DevMtg 2018-12-05

Developers Meeting on Weds, December 5, 2018

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)
    - Please note that all meetings are publicly logged

Agenda

Quick Reminders

Friendly reminders of upcoming meetings, discussions etc

- DSpace 7 Working Group (2016-2023): Next meeting was today (no meeting tomorrow). After that, we meet next week, Thursday, Dec 13 at 15:00 UTC.
- DSpace 7 Entities Working Group (2018-19): Next meeting on Tues, Dec 18 at 16:00 UTC

    - Yesterday's meeting notes/video at 2018-12-04 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.

Tim is only able to attend the first 30 minutes of this meeting. Others can continue discussion after that, as needed.

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
    a. PR https://github.com/DSpace/DSpace/pull/2058
4. DSpace Docker and Cloud Deployment Goals (old) (Terrence W Brady)
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

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 [2:00 PM]
@here: It's DSpace DevMtg time. Agenda at: https://wiki.duraspace.org/display/DSPACE/DevMtg+2018-12-05

Terry Brady [2:00 PM]
hello

Tim Donohue [2:00 PM]
Let's do a quick roll call to see who's here today

Mark Wood [2:00 PM]
Me!

Tim Donohue [2:01 PM]
As a reminder, I'm only "here" for the first 30mins. I can catchup with any discussions after that, but I'll
have to depart pretty promptly
So, we'll jump on in to topics.
DSpace 7 Entities WG met yesterday. Recording/notes at: https://wiki.duraspace.org/display/DSPACE/2018-12-
04+DSpace+7+Entities+WG+Meeting
DSpace 7 team met *today* (no meeting tomorrow). We mostly updated the Planning spreadsheet to detail which
features will be in the 7.0 "Preview" (in Jan/Feb) and which will wait for "Beta" (in April): https://docs.
google.com/spreadsheets/d/18brPF7cZy_UKyj97Ta44UJg5Z8OwJGi7PLoPJVz-g3g/edit#gid=0
That's it for DSpace 7 updates today. And I have no DSpace 6.x updates to speak of at this time.
So, we can move along to other discussion topics on our agenda
Next on the agenda is updates on our Solr Upgrade work: https://wiki.duraspace.org/display/DSPACE
/Upgrading+Solr+Server+for+DSpace (and PR#2058)
Any updates / discussion topics to share here, @mwood or @terrywbrady?

Mark Wood [2:05 PM]
I'm still trying to figure out that single test failure, which finds records that should not match the query.

Terry Brady [2:06 PM]
I have the following PR which provides a preliminary upgrade of the 4 schemas: https://github.com/DSpace-Labs
/DSpace-Docker-Images/pull/65

Tim Donohue [2:06 PM]
I recall there being a "clue" that date queries/facets work a bit differently now (from @terrywbrady on #dev)?
Did that clue help in any way?

Terry Brady [2:06 PM]
The search schema definitely needs more work.

Mark Wood [2:07 PM]
I'm still trying to work out *how* they work differently.

Tim Donohue [2:08 PM]
oh, ok. I was hoping maybe that'd be the start of the "breakthrough" we needed in this area

Terry Brady [2:08 PM]
I presume that either (1)date facet syntax has changed or (2)the json element names for facet results have
changed. I did not have a chance to look at it in any more detail.

Mark Wood [2:09 PM]
This is the query that is returning things which don't match the constraints. It's pretty simple.
        getClient().perform(get("/api/discover/search/objects")
                                .param("query", "dc.date.issued:2010-02-13"))

Terry Brady [2:10 PM]
That is a metadata search not a date faceting search so we may not have had a breakthrough...
I think the work that I did on the server side will be helpful. I will be looking for some guidance on when
/how that should be incorporated into the code base.
Sorry I missed the DSpace 7 meeting date change today. Is that team expecting to see these Solr changes?

Tim Donohue [2:12 PM]
I wonder if there's a way to test the query more manually? Take @terrywbrady's server side work to look at
date queries in general to validate no changes in behavior in metadata searches?

Mark Wood [2:13 PM]
I am looking for any way to get more information, so I'll try out the new schema.

Terry Brady [2:13 PM]
There is a change in the date datatype in the schema, so that could have some effect.

Tim Donohue [2:13 PM]
@terrywbrady: the DSpace 7 team is aware that Solr changes are happening. But, they are not "waiting on them" in any way. As of now, I'm hoping Solr changes will be mostly "invisible" to the REST API (and above) -- i.e. they'd just mostly affect the underlying Java API

Mark Wood [2:13 PM]
The new *Point matcher instead of the old *Trie matcher, yes. It *should* make no difference....

