## Infrastructure - Task #3069

### Migrate Gerrit backend to NoteDb and upgrade to version 3.x

2020-01-08 22:40 - David Ostrovsky

<table>
<thead>
<tr>
<th>Status:</th>
<th>Closed</th>
<th>Start date:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority:</td>
<td>Normal</td>
<td>Due date:</td>
<td>2020-02-08</td>
</tr>
<tr>
<td>Assignee:</td>
<td>Guilhem Moulin</td>
<td>% Done:</td>
<td>0%</td>
</tr>
<tr>
<td>Category:</td>
<td>Gerrit</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target version:</td>
<td>Q1/2020</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tags:</td>
<td></td>
<td>URL:</td>
<td></td>
</tr>
</tbody>
</table>

### Description

There are many advantages to upgrade to the recent Gerrit version:

- GWT UI is gone
- ReviewDb is gone and backend migration to NoteDb eliminates dependency on relational database
- PolyGerrit UI is more mature in Gerrit 3.x
- Many new core plugins were added in Gerrit 3.x: delete-project, gitiles, plugin-manager, and webhooks
- Gerrit 3.1 supports Git wire protocol v2 out of the box, that substantially accelerates fetch operations

To enable Git wire protocol v2, these lines should be added to the jgit configuration file in gerrit site:

```bash
/etc/jgit.config

[...]
[protocol]
version = 2
[...]
```

Moreover, the Git wire protocol v2 should be enabled on the client side as well. In my case, it is:

```bash
$ grep -1 protocol ~/.gitconfig

[protocol]
version = 2
```

And last but not least, the minimum required git-core version that supports wire protocol v2 is 2.18.0.

To test, that the git wire protocol communication takes place, one can do:

```bash
$ dacido@wizball:~/projects/reviewers (master %) $ GIT_TRACE_PACKET=1 git -c protocol.version=2 ls-remote
```

### History

#### #1 - 2020-01-08 22:51 - Guilhem Moulin

Upgrade to 3.0 is already scheduled and will happen shortly after FOSDEM. The only reasons why this wasn't done on Christmas days is because I didn't have time to thoroughly test the upgrade path, and accounts are easier to link (using our custom scripts) on ReviewDB (I want a grace period for those who missed the multiple announcements and are currently locked out). I'm testing the migration right now...

Upgrade to 3.1 is not scheduled yet. Might not happen in Q1, perhaps not even Q2 either.

#### #2 - 2020-01-08 23:01 - Guilhem Moulin

What's the status of Java 11 support, by the way? AFAICT it works just fine on the stage 2.16 instance but I've seen
https://bugs.chromium.org/p/gerrit/issues/detail?id=7843 and I'm reluctant to blindly upgrade JRE on the prod instance. The release notes for 2.14 and 2.15 read "Gerrit requires Java Runtime Environment (JRE) version 8; it is not compatible with JRE 9 or newer yet. For more information, see Issue 7843."; release notes for 2.16 don't have an explicit minimum JRE version and doesn't mention Issue 7843.

#### #3 - 2020-01-08 23:56 - David Ostrovsky

2022-03-21
Isn't accounts backend was unconditionally migrated to NoteDb in 2.15 release? So to link on 2.16.15 you would have to mutate NoteDb backend anyway. See this release notes:

https://www.gerritcodereview.com/2.15.html#release-highlights

Account data is stored in NoteDb.

What's the status of Java 11 support, by the way? [...] For more information, see Issue 7843. [...] It is fixed now, see: [1]. I updated the status of: [2].

Note though, that there were some issues reported, mostly related to the JDK 11 and LDAP: [3], [4] and [5].

There is also this related mailing list thread: [6].

Also note, that we are building Gerrit on master using JDK 11. That build would produce JDK 11 byte code and thus JDK 11 required to run it: [7]. The relevant Bazel options are:

$ bazel build
   --host_javabase=@bazel_tools//tools/jdk:remote_jdk11 
   --javabase=@bazel_tools//tools/jdk:remote_jdk11 
   --host_java_toolchain=@bazel_tools//tools/jdk:toolchain_java11 
   --java_toolchain=@bazel_tools//tools/jdk:toolchain_java11 
   release

It would be highly appreciated, if more gerrit users would start using JDK 11. What could possibly go wrong? In emergency case, admin could stop gerrit, downgrade to JDK 8, restart gerrit, done.

