

Alfresco Architecture

Search Server integration



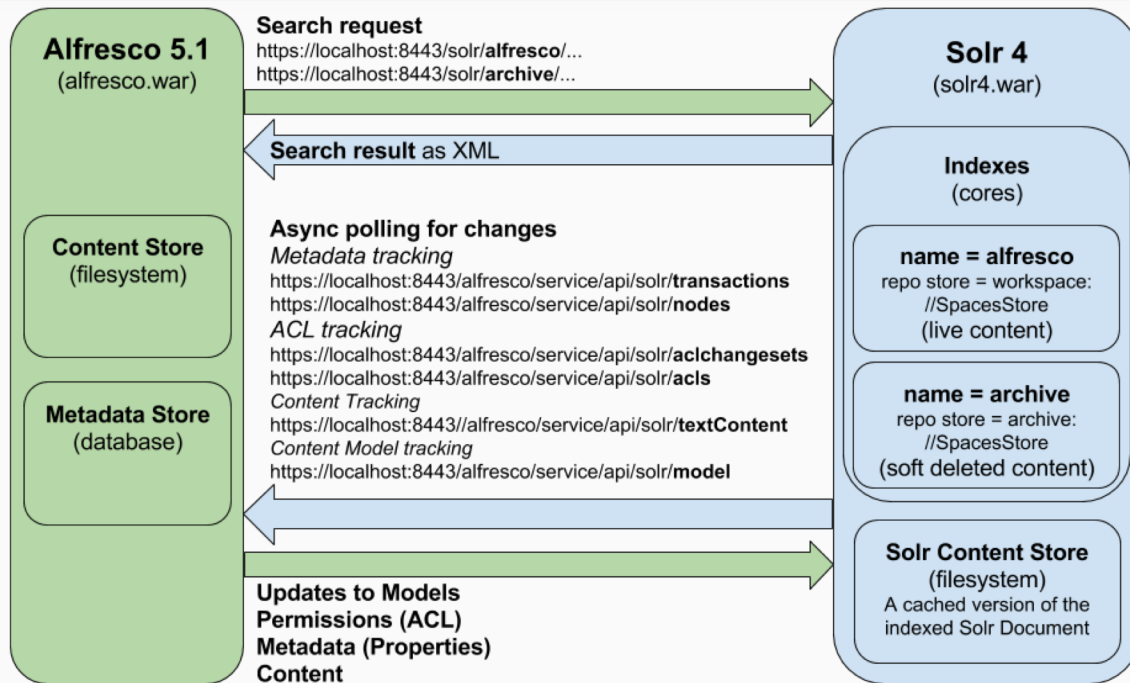
How does the integration work?

- The easiest way to put it is as follows:
 - Alfresco uses HTTP GETs to talk to Solr and search for content.
 - Solr updates the indexes by talking to Alfresco to get the number of transactions that have been committed since it last talked to Alfresco.
 - The custom components in Solr that talks to Alfresco are referred to as trackers.
- Solr talks to Alfresco every 15 seconds to see if any new transactions have been committed. So the index will eventually be up to date with the database.
 - This means that if you upload a document to Alfresco and then immediately execute a full text search, it's not 100% sure that you will hit the document just uploaded. An important thing to know is that unlike with Solr 1, Solr 4 indexing is split between metadata tracking, ACL tracking, and content tracking.

How does the integration work? continued

- Solr will also keep track of new custom content models that have been deployed and download them to be able to index the properties in these models.
- Permission settings will also be downloaded by Solr so it can do query time permission filtering.
- There are 2 cores (i.e. indexes):
 - **WorkspaceStore**: for searching all live content (i.e. stuff in `alf_data/contentstore`)
 - **ArchiveStore**: for searching soft deleted content (i.e. content in `alf_data/contentstore` that has been marked as deleted)

The big picture



To think about

- As the picture showed, Solr version 4 (or 6) is used and has been extended with code to talk to Alfresco.
 - So if you are going to install Solr on a separate dedicated search box, then you need to download it from Alfresco download site and not from Apache...
- If you wanted to use a newer version of Solr such as for example 6.5, then that would be a bit of a problem. It would not be a simple task to upgrade and patch it with Alfresco specific code for trackers etc, and Alfresco might not support you...
- Note the new Solr Content Store that has been added to store every indexed Solr Document, including metadata and content text (new for the Alfresco 5 - Solr 4 integration).
 - This way things like re-indexing and updates can be done in a quicker way.

So, when/where exactly is Solr used?

- Instant Search (search field in top right corner)
- Advanced Search
- Filters
- Tags
- Categories (implemented as facets)
- Dashlets such as the Recently Modified Documents, Site Content
- People Finder and Site Finder pages
- Wildcard searches for People, Groups, Sites (uses database search if not wildcard)