

Fulltext Search in Nextcloud

Volltextsuche für Nextcloud

Neue Vorgehensweise

Machen wir uns also zunächst an die Installation von Elasticsearch. Als erstes bringen wir das System wie immer auf den neusten Stand:

```
apt update && apt upgrade -V
```

Als nächstes müssen die Paketquellen für Elasticsearch auf dem System hinzugefügt werden, da wir Elasticsearch aus den Paketquellen des Herstellers installieren wollen. Dazu holen wir uns erst einmal den Repository-Key auf das System:

```
wget -O - https://artifacts.elastic.co/GPG-KEY-elasticsearch | gpg --dearmor -o /etc/apt/keyrings/elasticsearch-keyring.gpg
```

Dann werden die Paketquellen selbst hinzugefügt:

```
echo "deb [signed-by=/etc/apt/keyrings/elasticsearch-keyring.gpg] https://artifacts.elastic.co/packages/8.x/apt stable main" > /etc/apt/sources.list.d/elasticsearch.list
```

Nun kann Elasticsearch auch schon installiert werden:

```
apt update && apt install elasticsearch
```

Während der Installation werden euch ein paar Infos angezeigt:

```
----- Security autoconfiguration information -----
```

```
Authentication and authorization are enabled.
```

```
TLS for the transport and HTTP layers is enabled and configured.
```

```
The generated password for the elastic built-in superuser is : <password>
```

```
...
```

Das hier angezeigte Passwort könnt ihr euch mal an einem sicheren Ort notieren (z.B. in einem Passwort-Safe). Der dazugehörige Benutzer trägt dabei den Namen „elastic“. Seit Elasticsearch 8 ist die erweiterte Sicherheit standardmäßig aktiviert. Um daher mit Elasticsearch „sprechen“ zu können, werden diese Anmeldedaten benötigt.

Da wir Elasticsearch später so konfigurieren werden, dass der Zugriff nur lokal auf dem Server stattfinden kann, werden diese Sicherheitsfeatures nicht benötigt. Trotzdem würde ich empfehlen, dass ihr euch das hier angezeigte Passwort irgendwo notiert, falls ihr dieses später doch mal benötigen solltet.

I have managed to solved it on my own.

So exit code 137 means that Elasticsearch is taking up too much memory in my server.

My solution was to locate jvm.options in /etc/elasticsearch.

I have copied a **jvm.options** into **jvm.options.d** folder and uncommented a line to set this inside.

-Xms1g
-Xmx1g

```
#####  
## IMPORTANT: JVM heap size  
#####  
##  
## The heap size is automatically configured by Elasticsearch  
## based on the available memory in your system and the roles  
## each node is configured to fulfill. If specifying heap is  
## required, it should be done through a file in jvm.options.d,  
## which should be named with .options suffix, and the min and  
## max should be set to the same value. For example, to set the  
## heap to 4 GB, create a new file in the jvm.options.d  
## directory containing these lines:  
##  
-Xms1g  
-Xmx1g  
##  
## See https://www.elastic.co/guide/en/elasticsearch/reference/8.14/heap-size.html  
## for more information  
##
```

Hope it helps!

Ebenso sagt euch die Installation, dass Elasticsearch noch nicht gestartet wurde. Freundlicherweise werden aber im gleichen Atemzug die Befehle aufgeführt, mit denen ihr dies bewerkstelligen

könnt:

```
sudo systemctl daemon-reload  
sudo systemctl enable elasticsearch.service  
sudo systemctl start elasticsearch.service
```

Vorherige Vorgehensweise

“ Da die Signierung der vorherigen Vorgehensweise veraltet ist.

Ubuntu

```
apt install apt-transport-https sudo wget curl gnupg
```

```
echo "deb [signed-by=/etc/apt/trusted.gpg.d/elasticsearch.gpg] https://artifacts.elastic.co/packages/7.x/apt  
stable main" | \  
tee -a /etc/apt/sources.list.d/elastic-7.x.list > /dev/null
```

```
curl -fsSL https://artifacts.elastic.co/GPG-KEY-elasticsearch | \  
gpg --dearmor | tee /etc/apt/trusted.gpg.d/elasticsearch.gpg > /dev/null
```

```
apt update
```

```
apt install elasticsearch
```

```
/usr/share/elasticsearch/bin/elasticsearch-plugin install ingest-attachment
```

Nach dem installieren von Elasticsearch and die Plugins den Start aktivieren

```
systemctl start elasticsearch
```

```
systemctl enable elasticsearch
```

Weiter geht es, um auch PDF's durchsuchen zu können

Vorherige Vorgehensweise



Nextcloud mit Elasticsearch/Tesseract

Wir beginnen mit der Aktualisierung des Servers selbst.

```
sudo -s
```

```
apt update && apt upgrade
```