Moreover, with this my pending change: [8], Gerrit master is even compatible with JDK 13, and it should be trivial to adapt that change to 3.0.x release line.

[1] https://gerrit.googlesource.com/gerrit.git/+/da8a2ad40293c0070502ae28188f3685cd10caae6
[4] https://bugs.openjdk.java.net/browse/JDK-8217606
[5] https://bugs.openjdk.java.net/browse/JDK-8213202
[6] https://groups.google.com/d/topic/repo-discuss/VMQ8b8p22EU/discussion

#4 - 2020-01-09 00:08 - Guilhem Moulin

Isn't accounts backend was unconditionally migrated to NoteDb in 2.15 release? So to link on 2.16.15 you would have to mutate NoteDb backend anyway.

Sure, that's fine. I'm talking about the case when someone is locked out and create a new profile in SSO (with a different email address). Once they log to Gerrit they'll end up with a duplicate account. Reconciling accounts isn't hard in principle, but it becomes cumbersome when both have some changes attached to it. We have some SQL queries to process that case, and I'd rather avoid rewriting that to deal with NoteDB. Now that we have a single IdP we should see way less duplicates (in the past duplicate accounts were created when users logged in with an unusual/unlinked IdP).

It is fixed now, see: [1]. I updated the status of: [2].

Cool, thanks for confirming and spelling out the fixed versions.

#5 - 2020-01-09 00:32 - Guilhem Moulin

It would be highly appreciated, if more gerrit users would start using JDK 11.
A scary warning can be found in 2.16.15 docs, and it's still present in 3.1.2's: "Java SE Runtime Environment version 1.8. Gerrit is not compatible with Java 9 or newer yet." — https://gerrit-documentation.storage.googleapis.com/Documentation/3.1.2/linux-quickstart.html#_before_you_start.

Changing that line might help :-)

#6 - 2020-01-09 01:08 - David Ostrovsky

A scary warning can be found in 2.16.15 docs, and it's still present in 3.1.2's: "Java SE Runtime Environment version 1.8. Gerrit is not compatible with Java 9 or newer yet."

Indeed. Done in: [1]. Let us see the reviewers reaction ;-) 


#7 - 2020-01-10 04:26 - Guilhem Moulin

- File broken-changes.txt added
- File broken-revisions.txt added
- Due date set to 2020-02-08
- Target version set to Q1/2020

While testing the migration to NoteDb gerrit complained about 45 changes with unknown revisions. For instance https://gerrit.libreoffice.org/c/221 yields a 500 HTTP response code. For the most part these missing revisions date from the early days (2012) and I stumbled upon them before (they broke our Grimoire dashboard — https://dashboard.documentfoundation.org/app/kibana#/dashboard/Gerrit — so we had to blacklist them there).

broken-changes.txt shows the list of broken change IDs along with their owner, creation date, project, and the list of their revisions. That's 45 changes for 101 revisions (the 2 changes / 12 revisions from summer 2016 yield a 404 Not Found not 500 Internal Server Error so I guess it's a different issue), but “only” 57 of these 101 revisions can't be found. They are listed at broken-revisions.txt.

Not sure why there revisions are lost as they were created way before I joined the team, but I suspect an aggressive GC gone wrong. I also looked for some of them on some of our mirrors, to no avail. Anyway, this is nothing new, and we've been running just fine without them for years. David Ostrovsky, do you know if that's more likely to cause problems with NoteDB than with ReviewDB? I should add despite the warnings the offline migrator did set `noteDb.changes.disableReviewDb`.

Also FWIW the stage instance is now running 3.0, with rebased & rebuilt the custom plugins, etc.: https://gerrit-stage.documentfoundation.org.

#8 - 2020-01-10 13:39 - David Ostrovsky

David Ostrovsky, do you know if that's more likely to cause problems with NoteDB than with ReviewDB?

