### Big Data

Aspodinger Alexander Pilan

#### Delinitio

Unterschied Anwendung Datenherkunft Wachstum von

#### Entwicklunge

Schwerpunkte NoSQL JSON Map Reduce Hadoop

# 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

Schwerpunkte

JSON Map Reduce Hadoop Spark

### Definition

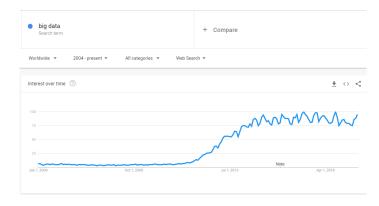
- Unterschied
- Anwendung
- Datenherkunft
- Wachstum von Daten

# 2 Entwicklungen

- Schwerpunkte
- NoSQL
- JSON
- Map Reduce
- Hadoop
- Spark
- R

### Big Data





## **Definition**

### Big Data

Isabella Aspodinger Alexander Pilan

#### Definition

Unterschied Anwendung Datenherkunft Wachstum von Daten

Schwerpunkte
NoSQL
JSON
Map Roduce

JSON Map Reduce Hadoop Spark 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

Unterschied
Anwendung
Datenherkunft
Wachstum von
Daten

Schwerpunkte
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 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

Datenherkunf Wachstum vo Daten

Schwerpunkte

JSON Map Reduce

Map Reduc Hadoop Spark

- Kundenanalyse
- Risikoanalyse
- Standortbasiertes Targeting
- Kampagneoptimierung
- Produktplatzierungsoptimierung
- Kriminalistik
- Medizin

### Datenherkunft

### Big Data

Isabella Aspodinger Alexander Pilan

Definition Unterschied Anwendung

Datenherkunft Wachstum von Daten

Entwicklunge Schwerpunkte NoSQL JSON Map Reduce 4 Aufzeichnungen verschiedenster Überwachungssysteme.

4 die Nutzung von Kunden- oder Bank- bzw. Bezahlkarten

die Nutzung eines Smartphones

Social-Media

Kraftfahrzeuge

o vernetzte Technik in Häusern

 von Behörden und Unternehmen erhobene und gesammelte Daten.

### Wachstum von Daten

Big Data

Isabella Aspodinger, Alexander Pilan

Definition

Unterschied

Anwendun

Dateillerku

Wachstum von Daten

Entwicklung

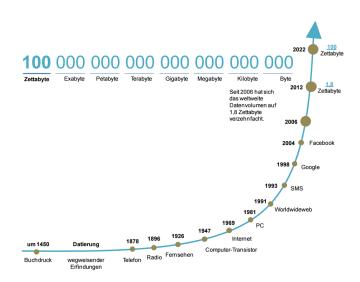
Schwerpunkt NoSOI

100042

Map Redu

Hadoop

R



# Entwicklungen

### Big Data

Isabella Aspodinge Alexander Pilan

#### Definiti

Unterschie

Anwendun

Datenherk

Wachstum vo

# Entwicklungen

NoSQL

Map Reduc

Hadoop

• NOSQL

- JSON
- Map Reduce
- Hadoop
- Spark
- R

# Schwerpunkte

### Big Data

Isabella Aspodinger Alexander Pilan

Definition
Unterschied
Anwendung
Datenherkunft
Wachstum von
Daten

Schwerpunkte
NoSQL
JSON
Map Reduce

### • 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 Aspodinge Alexander Pilan

### Definitio

Unterschied

Datanbark

Wachstum v

Daten

### Schwerpunkte

NoSQL

Map Reduc

Hadoop

Spai 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

Schwerpunkte
NoSQL
JSON

Map Reduce Hadoop Spark

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

# **NoSQL**

### Big Data

Isabella Aspodinger Alexander Pilan

Definition

Unterschied

Datashadaa

Wachstum w

Daten

#### Entwicklung

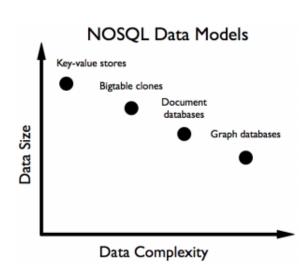
Schwerpuni

NoSQL

Map Redu

Hadoop

R



# **JSON**

### Big Data

Isabella Aspodinge Alexander Pilan

Definition Unterschied

Anwendung

Datenherkunf

Wachstum vo

Daten

Entwicklunger

Schwerpunkte NoSQL

JSON Map Reduc

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

# Map Reduce

### Big Data

Aspodinger Alexander Pilan

#### Definition

Unterschied

Anwendun

Datenherkui

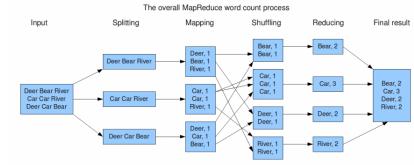
Wachstum v

### Entwicklunge

NoSQL

Map Reduce

Hadoop Spark



# Hadoop

### Big Data

Isabella Aspodinge Alexander Pilan

### Definitio

Anwendung

Datenherkunft

Wachstum von

### Schwernunkte

NoSQL JSON

Map Reduc

Spar R

# MapReduce (data processing)

Others (TEZ, Spark etc. for data processing)

### YARN

(cluster resource managment)

### **HDFS**

(redundant, reliable storage)

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

# **HDFS**

### Big Data

Isabella Aspodinger Alexander Pilan

Definition

Unterschied

Anwendur

Wachstum v

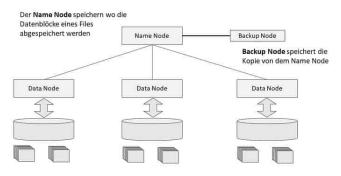
Entwicklung

Schwerpunkte

Map Reduc

Hadoop

- Spari



Data Nodes speichern die Daten und geben diese auf Anfrage zurück

# Spark

### Big Data

Isabella Aspodinger, Alexander Pilan

#### Definition

Unterschied

Datenherkun

Wachstum voi

#### Entwicklunger

Schwerpunkte

NoSQL

Map Reduc

Hadoop

Spa

### Big Data

Isabella Aspodinger Alexander Pilan

### Definition Unterschied

Anwendung
Datenherkunft
Wachstum von
Daten

#### Entwicklunge

NoSQL JSON Map Reduce Hadoop Spark 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)</pre>
```