

Metacat - Bug #4637

Metacat Harvester fails to catch some insert and update failures

12/17/2009 11:34 AM - Duane Costa

Status:	Resolved	Start date:	12/17/2009
Priority:	Normal	Due date:	
Assignee:	Duane Costa	% Done:	0%
Category:	harvester	Estimated time:	0.00 hour
Target version:	1.9.2	Spent time:	0.00 hour
Bugzilla-Id:	4637		

Description

Metacat Harvester is not catching all insert and update errors.

Recently at LTER, there have been a handful of documents that have been reported by Metacat Harvester as successful inserts or updates, but in fact the documents are not being successfully inserted or updated into Metacat.

The logical error is in method `HarvesterDocument.putMetacatDocument()`. The problem is that Harvester treats the absence of an exception as a success condition, when it should instead require hard confirmation of success from the Metacat client that the insert or update operation succeeded:

```
if (harvester.connectToMetacat()) {
    try {
        if (insert) {
            metacatReturnString = metacat.insert(docidFull, stringReader, null);
            inserted = true;
            harvester.addLogEntry(0, docidFull + " : " + metacatReturnString,
                "harvester.InsertDocSuccess",
                harvestSiteSchedule.siteScheduleID,
                null, "");
        }
        else if (update) {
            metacatReturnString = metacat.update(docidFull, stringReader, null);
            updated = true;
            harvester.addLogEntry(0, docidFull + " : " + metacatReturnString,
                "harvester.UpdateDocSuccess",
                harvestSiteSchedule.siteScheduleID,
                null, "");
        }
    }
    catch (MetacatInaccessibleException e) {
        logMetacatError(insert, metacatReturnString,
            "MetacatInaccessibleException", e);
    }
    catch (InsufficientKarmaException e) {
        logMetacatError(insert, metacatReturnString,
            "InsufficientKarmaException", e);
    }
    catch (MetacatException e) {
        logMetacatError(insert, metacatReturnString, "MetacatException", e);
    }
    catch (IOException e) {
        logMetacatError(insert, metacatReturnString, "IOException", e);
    }
}
```

Harvester does not check the value of the string returned by Metacat ('metacatReturnString' in the above code). In the cases where the insert/update operations have been failing, the return string is empty or null. Harvester should examine the return string to confirm that it contains the substring "<success>" or something similar.

The fact that no exception is thrown by Metacat could point to an additional problem in Metacat, since the insert/update operation completes without raising an exception even though the document is not inserted or updated. The documents that appear to trigger this condition are unusually large EML documents (currently there are three documents from CDR and one document from LUQ that trigger this bug).

After the Harvester bug is resolved, or as part of resolving it, further investigation should be done to determine whether there is also a Metacat bug involved here, and if there is, a separate bug entry should be entered for it.

History

#1 - 12/21/2009 12:14 PM - Duane Costa

Resolved with the following update:

Author: costa

Date: 2009-12-21 12:10:29 -0800 (Mon, 21 Dec 2009)

New Revision: 5169

Modified:

trunk/src/edu/ucsb/nceas/metacat/harvesterClient/HarvestDocument.java

Log:

Fix for Bug [#4637](http://bugzilla.ecoinformatics.org/show_bug.cgi?id=4637) [http://bugzilla.ecoinformatics.org/show_bug.cgi?id=4637] - Metacat Harvester fails to catch some insert and update failures. As per comments in the bug entry, the Metacat Harvester logic has been modified to examine the Metacat client return string to confirm that it contains the substring "<success>" following an insert or update operation. It no longer considers just the absence of an exception as indicative of a successful operation.

#2 - 03/27/2013 02:27 PM - Redmine Admin

Original Bugzilla ID was 4637