Tim Donohue [2:14 PM]
That said, as I noted in today's DSpace 7 meeting, it'd be *nice* to get these Solr changes ready in time for the 7.0 Preview release (in late Jan / early Feb), as the sooner we get Solr changes done the better
So, in terms of stuff to help DSpace 7, Solr upgrade is a high priority for these meetings.

Terry Brady [2:15 PM]
Also it will change the runtime configuration to support the external service.

Tim Donohue [2:16 PM]
@terrywbrady: yes, true. I was just noting the DSpace 7 meetings mostly concentrate on REST API & Angular UI, and I'm not sure Solr upgrade will affect either of those (if so, we'll need to pass it along to DSpace 7 team to prepare)
In any case, what can we do to move this forward? Is the next step to try and setup a Solr v7 instance, populate with some data and test out some date metadata queries? Do we know how to do that easily?

Mark Wood [2:18 PM]
Even if it isn't easy, I need to understand how the new arrangement will work in production.

Tim Donohue [2:18 PM]
True. I guess I'm just asking if we have a next step...or are there major "unknowns" here

Mark Wood [2:18 PM]
I think that's probably a good thing to do next: install and try out.

Terry Brady [2:19 PM]
My current thought is that we will prepare 4 empty repos in the new solr install. The only code override for those repos should be the schema file.
That is the process that the Dockerfile is performing.

Mark Wood [2:19 PM]
I was sidetracked for a while, trying to figure out why none of my logging config changes were helping.
There's a PR to show that I finally found *that*.

Terry Brady [2:20 PM]
If it is useful to you all, I can push up the dspace/dpsace-solr image to dockerhub.
If you are unlikely to use it, I will wait for the schemas to stabilize.

Tim Donohue [2:20 PM]
@mwood: if merging your logging config changes quickly helps, I think we can immediately merge this PR: https://github.com/DSpace/DSpace/pull/2279 It's very tiny/obvious

Mark Wood [2:21 PM]
Only needs one more approval....

Tim Donohue [2:22 PM]
Regarding docker stuff, I'd leave that up to you two. I won't be able to spend any significant time looking at this issue, so I'm considering it on your plate(s). I'm just here for support/encouragement / to bounce ideas off of

Terry Brady [2:23 PM]
I approved that PR.

Mark Wood [2:23 PM]
I'd like to be using the same schema changes, so anything that makes them easy to get would be helpful.
Thanks!

Terry Brady [2:24 PM]
I will help as I can on this. I have a small window of time between semesters to try to do our DSpace 5->6 upgrade. That is going to take up a good amount of my time this month.

Mark Wood [2:24 PM]
OK, 2279 is merged.

Terry Brady [2:25 PM]
Will our DSpace install process create the solr repo directories, or would we ask users to do that manually?

Mark Wood [2:25 PM]
OK, understood @terrywbrady. I'll be OOO from the 17th, but if I can answer a question, ask -- I'll be looking in.
I think that to start with we should just tell the installer (person) what to copy and where it goes (relative to the Solr installation).
Poking files into a separate Solr instance with Ant sounds like a bad idea. If we want to mess with Solr programmatically, we should use the APIs that it exposes.

Terry Brady [2:27 PM]
Will we have a recommended directory such as [dspace-install]/solr7 to make it clear that it is not the old repo dir?

Tim Donohue [2:27 PM]
I don't know that we've figured out the entire Solr install or upgrade process here. If there are opportunities to streamline some of it, we should take it. But, we'll have to figure out what is reasonable and what we just have to document as part of the Solr setup

Mark Wood [2:27 PM]
What would that directory be? I don't think we should tell folks where to install Solr.

Terry Brady [2:27 PM]
If you find some good API calls, I would be interested to know the following

1. create an empty repo via API
2. overwrite or modify the schema for that rep via API (edited)

Mark Wood [2:28 PM]
Collections API.
Schema API.
But I think we might leave that stuff for DSpace 8.

Tim Donohue [2:29 PM]
These are all good questions, but I think the first step is to get Solr 7 simply working with DSpace 7. Then we can figure out how to more easily install / configure it. Until we get the former setup, it's hard to streamline/test the latter.

Terry Brady [2:29 PM]
We are not saying where to install solr, but we might want to recommend where to host the repos.

