DevMtg 2019-01-30

Developers Meeting on Weds, January 30, 2019

Today's Meeting Times
ⓘ

- DSpace Developers Meeting / Backlog Hour: 20:00 UTC in #duraspace IRC or #dev-mtg Slack channel (these two channels sync all conversations)

Agenda

Quick Reminders

Friendly reminders of upcoming meetings, discussions etc

- DSpace 7 Working Group (2016-2023): Next meeting is Thurs, Jan 31 at 15:00 UTC
- DSpace 7 Entities Working Group (2018-19): Next meeting on Tues, Feb 5 at 16:00 UTC

    - Last meeting notes at 2019-01-22 DSpace 7 Entities WG Meeting
- DSpace Developer Show and Tell Meetings: On hold until interesting topics arise.

Discussion Topics

If you have a topic you'd like to have added to the agenda, please just add it.

1. (*Ongoing Topic*) DSpace 7 Status Updates for this week (from DSpace 7 Working Group (2016-2023))
2. *(Ongoing Topic)* DSpace 6.x Status Updates for this week

    a. 6.4 will surely happen at some point, but no definitive plan or schedule at this time. Please continue to help move forward / merge PRs into the dspace-6.x branch, and we can continue to monitor when a 6.4 release makes sense.
3. Upgrading Solr Server for DSpace (Status Updates?)
    a. PR https://github.com/DSpace/DSpace/pull/2058
4. DSpace Docker and Cloud Deployment Goals (old) (Terrence W Brady)
    a. Simplify invocation by using multiple fragments, auto load content on startup
        i. https://github.com/DSpace-Labs/DSpace-Docker-Images/pull/68
        ii. Summary page: https://github.com/DSpace-Labs/DSpace-Docker-Images/blob/helper_cmds/docker-compose-files/dspace-compose/ComposeFiles.md
    b. Speed up Docker builds
        i. https://github.com/DSpace/DSpace/pull/2307
    c. Add Docker build/push to Travis
        i. This make sense to consider after 2307 is merged
        ii. https://github.com/DSpace/DSpace/pull/2308
5. Brainstorms / ideas (*Any quick updates to report?*)
    a. (*On Hold, pending Steering/Leadership approval*) Follow-up on "DSpace Top GitHub Contributors" site (Tim Donohue): https://tdonohue.github.io/top-contributors/
    b. Bulk Operations Support Enhancements (from Mark H. Wood)
    c. Curation System Needs (from Mark H. Wood)
        i. PR 2180 improves reporting. Ready for review.
6. Tickets, Pull Requests or Email threads/discussions requiring more attention? (*Please feel free to add any you wish to discuss under this topic*)
    a. Quick Win PRs: https://github.com/DSpace/DSpace/pulls?q=is%3Aopen+review%3Aapproved+label%3A%22quick+win%22

Tabled Topics

These topics are ones we've touched on in the past and likely need to revisit (with other interested parties). If a topic below is of interest to you, say something and we'll promote it to an agenda topic!

1. Management of database connections for DSpace going forward (7.0 and beyond). What behavior is ideal? Also see notes at DSpace Database Access
    a. In DSpace 5, each "Context" established a new DB connection. Context then committed or aborted the connection after it was done (based on results of that request). Context could also be shared between methods if a single transaction needed to perform actions across multiple methods.
    b. In DSpace 6, Hibernate manages the DB connection pool. Each **thread** grabs a Connection from the pool. This means two Context objects could use the same Connection (if they are in the same thread). In other words, code can no longer assume each `new Context()` is treated as a new database transaction.
        i. Should we be making use of `SessionFactory.openSession()` for READ-ONLY Contexts (or any change of Context state) to ensure we are creating a new Connection (and not simply modifying the state of an existing one)? Currently we always use `SessionFactory.getCurrentSession()` in HibernateDBConnection, which doesn't guarantee a new connection: https://github.com/DSpace/DSpace/blob/dspace-6_x/dspace-api/src/main/java/org/dspace/core/HibernateDBConnection.java
    c. Bulk operations, such as loading batches of items or doing mass updates, have another issue: transaction size and lifetime. Operating on 1 000 000 items in a single transaction can cause enormous cache bloat, or even exhaust the heap.

i. Bulk loading should be broken down by committing a modestly-sized batch and opening a new transaction at frequent intervals. (A consequence of this design is that the operation must leave enough information to restart it without re-adding work already committed, should the operation fail or be prematurely terminated by the user. The SAF importer is a good example.)

ii. Mass updates need two different transaction lifetimes: a query which generates the list of objects on which to operate, which lasts throughout the update; and the update queries, which should be committed frequently as above. This requires *two* transactions, so that the updates can be committed without ending the long-running query that tells us what to update.

