

Hochschule fÄijr Technik, Wirtschaft und Kultur Leipzig

FakultÄd't Informatik, Mathematik und Naturwissenschaften

Masterstudiengang Informatik

Masterarbeit

zur Erlangung der akademischen Grades

Master of Science (M.Sc.)

Untersuchung und Optimierung verteilter Geografischer Informationssysteme zur Verarbeitung Agrartechnischer Kennzahlen

Eingereicht von: Kurt Junghanns

Matrikelnummer: 59886

Leipzig 9. Oktober 2014

Erstprüfer: Prof. Dr. rer. nat. Thomas Riechert

Zweitprüfer: M. Sc. Volkmar Herbst

Abstrakt

Danksagung

Vorwort

Abbildungsverzeichnis

Tabellenverzeichnis

Inhaltsverzeichnis

Abbildungsverzeichnis	v
Tabellenverzeichnis	vi
1 Einleitung	1
1.1 Motivation	1
1.2 Zielsetzung	1
2 Grundlagen	2
2.1 Datenbank	2
2.1.1 ACID	2
2.1.2 MVCC	2
2.1.3 BASE	2
2.1.4 weitere Begriffsdefinitionen	3
2.1.5 Indexstrukturen	3
2.1.6 Mehrrechner-Datenbanksystem	3
2.1.7 Verteiltes Datenbanksystem	3
2.1.8 Replikationsverfahren	3
2.2 geografische Datenverarbeitung	4
2.2.1 Bezugssysteme	4
2.2.2 Datenformate	4
2.2.3 GIS	4
2.2.4 PostGIS	4
2.3 NoSQL	4
2.3.1 Definition	4
2.3.2 Hadoop	4

Inhaltsverzeichnis

2.3.3	NoSQL GIS	4
2.3.4	MongoDB	4
2.3.5	CouchDB	4
2.3.6	Neo4J	4
2.3.7	Rasdaman	5
2.3.8	Spacebase	5
2.4	Leistungstests	5
3	Ausgangsszenario	6
4	Gegenüberstellung	7
4.1	Kosten	7
4.2	Umfang	7
4.3	Leistung	7
5	Fazit	8
5.1	Zusammenfassung	8
5.2	Wertung	8
5.3	Ausblick	8
	Literaturverzeichnis	I

1 Einleitung

1.1 Motivation

1.2 Zielsetzung

- Erarbeitung Grundlagen - Analyse vorhandener Systeme zum speichern und verarbeiten von räumlichen Daten - Besonderer Augenmerk auf NoSQL - Erarbeitung einer Empfehlung für das Szenario - Prototyp dazu erstellen und fehlende Teile implementieren

2 Grundlagen

Computer

2.1 Datenbank

2.1.1 ACID

Atomicity, Consistency, Isolation und Durability (ACID)

2.1.2 MVCC

Multi Version Currency Control (MVCC)

2.1.3 BASE

Basically Available, Soft state, Eventual consistency (BASE)

2.1.4 weitere Begriffsdefinitionen

2.1.5 Indexstrukturen

R-Baum

B-Baum

LSM-Baum

2.1.6 Mehrrechner-Datenbanksystem

2.1.7 Verteiltes Datenbanksystem

2.1.8 Replikationsverfahren

Synchron

Asynchron

Kaskadiert

2.2 geografische Datenverarbeitung

2.2.1 Bezugssysteme

2.2.2 Datenformate

Punkte

Vektoren

Raster

Shapefile

2.2.3 GIS

2.2.4 PostGIS

2.3 NoSQL

2.3.1 Definition

2.3.2 Hadoop

2.3.3 NoSQL GIS

2.3.4 MongoDB

2.3.5 CouchDB

2.3.6 Neo4J

2.3.7 Rasdaman

2.3.8 Spacebase

2.4 Leistungstests

- siehe BA - in Absprache mit Prof. Riechert

3 Ausgangsszenario

4 Gegenüberstellung

4.1 Kosten

4.2 Umfang

4.3 Leistung

5 Fazit

5.1 Zusammenfassung

5.2 Wertung

5.3 Ausblick

Literaturverzeichnis

Eidesstatliche Erklärung

Ich versichere, dass die Masterarbeit mit dem Titel „...“ nicht anderweitig als Prüfungsleistung verwendet wurde und diese Masterarbeit noch nicht veröffentlicht worden ist. Die hier vorgelegte Masterarbeit habe ich selbstständig und ohne fremde Hilfe abgefasst. Ich habe keine anderen Quellen und Hilfsmittel als die angegebenen benutzt. Diesen Werken wörtlich oder sinngemäß entnommene Stellen habe ich als solche gekennzeichnet.

Leipzig, 9. Oktober 2014

Unterschrift