

2016-09-26 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

- [Nick Ruest](#)
- [Andrew Woods](#)
- [Yinlin Chen](#)
- [Daniel Lamb](#)
- [Colin Gross](#)
- [Danny Bernstein](#) ★
- [Bill Branan](#)

Agenda

1. Status of current testing
2. Create graphs and summaries of completed tests

```
https://wiki.duraspace.org/display/FF/2016-08-15+Performance+-+Scale+meeting#id-2016-08-15Performance-Scalemeeting-CurrentSummaries
```

3. Next steps
 - a. Finalize remaining tests?
 - b. Investigate other features: versioning? batch-ops?
 - c. Make call to community?

Minutes

- Status of Current testing
 - Nick will update tests results. 2 month test appears to have failed. Will run tests again on new equipment.
 - He will run RDF serialization improvements from Aaron Coburn on his new hardware.
 - Yinlin - 100K Items - 230MB files - 20 Mbs per client. Takes about 1 week.
- General agreement on the value of aggregating and summarizing the results of the tests that we have run so far.
- There is general agreement that it would be good to summarize relative improvements between runs of a given test (graphs in addition to any other details/observations)
 - Factors would be good to include in the summary:
 - Hardware specifics
 - Total execution time
 - Average response time over the course of the execution
 - Fedora version
 - Database type/specs
 - Client count
- Colin suggested it might be helpful to have a basic test to establish baseline conditions in the environment to account for variations in network performance characteristics, disk performance, etc.
 - Example script for gathering system info: <https://github.com/fcrepo4-archive/ff-jmeter-testResults/blob/master/gatherSystemInfo.sh>
- The team sees promise in expending effort to develop an automated system for performance tests that would
 - enable us to perform tests against on a consistent set of hardware hardware and network resources
 - automatically run the test suite against new tags / branches / forked repos?
 - focus on time-limited tests with known inputs and expected execution time framef.
- Aaron Coburn would like a test for understanding how memory is affected by specific kinds of serializations (Turtle and N-Triples) of RDF Sources and differing degrees of concurrency.

Actions

- Colin will look into putting together a script for baselining hardware and network characteristics to be factored into each test run.

- Nick: Add the log files to be added to the performance test results on the Test Plan page (<https://wiki.duraspace.org/display/FF/Performance+and+Scalability+Test+Plans>)
- Aaron will create a summary of what he would like to see out of a test.
- Danny: Summarize the existing test data.