Infrastructure - Story #7716

How to facilitate resubmission of sync Failures?

2016-04-12 18:51 - Jing Tao

Status:	In Progress		**Start date:**	2016-05-17
Priority:	Normal		**Due date:**	
Assignee:	Dave Vieglais		**% Done:**	30%
Category:	d1_synchronization		**Estimated time:**	0.00 hour
Target version:	CCI-2.4.0			
Story Points:				

Description

Currently, when CNs harvest objects from MNs, they only compare the last modified time stamp on the system metadata with the last harvest date. If the last modified stamp is earlier than the last harvest date, the object wouldn't be put on the synchronization queue.

This causes an issue: if the synchronization for this object failed in the previous synchornizations and the issue of the failure was corrected (e.g. a misconfiguration on the MN and the last modified time stamp should NOT be changed), the harvest will not pick up the object. We have to manually to either reset the harvest date, or submit a v2.CN.synchronize(PID,NodeID) in order to pick up the object.

MemberNode operators are not well versed in how to submit these requests, so some sort of monitoring and management tool for MemberNode operators (and DataONE administrators) is proposed to simplify this type of action.

We are thinking to persist the information for the failed objects on a database table -e.g. failedSynchronization and it contains: identifier, mn_id, last_failed_time, message.

Message can be separated to another table to use the foreign key.

Once we have the failure information, we can work on the monitoring and management tool.

Subtasks:

Task # 7813: create sync task log outcome messages		**Closed**
Task # 7814: create queue state log events		**Closed**
Story # 8038: connect logging output to a log analysis tool		**In Progress**

Related issues:

Related to Infrastructure - Bug #7955: Synchronization should check dateSysMe...	**Closed**	**2016-12-21**

History

#1 - 2016-04-12 20:18 - Rob Nahf

- Category set to d1_synchronization

- Description updated

- Subject changed from Tracking objects that failed during synchronization to Addressing synchronization failures

I also propose adding the CN nodeId for when we redistribute synchronization to all CNs.

We will also need record of resolution so that MN operators can determine how many (and which) objects are still unsynchronized.

The format of the table should be actionable by any automated auditing process, and be able to be analyzed by a tool like Kibana (https://www.elastic.co/products/kibana). That's to say, it might best be log-oriented.

#2 - 2016-05-06 22:40 - Rob Nahf

it is somewhat likely that synchronization will transition into a multiple queue-based processing system, in which case, sticking synchronizationFailed items on a separate queue would make sense, and leave open the opportunity for daily processing of failed items.

A system like that would need to make sure that TransferObjectTask removes items from the syncFailed queue upon the starting to sync it. (an item could be on both queues).

#3 - 2016-12-21 09:09 - Dave Vieglais

- Related to Bug #7955: Synchronization should check dateSysMetaModified when adding tasks to processing queue added

#4 - 2016-12-21 09:09 - Dave Vieglais

- Target version changed from CCI-2.3.0 to CCI-2.3.2

#5 - 2017-03-03 23:48 - Rob Nahf

- Subject changed from Addressing synchronization failures to How to facilitate resubmission of sync Failures?

#6 - 2017-03-04 06:55 - Rob Nahf

- % Done changed from 0 to 30

- Status changed from New to In Progress

Implemented the logging portion, but do we still need a re-submission tool?

#7 - 2017-03-07 20:40 - Rob Nahf

- Target version changed from CCI-2.3.2 to CCI-2.4.0

#8 - 2018-01-17 19:22 - Dave Vieglais

- Assignee changed from Rob Nahf to Dave Vieglais

#9 - 2018-01-17 19:22 - Dave Vieglais

- Sprint set to Infrastructure backlog