Ticket Summaries

1. Help us test / code review! These are tickets needing code review/testing and flagged for a future release (ordered by release & priority)

key	summary	type	created	updated	assignee	reporter	priority	status	fixversions

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

2. Newly created tickets this week:

key	summary	type	created	assignee	reporter	priority	status

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

3. Old, unresolved tickets with activity this week:

key	summary	type	created	updated	assignee	reporter	priority	status

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

4. Tickets resolved this week:

key	summary	type	created	assignee	reporter	priority	status	resolution

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

5. Tickets requiring review. This is the JIRA Backlog of "Received" tickets:

key	summary	type	created	updated	assignee	reporter	priority

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

Meeting Notes

Meeting Transcript

Log from #dev-mtg Slack (All times are CST)

```
Tim Donohue [2:01 PM]
@here: it's DevMtg time.  Agenda is at: https://wiki.duraspace.org/display/DSPACE/DevMtg+2019-01-30
Let's do a brief roll call to see who's able to join today

James Creel [2:01 PM]
Greetings

Mark Wood [2:01 PM]
Hello.

Terry Brady [2:01 PM]
hello

Tim Donohue [2:02 PM]
Hello usual crew :wink:  It seems like this meeting is mostly only the four of us these days (though, there are
a lot of DSpace meetings nowadays -- and plenty of others in other mtgs)
So, as noted in #dev, this agenda has been very static as of late.  I'd welcome folks to bring topics for
discussion (either today or in coming weeks).  This is often the last agenda I am able to even think about in a
given week.
In any case, we can see where discussion brings us today.
I'll admit, on the DSpace 7 side, I don't have much to say today.  Development is ongoing & very active. There
will be much more to discuss in tomorrow's DSpace 7 meeting (agenda forthcoming)
I did get a chance last week to meet with the core contributors (Atmire & 4Science), and I'll be reporting out
about that tomorrow
Ditto on the DSpace 6 side. I don't have any updates myself
Are there any updates / questions anyone would like to share on these first two?  Otherwise, we can move
quickly along to some of the ongoing hot topics (#3 and #4 on the agenda)
Ok, no one typing, so I'll assume no questions/comments
Moving right along then to Solr Server upgrade: https://wiki.duraspace.org/display/DSPACE
/Upgrading+Solr+Server+for+DSpace
Any updates to share this week @mwood?

Mark Wood [2:08 PM]
I've been covered up with other stuff recently, but today I verified that the current code builds, 'ant
fresh_install' succeeds, and the result "looks right".

Tim Donohue [2:08 PM]
:tada:

Mark Wood [2:08 PM]
I'm at home due to weather closure, and discovering all the stuff I don't yet have installed, which would be
useful for testing this.
At the moment I have spring-rest running, and now I need to get the GUI up and see if it can actually talk to
Solr.
Oh, I also verified that one can simply copy the Solr cores to wherever standalone Solr will look for them, and
they'll start.

Tim Donohue [2:10 PM]
Enjoy your frigid temps at home, @mwood.  I'm betting you are in the same realm as I am...currently -17 F (-27
C) outside where I am.

Terry Brady [2:10 PM]
Yikes!

Tim Donohue [2:10 PM]
It's good to hear it's "working"

Mark Wood [2:10 PM]
Ow!  We're actually up to +1 F.
```

Terry Brady [2:10 PM]
We are having an unusually sunny winter here

James Creel [2:11 PM]
50 F and cloudy in central TX

Tim Donohue [2:12 PM]
We've set a record today in temps...and aren't getting back to 0 F until sometime tomorrow. Strangely, we're
supposed to get to nearly 50 F by the weekend...so, yes, on a roller coaster :wink:
In any case...back to Solr!

Mark Wood [2:12 PM]
There is an issue that we'll need to at least mention in the setup doco: Your Solr admin. (if that's not you)
may want you to name your cores specifically, and that will affect core.properties and various DSpace
configuration properties.
(Think: a large Solr install, even a cluster, with many cores belonging to different applications.)
In the simple case, though, you can just copy [DSpace]/solr to somewhere that Solr looks at, and it should be
enough.
Also get used to seeing Solr on port 8983 instead of 80 or 8080.
Anyway I'll write up something on "if your Solr admin. tells you what core names are wanted, here are the
places you'll need to edit."

Tim Donohue [2:16 PM]
Could we document where/how Core names affect DSpace configuration? Doing so could give us a better picture of
whether to split out some new configurations to make this easier -- e.g. we could have a "discovery.solr.core =
[name]", etc.
Oh, you just answered my question :wink:

Mark Wood [2:17 PM]
Generally we'll need to help DSpace installers translate between DSpace concepts and Solr concepts. It
shouldn't be difficult or lengthy.
That's all I have for now, unless there are questions.

