

Technologierecherche ElasticSearch

Stephan Lex

8. Mai 2015

EINLEITUNG

ElasticSearch stellt eine auf Apache Lucene¹ basierende Serverkomponente mit integrierter Search-Engine für u.a. Volltextsuche dar. Am häufigsten findet ElasticSearch Anwendung bei Problemen, die einen komplizierteren und verlässlichen Algorithmus, Daten zu suchen und analysieren, voraussetzen.

FUNKTIONSWEISE

ElasticSearch arbeitet wie andere NoSQL Lösungen ebenfalls verteilt in einem Servercluster, der aus einem oder mehreren Nodes bestehen kann. Ebenfalls kommt hier das Master-Slave Prinzip zum Einsatz, das man z.B. von Hadoop schon kennt, wobei allerdings jeder Knoten eines ElasticSearch Clusters beide Rollen übernehmen kann. Die Daten, die in einem ElasticSearch Cluster gespeichert werden wollen, werden von diesem indiziert. Ein Index bezeichnet dabei eine kontextabhängige Gruppierung vom Aufbau und Inhalt ähnlicher Daten. Auf die Daten im Cluster können mithilfe der integrierten Search-Engine umfangreiche Suchanfragen ausgeführt werden. Die Client-Server-Kommunikation findet mithilfe von REST statt, kann aber auch mit der enthaltenen Java API umgesetzt werden.

HANDHABUNG

Im Vergleich zu anderen größeren NoSQL Lösungen bietet ElasticSearch einen äußerst simplen und unkomplizierten Einstieg an. So ist es möglich mithilfe cURL² und einer Konsole in unter 10 Minuten einen voll funktionstüchtigen Cluster zu erstellen. Da ElasticSearch

¹<https://lucene.apache.org/>

²<http://curl.haxx.se/>

einen dokumentnahen Ansatz mithilfe vieler JSON Integrationen verfolgt, ist es sehr simpel die Nodes im Cluster mit Daten zu füttern. Dies kann man mithilfe eines browserintegrierten RESTClients oder mithilfe von cURL tun. Um die im Cluster gespeicherten Daten nun suchen und analysieren zu können findet die von Elasticsearch entwickelte QueryDSL Anwendung. Sie stellt eine an das JSON Format angelehnte Querying Language dar, mit der es möglich ist auch kompliziertere Anfragen relativ verständlich zu verfassen.