

Metacat - Bug #101

generate data set usage metadata/ provide access log

08/24/2000 10:27 AM - Matt Jones

Status:	Resolved	Start date:	08/24/2000
Priority:	Immediate	Due date:	
Assignee:	Matt Jones	% Done:	0%
Category:	metacat	Estimated time:	0.00 hour
Target version:	1.4	Spent time:	0.00 hour
Bugzilla-Id:	101		
Description			
<p>Tracking data set usage is an important part of running a metadata and data archive. We need a new feature in the metacat server that tracks usage statistics for each data-set and creates metadata records that document the usage of data sets. This information is currently encoded in the "data_request", "review_history", and "remarks" fields of eml-supplement.dtd. These fields need to move forward as we revise these standards, and the metacat servlet needs to automatically generate new records of this type as needed for a data set, as it is accessed. We'll need to decide whether simply viewing the metadata should be tracked, or whether a download is required to trigger a log entry.</p>			
Related issues:			
Is duplicate of Morpho - Bug #1104: REOT feedback: attach strings to "public"...		Resolved	06/06/2003

History

#1 - 01/23/2002 09:03 AM - Matt Jones

This information needs to be accessible to the data set owner (and possibly others) through the Morpho interface and other interfaces.

#2 - 02/12/2002 04:03 PM - Matt Jones

Moved issues that we do not intend to address during the current release to the 'Postpone' milestone.

#3 - 06/06/2003 04:21 PM - Matt Jones

- Bug 1104 has been marked as a duplicate of this bug. ***

#4 - 06/06/2003 04:23 PM - Matt Jones

Increasing priority and milestone because this issue has resurfaced a lot recently, as evidenced by the recent duplicate bug [#1104](#) that was filed.

#5 - 04/02/2004 01:53 PM - Matt Jones

Basic usage tracking is now implemented. Still need to test on oracle, but basic functionality is working on postgres. Every time the user does one of the following actions, it is logged in the access_log table, along with their uid, ip number, the docid, and the date. I have eye to build a reporting function for this, but it shouldn't be too difficult. The one problem I foresee is this database growing too large too quickly, so one item might be a mechanism for archiving the access_log table outside of the database on a regular basis. But we'll see if that actually becomes an issue.

Still to do:

- 1) query/report functionality
- 2) make logging asynchronous to decrease the number of jdbc immediate calls
- 3) possibly write code to archive access_log data periodically
- 4) test on oracle

The access_log table records look like this, FYI:

```
entryid | ip_address | principal |
```

docid	event	date_logged
5	public	
knb-lter-gce.180.1	read	2004-04-02 12:31:46
6	public	
obfs.45345.3	read	2004-04-02 12:32:19
7	public	
obfs.45345.3	read	2004-04-02 12:32:36
8	uid=jones,o=NCEAS,dc=ecoinformatics,dc=org	
obfs.45345.4	update	2004-04-02 12:33:41
9	uid=jones,o=NCEAS,dc=ecoinformatics,dc=org	
obfs.45345.4	read	2004-04-02 12:35:05
10	uid=jones,o=NCEAS,dc=ecoinformatics,dc=org	
obfs.45345.4	delete	2004-04-02 12:35:06
11	public	
obfs.45346.1	read	2004-04-02 12:36:23
12	uid=jones,o=NCEAS,dc=ecoinformatics,dc=org	
obfs.45350.1	insert	2004-04-02 12:39:24
13	uid=jones,o=NCEAS,dc=ecoinformatics,dc=org	
knb-lter-gce.180.1	read	2004-04-02 12:45:08

#6 - 09/21/2004 03:49 AM - Matt Jones

The query/report functionality (1) is now complete. Use the 'getlog' servlet action to get an XML report of the events if you are in the administrators group.

The function has been tested on oracle and is working so (4) is complete.

Have decided to indefinitely postpone asynchronous logging (2) and archiving of log data (3) as its not clear that they are needed. Closing the bug as all features are now complete.

#7 - 03/27/2013 02:13 PM - Redmine Admin

Original Bugzilla ID was 101