

Infrastructure - Story #8702

Indexing Refactor Strategy

2018-09-21 22:42 - Rob Nahf

Status: New	Start date: 2018-09-24
Priority: Normal	Due date:
Assignee: Rob Nahf	% Done: 0%
Category: d1_indexer	Estimated time: 0.00 hour
Target version: CCI-2.4.0	
Story Points:	
Description	
Indexing is non-performing and has some inconsistency problems.	
A solution was developed that addresses the main issues, and involves the creation of a separate solr core for relationships (the resource maps). Initially, the solution will create the separate core as a behind the scenes reference for the main search index. Relationships (resource_map, documents, isDocumentedBy) will still be copied into the main search record.	
Additionally, archived objects will not be removed from the index, but the field archived will be added to the schema.	
The new logic for processing resource maps and archiving objects should remove many of the inefficient checks that cause records to be reindexed.	
The main phases for development will be:	
1. refactor out the custom solr client for use of the standard org.apache.solrj-client.	
2. migrate the schema to include archived field & introduce relationships core. Refactor the resourcemap subprocessor to use it, and trigger relationship tasks.	
3. refactor the delete subprocessor (for archived records) & add the search handler.	
Subtasks:	
Task # 8703: test the cleaned up indexer in DEV	New
Related issues:	
Related to Infrastructure - Bug #8696: double indexing of a resource map and ...	New 2018-09-12
Related to Infrastructure - Bug #3675: package relationships not available fo...	New 2013-03-20
Related to Infrastructure - Bug #8542: resource Map not indexed when the reso...	Closed 2018-04-14
Related to Infrastructure - Bug #8536: resource Map update when metadata SID ...	In Progress 2018-04-09
Related to Infrastructure - Story #8537: indexer doesn't populate SID-defined...	New 2018-04-09
Related to Infrastructure - Story #8363: indexer shutdown generates index tasks	New 2018-02-12
Related to Infrastructure - Story #8172: investigate atomic updates for some ...	In Progress 2017-09-01
Related to Infrastructure - Story #8082: implement SolrCloudClient to replace...	New 2017-04-25
Related to Infrastructure - Bug #8182: relationships defined using SIDs in re...	New 2017-09-11
Related to Infrastructure - Bug #8093: resource map index processing is ineff...	New 2017-05-10

History

#1 - 2018-09-21 22:51 - Rob Nahf

- Related to Bug #8696: double indexing of a resource map and another not processed because of resource contention (lock) on member added

#2 - 2018-09-21 22:52 - Rob Nahf

- Related to Bug #3675: package relationships not available for archived objects added

#3 - 2018-09-24 15:17 - Rob Nahf

- Related to Bug #8542: resource Map not indexed when the resourceMap SID is used in triples. added

#4 - 2018-09-24 15:17 - Rob Nahf

- Related to Bug #8536: resource Map update when metadata SID is used is not indexed added

#5 - 2018-09-24 15:20 - Rob Nahf

- Related to Story #8537: indexer doesn't populate SID-defined relationships unless new resourceMap is submitted added

#6 - 2018-09-24 15:22 - Rob Nahf

- Related to Story #8363: indexer shutdown generates index tasks added

#7 - 2018-09-24 15:28 - Rob Nahf

- Related to Story #8172: investigate atomic updates for some solr updates added

#8 - 2018-09-24 15:29 - Rob Nahf

- Related to Story #8082: implement SolrCloudClient to replace HttpService to allow concurrent updates of the solr index from differen machines added

#9 - 2018-09-24 15:30 - Rob Nahf

- Related to Bug #8182: relationships defined using SIDs in resource maps don't stay current added

#10 - 2018-09-24 15:32 - Rob Nahf

- Related to Bug #8093: resource map index processing is inefficient (2sec / referenced object) added