

## Member Nodes - Task #7348

MNDeployment # 3238 (Operational): Idaho Northwest Knowledge Network member node

### NKN resource map indexing problem

2015-09-14 18:53 - Dave Vieglais

<b>Status:</b>	Closed	<b>Start date:</b>	2015-09-14
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Mark Servilla	<b>% Done:</b>	100%
<b>Category:</b>		<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	Operational		
<b>Story Points:</b>			

#### Description

NKN is still having issues with correctly building the ORE resource map: apparently, the ORE builder in GMN requires objects to be listed in sequence, beginning with the science metadata object; method signature: `def simple_generate_resource_map(self, resource_map_pid, science_metadata_pid, science_data_pids)`

Skye said that one data object (metadata) failed, so the whole package wouldn't index. Re-indexing with a new ORE resulted in the full package being available on cn-sandbox. NKN was going to test additional content.

--- UPDATE --- latest from Ed is that he's seeing a lot of error traffic on his sandbox node; he's investigating.

Ed (NKN) - latest data package harvested correctly but it hasn't indexed yet. Timing of the harvest/index process is driven by the synchronization schedule identified in the node services registration document:

<https://cn-sandbox.test.dataone.org/cn/v1/node/urn:node:mnTestNKN> (every three minutes)

Here's the identifier for this object: 11adf019-187d-4587-a6ca-0a991fee7cb8

See also: <https://dataone-dev.nkn.uidaho.edu/mn/v1/object/11adf019-187d-4587-a6ca-0a991fee7cb8>

At this time (2:14pm), Dave can see the sysmeta and data (2) on the MN but NOT on the CN. This means that, for some reason, this content hasn't synchronized to the CNs yet. Dave looking through the logs... CN synch log indicates that the item was placed in the queue but no further action.

From `/var/log/dataone/synchronize/cn-synchronization.log`

```
[DEBUG] 2015-09-03 17:39:00,337 (ObjectListHarvestTask:call:188) placed on hzSyncObjectQueue-urn:node:mnTestNKN:71c7693e-8050-4f4c-a28f-69fe0856efec
[DEBUG] 2015-09-03 17:39:00,337 (ObjectListHarvestTask:call:188) placed on hzSyncObjectQueue-urn:node:mnTestNKN:11006067-27b6-438a-b6f4-d50982fcb6fc
[DEBUG] 2015-09-03 17:39:00,338 (ObjectListHarvestTask:call:188) placed on hzSyncObjectQueue-urn:node:mnTestNKN:8e8d9e52-c6cd-4daf-b893-e226cba377f7
[DEBUG] 2015-09-03 17:39:00,339 (ObjectListHarvestTask:call:188) placed on hzSyncObjectQueue-urn:node:mnTestNKN:ef4beeb2-2b34-4703-9d68-782f99c02cd6
[DEBUG] 2015-09-03 17:39:00,339 (ObjectListHarvestTask:call:188) placed on hzSyncObjectQueue-urn:node:mnTestNKN:0d82882b-a8b4-44e8-9985-63edacfadcf3
[DEBUG] 2015-09-03 17:39:00,340 (ObjectListHarvestTask:call:188) placed on hzSyncObjectQueue-urn:node:mnTestNKN:11adf019-187d-4587-a6ca-0a991fee7cb8
```

Check these PIDs (bash script follows):

```
P=(71c7693e-8050-4f4c-a28f-69fe0856efec 11006067-27b6-438a-b6f4-d50982fcb6fc 8e8d9e52-c6cd-4daf-b893-e226cba377f7 ef4beeb2-2b34-4703-9d68-782f99c02cd6 0d82882b-a8b4-44e8-9985-63edacfadcf3 11adf019-187d-4587-a6ca-0a991fee7cb8)
export NODE="https://dataone-dev.nkn.uidaho.edu/mn"
for p in ${P[@]}; do echo ${p}; d1sysmeta "${p}"; done
```

```
export NODE="https://cn-sandbox.test.dataone.org/cn"
for p in ${P[@]}; do echo ${p}; d1sysmeta "${p}"; done
#
```

Systemeta on MN ( <https://dataone-dev.nkn.uidaho.edu/mn> )? All OK

Systemeta on CN ( <https://cn-sandbox.test.dataone.org/cn> )? All failed (404 error) (approx 14:45 ET) All OK at approx 15:00ET

Indexed on CN?

71c7693e-8050-4f4c-a28f-69fe0856efec ( FGDC-STD-001-1998 metadata)

11adf019-187d-4587-a6ca-0a991fee7cb8 (<http://www.openarchives.org/ore/terms> resource map)

Not in index, others are OK.

Seems like the issue is unresponsiveness of a Member Node mn-sandbox-ucsb-2.test.dataone.org. Connections to this node hang for a very long time and seem to be causing a slowdown in the CN processing. This *may* be why indexing has not processed the metadata or resource map yet. Restarting the MN service on mn-sandbox-ucsb-2 to see if that helps. (Note: had to reboot node)

The basic problem appears to be:

[ INFO] 2015-09-03 19:39:21,372 (IndexTaskProcessor:isObjectPathReady:263) Object path for pid:

71c7693e-8050-4f4c-a28f-69fe0856efec is not available. Task will be retried.

[ INFO] 2015-09-03 19:39:21,377 (IndexTaskProcessor:getNextIndexTask:164) Task for pid:

71c7693e-8050-4f4c-a28f-69fe0856efec not processed.

which indicates the science metadata (fgdc) is not available on the filesystem. So... failed to pass validation with Metacat??

## History

---

### #1 - 2015-10-09 07:30 - Laura Moyers

- Status changed from *New* to *In Progress*

- % Done changed from 0 to 30

Ed has installed the (intermediate) GMN update which seems to have resolved the replication queue issue (trying to start too many jobs). We think NKN's "indexing" issue was really the GMN replication queue issue.

### #2 - 2015-10-28 17:16 - Laura Moyers

- % Done changed from 30 to 50

- Status changed from *In Progress* to *Testing*

### #3 - 2015-10-28 17:16 - Laura Moyers

- % Done changed from 50 to 80

- Status changed from *Testing* to *In Review*

### #4 - 2015-10-28 17:17 - Laura Moyers

- Status changed from *In Review* to *Closed*

- translation missing: *en.field\_remaining\_hours* set to 0.0

- % Done changed from 80 to 100