

MongoDB



Dipl.-Math. Tobias Trelle ist Senior IT Consultant bei der code-centric AG, Solingen. Er ist seit knapp 20 Jahren im IT-Business unterwegs und interessiert sich für Softwarearchitekturen und skalierbare Lösungen. Zu diesen Themen veröffentlicht Tobias Artikel in Fachzeitschriften und hält Vorträge auf Konferenzen und Usergruppen. Außerdem organisiert er die Düsseldorfer MongoDB-Usergruppe.

Tobias Trelle

MongoDB

Der praktische Einstieg



dpunkt.verlag

Tobias Trelle
tobias.trelle@codecentric.de

Lektorat: René Schönfeldt
Copy-Editing: Sandra Gottmann, Münster
Herstellung: Birgit Bäuerlein
Umschlaggestaltung: Helmut Kraus, www.exclam.de
Druck und Bindung: M.P. Media-Print Informationstechnologie GmbH, 33100 Paderborn

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

ISBN 978-3-86490-153-9

1. Auflage 2014
Copyright © 2014 dpunkt.verlag GmbH
Wieblinger Weg 17
69123 Heidelberg

Die vorliegende Publikation ist urheberrechtlich geschützt. Alle Rechte vorbehalten. Die Verwendung der Texte und Abbildungen, auch auszugsweise, ist ohne die schriftliche Zustimmung des Verlags urheberrechtswidrig und daher strafbar. Dies gilt insbesondere für die Vervielfältigung, Übersetzung oder die Verwendung in elektronischen Systemen.

Es wird darauf hingewiesen, dass die im Buch verwendeten Soft- und Hardware-Bezeichnungen sowie Markennamen und Produktbezeichnungen der jeweiligen Firmen im Allgemeinen warenzeichen-, marken- oder patentrechtlichem Schutz unterliegen.

Alle Angaben und Programme in diesem Buch wurden mit größter Sorgfalt kontrolliert. Weder Autor noch Verlag können jedoch für Schäden haftbar gemacht werden, die in Zusammenhang mit der Verwendung dieses Buches stehen.

5 4 3 2 1 0

❖❖❖ Für Vera und Kelian ❖❖❖

Inhaltsverzeichnis

1	Einleitung	1
1.1	Big Data	1
1.2	NoSQL	2
1.3	Dokumentenorientierte Datenbanken	4
1.4	Verteilte Systeme und das CAP-Theorem	5
1.5	ACID vs. BASE	6
1.6	Zusammenfassung	8
2	MongoDB in 21 Minuten	9
2.1	Installation	9
2.2	Server starten	9
2.3	Mongo Shell	11
2.4	Erste Schritte	12
2.4.1	Collections	12
2.4.2	Dokumente	13
2.4.3	CRUD-Operationen	13
2.5	Daten importieren	16
2.6	Schemafreiheit	17
2.7	Zusammenfassung	18
3	Grundlegende Konzepte	19
3.1	Konnektivität	19
3.2	Datenhaltung	21
3.3	Datenbanken	23
3.4	Collections	25
3.4.1	System Collections	30
3.4.2	Capped Collections	31

3.5	Dokumente	32
3.5.1	BSON	33
3.5.2	Datentypen	34
3.5.3	ObjectId	36
3.5.4	Primärschlüssel	37
3.5.5	Eingebettete Dokumente	37
3.5.6	Arrays	38
3.5.7	Dokumentreferenzen	38
3.6	Indizes	40
3.7	Namespaces	42
3.8	Abstraktionsebenen	42
3.9	Zusammenfassung	43
4	Replikation	45
4.1	Das Oplog	45
4.2	Master-Slave-Replikation	47
4.3	Replica Sets	51
4.3.1	Erste Schritte	53
4.3.2	Zugriff durch Anwendungen	58
4.3.3	Konfiguration im Detail	61
4.3.4	Grenzen der Replikation	66
4.4	Zusammenfassung	69
5	Sharding	71
5.1	Grundlagen	71
5.2	Der Schlüssel zum Sharding	72
5.3	Konfiguration	74
5.4	Fortgeschrittene Themen	81
5.4.1	Chunk-Verwaltung	81
5.4.2	Shards hinzufügen	83
5.4.3	Shards entfernen	84
5.4.4	Tag-basiertes Sharding	85
5.4.5	Produktive Sharding-Systeme	87
5.5	Zusammenfassung	91

