

Elasticsearch und die Oracle Datenbank

Die komfortable Suchmaschine in der Praxis

1. Einleitung
2. Installation
3. Daten aus der Oracle Datenbank
4. praktische Beispiele
5. Fragen & Anregungen

Was ist Elasticsearch?

→ Dokumentation: <https://www.elastic.co>

In Java geschriebene Suchmaschine auf der Basis von Apache Lucene [<http://lucene.apache.org/>]

Erzeugt Indizes beliebiger Art, die hochverfügbar als „shard“ auf mehrere Server verteilt werden können.

Speichert die Daten im JSON Format und gibt diese über eine RESTful Schnittstelle wieder als JSON aus

Kommunikation mittels cURL ist weit verbreitet und gut dokumentiert

Unabhängig von installierter Middleware, benötigt weder Tomcat noch WebLogic, lediglich ein aktuelles Java muss auf dem Server installiert sein.

Kann man sicher auch mit Oracle Text machen, 😊 Elasticsearch hat aber wenig Anforderung an Hardware, kommt ohne DB-Connect aus und bietet viele Erweiterungsmöglichkeiten

Was ist Elasticsearch?

→ Dokumentation: <https://www.elastic.co>

Daten werden an Elasticsearch gesendet und dort zu einem Index aufbereitet

Die Daten werden analysiert und ähnlich wie in Oracle Text zerschnitten, so wird aus dem Begriff „Oracle Text“ z.B. „Oracle“ und „Text“. Leerzeichen, Bindestriche, Doppelpunkte sind typische „Bruchstellen“. Großbuchstaben werden in Kleinbuchstaben gewandelt.

Installation aktuelle Version ist 1.7

auf Linuxbasierten Systemen sehr einfach

Public Key für das Elasticsearch Repository installieren

→ `sudo rpm --import http://packages.elasticsearch.org/GPG-KEY-elasticsearch`

Repository Eintrag erstellen

→ `sudo vi /etc/yum.repos.d/elasticsearch.repo`

```
[elasticsearch-1.5]
name=Elasticsearch repository for 1.5.x packages
baseurl=http://packages.elastic.co/elasticsearch/1.5/centos
gpgcheck=1
gpgkey=http://packages.elastic.co/GPG-KEY-elasticsearch
Enabled=1
```

Elasticsearch installieren

→ `sudo yum -y install elasticsearch`

Installation aktuelle Version ist 1.7

Konfigurationsdatei anpassen

- wichtig für Produktionssysteme, schliessen des Port 9200
- vi /etc/elasticsearch/elasticsearch.yml

```
network.bind_host: 127.0.0.1
network.publish_host: 127.0.0.1
network.host: 127.0.0.1
http.port: 9200
```

Plugin für Datenbankverbindungen installieren

- ./bin/plugin --install jdbc --url <http://xbib.org/repository/org/xbib/elasticsearch/plugin/elasticsearch-river-jdbc/1.5.0.5/elasticsearch-river-jdbc-1.5.0.5-plugin.zip>
- check, ob die richtige ojdbc6.jar im Verzeichnis /lib der Elasticsearchinstallation liegt.
- für MySQL den aktuellen mysql-connector in das /lib Verzeichnis kopieren

Elasticsearch starten → service elasticsearch start

Mein erster Elasticsearch-Index

Welche Daten will ich im Index haben?

Mein Tipp: bauen Sie sich eine View in der DB und erzeugen Sie den Index auf die View!

```
curl -XPUT 'localhost:9200/_river/MyIndex/_meta' -d '{
  "type" : "jdbc",
  "jdbc" : {
    "index" : "MyIndex",
    "url"   : "jdbc:oracle:thin:@//localhost:1521/MySID",
    "user"  : "USER",
    "password" : "PASSWORD",
    "sql"   : "select * from V_My_View"
  }
}'; echo
```

Was mache ich jetzt damit? → Demo

- Wie sieht das JSON aus?
- Sortierung der Daten für verschiedene Anwendungen
- Daten nicht analysieren
- eigene Analyzer → z.B. ein Komma
- Facetten

Daten nicht analysieren

```
"mappings" : {  
  "jdbc" : {  
    "_source" : { "enabled" : true },  
    "properties" : {  
      "AUDIENCE" : { "type" : "string", "index" : "not_analyzed" },  
      "AUDIENCE_EN" : { "type" : "string", "index" : "not_analyzed" },  
      .....  
    }  
  }  
}
```

Eigener Analyzer

```
curl -XPUT 'localhost:9200/jland' -d '{  
  "settings": {  
    "analysis": {  
      "analyzer": {  
        "komma": {  
          "type": "pattern",  
          "pattern": ",",  
        },  
        .....  
      }  
    }  
  }  
}
```

Mapping mit eigenem Analyzer

```
"mappings" : {  
  "jdbc" : {  
    "_source" : { "enabled" : true },  
    "properties" : {  
      "AUDIENCE" : { "type" : "string", "index" : "not_analyzed" },  
      "BEGINN" : { "type" : "string", "index" : "not_analyzed" },  
  
      .....  
  
      "KEYWORDS" : { "type" : "string", "analyzer" : "komma" }  
    }  
  }  
}
```

Interessante Projekte & Literatur

→ facetview2.js – ein komplettes JavaScript Frontend für Elasticsearch

<https://github.com/CottageLabs/facetview2>

→ Monitoring OBIEE with Elasticsearch, Logstash, and Kibana →

<http://www.rittmanmead.com/2014/10/monitoring-obiee-with-elasticsearch-logstash-and-kibana/>

<https://paulluaq.com/oracle-traffic-director-access-log-analysis-using-elasticsearch-logstash-and-kibana-2/>

→ <http://www.heise.de/developer/artikel/Volltextsuche-mit-ElasticSearch-1920454.html>

→ <http://frommyworkshop.blogspot.de/2015/08/ingest-data-from-oracle-database-to.html>

Blogbeitrag zur Synchronisation Oracle ↔ Elasticsearch

Elasticsearch und die Oracle Datenbank

Die komfortable Suchmaschine in der Praxis