

Preliminary Testing or Exploring for Migration-utils

DataStreamStores and ObjectStores for the Testing

Testing for Migration-utils happened on two different repositories (production and development) with different structures for data stream store and object store. Following are how those data stream stores and object stores organized in those two repositories:

Repository A

Under dataStreamStore/objectStore, data is organized as:

- First level-- year: for example: 2005, 2006, ... 2015
- second level-- DDMM (date), for example, under 2005, it has 0930, 1014, 1102...
- third level and fourth level-- two level hash directories, for example 09/49,...
- under the four levels: there are the actual datastream, for example ~datastreams/2015/0316/09/49/iudl_1897784+ITEM-METADATA+ITEM-METADATA.

Repository B

It is the same as the default setting that migration-utils has:

- two-layer hash directories, for example 09/49,...
- under the two layers , there are the actual datastream/object, for example ~datastreams/2015/0316/09/49/iudl_1897784+ITEM-METADATA+ITEM-METADATA.

Fails in the Tests

Fails on Repository A

NullPointerException, detailed as following:

```
INFO 15:27:18.336 (InternalIDResolver) Building an index of all the datastreams in "/srv/fedora/datastreams"...
INFO 15:31:58.187 (Migrator) Processing "iudl:1135374"...
DEBUG 15:31:58.268 (BasicObjectVersionHandler) Considering object iudl:1135374 version at 2012-09-27T12:58:36.220Z.
DEBUG 15:31:58.844 (BasicObjectVersionHandler) Considering changed datastream version AUDIT.0
DEBUG 15:31:59.535 (BasicObjectVersionHandler) Considering changed datastream version DC1.0
DEBUG 15:31:59.551 (BasicObjectVersionHandler) Adding http://purl.org/dc/elements/1.1/identifier value iudl:1135374
DEBUG 15:31:59.551 (BasicObjectVersionHandler) Adding http://purl.org/dc/elements/1.1/title value Union Board records 1912-2010, bulk 1922-2010. 13
December 1955. (Minutes
, agendas and working papers, 1912-2009, undated, "Official" Minutes of the Union Board, 1912-2009, Bound Volume: Minutes of the Union Board: April
1955 – 24 April 1956)
DEBUG 15:31:59.801 (BasicObjectVersionHandler) Considering object iudl:1135374 version at 2012-09-27T12:58:36.734Z.
DEBUG 15:31:59.801 (BasicObjectVersionHandler) Considering changed datastream version RELS-EXT.0
DEBUG 15:31:59.856 (BasicObjectVersionHandler) Considering object iudl:1135374 version at 2012-09-27T12:58:37.067Z.
DEBUG 15:31:59.856 (BasicObjectVersionHandler) Considering changed datastream version DC.1
DEBUG 15:31:59.862 (BasicObjectVersionHandler) Adding http://purl.org/dc/elements/1.1/identifier value http://purl.dlib.indiana.edu/iudl/archives/AC3408-01027
DEBUG 15:31:59.862 (BasicObjectVersionHandler) Adding http://purl.org/dc/elements/1.1/identifier value iudl:1135374
DEBUG 15:31:59.862 (BasicObjectVersionHandler) Adding http://purl.org/dc/elements/1.1/title value Union Board records 1912-2010, bulk 1922-2010. 13
December 1955. (Minutes
, agendas and working papers, 1912-2009, undated, "Official" Minutes of the Union Board, 1912-2009, Bound Volume: Minutes of the Union Board: April
1955 – 24 April 1956)
DEBUG 15:31:59.909 (BasicObjectVersionHandler) Considering object iudl:1135374 version at 2012-09-27T12:58:37.286Z.
DEBUG 15:31:59.909 (BasicObjectVersionHandler) Considering changed datastream version METADATA.0
[WARNING]
java.lang.reflect.InvocationTargetException
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:606)
at org.codehaus.mojo.exec.ExecJavaMojo$1.run(ExecJavaMojo.java:293)
at java.lang.Thread.run(Thread.java:744)
Caused by: java.lang.NullPointerException
at org.fcrepo.migration.foxml11.Foxml11InputStreamFedoraObjectProcessor$Foxml11DataStreamVersion.getContent
(Foxml11InputStreamFedoraObjectProcessor.java:370)
at org.fcrepo.migration.handlers.BasicObjectVersionHandler.processObjectVersions(BasicObjectVersionHandler.java:150)
at org.fcrepo.migration.handlers.VersionAbstractionFedoraObjectHandler.processObject(VersionAbstractionFedoraObjectHandler.java:90)
at org.fcrepo.migration.handlers.ObjectAbstractionStreamingFedoraObjectHandler.completeObject(ObjectAbstractionStreamingFedoraObjectHandler.java:
62)
at org.fcrepo.migration.foxml11.Foxml11InputStreamFedoraObjectProcessor.processObject(Foxml11InputStreamFedoraObjectProcessor.java:115)
at org.fcrepo.migration.Migrator.run(Migrator.java:69)
at org.fcrepo.migration.Migrator.main(Migrator.java:29)
... 6 more
```

