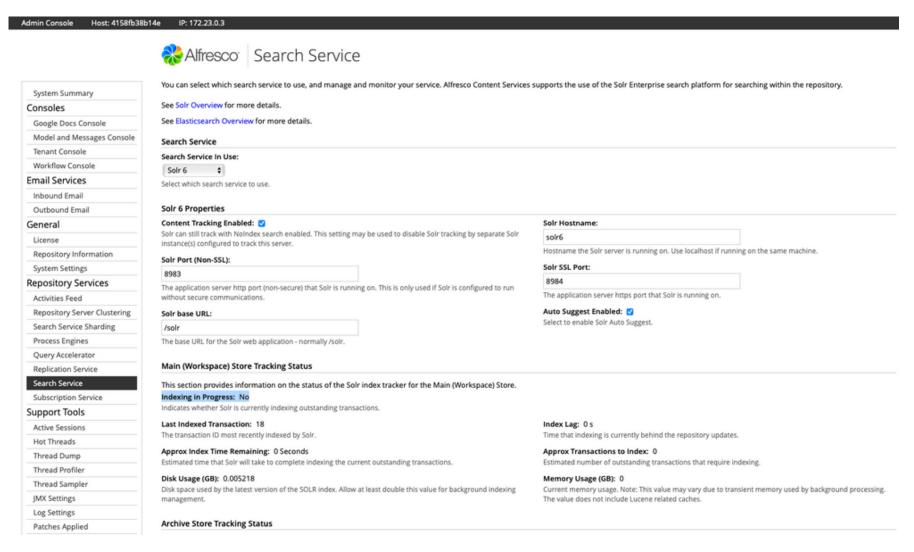**angelborroy**   👍2 💬1 969
Alfresco Employee

13 May 2022 10:4

Our colleague @aitseitz reached me to point out a missing feature related to **Content Indexing**. Despite we are still considering the addition of this feature to the product, this blog post describe alternative approaches to get required information.

Alfresco Search Services indexes metadata and content using different trackers. Once the metadata has been indexed, some components (like the Enterprise Admin Web Console for the Enterprise version) will report that no indexation is happening.

Alfresco | Search Service

System Summary

Consoles

Google Docs Console

Model and Messages Console

Tenant Console

Workflow Console

Email Services

Inbound Email

Outbound Email

General

License

Repository Information

System Settings

Repository Services

Activities Feed

Repository Server Clustering

Search Service Sharding

Process Engines

Query Accelerator

Replication Service

Search Service

Subscription Service

Support Tools

Active Sessions

Hot Threads

Thread Dump

Thread Profiler

Thread Sampler

JMX Settings

Log Settings

Patches Applied

You can select which search service to use, and manage and monitor your service. Alfresco Content Services supports the use of the Solr Enterprise search platform for searching within the repository.

See Solr Overview for more details.

See Elasticsearch Overview for more details.

Search Service

Search Service In Use:

Solr 6

Select which search service to use.

Solr 6 Properties

Content Tracking Enabled: ☑

Solr can still track with NoIndex search enabled. This setting may be used to disable Solr tracking by separate Solr instance(s) configured to track this server.

Solr Port (Non-SSL):

8983

The application server http port (non-secure) that Solr is running on. This is only used if Solr is configured to run without secure communications.

Solr base URL:

/solr

The base URL for the Solr web application - normally /solr.

Solr Hostname:

solr6

Hostname the Solr server is running on. Use localhost if running on the same machine.

Solr SSL Port:

8984

The application server https port that Solr is running on.

Auto Suggest Enabled: ☑

Select to enable Solr Auto Suggest.

Main (Workspace) Store Tracking Status

This section provides information on the status of the Solr index tracker for the Main (Workspace) Store.

Indexing in Progress: No

Indicates whether Solr is currently indexing outstanding transactions.

Last Indexed Transaction: 18

The transaction ID most recently indexed by Solr.

