Arvados - Story #11807

[API] Migrate old serialized database content from YAML to JSON

06/05/2017 03:41 PM - Tom Clegg

<table>
<thead>
<tr>
<th>Status:</th>
<th>Resolved</th>
<th>Start date:</th>
<th>06/05/2017</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>2017-07-19 sprint</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Description

Test converting the Jobs or Pipeline Instances table in a representative database (4xp or qr1hi) to JSON to smoke out any potential problems as a preliminary investigation. If this shows substantial problems, spawn separate stories for them.

The Logs table is excluded and will be dealt with separately.

Collection provide performance information to Ops to help calculate how long the migration will take and help determine migration strategy.

Subtasks:

- Task # 11921: Update "upgrading" wiki with dev cluster times
- Task # 11892: Review 11807-yaml-to-json
- Task # 11914: Migrate jobs table
- Task # 11923: Migration for other serialized columns
- Task # 11924: Load fixtures as JSON

Resolved

Related issues:

- Related to Arvados - Bug #11168: [API] Use JSON instead of YAML for serialize... Resolved 02/24/2017
- Blocks Arvados - Story #11908: Migrate Collections.properties to JSONB Resolved 06/27/2017

Associated revisions

Revision 55aafbb0 - 06/30/2017 06:30 PM - Tom Clegg

Merge branch '11807-yaml-to-json'

refs #11807

Arvados-DCO-1.1-Signed-off-by: Tom Clegg <tom@curoverse.com>

History

#1 - 06/06/2017 06:36 PM - Tom Morris

- Tracker changed from Bug to Story

#2 - 06/06/2017 06:56 PM - Tom Morris

- Description updated
- Story points set to 2.0

#3 - 06/21/2017 07:08 PM - Tom Morris

- Target version set to 2017-07-19 sprint

#4 - 06/21/2017 07:12 PM - Tom Morris

- Target version changed from 2017-07-19 sprint to 2017-07-05 sprint

#5 - 06/21/2017 07:19 PM - Tom Morris

- Target version changed from 2017-07-05 sprint to 2017-07-19 sprint

#6 - 06/21/2017 07:37 PM - Tom Morris

- Target version changed from 2017-07-19 sprint to 2017-07-05 sprint
This migration can be run on a live database while an old apiserver is still running, if a sysadmin is sufficiently motivated to avoid downtime to arrange this, instead of doing the usual "apt-get upgrade" procedure.

I was able to verify that the migration is working correctly in my test database.

LGTM

I noticed this during my testing of this branch: I migrated my arvados_test db and jobs text fields are now json formatted. But when I run of the tests, the jobs objects are recreated from test fixtures and they now have the yaml encoding for components etc. Wondering if this would cause a problem if were to change the db column type to json or jsonb later? Do the object creations from fixtures need to use this new serialization strategy?

this is a deploy in su92l

```
su92l:~# tail -F /tmp/pupp*
Reading package lists...  
Reading state information...  
The following packages will be upgraded:  
  arvados-api-server
1 upgraded, 0 newly installed, 0 to remove and 35 not upgraded.
Need to get 27.5 MB of archives.
After this operation, 75.8 kB of additional disk space will be used.
Get:1 http://apt.arvados.org xenial/main amd64 arvados-api-server amd64 0.1.20170705132428.ad77601-8 [27.5 MB]
Fetched 27.5 MB in 1s (18.0 MB/s)
Preparing to unpack .../arvados-api-server_0.1.20170705132428.ad77601-8_amd64.deb ...
Unpacking arvados-api-server (0.1.20170705132428.ad77601-8) over (0.1.20170503164843.1885991-7) ...
Setting up arvados-api-server (0.1.20170705132428.ad77601-8) ...
Assumption: nginx is configured to serve Rails from
    /var/www/arvados-api/current
Assumption: nginx and passenger run as www-data
Creating symlinks to configuration in /etc/arvados/api ...... done.
Running bundle install... done.
Ensuring directory and file permissions ...... done.
Running db:migrate...Defaulting to memory cache, because /var/www/arvados-api/current/tmp/cache owner (uid=33) is not me (uid=0)

DEPRECATION WARNING: The configuration option `config.serve_static_assets` has been renamed to `config.serve_static_files` to clarify its role (it merely enables serving everything in the `public` folder and is unrelated to the asset pipeline). The `serve_static_assets` alias will be removed in Rails 5.0. Please migrate your configuration files accordingly. (called from block in <top (required)> at /var/www/arvados-api/current/config/environments/production.rb:12)

DEPRECATION WARNING: You did not specify a `:log_level` in `production.rb`. Currently, the default value for `:log_level` is `:info` for the production environment and `:debug` in all other environments. In Rails 5 the default value will be unified to `:debug` across all environments. To preserve the current setting, add the following line to your `production.rb`:

    config.log_level = :info
```

. (called from block in tsort_each at /usr/local/rvm/rubies/ruby-2.3.3/lib/ruby/2.3.0/tsort.rb:228)
Called `load` without the :safe option -- defaulting to safe mode.
You can avoid this warning in the future by setting the SafeYAML::OPTIONS[:default_mode] option (to :safe or :unsafe).

== 20170628185847 JobsYamlToJson: migrating ===============================
== 20170628185847 JobsYamlToJson: migrated (73.9702s) ====================

done.

Checking application.yml for completeness...Defaulting to memory cache, because /var/www/arvados-api/current/tmp/cache owner (uid=33) is not me (uid=0)

DEPRECATION WARNING: The configuration option `config.serve_static_assets` has been renamed to `config.serve_static_files` to clarify its role (it merely enables serving everything in the `public` folder and is unrelated to the asset pipeline). The `serve_static_assets` alias will be removed in Rails 5.0. Please migrate your configuration files accordingly. (called from block in <top (required)> at /var/www/arvados-api/current/config/environments/production.rb:12)

DEPRECATION WARNING: You did not specify a `log_level` in `production.rb`. Currently, the default value for `log_level` is :info for the production environment and :debug in all other environments. In Rails 5 the default value will be unified to :debug across all environments. To preserve the current setting, add the following line to your `production.rb`:

    config.log_level = :info

. (called from block in tsort_each at /usr/local/rvm/rubies/ruby-2.3.3/lib/ruby/2.3.0/tsort.rb:228)

Called `load` without the :safe option -- defaulting to safe mode.
You can avoid this warning in the future by setting the SafeYAML::OPTIONS[:default_mode] option (to :safe or :unsafe).

AppVersion (discovered) ad77601-8
action_controller.perform_caching true
admin_notifier_email_from support@curoverse.com
arvados_docsite https://doc.arvados.org
arvados_theme default
assets.compile false
assets.compress true
...

#15 - 07/06/2017 05:44 PM - Nico César
and qr1hi...

== 20170628185847 JobsYamlToJson: migrating ===============================
== 20170628185847 JobsYamlToJson: migrated (269.9086s) ====================

#16 - 07/19/2017 07:00 PM - Tom Clegg
- Status changed from In Progress to Resolved