Tim Donohue [2:18 PM]
I don't think I have specifics, other than...what are the next steps here? Is this ready to review & think
about merging? Are there more TODOs?
Did you get the Solr Schema stuff merged in (from @terrywbrady)?

Mark Wood [2:19 PM]
Ah, no, that is still to be done.
That's next.

Tim Donohue [2:20 PM]
Ok, so, it sounds like you have some docs to update & schema updates. After that, I'm assuming we might be
ready to look for reviewers?
(by docs, I mean some general notes about migration)

Mark Wood [2:20 PM]
Yes. I think we agreed that this PR is done when fresh_install works and the schema updates are in.

Tim Donohue [2:22 PM]
Ok, sounds good. Feel free to ping us then when this is ready for review. Any general notes/docs you can draft
would also be helpful (so we can all start to get our minds around the big picture for the migration)

Mark Wood [2:22 PM]
OK.

Tim Donohue [2:23 PM]
Sounds like we're done with that topic then. Moving along...
Any Docker updates to share, @terrywbrady? https://wiki.duraspace.org/display/~terrywbrady
/DSpace+Docker+and+Cloud+Deployment+Goals

James Creel [2:24 PM]
I mean to review a PR this week -although I'm new to Docker (having used it only at a Fedora camp once)

Terry Brady [2:24 PM]
I am still waiting on reviews. I really want to merge the docker config changes before we plan the webinar.
Thanks @jcreel256

James Creel [2:24 PM]
The site wants me to register - is that legit?

Terry Brady [2:25 PM]
I created a dockerhub account. Eventually you will want to push images, so it is legit

James Creel [2:25 PM]
Ok, understood

Terry Brady [2:25 PM]
Give me a shout if you want a walk through. Ultimately, I think the new instructions will be much easier to follow.
I have one tiny PR that will make the DSpace 7 images easier to use: https://github.com/DSpace/DSpace/pull/2337

Tim Donohue [2:27 PM]
Ok, so, sounds like no other major updates there.
I can give that PR a :+1:

Terry Brady [2:27 PM]
Thanks @tdonohue
This issue led to my questions about setting the env variables with dots in their names.

Tim Donohue [2:29 PM]
Ok, so it sounds like we've wrapped up our topics for today. Are there other topics for discussion that anyone would like to bring up?

Terry Brady [2:29 PM]
I had hoped to resolve the issue by setting an env variable. If I get some time, I want to see why apache commons is not reading those vars correctly.
No other issues here

Mark Wood [2:30 PM]
If a shell ever gets to handle those names on their way from Docker to DSpace, it may be doing Bad Things to them.

Terry Brady [2:31 PM]
@jcreel256, this presentation might also help with the Docker PR: https://gitpitch.com/DSpace-Labs/DSpace-Docker-Images/helper_cmds
GitPitch
GitPitch Slide Deck
Modern Slide Decks for Developers on GitHub, GitLab, and Bitbucket.

Terry Brady [2:31 PM]
(The presentation itself is now part of the PR)
This will eventually become the presentation that will be used at the DuraSpace webinar.

Tim Donohue [2:33 PM]
Any other discussion topics / questions for today? If not, I'm wondering if we should wrap up early & give you back some of your day

Terry Brady [2:34 PM]
Sorry... one quick item
The community leader for Eclipse Che reached out to both @Patrick Trottier and I to learn more about how DSpace uses the Che editor.
That is the "eclipse for the cloud" that was part of Codenvy and is now part of openshift.io.
I indicated that I had hit a dead end with Codenvy, but I said I would be eager to share more after I get a chance to try out the newer version of Che. If you hear of anyone else who is looking at cloud based IDE's for DSpace, let me know.

Tim Donohue [2:37 PM]
Sounds good, @terrywbrady. I'd be interested in hearing updates on your discussions with them.
Has the payment model for Eclipse Che changed after moving to openshift.io? Just curious if things are mostly the same as at Codenvy, or if we know

Terry Brady [2:40 PM]
I think Che is open source, but you need to host it somewhere. There is a free tier at che.openshift.io. I will be curious to see how well it handles DSpace. On Codenvy, the size of our generated application killed the IDE server once all the jars were loaded. (edited)
The free tier had 3 license agreement docs that I have not yet digested.

Tim Donohue [2:41 PM]
I see. OK
Final call then, any other topics for today?

Ok, not hearing any. As noted at the beginning of the meeting, feel free to bring topics to this meeting (or pass them my way). I'm always glad to get something new on the agenda, or dig deeper on a topic
Thanks all! We'll consider the meeting complete. Have a good rest of the week!

Mark Wood [2:43 PM]
Thanks!

James Creel [2:44 PM]
Take care, all

Terry Brady [2:44 PM]
have a good week