I don't think so. I would guess, that those changes also cannot be reindexed with NoteDb backend, like it is already the case with ReviewDb backend. Given that those changes are broken anyway, one way forward is to remove them from the database, in that way, they will be entirely skipped/ignored during the migration process. Also in most recent Gerrit version there is new feature: delete change (Administrator permission is required to remove changes), but I think, that because those changes are broken, they cannot be probably cleanly removed in Gerrit from the UI as well.

Also FWIW the stage instance is now running 3.0, with rebased & rebuilt the custom plugins, etc.: https://gerrit-stage.documentfoundation.org.

Great! Though, it would be trivial to also skip 3.0.x altogether and just jump directly to 3.1.x, in which case, Git wire protocol v2 can be activated, right?

#9 - 2020-01-10 14:54 - Guilhem Moulin

Given that those changes are broken anyway, one way forward is to remove them from the database, in that way, they will be entirely skipped/ignored during the migration process.

OK thanks, I'll try that.

Though, it would be trivial to also skip 3.0.x altogether and just jump directly to 3.1.x, in which case, Git wire protocol v2 can be activated, right?

One thing at a time please, see msg#1. Any regressions on our instance will end up on my shoulders not yours. As an upstream dev you probably want people to upgrade often, but please accept that admins are more conservative about upgrades. And this is not FUD, despite your explicit
blessing ("do this, do that, done") and tests on the stage instance we did hit some regressions in the past. We don't have resources to follow gerrit's fast moving target anyway.

I have looking forward to wire protocol v2, by the way (learnt the hard way JGit didn't support it yet at the time, cf. https://lists.freedesktop.org/archives/libreoffice/2018-October/081249.html ). But I care about stability even more :-P. It'll come in due time, no need to rush.

#10 - 2020-01-10 23:04 - David Ostrovy

Unrelated to this issue.

 […] Grimoire dashboard —

https://dashboard.documentfoundation.org/app/kibana#/dashboard/Gerrit […]

Interesting. I asked my friends from GerritForge last year, to set up analytics dashboard for LibreOffice projects (core and a couple of other projects). The result is here: [1].

This pipeline to extract the analytics is 100% open source and consists of Gerrit analytics plugin, Spark ETL job and Elastic/Kibana. In this article the pipeline is described: [2].


#11 - 2020-01-10 23:27 - Guilhem Moulin

Thanks for the link, wasn't aware we had external dashboards :-) Our dashboard is also entirely Free Software and ES/Kibana-based. It was first announced at https://blog.documentfoundation.org/blog/2017/08/02/dashboard-announcement/, is at least 1 year older (was already in place when I started). The maintenance was handed over from Bitergia upstream to the infra team in 2018, cf. #1836.

#12 - 2020-01-11 03:32 - Guilhem Moulin

Given that those changes are broken anyway, one way forward is to remove them from the database, in that way, they will be entirely skipped/ignored during the migration process.

OK thanks, I'll try that.

So I deleted these 45 broken changes on the stage ReviewDB (using DELETE FROM changes WHERE change_id IN ([…]);, so excluding comments and patchsets), and ran migrate-to-note-db again. This time it didn't produce any warning and visiting URLs such as /c/221 yields a 404 rather than 500 HTTP response code (hence no longer pollutes the error_log with large Java stacktraces). Thanks!

I'll do the same before migrating the production ReviewDB. The changes metadata won't be migrated to NoteDB, but their 101-57=44 known revisions will still reachable by ref. The metadata for these changes, which is currently visible using gerrit query change:$/CHANGE (but broken in the web UI), will be lost though (beside SQL dumps). I guess that's fair.

#13 - 2020-02-09 06:10 - Guilhem Moulin

- Status changed from New to Closed

Upgraded to 3.0.7.

Files

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Date</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>broken-changes.txt</td>
<td>10.5 KB</td>
<td>2020-01-10</td>
<td>Guilhem Moulin</td>
</tr>
<tr>
<td>broken-revisions.txt</td>
<td>3.45 KB</td>
<td>2020-01-10</td>
<td>Guilhem Moulin</td>
</tr>
</tbody>
</table>