

Infrastructure - Task #2624

Story # 2622 (Closed): d1_replication should prioritize MN replication tasks based on load, failures, and bandwidth factors

Modify ReplicationManager.createAndQueueTasks() to limit replication tasks based on current MN load

2012-04-20 23:57 - Chris Jones

Status:	Closed	Start date:	2012-04-20
Priority:	Normal	Due date:	
Assignee:	Chris Jones	% Done:	100%
Category:	d1_replication	Estimated time:	0.00 hour
Target version:	Sprint-2012.19-Block.3.2	Story Points:	
Milestone:	CCI-1.0.0		
Product Version:	*		

Description

As a first pass on trying to throttle replicate() calls to possibly bogged-down MNs, change createAndQueueTasks() such that we evaluate how many currently pending tasks there are for a given node, and cap the calls based on a configurable limit.

See the epad for algorithm details:

<http://epad.dataone.org/20120420-replication-priority-queue>

History

#1 - 2012-05-07 16:57 - Chris Jones

- Status changed from New to In Progress

Added the prioritizeNodes() method to implement the new prioritization algorithms. Needs testing.

#2 - 2012-05-21 16:21 - Chris Jones

- Status changed from In Progress to Closed

Added the getRequestFactors(), getFailureFactors(), and getBandwidthFactors() methods that feed into prioritizeNodes(). Skye moved these methods into a new ReplicationPrioritizationStrategy class for modularity and independent testing. Testing showed that the prioritization scheme works fine now that we give a 5 try grace period for MNs before calculating scores.