2013-12-05 - Fedora Committer 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)
- ReadyTalk:
  - U.S.A/Canada toll free: 866-740-1260, participant code: 2257295
  - International toll free:
    - http://www.readytalk.com/intl
    - Use the above link and input 2257295 and the country you are calling from to get your country's toll-free dial-in number
    - Once on the call, enter participant code 2257295
- IRC:
  - Join the #duraspace-ff chat room via Freenode Web IRC (enter a unique nick)
  - Or point your IRC client to #duraspace-ff on irc.freenode.net

Attendees

- Andrew Woods
- Benjamin Armintor
- Osman Din
- Eric James
- Michael Durbin
- Unknown User (escowles@ucsd.edu)
- Scott Prater
- Greg Jansen
- Jonathan Green
- Nigel Banks ⭐
- Chris Beer

Agenda

1. Holiday release goals and status
   a. Wait on LevelDB/ModeShape bugfix?
2. Reality check on recommended node type usage (if time permits) (Scott, Nigel, Adam?, Chris?)
   a. Content Modeling directions
3. Code4Lib pre-conference plans

Previous Actions

- ☐

Status

Osman

Esme

- Testing maximum upload/download file size in REST API.
- Setting up clustering on UCSD VMs to test 2-node ingest/etc.
- Repeating F3/F4 comparison tests with updated benchtool.

Mike

- Finished round one versioning tickets
- Finishing documentation
- Will move on to other tickets, enhancements or release testing

Nigel

- Working with Scott on instructions for migrating content models to node types: Converting an Fedora 3 Islandora Content model to Fedora 4 by Example
- Wiki clean up

Eric

- investigating and implementing external indexing (design plan, fcrepo-jms-ingdexer-pluggable,solr standalone,fcrepo-jms,fcrepo-transform)

Frank

Adam

Chris

Ben

Scott

- Working with Nigel on instructions for migrating content models to node types: How to Migrate a Fedora 3 Content Model to a Fedora 4 Node Type
- On the back burner: continuing with Authz/No AuthZ performance testing (will finish up after finishing first draft of CModel instructions)

Greg

Andrew

- Reviewing completed tickets
- Spoke with Shared IR regarding process and collaboration
- Fixed fcrepo-jms-indexer-pluggable build failure
- Starting user docs, targeting release functionality

Minutes

Roll call and attendance.

Mike: Scott Testing for Authz in good shape?
Scott: Yup, rerun the tests, should be done in a day or so.
Mike: Work on content modelling we will talk later.
Andrew: Authz Greg? Can we pull this into the Web UI.
Greg: Nothing done as of yet, but will talk with Chris about it.
Andrew: Lets focus on getting that in soon.
Mike: Feature Freeze this weekish?
Mike: Hows large files going?
Frank: No, when you use single node, we still get out of memory errors, we're waiting on a fix for the next release of ModeShape.
Mike: Will this happen on Dec 17th? Are we reasonably sure?
Frank: Yes. We should be able to use Modeshape 3.7 before christmas.
Andrew: There shouldn't be too much trouble in the upgrade.
Mike: Are we trying to roll this release after bumping this version?
Andrew: Maybe we put this question off for a bit, till after the features meeting.
Mike: External search? Can we do this without Adam's import?
Andrew: Seems like Eric and Adam are making good progress, and everything is under control. Eric can you move forward while Adam is gone?
Eric: Yes, I've got some work I expect to hand off to him later.
Andrew: Sounds great but maybe ambitions to get this wrapped up for tomorrow? It will probably bleed into next week?
Eric: Yes most likely. The transform modules is presenting some problems at the moment. There is a line in the Fedora Transform class, where the session is already injected, but it's creating additional sessions which default to null.
Andrew: Could you raise the problem on IRC, later?
Eric: Yup.
Mike: Versioning will be ready for acceptance testing after the next pull request, gets merged. Any one want to do the acceptance testing?
Scott: I can do some of the acceptance testing next week.
Andrew: Jonathan, Is that something that anyone at DGI wants to poke around with?
Jonathan: Ya probably, I can find someone.
Andrew: It would be a combination of reading the wiki and checking if it make sense, then a test of the stated functionality.
Jonathan: Ok.
Mike: There is still some work to be done on the easy deployment, has any one got the time to do it?
Andrew: No one has grabbed them.
Mike: We need to do that soon.
Andrew: They should be low hanging fruit.
Mike: Single node performance, there has been some reorganization of the notes and the testing is winding down.
Esme: Yup I'm now focusing on the large file testings.
Greg: I'm working on the cluster testing atm.
Andrew: I'm curious about the wide gap in results, is this due to the recent changes to the bench tool? We need to get a consolidated view of what the actual results are.
Esme: I think the update have given us better functionality, but also have resulted in people getting different results, also if the tool is run on the same machine as the client.
Andrew: Osman could you run your tests again?
Osman: Yup.
Mike: I think Ben is working on improving single node performance atm.
Andrew: Are you still looking into that?

