

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
2.2.5 GeoTools	4
2.3 NoSQL	4
2.3.1 Definition	4

Inhaltsverzeichnis

2.3.2	Kategorisierung	4
2.3.3	Hadoop	4
2.3.4	Accumulo	4
2.3.5	NoSQL GIS	5
2.3.6	MongoDB	5
2.3.7	CouchDB	5
2.3.8	Neo4J	5
2.3.9	Rasdaman	6
2.3.10	Spacebase	6
2.3.11	Geomesa	6
2.4	Leistungstests	7
3	Ausgangsszenario	8
4	Gegenüberstellung	9
4.1	Kosten	9
4.2	Umfang	9
4.3	Leistung	9
5	Fazit	10
5.1	Zusammenfassung	10
5.2	Wertung	10
5.3	Ausblick	10
	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

Geohash

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

2.2.5 GeoTools

2.3 NoSQL

2.3.1 Definition

2.3.2 Kategorisierung

2.3.3 Hadoop

2.3.4 Accumulo

https://en.wikipedia.org/wiki/Apache_Accumulo

2.3.5 NoSQL GIS

2.3.6 MongoDB

2.3.7 CouchDB

2.3.8 Neo4J

2.3.9 Rasdaman

http://live.osgeo.org/de/overview/rasdaman_overview.html :

- Array-Datenbanksystem - PostgreSQL Aufsatz - Multi-Dimensionalität - eigene Anfragesprache - skalierend - unterstützt WCS Core und WCPS - Implementierte Standards: OGC WMS 1.3, WCS 2.0, WCS-T 1.4, WCPS 1.0, WPS 1.0 - Lizenz: Clients und APIs: GNU Lesser General Public License (LGPL) version 3; Server-Engine: GNU General Public License (GPL) version 3 - Unterstützte Plattformen: Linux, MacOS, Solaris - APIs: rasql, C++, Java

<http://www.rasdaman.org/> :

- open-source - "extends standard relational database systems with the ability to store and retrieve multi-dimensional raster data"

<http://www.rasdaman.de/> :

- "erlaubt die Ablage von unbeschränkt grossen multi-dimensionalen Arrays ("Rasterdaten") in einer konventionellen Datenbank"

2.3.10 Spacebase

<http://docs.paralleluniverse.co/spacebase/> :

- serverseitig - in-memory - spatial data store - ausgelegt für viele rechner und hohen Durchsatz (real-time) - 2D und 3D Objekte mit 3D bbox - load balancing enthalten - spatial queries möglich - benötigt JVM - API für Java, Ruby, Python, Node.js, C++, Erlang - API stellt nur elementare spatial queries zur verfügung: intersect oder contains - eigene spatial queries können definiert werden

2.3.11 Geomesa

http://www.eclipse.org/community/eclipse_newsletter/2014/march/article3.php :

- open-source - build on Accumulo and Hadoop - Supporting the GeoTools API - GeoServer Plugin - geohash for indexing

2 Grundlagen

<https://www.locationtech.org/proposals/geomesa> :

- outperforming postgis with geoserver

<http://de.slideshare.net/CCRinc/location-techdc-talk2-28465214> - Verwendung fraktaler Kurven - mit Spark und Scalding wesentlich schneller als PostGIS

<https://docs.google.com/presentation/d/1N00ppk8MfDs8Q-QcUIdZCSZK7YYwd9RjJoHV1V4Yqw/edit?pli=1#slide=id.p> :

-

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