

CN Index - Task #3766

Story # 3760 (Closed): Index Maintenance Story

Speeding up index build time

2013-05-16 17:03 - Skye Roseboom

Status:	Closed	Start date:	2016-04-01
Priority:	Normal	Due date:	
Assignee:	Andrei Buium	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:			
Story Points:			

Description

Investigate alternative solr commit strategies while building the index to decrease time to takes to build index.
- For example 'commitWithin' seconds instead of hard commits on every update.

Possibly also substitute out the hazelcast hzIdentifiers iteration strategy for looping over all documents and move to a DAO/postgres query paging solution. Decrease network traffic/wait time.

Associated revisions

Revision 17774 - 2016-03-14 19:30 - Andrei Buium

Committed a change to d1_cn_index_tool to allow it to output a performance log file with the same logging statements as the index task processor.
refs #3766

Revision 17774 - 2016-03-14 19:30 - Andrei Buium

Committed a change to d1_cn_index_tool to allow it to output a performance log file with the same logging statements as the index task processor.
refs #3766

History

#1 - 2013-05-16 17:03 - Skye Roseboom

- Parent task set to #3760

#2 - 2016-02-25 14:13 - Dave Vieglais

- Assignee changed from Skye Roseboom to Andrei Buium

reassigning for review.

#3 - 2016-04-01 19:38 - Andrei Buium

- Start date set to 2016-04-01

- % Done changed from 0 to 30

- Status changed from New to In Progress

Adding support to do batch adds from the index build tool has sped up tests significantly.

We also had commit=true hard-coded into the solr update url. Removing this speeds up performance as well, but it appears autoCommit isn't taking over in its absence. This needs investigating, because having the commit happen outside of the index tool code execution path would be ideal.

#4 - 2017-03-28 16:10 - Dave Vieglais

- % Done changed from 30 to 100

- Status changed from In Progress to Closed