

Andrea Wist

Entwicklung von Big-Data-Anwendungen

Masterarbeit

Bibliografische Information der Deutschen Nationalbibliothek:

Bibliografische Information der Deutschen Nationalbibliothek: Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de/> abrufbar.

Dieses Werk sowie alle darin enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsschutz zugelassen ist, bedarf der vorherigen Zustimmung des Verlanges. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen, Auswertungen durch Datenbanken und für die Einspeicherung und Verarbeitung in elektronische Systeme. Alle Rechte, auch die des auszugsweisen Nachdrucks, der fotomechanischen Wiedergabe (einschließlich Mikrokopie) sowie der Auswertung durch Datenbanken oder ähnliche Einrichtungen, vorbehalten.

Copyright © 2021 Diplom.de
ISBN: 9783961164363

Andrea Wist

Entwicklung von Big-Data-Anwendungen

Inhaltsverzeichnis

Abkürzungsverzeichnis	III
Formelverzeichnis	IV
1 Die Bedeutung von Big Data	1
1.1 Motivation und Ziel der Arbeit	1
1.2 Methodischer Aufbau der Arbeit	3
2 Aspekte der Datenerzeugung und Datenhaltung	5
2.1 Daten und Informationen	5
2.1.1 Strukturierte Daten	6
2.1.2 Unstrukturierte Daten	6
2.1.3 Datenklassen	7
2.2 Speichermedien und Datenhaltungssysteme	8
2.2.1 Speicherkapazität	8
2.2.2 Computer Cluster	10
2.2.3 Datenbanksysteme	11
3 Anforderungen an Big-Data-Anwendungen	13
3.1 Die 5 V's	13
3.1.1 Volume	13
3.1.2 Veracity	14
3.1.3 Variety	15
3.1.4 Velocity	16
3.1.5 Value	17
3.2 Notwendige Eigenschaften jeder Anwendung	17
3.2.1 Allgemeingültigkeit	18
3.2.2 Belastbarkeit	19
3.2.3 Skalierbarkeit und Erweiterbarkeit	19
3.2.4 Fehlertoleranz	20
3.2.5 Lesen und Aktualisieren mit geringen Latenzzeiten	21
3.2.6 Ad-hoc-Abfragen und Echtzeitauswertungen	21
4 Umsetzung von Big-Data-Anwendungen	22
4.1 Lambda-Architektur	22
4.1.1 Batch-Layer	23
4.1.2 Serving-Layer	24
4.1.3 Speed-Layer	25

4.2	Datenspeicherung	27
4.2.1	Rohdaten werden zu Stammdaten	27
4.2.2	Stammdaten speichern	28
4.2.3	Datenverteilung	29
4.3	Big-Data-Analysen	30
4.3.1	MapReduce Algorithmus	30
4.3.2	Korrelation und Kausalität	32
4.3.3	Marktforschung	34
4.3.4	Verkehrssteuerung	35
4.3.5	Big Data in der Medizin	36
4.4	Rechtliche Grundlagen	38
4.4.1	Privatsphäre	39
4.4.2	Zweckbindung der Daten	40
4.4.3	Datensparsamkeit	41
4.4.4	Social Media Analysen	43
5	Kernbausteine von Big-Data-Anwendungen	45
5.1	Technische Notwendigkeiten	45
5.1.1	Skalierbare Datenverarbeitung	46
5.1.2	Redundante Datenhaltung	46
5.1.3	Daten sind unveränderlich	47
5.1.4	Parallelisierung der Operation	48
5.1.5	Optimierung der Datenstrukturen	49
5.2	Konzeptionelle Notwendigkeiten	50
5.2.1	Datenqualität	50
5.2.2	Zusammenhänge richtig deuten	51
5.2.3	Privacy by Design	52
6	Zusammenfassung und Ausblick	55
6.1	Zusammenfassung	55
6.2	Ausblick	57
	Literatur	V

Abkürzungsverzeichnis

API Application Programming Interface

BDSG Bundesdatenschutzgesetz

BGB Bürgerliches Gesetzbuch

B2C Business-to-Consumer

CPU Central Processing Unit

CT Computertomographie

DFS Data Filesystem

DSGVO Datenschutz-Grundverordnung

ETL Extract, Transform, Load

GPS Global Positioning System

HDD Hard Disk Drive

HDFS Hadoop Distributed File System

I/O Input/Output-Durchsatz

IT Informationstechnik

MS Massenspektrometrie

NoSQL Not only SQL

RAM Random Access Memory

SPOF Single Point of Failure

SSD Solid State Disk

SQL Structured Query Language

IoT Internet of Things

XML Extensible Markup Language

Formelverzeichnis

1	Abfrage aus der gesamten Datenmenge	22
2	Erstellung vom Batch-View	23
3	Abfrage aus dem Batch-View	23
4	Erstellung der Echtzeit-View	26
5	Abfrage aus alle Daten in Echtzeit	26