Arvados - Story #11908

Migrate Collections.properties to JSONB

06/27/2017 06:55 PM - Tom Morris

<table>
<thead>
<tr>
<th>Status:</th>
<th>Resolved</th>
<th>Start date:</th>
<th>06/27/2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority:</td>
<td>Normal</td>
<td>Due date:</td>
<td></td>
</tr>
<tr>
<td>Assigned To:</td>
<td>Tom Clegg</td>
<td>% Done:</td>
<td>100%</td>
</tr>
<tr>
<td>Category:</td>
<td>API</td>
<td>Estimated time:</td>
<td>0.00 hour</td>
</tr>
<tr>
<td>Target version:</td>
<td>2017-12-20 Sprint</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description
Handle (deserialize) both YAML and JSON and convert column type to JSONB

Subtasks:
Task # 11933: Review 11908-properties-column-json Resolved

Related issues:
Related to Arvados - Story #11884: Convert Collection properties column type ... Duplicate
Blocked by Arvados - Story #11807: [API] Migrate old serialized database cont... Resolved 06/05/2017
Blocks Arvados - Story #4019: [API] Support query of "properties" field on ob... Resolved 12/12/2017

Associated revisions
Revision 8f987a92 - 12/08/2017 06:49 PM - Tom Clegg
Merge branch '11908-properties-column-json'
refs #11908
Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tclegg@veritasgenetics.com>

History
#1 - 06/27/2017 07:04 PM - Tom Morris
- Description updated
- Target version set to Arvados Future Sprints
- Story points set to 2.0

#2 - 07/05/2017 06:32 PM - Tom Morris
- Target version changed from Arvados Future Sprints to 2017-07-19 sprint

#3 - 07/05/2017 07:21 PM - Tom Morris
- Assigned To set to Tom Clegg

#4 - 07/06/2017 09:33 PM - Tom Clegg
- Status changed from New to In Progress

Changing the column type in-place is trivial ("alter table foo alter column bar type jsonb using bar::jsonb"). Unfortunately, the fulltext index includes the old column so it gets dropped in the process, and has to be rebuilt.

The good news is that I ran into the fulltext index migration bug (existing index not detected, migration fails) and fixed it.

ActiveRecord::StatementInvalid: PG::DuplicateTable: ERROR: relation "collections_full_text_search_idx" already exists

11908-properties-column-json @ bb821d03eb10ddcc7822f7c51a56d1a1182ebc

#5 - 07/07/2017 06:50 PM - Radhika Chippada

- Do we not need to drop the full text index on collection before changing column type back in down migration?
- Do we not want to update workflows -> definition column?
Radhika Chippada wrote:

- Do we not need to drop the full text index on collection before changing column type back in down migration?

On my system the down-migration worked without dropping/recreating the index. I added a comment to the up-migration with the postgresql error it avoids.

- Do we not want to update workflows -> definition column?

Not here/now (this is just collections.properties, to support tags) and perhaps not ever (IIRC we decided to store literal YAML there, instead of using a serialized field, in order to preserve key order, formatting, comments).

11908-properties-column-json @ ebc65675cecd25ca11a86f789bf23b600875b8

Radhika Chippada

On my system the down-migration worked without dropping/recreating the index. I added a comment to the up-migration with the postgresql error it avoids.

Down migration worked for me as well. I also switched to master branch after down migration and added tags to a collection and everything worked fine.

LGTM

#8 - 07/19/2017 06:58 PM - Tom Clegg

- Category set to API
- Target version changed from 2017-07-19 sprint to Arvados Future Sprints
- Story points changed from 2.0 to 0.5

Merge is blocked on PostgreSQL 9.4 dependency.

#9 - 10/31/2017 03:15 PM - Tom Morris

- Target version changed from Arvados Future Sprints to 2017-11-08 Sprint

#10 - 11/08/2017 08:00 PM - Tom Clegg

- Target version changed from 2017-11-08 Sprint to 2017-11-22 Sprint

#11 - 11/17/2017 05:25 PM - Tom Clegg

As a reminder, in July #11807 did (what we expect to be) the slowest part of the yaml-to-json migration, i.e., the jobs table, and that took 3.5 minutes on qr1hi.

This branch does a smaller part of the yaml-to-json migration. It also changes the column type of a single column (collections.properties) that is typically null in most rows, and then it has to regenerate the fulltext index. We don't having timing estimates for that yet.

#12 - 11/22/2017 07:33 PM - Tom Clegg

- Target version changed from 2017-11-22 Sprint to 2017-12-06 Sprint

#13 - 12/06/2017 08:00 PM - Tom Clegg

- Target version changed from 2017-12-06 Sprint to 2017-12-20 Sprint

#14 - 12/06/2017 08:15 PM - Peter Amstutz

- Related to deleted (Story #4019: [API] Support query of "properties" field on objects)

#15 - 12/06/2017 08:16 PM - Peter Amstutz

- Blocks Story #4019: [API] Support query of "properties" field on objects added

#16 - 12/07/2017 03:30 PM - Tom Clegg

Rebased.
#17 - 12/11/2017 09:29 PM - Tom Clegg
- Status changed from In Progress to Resolved