

Proposal: Create Cloud Instances for DSpace Active Branches

- [Summary](#)
- [What is required?](#)
- [Related Conversations](#)
- [Truncated conversation from the DSpace Developer Meeting 2017-08-09](#)

Summary

The DSpace Demo Site (demo.dspace.org) should always run the latest, stable version of DSpace since this site is used for evaluating DSpace as a platform.

There would be significant benefit to running a cloud instance of the actively supported branches of DSpace (currently `dspace-4_x`, `dspace-5_x`, `dspace-6_x`). This would ensure that anyone in the community would have access to latest code under consideration for a release.

In particular, this would allow the DCAT team to conduct a testathon on a code base prior to the creation of a release/candidate.

The DSpace 5.7 and DSpace 6.1 releases incorporated last-minute changes that unexpectedly introduced bugs into the system that were only discovered by running end to end functionality in DSpace. Because of these bugs, new releases will need to be created. The availability of a widely-accessible stable test environment might have prevented these bugs.

In the past, for major DSpace releases, a release candidate is generated. In order for the release candidate to be tested, host institutions needed to install, deploy and test the release candidate on their own network. This process can provide a valuable verification of installation and migration instructions, but it also imposes a significant burden on each host institution that participates in the test. This burden lengthens the window of time required to conduct system testing.

What is required?

- Donor to fund the creation of 3 new publicly-accessible DSpace instances
 - If cost is prohibitive, we could prioritize the DSpace 6 instance.
 - How much would this cost?
 - What constraints might need to be placed on the instances (# items, repository size)?
- Volunteer to manage the cloud instances within the operational budget (bandwidth, compute, etc)
- Creation of an automated code deployment process to update the instance based on changes to an active code branch.
 - Changes could deploy on merge or could be re-deployed on a set schedule
- Creation of an automated data refresh process to keep content fresh and stable
 - Similar to what is done for demo.dspace.org
- Creation of a mechanism for executing command line processes
 - In the most recent 6.1 release, bugs were introduced into these batch processes. As part of a testathon, users should be able to validate the results of specific command line tasks.

Related Conversations

- [DCAT Meeting August 2017](#)
- [DSpace Developer Meeting 2017-08-09](#)
 - The log of this conversation was truncated. See notes below.

Truncated conversation from the DSpace Developer Meeting 2017-08-09

Tim Donohue [9:00 AM]

We are at the top of the hour here.. so, I'd like to call the end to the "formal meeting" (and it sounds like we've hit all the main next steps on a 6.2).

That said, I saw @terrywbrady noted it'd be good to talk about "creating a cloud instance running the latest release" (discussion from DCAT this week)

Terry Brady [9:01 AM]

Would it make sense to save that for next week's meeting

Tim Donohue [9:01 AM]

If folks are available a bit longer, we could continue that discussion for a bit. Otherwise, we can touch back on this next week?

Hardy Pottinger [9:01 AM]

I'd like a definition of terms, what do we mean?

Terry Brady [9:02 AM]
demo.dspace.org will run the latest, stable release.

Mark Wood [9:02 AM]
Please carry agenda item 2 (use of DB connections) over for next week.

Tim Donohue [9:02 AM]
Here's what DCAT / Bram is talking about: <https://wiki.duraspace.org/display/cmygp/DCAT+Meeting+August+2017?focusedCommentId=87468402#comment-87468402>

kompewter (IRC) APP [9:02 AM]
[DCAT Meeting August 2017 - Community Groups - DuraSpace Wiki] - <https://wiki.duraspace.org/display/cmygp/DCAT+Meeting+August+2017?focusedCommentId=87468402#comment-87468402>

Terry Brady [9:02 AM]
We will have a new cloud instance that runs the latest active branch such as dspace-6_x

Mark Wood [9:02 AM]
Oooh, continuous delivery.

Terry Brady [9:02 AM]
It would be available for pre-release testing, DCAT testathons, etc

[9:03]
The tricky issue is do we need to support multiple branches (5x, 6x, 7x) or do we start with one instance

[9:04]
Ideally, this could help us catch the last-minute bugs introduced into 5.7 and 6.1

Tim Donohue [9:04 AM]
Yes, a form of continuous delivery specifically as a test environment for latest changes, etc

Hardy Pottinger [9:04 AM]
OK, so, you'd like a set of demos, one for each maintenance branch, with CD tooling to stand them up on every commit to the maintenance branch?

Tim Donohue [9:04 AM]
And keeping that *separate* from demo.dspace.org (which would be kept "stable" for demos/trying out DSpace)... the testing would be elsewhere, so that demo.dspace.org would never be perceived as "unstable"

Terry Brady [9:05 AM]
On every commit or on a regular schedule

Mark Wood [9:05 AM]
Multiple instances is easy once you work it out. I have some shell scripts that help.

Terry Brady [9:06 AM]
I know this would not be cheap or trivial to implement, but I feel like we pay the cost over and over again by not having this infrastructure

Hardy Pottinger [9:06 AM]
we're doing work on this kind of thing at UCLA Library, and I have an idea of how to do it with DSpace, though

what I'm working on will use GitHub tags as a trigger, but you could always make it trigger on commit

Tim Donohue [9:07 AM]

As I noted in both the DCAT call, and in Bram's comments, I fully support this idea. But, it's not something I have any expertise in (nor do I see time for it in my immediate future). So, I'd need someone else to take the lead here. (Similarly, see my comments, Fedora has U of Maryland providing this idea for their project via a hosted Jenkins & Sonar)

Hardy Pottinger [9:08 AM]

here's the gist of it: Travis-CI builds *all* the artifacts, and they just sit out there in Travis-land, Travis *can* write to Amazon S3, and Ant can sync from Amazon S3... so... make Travis dump all the artifacts into a bucket, and make Ant sync the latest bucket before it runs Ant update. (edited)

Tim Donohue [9:09 AM]

As of yet, DuraSpace unfortunately doesn't have a shared infrastructure for this sort of thing (I will admit we've talked about it internally, but there's no funding behind it yet)

Terry Brady [9:09 AM]

Unfortunately, I do not have the expertise for this either. Since this would cost money to implement, it is beyond my pay grade. I plan to ask my boss to suggest this to the DSpace steering committee.

Hardy Pottinger [9:11 AM]

I'm already working on the S3 deploy idea I just outlined, would be happy to contribute it back, so mostly we're just talking about an S3 bucket and three micro instances on AWS

Tim Donohue [9:12 AM]

So, I guess the summary is here... we are looking for folks with ideas and/or proposals to make this happen (and feel free to draft something up on wiki). Currently, as of today, there's no funding behind this (unless DSpace Steering makes that happen somehow) and no centralized DuraSpace infrastructure.

[9:13]

So, we'd be reliant on finding someone else to help host this. I'd be glad though to help make this happen (i.e. give some time to help do setup, etc as needed), but likely couldn't take the lead.

Terry Brady [9:13 AM]

Good summary. I will start a proposal page on the wiki.

Tim Donohue [9:13 AM]

Thanks @terrywbrady!