

## Metacat - Bug #2060

### Documents not indexed because of error generated during indexing of documents

04/04/2005 05:20 PM - Saurabh Garg

<b>Status:</b>	Resolved	<b>Start date:</b>	04/04/2005
<b>Priority:</b>	Immediate	<b>Due date:</b>	
<b>Assignee:</b>	Saurabh Garg	<b>% Done:</b>	0%
<b>Category:</b>	metacat	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	1.6	<b>Spent time:</b>	0.00 hour
<b>Bugzilla-Id:</b>	2060		

#### Description

Sometimes given below errors have been seen in the log files. These are generated because of lack of coordination between the thread updating the xml\_nodes and the thread updating the xml\_index table. The result is that sometimes you have documents which are not indexed.

MetaCat: Error in DBSAXHandler.checkDocumentTable Couldn't find the docid for index build in reseasonable time!

MetaCat: SQL Exception while inserting path index in DocumentImpl.buildIndex for document test.200594171957

MetaCat: ORA-02291: integrity constraint (SGARG.XML\_INDEX\_DOCID\_FK) violated - parent key not found

java.sql.SQLException: ORA-02291: integrity constraint (SGARG.XML\_INDEX\_DOCID\_FK) violated - parent key not found

```
at oracle.jdbc.driver.DatabaseError.throwSQLException
(DatabaseError.java:125)
    at oracle.jdbc.driver.T4CTTIoer.processError(T4CTTIoer.java:305)
    at oracle.jdbc.driver.T4CTTIoer.processError(T4CTTIoer.java:272)
    at oracle.jdbc.driver.T4C8Oall.receive(T4C8Oall.java:623)
    at oracle.jdbc.driver.T4CPreparedStatement.doOall8
(T4CPreparedStatement.java:181)
    at oracle.jdbc.driver.T4CPreparedStatement.execute_for_rows
(T4CPreparedStatement.java:543)
    at oracle.jdbc.driver.OracleStatement.doExecuteWithTimeout
(OracleStatement.java:1028)
    at oracle.jdbc.driver.OraclePreparedStatement.executeInternal
(OraclePreparedStatement.java:2888)
    at oracle.jdbc.driver.OraclePreparedStatement.executeUpdate
(OraclePreparedStatement.java:2960)
    at edu.ucsb.nceas.metacat.DocumentImpl.updateNodeIndex
(DocumentImpl.java:1345)
    at edu.ucsb.nceas.metacat.DocumentImpl.buildIndex
(DocumentImpl.java:1214)
    at edu.ucsb.nceas.metacat.DBsAXHandler.run(DBsAXHandler.java:444)
    at java.lang.Thread.run(Thread.java:534)
```

MetaCat: SQL Exception while inserting path index in DocumentImpl.buildIndex for document test.20059417224

MetaCat: ORA-00001: unique constraint (SGARG.XML\_INDEX\_PK) violated

java.sql.SQLException: ORA-00001: unique constraint (SGARG.XML\_INDEX\_PK) violated

```
at oracle.jdbc.driver.DatabaseError.throwSQLException
(DatabaseError.java:125)
    at oracle.jdbc.driver.T4CTTIoer.processError(T4CTTIoer.java:305)
    at oracle.jdbc.driver.T4CTTIoer.processError(T4CTTIoer.java:272)
    at oracle.jdbc.driver.T4C8Oall.receive(T4C8Oall.java:623)
    at oracle.jdbc.driver.T4CPreparedStatement.doOall8
(T4CPreparedStatement.java:181)
```

```
    at oracle.jdbc.driver.T4CPreparedStatement.execute_for_rows
(T4CPreparedStatement.java:543)
    at oracle.jdbc.driver.OracleStatement.doExecuteWithTimeout
(OracleStatement.java:1028)
    at oracle.jdbc.driver.OraclePreparedStatement.executeInternal
(OraclePreparedStatement.java:2888)
    at oracle.jdbc.driver.OraclePreparedStatement.executeUpdate
(OraclePreparedStatement.java:2960)
    at edu.ucsb.nceas.metacat.DocumentImpl.updateNodeIndex
(DocumentImpl.java:1345)
    at edu.ucsb.nceas.metacat.DocumentImpl.buildIndex
(DocumentImpl.java:1214)
    at edu.ucsb.nceas.metacat.DB SAXHandler.run(DB SAXHandler.java:444)
    at java.lang.Thread.run(Thread.java:534)
```

## History

---

### #1 - 07/25/2005 12:59 PM - Saurabh Garg

Changin severity of teh bug to Critical because with the new changes if the indexing thread fails, the paths will not be indexed in xml\_path\_index. This means the document wont show up in the search for specified paths.

### #2 - 09/08/2005 04:09 PM - Saurabh Garg

Fixed.

Moved the call to starting of indexing thread from endDocument to DocumentImpl after commit has been done. This way when ever a document is indexed it has already been entered in xml\_nodes and xml\_documents.

Closing the bug.

### #3 - 10/10/2005 01:25 PM - Saurabh Garg

Still getting the same error...

```
Metacat: [ERROR]: Error in DB SAXHandler.checkDocumentTable Couldn't find the
docid for index build in reseasonable time! [edu.ucsb.nceas.metacat.DB SAXHandler]
Metacat: [ERROR]: SQL Exception while inserting path index in
DocumentImpl.buildIndex for document sgarg.1130
[edu.ucsb.nceas.metacat.DocumentImpl]
Metacat: [ERROR]: ERROR: insert or update on table "xml_index" violates foreign
key constraint "xml_index_nodeid_fk" [edu.ucsb.nceas.metacat.DocumentImpl]
```

### #4 - 10/28/2005 12:20 PM - Saurabh Garg

Increasing severity to Blocking. This is after the update on KNB. The build index thread is failing on a more regular basis. Maybe because of the size of the current knb database.

The result is after a document is inserted or updated, it is not searchable using the new web searches. The modified web searches use the new table - xml\_path\_index. If the indexing of document fails, document doesnt show up in the xml\_path\_index. Hence when a search is done on xml\_path\_index, the document is not found.

If I am not able to figure out the reason for the failure of buildindex search (the indexing thread is now started after the document has been inserted into the database. but the indexing thread complains that the document was not found in xml\_documents. so the changes to xml\_documents are not committed - this is what i am not able to figure out), I suggest an alternative approach.

Metacat can run a seperate thread which does indexing after set amount of time. This could be 2 minutes. So every 2 minutes the following query is run:

```
SELECT d.docid FROM xml_documents d, xml_index i WHERE d.docid = i.docid(+) AND
i.docid is NULL and (d.doctype like 'eml://ecoinformatics.org/eml-2.0.0' or
d.doctype like 'eml://ecoinformatics.org/eml-2.0.1');
```

The above query is fast as it is a left outer join. (The syntax for postgres is different). It is even faster if we have an index on xml\_documents.doctype

The documents found in the above query can be indexed.

The downside of the above approach is that there is a delay of 2-3 minutes between when a document is inserted and when it is searchable.

As indexing fails for updating of documents also, for the above change to work, the indexing information in `xml_index` and `xml_path_index` will have to be deleted for a document after it has been updated.

**#5 - 12/08/2005 10:24 AM - Saurabh Garg**

Closing for now. The code seems to be working for now.

**#6 - 03/27/2013 02:19 PM - Redmine Admin**

Original Bugzilla ID was 2060