2016-04-14 - Fedora Tech Meeting

Time/Place

This meeting is a hybrid teleconference and IRC chat. Anyone is welcome to join...here's the info:

- Time: 11:00am Eastern Daylight Time US (UTC-4)
- Dial-in Number: (712) 775-7035
  - Participant Code: 479307#
  - International numbers: Conference Call Information
- IRC:
  - Join the #fcrepo chat room via Freenode Web IRC (enter a unique nick)
  - Or point your IRC client to #fcrepo on irc.freenode.net

Attendees

- A. Soroka
- Jared Whiklo
- Bethany Seeger
- Yinlin Chen
- Doron Shalvi
- Andrew Woods
- James R. Griffin III 🌟
- Martin Haye (California Digital Library)

Agenda

1. 4.5.1 RC-2 testing status
2. Backup/Recovery recommendations and recipes
3. A complementary service to fcrepo-serialization?
4. Strange behavior of Binaries and NonRdfSourceDescriptions (see gist)
5. ...
6. Status of “in-flight” tickets

<table>
<thead>
<tr>
<th>key</th>
<th>summary</th>
<th>type</th>
<th>created</th>
<th>updated</th>
<th>due</th>
<th>assignee</th>
<th>reporter</th>
<th>priority</th>
<th>status</th>
<th>resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unable to locate Jira server for this macro. It may be due to Application Link configuration.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ticket Summaries

1. Please squash a bug!

<table>
<thead>
<tr>
<th>key</th>
<th>summary</th>
<th>type</th>
<th>created</th>
<th>updated</th>
<th>due</th>
<th>assignee</th>
<th>reporter</th>
<th>priority</th>
<th>status</th>
<th>resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unable to locate Jira server for this macro. It may be due to Application Link configuration.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2. Tickets resolved this week:

<table>
<thead>
<tr>
<th>key</th>
<th>summary</th>
<th>type</th>
<th>created</th>
<th>updated</th>
<th>due</th>
<th>assignee</th>
<th>reporter</th>
<th>priority</th>
<th>status</th>
<th>resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unable to locate Jira server for this macro. It may be due to Application Link configuration.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Tickets created this week:

<table>
<thead>
<tr>
<th>key</th>
<th>summary</th>
<th>type</th>
<th>created</th>
<th>updated</th>
<th>due</th>
<th>assignee</th>
<th>reporter</th>
<th>priority</th>
<th>status</th>
<th>resolution</th>
</tr>
</thead>
</table>

Unable to locate Jira server for this macro. It may be due to Application Link configuration.

Minutes

4.5.1 RC-2 testing status

- B. Seeger: A. Coburn could not attend this meeting
- Unable to locate Jira server for this macro. It may be due to Application Link configuration.
- J. Whiklo: Has "Verify reindexing to triplestore works failure" been resolved with the release of fcrepo-camel-toolbox 4.5.1
- Woods: This probably has been resolved with the latest release
- Whiklo: Manual tests all seem to work
- No other tickets for RC3 beyond FCREPO-1985 and FCREPO-1986
- A. Woods: Shoot for the release of RC3 this week

Backup/Recovery recommendations and recipes

- Woods: Increasingly clear that one of the primary features for fcrepo is to make it easy to get content into and out of the repository
- We have tooling which helps with aspects of this
- There are no clear recommendation for what are the recommended practices for maintaining a backup state for the repository, and how to restore backups
- 4.5.1 release introduces the possibility of MySQL or PostgreSQL rather than LevelDB
- A case would be moving from an existing fcrepo4 instance into one backed by MySQL
- There current exist backup and restore endpoints (all of these relate to RDF serialization, and are JCR-compliant)
- Would want to retain these for future releases
- A. Soroka: Confirms that GETting RDF serializations and PUTting them back were scoped some time ago

A complementary service to fcrepo-serialization?

- Woods: B. Seeger's work with serialization does a strong job of the GET side
- Iterates over repo, serializes the RDF to disk, offers the option to serialize binaries
- Whiklo: This is more of a migration than backup
- Soroka: Migrations are backups (or, should be)
- This is the case right now; might change in the future; would prefer to keep this model