Fails on Repository B

Redirect link migration error

some of purl_redirects were migrated right. But some of them weren't.

```
DEBUG 09:03:41.472 (BasicObjectVersionHandler) Considering changed datastream version PURL_REDIRECT.1
ERROR 09:03:41.480 (FedoraRepositoryImpl) error creating resource http://birch.dlib.indiana.edu:8080/fedora4f/rest/migration\_6/migrated-fedora3/iudl/49/87/PURL\_REDIRECT:204NoContent
ERROR 09:03:41.484 (FedoraRepositoryImpl) Error making or building PUT request.
org.fcrepo.client.FedoraException: error creating resource http://birch.dlib.indiana.edu:8080/fedora4f/rest/migration\_6/migrated-fedora3/iudl/49/87/PURL\_REDIRECT:204NoContent
at org.fcrepo.client.impl.FedoraRepositoryImpl.createOrUpdateRedirectDatastream(FedoraRepositoryImpl.java:181) ~[fcrepo-client-impl-4.1.2-SNAPSHOT.jar:na]
at org.fcrepo.migration.handlers.BasicObjectVersionHandler.processObjectVersions(BasicObjectVersionHandler.java:146) [classes:/na]
at org.fcrepo.migration.handlers.VersionAbstractionFedoraObjectHandler.processObject(VersionAbstractionFedoraObjectHandler.java:90) [classes:/na]
at org.fcrepo.migration.handlers.ObjectAbstractionStreamingFedoraObjectHandler.completeObject(ObjectAbstractionStreamingFedoraObjectHandler.java:62) [classes:/na]
at org.fcrepo.migration.foxml11.Foxml11InputStreamFedoraObjectProcessor.processObject(Foxml11InputStreamFedoraObjectProcessor.java:115) [classes:/na]
at org.fcrepo.migration.Migrator.run(Migrator.java:69) [classes:/na]
at org.fcrepo.migration.Migrator.main(Migrator.java:29) [classes:/na]
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method) ~[na:1.7.0_45]
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57) ~[na:1.7.0_45]
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43) ~[na:1.7.0_45]
at java.lang.reflect.Method.invoke(Method.java:606) ~[na:1.7.0_45]
at org.codehaus.mojo.exec.ExecJavaMojo$1.run(ExecJavaMojo.java:293) [exec-maven-plugin-1.4.0.jar:na]
at java.lang.Thread.run(Thread.java:744) [na:1.7.0_45]
[WARNING]
java.lang.reflect.InvocationTargetException
at sun.reflect.NativeMethodAccessorImpl.invoke0(Native Method)
at sun.reflect.NativeMethodAccessorImpl.invoke(NativeMethodAccessorImpl.java:57)
at sun.reflect.DelegatingMethodAccessorImpl.invoke(DelegatingMethodAccessorImpl.java:43)
at java.lang.reflect.Method.invoke(Method.java:606)
at org.codehaus.mojo.exec.ExecJavaMojo$1.run(ExecJavaMojo.java:293)
at java.lang.Thread.run(Thread.java:744)
Caused by: java.lang.RuntimeException: org.fcrepo.client.FedoraException: org.fcrepo.client.FedoraException: error creating resource http://birch.dlib.indiana.edu:8080/fedora4f/rest/migration\_6/migrated-fedora3/iudl/49/87/PURL\_REDIRECT:204NoContent
at org.fcrepo.migration.handlers.BasicObjectVersionHandler.processObjectVersions(BasicObjectVersionHandler.java:165)
at org.fcrepo.migration.handlers.VersionAbstractionFedoraObjectHandler.processObject(VersionAbstractionFedoraObjectHandler.java:90)
at org.fcrepo.migration.handlers.ObjectAbstractionStreamingFedoraObjectHandler.completeObject(ObjectAbstractionStreamingFedoraObjectHandler.java:62)
at org.fcrepo.migration.foxml11.Foxml11InputStreamFedoraObjectProcessor.processObject(Foxml11InputStreamFedoraObjectProcessor.java:115)
at org.fcrepo.migration.Migrator.run(Migrator.java:69)
at org.fcrepo.migration.Migrator.main(Migrator.java:29)
... 6 more
