

2014-03-06 - Fedora Committer Meeting

Time/Place

This meeting is a hybrid teleconference and IRC chat. Anyone is welcome to join...here's the info:

- Time: [11:00am Eastern Daylight Time US \(UTC-4\)](#)
- U.S.A/Canada toll free: 866-740-1260, participant code: 2257295
- International toll free: <http://www.readytalk.com/intl>
 - Use the above link and input 2257295 and the country you are calling from to get your country's toll-free dial-in number
 - Once on the call, enter participant code 2257295
- IRC:
 - [Join the #duraspace-ff chat room via Freenode Web IRC](#) (enter a unique nick)
 - Or point your IRC client to #duraspace-ff on irc.freenode.net

Attendees

- [Andrew Woods](#)
- [frank asseg](#)
- [Ed Fugikawa](#)
- [Stefano Cossu](#)
- [Nikhil Trivedi](#)
- [Greg Jansen](#)
- [Michael Durbin](#) ★
- [Mike Daines](#)
- [Ben Pennell](#)
- [A. Soroka](#)
- [Kevin S. Clarke](#)
- [Martin Dow](#)
- [Nigel Banks](#)

Agenda

1. Clustering status
 - a. load balancer routing and transactions
 - b. puppet-fcrepo planning
2. Authorization refactoring plan
3. Fedora 3 authorization to Audit datastream profile question
4. other...

Minutes

Recent work around clustering has been to support the goal of determining performance and improve performance by adding nodes. (higher ingest performance, greater user load)

Currently we've just been working to get clusters up and running in a consistent way.

Greg reports that...

- UNC's cluster is up and running, some details added to the wiki
 - Was using file stores for back end storage, ran several performance tests (though not with a larger number of files)
 - behind load balancer
 - when not using transactions, everything was fine
 - transactions didn't work...
 - possible solutions include:
 - hacking sessions using cookies or fancy routing and rewrite rules
 - use JTA transactions rather than sharing a session
 - hitting the first node in the cluster
 - transactions and threads all worked
 - eventually things broke and everything returns an error
 - Is the error reproducible? (Andrew)
 - Yes... in that the errors are constant now (possibly when using transaction)

Andrew summarized... number of issues around clustering

- transactions
- fixity (when segmented)
- setting it up is still hard (Esme, Scott are having trouble just getting it up)
 - Greg identified TCP issues and documented them, likely making future setups easier
 - Greg will reconfigure his cluster for levelDB, then go from TCP to UDP (which is more efficient)

- Greg asked whether the levelDB storage is supposed to be shared or independent between nodes, assuming the later.

Frank's cluster:

- setup is essentially like Greg's except that he's using UDP instead of TCP for discovery.
- He's been able to run a great number of ingests in sequence against a single node.
- Stability is not an issue, but performance is still not where it needs to be.
- has not tested with transactions
- contains 6 nodes
- network isn't a bottleneck
- Disk writes seem to be the bottleneck... though the amount of data written is quite low.
- Used load balancer (apache, modjk) in the past but last test was a single node with lots of threads.

Andrew wonders about the theory that performance could improve without sending requests to different nodes.

Frank found performance was worse in a previous test using a load-balancer

Puppet-fcrepo planning

Andrew wants us to get to a place where we have scripts (puppet) to easily bring up a cluster as a prerequisite for more cluster testing.

Greg feels we need requirements for what should be configured by the scripts
setting up TCP ping vs UDP ping needs work..

Scripts right now DO include install of java and tomcat, Scott was going to make it configurable whether it uses package management or downloads those resources directly...

Greg thinks we should make the master simple (using package management to require installation of java/tomcat) and if Scott wants to work on a branch to handle it specially.

Authorization Refactoring Plan

- Call for comments on the proposal, sign off on it?
- Make more information available through session attributes to PEP:
 - IP Address
 - a framework is in place to get stuff from HTTP Headers, though they still have to write a small class...
- Was some concern about high level (HTTP) concepts finding their way into lower level code... Andrew asks whether its possible to have the lower level classes/interface not depend on HTTP assumptions?
 - FedoraUserSecurityContext could just take an instance of Principle which would address this concern.
- This will be a Pull request soon... will lay the groundwork and future work will implement solutions to required use cases.
- Will write up draft of documentation to help users create their own implementations.

Fedora 3 authentication and audit...

User wishes to allow anonymous access to certain APIM methods... using the spring configuration.

The audit datastream doesn't seem to have a responsibility, is there a way to inject one... other than using the datastream dissemination end point.

Andrew's minor news update... Fedora users group in Washington DC on Monday the 10th.

[A. Soroka](#), [Andrew Woods](#), [David Wilcox](#), [Stefano Cossu](#) and [Michael Durbin](#) will be there.

Actions

-