Current state

A blueprint to incorporate Elasticsearch into the Folio platform has been created. The Shanghai Public Libraries group has implemented portions of that blueprint in their solution and diverged in other areas. Their solution can be considered as a proof of concept for the Folio community solution and serve as a starting point for further work.

Implementation phases:

The suggested approach is to build upon the SPL PoC and build out additional parts of the blueprint using a thin thread approach. We will focus the development in one of the many use cases for Elasticsearch and build out end-to-end, production grade, capabilities. It is proposed to focus on the Inventory domain as the first case.

Phase 1

Provides the ability to search Inventory through Elasticsearch.

Inventory and SRS Changes (deliverable 1)

Provide Inventory support for ES indexing

- Sending update notification messages from inventory and source record storage (SRS)
- Providing Inventory and SRS APIs for fetching view for indexing by ids
- *Extract common library for using it in other modules*

Search (deliverable 2)

Provide capabilities to consume the ES indexes

- Create Search Indexing module
  - Design and create metadata tables
  - Implement creation of search indexes' settings and mappings based on metadata
  - Implement embedding of the linked entities in order to allow searching over attributes of linked entities (e.g. adding items and holdings attributes to instances)
  - Implement receiving of notification message
- Create Search Query module
  - Implement simple searching API (CQL or REST?)

Search User Experience

- First pass will replicate existing search experience in Inventory?
- Update Stripes Components to support above search API
- Update Inventory UI with new search components

- Provide "switcher" to allow selectively using ES search or Legacy search (Inventory)
  - Possibly expose ES as "Full Text Search" on existing user experience

Infrastructure (deliverable 3)

Provide the infrastructure necessary to support Elasticsearch
- Add Elasticsearch cluster to CI/CD and setup it on environments (k8s conf)
- Check configuration of existing Kafka cluster

Phase 2

Elasticsearch for logs aggregation

Elasticsearch would be used for application logs aggregation. The same ES cluster can be leveraged for both logs and instances search. Nevertheless the business logic for instances search and logs aggregation has significant differences and therefore only a small part of it (or even nothing) can be extracted as a common part.

Phase 3

Possible implementation of rich search capabilities

Based on requirements provided by POs the following functions can be implemented:
- Facets
- Advanced relevance
- Missing words strikeout
- Autocomplete
- Didyoumean

Implementation for other modules should be done considering the data amount and requirements for each module. Possible candidates are:
- Acquisitions
- Circulation
- Users
- Others