

## Infrastructure - Feature #4038

### change default replication implementation

2013-10-04 06:26 - Matthew Jones

<b>Status:</b>	New	<b>Start date:</b>	
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Matthew Jones	<b>% Done:</b>	0%
<b>Category:</b>	d1_replication	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>		<b>Story Points:</b>	
<b>Milestone:</b>	CCI-2.0		
<b>Product Version:</b>	*		

#### Description

For objects less than a threshold size, change the default number of replicas for objects to two (instead of the current 0) within the replication management system on CNs.

During today's MN Forum call, the MN operators discussed the impact of the government shutdown on data availability and noted that DataONE's replication system could effectively improve data availability in the midst of massive service shutdowns like we have seen this week. We discussed the pros and cons of changing the default replication policy so that, in the absence of an explicit ReplicationPolicy for an object, DataONE Coordinating Nodes would create two replicas of any object that was smaller than a certain threshold size, probably around 500 MB today. This threshold would increase over time as network bandwidth improved, so should be configurable on the CN without code changes. This would ensure that the replication capabilities of DataONE are more fully utilized, while keeping the control of replication firmly in the hands of Member Nodes (if they don't want replicas, they simply set replicationAllowed=false in the replication policy in system metadata).

I have created a design document that describes our envisioned system here:

<http://mule1.dataone.org/ArchitectureDocs-current/design/ReplicationOverview.html>

#### History

##### #1 - 2013-10-04 06:29 - Matthew Jones

- Description updated

##### #2 - 2013-10-04 11:50 - Dave Vieglais

My preference would be to make no changes and use a tool to assist MNs with updating the replication policy on their content through the CNReplication.setReplicationPolicy method. If it doesn't do so already, the CLI would seem to be the right tool for a generic implementation.

##### #3 - 2014-01-06 18:16 - Dave Vieglais

- Due date changed from 2013-10-05 to 2014-02-01

- Target version changed from 2013.39-Block.5.3 to 2014.4-Block.1.2

##### #4 - 2014-03-14 18:09 - Chris Jones

- Target version changed from 2014.4-Block.1.2 to 2014.14-Block.2.3

- Due date changed from 2014-02-01 to 2014-04-12

- Assignee changed from Chris Jones to Matthew Jones

This policy decision needs consensus from MN operators. Reassigning to Matt for now for discussion at the CCIT 2014 meeting.

##### #5 - 2014-06-26 22:04 - Robert Waltz

- Due date deleted (2014-04-12)

- Target version deleted (2014.14-Block.2.3)

- Start date deleted (2013-10-04)

- Milestone changed from None to CCI-2.0

**#6 - 2014-10-01 21:24 - Dave Vieglais**

- Due date set to 2014-10-01
- Target version set to Maintenance Backlog
- Start date set to 2014-10-01

**#7 - 2014-10-01 21:26 - Dave Vieglais**

- Start date deleted (2014-10-01)
- Target version deleted (Maintenance Backlog)
- Due date deleted (2014-10-01)