Im Anschluss daran werden die Softwareanforderungen an die Volltextsuche sichergestellt und erfüllt.

```
apt install apt-transport-https ca-certificates  
apt install openjdk-8-jre
```



Jetzt laden wir den elasticsearch-Key des Software-Repositories herunter und publizieren diesen im System

```
wget -qO - https://artifacts.elastic.co/GPG-KEY-elasticsearch | sudo apt-key add -
```

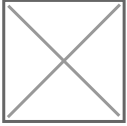
um das hinzuzufügende Repository nutzen zu können:

```
echo "deb https://artifacts.elastic.co/packages/7.x/apt stable main" | sudo tee -a  
/etc/apt/sources.list.d/elasticsearch7.list
```



Elasticsearch kann nun bereits installiert werden:

```
apt update && apt install elasticsearch -y
```



Die Installation ist abgeschlossen – konfigurieren wir Elasticsearch noch für den automatischen Start:

```
systemctl daemon-reload  
systemctl enable elasticsearch  
systemctl start elasticsearch
```



Um auch PDF Dokumente inhaltlich durchsuchen zu können installieren wir noch ein Plugin für Elasticsearch:

```
/usr/share/elasticsearch/bin/elasticsearch-plugin install ingest-attachment
```



Abschließend editieren wir noch die Konfiguration hinsichtlich des network.host:

```
nano /etc/elasticsearch/elasticsearch.yml
```

Tragen Sie Folgendes ein:

```
network.host: 127.0.0.1
```



Nach einem Neustart von Elasticsearch ist die Volltextsuche für Nextcloud bereits vorbereitet und einsatzfähig.

```
service elasticsearch restart
```

Um Text aus Bildern lesbar und somit auch durchsuchbar zu machen benötigen wir noch Tesseract (OCR).

```
apt-get install tesseract-ocr tesseract-ocr-deu tesseract-ocr-eng
```



Nach einem Neustart des elasticsearch-Services

```
service elasticsearch restart
```

können wir mit der Einrichtung, also der Aktivierung und Konfiguration der Apps in Nextcloud fortfahren.



Aktivieren Sie mindestens diese Apps:

- Full text search – Diese App bietet die Grundfunktionen zur Volltextsuche
- Full text search – Elasticsearch Platform: Dies stellt die Verbindung zur Suchmaschine her.
- Full text search – Files: Erweitert die Dateien-App um die Volltextsuche.
- Full text search – Files – Tesseract OCR: Verbindung zu Tesseract.



Bevor die Volltextsuche in Nextcloud genutzt werden kann muss diese noch konfiguriert werden. Dies geschieht in den Nextcloud-Administratoreinstellungen unter Volltextsuche:





Die Werte werden automatisch gespeichert! Nun erzeugen wir noch initial den Volltextindex über das Nextcloud Kommandozeilentool occ:

```
sudo -u www-data php /var/www/nextcloud/occ fulltextsearch:index
```



“ **Achtung:**

Je nach Größe des Datenbestands kann dieser Befehl Minuten bis Stunden dauern.

Nach dem initialen Anlegen des Volltextindexes muss der zuvor ausgeführte Befehl **nicht** mehr regelmäßig ausgeführt werden. Die Aktualisierung des Indexes erfolgt dann im Rahmen des Nextcloud-Cronjobs.

Nach der Indizierung finden Sie über die Suche Dokumente sowohl über die Dateinamen, als auch über Inhalte.



Sofern Updates eingespielt werden muss das Plugin aktualisiert werden. Ein Blick in das Logfile lohnt sich und gibt entsprechende Auskunft:

```
nano /var/log/elasticsearch/elasticsearch.log
```

“ „... org.elasticsearch.bootstrap.StartupException:
java.lang.IllegalArgumentException: Plugin [ingest-attachment] was built for
Elasticsearch version...”

Update Plugin nach Update Nextcloud oder Rechner

Wir entfernen zuerst das bestehende Plugin,

```
/usr/share/elasticsearch/bin/elasticsearch-plugin remove ingest-attachment
```

um es anschließend in der neuen Version zu installieren:

```
/usr/share/elasticsearch/bin/elasticsearch-plugin install ingest-attachment
```

Nach einem Neustart des Dienstes wird die Volltextsuche wieder funktionieren.

Die Installation und Konfiguration der Volltextsuche für Ihren Nextcloudserver wurde erfolgreich abgeschlossen und so wünsche ich Ihnen viel Spaß mit Ihren Daten in Ihrer privaten Cloud. Über Ihre Unterstützung (diese wird ordnungsgemäß versteuert!) würden sich meine Frau, meine Zwillinge und ich sehr freuen!

Revision #9

Created 21 March 2023 10:10:37 by Hermann

Updated 13 January 2025 13:14:58 by Hermann