

**Florian Hopf**

# **Elasticsearch**

**Ein praktischer Einstieg**



**dpunkt.verlag**

Florian Hopf

Lektorat: René Schönfeld

Copy-Editing: Sandra Gottmann, Münster-Nienberge

Satz: Da-TeX, Leipzig

Herstellung: Frank Heidt

Umschlaggestaltung: Helmut Kraus, [www.exclam.de](http://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:

Buch 978-3-86490-289-5

PDF 978-3-86491-826-1

ePub 978-3-86491-827-8

mobi 978-3-86491-828-5

1. Auflage

Copyright © 2016 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

---

# Inhaltsverzeichnis

<b>1</b>	<b>Einführung</b> .....	<b>1</b>
1.1	Motivation .....	1
1.2	Geschichte von Elasticsearch .....	1
1.3	Ein erstes Beispiel .....	3
1.4	Anwendungsfälle .....	5
1.5	Wann Elasticsearch? .....	6
1.6	Über dieses Buch .....	7
1.7	Danksagung .....	9
<b>2</b>	<b>Eine Suchanwendung entsteht</b> .....	<b>11</b>
2.1	Die Beispielanwendung .....	11
2.2	Dokumente indizieren .....	12
2.3	Der invertierte Index .....	16
2.4	Über die Query-DSL zugreifen .....	19
2.5	Die Indizierung über das Mapping konfigurieren .....	23
2.6	Suchergebnisse sortieren und paginieren .....	28
2.7	Facetten für Suchergebnisse .....	30
2.8	Die Anwendung vereinfachen .....	33
2.9	Zusammenfassung .....	35
<b>3</b>	<b>Textinhalte auffindbar machen</b> .....	<b>37</b>
3.1	Analyzing und der invertierte Index .....	37
3.2	Sprachspezifische Verarbeitung durch Stemming .....	40
3.3	Teilbegriffe finden .....	42
3.4	Ähnliche Begriffe mit der Fuzzy-Query finden .....	48
3.5	Mit mehrsprachigen Inhalten arbeiten .....	49
3.6	Die Suche verbessern .....	51
3.7	Hervorheben von Suchbegriffen im Auszug .....	57
3.8	Autovervollständigung .....	59
3.9	Zusammenfassung .....	64

---

<b>4</b>	<b>Relevanz verstehen und beeinflussen</b> .....	<b>65</b>
4.1	Relevanz für die Nutzer .....	65
4.2	Berechnung der Relevanz .....	66
4.3	Einfluss von Abfragen auf die Relevanz .....	69
4.4	Relevanz durch Boosting beeinflussen .....	74
4.5	Funktionen zur Ergebnissortierung .....	76
4.6	Relevanz im verteilten System .....	80
4.7	Relevanz verstehen .....	81
4.8	Zusammenfassung .....	82
<b>5</b>	<b>Daten indizieren</b> .....	<b>83</b>
5.1	Indizierungsstrategien .....	83
5.2	Dokumente einzeln indizieren .....	85
5.3	Dokumente gesammelt indizieren .....	87
5.4	Externe Datenquellen anbinden .....	89
5.5	Partial Updates – Dokumente aktualisieren .....	94
5.6	Interna zur Indizierung .....	96
5.7	Zusammenfassung .....	102
<b>6</b>	<b>Elasticsearch als verteiltes System</b> .....	<b>103</b>
6.1	Shards und Replicas .....	103
6.2	Suche im verteilten System .....	115
6.3	Kommunikation im Cluster .....	122
6.4	Indizierung im verteilten System .....	131
6.5	Zusammenfassung .....	132
<b>7</b>	<b>Daten modellieren</b> .....	<b>133</b>
7.1	Einsatzfelder für Elasticsearch .....	133
7.2	Gestaltung der Indexstruktur .....	136
7.3	Mapping-Optionen .....	142
7.4	Beziehungen zwischen Dokumenten .....	146
7.5	Zusammenfassung .....	151
<b>8</b>	<b>Daten aggregieren</b> .....	<b>153</b>
8.1	Einführung .....	153
8.2	Aggregationen .....	153
8.3	Bucket-Aggregationen .....	159
8.4	Metric-Aggregationen .....	163
8.5	Aggregationen im Praxiseinsatz .....	167
8.6	Zusammenfassung .....	170

<b>9</b>	<b>Zugriff auf Elasticsearch</b> .....	<b>171</b>
9.1	Zwischenschicht zum Zugriff .....	171
9.2	Der Java-Client .....	172
9.3	Der JavaScript-Client .....	176
9.4	Client-Bibliotheken auswählen .....	177
9.5	Zusammenfassung .....	178
<b>10</b>	<b>Elasticsearch in Produktion</b> .....	<b>179</b>
10.1	Installation .....	179
10.2	Elasticsearch dimensionieren .....	182
10.3	Elasticsearch konfigurieren .....	184
10.4	Das Betriebssystem für Elasticsearch konfigurieren .....	187
10.5	Mapping-Optionen zur Kontrolle der gespeicherten Inhalte .....	188
10.6	Caches .....	191
10.7	Monitoring .....	194
10.8	Datensicherung .....	197
10.9	Zusammenfassung .....	200
<b>11</b>	<b>Zentralisiertes Logging mit Elasticsearch</b> .....	<b>201</b>
11.1	Warum zentralisiertes Logging? .....	201
11.2	Der ELK-Stack .....	202
11.3	Logstash .....	202
11.4	Kibana .....	211
11.5	Skalierbares Setup .....	217
11.6	Curator zur Indexverwaltung .....	221
11.7	Alternative zur Loganalyse: Graylog .....	222
11.8	Zusammenfassung .....	227
<b>12</b>	<b>Ausblick</b> .....	<b>229</b>
<b>A</b>	<b>Daten neu indizieren</b> .....	<b>233</b>
A.1	Neuindizierung ohne Änderungen .....	234
A.2	Neuindizierung mit Änderungen .....	235
A.3	Ausblick .....	236
<b>B</b>	<b>Der Twitter-River</b> .....	<b>237</b>
	<b>Literaturverzeichnis</b> .....	<b>239</b>
	<b>Index</b> .....	<b>251</b>