

Infrastructure - Bug #8185

Index-processor leaves lots of "IN PROCESS" index tasks in the index table

2017-09-14 19:19 - Jing Tao

Status:	Closed	Start date:	2017-09-14
Priority:	High	Due date:	
Assignee:	Jing Tao	% Done:	100%
Category:	d1_indexer	Estimated time:	0.00 hour
Target version:	CCI-2.3.6	Story Points:	
Milestone:	None		
Product Version:	*		

Description

Dave found this one the production cn:  
[cn-orc-1.dataone.org] run: psql -h localhost -U dataone\_readonly d1-index-queue -P pager=off --single-transaction -o /tmp/sql\_150532373616459.txt -c "SELECT status,deleted,COUNT(\*) AS cnt,to\_timestamp(min(datesysmetamodified)/1000) as minTmod,to\_timestamp(max(datesysmetamodified)/1000) as maxTmod FROM index\_task GROUP BY status,deleted ORDER BY status;"  
[cn-orc-1.dataone.org] download: /tmp/qcount\_150532373616438.txt <- /tmp/sql\_150532373616459.txt  
[cn-orc-1.dataone.org] run: rm /tmp/sql\_150532373616459.txt  
status | deleted | cnt | mintmod | maxtmod  
-----+-----+-----+-----+-----  
FAILED | f | 6418 | 2011-05-16 20:33:32+00 | 2017-09-13 16:12:13+00  
IN PROCESS | f | 28227 | 2012-06-15 02:39:52+00 | 2017-09-06 13:06:12+00  
(2 rows)  
  
I looked at the orc-sandbox (the test server) first and found it also had 18511 index tasks with the status "IN PROCESS" in sandbox. I also wrote a script file to check if the pids already have the solr doc in the solr server. The result shows that 18057 objects have a slor record while 454 objects don't have. I randomly chose ten pids from the list which has the solr index and compared the dateModified in the solr record with the one in its system metadata. All of ten objects' solr records match their system metadata value. This means the solr index was built upon the newest system metadata. This is good.  
  
The format id of those objects varies. It can be a meta data, data or resource map object. The only thing I can image is the index processor didn't remove or change the status of index tasks after it finished indexing in some special circumstances. But I don't find it on the code at the first glance.

History

#1 - 2017-10-04 20:28 - Jing Tao

- Priority changed from Normal to High

It happened again, we need to figure it out soon.

#2 - 2017-10-04 20:50 - Jing Tao

Chris's comment:  
The idea that tasks are staying in the IN PROCESS status is concerning. They really should be deleted once complete. So, I'm wondering if we might have a saturated JDBC connection pool or something, and perhaps we aren't handling the exception correctly when say a delete call fails to get a connection. I'm not sure, but we might want to look at the exceptions to see if anything like that is happening

#3 - 2017-10-20 16:16 - Jing Tao

I did a test: Made an CN index-processor index tasks and more than seven thousands index tasks were marked with the status of "IN PROCESS" in the db. At this time, I stopped the index processor. I checked the db and found those seven thousand tasks were still with the status of "IN PROCESS" while the index processor not working. Their status never changed at all.  
So the stop process of the index-processor has an issue to handle those queued index tasks.

#### **#4 - 2017-10-30 19:05 - Jing Tao**

- *Status changed from New to Closed*

- *% Done changed from 0 to 100*

We did two things:

First to make sure the `IndexTaskProcessor.shutdown` method will be called.

Second, in that method, a try-catch clause was added to the loop in order to avoid one failure causing entire stop.