

Utilities - Bug #324

perl implementation of harvester api

11/08/2001 11:34 AM - Matt Jones

Status:	In Progress	Start date:	11/08/2001
Priority:	Immediate	Due date:	
Assignee:	David Blankman	% Done:	0%
Category:	Harvester	Estimated time:	0.00 hour
Target version:	---	Spent time:	0.00 hour
Bugzilla-Id:	324		
Description			
Need to implement registry and harvester.			
Related issues:			
Blocked by Metacat - Bug #162: need harvest/batch load for metacat		Resolved	10/24/2000
Blocked by Metacat - Bug #325: create site filters to convert site metadata t...		Resolved	11/08/2001

History

#1 - 03/06/2002 12:16 PM - James Brunt

Matt Jones and David Blankman can provide the start of the requirements for this program based on discussions we had at the October meeting and a flow diagram of communication with metacat that we produced. This development should be done in close communication with Matt and the other KNB developers. There are also some packaging related problems that will need to be addressed but alas that is another bug.

#2 - 04/16/2002 11:02 AM - Owen Eddins

April 11 David, Owen and James had an impromptu Harvester meeting and decided that for purposes of 'Harvesting' .xml files that the web service architecture did not make any sense. To much overhead to try and manage arbitrary files on a file system. What we were trying to do with the current design was create a lightweight dbms with out providing any controls on how files got in or out of the 'dbms'. What we decided was we needed a one time metacat loader tool to upload in to Metacat a .xml file. See cvs pubs for old design of Harvester. A new design doc will be placed in pubs directory with metacat loader design for further discussing. It was decided that all the ideas we had been discussing for the web service were appropriate but only for harvesting metadata out of RDBMS.

#3 - 08/27/2002 08:26 AM - Owen Eddins

Just an update. Currently looking in to Morpho classes and how to adapt them to create a generalize client interface to Metacat. From this we're going to create a command line tool for one time uploading of .xml files and datafiles in to metacat.

#4 - 08/27/2002 08:37 AM - Matt Jones

Excellent. You might also consider looking at the Metacat.pm perl module that I wrote. It is a perl client interface to metacat, and I think it is the easiest way to access the metacat system from the command line. I wrote it to use in the NRS and OBFS data registries, which use perl as a backend to their HTML form submission pages. Just a thought. You can find it in the webmdentry module on cvs.nceas.ucsb.edu (I think the path is webmdentry/src/Metacat). There are some examples of its use in there as well (load-dataset.cgi and load-dataset.pl).

#5 - 08/27/2002 05:26 PM - Owen Eddins

Matt,

Thanks for responding. Yes, I looked in to using your perl scripts to do this and I agree that they probably are a quicker solution for a command line tool. It's essentially written already for pre beta 9 EML. I downloaded it and installed it and got it working easily. The problem is that we also are working on creating an automated data replication engine (essentially the

Harvester with a cooler name) and a web service interface to the Metacat that implements a WSDL I have hacked up to define a very light weight 'API' for EML. I spent some time last week looking at the Morpho classes and they look like they can lend themselves easily to being turned in to a generic client side set of classes. Since we are looking in to other methods of connecting to Metacat other than Morpho generalizing these classes seems a worth while effort. The ClientFramework class is the big one but other than that most of the others seem pretty general. Most of the methods in MetacatDataStore seem relevant to the 'API' I'm thinking of. What do you think?

#6 - 03/27/2013 02:13 PM - Redmine Admin

Original Bugzilla ID was 324