

2021-07 DuraCloud Community Sprint Notes

Call Details

- Time: 11:00pm ET / 10:00am CT / 9:00am MT / 8:00am PT
- Day: Tue/Thurs
- Join Zoom meeting: <https://lyrasis.zoom.us/j/92975739532?pwd=M2pkTVhtaEJtWmZmbTdGRk1lSCs2QT09&from=addon>
- ★ - Indicates who will be taking minutes

Participants

- Bill Branan ★
- Nicholas Woodward
- Andy Foster
- Danny Bernstein

Activity

- Kanban Board
- Backlog

Notes

Day 2 (2021-07-06)

- Interest in being able to handle content already in S3/Swift content
- DURACLOUD-1316
 - Interest:
 - Andy - definite interest, but it'll be challenging to ensure everything works.
 - How do we prevent overwrites (or simultaneous writes)?
 - Perhaps some type of locking mechanism?
 - Could have a toggle that only allows writes via either S3 or DuraCloud API
 - S3 does provide a way to send SNS notifications when content lands. Could use this to add a message to the SQS monitored by the Mill
 - Fedora has a similar need to allow direct writes to OCFL then notifying Fedora
 - Would need a way to ensure we get checksums properly and that the manifest/audit logs
 - Would need to validate
- Top priorities
 - Andy: 1258 is important, Mill using quite an old version. 1257 is essentially done.
 - Danny: 1290 (docker/terraform), setting us up for long-term maintenance
 - Nick: Snapshot totals (most questions), 1150, 1279, 1301
 - Bill: Snapshot totals
- Plans for this week
 - Andy: 1257, 1258, 1290
 - Nick: 1069, 1268
 - Danny: 1290
 - Bill: 1292, 1278
- Github Actions
 - In DuraCloud, we only merge to master on release
 - The push to sonatype can push it all the way through
 - Most projects tend to use tags
- Should we consider changing our branch structure? Have master being the leading edge?
 - One option: develop main and master release
 - It's helpful to have a branch where we're able to merge everything in and do testing
 - Rename master to main?
 - Develop: The integration point for development
 - Master/Main: Merges from develop after testing (usually a release)

Day 4 (2021-07-08)

- Status
 - Bill: 1316 and 1292
 - Danny:
 - Nick brought up issue on the Mill. Danny added fix PR, has been merged.
 - Focused on Terraform/Docker for the Mill. It's coming along
 - Plan to have one Dockerfile in the mill that will get built and pushed on release. Will include all the java jars related to the mill. A script will run the type of node that we want.
 - A configuration setting (node type) will be passed in to say what that node will do
 - Will also be an AWS bucket that includes all the mill configuration

- Outside of the docker container the mounting for EFS will be handled. Terraform will handle mounting EFS and set up the path mapping.
 - There will be a script to push to dockerhub, creating the image with Github Actions
 - Nick:
 - Working on 1268, Retrieval Tool failing.
 - Preferred solution: When dealing with files list, get the space manifest
 - Andy
 - Made a token start on updating the methods for the mill
- Topics
 - Make the move from master to main - Action for Bill
 - Check on whether there can be a symlink? Are these auto-forwarded?
 - Github actions to push release artifacts should work off of tags (rather than merges to main)

Day 7 (2021-07-13)

- Status
 - Danny:
 - Have a docker container working for the workman piece of the Mill
 - Structure in place for running the various node types
 - Will continue to use cloud-init for launch
 - The program will be invoked by terraform, passing in user parameters, which will generate the cloud-init scripts
 - Terraform will generate the auto-scaling groups
 - cloud-init has been pared down considerably (attach EFS, run docker)
 - Each node type will run have a distinct cloud-init
 - Within the docker container there will be scripts based on the node type
 - Most of the work so far has been in the mill project, will push so Andy can take a look
 - Feedback welcome!
 - Should be able to run the Docker container independently (locally)
 - Andy
 - Crossed off 1257 and 1258
 - Picking up 1304
 - Can't make Thursday call
 - Mohana focused elsewhere - not much traction on DuraCloud
 - Nick
 - PR for 1268 in progress (Retrieval Tool for list file)
 - Probably worth adding a comment in the docs to say how it handles
 - 1069 in process
 - Looking at an approach
 - 1150 also in process - snapshot totals in the Bridge API
 - Do we need to have a client parameter
 - Bill
 - Continued work on dependencies - 1292

Day 9 (2021-07-15)

- Status
 - Danny:
 - The Terraform piece is mostly done at this point
 - Project (duracloud-aws-deployer), with shared and mill projects
 - shared brings up vpc, subnets, and db
 - mill does the mill setup
 - queues are being set up, along with the audit worker
 - still need to publish the docker image
 - each auto-scaling group has a launch config and auto scaling policies (all this done for one so far)
 - Publishing to DockerHub, requires setting up an org in DockerHub (Bill to look at this), maybe use ECR
 - Terraform configs sit in a new repo
 - Potential to import existing AWS components rather than creating everything new with Terraform
 - Perhaps just import the database. May need to also bring in the VPC and subnets into
 - Nick
 - 1268, test is taking some time
 - 1150, did total number of snapshots (1150 has work on snapshot and duracloud, two branches)
 - Looking to do the required and optional endpoints
 - using /snapshot/total/(endpoint)
 - Bill
 - Almost done with dependencies
- Release planning
 - Nick may do some work next week
 - Planning to do release work for 7.1 week of August 16
- Sprint retrospective
 - Danny
 - Plus: Having the sprint time blocked out helped to focus and get stuff done
 - Plus: Didn't need a lot of support from the team
 - Having this on the calendar to be able to block off the time
 - Deltas: Conflicts with Fedora meetings, need to make sure there aren't overlaps.
 - Delta: Better to kick off on Monday, wrap up on Friday. Maybe: Mon, Wen, Mon, Fri
 - Nick
 - Delta: Meetings on Mon/Wed/Fri would be better
 - Plus: Blocking off the time well in advance is incredibly helpful
 - Delta: Becoming more familiar with the application. To keep information fresh, would be good to get into the code more often

- Bill
 - Plus: Two week sprint is better than one (couldn't really be more than two)
- Next sprint
 - Delta: Having everyone focused
 - Delta: More conflicting meetings than I would like
 - One more this year?
 - Feels like this is a possibility
 - Sprint: Late-Oct, Early-Nov
 - One week sprint? Maybe two?
 - Plan to get this on the calendar during the August contributor call

Actions