6	Queries	93
6.1	Grundlegendes	93
6.2	Cursor	95
6.2.1	pretty()	95
6.2.2	limit()	96
6.2.3	skip()	96
6.2.4	sort()	97
6.2.5	batchSize()	99
6.2.6	objsLeftInBatch()	99
6.2.7	readPref()	99
6.2.8	snapshot()	99
6.2.9	showDiskLoc()	100
6.2.10	addOption()	100
6.3	Suchkriterien	100
6.3.1	Gleichheit	100
6.3.2	Logische Meta-Operatoren	101
6.3.3	Vergleichsoperatoren	102
6.3.4	Eingebettete Dokumente	102
6.3.5	Arrays	103
6.3.6	Reguläre Ausdrücke	106
6.3.7	\$where	107
6.3.8	Schemafreiheit	108
6.3.9	Sonstiges	110
6.4	Indizes	110
6.4.1	Zusammengesetzte Indizes	114
6.4.2	Eingebettete Dokumente	114
6.4.3	Arrays	114
6.4.4	explain() und hint()	115
6.4.5	Covered Query	119
6.4.6	TTL-Index	120
6.4.7	Hashed Index	122
6.5	Profiling	124
6.5.1	Performance-Optimierung	127
6.5.2	Tools	130
6.6	Geodaten-Suche	132
6.6.1	Übersicht	133
6.6.2	Planare Koordinaten	133
6.6.3	Sphärische Koordinaten	138

6.7	Volltextsuche	144
6.7.1	Erstes Beispiel	145
6.7.2	Volltext-Indizes	146
6.7.3	Volltextsuche	146
6.7.4	Kombination mit anderen Suchkriterien	147
6.7.5	Sprachen und Stoppwortlisten	148
6.7.6	Negation und Suche nach Phrasen	149
6.7.7	Gewichtete Suche	150
6.7.8	Mehrsprachige Collections	151
6.8	Zusammenfassung	153
7	Manipulation von Dokumenten	155
7.1	Insert – Dokumente einfügen	155
7.1.1	Write Concern – der Konsistenzgrad	156
7.1.2	Mehrere Dokumente einfügen	158
7.2	Update – Änderungen an Dokumenten	159
7.2.1	Vollständige Ersetzung	160
7.2.2	Partielle Änderungen	160
7.2.3	Upsert – Anlegen und Ändern	168
7.2.4	FindAndModify – Suchen und Ändern	169
7.2.5	Änderungen an mehreren Dokumenten	170
7.2.6	Optimistisches Sperren	171
7.3	Save – Einfügen bzw. Ändern	173
7.4	Remove – Löschen von Dokumenten	173
7.5	Zusammenfassung	175
8	Schema-Design	177
8.1	Einleitung	177
8.2	Analyse und Modelle	178
8.2.1	Diagramme	181
8.2.2	Design by Example	182
8.3	Beziehungen	183
8.3.1	1:1-Beziehungen	184
8.3.2	1:n-Beziehungen	186
8.3.3	m:n-Beziehungen	188
8.4	Vererbung	188
8.5	Weitere Muster	191
8.5.1	Variable Dokumenteneigenschaften	191
8.5.2	Temporale Datenhaltung	193
8.5.3	Schema-Migration	193
8.6	Zusammenfassung	194

9	Aggregation von Daten	195
9.1	Abfragemethoden zur Datenaggregation	196
9.1.1	count()	196
9.1.2	distinct()	197
9.1.3	group()	197
9.2	Das Aggregation Framework	200
9.2.1	Grundkonzepte	200
9.2.2	Wie starte ich eine Aggregation?	202
9.2.3	Die Pipeline-Operatoren	203
9.2.4	Expressions	215
9.3	MapReduce	221
9.4	Zusammenfassung	238
10	Weiterführende Themen	239
10.1	GridFS	239
10.1.1	Zugriff über Treiber	241
10.2	REST-Schnittstelle	243
10.2.1	Server-Status ermitteln	243
10.2.2	Collections lesen	245
10.3	Sicherheit	246
10.3.1	Authentifizierung und Autorisierung	246
10.3.2	Netzwerk	251
10.4	Zusammenfassung	251
11	Softwareentwicklung mit MongoDB	253
11.1	Programmierbeispiele	254
11.1.1	Ruby	255
11.1.2	Java	259
11.2	Persistenz-Frameworks	263
11.2.1	Ruby	263
11.2.2	Java	263
11.3	Zusammenfassung	264

Anhang

A	Referenz der Kommandozeilenoptionen	267
A.1	Mongo – die Mongo Shell	267
A.2	mongod – Datenbank-Server	268
A.3	mongos – Router fürs Sharding	272
	Index	273