Big Data

Aspodinger Alexander Pilan

Definition

Unterschied Anwendung Datenherkunf

Entwicklunge

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

Definitio

Anwendung Datenherkui

Datenherkunft Wachstum von Daten

Entwicklunge

JSON Map Reduce Hadoop

Definition

- Unterschied
- Anwendung
- Datenherkunft
- Wachstum von Daten

2 Entwicklungen

- NoSQL
- JSON
- Map Reduce
- Hadoop
- Spark

Definition

Big Data

Isabella Aspodinger Alexander Pilan

Definition

Unterschied Anwendung Datenherkunft Wachstum von Daten

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)

Unterschied

Big Data

Isabella Aspodinger, Alexander Pilan

Unterschied
Anwendung
Datenherkunft
Wachstum von
Daten

Entwicklunger
NoSQL
JSON
Map Reduce

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

Entwicklunge NoSQL JSON Map Reduce 4 Aufzeichnungen verschiedenster Überwachungssysteme.

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

die Nutzung eines Smartphones

Social-Media

Kraftfahrzeuge

vernetzte Technik in Häusern

von Behörden und Unternehmen erhobene und gesammelte Daten.

Wachstum von Daten

Big Data

Isabella Aspodinger, Alexander Pilan

Definition

D 0.....

Anwendur

Datennerku

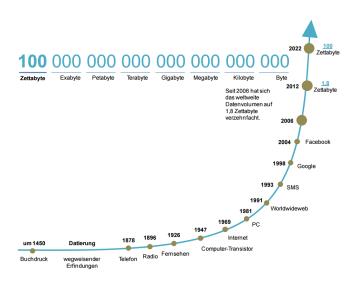
Wachstum von Daten

Entwicklunge

NoSQL

Map Redu

Hadoop Spark



Entwicklungen

Big Data

Isabella Aspodinger Alexander Pilan

Definition
Unterschied
Anwendung
Datenherkunft
Wachstum von
Daten

Entwicklungen
NoSQL
JSON

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
Unterschied
Anwendung
Datenherkunft
Wachstum von
Daten

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

Aspodinge Alexande Pilan

Definitio

Datenherkunft Wachstum von Daten

Entwicklunge

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

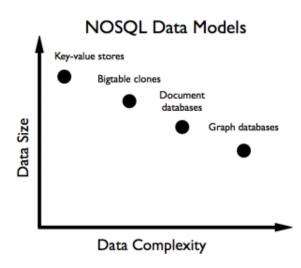
JSON Map Reduce Hadoop

- 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

NoSQL



JSON

Big Data

Isabella Aspodinge Alexander Pilan

Definitio

Anwendung
Datenherkunf
Wachstum vo
Daten

Entwicklunger

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

Map Reduce

Big Data

Aspodinger Alexander Pilan

Definition

Anwendun

Datenherku

Wachstum vo Daten

Entwicklunge

NoSQL

Map Reduce

The overall MapReduce word count process Input Splitting Mapping Shuffling Reducing Final result Bear, 2 Bear, 1 Deer, 1 ► Bear, 1 Deer Bear River Bear, 1 River, 1 Car, 1 Car. 3 Bear, 2 Car. 1 Deer Bear River Car, 1 Car, 3 Car, 1 Car Car River Car Car River Car. 1 Deer, 2 Deer Car Bear River, 1 River, 2 Deer, 1 Deer, 2 Deer, 1 Deer, 1 Deer Car Bear Car, 1 River, 2 Bear, 1 River, 1 River, 1

Hadoop

Big Data

Isabella Aspodinger Alexander Pilan

Definition

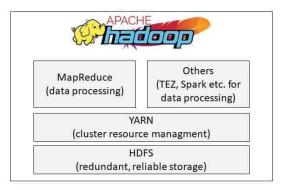
Unterschie

Datenherkur

Wachstum vo Daten

Entwicklunger

JSON Map Reduc Hadoop



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

HDFS

Big Data

Isabella Aspodinger Alexander Pilan

Definition

Haramakina.

Anwendur

Datennerkun

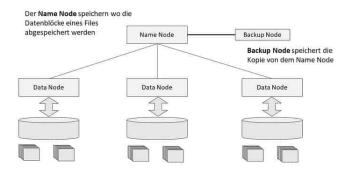
Wachstum vo Daten

Entwicklunge

NoSQL JSON

Map Redu Hadoop

Spark



Data Nodes speichern die Daten und geben diese auf Anfrage zurück

Spark

Big Data

Isabella Aspodinger, Alexander Pilan

Definition

Unterschied

Datenherkun

Wachstum vo

Entwicklungen

NoSQL

Map Reduc

Spark