Caused by: org.fcrepo.client.FedoraException: org.fcrepo.client.FedoraException: error creating resource http://birch.dlib.indiana.edu:8080/fedora4f/rest/migration\_6/migrated-fedora3/iudl/49/87/PURL\_REDIRECT:204NoContent
at org.fcrepo.client.impl.FedoraRepositoryImpl.createOrUpdateRedirectDatastream(FedoraRepositoryImpl.java:186)
at org.fcrepo.migration.handlers.BasicObjectVersionHandler.processObjectVersions(BasicObjectVersionHandler.java:146)
... 11 more
Caused by: org.fcrepo.client.FedoraException: error creating resource http://birch.dlib.indiana.edu:8080/fedora4f/rest/migration\_6/migrated-fedora3/iudl/49/87/PURL\_REDIRECT:204NoContent
at org.fcrepo.client.impl.FedoraRepositoryImpl.createOrUpdateRedirectDatastream(FedoraRepositoryImpl.java:181)
... 12 more
[INFO] -----
[INFO] BUILD FAILURE
[INFO] -----
[INFO] Total time: 51.350s
[INFO] Finished at: Tue Apr 21 09:03:41 EDT 2015
[INFO] Final Memory: 33M/290M
[INFO] -----
[ERROR] Failed to execute goal org.codehaus.mojo:exec-maven-plugin:1.4.0:java (default-cli) on project migration-utils: An exception occurred while executing the Java class. null: InvocationTargetException: org.fcrepo.client.FedoraException: org.fcrepo.client.FedoraException: error creating resource http://birch.dlib.indiana.edu:8080/fedora4f/rest/migration\_6/migrated-fedora3/iudl/49/87/PURL\_REDIRECT:204NoContent -> [Help 1]
]
[ERROR]
[ERROR] To see the full stack trace of the errors, re-run Maven with the -e switch.
[ERROR] Re-run Maven using the -X switch to enable full debug logging.
[ERROR]
[ERROR] For more information about the errors and possible solutions, please read the following articles:
[ERROR] [Help 1] http://cwiki.apache.org/confluence/display/MAVEN/MojoExecutionException
```

Policy datastream migration

There is no migration error, but the content of the policy was not migrated.

No master content

This error is due to the master data stream has no content. But I would like to see the migration-utils log the error and sequence number of the record for future investigate, and keep migration.

Expected Features for Future Version

1. Feature to skip the top n objects for migration: since fedora repository randomly hashed the data, we can not control what data we want to migrate. This feature allows the users to skip certain number of records, therefore, if error happens during the migration, the users don't need to re-migrate the data which already migrated.
2. Feature to overwrite the migrated objects: right now, if the objects exist in the migration path, then the migration-utils just stopped with errors. This feature allows user to overwrite the migrated data. For example, during the migration, some errors happened, only part of the data streams has been migrated, this feature allows user to investigate the data, re-migrate and overwrite the previous one.
3. Feature to skip no-content datastream, and keep migration, but record the no-content datastream in log with the sequence of the migration. For example, some objects may have no-content data stream due to some reasons, we can migrate them first, then we either remove the data stream later or re-migrate this one (here needs the sequence information, so that we know where to start from)
4. Feature to leverage fedora 4 hierarchical structure? For example, create the migration path based on another identifier (such as, part of the purl link), not the pid(hash directories).

[Related JIRA Issues:](#)