

2017-03-20 Performance - Scale meeting

Time/Place

- Time: 11:00AM Eastern Time US (UTC-4)
- Dial-in Number: (712) 775-7035
 - Participant Code: 479307#
 - International numbers: [Conference Call Information](#)
 - Web Access: <https://www.freeconferencecallhd.com/wp-content/themes/responsive/flashphone/flash-phone.php>

Attendees

- [Andrew Woods](#)
- [Yinlin Chen](#)
- [Colin Gross](#)
- [Danny Bernstein](#) ★
- [Trey Pendragon](#)
- [Longshou Situ](#)
- [Joshua Westgard](#)

Agenda

1. Review last meeting actions: See a
2. Status of "many members" testing:

Unable to locate Jira server for this macro. It may be due to Application Link configuration.
3. Input on [AWS Credits Proposal](#) ([guide](#))
4. [Documentation for running tests and creating graphs](#)
5. Define tests of known pain points, e.g.
 - a. [Performance of "many members"](#)
 - b. Ugliness of pairtree resources
 - c. migration of PIDs from fcrepo3 to fcrepo4
 - d. Migration from fcrepo3 to fcrepo4 in general and in a reasonable amount of time for 1M+ objects
 - e. Loss of system dates, the migration of creation/modification times from fcrepo3 to fcrepo4
 - f. Intra-repository referential integrity can be a pain point, but it has benefits too

Minutes

Regarding the performance test summary report: [Nick Ruest](#) was waiting on [Colin Gross](#) to get the R scripts together. But Colin has generated the data. Analysis and summary of graphs is needed. We should come up with some questions contextualize the graphs? For example: How does performance change as resources increase? Then we could drill down onto each kind of resource - size, type, etc. Long term it would be helpful to create a template for the summary so that we can generate the summary automatically based on the summaries.

Questions / Advice for a State of Performance and Scale page:

[Joshua Westgard](#) would be interested in best practices for performance wiki page :

What are the different ways to set up a repository and what are the performance characteristics and performance mitigation strategies for each type.

How to protect against performance degradation as repository grows?

[Esmé Cowles](#) seconded that thought.

[Colin Gross](#) is working on a mock file system that will enable testing of very large files. The file system path of an item is based on the SHA-1. The metadata cannot go into the mock system. [Colin Gross](#) 's python based FUSE file system is not quite ready. He will let us know when it reaches 0.1 he will let us know.

[Esmé Cowles](#) is has tested Danny's performance changes. He has not seen the performance improvements. But thinks it may have something to do with large numbers of hash-uris stored on a node and possibly inefficiencies in hydra updates. Developing...

Actions

- [Nick Ruest](#) and [Yinlin Chen](#) to take above R-script output as starting point for performance test summary report. (Isn't it covered by https://github.com/fcrepo4-labs/fcrepo_perf_analysis/tree/master/dist already?)
- [Colin Gross](#) will finish up current GitHub issues: https://github.com/fcrepo4-labs/fcrepo_perf_analysis/issues
- [Andrew Woods](#) and [Yinlin Chen](#) will work on AWS grant proposal, deadline 03/31/17, will have a draft for review on next meeting: <https://aws.amazon.com/research-credits/>
- [Danny Bernstein](#) and [Longshou Situ](#) will have a [wiki page](#) about a documented process for end-to-end testing and reporting

- [Danny Bernstein](#) will create an outline of Dos and Don'ts and put out a message to solicit community input (cc: fedora-community list + Colin, Josh, Esme)