

Theme: Shanghai library FOLIO project

Time: November 10, 2020 07:00pm (EST) / November 11, 2020 08:00am (GMT+8)

Attendees:

Vincent Bareau (Enterprise Architect, EBSCO)

Gang Zhou (Project manager, Shanghai library)

Sha Jiang (Technical Director, Jiata)

Lucy Liu (Product Owner, Folio China)

Notes:

1. Problem with data import

- **Gang Zhou** reported that SHL tested user import in two ways: one used user import API and the other - copied data directly to database. The second way was much faster.

Vince commented that the test results indeed showed very different times. The second way was almost a hundred times faster. Vince will pass the feedback to the team/PO that is responsible for the performance of that tool and then see if they're going to improve it or what the plans are.

- **Vince:** It was stuck when Shanghai tried to import 20k users using user-import. How did that happen? You imported 10k, then it was stuck?

Gang Zhou: When we imported 10k, it was ok. When we tried 20k, there was no response. We had to stop the process manually. We can import 10k each time and do multiple batches.

Vince: If you have 6 million users, then you have to do 600 batches. It is taking about 5-6 minutes each batch. That's too long. It's helpful to have this information because we do need a solution for folio. It's not acceptable to have slow times like this and also to require people to do database access. That shouldn't be necessary.

- **Lucy's note:** Gang Zhou shared the numbers on Slack before the meeting:

- Time-consuming

User-import : 337s

copy to database: 3.6s

It is stuck when importing 20K users using user-import.

- Environment

cpu: i5

memory: 16G

JDK: jdk11.0.9_11

okapi: 2.34

mod-user-import: 3.2.1

2. Are there any recommended settings/configurations for using ELK to collect logs in K8S cluster, for example, log templates, etc.?

- **Vince** looked into the ELK question and shared the findings on Slack. There is one case of gathering logs into ELK (or ELK-like stack) from K8S clusters. This is with the “scratch dev” environments. A log aggregation stack was set up which consists of Elasticsearch + Filebeat + Kibana. It is created utilizing Helm charts from [Elastic.co](https://elastic.co) for each component. There are other Helm charts that include Fluentd as the aggregator instead of Filebeat. Additionally, there were a couple of customizations to provide GitHub OAUTH2 for authentication and to provide a Lifecycle Index Policy (to rotate the log indexes). Vince is looking to locate the actual charts that are being used there and will share more info later.

- **Sha Jiang:** Is the format of the log consistent across the platform?

Vince: Which logs are you referring to?

Sha Jiang: We are looking into logs from Okapi and mod-circulation.

Vince: There is no standard format for the logs. There have been discussions about doing that in the past. But I don't think that was adopted. There was a project whose purpose was to create a standard for the log formats and also to implement a Elasticsearch, Fluentd, and Kibana (EFK) stack for reporting and collecting logs. But that project was not prioritized and was not completed. So it means that the logs are most likely not the same between the different components of the platform.

- **Lucy** shared the link (Folio logging solution) to the logging project that Vince mentioned.

Sha Jiang: Is there a timeline or roadmap for this project?

Vince: This is a proposal. When we go forward with this proposal, this is the format that was suggested and was agreed upon. There is a link to [RMB-709](#) - structured logging in RMB, which is blocked by the release and migration to *Vert.x 4*. When the RMB migration to *Vert.x 4* is complete, this will be resumed.

Lucy: Then in this case, can the Shanghai team look for other solutions on their own? You know they have a deadline for trial run and go-live.

Vince: Do you mean they will change the logging mechanism or aggregate the logs separately?

Sha Jiang: We mean the second one.

Vince: The currently linked Jira issue is the standardizing process. It does not offer a particular solution to how to aggregate the logs. So how you collect logs is not part of this and is a separate issue. I don't know that there is a current effort to do that. So the Shanghai team should probably implement what they need to meet deadlines for collecting logs.

3. Progress made or difficulties encountered by the Shanghai team in the development

Sha Jiang: We have set up the K8S clusters for SHL and are working on the logging collection.

Lucy: Will SHL or the consortium start the trial run with one branch in December?

Gang Zhou: We haven't set a date yet. We will start with one branch from the consortium.

Vince: How big is that branch, I mean, the one that will start the trial run?

Gang Zhou: We are still discussing it.