

CN REST - Bug #7161

TERN object fails to be indexed by Solr, but successfully synchronized

2015-06-05 18:02 - Mark Servilla

| | | | |
|------------------------|--------------------|------------------------|------------|
| Status: | New | Start date: | 2015-06-05 |
| Priority: | Normal | Due date: | |
| Assignee: | | % Done: | 0% |
| Category: | d1_synchronization | Estimated time: | 0.00 hour |
| Target version: | | | |
| Story Points: | | | |

Description

TERN object aekos.org.au/collection/nsw.gov.au/nsw_atlas/vis_flora_module/V_ILLAWDB3.20150515 fails to be indexed by Solr on cn.dataone.edu (cn-ucsb-1) even though it was successfully synchronized. Investigation indicates it was not added to the HazelCast ObjectPathMap structure according to Skye Roseboom:

Im not sure why this pid is not appearing in the hazelcast ObjectPathMap structure.

I looked into the metacat database schema a bit and noticed the pid does not seem to appear in the 'identifier_mapping' table:

```
select * from identifier_mapping where
guid='aekos.org.au/collection/nsw.gov.au/nsw_atlas/vis_flora_module/V_ILLAWDB3.20150515';
```

| | | |
|------|-------|-----|
| guid | docid | rev |
|------|-------|-----|

(0 rows)

Without the 'docid' or 'localid' as metacat calls them, Im don't thing the pid could be added to the objectPathMap in hazelcast.

Related issues:

| | | |
|--|---------------|-------------------|
| Related to CN REST - Bug #7222: SEAD object only partially synchronized - mis... | Closed | 2015-06-18 |
|--|---------------|-------------------|

History

#1 - 2015-06-05 18:04 - Mark Servilla

The following is excerpted from the email thread between Mark Servilla, Skye Roseboom, and Chris Jones:

On Wed, Jun 3, 2015 at 3:46 PM, Mark Servilla mark.servilla@gmail.com wrote:
Hi Chris/Skye,

I'm just throwing this information out to you guys and seeing if any of it sticks (i.e., makes sense).

1. The autogen identifier for this object is: `autogen.2015052119280874140` - the file does not exist on the filesystem in `/var/metacat/documents` (or other metacat directories, for that matter).
2. There is only one entry for the PID in `hazelcast-storage.log.5`: `[DEBUG] 2015-05-22 02:28:08,131 (FactoryImpl$8:process:870) [128.111.54.80]:5701 [DataONE] Instance created ProxyKey {name='lock', key=aekos.org.au/collection/nsw.gov.au/nsw_atlas/vis_flora_module/V_ILLAWDB3.2015051}`
3. There are no entries for the PID in any of the Metacat log files in `/var/metacat/logs`.
4. There are no entries for the `autogen.2015052119280874140` identifier in any of the Metacat or Hazelcast (or Replicate) log files in `/var/metacat/logs`.
5. I've reviewed the Metacat log file `metacat.log.1` around the `d1_sync` time, but do not see any `ERROR` labels in that time period.

My head now hurts. :-)

Mark

Mark Servilla
mark.servilla@gmail.com

On Wed, Jun 3, 2015 at 1:25 PM, Christopher Jones cjones@nceas.ucsb.edu wrote:
Hi Mark,

Well, hmm.

I'd also check if the autogen file for that pid is present in /var/metacat/documents since it's an EML doc and should be added to the CN. If it's not in that directory, the problem would be on the Metacat side, and I'd look through /var/metacat/logs/metacat.log* around the create() timestamp in d1_sync to see what happened.

The last thing I'd look at is the index processor log file around the timestamp (or maybe a bit after) the CN.create() was called for that pid in d1_sync. Maybe there was some other error specific to indexing?

Cheers,
C

On Jun 3, 2015, at 12:54 PM, Mark Servilla mark.servilla@gmail.com wrote:

Hi Chris/Skye,

There are entries for PID "aekos.org.au/collection/nsw.gov.au/nsw_atlas/vis_flora_module/V_ILLAWDB3.20150515" in both the Metacat systemmetadata and identifier tables on cn-ucsb-1.dataone.org - no obvious ommisions in the tuple attribute fields. I didn't see any obvious error messages in the synchronization log (cn-synchronize.log.13). Any other suggestions/thoughts? Thank you!