- Whiklo: This would need to be a core module?
- Soroka: Serialization is handy...tools for constant serialization, or, use a script to walk the repo
- Similarly, might have scripts which can take pile of serializations and place them into repo
- Would likely just be BASH (or Python) scripts
- Seeger: Apache Camel would not a good fit; in agreement
• Soroka: Start and the end of a stream of events...restore is just for an event starting from the backup
• Woods: We do have this already...using the various ways of having serialization on disk
• Seeger has implemented both with fcrepo-serialization
• Soroka: Reassuring to still offer tools for creating snapshots; Maybe this can be baked into fcrepo-camel-toolbox?
• Woods: We already have that
• Seeger: It can exist and wouldn’t be difficult if it doesn’t
• Woods: fcrepo-serialization works off of messages in a queue; reindexer walks the repo or a subtree, putting messages on the appropriate queue(s)
• Has tested this with the reindexer, serializer serializes based upon the messages
• Note that this also offers the ability to specify the MIME type of the RDF being serialized
• Woods: A. Coburn sent an e-mail to the list warning about relationships created by direct and indirect containers...not persisting those directly into the repository
• Soroka: What kind of headers and strictness should be supported for the serialization and restoration endpoint?
• Ensure that headers and tools are available for the ordinary PUT/POST to alleviate these concerns
• M. Haye: (Supports the idea of releasing additional scripts)
• Whiklo: Suggests that they try serializing part of an fcrepo instance...and outline approach to serialize and restore
• Look to automate the process (using scripts)
• Whiklo: Suggests that a JIRA Issue be created, directing someone who is interested in such a task
• Might be good for someone with large set of sample data?
• Woods: Should this be in a Java utility? Should these scripts be in a specific language?
• Woods: Notes that while the import/export features (a smaller version of backup/restore features) have been deprecated
• Confirm that they have not deprecated backup/restore
• Woods will create the ticket - done: Unable to locate Jira server for this macro. It may be due to Application Link configuration.

Strange behavior of Binaries and NonRdfSourceDescriptions (see gist)

• Testing was performed by Coburn (who was not able to attend this meeting)
• Seeger: They prefer that this discussion be deferred until they are present
• Whiklo: They offer a lot of information in gist, and this should be reviewed by others

Status of "in-flight" tickets

• (High priority issues are in review)
• (No specific questions about issues)
• Woods: Regarding in-flight tickets, several issues in review
• Woods: The second Fedora Camp was held earlier this week (04/11/16 - 04/13/16) in Pasadena, CA
• Invites members of the community to seek to co-present at future Fedora Camps
• Might schedule the next during the Fall
• M. Haye: Raises the issue of getting blank nodes back on the table, and fixing them
• Haye: CDL and he understands that this is a political topic, but feels that supporting bnodes is the best approach
• Soroka: Does this involve changing fcrepo itself to support this?
• Haye: When a node created, fcrepo generates an ID, rendering it impossible to structure these as an RDF list within JSON-LD...
• Soroka: So fcrepo should be trying to reproduce bnodes rather than skolemizing them?
• Opposes the production of bnodes in responses
• Can offer better solution regarding the problem of ordering
• Ordering is a reasonable desire for fcrepo, and it is reasonable to offer use cases
• However, this is not the same as reproducing bnodes
• Soroka: Historically, considerable time and institutional was lost due to confusion between use cases involving ordering (as well as hierarchy, and other organizational schemes) and those with users looking at specific implementations of functionality which (shouldn’t) involve bnodes
• D. Shalvi: (Providing updates involving efforts to migrate between instances of fcrepo)
• Resetting up environment, couple of hundred thousand objects migrated on 4.3 some months ago
• Shalvi: attempting this again with MySQL
• They are still reestablishing where they were with migration tests
• There are around a few million objects
• They are attempting to profile memory consumption
• Woods: Any discussion involving G1/GC flag or Java options in this second attempt?
• D. Pino has been advocating for discussions involving garbage collection options
• Whiklo: Java option is -XX:+UseG1GC
• Soroka: Much of this conversation was in IRC
• There are no new ideas for garbage collection; there are only ideas for experimentation
• Shalvi: Much of the issues encountered seemed like a memory leaks at the time
• Now, looking to inspect the heap more carefully

• Woods: (Announces that the next) Performance and Scalability meeting is scheduled for Monday (04/18/16) at 11:00EDT
  ○ 2016-04-18 Performance - Scale Meeting

• Moving those tests forward
• N. Ruest has done more testing
• Likely encounters memory leaks
• (Re)cycles server, quickens again

• Seems likely that fcrepo performs well
• The leakage seems to arise when the objects are ingested
• This is likely originating lower in the stack (Modeshape layer?)

• (Regarding the topic of the default garbage collection)
• Was the default garbage changed to the default in Java 7 or Java 8 releases recently?
• D. Pino and J. Coyne did raise this discussion in the past
• (No parties could confirm this)