Fulfill a GET request for a collection by PDH by querying the federation.

1. Match GET request for collection by PDH
2. Send request to the local cluster
3. If success, return result
4. If 404, send parallel requests to list of federated sites defined in discovery document.
5. On the first successful result, alter the signature hints in the returned record so keepstore will be able to get the data from the remote cluster (see Federated collections wiki and #13993)
6. Return first result
7. If all requests fail / time out, return error. Either 404 (if all sites return 404) 502 (Bad Gateway) or 504 (Gateway timeout) depending on the type of failure (return 5xx error only when it makes sense for the client to retry).

Subtasks:
Task # 14174: Review 14087-federated-collection-by-pdh
Resolved

Related issues:
Related to Arvados - Feature #13993: [API] Fetch remote-hosted collection by UUID
Resolved 09/07/2018

Associated revisions
Revision e0ecd284 - 09/17/2018 03:58 PM - Peter Amstutz
Merge branch '14087-federated-collection-by-pdh' refs #14087
Arvados-DCO-1.1-Signed-off-by: Peter Amstutz <pamstutz@veritasgenetics.com>

History
#4 - 08/21/2018 07:51 PM - Peter Amstutz
- Related to Feature #13993: [API] Fetch remote-hosted collection by UUID added

#5 - 08/21/2018 07:57 PM - Peter Amstutz
- Description updated

#6 - 08/21/2018 08:24 PM - Peter Amstutz
- Description updated
- Target version changed from To Be Groomed to Arvados Future Sprints
- Story points set to 2.0

#7 - 09/05/2018 03:44 PM - Peter Amstutz
- Assigned To set to Peter Amstutz
- Target version changed from Arvados Future Sprints to 2018-09-19 Sprint

#8 - 09/10/2018 08:40 PM - Peter Amstutz
14087-federated-collection-by-pdh @ f7fe4c8c4ba93eb1746c5392a820af78c2ad2562

#9 - 09/10/2018 10:18 PM - Peter Amstutz
Rebased, now a4d77e7f0460e85a4dd8c3730c65fea60a8d212f

#10 - 09/11/2018 02:01 PM - Peter Amstutz
Added rate limiting for parallel requests 61bad45c80df09b92cad96d4541421009183c142

#11 - 09/11/2018 03:11 PM - Peter Amstutz

Added rate limiting, do 4 parallel requests at a time

Rebased again @ 3539e6e3e951e5a5e08bb2f10dc71be5841827b294


#12 - 09/12/2018 06:28 PM - Lucas Di Pentima

Sorry for the delay, here're my questions/comments:

- federation.go
  - Line 205: Comment typo: s/we want/we/
  - Lines 234-325: Could this be written in some other way to avoid having such a large conditional block? I found it somewhat difficult to follow.

- Would it be a good idea to test the case with more than 1 remote, some returning 404 and one returning what's requested?

#13 - 09/14/2018 02:15 AM - Peter Amstutz

"Content-Encoding" needs to be added to the blacklist of headers not to be propagated by the proxy.

#14 - 09/14/2018 03:29 PM - Peter Amstutz

Lucas Di Pentima wrote:

Sorry for the delay, here're my questions/comments:

- federation.go
  - Line 205: Comment typo: s/we want/we/

Fixed, thanks.

- Lines 234-325: Could this be written in some other way to avoid having such a large conditional block? I found it somewhat difficult to follow.

I refactored the code a bit. The shorter, non-PDH search case now comes first. I also adjusted the logic that decides when to skip federation and made it wrap the federation endpoint routing.

- Would it be a good idea to test the case with more than 1 remote, some returning 404 and one returning what’s requested?

That's already more or less happening, the test harness sets up two remotes, one called "zzzzz" and the other called "zmock" which always returns 200 OK with an empty response. So that tests the failure case where one of the remotes returns a non-parsable response.

now 14087-federated-collection-by-pdh @ cf1f57e51578de2ad9121e300f55b816b74938684


#15 - 09/14/2018 06:44 PM - Lucas Di Pentima

cf1f57e51578de2ad9121e300f55b816b74938684 LGTM, thanks!

#16 - 09/17/2018 03:42 PM - Peter Amstutz

Peter Amstutz wrote:

"Content-Encoding" needs to be added to the blacklist of headers not to be propagated by the proxy.

And also "Accept-Encoding"

#17 - 09/19/2018 03:21 PM - Peter Amstutz

- Status changed from New to Resolved

#18 - 11/13/2018 09:00 PM - Tom Morris

- Release set to 14