Advanced Search White Paper with Elasticsearch 7.7.1

(06/07/2021)

Contents

Overview

Elasticsearch is a tool for indexing data, particularly text data, for high performance searching. LawMaster integrates with an Elasticsearch service to store a *copy* of folio data for searching. The LawMaster-Elasticsearch integration includes Apache Tika as a tool to extract text data from documents.

Searching

If Elasticsearch is enabled, the "File / Word in Text" folio search type uses the Elasticsearch index, regardless of whether "keywords" are specified or not. The search criteria behaves as before, but the "keywords" option now has different syntax considerations. Input into the "keywords" field must comply with the Elasticsearch "Query String" syntax, the detail of which is available at http://www.elastic.co/guide/en/elasticsearch/reference/7.7/query-dsl-query-string-query.html#query-string-syntax.

Some key points for consideration include:

- Operators (AND, OR, and NOT) must be uppercase.
- Parentheses for advanced Boolean operator use is permitted. For example, (big AND cat) OR tiger OR lion.
- Wildcards and regular expressions are supported.
- Fuzzy searching is supported. For example, "quikc~" matches the word "quick".
- There are reserved characters that are *not* searched for unless escaped with a leading backslash. Reserved characters are:
  ```
  + - = && || > < ! ( ) { } [ ] ^ " ~ * ? : \ /
  ```
  For example, to search for (1+1)=2, write the query as \(1\+1\)=2.

Topology
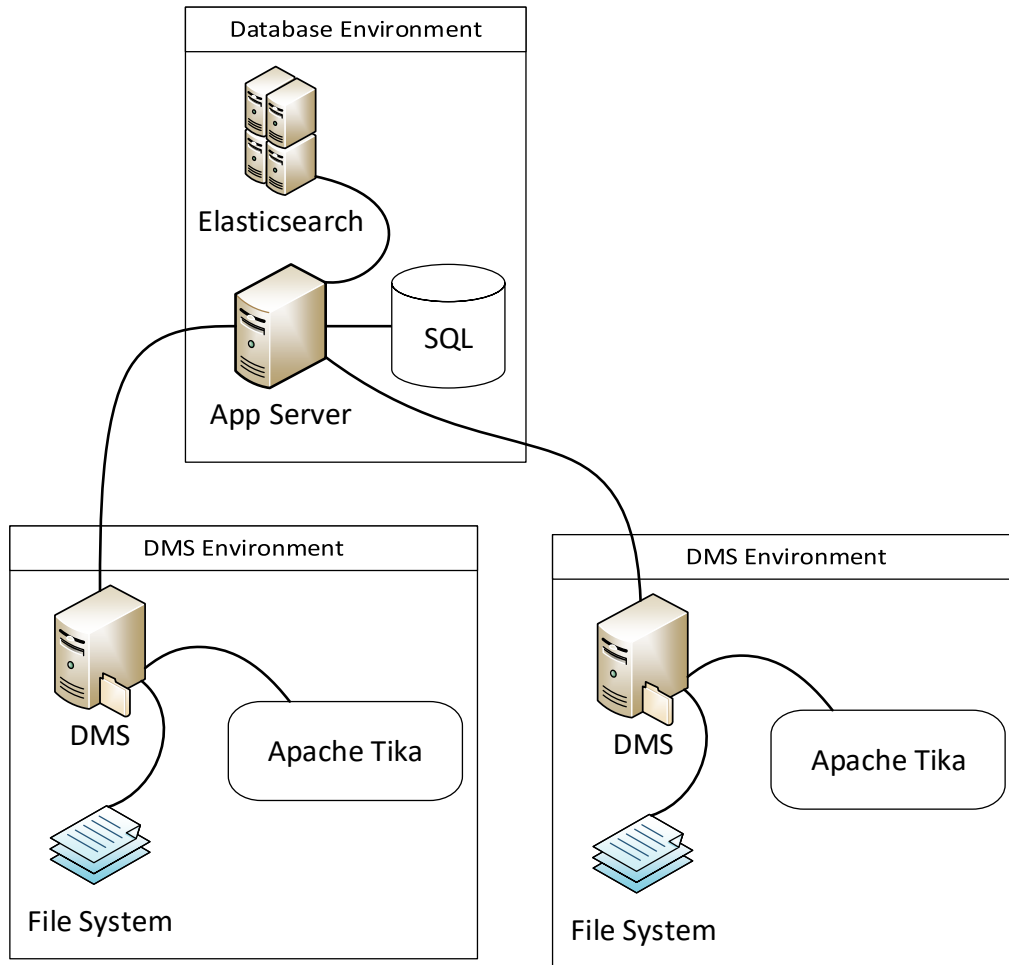
This diagram above depicts a "scaled out" topology, where each Document Management Server (DMS) runs on its own server, and Elasticsearch runs on a cluster of servers apart from the LawMaster application service / app server. In a smaller scale implementation, it's reasonable for all services to occupy the same server.

