

Metacat - Bug #4367

metacat didn't update xml_path_index table while a document was updated

09/04/2009 10:09 AM - Jing Tao

Status:	Resolved	Start date:	09/04/2009
Priority:	Immediate	Due date:	
Assignee:	Michael Daigle	% Done:	0%
Category:	metacat	Estimated time:	0.00 hour
Target version:	1.9.2	Spent time:	0.00 hour
Bugzilla-Id:	4367		
Description			
Hi Jing:			
Marissa Bauer, working on the POD project, created a new data package using Morpho - she has created about 25 of them so far. The latest accession # is mbauer.42.9.			
Unlike the rest of the packages, this one is not included in the results of an NCEAS Data Repository search on the string 'POD!'. KNB portal and Morpho searches both to find the package.			
Initially, Marissa did not include an organizationName field with NCEAS: 12192 string in it, so I added that with the Morpho editor.			
In any case, can you take a look at the DP (mbauer.42.9) and let me know what might be missing?			
Thanks,			
Rick R			
Hi, Rick:			
Since you can find the mbauer.42.9 by searching in knb page and morpho, I thought this is a nceas skin filter issue. I looked the mbauer.42.9 and found it has the filter - organizationName="National Center for Ecological Analysis and Synthesis". I am confused - NCEAS skin should find it.			
I carefully checked the nceas skin sql query and the metacat database, it turned out that xml_path_index table hasn't been updated since mbauer.42.5. When I run "select * from xml_path_index where docid like 'mbauer.42';", I got:			
4196745 mbauer.42 @packageId i2% i2% i2% i2% i2% i2% i2% i2% 0 i2% i2%304886418 mbauer.42.5			
.....			
4196773 mbauer.42 organizationName i2% i2% i2% i2% i2% i2% i2% i2% 0 i2% i2%304886527 NCEAS			
.....			
You see, in xml_path_index table, the organizationName is "NCEAS", rather than "National Center for Ecological Analysis and Synthesis", so the nceas skin query couldn't catch the docid. Also we can see the package id is still mbauer.42.5. But the package id is mbauer.42.9 in document mbauer.42.9.			
So I think it is a metacat bug - the xml_path_index table wasn't updated while the document was updated. In your case, the current document is mbauer.42.9, but the in xml_path_index is still mbuaser.42.5.			
Related issues:			
Blocks Metacat - Bug #4680: Geoserver StringIndexOutOfBoundsException		Resolved	01/20/2010

History

#1 - 01/11/2010 02:25 PM - Michael Daigle

I haven't been able to reproduce this using the dev skin in metacat. I'll try using morpho.

#2 - 01/11/2010 03:48 PM - Michael Daigle

So far I haven't been able to reproduce this with the dev skin or morpho.

#3 - 01/13/2010 02:18 PM - Michael Daigle

I should test this by reindexing doc via api and looking for errors

#4 - 01/20/2010 10:52 AM - Michael Daigle

Forcing index via the API causes a `StringIndexOutOfBoundsException` when processing spatial coordinates. See bug 4680 which addresses this issue.

In the meantime, I added a catch in `DocumentImpl.buildIndex` that logs the exception and continues on.

#5 - 01/20/2010 11:24 AM - Michael Daigle

Hmm, there is also a foreign key constraint violation. Inserting into `xml_index` violates a constraint on `xml_nodes.nodeid`. This may be a race condition. I've seen it happen during unit testing and then go away.

I think the geoserver issue is unrelated.

#6 - 01/20/2010 03:16 PM - Michael Daigle

It looks like this is a race condition. The `xml_nodes` insertion commit must still be going on when the indexing kicks off. However, the failed index should get readded to the indexing queue for retry (up to 25 times).

In this case the `SQLException` was being caught and reported, but not passed on. I rethrow the exception in `DocumentImpl.buildIndex()` so it gets readded to the indexing queue and retried.

#7 - 03/27/2013 02:26 PM - Redmine Admin

Original Bugzilla ID was 4367