2018-06-26 DuraCloud Contributor Call

Call Details
- Time: 11:00am Eastern Time (find your time here)
- Join Zoom meeting: https://duraspace.zoom.us/j/814345256
- Or iPhone one-tap: US: +16468769923, 814345256# or +16699006833, 814345256#
- Or Telephone: US: +1 646 876 9923 or +1 669 900 6833 or +1 408 638 0968 (Meeting ID: 814 345 256)
- International numbers available: https://zoom.us/u/YQRCN
- * Indicates who will be taking minutes

Attendees
- Andrea Bollini (4Science)
- Bill Branan
- Courtney C. Mumma
- Danny Bernstein
- Heather Greer Klein *
- Nicholas Woodward
- (your name here)

Agenda
(If you have an agenda suggestion/addition, please leave a comment!)

a. Communicating as a group: Slack invitations sent out. Anyone else that should be invited to the channel?
b. Status of current/planned development
   i. DuraSpace
      1. Multipart upload
      2. Streaming (July-Aug) and transcoding (Aug-Sept)
      3. New Kanban board to track activity
   ii. 4Science
      1. S3 One as a distinct storage provider
      2. Other storage providers (ExoScale, CynnySpace, others?)
      3. Cross region duplication
   iii. TDL
   c. Secondary storage providers
      i. Secondary provider use cases - use of secondary providers now and user expectations
      ii. Secondary provider vs primary provider (which providers fit into each category): S3, Glacier, Chronopolis, DPN, S3 One, ExoScale, CynnySpace
      iii. Considerations/Requirements when adding a new provider
   d. TDL and 4Science experience with use and implementation of DuraCloud
      i. Pain points
      ii. Customer feedback/response
      iii. How could this be made easier in the future?
   e. Customer requested features (DuraSpace, TDL)
      i. Features that have been requested
      ii. For each feature, assign someone to ensure the feature request is captured in JIRA
   f. Operational needs
      i. From last time: better release notes, more informative error handling, better sync tool tooltips, options to customize emails for registration in management console.
   g. Approach for reviewing and prioritizing JIRA list
   h. Activity planning for community sprint in October

Reminders
1. Sprint planning day: October 5th
2. Development sprint: October 15-26

Minutes
1. Communicating as a group: Slack invitations sent out. Anyone else that should be invited to the channel?
   a. Everyone usually on the calls is there.
2. Status of current/planned development
   a. DuraSpace
      i. Multipart upload: started primarily on the retrieval side, added ability to pull content in pieces. Still quite a bit of work to be done on the API for multi-part. Bill has had limited time for it
      ii. Streaming (July-Aug) and transcoding (Aug-Sept) sprints coming up for customer paying for streaming updates. Will ensure they can integrate their systems as well.
iii. **New Kanban board to track activity** to show work in progress. Different from other Kanban boards even in JIRA, separates backlog to another page. Backlog is in menu on the lefthand side. Can pull from backlog to in progress. Can add tickets or pull in existing tickets for work that is being done elsewhere (4Science, TDL, etc.)

b. **4Science**
   1. 4Science working on S3-1, it is working in DuraCloud Europe.
   2. Some work is overlapping with what DuraSpace is doing to remove old providers.
   3. Need to determine what needs to be repaired that might have been missed with removing old providers. DuraSpace should wait for 4Science to finish their work first.
   4. New class, S3 secondary provider, to allow another provider in a different region (for DC EU, S3 Ireland is replicated to S3 Paris).
   5. Working on an S3 compatible provider. Still working through issues for this, figure out a way to replace cloudwatch, another provider does not have equivalent.
      a. Bill: Glacier has the same issue, it is pulled from the mill manifest
      b. Andrea working to come up with another solution
      c. **Andrea will create tickets** ([https://jira.duraspace.org/browse/DURACLOUD-1198](https://jira.duraspace.org/browse/DURACLOUD-1198), [https://jira.duraspace.org/browse/DURACLOUD-1199](https://jira.duraspace.org/browse/DURACLOUD-1199), [https://jira.duraspace.org/browse/DURACLOUD-1200](https://jira.duraspace.org/browse/DURACLOUD-1200))

c. **TDL**
   i. TDL is working on a new interface for their limited service. Another model for replication task suite. Can be shared.
      1. Have been working to set up sprints now that they have a full staff. Nick will work on sprint in the fall.
      2. **TDL wrote a letter of support** for a Canadian consortium requesting a grant to implement DuraCloud in Canada. Would likely provide code for alternative implementations.

3. **Secondary storage providers**
   a. Secondary provider use cases - use of secondary providers now and user expectations
      i. Primary at this point is always S3. DPN or Chronopolis can also be primary. One DC customer uses both Chronopolis and S3.
      ii. Glacier is pretty much always set up to go into secondary storage.
      iii. Not easy to restore content from Glacier. Not notified when the content is available. A different Glacier provider, Glacier Vault, uses different API. In Europe it is challenge to find different providers from Amazon, and this is why they are working on support for S3 in different regions.
   b. **Secondary provider vs primary provider** (which providers fit into each category): S3, Glacier, Chronopolis, DPN, S3 One, ExoScale, CynnySpace
      i. S3-1 can work as both primary and secondary storage. Thinking of having S3-1 for users who have other replication elsewhere.
      ii. Makes sense to have more options. DuraCloud has not had primary outside of Amazon. Initial drop of content is always in S3.
      iii. Libraries for open stack have not been good in our experience
      iv. Costs of streaming content would come through outbound from EC2 before it lands in CynnySpace
      v. TDL: users who have Glacier set up as primary storage providers and add DPN, etc run into issues creating spaces
      vi. Not everyone wants the replication. Some want the concept of a dual primary storage area. Something that could be made optional.
      vii. Likely that TDL will have more members who want Chronopolis for one set of content and DPN for another. Would be useful to manage this administratively.
      viii. **TDL folks will create a ticket to keep track of this need**
      ix. May be a good focus for the sprint in October

4. **TDL and 4Science experience with use and implementation of DuraCloud**
   a. **Pain points**
      i. TDL has compiled a list. Secondary storage provider issue; open sourcing the mill deployment scripts; sync tool error handling and clearing stored configs, tool tips; customization of emails; upgrade dependencies (Javascript library updating).
   b. **Client Javascript code is pretty old. Could take a lot of time for little value.**

5. **Approach for reviewing and prioritizing JIRA list**
   a. Doing work for our own environments, but there are existing items that would be a problem for all of us. How to resource getting that work done?
   b. How to prioritize? In this meeting, or outside this meeting?
   c. Voting on issues?
   d. Can ID higher priority issues that are on your list
   e. Need to know what items are most important
   f. **Courtney: this is similar to Datapoint, and how they work**
   g. Help developers know what is effecting customers
   h. Can break out tickets or drag to the top of the backlog? Tag/label tickets?
      i. Can sort by the tag/label
      j. Other calls in the future will likely be shorter (30 min or so)

**Actions**
- Andrea will create tickets for existing 4Science work
- TDL folks will create a ticket to keep track of the need for dual primary storage and dual Chronopolis/DPN storage
- Bill will determine if creating labels would be an effective approach for prioritization