Elasticsearch *may* be serviced by a cluster of servers. That is, multiple servers, each hosting a single node of a single Elasticsearch cluster. A single index can be spread across the nodes in the cluster to optimise performance.

Inner Workings

There are two distinct elements to the Elasticsearch integration:

- Indexing – The process of copying folio data from the primary data stores (the SQL database and the file system for attachments) into the Elasticsearch index.
- Searching – Querying the Elasticsearch index.

Indexing

The LawMaster Application Service runs a background job for indexing. Indexing is the process of copying folio data into the Elasticsearch index. You can see the status of the indexing job in Parameters > Records Management > Indexing.

The indexing job fetches folios from oldest to newest (according to a dedicated timestamp value) and sends their content to Elasticsearch. If a folio has an attachment on a DMS with full text searching enabled, LawMaster sends a request to the corresponding DMS to extract the text content of the folio attachment. The DMS responds by using the Apache Tika service to extract text from the document and returns it to the indexing job for sending to Elasticsearch.

Folio timestamps are stored in the Matter_Folios_RowVer SQL table. In this table is the mforv_row_ver column. Whenever a folio (Matter_Folio table) or data related to its attachment (Folios_Storage table) is modified, the mforv_row_ver value is updated such that it will be the *first* row if sorted in descending order. When sending data to Elasticsearch, this timestamp is also stored in Elasticsearch as a way of identifying which folios are indexed (all those with a lesser timestamp value), and those that are *not* indexed (those with a greater timestamp value).

Eventual Consistency

Folio data is copied from the primary storage (the SQL database for data, and the file system for documents) to a secondary storage (Elasticsearch) using an *eventual consistency* pattern. This means that the data in Elasticsearch is not necessarily the same as in the primary storage, but given enough time, *eventually* will be.

At an extreme case, there might be *no* data in Elasticsearch. This occurs when the LawMaster database is first brought online (for example, from migrated data). Routine folio maintenance (adds, modifies, and deletes) is never immediately represented in Elasticsearch, so there is a latency between the data maintenance, and that maintenance being reflected in the search results.

This contrasts with a *transactional consistency* pattern, where a folio maintenance task is not acknowledged in the primary storage until it is also updated in the secondary storage. Such an approach could have a significant impact on data maintenance performance, especially in distributed systems. Hence, eventual consistency is chosen in favour of transactional consistency, as a way of prioritising primary data maintenance performance.

Searching

An Elasticsearch query goes through the following steps:

1. The user sends a query from the client application to the Application Service.
2. The Application Service combines query content with the user's security profile.
3. The Application Service sends the query to Elasticsearch.
4. The Application Service receives a list of folio numbers from the Elasticsearch search result, and uses these numbers to retrieve the full set of folio data from the SQL database.
   Because the Elasticsearch index might not be consistent with the database, it's possible the search result from Elasticsearch includes folios that no longer exist in the database. If this is the case, the Application Service might request *more* results from Elasticsearch to ensure the "max records" specified by the user is fulfilled.
5. The search result is returned to the client application.

System Requirements

Elasticsearch Service

- It's recommended that this is the same server as running the LawMaster Application Service, or if it's another server, one with a strong network link to the LawMaster Application Service.
- Elasticsearch 7.7.1.
- RAM / System Memory. Elasticsearch is a memory-intensive application. Minimum 2GB of RAM should be allocated.
- Disk space. Testing indicates that the Elasticsearch index is 4% of the size of the files in the DMS that it indexes. For example, a 25 GB DMS results in an approximately 1 GB Elasticsearch index.

