

## Infrastructure - Story #7939

### Indexing is too slow, especially with large packages

2016-11-25 19:03 - Dave Vieglais

<b>Status:</b>	Rejected	<b>Start date:</b>	2016-11-25
<b>Priority:</b>	High	<b>Due date:</b>	
<b>Assignee:</b>		<b>% Done:</b>	30%
<b>Category:</b>	d1_indexer	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	CCI-2.4.0		
<b>Story Points:</b>			

#### Description

It appears that the indexing process is far too slow to keep up with content additions and changes. Since the version 2.3 upgrade which includes support for multiple indexing threads, the performance appears improved, but it falls far short of what is needed to provide reasonable currency.

In particular, it appears that large resource maps such as those provided by the ARCTIC node are very slow to evaluate.

Some optimization may be possible without major refactoring of the indexing process.

A few possible options:

1. Check that changes to properties such as ownership do not trigger an entire re-index of the package. If permissions change, then there is no need to reindex the entire package since other properties are unchanged. This should be in place now since content is immutable, and only mutable metadata fields should be updated.
2. Dedicate a single thread to resource map processing, expanding to more threads when there is no backlog of other content. This would allow efficient processing of content on which the resource map indexing may depend.
3. Refactor the index so that resource maps may be processed independently, without the need for all other objects to be loaded and processed.
4. Refactor the indexing of resource maps so that a partially processed resource map is persisted so that processing may continue as content becomes available rather than starting from scratch each time.

#### Related issues:

Related to Infrastructure - Story #8175: Solr crashing with out of memory error **Closed** **2017-09-05**

#### History

##### #1 - 2016-12-01 23:51 - Dave Vieglais

- Target version changed from CCI-2.3.1 to CCI-2.4.0

##### #2 - 2017-09-05 16:24 - Dave Vieglais

- Category set to d1\_indexer

##### #3 - 2017-09-05 16:33 - Dave Vieglais

- Related to Story #8175: Solr crashing with out of memory error added

##### #4 - 2018-01-17 18:57 - Dave Vieglais

- Sprint set to Infrastructure backlog

##### #5 - 2018-01-17 18:58 - Dave Vieglais

- % Done changed from 0 to 30

- Status changed from New to In Progress

**#6 - 2018-01-17 18:58 - Dave Vieglais**

*- Status changed from In Progress to Rejected*