

Member Nodes - Support #8020

tDAR - Resolve MN / CN Inconsistencies

2017-02-14 20:17 - Monica Ihli

Status:	Closed	Start date:	2017-02-14
Priority:	Normal	Due date:	
Assignee:	Monica Ihli	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:			
Story Points:			
Description			
Inconsistencies exist between CN and MN. They include both orphaned objects on CN and failed syncs. There may be several things going on at once.			
Attached difference report was created on 2/13:			
* urn:node:CN: Has 637 objects that are not on urn:node:TDAR			
* urn:node:TDAR: Has 602 objects that are not on urn:node:CN			
Related issues:			
Related to Member Nodes - MNDeployment #6485: The Digital Archaeology Record ...		Operational	2015-04-30

History

#1 - 2017-02-14 20:17 - Monica Ihli

- Related to MNDeployment #6485: The Digital Archaeology Record (tDAR) added

#2 - 2017-02-15 01:09 - Monica Ihli

- File Need_to_Archive.txt added

- File Failed_Syncs_Due_to_Checksums.xlsx added

- File Originally_501_Error_Then_Attempted_Resync.txt added

Some Notes on Background & Technology:

* tDAR's repository is built on dSpace.

* nginx webserver on tomcat.

* tDAR is using webserver to proxy application server.

* tDAR repository does not natively keep versioned records. (If something changes, it's processed as an update to the original record rather than as a new version of the same record which obsoletes the old.)

* tDAR's DataONE code can be found at:

https://bitbucket.org/tDar/tDar_src/src/7936aeaf30192de5dc7160219577394abdeb5e7e/web/service/dataone/?at=default

* tDAR member node supports version 2 services.

* Some implementation history is documented at: <https://epad.dataone.org/pad/p/20150429-DigitalAntiquity>

* Note on dates: in this system the dateuploaded is by default the date system metadata modified.

Specific Priorities within the Context of the Larger Issue of Consistency:

1. *PIDs Needing Manual Archival* – Confirm the status of 37 PIDs that were part of the v1 and v2 testing process that got into production and need to be manually archived in production. These must be archived if it hasn't already been done. Confirm if these account for any of the discrepancies identified in the difference report. The list of 37 PIDs has been attached as Need_to_Archive.txt

2. *PIDS with Encoding Issues* – There were recently a few PIDs that had a character encoding issue. Adam has cleaned these up in repository and initiated resync on these. Should check and ensure that they have correctly synced now.

3. *PIDS with Failed Sync Due to Checksum Issue* – Identified a list of 80 PIDs from logs which failed to sync due to checksum mismatches. These are listed in the attached file Failed_Syncs_Due_to_Checksums.xlsx which was created on 2/14/2017. Adam believes that this may be due to an unaccounted for consequence of people making changes in their local repository. He is testing a fix for this which will essentially trigger obsoleting a changed record to avert the checksum issue, which may go into production as soon as tomorrow (2/15/2017). Two action items needed here are to (A) confirm if the new fix prevents future occurrences, and (B) ensure that the sync of the current version is attempted on these 80 and that the sync is successful.

4. *Verify Success of New Sync Attempt for Previously 501 Errors* – A list of 768 PIDs were attempted to resync using curl due to initial attempts resulting in a 501 error returned. Node operator is waiting on confirmation that these have been successfully resolved and no longer show up in the latest report of differences. The 768 PIDs are attached as Originally_501_Error_Then_Attempted_Resync.txt.

