

Big Data

Isabella
Aspodinger,
Alexander
Pilan

Definition

Unterschied

Anwendung

Datenherkunft

Wachstum von
Daten

Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

R

Big Data

Isabella Aspodinger, Alexander Pilan

Paris Lodron Universität Salzburg

22. November, 2019

Big Data

Isabella
Aspodinger,
Alexander
Pilan

Definition

Unterschied

Anwendung

Datenherkunft

Wachstum von
Daten

Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

R

- 1 Definition
 - Unterschied
 - Anwendung
 - Datenherkunft
 - Wachstum von Daten

- 2 Entwicklungen
 - NoSQL
 - JSON
 - Map Reduce
 - Hadoop
 - R

Big Data

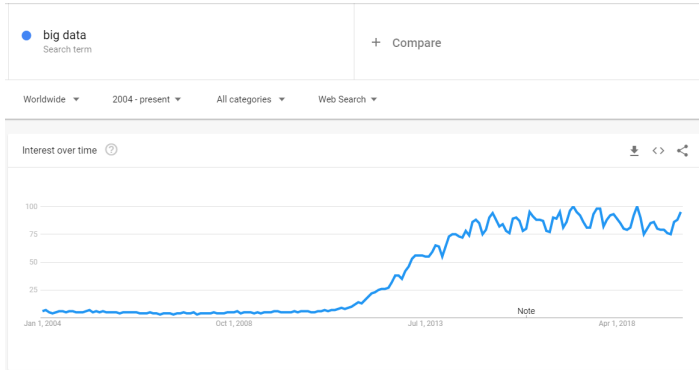
Isabella
Aspodinger,
Alexander
Pilan

Definition

Unterschied
Anwendung
Datenherkunft
Wachstum von
Daten

Entwicklungen

NoSQL
JSON
Map Reduce
Hadoop
R



Definition

Big Data

Isabella
Aspodinger,
Alexander
Pilan

Definition

Unterschied
Anwendung
Datenherkunft
Wachstum von
Daten

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

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

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

Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

R

- 1 Aufzeichnungen verschiedenster 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