Approx Index Time Remaining: 0 Seconds

Estimated time that Solr will take to complete indexing the current outstanding transactions.

Disk Usage (GB): 0.005218

Disk space used by the latest version of the SOLR index. Allow at least double this value for background indexing management.

Index Lag: 0 s

Time that indexing is currently behind the repository updates.

Approx Transactions to Index: 0

Estimated number of outstanding transactions that require indexing.

Memory Usage (GB): 0

Current memory usage. Note: This value may vary due to transient memory used by background processing. The value does not include Lucene related caches.

Archive Store Tracking Status

However, despite all the metadata has been created in SOLR Index, the content of the documents is still being indexed by the *ContentTracker* component. While we improve the feature, this blog post will give you some tools to check the status of the content indexation.

Search Services 2.0.x

From Search Services 2.0 there is no content store in SOLR Side, so two new fields are required to identify if the content of a SOLR Document needs to be indexed or updated:

- LATEST_APPLIED_CONTENT_VERSION_ID: it corresponds to the identifier of the latest applied content property ID (it can be null)
- LAST_INCOMING_CONTENT_VERSION_ID: If the field has the same value of the previous one (or it is equal to SolrInformationServer.CONTENT_UPDATED_MARKER), then the content is supposed to be in sync. Otherwise, the content is intended as outdated and therefore it will be selected by the *ContentTracker*.

Additional details can be found in:

https://github.com/Alfresco/SearchServices/blob/master/search-services/alfresco-search/doc/architect...

The following SOLR Query will return the number of SOLR Documents with a different value in those *_CONTENT_VERSION_ID fields:

```
http://localhost:8983/solr/alfresco/select
?q=*
&fq={!frange l=1 u=1 v=$equals}
&equals=if(not(eq(LATEST_APPLIED_CONTENT_VERSION_ID,LAST_INCOMING_CONTENT_VERSION_ID)),1,0)
&indent=on
&wt=json

{
  "responseHeader":{ ... },
  "_original_parameters_":{ ... },
  "lastIndexedTx":574,
  "lastIndexedTxTime":1652429863621,
  "txRemaining":0,
  "response":{"numFound":76,"start":0,"docs":[ ... ]},
  "processedDenies":false
}
```

In this sample, the content of *76 documents* is still pending to be indexed or updated. While the lastIndexedTx points the latest TX in DB and the txRemaining value indicates there is no metadata pending to be indexed.

If you want to know the status of the nodes pending to be indexed or updated, you can add those *_CONTENT_VERSION_ID fields to the "fl" parameter:

```
http://localhost:8983/solr/alfresco/select
?q=*
&fl=[cached]LATEST_APPLIED_CONTENT_VERSION_ID,LAST_INCOMING_CONTENT_VERSION_ID
&fq={!frange l=1 u=1 v=$equals}
&equals=if(not(eq(LATEST_APPLIED_CONTENT_VERSION_ID,LAST_INCOMING_CONTENT_VERSION_ID)),1,0)
&indent=on
&wt=json

{
  "responseHeader":{ ... },
  "_original_parameters_":{ ... },
  "lastIndexedTx":574,
  "lastIndexedTxTime":1652429863621,
  "txRemaining":0,
  "response":{"numFound":7,"start":0,"docs":[
      {
        "LATEST_APPLIED_CONTENT_VERSION_ID":591,
        "LAST_INCOMING_CONTENT_VERSION_ID":-10},
      {
        "LATEST_APPLIED_CONTENT_VERSION_ID":393,
        "LAST_INCOMING_CONTENT_VERSION_ID":-10},
      {
        "LATEST_APPLIED_CONTENT_VERSION_ID":573,
        "LAST_INCOMING_CONTENT_VERSION_ID":-10},
      {
        "LATEST_APPLIED_CONTENT_VERSION_ID":579,
        "LAST_INCOMING_CONTENT_VERSION_ID":-10},
      {
```

```
        "LATEST_APPLIED_CONTENT_VERSION_ID":606,

        "LAST_INCOMING_CONTENT_VERSION_ID":-10},

      {

        "LATEST_APPLIED_CONTENT_VERSION_ID":582,

        "LAST_INCOMING_CONTENT_VERSION_ID":-10},

      {

        "LATEST_APPLIED_CONTENT_VERSION_ID":585,

        "LAST_INCOMING_CONTENT_VERSION_ID":-10}]

    },

    "processedDenies":false

  }
```

