

Theme: Shanghai library FOLIO project

Time: November 23, 2022 07:30 pm (EST) / November 24, 2022 08:30 am (GMT+8)

Attendees:

Vincent Bateau (Enterprise Architect, EBSCO)

Gang Zhou (Project manager, Shanghai library)

Sha Jiang (Technical Director, Jiata)

Lucy Liu (Product Owner, Folio China)

Notes:

1. **I noticed that minio was used in FOLIO. Why did the community choose this? Do you have any suggestions on the maintenance and deployment of minio? (Sha Jiang)**

Vince: The Minio **client** is used in Folio. This is because it is compatible with both the Minio APIs and the AWS S3 APIs. In fact, Minio Apis duplicates the AWS S3 APIs. As an open source project, Folio cannot include a hard dependency to a commercial close-source service such as S3. However, since Minio is open source, it can be included and it allows hosting providers to still use AWS S3 in their operations.

Sha Jiang: If some host providers provide other storage that is not compliant with minio. What should we do?

Vince:

- I think minio and external storages can be used in a limited number of places. The thing to do is to replace the client in those places that you want to use something else with another client.
- The thing you have to be careful about is that when you start using external storage, you have to be sure you don't lose the ability to multi-tenancy or you have locked in a particular server. If you want to run scalability on your servers, you want to create a number of servers, and you have run them in clusters for greater capacity. What you don't want to do is to run your storage which is tied to the physical storage of that server because now you have a problem when you want to add another server and you don't know which server's storage to use. That's why minio and S3 are used because they are not dependent on a specific server. They are not local storage servers.

Sha Jiang: How do you consider the uploading and downloading time consumed by minio? Will there be a long time http connection with Okapi? Can this be a problem with http connection leaking with Okapi?

Vince:

- It could be. There is a risk. That's a question of how it's deployed. I don't know the details of the specific implementation.
- I don't believe the current deployment has that kind of dependency. It's kind of the opposite. If you end up scaling your server configuration, and you're supposed to be using storage that belongs to one of the servers, then you're not really going to be able to scale the storage because now it's all going to the same server.
- So in the minio and S3 implementation, one call will be handled by the server, then maybe the next call will end up on a different server because of the load balancing, we have to ensure that all that ends up in the same storage. If we had separate storage on each server, then each one would get a mix or partial sets of data. But because of the way it's written, data is coming incrementally and is going back to the server each time, so it's not persisting with http connection. So if it was written differently, if it maintained http connection for a long time, we would not be able to load balance between multiple independent storage locations.

Sha Jiang: Can we directly connect the storage server from the UI interfaces, i.e, skip the http, so the actual connection won't be a problem?

Vince:

- The connection time means how long you need to keep the connection open. You don't mean the latency, right? - Yes.
- Maybe. I don't think the solution is to use a dedicated instance like that. I think the solution is to ensure you are not requiring a connection to be open the whole time so that you can establish the connection every time and incrementally push it out. If there's too much connection overhead in the opening and closing connections, that can be solved by the connection pooling and can be solved by an abstraction layer, which is the load balancer which will subscribe to that for you.

2. Did the community decide to move to Open Search? Is there migration guidance for projects that are already using ES? (Sha Jiang)

Vince: The issue about Opensearch vs. Elasticsearch is similarly related to licensing. Elasticsearch changed its licensing model in January 2021 which made its client no longer compatible for inclusion in Folio which has an Apache v2 OSS License. As a consequence Folio remained on an older version of Elasticsearch which was still under the Apache v2 License. However, there was a need to update to a newer version in order to include defect and security fixes. AWS Opensearch is a fork of Elasticsearch and it remains under Apache v2 compatible licensing. Therefore, the Folio project had no choice but to move to the Opensearch client. Since Opensearch is a fork, the migration should be simple.

Sha Jiang: We are considering creating a module or an app to manage the middleware storage or search engines, i.e. the index in the search engine and the

items in the external storage. Does the community have any plan for this purpose? I found the community has already implemented this in FOLIO. But every module using this middleware just holds the client for this middleware. There seems to be no management module for search engines and external storage. Is there any idea for this? (Sha Jiang)

Vince:

- No, not really. The FOLIO project doesn't take responsibility for the 3rd party tools. So it doesn't distribute or deeply integrate with any of those tools. It does have the client which is used to facilitate the integration. But it doesn't attempt to manage the indexes on the search engine. The tuning of the search engine, how you do your indexes, is not necessarily going to be managed by FOLIO. It's done externally. The responsibility of the search engine and the tools comes with it to let the hosting providers tune that as needed for different conditions.
- There's a mod-search application. It is responsible for creating or consuming domain events for Kafka, which is a middleware to provide the index. The entire integration is done by mod-search. If someone wants to use a different search engine, they can basically recreate a new mod-search. And this will provide them with the integration with whatever they want to do. They need to recreate APIs for mod-search.

3. So it is recommended to use Open search and Minio in FOLIO platform, isn't it? (Gang Zhou)

Vince: In both cases it is the **client** for Opensearch or Minio that is provided with the Folio platform. The installation or configuration of the search engine and external storage is done outside of Folio by the Folio hosting organization. The Opensearch client has been compatible with Opensearch server and with older versions of Elasticsearch (not sure about the latest ones). The Minio client is compatible with either AWS S3 or Minio server.

4. Is there any recent or ongoing work related to multi-tenancy in the FOLIO community? (Lucy)

Vince: We have multi-tenancy support already in FOLIO. But there is no work being done to extend that in any way.

5. Mike Taylor proposed changes to mod-configuration. What are your thoughts on that proposal? (Lucy)

Vince:

- We briefly discussed this in the TC meeting today. Generally speaking, there's a need to improve the security context. Configuration is not well protected. It's a federated form of configuration and values are put into the configuration, which could include credentials connecting to external systems.

- The TC decision is that Mike can continue his work because he has urgency in the implementation of LDP. We would have to recognize it as a technical debt. In the meanwhile, the tech council will begin a conversation regarding distributed configuration vs. central configuration. If we implement central configuration, we could have a more robust security model. A working group will start working on that. Mike will be invited to that group. The subgroup is to be finalized in the next TC meeting.
- We will have to go through the RFC process and firmly document it (dependencies, consequences, etc.) before we can do anything or update any part of the project.

6. When can the architectural blueprint be prioritized? Is TC working on it? (Lucy)

Vince: Not sure. I haven't heard it mentioned for a while. Not on the agenda recently.

7. Are there any updates on the LoC project? (Gang Zhou)

Vince: It's ongoing. We are identifying what we have to provide, especially in regards to the external systems that need to be connected. LoC has a lot of external systems. And they need to be connected with their version of FOLIO. We are meeting regularly to understand things that need to be done.