Mark Wood [2:29 PM]
That might be difficult. Solr thinks it knows where they all are.
There *may* still be a way to tell it to look elsewhere, but it's definitely not best practice anymore.

Terry Brady [2:30 PM]
Is "Collections API" the name of a solr api? Do you have a link for it?

Mark Wood [2:31 PM]
Yes. Not at the moment; I've been using the PDF version of the Solr doco.

Tim Donohue [2:31 PM]
Unfortunately, I'm going to have to leave this discussion shortly. But, it sounds like we know the immediate next steps -- try and get a Solr 7 setup, test out some "date" related queries to see if we can figure out what is going on with our Unit Tests

Mark Wood [2:31 PM]
Yes.

Tim Donohue [2:31 PM]
Solr Collections API: https://lucene.apache.org/solr/guide/7_3/collections-api.html

Terry Brady [2:32 PM]
Nice... I will want to replace the logic in the dockerfile to use that to create the repos

Tim Donohue [2:32 PM]
Solr Schema API: https://lucene.apache.org/solr/guide/6_6/schema-api.html

Mark Wood [2:32 PM]

I think Collections is only available in SolrCloud mode.

Tim Donohue [2:32 PM]
(Those are the two Solr APIs mentioned above)

Terry Brady [2:33 PM]
I'll do a bit of experimentation and let you all know what I find.

Tim Donohue [2:34 PM]
Ok, I'm heading out now. Will catch up later on any further discussion. Sorry @terrywbrady that we didn't get to Docker stuff (https://wiki.duraspace.org/display/~terrywbrady/DSpace+Docker+and+Cloud+Deployment+Goals), but I'm glad to see folks already adding comments to that wiki page

Mark Wood [2:34 PM]
OK, @tdonohue, enjoy the rest of the day!

Tim Donohue [2:34 PM]
:wave:

Terry Brady [2:34 PM]
Take care @tdonohue.   We can discuss that at the next mtg.

Mark Wood [2:36 PM]
The basic idea behind Collections and the managed schema seems to be that you forget Solr even has files. Configuration sets go into Apache Zookeeper, and Solr puts cores where it wants them when you ask it to create a collection.

DSpaceSlackBot (IRC) APP [2:36 PM]
tdonohue has quit the IRC channel

Terry Brady [2:37 PM]
For your testing, it should be possible to spin up a Solr 7 instance with the new schema. Retrieve a json representation of a solr doc. Add it to Solr7 via Solr Admin. Compare query results between the 2 instances.

Mark Wood [2:38 PM]
That sounds reasonable. I actually have a totally empty Solr 7 instance running right now.

Terry Brady [2:38 PM]
It makes sense to pretend that the filesystem does not exist.
One frustration I had with Solr7 Admin is that if you try to define a new core, it asks you do provide a directory. I wanted to simply tell Solr admin the name of my new core and have it do all the magic.
If you find a way around that, let me know.

Mark Wood [2:39 PM]
Are you running a standalone instance or a (1-node) Cloud? That may be the difference.
I think Cloud gets upset if you *do* try to mess with its files.

Terry Brady [2:40 PM]
I am running a standalone instance built from the base docker image I found.
https://github.com/DSpace-Labs/DSpace-Docker-Images/blob/extsolr/dockerfiles/dspace-solr/Dockerfile#L1
Which uses: https://hub.docker.com/_/solr/

Mark Wood [2:45 PM]
I don't really know enough yet to talk about why Solr does this or that. I only just received a book newer than Solr 1.4, and a lot has changed....

Terry Brady [2:46 PM]
I just grabbed a solr image from Docker and started hacking. I am unclear if that image can be run in a SolrCloud mode.

Mark Wood [2:46 PM]
IIRC there's an option you give to the startup script that does that.

Terry Brady [2:47 PM]
I will look around for that.

Mark Wood [2:47 PM]
bin/solr -h will show a (rather complex) example.
Be warned that all the advice I've found on the web so far seems to assume we're running a giant server farm, not a 1-node cloud.

Terry Brady [2:51 PM]
Everything in the IT world seems to have that same assumption...
I am going to sign off for today. i will let you know what I learned. Would it be worthwhile to have a
dedicated channel in Slack for the solr work?

Mark Wood [2:53 PM]
There probably isn't enough traffic for it. We can chat in #dev ?

Terry Brady [2:54 PM]
That works for me. If we start to generate lots of chatter, we can reconsider.

Mark Wood [2:55 PM]
OK.
I guess we should declare the meeting closed. If there is no other business?
Meeting adjourned.