In this sample, we can see that the content of the *7 documents* is still pending to be indexed, since the LAST_INCOMING_CONTENT_VERSION_ID value is set to SolrInformationServer.CONTENT_OUTDATED_MARKER (-10)

Alternatively, Search Services Admin REST API can be used to get this information.

http://localhost:8983/solr/admin/cores?action=summary&core=alfresco

```
<lst name="FTS">

   <long name="Node count whose content is in sync">170</long>

   <long name="Node count whose content needs to be updated">7</long>

</lst>
```

Search Services 1.4.x

When using Search Services 1.3.x / 1.4.x the logic is quite different.

There is a field in the Solr schema called FTSSTATUS that could have the following domain:

- *Clean*: the text content of the document is in sync, no update is needed
- *New*: the document has been just created, it has to be updated with the corresponding text content
- *Dirty*: the text content of the document changed, the new content needs to be retrieved and the document updated

The following SOLR Query will return the number of SOLR Documents faceted by FTSSTATUS field:

```
http://localhost:8080/solr/alfresco/select?facet.field=FTSSTATUS&facet=on&indent=on&q=*&wt=json


{
   "responseHeader":{ ... },
   "_original_parameters_":{ ... },
   "lastIndexedTx":136,
   "lastIndexedTxTime":1652430806500,
   "txRemaining":0,
   "response":{"numFound":874,"start":0,"docs":[ ... ]},
   "facet_counts":{
     "facet_fields":{
       "FTSSTATUS":[
         "New",152,
         "Clean",125,
         "Dirty",3]},
   "processedDenies":false
}
```

In this sample, we can see the content of 152 *New* Documents and 3 *Dirty* Documents need to be indexed. The content of 125 Documents is in sync (*Clean*) and it's not required.

Alternatively, <u>Search Services Admin REST API</u> can be used to get this information.

http://localhost:8983/solr/admin/cores?action=summary&core=alfresco

```
<lst name="FTS">
  <long name="Node count with FTSStatus Clean">125</long>
  <long name="Node count with FTSStatus Dirty">3</long>
  <long name="Node count with FTSStatus New">152</long>
</lst>
```
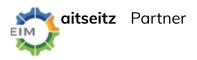
Labels :

Search Services 1.4.x   **Search Services 2.0.x**

👍

| 2 Kudos |

Comments

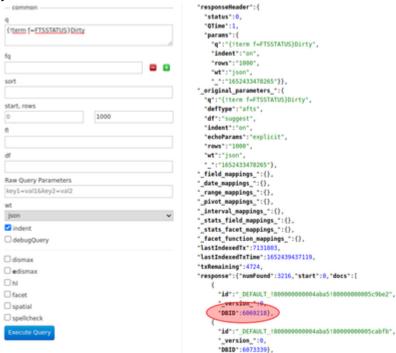**aitseitz** Partner   13 May 2022
                                                                    4:18 PM

Muchas Gracias @angelborroy for this article.

Additional information for ASS 1.4:

In case you would like to know which documents are in the FTSStatus "Dirty" state
http://localhost:8983/solr/#/alfresco/query?q=%7B!term%20f%3DFTSSTATUS%7DDirty

you can use the DBID

and fire up a query in the Alfresco Node Browser (Admin Console oder Share Administration Tools) with @sys:node-dbid:<DBID>

to get more information about the document in the "dirty" state.