Metacat - Bug #331

metacat data replication feature

11/29/2001 02:01 PM - Matt Jones

Status:	Resolved		**Start date:**	11/29/2001
Priority:	Normal		**Due date:**	
Assignee:	Jing Tao		**% Done:**	0%
Category:	metacat		**Estimated time:**	0.00 hour
Target version:	1.1		**Spent time:**	0.00 hour
Bugzilla-Id:	331			

Description

Metacat currently does not replicate data objects. Agreed at the 2001 annual meeting that it should be an option to replicate data just as metadata works. This should be configurable by each metacat administrator, and should be independent of the metadata replication option (ie, metacat admins can choose to replicate nothing, metadata only, or metadata and data.

If data is not replicated between servers, but metadata is, some queries will return packages that refer in the triples to data entities that don't reside on the server that is queried. The queried server has the information it needs to retrieve the data set from the home server (identifier + home server code), and so we need a mechanism to do this in metacat. Thus, if a user requests a download of lter.4.1 from the NCEAS metacat, but the data hasn't been replicated from the lter metacat, then the NCEAS metacat will be able to forward the request to the LTER metacat, get the data object, and then satisfy the original request made by the user. All of this happens without the user knowing that the data object was located on another server.

Note this feature will be influenced by the "hub" concept, described in another bug.

Related issues:

Blocked by Metacat - Bug #332: hub replication feature	**Resolved**	**11/29/2001**

History

#1 - 04/26/2002 05:33 PM - Jing Tao

Now, administrator can configure the metacat to replicate nothing, meta data only or meta data and data file by set two properties in build.xml file.

If replication="off", metacat replicate nothing (no matter replicationdata's value).
If replication="on" and replicationdata="off", replicate meta data only.
If replication="on" and replicationdata="on", replicate both meta data and data file.

Metacat can replicate data file by both force replication and delta T replication. The following classes were revised:
MetacatReplication.java
ReplicationHandler.java
ReplMessageHandler.java
ForceReplicationHandler.java
DocumentImpl.java
MetacatServlet.java

#2 - 05/10/2002 03:58 PM - Jing Tao

Replication data will be considered two possibility. One is for accept data, the other is send data. So administer can configure Metacat:
1) No Replication at all.
2) Only replication xml (set accept data and send data off).
3) Replicate xml and only send data file.
4) Replicate xml and only accept data file
5) Replicate xml , send and accept data file

#3 - 05/10/2002 06:18 PM - Jing Tao

replicationsenddata and replicationacceptdata were added to build.xml.
Replication code was changed too.

So administrator can configure metacat to send or accept data now.

Six combinations of possibilities were test for both force replication and
delt T replciation. They worked fine.

#4 - 06/06/2002 08:44 AM - Jing Tao

Change data file replication from node base to server base.
Add a feature that for requesting data file to home server if local metacat
server doesn't have the data file.

#5 - 07/05/2002 10:45 AM - Jing Tao

Two new class, ReplicationServer and ReplicationServerList were created. In an
object of ReplicationServer, it has two fields, replicate and dataReplicate.
The data type is boolean and values read from xml_replication table. If
replicate=true and datareplicate=true, local metacat can replicate both xml
and data file to the is remote server. If replicate=ture but
datareplicate=false, local metacat can only replicate xml documents to this
remote server. If replicate =false, local metacat would not replicate anything.

To my understanding, the feature about the retrive data file from remote home
server(Because it is not reside in local server) should depend on xml_relation
table. MetaCat is document orientated, not package orientated. If a data file
is not in local server, there is no entry for it in xml_document table. But
becuase xml_doucments were replicated here, and we can found a entry in
xml_relation table for this data file. So from the dataset documents, we can
find the home server of the data file. Then retrive it from home server.

The retrive invovle tow action, read and export. Does it have more?

Comment and suggestion for this feature?

#6 - 07/12/2002 06:06 PM - Jing Tao

An new class was created to handle to read remote data file. Its name is
RemoteDocument. In read and export action, if a McdbNotFoundException happend,
the docid and revision will be stored in this exception. In catch clause for
this exception an object of RemoteDocument will create to hanlde read this
document from remote server (data file home server). If remote server server
couldn't find this document either, error message be send to local metacat and
local metatcat will send it to client.

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

Original Bugzilla ID was 331