

# Migrate from Autonomy IDOL to Elasticsearch

Switch your FDD system from using Autonomy IDOL to Elasticsearch (ES) for full text search.

## What do I tell my users?

Steps that will affect your end users are marked with a red star ★ and have an explanation to the right.

- 1. Before you Begin, Review Components
- 2. Install and Configure Elasticsearch
- 3. Configure UTR for Elasticsearch
- 4. Test Small Set of Documents with Elasticsearch
  - a. Choose Test Set
  - b. Send to UTR
  - c. Wait for UTR
  - d. Test Full Text Search
- 5. Add Documents to Elasticsearch
  - a. How Long Will UTR Take?
  - b. Process Bins and File Cabinets
  - c. Wait for UTR
  - d. Test Full Text Search
- 6. User Acceptance Testing
- 7. Hide IDOL
- Related articles

## 1. Before you Begin, Review Components

Before migrating, you need the correct versions of the Feith applications that support Elasticsearch with the Feith Authentication Plugin that uses digest authentication.

We do not recommend prior versions of Feith applications with out-moded implementations of ES.

The minimum required versions you should use for full text searching in ES:

- FDD Schema 9.2.2
- Elasticsearch 1.7.4, the Feith installer for ES which includes the Feith Authentication Plugin
- FCP 9.2.2.6
- UTR 9.2.3.3
- FDD 9.2.0.5 (2/18/2016)
- WebFDD 9.2.1.9 (still in progress; significant bugs exist; see [WebFDD Full Text](#))

## 2. Install and Configure Elasticsearch

To install and configure Elasticsearch

1. Install Elasticsearch on a server that has the necessary hardware.  
See:
  - [ElasticSearch Hardware Guidelines](#)
  - [Feith Install Guide for Elasticsearch in \\thor\projects\3rdparty\elasticsearch\doc](#)
2. Start the Elasticsearch service, if you have not already done so.
3. Add the FDD server entry for Elasticsearch in FCP's **Server s** module. ★

### ★ Tell your users:

- They can full text search in WebFDD only.
- In order to full text search, they must select the IDOL server in the new **Servers** drop-down list they will be given next to the **Full Text Search** box. The IDOL server has the port **9000** in the name, e.g. myidolserver:9000.

Why: At this point, both an IDOL server entry and ES server entry exist in your FDD system. Feith applications will be affected as follows:

- Any FDD Client instance started after this point will favor Elasticsearch and ignore IDOL, and therefore full text searches in FDD will return few or no results.
- Any WebFDD instance started after this point will present you with a drop-down list to select whether you want to full text search in IDOL or ES.

### 3. Configure UTR for Elasticsearch

To configure Feith Universal Text Recognition (UTR) Engine for Elasticsearch:

1. In UTR's initialization file (utr.ini), set the **ESUser** and **ESEncryptedPassword** settings for authentication. See the UTR Manual in \\thor\projects\docs\public\UTR.
2. Restart the UTR service. UTR finds the Elasticsearch server and will use it from now on; the IDOL server is ignored. ★

★ Tell your users: If they want to find documents recently added to full text search, they can:

- Search in FDD Client.
- Search in WebFDD and select the ES server in the **Servers** drop-down list. The server has the port **9200** in the name, e.g. myesserver:9200.

Why: From this point forward, UTR will put all new documents into Elasticsearch instead of IDOL. Therefore, documents recently added to full text search must be searched for in ES.

### 4. Test Small Set of Documents with Elasticsearch

Before adding all of your full-text-searchable documents to ES, test the configuration with a small set of documents first.

#### a. Choose Test Set

Find a few of your existing documents to test that have the following types of pages:

- A page with a text layer i.e. text you can highlight with the mouse. Examples: TXT, DOCX, XLSX, PDF.
- An image. Examples: TIF, JPG.
- Feith Forms iQ form.
- Email (EML), such as in the Mail iQ2 file cabinet.

Not all of the above page types may apply to your organization.

#### b. Send to UTR

Send the test set of documents to UTR to be added into ES. To do this you can either:

- Right-click the document in FDD Client and select **Full Text>Add**.

This FDD feature should be used for testing and troubleshooting, such as this case. It should not be used as the standard way to add documents to Elasticsearch; the best way to get documents into ES is to use the **Full-Text Index** option set at the file cabinet level in FCP.

or

- Insert a command into the **fdd.actions** table for each document. UTR will process the command. You will need to get the document's internal doc\_id and use the SQL:

```
insert into fdd.actions
(doc_id, action) values
(456, 'UTR')
```

### c. Wait for UTR

Wait for UTR to process the test set into ES. You can speed up the process by opening the UTR Monitor and using the **Process new actions** and **Process new index data** actions. See the UTR Manual in \\thor\projects\docs\public\UTR.

### d. Test Full Text Search

Use FDD Client or WebFDD to full text search for your test set of documents. Confirm they are returned in the search results as expected. If so, you can proceed to migrating your documents to ES.