#3 - 2017-02-20 14:57 - Monica Ihli

20 Additional failed syncs between 2-16 and 2-17:

```
TDAR doi:10.6067:XCV81G0P79_format=d1rem1483726160322
TDAR doi:10.6067:XCV81G0P79_meta$v=1483726160322
TDAR doi:10.6067:XCV8DR2XJQ_format=d1rem1487225733953
TDAR doi:10.6067:XCV8DR2XJQ_meta$v=1487225733953
TDAR doi:10.6067:XCV8F47R69_format=d1rem1487039168917
TDAR doi:10.6067:XCV8F47R69_format=d1rem1487214148105
TDAR doi:10.6067:XCV8F47R69_format=d1rem1487214148105
TDAR doi:10.6067:XCV8F47R69_meta$v=1487039168917
TDAR doi:10.6067:XCV8F47R69_meta$v=1487214148105
TDAR doi:10.6067:XCV8F47R69_meta$v=1487214148105
TDAR doi:10.6067:XCV8FN183H_format=d1rem1468436505149
TDAR doi:10.6067:XCV8FN183H_meta$v=1468436505149
TDAR doi:10.6067:XCV8KW5HXXH_format=d1rem1478619831789
TDAR doi:10.6067:XCV8KW5HXXH_meta$v=1478619831789
TDAR doi:10.6067:XCV8PC34GH_format=d1rem1487225733773
TDAR doi:10.6067:XCV8PC34GH_meta$v=1487225733773
TDAR doi:10.6067:XCV8PV6N8S_format=d1rem1480081416608
TDAR doi:10.6067:XCV8PV6N8S_meta$v=1480081416608
TDAR doi:10.6067:XCV8V988Z0_format=d1rem1481636591352
TDAR doi:10.6067:XCV8V988Z0_meta$v=1481636591352
```

These seem to all have the same kinds of error messages recorded:

```
cn-synchronization.log.1:[ERROR] 2017-02-16 21:54:43,986 (V2TransferObjectTask:processTask:416)
Task-urn:node:TDAR-doi:10.6067:XCV8FN183H_format=d1rem1468436505149 - NotAuthorized to claim the seriesId - NotAuthorized - Submitter
does not have CHANGE rights on the SeriesId as determined by the current head of the Sid collection, whose pid is:
org.dataone.service.types.v1.Identifier@2cf9dd65
```

#4 - 2017-02-20 21:48 - Monica Ihli

- File *tDAR_example.pdf* added

Attached file *tDAR_example.pdf* shows the differences between MN and CN for a particular seriesId. The analysis shows that at this point, the relationships within the obsolescence chain have been correctly encoded within MN system metadata, and the current versions of system metadata need only be synchronized to the Coordinating Nodes. However, there appears to be a bug in how CN is processing seriesId in some cases. A bug report will be created and treated as a high priority. In this situation, the best we can do is to keep track of affected PIDs (and probably their chains as a whole) in order to manually resync when the issue is resolved.

#5 - 2017-03-02 21:20 - Monica Ihli

- % Done changed from 0 to 30
- Status changed from New to In Progress

Patch 2.3.1 appears to have resolved the submitter-related not authorized problem which interfered with syncing in some cases. Moving forward, a list of 637 PIDs which are orphaned on the CN node, per the latest difference report run on 2-28-17, has been provided to the node operator for obsolete. The next step after that will be to re-attempt synchronization on any PIDs which previously failed due to the issue fixed in 2.3.1.

The node operator Adam will be traveling most of the month of March so will have limited availability during this time. He wishes to proceed cautiously in dealing with the remaining objects which have not synchronized to the CN from his MN yet.

#6 - 2017-03-31 12:53 - Monica Ihli

We are presently attempting to address the 600ish objects that are orphaned on the CN. The quandary lies in that a node supporting v2 services is expected to call archive method on the object living on the MN, after which the updates will be synchronized to the CN. This is not possibly in a situation where objects no longer exist on the MN.

The proposed resolution is to temporarily disable v2 services on the MN in favor of v1. The CN is expected to then recognize the node as v1, which will permit the MN to utilize archive method in the v1 context-- meaning that archive will be permitted to be called on the CN version of the data.

#7 - 2017-04-10 19:57 - Monica Ihli

Orphaned nodes were successfully archived on the CN. The tDAR MN has reverted its node registration back to v2.

#8 - 2017-08-23 16:04 - Monica Ihli

- Status changed from In Progress to Closed
- % Done changed from 30 to 100

Files			
tDAR-diff-2017-02-13.txt	161 KB	2017-02-14	Monica Ihli
Originally_501_Error_Then_Attempted_Resync.txt	35.1 KB	2017-02-15	Monica Ihli
Failed_Syncs_Due_to_Checksums.xlsx	10.6 KB	2017-02-15	Monica Ihli
Need_to_Archive.txt	1.03 KB	2017-02-15	Monica Ihli
tDAR_example.pdf	89.1 KB	2017-02-20	Monica Ihli