

CN REST - Story #8582

Story # 8756 (New): Ensure replica auditor is effective

Replica Auditing service is throwing errors

2018-05-01 19:15 - Chris Jones

<b>Status:</b>	New	<b>Start date:</b>	2018-05-01
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Jing Tao	<b>% Done:</b>	0%
<b>Category:</b>	d1_replication_auditor	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>			
<b>Story Points:</b>			
<b>Description</b>			
Replica auditing should be auditing objects every 90 days for fixity, and setting the replicaStatus appropriately. The /var/log/dataone/cn-replication-audit.log* files are showing many errors:			
<pre>cjones@cn-ucsb-1:replicate\$ grep ERROR cn-replication-audit.log*   grep "Cannot update replica status"   wc -l 437601</pre>			
Determine if this is a configuration issue or a code issue and fix it as needed. Also, fix the code to call Identifier.getValue() when logging these errors to avoid printing the memory location of the object like org.dataone.service.types.v1.Identifier@7e90f2e8. There are multiple places where getValue() needs to be added.			

History

#1 - 2018-05-01 20:33 - Rob Nahf

additional logging information is in cn-replication.log (from ReplicationService, which performs setting the INVALID status).

```
cn-replication.log.1-[DEBUG] 2018-04-27 21:41:55,946 [pool-11-thread-3] (ReplicationService:updateReplication
Metadata:3
58) org.dataone.service.exceptions.NotFound: No record found for: http://dx.doi.org/10.5061/dryad.64 : null
cn-replication.log.1-[ERROR] 2018-04-27 21:41:55,946 [pool-11-thread-3] (ReplicationService:updateReplication
Metadata:3
60) Error in calling updateReplicationMetadata(): No record found for: http://dx.doi.org/10.5061/dryad.64 : nu
ll
cn-replication.log.1:[ERROR] 2018-04-27 21:41:55,946 [pool-11-thread-3] (ReplicaAuditingDelegate:updateInvali
dReplica:1
10) Cannot update replica status to INVALID for pid: org.dataone.service.types.v1.Identifier@37cca400 on MN: u
rn:node:DR
YAD
```

The NotFound comes from either cn.getSystemMetadata or cn.updateReplicationMetadata.

Maybe it has something to do with the MemberNode of the Replica being updated?

## #2 - 2018-05-01 21:04 - Rob Nahf

cn.updateReplicationMetadata in Metacat returns a string formatted like what's in the logs, and it looks like it does it for any issue with the Hazelcast.getInstance().getSystemMetadataMap().get() call. The method returns a NotFound, but both HZ client errors (resulting in .getSystemMetadataMap() returning null) and nulls returned from the call (indicating a true NotFound) are handled by catching NPEs, and returned as NotFound.

It would be better to split out these two root causes for throwing an exception from the method.

```
} catch (RuntimeException e) { // Catch is generic since HZ throws RuntimeException
    throw new NotFound("4854", "No record found for: " + pid.getValue() +
        " : " + e.getMessage());
}
```

## #3 - 2018-05-01 21:10 - Rob Nahf

In this case, the NotFound is a true not found, because calling /meta from a browser returns a not found. Are we trying to audit pids that don't exist anymore? <http://dx.doi.org/10.5061/dryad.64>, for example?

## #4 - 2019-01-12 20:03 - Chris Jones

Well, somewhere in the CN processing pipeline, we are truncating the pid values that have special characters like ?. So, the NotFound errors are due to us not looking for the correct pid string.

## #5 - 2019-01-12 23:22 - Chris Jones

After looking at the auditing code, the results of each audit task are being written to the /var/log/dataone/replicate/cn-replication-audit-splunk.log files. These files are nice in that they uniformly document the AuditLogEntry results. One thing to note is that every log entry is an ERROR entry (despite most being successful), so pretty much ignore that status.

The general logging debug information is stored in the /var/log/dataone/replicate/cn-replication-audit.log files. Note that the splunk logs are rotated out frequently, and Splunk is no longer persisting them, so we should consider creating file beat forwarders and persisting these entries in Elastic Search on log-proc-ucsb-1.dataone.org to help us understand the actual frequency of replica auditing over time.

## #6 - 2019-01-14 16:40 - Chris Jones

- Parent task set to #8756