Apache Tika (the text extraction service)

- Is required on each server hosting a DMS configured for full text indexing.
- Apache Tika Server 1.24.1.
- Java OpenJDK8U.
- Apache Commons Daemon 1.2.2.

Network

- The server running the LawMaster Application Service needs to communicate with all DMS instances configured for full text indexing.
- The link between the LawMaster Application Service and Elasticsearch (or the primary node in a cluster) is ideally a strong link, but this isn't critical.
- Links between multiple nodes in an Elasticsearch cluster *must* be strong. Sub-standard links could result in outages and errors. As a minimum, a 100 Mbps LAN link is recommended.

IIS Application Service

- If you are hosting the LawMaster Application Service in Internet Information Services (IIS), the "Application Initialization" module is necessary to ensure the background indexing job runs uninterrupted.

Download Links

Complete Set

A single download containing all necessary components is available from
http://www.lawmaster.com.au/memberswp/wp-content/downloads/other/LM-ElasticSearch-7.7.1.exe.

The download includes:

- Elasticsearch 7.7.1
- Java OpenJDK8U-jdk_x64_windows_hotspot_8u252b09
- Tika 1.24.1
- Apache Commons Daemon 1.2.2.

Individual Downloads

- Elasticsearch 7.7.1: https://artifacts.elastic.co/downloads/elasticsearch/elasticsearch-7.7.1.msi
- Java AdoptOpenJDK8: https://github.com/AdoptOpenJDK/openjdk8-binaries/releases/download/jdk8u252-b09.1/OpenJDK8U-jdk_x64_windows_hotspot_8u252b09.msi
- Apache Tika Server 1.24.1: https://www.apache.org/dyn/closer.cgi/tika/tika-server-1.24.1-bin.zip
- Apache Commons Daemon (for Windows) 1.2.2: http://apache.mirror.amaze.com.au/commons/daemon/binaries/windows/commons-daemon-1.2.2-bin-windows.zip

    Note: Unzip only the necessary files from Apache Commons Daemon 1.2.2 to the same folder as the Tika server, as follows:
    - prunmgr.exe – **Rename** to tika-service-mgr.exe
    - prunsrv.exe – **Rename** to tika-service-x86.exe
    - amd64\prunsrv.exe – **Rename** to tika-service-x64.exe

Installation

Elasticsearch

Elasticsearch is usually installed on the same server as the LawMaster Application Service. This is because there is:

- No need to isolate it.
- Heavy network communication between the Application Service and Elasticsearch. Keeping both services on the same server optimises this communication, and simplifies security by configuring Elasticsearch to accept local connections only.

Important notes:

- Elasticsearch 1.4.4 and 7.7.1 **cannot** be run on the same server.
- Elasticsearch 1.4.4 **does not** work with Portugal and future releases of LawMaster.
- Elasticsearch 1.4.4 **does work** with Tika 1.24.1.
- Elasticsearch 7.7.1 **does not** work with releases prior to Portugal.

There are some scenarios where installing Elasticsearch on a separate server might be necessary. For more information, see [Scaling Out](#).

Installation Instructions for Elasticsearch 7.7.1

1. Unzip LM-ElasticSearch-7.7.1.exe to C:\.
2. From Computer > Properties > Advanced System Settings > Environment Variables > System Variables, remove:
    - The JAVA_HOME environment variable to the Java path. For example, C:\Program Files\Java\jdk1.8.0_45
    - C:\ProgramData\Oracle\Java\javapath from the PATH variable.

