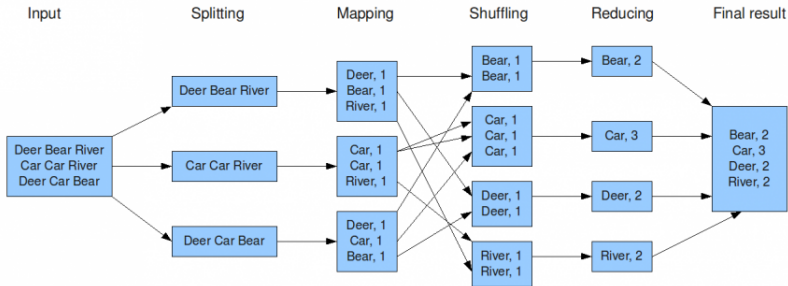
## Definition

Anwendung  
Datenherkunft  
Wachstum von  
Daten  
Unterschied

## Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
Spark

The overall MapReduce word count process



# Hadoop

## Big Data

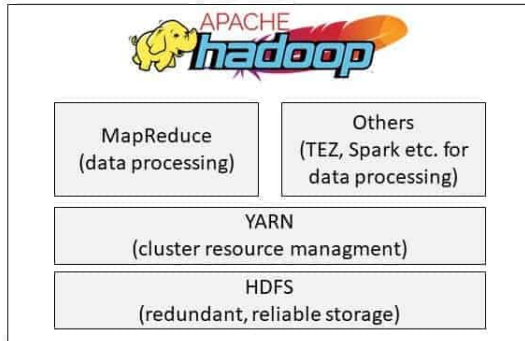
Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Anwendung  
Datenherkunft  
Wachstum von  
Daten  
Unterschied

## Entwicklungen

NoSQL  
JSON  
Map Reduce  
**Hadoop**  
Spark



- HDFS (Hadoop Distributed File System)
- YARN
- Map Reduce

# HDFS

## Big Data

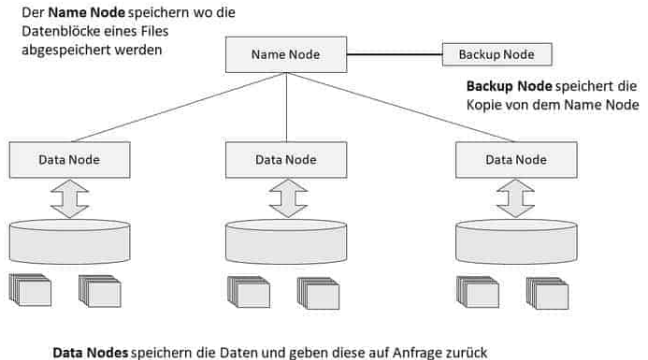
Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Anwendung  
Datenherkunft  
Wachstum von  
Daten  
Unterschied

## Entwicklungen

NoSQL  
JSON  
Map Reduce  
**Hadoop**  
Spark





# Spark

## Big Data

Isabella  
Aspodinger,  
Alexander  
Pilan

## Definition

Anwendung  
Datenherkunft  
Wachstum von  
Daten  
Unterschied

## Entwicklungen

NoSQL  
JSON  
Map Reduce  
Hadoop  
**Spark**