

2017-02-20 Performance - Scale meeting

Time/Place

- Time: 11:00AM Eastern Time US (UTC-5)
- 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](#)

Agenda

1. Review last meeting actions:
 - [Colin Gross](#) to transfer [fcrepo_perf_analysis](#) to fcrepo4-labs
 - [Andrew Woods](#) to post test outputs to S3
 - [Colin Gross](#) to push outputs from R-scripts to GitHub
https://github.com/fcrepo4-labs/fcrepo_perf_analysis/tree/feature/dist-upload/dist
 - [Nick Ruest](#) and [Yinlin Chen](#) to take above R-script output as starting point for performance test summary report
2. Review current GitHub issues: https://github.com/fcrepo4-labs/fcrepo_perf_analysis/issues
3. Infrastructure options - Research Grant
 - a. AWS Cloud Credits for Research <https://aws.amazon.com/research-credits/> Rewards varies from 1 to 2 years.
4. Reproduce new performance report?
 - a. If anyone what to know current Fedora 4.7 performance, we can refer him to see the performance report page.
 - b. [Performance and Scalability Test Plans](#)
 - c. Baseline performance on these six scenarios in a:
 - i. Select an instance type in AWS
 - ii. other options?
 - d. An Ansible tool to install Fedora 4 on the local machine or AWS.
 - i. <https://github.com/VTUL/fcrepo4-ansible>
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

Previous actions

1. Updated scripts to use perf.log instead of the csv
 - GitHub feature branch contains the generated reports
 - Hydra: would like a concise report
 - version of fedora
 - configuration of fedora
 - how tests relate to real-world questions
 - what these tests mean to the fedora development team

GitHub Issues

1. Colin has it covered

Infrastructure options

1. Need estimates of storage and instances
 - sometimes need 20-TB of data, for example
 - i/o latency
 - writes to /dev/null

- reads are based on hash of path of resource
- + Self and Yinlin to put in proposal <https://aws.amazon.com/research-credits/>

New performance reports

1. Running against 4.7.1
 2. Waiting on open actions from fcrepo_perf
 3. Need documented process for end-to-end testing and reporting
- + create new wiki page : danny

Actions

- [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