3. Decide on a maximum memory heap size for Elasticsearch.
   The recommendation is 50% of system memory if the system is *only* running Elasticsearch. Exceeding 50% is not recommended because Elasticsearch relies heavily on system cache. That is, some memory needs to be left for the operating system. In an environment where a LawMaster Application Service is also running on the server, we recommend a limit of 25% of system memory. A value higher than 32 GB should be avoided. For more information, see the Elasticsearch guide on heap sizing, at
   https://www.elastic.co/guide/en/elasticsearch/guide/master/heap-sizing.html.
4. Install Elasticsearch using C:\ES-7.7.1-Tika-1.24.1\ElasticSearch-7.7.1\elasticsearch-7.7.1.msi.
   - Use the default directories, but if you prefer your Elasticsearch data to be on another drive, change it here.
     - DO NOT change the config location.
   - Install as a service:
     - Use Local System account
     - Start the service after this installation is complete
     - Start the service when Windows Starts (Automatic)
   - Change ClusterName to LM-elasticsearch
     - Change Node name to <Server Name>
     - Change memory as required. (for example, 500 MB for a small number of users)
     - Leave HTTP port as 9200
     - Leave Transport port as 9300
   - Select Basic License
5. Validate Elasticsearch is running. Using a web browser, navigate to http://localhost:9200, which initiates a file download. Open the downloaded file in Notepad.
   For example:

```
{
  "name" : "SUPSQL2019",
  "cluster_name" : "LM-elasticsearch",
  "cluster_uuid" : "ZlaKGgVGTKmfi_eAmOs8PQ",
  "version" : {
    "number" : "7.7.1",
    "build_flavor" : "unknown",
    "build_type" : "unknown",
    "build_hash" : "81a1e9eda8e6183f5237786246f6dced26a10eaf",
    "build_date" : "2020-05-12T02:01:37.602180Z",
    "build_snapshot" : false,
    "lucene_version" : "8.5.1",
    "minimum_wire_compatibility_version" : "6.8.0",
    "minimum_index_compatibility_version" : "6.0.0-beta1"
  },
  "tagline" : "You Know, for Search"
}
```

Verify the heap setting by navigating to http://localhost:9200/_nodes/stats and look for the "heap_max_in_bytes" value.

Uninstall Instructions for Elasticsearch 1.4.4

- Run as Administrator in the Command Prompt window (CMD):

```
Cd C:\Elasticsearch\elasticsearch-1.4.4\bin\

service.bat remove
```

Apache Tika (On DMS Servers)

Note: Elasticsearch 1.4.4 and 7.7.1 **does work** with Tika 1.24.1

Uninstall Instructions for Tika 1.7 or 1.9.

- Run as Administrator in the Command Prompt window (CMD):

```
Cd C:\Elasticsearch\TikaService

tika-service-x64.exe //SS//tika-service

tika-service-x64.exe //DS//tika-service
```

Installation Instructions for Tika 1.24.1

Note: Earlier versions of Tika **must** be removed before installing version 1.24.1.

1. Unzip LM-ElasticSearch-7.7.1.exe to C:\.
2. Remove from "Computer-Properties-Advanced System Settings-Environment Variables-System variables":
   - The JAVA_HOME environment variable to the Java path. For example, C:\Program Files\Java\jdk1.8.0_45
   - C:\ProgramData\Oracle\Java\javapath from the PATH variable.
3. Install the JDK using C:\ES-7.7.1-Tika-1.24.1\Tika-1.24.1\OpenJDK8U-jdk_x64_windows_hotspot_8u252b09.msi.
   - Unselect **all** features.
4. Run as Administrator in the Command Prompt window (CMD):

```
Cd C:\ES-7.7.1-Tika-1.24.1\Tika-1.24.1\

tika-service-x64.exe //IS//tika-service --DisplayName "Tika 1.24.1
Service" --Description "Tika 1.24.1 Windows Service for text
extraction" --StartMode exe --StartImage "C:\Program
Files\AdoptOpenJDK\jdk-8.0.252.09-hotspot\bin\javaw.exe" --StartPath
"C:\Program Files\AdoptOpenJDK\jdk-8.0.252.09-hotspot\bin"
++StartParams -jar;"C:\ES-7.7.1-Tika-1.24.1\Tika-1.24.1\tika-server-
1.24.1.jar" --Startup auto
```

Note: An uninstall can be performed as follows:

```
tika-service-x64.exe //SS//tika-service

tika-service-x64.exe //DS//tika-service

net start "Tika 1.24.1 Service"
```

5. Validate Tika is running. Using a web browser, navigate to http://localhost:9998.

Upgrade from 1.4.4 (Pre-Portugal to Portugal or later)

Elasticsearch Migration from 1.4.4 to 7.7.1

Elasticsearch 1.4.4 needs to be running on one machine while Elasticsearch 7.7.1 is running on another so the index can be imported into Elasticsearch 7.7.1.

