Infrastructure - Task #3407

Fix indexing conflicts on CNs that cause Metacat systemmetadata tables get out of sync

2012-11-26 16:56 - Chris Jones

Status:	Closed	**Start date:**	2012-11-26
Priority:	Normal	**Due date:**	
Assignee:	Chris Jones	**% Done:**	100%
Category:	Metacat	**Estimated time:**	0.00 hour
Target version:	Sprint-2012.44-Block.6.2		
Milestone:	CCI-1.1	**Story Points:**	
Product Version:	*		

Description

During testing of d1_replication, we've seen the smreplicationstatus tables get out of sync across CNs. For instance, two CNs may have a status of COMPLETED for a replica, and the third CN still says REQUESTED. This looks to coincide with SQL errors showing a foreign key violation in the xml_documents table during calls to CNodeService.archive(). After chatting with Ben, the current thought is that, due to the asynchronous nature of the indexing process in Metacat to populate xml_index is causing the issue. When an object gets created, the index process is queued, but the object may quickly get archived before the indexing is complete. Or, the indexing may complete, causing the archive() method to through a SQL exception because the reference to the docid is still in the xml_index table as a foreign key.

These issues may be alleviated if we don't use Metacat's indexing feature on the CNs. Determine if we can set a flag to not push docids into the IndexerQueue. Test this new Metacat code in the sandbox environment with high-transaction rates that frequently call the archive() method.

Ben, Matt - thoughts on the consequeunces of turning off Metacat indexing on the CNs?

History

#1 - 2012-12-03 16:03 - Chris Jones

- Status changed from New to In Progress

Ben has changed the IndexingQueue handling in Metacat to remove indexing tasks when delete() and archive() have been called, and replication runs so far show no foreign key violation SQLExceptions.

#2 - 2012-12-03 16:04 - Chris Jones

- Status changed from In Progress to Testing

#3 - 2013-01-08 00:06 - Chris Jones

- Status changed from Testing to Closed

- translation missing: en.field_remaining_hours set to 0.0

We've also changed calls to update (not just delete) systemmetadata to ensure that there are no outstanding indexing tasks in the queue that would cause a subsequent SQL exception. This has been tested in sandbox and stage, and we are not seeing the exceptions anymore.