## 5. Add Documents to Elasticsearch

### a. How Long Will UTR Take?

Depending on how many documents you want to be full-text-searchable, it could take UTR hours or days to get those documents into Elasticsearch so users can look for them. The time UTR takes highly depends on your data - how many documents and pages you have - and whether your documents have been full-text-indexed before.

#### Example 1: Already have pages' text

If your full-text-searchable documents were processed through UTR before, when they were added to Autonomy IDOL, all of the text from their pages is already stored in the FDD database. No OCRing or other work is necessary to get the text since you already have it and the migration will go more quickly. This would normally be the case when migrating from IDOL to ES.

We did a migration like this for **3.390 million documents** containing **9.198 million pages**. UTR completed processing in **3 days and 3 hours**. The rate per hour was **123,000 pages per hour** and **45,200 documents per hour**. The rate per day was **2,943,000 pages per day** and **1,085,000 documents per day**. In this case, UTR ran with **8** helpers.

The number of UTR notes that need to be re-processed during the migration can be estimated with the SQL

#### Count of UTRed Documents

```
select count(*) from
fdd.doc_note_info where
note_type='U' ;
```

### Example 2: Fresh start with no page text

If your documents have not been processed through UTR before, then you are starting fresh and the pages need to be OCR'd and such in order to get the text. Since you are starting from scratch, this migration will take more time. This would normally be the case if you did not use full text search before.

We did a migration like this for **3.364 million documents** containing **9.115 million pages**. UTR completed processing in **12 days and 7 hours**. The rate per hour was **30,900 pages per hour** and **11,400 documents per hour**. The rate per day was **742,000 pages per day** and **274,000 documents per day**. In this case, UTR ran with **8 helpers**.

## b. Process Bins and File Cabinets

Take bins and file cabinets that have the **Full-Text Index** turned on and send them to UTR to be added to Elasticsearch. The below SQL will find the bins and file cabinets that you have full-text indexed and creates the actions for UTR to process.

### File Cabinets

```
insert into fdd.actions
(action, fc_id) select 'UTR',
fc_id from fdd.filecabinets
where utr='Y';
```

### Bins

```
insert into fdd.actions
(action, bin_id) select
'UTR', bin_id from fdd.bins
where utr='Y';
```

Alternatively, you can un-check and re-check the **Full-Text Index** option set on each bin and file cabinet in FCP. The result is the same but you have to un-check and re-check the option for each applicable file cabinet/bin. The above SQL does all applicable file cabinets or all applicable bins at once.

## c. Wait for UTR

Wait for UTR to process the documents into ES. As stated previously, this could take hours or days depending on your data. You can use the UTR Monitor to check on the progress, but when the Monitor is not needed you should disconnect to save UTR from sending information to the Monitor. For example, you could check on progress in the Monitor once a day.

Another way to check progress is to see how many documents are left for UTR to process using the SQL below.

### Number of remaining docs to process

```
select count(*) from
fdd.actions where
action='UTR' ;
```

This number is only accurate once UTR converts the fc\_id and bin\_id actions into the corresponding doc\_id actions.

#### d. Test Full Text Search

Use FDD Client or WebFDD to full text search for documents you expect to be full-text-searchable. Confirm they are returned in the search results as expected.

## 6. User Acceptance Testing

Instruct users to use Elasticsearch when full text searching:

- Search in FDD Client.
- Search in WebFDD and select the ES server in the **Servers** drop-down list. The server has the port **9200** in the name, e.g. myesserver:9200.

You can ask users to intentionally test full text search and/or allow them to search according to their normal usage/workload. Follow up with users to confirm they have done full text searches and the results come back as expected.

## 7. Hide IDOL

After users are happy with Elasticsearch, decide when you would like to hide Autonomy IDOL. Reasons to do this include:

- IDOL does not have recently-added documents. When UTR was switched over to ES, UTR stopped adding documents to IDOL.
- Users in WebFDD no longer have to select the ES server in the **Servers** drop-down list. ES is always used.

To hide IDOL:

1. Go into FCP's **Servers** module.
2. Delete the server of type **Autonomy IDOL Querying**. ★

If this delete fails, there are still some documents referencing the IDOL querying server. You can query the database to investigate and move any remnants from IDOL to ES or remove them, as appropriate. Once all references to the IDOL querying server are removed, you will be able to delete it.

★ Tell your users: In WebFDD they will no longer be able to select the IDOL server (with port **9000**) and the **Servers** drop-down list will disappear.

Why: WebFDD no longer detects an IDOL server for full text search and therefore no longer provides the option.

### 3. Delete the server of type **Autonomy IDOL Indexing**.

Deleting the server entries simply hides IDOL from the Feith applications. The IDOL database itself still contains all the documents it had and may sit dormant on its server.

At some later time it will be appropriate to turn IDOL off and eventually uninstall it.

## Related articles

- [Configure ES to log queries](#)
- [Autonomy IDOL URLs](#)
- [Set up Full Text Search](#)
- [Migrate from Autonomy IDOL to Elasticsearch](#)
- [Autonomy IDOL timeout settings in WebFDD](#)