Therefore, the upgrade process for each client site is:

1. Uninstall Tika 1.7 or 1.9. See Uninstall instructions for Tika 1.7 or 1.9.
2. Install Tika 1.24.1. See Installation Instructions for Tika 1.24.1.
3. Setup the new Elasticsearch 7.7.1 machine. See Installation Instructions for Elasticsearch 7.7.1.
4. Setup the Elasticsearch machines for migration. See Setup Elasticsearch Machines for Migration.
5. Installation Instructions for Elasticsearch 7.7.1
6. Login to the non-Production databases and take note of the Folios in Index.
    a. Login into the database.
    b. Select Parameters > Set Parameters > Records Management > Indexing and take note of the value in "Folios in Index".
    c. Logout of the database.
7. Migrate Elasticsearch Indexes for these DB's to version 7.7.1. See Migrate Index Process.
8. Upgrade non-Production databases to Portugal (or later) version using the usual upgrade process.
9. Edit Elasticsearch Address setting in these databases. See Change the Setting in the LawMaster Database.
10. Add a Folio Attachment and confirm it is now searchable.
11. Repeat steps 4 to 9 for each database and for Production when ready.
12. After all indexes are migrated:
    a. Elasticsearch 1.4.4 can be uninstalled on the "old" Elasticsearch machine. See Uninstall Instructions for Elasticsearch 1.4.4.
    b. Elasticsearch 7.7.1 can be installed on the "old" Elasticsearch machine. See Installation Instructions for Elasticsearch 7.7.1.
    c. Stop Elasticsearch 7.7.1 on the "new" Elasticsearch machine, using Windows Services.
    d. Stop Elasticsearch 7.7.1 on the "old" Elasticsearch machine, using Windows Services.
    e. Copy indexes from the "new" Elasticsearch machine to the "old" Elasticsearch machine.
       That is, copy the indexes from
       \\<<new machine>>\c$\ProgramData\Elastic\Elasticsearch\data\nodes\0\indices
       to
       \\<<old machine>>\c$\ProgramData\Elastic\Elasticsearch\data\nodes\0\indices
    f. Start Elasticsearch 7.7.1 on the "old" Elasticsearch machine, using Windows Services.
    g. For each database, update the Elasticsearch address back to its original value. That is, http://localhost:9200. See Change the Setting in the LawMaster Database.
    h. Reboot the "old" Elasticsearch machine to restart all Application Services.
    i. For each database:
        i. Log into the database.
        ii. Add a Folio Attachment to each database and confirm they are searchable.
        iii. Select Parameters > Set Parameters > Records Management > Indexing and confirm the value in "Folios in Index" is what you expect.
    j. Turn off the "new" Elasticsearch machine.

The upgrade process for this site is finished.

Setup Elasticsearch Machines for Migration

Setup the Elasticsearch machines so they can communicate.

1. On the Elasticsearch 7.7.1 machine:
    a. Edit C:\ProgramData\Elastic\Elasticsearch\config\elasticsearch.yml on the new server to add the following line:
    `reindex.remote.whitelist: oldESservername:9200`

    Note: `oldESservername` is the local name of the machine with Elasticsearch 1.4.4 installed on it.
    b. Restart the Elasticsearch (7.7.1) service.
    c. If this machine is **not** the LawMaster Application Service Machine, open the incoming firewall port 9200 on the Elasticsearch 7.7.1 machine so the LawMaster Application Service machine can communicate with it.
2. Open the firewall on port 9200 on the Elasticsearch 1.4.4 machine:
    a. Comment out following line in C:\Elasticsearch\elasticsearch-1.4.4\config\elasticsearch.yml on the old server:
    `network.host: 127.0.0.1`
    b. Open the incoming firewall port 9200 on Elasticsearch 1.4.4 machine so the Elasticsearch 7.7.1 machine can communicate with it.
    c. Restart the Elasticsearch 1.4.4 service.
3. On the Elasticsearch 7.7.1 server, run the following powershell code to confirm access to the old and new versions of Elasticsearch:

```
$ipsOLD = "oldESservername"
$ips = "Localhost"

#GET OLD Server DB list
$Uri = "http://" + $ipsOLD + ":9200/_all"
$r = Invoke-WebRequest -Uri $Uri -Method Get
    $allIndexes = (ConvertFrom-Json -InputObject $r)
    $allIndexes

#GET New Server DB list
$Uri = "http://" + $ips + ":9200/_all"
$r = Invoke-WebRequest -Uri $Uri -Method Get -ContentType
'application/json'
    $allIndexes = (ConvertFrom-Json -InputObject $r)
    $allIndexes
```

