



**Hochschule
Bonn-Rhein-Sieg**
University of Applied Sciences

Semester-Projekt

Aufbau einer NoSQL-Datenbank mit Apache Hadoop und Apache HBase für das Million Song Dataset

- Schemalose Datenbanken -

Fachbereich Informatik
Referent: Prof. Dr. Harm Knolle

eingereicht von:
Moritz Kemp, Johann Martens
Matr.-Nr.

Sankt Augustin, den 12.10.2016

Zusammenfassung

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugait nulla facilisi.

Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis.

Vorwort

Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet. Lorem ipsum dolor sit amet, consetetur sadipscing elitr, sed diam nonumy eirmod tempor invidunt ut labore et dolore magna aliquyam erat, sed diam voluptua. At vero eos et accusam et justo duo dolores et ea rebum. Stet clita kasd gubergren, no sea takimata sanctus est Lorem ipsum dolor sit amet.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugait nulla facilisi. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat.

Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat. Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis at vero eros et accumsan et iusto odio dignissim qui blandit praesent luptatum zzril delenit augue dui dolore te feugait nulla facilisi.

Nam liber tempor cum soluta nobis eleifend option congue nihil imperdiet doming id quod mazim placerat facer possim assum. Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper suscipit lobortis nisl ut aliquip ex ea commodo consequat.

Duis autem vel eum iriure dolor in hendrerit in vulputate velit esse molestie consequat, vel illum dolore eu feugiat nulla facilisis.

Inhaltsverzeichnis

Zusammenfassung	II
Vorwort	III
1 Expose	1
1.1 Ziel	1
1.2 Vorgehensweise	1
A Anhang	2
A.1 Erster Teil	2
A.2 Zweiter Teil	2
B Anhang	3
Abbildungsverzeichnis	4
Tabellenverzeichnis	5
Listings	6
Abkürzungsverzeichnis	6
Eidesstattliche Erklärung	8

1 Expose

1.1 Ziel

Das Ziel der Projektarbeit zum Thema *Hadoop/Hbase* ist es, eine Datenbank mittels Hadoop/Hbase zu implementieren, die eine Lieder-Datenbank darstellt. Dabei soll die Datenbank so installiert werden, dass sie ihre Stärken mit sehr großen Datensätzen, in diesem Falle dem Million-Song-Datensatz, ausspielen kann.

1.2 Vorgehensweise

Das Team wird sich zuerst mit den Grundlagen von NoSQL-Datenbanken befassen, vor allem explizit mit den Grundlagen von Hadoop/Hbase. Darauf aufbauend lässt sich eine Argumentation formulieren, warum die Behandlung von derart großen Datensätzen wie der Million-Song-Datensatz sich mit Hadoop/Hbase effektiver gestalten lässt als gegenüber den klassischen relationalen Datenbanken. Dieser Teil beinhaltet also im wesentlichen Grundlagen zu den Wide-Column-Datenbanken sowie Grundlagen zu Hadoop und Hbase.

Im zweiten Teil soll dann darauf aufbauend die praktische Umsetzung durchgeführt und beschrieben werden. *Johann, your part! Whoop whooop*

Die geplante Gliederung sieht wie folgt aus:

- Grundlagen von Wide-Column-Datenbanken (only MapReduce, 6C)
- Grundlagen Hadoop/Hbase (6D1/2)
- Anforderungen an die Anwendung
- Aufbau der Datenbank (6F)
- Installation
- Cluster-Betrieb
- Benchmarks

1.3 Meilensteine

A Anhang

Appendix

A.1 Erster Teil

A.2 Zweiter Teil

B Anhang

Appendix

Abbildungsverzeichnis

Tabellenverzeichnis

Listings

Abkürzungsverzeichnis

API	Application Programming Interface
BASE	Basically Available, Soft State, Eventual Consistency
BG	Barahmand Ghandeharizadeh
CAP	Consistency Availibiltiy Partition Tolerance
CLI	Command-Line Interface
CPU	Central Processing Unit
CRUD	Create, Read, Update, Delete
DBA	Datenbankadministrator
DBS	Datenbanksystem
DNS	Domain Name System
DTD	Document Type Definition
GUI	Graphical User Interface
HDD	Hard Disk Drive
IP	Internet Protocol
JDBC	Java Database Connectivity
JSON	JavaScript Object Notation
NoSQL	Not only SQL
OLTP	Online Transaction Processing
RDBMS	Relational Database Management System
RFC	Request For Comments
SLA	Service Level Agreement
SPEC	Standard Performance Evaluation Corporation
SQL	Structured Query Language
SSD	Solid State Drive
TPC	Transaction Processing Performance Council
UML	Unified Markup Language
XML	Extensible Markup Language
YCSB	Yahoo! Cloud Serving Benchmark

Literaturverzeichnis

Eidesstattliche Erklärung

Ich versichere an Eides Statt durch meine eigenhändige Unterschrift, dass ich die vorliegende Arbeit selbstständig und ohne fremde Hilfe angefertigt habe. Alle Stellen, die wörtlich oder dem Sinn nach auf Publikationen oder Vorträgen anderer Autoren beruhen, sind als solche kenntlich gemacht. Ich versichere außerdem, dass ich keine andere als die angegebene Literatur verwendet habe. Diese Versicherung bezieht sich auch auf alle in der Arbeit enthaltenen Zeichnungen, Skizzen, bildlichen Darstellungen und dergleichen.

Die Arbeit wurde bisher keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

den 18. Oktober 2016
Ort, Datum

Moritz Kemp, Johann
Martens