Entwicklungen

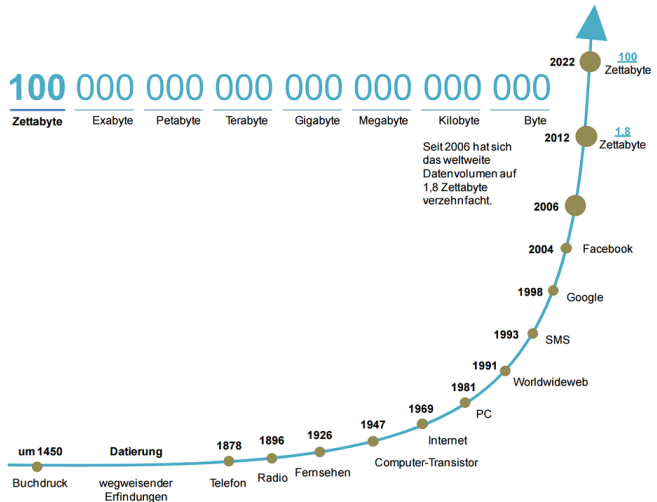
NoSQL

JSON

Map Reduce

Hadoop

R



Entwicklungen

Big Data

Isabella
Aspodinger,
Alexander
Pilan

Definition

Unterschied

Anwendung

Datenherkunft

Wachstum von
Daten

Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

R

- NOSQL
- JSON
- Map Reduce
- Hadoop
- Spark
- R

Analysemethoden

Big Data

Isabella
Aspodinger,
Alexander
Pilan

Definition

Unterschied
Anwendung
Datenherkunft
Wachstum von
Daten

Entwicklungen

NoSQL
JSON
Map Reduce
Hadoop
R

- Kopplungsanalyse
- Predictive Analytics
- Data Mining

NoSQL

Big Data

Isabella
Aspodinger,
Alexander
Pilan

Definition

Unterschied

Anwendung

Datenherkunft

Wachstum von
Daten

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

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

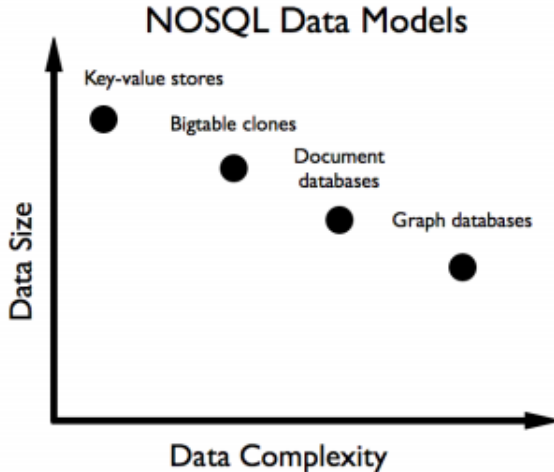
Isabella
Aspodinger,
Alexander
Pilan

Definition

Unterschied
Anwendung
Datenherkunft
Wachstum von
Daten

Entwicklungen

NoSQL
JSON
Map Reduce
Hadoop
R



JavaScript Object Notation

Big Data

Isabella
Aspodinger,
Alexander
Pilan

Definition

Unterschied

Anwendung

Datenherkunft

Wachstum von
Daten

Entwicklungen

NoSQL

JSON

Map Reduce

Hadoop

R

Map Reduce

Big Data

Isabella
Aspodinger,
Alexander
Pilan

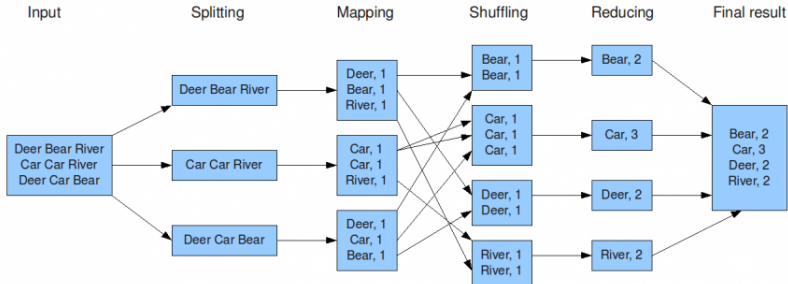
Definition

Unterschied
Anwendung
Datenherkunft
Wachstum von
Daten

Entwicklungen

NoSQL
JSON
Map Reduce
Hadoop
R

The overall MapReduce word count process



Hadoop

Big Data

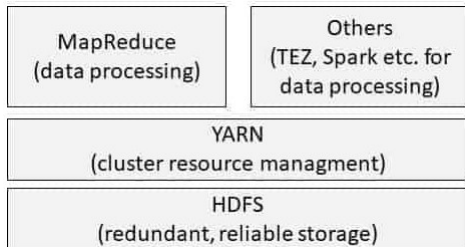
Isabella
Aspodinger,
Alexander
Pilan

Definition

Unterschied
Anwendung
Datenherkunft
Wachstum von
Daten

Entwicklungen

NoSQL
JSON
Map Reduce
Hadoop
R



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

HDFS

Big Data

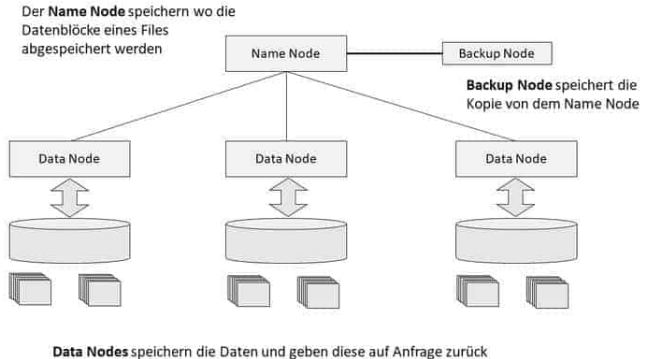
Isabella
Aspodinger,
Alexander
Pilan

Definition

Unterschied
Anwendung
Datenherkunft
Wachstum von
Daten

Entwicklungen

NoSQL
JSON
Map Reduce
Hadoop
R



Big Data

Isabella
Aspodinger,
Alexander
Pilan

Definition

Unterschied
Anwendung
Datenherkunft
Wachstum von
Daten

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)
```