Migrate Index Process

1. Before migrating an Elasticsearch index you should check the status of the ES Index.
    a. Log in to the database.
    b. Select Security > Show User Logins.
        i. Right Click, and Select Prevent Login
        ii. Right Click, and Select Disable All
    c. Select Parameters > Set Parameters > Records Management > Indexing.
        i. If "Folios in Index" = 0, then you don't currently have an Elasticsearch index setup for this database. If you want to test the Elasticsearch Index migration process on this database, then please get in touch with the LawMaster HelpDesk so we can discuss the best way for you to test the process.
        ii. If "Folios to Index" > 0, then you will need to wait until this number has gone down to zero to ensure all folios are indexed at the time of the Elasticsearch Index migration.

iii. "Status" should show "Idle". If it does not, you will need to wait until it does. If you are uncertain of the meaning of a different "Status" value, then please get in touch with the LawMaster HelpDesk for advice.

iv. Take note of the number in "Folios in Index".

2. Run the following code on the same machine (to create the empty index). Change $dbnameold according to the database associated with the Elasticsearch index you are migrating. That is, lawmasterproduction, lawmastertraining, and so on.

```
$dbnameold = "lawmastertraining"
$dbnamenew = $dbnameold

#Create Empty Index
$json = '{
        "mappings": {
            "properties": {
                "author": {
                    "type": "text",
                    "fields": {
                        "keyword": {
                            "type": "keyword",
                            "ignore_above": 256
                        }
                    }
                },
                "dateOfWriting": {
                    "type": "date"
                },
                "docTemplate": {
                    "type": "text",
                    "fields": {
                        "keyword": {
                            "type": "keyword",
                            "ignore_above": 256
                        }
                    }
                },
                "documentText": {
                    "type": "text",
                    "fields": {
                        "keyword": {
                            "type": "keyword",
                            "ignore_above": 256
                        }
                    }
                },
                "fileInd": {
                    "type": "text",
                    "fields": {
                        "keyword": {
                            "type": "keyword",
                            "ignore_above": 256
                        }
                    }
                },
                "fileNo": {
                    "type": "long"
                },
                "id": {
                    "type": "long"
                },
                "overrideFileName": {
                    "type": "text",
                    "fields": {
                        "keyword": {
                            "type": "keyword",
                            "ignore_above": 256
```

```
                                }
                            }
                        },
                        "precis": {
                            "type": "text",
                            "fields": {
                                "keyword": {
                                    "type": "keyword",
                                    "ignore_above": 256
                                }
                            }
                        },
                        "registerType": {
                            "type": "text",
                            "fields": {
                                "keyword": {
                                    "type": "keyword",
                                    "ignore_above": 256
                                }
                            }
                        },
                        "rowVersion": {
                            "type": "text",
                            "fields": {
                                "keyword": {
                                    "type": "keyword",
                                    "ignore_above": 256
                                }
                            }
                        },
                        "securityLevel": {
                            "type": "text",
                            "fields": {
                                "keyword": {
                                    "type": "keyword",
                                    "ignore_above": 256
                                }
                            }
                        },
                        "toFrom": {
                            "type": "text",
                            "fields": {
                                "keyword": {
                                    "type": "keyword",
                                    "ignore_above": 256
                                }
                            }
                        }
                    }
                }
            }
        }'
$Uri = "http://" + $ips + ":9200/" + $dbnamenew + ""
$r = Invoke-WebRequest -Uri $Uri -Method PUT -ContentType
'application/json' -Body $json
$conv = (ConvertFrom-Json -InputObject $r)
$conv
```

