

Metacat - Bug #5199

OAI-PMH: Improve memory management of data provider catalog metadata

10/11/2010 09:16 AM - Duane Costa

Status:	New	Start date:	10/11/2010
Priority:	Normal	Due date:	
Assignee:	Chad Berkley	% Done:	0%
Category:	metacat	Estimated time:	0.00 hour
Target version:	2.x.y	Spent time:	0.00 hour
Bugzilla-Id:	5199		

Description

On October 6, 2010, Marco Fahmi wrote:

What scares me about this code is that it stores the whole metadata catalog in memory (static member fields docTypeMap and dateMap) at class-load time, which is on the first call to the OAIHandler servlet.

So that check for the refreshDate is checking whether the entire catalog needs to be reloaded into memory! I can't see that scaling to any extent.

On October 11, 2010, Duane Costa and Mark Servilla wrote in a reply to Marco Fahmi:

With regard to the memory management issue -- the OAI-PMH data provider code has been tested on a repository of 375 EML documents with no apparent problem regarding memory. We will soon be applying the code to a much larger repository (~10,000 documents) so we'll have a better sense of whether the current implementation scales. One distinction we'd like to make is that when you state that "the entire catalog" is loaded into memory, please note that only the documents' identifiers, their EML versions, and their revision dates are loaded into memory; the full contents of the documents themselves are not stored in memory but are instead retrieved from the Metacat database only as needed. In any case, we agree that it's likely that a better memory management solution should ultimately be implemented for storing the subset of catalog metadata that is currently held in memory.

History

#1 - 03/27/2013 02:29 PM - Redmine Admin

Original Bugzilla ID was 5199