2013-12-02 Sprint Planning Meeting

Date

Dec 02, 2013

Attendees

- Michael Durbin ⭐
- Andrew Woods
- A. Soroka
- Benjamin Armintor
- Chris Beer
- Eric James
- Unknown User (escowles@ucsd.edu)
- frank asseg
- Nigel Banks
- Osman Din
- Scott Prater

General

- Indicates who took minutes - ⭐
- Call-in: Google-hangout at:
  https://plus.google.com/hangouts/_/calendar
  /eW91cm1lZGlhc2hlbGGYuY29tXzVlYzdpNXQ2Z282dTdidHI4aTVrbGJxOTUwQGdyb3VwLmNhbGVuZGFyLmdvb2dsZS5jb20.
  opn3ai06k1ge0no877ei3cku4o

Minutes

Sprint Priorities and Status

- AuthZ
    - Scott Prater has started performance testing (has completed the control test without AuthZ enabled)
    - Indicated that there's a need for documentation about how to set a fedora admin username and password
- Content Modeling
    - Scott Prater wants to create a wiki page documenting how to represent a fedora 3 content model as a fedora 4 CND, figure out how to ingest it and add objects.
    - Validation is likely a separate concern.
- Large Files
    - ModeShape 3.7 (which might become available the 17th of Decmeber) will include fixes that allow ingest of files larger than the application's allocated ram.
    - For Federated "ingest" of large files...
        - frank asseg will write up and share the configuration he used for testing
        - Unknown User (escowles@ucsd.edu) volunteered to do a cluster test.
        - Fixity bug exists because fixity checking is done against infinispan store rather than the federated files
        - We need to test adding properties to filesystem federated content
- External Search
    - This is blocked, pending a discussion on changes to the JMS message format
        - Are we committed to the fedora 3 messaging design? (no)
- Versioning
    - Michael Durbin needs to finish up a few tickets
    - Scott Prater may have time to do acceptance testing on this
- Easy Deployment
    - There are two or three tickets relating to this that should be completed for the release.
    - Unknown User (escowles@ucsd.edu) suggested that the default configurations be called "default".
- Performance Testing
    - Single Node
        - Unknown User (escowles@ucsd.edu) reported that fedora 3 read performance is better than fedora 4 and we need to figure out why..
        - .. and that if you run the benchtool client on the same machine fedora 4 suffers more
    - Clustered Performance Testing
        - The SCC cluster was limited by IO performance and useless for benchmarking
        - Greg Jansen is working on a UNC cluster set up
        - Scott Prater at UW has sysadmins setting up a cluster that may be completed by the new year
        - frank asseg can get 6 machines in his office but it's unclear when that might be completed
        - Unknown User (escowles@ucsd.edu) can repurpose the three machines he's done testing on to be a two-node cluster and an ingest machine for testing
            - These are VMs but should be unaffected by external load
        - Chris Beer is dealing with institutional networking issues in setting up his cluster
        - Greg Jansen has been working on shell scripts for deploys of fedora for testing across clusters, Scott Prater will be using these to configure his cluster

Work Assignments

- A. Soroka: will work on the needs for messaging and indexing aparatus once he gets clarity on the design moving forward
- Andrew Woods: will continue to review pull requests and do other project work, emphasized that it's very important to have strong walkthrough documentation for one-click run. Will e-mail advistors, steering group to request development resources through June of 2014. Will be on vacation starting next Sunday but will still roll the release and be somewhat available.
- Benjamin Armintor: identified that for the single node configuration if configuration doesn't make right assumptions it does more work. Collected some numbers on last branch (reducing node lookups) and is moving on to trying to figure out why there are some exploding instantiation counts. Will want to profile access numbering with YourKit.
    - Andrew Woods asked if there were other issues in the level-DB issues. Are any of those issues irreconcilable... should we change to File? Or have those issues past.
    - Unknown User (escowles@ucsd.edu) can test minimal against default. Esme was able to demonstrate one of the issues (out of memory, which would be resolved by modeshape 3.7) Minimal profile is significantly faster.
- Chris Beer: Clustering HTTP API tickets... Andrew suggested that it be documented in a user-digestible format. As well as identifying any problems.
- Eric James: Create property mappings to SOLR. But before that, wire up external index with JMS pluggable application. Adam suggested because the tests don't even run for the indexing framework it might not be good to thrash on that right now. Stand up JMXIndexerPlugin webapp and post documents to SOLR.... before worrying about mapping of properties and connecting... maybe just getting the codebase to build is important. Identify the problems (like deletes not working). Recognize that messages themselves are in a state of flux. Assume that messages are received, but make sure build works with assumption that message part is in flux.
- Unknown User (escowles@ucsd.edu): Has already volunteered to do a weeks's worth of testing. Editable views... gets it working but not reliably. UI enhancements from CARL? Will start with those.
- frank asseg: benchtool, then fixity issue. Bug to fix on another project. Frank out next monday and tuesday.
- Michael Durbin: will work on finishing up versioning tickets before moving on to something else
- Nigel Banks: interested in helping with the indexing issues, testing documentation and possibly islandora documentation
- Osman Din: will work on leftover task about modeshape artifact layout.
- General discussion:
    - Do people want another datapoint for single node performance testing?
    - Performance of ingest and server on same tree.
    - Deeply nested structure of objects haven't been tested. (this is partly covered in the SALT testing at stanford, chris mentioned) Could still do it in bench-tool.
    - Documenting persistance format for levelDB and file backend connector would be quite useful.
- Scott Prater wanted to finish up work from last sprint: AuthZ testing with benchtool (once update occurs), content modeling documentation, won't worry about validation of CND at this time.

Goals

- There will be enough examples and integration tests to create objects and content models as well as documentation to guide fedora 3 content modelers into the fedora 4 mechanics.
    - https://github.com/futures/fcrepo-content-model-examples
- Create cluster configuration recipes that real users could use
- Demonstrate that in a clustered environment, performance (throughput, accesses served, etc.) can be measurably increased by adding server nodes