3. After completion, run the following to migrate the data from the old to the new index:

```
#Transfer data from OLD ES version to New ES version
$json = '{
   "source": {
      "remote": {
         "host": "http://' + $ipsOLD + ':9200"
      },
      "index": "' + $dbnameold + '",
      "size": 1
```

```
        },
        "dest": {
            "index": "' + $dbnamenew + '"
        }
    }'
    #Wait for completion
    $Uri = "http://" + $ips + ":9200/_reindex?wait_for_completion=false"
    $returnedJson = Invoke-RestMethod -Uri  $Uri -Method POST  -
    ContentType 'application/json' -Body $json
    "Started Time: " + [string](date)
    Start-Sleep -Seconds 60
    $Uri = "http://" + $ips + ":9200/_tasks/" + $returnedJson.task
    $r = Invoke-WebRequest -Uri $Uri -Method Get
    $conv = (ConvertFrom-Json -InputObject $r)
    $conv.task.status
    do
    {
        #Display Progress
        Start-Sleep -Seconds 60
        $Uri = "http://" + $ips + ":9200/_tasks/" + $returnedJson.task
        $r = Invoke-WebRequest -Uri $Uri -Method Get
        $conv = (ConvertFrom-Json -InputObject $r)
        [string]($conv.task.status.created/$conv.task.status.total*100)
    +        " % completed"
    } until ($conv.completed)
    $conv.task.status
    "Completed: " + $conv.completed
    "Failure Count : " + $conv.response.failures.Count
    "Finished Time: " + [string](date)
```

Change the Setting in the LawMaster Database

You now need to change your Elasticsearch setting in LawMaster.

Note: The database **must** be running **10.14 (Portugal) or later**.

1. Log in to the database.
2. Select Parameters > Set Parameters > Records Management > Indexing, and then select Edit.
   a. Select both the "Index Folios in Elasticsearch" and "Search Folios using Elasticsearch" options.
   b. In "Elasticsearch Address" change http://localhost:9200 to http://<<Elasticsearch7.7.1 machine name>>:9200
   c. Click Save.
   d. Click Restart for the Application Service to locate and connect to the new Elasticsearch service.
   e. Wait for 10 seconds.
   f. Click on the Refresh button to confirm the value in the "Folios in Index" is what you expect. (Pre-Migration value minus one, or greater.)
3. Select Security > Show User Logins.
   a. Right Click, and Select Allow Login

Configuration Within LawMaster

DS – "Document Management Servers" Parameter

As before, the 'DS' parameter includes a "Full Text Indexing" option to specify whether the content of folio attachments is indexed.

Note**:** If folios from a DMS are already indexed *without* the "Full Text Indexing" option, selecting that option does *not* retro-actively index the folio attachment content. If these folios are to be re-indexed, actions described in the "re-indexing" topic need to be undertaken *after* enabling the "Full Text Indexing" option.

Indexing

There is a parameter screen in Records Management > Indexing, which indicates Elasticsearch indexing status in the "Status" panel, settings related to indexing in the "Configuration" panel, and buttons to interact with the indexing job.

Configuration

Use Elasticsearch for Folio Indexing:
By default, this option is disabled, allowing for configuration to be refined *before* enabling the Elasticsearch indexing job. If transitioning from disabled to enabled, it is necessary to restart the Application Service, or click the "Restart" button to start the indexing job.

Elasticsearch Address:
By default, Elasticsearch runs on port 9200, on the same server as the LawMaster application service, so a default address of http://localhost:9200 is used.

Index Name:
A single Elasticsearch instance can store many indexes. Be careful to *not* point a training database to the production Elasticsearch index, or other inappropriate index. The index name must be lower case because of constraints in Elasticsearch. Note that when the database type is changed (for example, from Production to Training), the value of this field is set to blank to prevent cross-talk. For more information on cross-talk prevention, see Cross-Talk Prevention.

By default, set Index Name to be the Database Name. For example, lawmasterv5626development.

Indexing Lag:
After the indexing job indexes all folios for the database, it enters an "Idle" state. The length of time it remains in this idle state before checking for more folios to index, is determined by this parameter. That is, if the indexing job is idle, and a user changes a folio, this is the maximum amount of time the indexing job remains idle before starting to index any folios that need indexing.

Buttons

Buttons on the indexing screen control the indexing job. It can be paused, resumed (un-paused), stopped, and restarted. There's also a button to delete the Elasticsearch index, which might be useful as a maintenance task, or cause all data to be re-indexed.

