Elastic Search

Overview

- Search engine
- Front end to Lucene
- Can be distributed over multiple servers
- Scalable to handle many simultaneous transactions
- NoSQL database
- Http interface and api's for several languages
- Built in redundancy in case of server failure

Lucene

- Search engine written by Doug Cutting
- He wrote it in java
- His fifth search engine
- Adopted by the Apache Software Foundation
- Open source
- Ported to Perl, C#, C++, Python, Ruby and PHP
- Full text searching

Storage

- Chunks of information are called documents
- Eg one document per film
- Documents are composed of fields
- Eg title, producer, summary
- Fields are searchable
- Documents are indexed

Elastic Search Head

- An open source inspection tool for elastic search

Indexing

- Is the key to fast retrievals
- Uses a tokenizer to transform the text
- Removes endings eg running → run
- Removes articles eg 'a' → ''
- This is controllable through the api
- They are language based
- Koha uses a European language tokenizer
- Can have different tokenizers for different language fields

Matching

- Queries are processed by the same tokenizer
- Matching is done by simple character comparison
- If the tokens are not identical – no match
- A weighting system is used to control match relevance ie how the results are sorted.
- Can modify the weighting for your application

Elastic Search in Koha

- It is an alternative to zebra search engine
- Working now since the 19.05 release
- Was partly implemented in older releases
- Uses version 5 but there are patches for 6.4
- There are two indexes in koha: bibliographic and authorities.
- Some item information is add to each document too

Kibana

- A tool made by Elastic Corp. for inspecting indexes

Configuring Elastic Search

- From the admin page