Sincerely,
Mark

Mark Servilla
mark.servilla@gmail.com

On Wed, Jun 3, 2015 at 10:20 AM, Christopher Jones cjones@nceas.ucsb.edu wrote:
Hi Skye,

The table you should look at is the 'identifier' table. the 'identifier_mapping' table isn't part of the core metacat schema, and I'd guess it was part of either an upgrade process or our efforts to re-sync the CNs due to HZ split brain - I can't quite remember.

Anyway, is there an entry in the systemmetadata table for that pid? Do you see any d1_sync errors during the time it was harvested? I'm not sure why it wouldn't have a guid (pid) to docid mapping in the identifier table if the call to CN.create() worked. I'd check if that call failed first.

Cheers,
Chris

On Jun 3, 2015, at 10:01 AM, Skye Roseboom sroseboo@epscor.unm.edu wrote:

Hey Mark, Chris

I was on cn-ucsb-1.dataone.org this morning so I tried pushing this object (aekos.org.au/collection/nsw.gov.au/nsw_atlas/vis_flora_module/V_ILLAWDB3.20150515) through the index but I am getting an error regarding its 'object path' value not being present in hazelcast ObjectPathMap:

```
[ INFO] 2015-06-03 15:46:26,969 (IndexTaskProcessor:getNextIndexTask:155) Start of indexing pid:
aekos.org.au/collection/nsw.gov.au/nsw_atlas/vis_flora_module/V_ILLAWDB3.20150515
[ INFO] 2015-06-03 15:46:27,883 (IndexTaskProcessor:isObjectPathReady:263) Object path for pid:
aekos.org.au/collection/nsw.gov.au/nsw_atlas/vis_flora_module/V_ILLAWDB3.20150515 is not available. Task will be retried.
[ INFO] 2015-06-03 15:46:27,891 (IndexTaskProcessor:getNextIndexTask:164) Task for pid:
aekos.org.au/collection/nsw.gov.au/nsw_atlas/vis_flora_module/V_ILLAWDB3.20150515 not processed.
```

Im not sure why this pid is not appearing in the hazelcast ObjectPathMap structure.

I looked into the metacat database schema a bit and noticed the pid does not seem to appear in the 'identifier_mapping' table:

select * from identifier_mapping where guid='aekos.org.au/collection/nsw.gov.au/nsw_atlas/vis_flora_module/V_ILLAWDB3.20150515';

| guid | docid | rev |
|------|-------|-----|
|------|-------|-----|

(0 rows)

Without the 'docid' or 'localid' as metacat calls them, Im don't thing the pid could be added to the objectPathMap in hazelcast.

Chris - is identifier_mapping the correct table in metacat where the 'guid/pid' is mapped to the localid/docid - in order to find it on the file system? A call to /object for this pid appears to work but this value does not seem to appear in the hazelcast ObjectPathMap. Any ideas?

Thanks!
-s

On Jun 2, 2015, at 10:16 AM, Mark Servilla mark.servilla@gmail.com wrote:

Hi Skye,

There is a single science metadata object from the urn:node:TERN MN that appears not to have been indexed into Solr: aekos.org.au/collection/nsw.gov.au/nsw_atlas/vis_flora_module/V_ILLAADB3.20150515. I don't see any obvious errors with this object (nothing in the index log files) and it does appear to have synchronized without issue (note that 1188 of 1189 objects did successfully index into Solr). I am fine with re-indexing this particular object, but I do want to confirm that there will be no negative side-effects on production when shutting down the "d1-index-task-processor" task to perform the re-index. Is this correct?

Thanks,
Mark

Mark Servilla
mark.servilla@gmail.com

#2 - 2015-06-18 01:44 - Mark Servilla

- Related to Bug #7222: SEAD object only partially synchronized - missing autogen.2015061616000265251 document from /var/metacat/documents added