Big Data

Big Data

Isabella Aspodinger, Alexander Pilan

Paris Lodron Universität Salzburg

22. November, 2019

Inhalt

Big Data

Isabella Aspodinger Alexander Pilan

Definitio

Anwendung Datenherku

Datenherkunft Wachstum voi Daten

Entwicklunger

JSON Map Reduce

R

Definition

- Unterschied
- Anwendung
- Datenherkunft
- Wachstum von Daten

2 Entwicklungen

- NoSQL
- JSON
- Map Reduce
- Hadoop
- R

Big Data

Isabella Aspodinger Alexander Pilan

Definitio

Unterschied

. .

Datenherkun

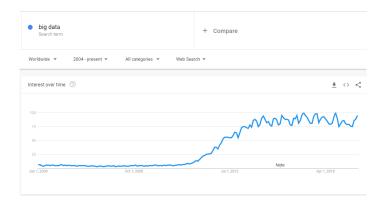
Wachstum vo

Liitwickiunge

NoSQL ISON

Map Reduc

R



Definition

Big Data

Isabella Aspodinger Alexander Pilan

Definition

Unterschied Anwendung Datenherkunft Wachstum von Daten

NoSQL JSON

JSON Map Reduce Hadoop

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

Unterschied

Big Data

Isabella Aspodinger, Alexander Pilan

Unterschied Anwendung Datenherkunft Wachstum von Daten

NoSQL JSON Map Reduce Hadoop

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

Anwendung

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

Datenherkunft

Big Data

Isabella Aspodinger Alexander Pilan

Definition Unterschied Anwendung

Datenherkunft Wachstum von Daten

NoSQL JSON 4 Aufzeichnungen verschiedenster Überwachungssysteme.

② die Nutzung von Kunden- oder Bank- bzw. Bezahlkarten

die Nutzung eines Smartphones

Social-Media

Kraftfahrzeuge

o vernetzte Technik in Häusern

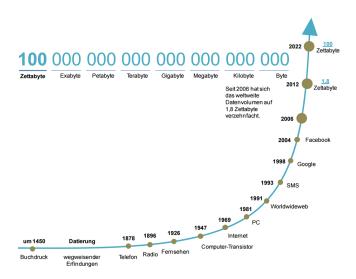
von Behörden und Unternehmen erhobene und gesammelte Daten.

Wachstum von Daten

Big Data

Wachstum von

Daten



Entwicklungen

Big Data

Isabella Aspodinge Alexander Pilan

Definiti

Unterschie

Anwendun

Datennerkui

Wachstum vo Daten

Entwicklungen

JSON

Map Reduc

R

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

Analysemethoden

Big Data

Isabella Aspodinger Alexander Pilan

Deminio

Unterschied

Anwendun

Datanbark

Wachstum v

Entwicklungen

Littwicklunge

JSON

Map Reduc

К

- Kopplungsanalyse
- Predictive Analytics
- Data Mining

NoSQL

Big Data

Aspodinge Alexande Pilan

Definitio

Anwendung

Datenherkunft

Wachstum von

Entwicklunger

NoSQL JSON Map Reduce Hadoop 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

- 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

Definitior

_ -----

Ontersenice

Datonbork

Wachstum v

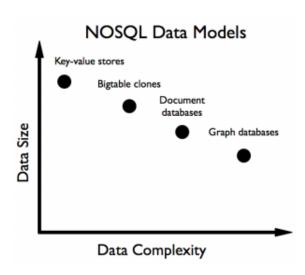
Entwicklung

NoSQL

JSON

Map Reduc

₹



JavaScript Object Notation

Big Data

Isabella Aspodinger, Alexander Pilan

Definition

Unterschied

Datenherkur

Wachstum vo

Daten Von

Entwicklunger

.

JSON

Map Reduc

--

Map Reduce

Big Data

Isabella Aspodinge Alexander Pilan

Definition

Anwendun

Datenherku

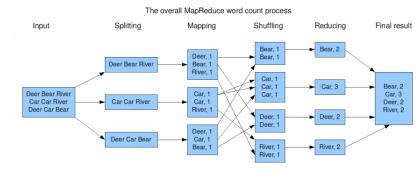
Wachstum vo

Entwicklunge

NoSQL

Map Reduce

Hadoop



Hadoop

Big Data

Isabella Aspodinger Alexander Pilan

Definition

Anwendung

Datenherkunft

Wachstum von

Entwicklunge

JSON Map Reduce Hadoop 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

2

Anwendu

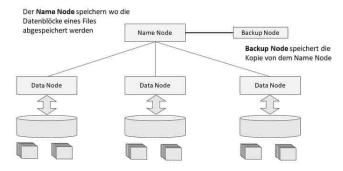
Wachstum vo

Entwicklunge

NoSQL

Map Reduce

R



Data Nodes speichern die Daten und geben diese auf Anfrage zurück

Big Data

Isabella Aspodinger, Alexander Pilan

Definitio Unterschied

Anwendung Datenherkunft Wachstum von Daten

Entwicklunge

JSON Map Reduce Hadoop

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>
```