2014-02-06 - 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)
- Google-Hangout
    - https://plus.google.com/hangouts/_/event/c1glu6soq43r1rr6ou17qtobug8
- 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

- frank asseg ⭐
- Ed Fugikawa
- A. Soroka
- Michael Durbin
- Andrew Woods
- Stefano Cossu
- Martin Dow
- Nikhil Trivedi

Agenda

1. Art Institute of Chicago updates
    a. What are the indexing needs?
    b. Can we document the artic case study?
2. Removing Jena types from kernel (new update pattern)
3. Cluster, how do we get over the hump?

Minutes

Introduction Martin Dow

1. Martin Dow of Acuity Unlimited introduced himself
2. We had a short discussion about the attendance of the committers at Code4Lib 2014
3. There will be no Fedora House at Code4Lib

Art Institute of Chicago updates

1. Stefano gave a short description of the indexing use case at the AIC
2. The need for accessing all JCR supertypes of an object exists and Adam suggested using the builtin and cheap JCR mechanism to access those supertypes instead of exposing them costly over an RDF endpoint
3. We discussed how various triple stores should be abled to be fed data by Fedora 4. A declarative approach for configuring arbitrary Triple Stores would be nice to have but might be hard in light of propriary APIs and non standard SPARQL implementations (e.g 4Store as mentioned by Andrew)
    a. Martin noted that we should take a good look at the features and API designs of various triple store impls in order to assess our needs and design access patterns
    b. Adam reminded us that Fedora 4 is not supposed to be a full implementation of a triple store, there's just a SPARQL Update endpoint in fcrepo4, but all the other features of a full blown semantic store are not and will not be exposed.
    c. Martin will collect and make data available about his ongoing triple store research
4. After an inquiry by Stefano it was made clear why blank rdf nodes are not supported in Fedora 4: Since blank nodes have no JCR node to be bound to they can not be persisted as Properties of a JCR node, which is indeed the way triples are stored in Fedora 4.

Removing Jena types from kernel (new update pattern)

1. We skipped this agenda point

Cluster, how do we get over the hump?

1. Frank reported about the long ingest duration issues he's facing ingesting the SCAPE test data which is due to the large number of children objects created. Approaches for mitigating this problem could be:
    a. Since the cluster only starts syncing data across nodes when the transaction is committed, in a cluster of n nodes n-1 nodes are waiting on data as long as the single node is working on the ingest. Only when the node committs the other n-1 nodes start work. Some kind of bundling/pipelining mechanism by mapping a number of jcr tx into one fedora tx might help mitigate this behaviour.

b. Another method to cut request time in half could be returning a 201 to the user before the data is replicated across the server and handle exceptions as a kind of internal fedora replication error indicated by the fixity service. This way we could further cut down ingest time to the duration needed by the ingest node to accept the data, and don't include the replication duration

2. It was generally agreed that horizonatal scalability is an important core feature of Fedora 4 for many use cases

New Actions

-