Ben: strange computer nosies
Andrew: Hmm
From IRC: Ben I think I need to reboot.
Mike: Any results on clustering Chris?
Chris: Hopefully, soon.
Mike: Esme are you having problems with clustering as well?
Esme: Ya. Chris could you update the wiki page.
Chris: Yup, but it may not be helpful for you.
Mike: Greg may have a tool.
Greg: Yup, its just a maven project and a shell script. It syncs the data and properties across nodes. It's working on my machine currently. You just have to customize whats in settings.xml.
Esme: I'll try that, it sounds more promising.
Greg: Right now it doesn't do a separate client machine or a load balancing, it's something we could add later. Maybe we could work together especially on the load balancer.
Esme: Maybe, I could help.
Andrew: I think Frank and Osman have set that up as well.
Mike: Hopefully we'll have some numbers real soon.
Andrew: We need to show the behaviour of ingest/etc as you increase the number of nodes in a cluster.
Scott: The systemadmins are setting up a cluster this week the "uw madison cluster" so hopefully I can pick away at that. https://wiki.duraspace.org/display/FF/Test+-+Platform+Profile%3A+Cluster+at+University+of+Wisconsin
Andrew: Osman, Persistence to disk?
Osman: I wrote a program, based on infinispan configurations are just serialize objects, if you use a particular back end stores.
Andrew: Useful to have a sample of what a fedora serialized node looks like given a particular backend for infinispan in the wiki for preservation?
Mike: It would be nice to see the full path of an object to disk when ingested though the rest interface.
Scott: ++
Osman: I could do some of what Andrew mentioned, but there are many types of serialization with inifinspan it may be very time consuming.
Andrew: I think you could write up the documentation about what Mike mentioned?
Scott: Like to know about migrations, so we can confirm that what we've migrated is what we expected.
Osman: I think some of that is covered in the documentation, we're on the same page.
Andrew: Lets talk about External search and large files, the other features are going well.
Andrew: For this release for supporting large files, I thought a clear description when ingesting via the REST API and Federating, documenting where things are and serving up large files (REST / etc), would be nice to have. There is also this bug around fixity, which is related but not really a problem with large files per-se.
There is some value in connecting the file-system connector to a much larger set of files to see what. Esme as demonstrated File backend for REST ingest, seems to work for Gigabyte level, we are currently testing the Terabyte level. We have been using Leveldb for performance testing. All the performance has been on leveldb, but for ingest we are using file.
Frank: Leveldb is quick, file slow.
Esme: I've been using file, but I haven't noticed much difference between them.
Andrew: Is anyone using anything else other than file/leveldb.
Greg: Whats the base line.
Andrew: Leveldb.
Scott: I'm using infinispan basic, is that leveldb?
Andrew: It's defined in your repository.json file.
Andrew: It's probably important and for people who are using fedora will need to decide between the two we should document the tradeoffs. The single node thus far has been mostly with leveldb. We should pull these numbers together.
Andrew: Ingesting files via the Rest API if your using the leveldb backend is broken, Modeshape is putting in a fix for that.
Frank: It seems that Modeshape writes a lot into Memory. I would like to see when Esme is ingesting TB of data, if you could look at the stack/heap that would be interesting.
Esme: I run with 2 GB of heap so I wouldn't noticed that before. Tomorrow I'll run the test and see what happens.
Andrew: Fixes for leveldb are coming into ModeShape. THe other implication is that Modeshape will be writing less.
Frank: At the moment they have the transaction associated with the binary stores, so they are breaking up transactions.
Andrew: Release planning. The plan is to get the release out this sprint, or at the end of this sprint. Including user documentation. There are people who are scheduled to leave at the end of next week, so it be nice to get it done before they go. Also there is a general value in getting a release out this year. One year anniversary, there is political value in that. Based on this ModeShape issue, it might be worth pushing this release out? Id like to see regular release every month rather than by external factors as a philosophy. How do people feel?
Frank: I would like to put a disclaimer, yes you can use it but watch out for heap explosions. I am available after the 13th so I could help with the release.
Andrew: So you are around for all of December.
Frank: I can do 4-5 hours a day.
Andrew: Is this ModeShape bug fix worth waiting for? Or can it wait till a release in January. Is anyone else available to help?
Mike: I can from the 16th-20th, the 23rd and 30th probably.
Mike: We shouldn't postponed the release. We already have a disclaimer to not use it in production.
Andrew: I think we shouldn't as well, we could get valuable user feedback.
Chris: We should push out the release.
Andrew: ?
Chris: It's a choice between speed, and working large files.
Andrew: I would like to pull the trigger on a release for next week. I think it's important to do a release before christmas, lets prepare to do a release before christmas and if ModeShape gets it's version bumped on time, then we'll included it in the release if not, then we won't.
Andrew: Lets assume that ModeShape will release on the 17th, than Frank Chris, etc, can help roll out the release after that point. We could do our release on the 19th. If ModeShape doesn't come out on the 17th, then we still release on the 19th.
Mike: ++.
Esme: ++.
Greg: ++.
Nigel: I'd like to see some release testing.
Andrew: I agree. This week is features, next week should be all about testing and documentation.
Andrew & Scott: Documentation ++.
Andrew: I think we have a plan, in summary we wrap the features as much as we can this week, with minor bleed over. Next week the majority of the team will focus on human level testing and documentation. Banging on the system from a users perspective. Getting that wrapped up by the end of next week. So that we will be ready before ModeShape comes out, then we can do the upgrade of ModeShape.

New Actions

.