Licensing

LawMaster Elasticsearch integration uses the "Advanced Document Management" (DMA) license module.

Cross-Talk Prevention

Cross-talk occurs when a LawMaster database, and indexing job, points to the *wrong* Elasticsearch index. For example, a database backup is copied from Production to Training, documents are changed in Training, but it is the *Production* Elasticsearch index that is updated.

To protect against cross-talk, when you change a LawMaster database type, for example, from Production to Training, LawMaster clears the "Index Name" field of the indexing configuration. When the indexing job, or a search, attempts to connect to an Elasticsearch index, it won't find *any* index to communicate with, so will fail.

Note that the index name is cached in the Application Server shortly after it starts. To reset this cache, re-save the indexing configuration in parameters.

Re-Indexing

Various circumstances might arise where it is necessary to re-index folio data, such as:

- Data consistency errors. For example, something has damaged the Elasticsearch index.
- A DMS not previously indexed has had the "Full Text Indexing" option enabled. (Document text won't be automatically indexed because the rest of the folio data is already indexed).
- Elasticsearch clustering requirements are changed. For example, more nodes are needed but the data is in too few shards to support further scaling out.

The indexing jobs dependency on folio timestamps (per the "Indexing" section of the "Inner Workings" topic) means there are two ways to mark folios for re-indexing. Modify the:

- "mfrt_row_ver" value in the "Matter_Folios_RowVer_Tracking" table. That is, decrement it.
- Timestamp value of one or more folio records. That is, increment them.

The simplest option for modifying the timestamp stored in Elasticsearch is to delete the index, leading to it being entirely re-created. This is accomplished by clicking the "Delete Index" button in Parameters > Records Management > Indexing.

Note: Large indexes could take an extremely long time to re-index.

Timestamp values of folio records are incremented by making any change (including setting a value to its existing value) to a row in the Matter_Folios or Folios_Storage table. For example, mark all folios with an attachment in the "MAIN" DMS for re-indexing using the following query:

```
UPDATE Folios_Storage SET fs_storage = fs_storage WHERE fs_server = 'MAIN'
```

Equally, any update to a folio through the LawMaster client application increments the folio timestamp, thus marking it for re-indexing.

Scaling Out

Elasticsearch is usually installed on the same server as the LawMaster Application Service, but there are two reasons why it might be installed on a separate server: isolation and scalability

Isolation

LawMaster does not recommend isolating Elasticsearch, but supports this as an option for clients that prefer it.

Scalability

In situations where Elasticsearch would share a constrained resource with the LawMaster Application Server, or other services running on the server, moving the Elasticsearch instance to another server is recommended. That is, if the server does not have enough resources to effectively host both the LawMaster Application Service and Elasticsearch, then separate them.

Understanding whether a resource is likely to be constrained depends on an understanding of the resources that Elasticsearch and the LawMaster Application Service consume.

Network Connections

Minimal. Only the Application Service is communicating with Elasticsearch.

Network Bandwidth

Solely communication between the Application Service and Elasticsearch. There's no scalability argument here for separating the Elasticsearch service from the AppService.

CPU

Elasticsearch consumes relatively little CPU resources. Statistical analysis from our environments show usage below one tenth of the LawMaster Application Service, but specific scenarios (large queries or large indexing operations) might present an increase. If there are any concerns, monitoring is recommended to get a better understanding of CPU consumption within a specific environment.

Memory

Elasticsearch is relatively memory-intensive, dependant on the size of index. Configuration includes setting a maximum memory heap size to define an upper limit on RAM consumption. An Elasticsearch instance running short of RAM will not only see a decrease in performance but might also reject queries and produce other errors. Minimum 2GB of RAM should be allocated.

Beyond 32 GB, Elasticsearch should be spread across multiple nodes in a cluster. That is, multiple servers running Elasticsearch.

Disk I/O

LawMaster recommends monitoring disk activity if this is a concern, to get an impression of usage within the specific environment.

Additional Resources

The Elasticsearch Setup Reference, available at https://www.elastic.co/guide/en/elasticsearch/reference/7.7/setup.html, provides recommended settings for configuration (heap size, node names, and so on.) as well as various other best practice "do's and don'ts".