

Arvados - Support #16558

Document FS structure & supported features of WebDAV and S3 compatible APIs

06/23/2020 05:12 PM - Peter Amstutz

Status:	Resolved	Start date:	10/15/2020
Priority:	Normal	Due date:	
Assigned To:	Peter Amstutz	% Done:	100%
Category:	Documentation	Estimated time:	0.00 hour
Target version:	2020-10-21 Sprint		
Description			
For the audience of people trying to automate getting data into and out of keep using our WebDAV or S3 APIs.			
Currently, there is a lot of info at https://godoc.org/git.arvados.org/arvados.git/services/keep-web but it needs to be more accessible from the SDK section of https://doc.arvados.org/ .			
Subtasks:			
Task # 16967: Review 16558-keep-web-api			Resolved

Associated revisions

Revision e46e9a2a - 10/16/2020 09:26 PM - Peter Amstutz

Add WebDAV and S3 API documentation

Remove sections of keep-web/doc.go are now part of the API documentation

refs #16558

Arvados-DCO-1.1-Signed-off-by: Peter Amstutz <peter.amstutz@curii.com>

Revision 948c9603 - 10/16/2020 09:27 PM - Peter Amstutz

Add WebDAV and S3 API documentation

Remove sections of keep-web/doc.go are now part of the API documentation

refs #16558

Arvados-DCO-1.1-Signed-off-by: Peter Amstutz <peter.amstutz@curii.com>

History

#1 - 06/23/2020 05:12 PM - Peter Amstutz

- Tracker changed from Bug to Support

#2 - 07/01/2020 03:42 PM - Peter Amstutz

- Subject changed from Document Keep APIs (keepstore, webdav, upcoming s3) to Document Keep APIs (keepstore, webdav)

#3 - 07/01/2020 03:50 PM - Peter Amstutz

- Description updated

- Subject changed from Document Keep APIs (keepstore, webdav) to Document webdav

#4 - 07/01/2020 04:09 PM - Peter Amstutz

- Target version changed from 2020-07-15 to 2020-08-12 Sprint

#5 - 07/15/2020 03:57 PM - Tom Clegg

- Description updated

#6 - 07/15/2020 04:27 PM - Peter Amstutz

- Target version changed from 2020-08-12 Sprint to 2020-08-26 Sprint

#7 - 08/04/2020 08:57 PM - Peter Amstutz

- Target version changed from 2020-08-26 Sprint to 2020-09-09 Sprint

#8 - 08/26/2020 04:23 PM - Peter Amstutz

- Target version changed from 2020-09-09 Sprint to 2020-09-23 Sprint

#9 - 09/09/2020 04:23 PM - Peter Amstutz

- Target version changed from 2020-09-23 Sprint to 2020-10-07 Sprint

#10 - 09/23/2020 04:21 PM - Peter Amstutz

- Target version changed from 2020-10-07 Sprint to 2020-10-21 Sprint

#11 - 10/06/2020 07:22 PM - Peter Amstutz

- Subject changed from Document webdav to Document FS structure & supported features of WebDAV and S3 compatible APIs

#12 - 10/06/2020 07:23 PM - Peter Amstutz

- Description updated

#13 - 10/07/2020 02:02 PM - Peter Amstutz

- Category set to Documentation

#14 - 10/07/2020 04:07 PM - Peter Amstutz

- Assigned To set to Peter Amstutz

#15 - 10/09/2020 09:22 PM - Peter Amstutz

- Status changed from New to In Progress

#16 - 10/15/2020 02:10 PM - Peter Amstutz

16558-keep-web-api @ [6c8ca0862ed8a66dd42a85ec498fd6e32fad310f](#)

First pass at documenting WebDAV and S3. Focus is on documenting supported features and authorization methods.

#17 - 10/15/2020 08:16 PM - Tom Clegg

Yikes, moving the keep-web commentary out of godoc is long overdue. Thanks.

s3

For GetBucketVersioning, "this is a stub to avoid breaking clients" seems like a weird way to put it, and sort of implies it's incomplete, which it isn't. Maybe "Bucket versioning is not supported, so the response from this API always indicates bucket versioning has never been enabled."? (I don't think we need to justify "why have an API that always returns the same thing? to avoid breaking clients" -- in this context it seems implicit/obvious enough.)

Accepts AWS Signature Version 4 format Authorization: AWS AWS4-HMAC-SHA256 Credential=...,SignedHeaders=...,Signature=...

Delete "AWS" there (should be Authorization: AWS4-HMAC-SHA256 ...) ... but see below

The "Credential" is the Arvados token uuid,

Credential is {AccessKey}/{Scope}, not just AccessKey ... but ...

Maybe we should back off a bit from detailing the V4 Authorization header format here. Better to just explain how to map token UUID/secret to AccessKey/SecretKey, and defer to AWS docs / client libraries for the signing methods?

the signature can be anything as keep-web will ignore it.

Similarly, describing the header instead of the AccessKey/SecretKey mapping is kind of awkward because client libraries ask for a SecretKey, not a signature -- "the SecretKey can be anything" is probably the information they need.

Perhaps something along these lines would be easiest to follow?

- If your client uses V4 signatures exclusively, use the Arvados token's UUID part as AccessKey, and its secret part as SecretKey. This is more common, and recommended.

- If your client uses V2 signatures -- or a combination of V2 and V4 -- use the Arvados token's secret part as both AccessKey and SecretKey.

(As an aside, I think we *could* improve the V2 code so we accept UUID/secret for both V2 and V4 ... it's just that it doesn't seem too urgent because clients generally support V4. It sure would simplify this section for people who don't know/care which kind of signature their client is using, though.)

webdav

The first part of the webdav page looks pretty good.

The "It can be installed anywhere" part seems out of place. I think we already deal with install concerns in the install docs.

"Fetching the root path ... return a 401" sounds a bit weird. Maybe better to start off with "Requests can be authenticated using Basic or Bearer", then (if needed?) mention the RFC 7617 401 thing. This point should probably be in an "authentication" section instead of "browsing". (Also, I'm not 100% sure but I think we use the anonymous token instead of returning 401 at the root URL, if anonymous token is configured.)

Perhaps worth mentioning that webdav also works with /c=foo/t=bar/ and {uuid-or-pdh}.configured.domain.example?

keep-web

Starting at the "URL structure" heading, the webdav page isn't really about webdav any more. Maybe the "other stuff you can do with keep-web" part should be a separate page?

"Authorization mechanisms" heading should really be "Authentication mechanisms", right? (my bad)

There's a spurious "<" at the end of the query string api_token example.

Maybe we should use v2 tokens in these examples so they look more familiar.

We should probably delete the godoc copy of the "stuff keep-web can do" essay, and replace with a link to the new doc page. Keeping divergent copies seems iffy.

#18 - 10/16/2020 08:33 PM - Peter Amstutz

16558-keep-web-api @ [817d7985e27cf188195fed1c7c3e412fb53beb5b](#)

Tweak S3 docs and move keep-web URL patterns to its own page

(Also, I'm not 100% sure but I think we use the anonymous token instead of returning 401 at the root URL, if anonymous token is configured.)

I'm not sure if this is true, in my experience if you arrived at a keep-web link unauthorized, it generally pops up the basic authentication dialog box.

#19 - 10/16/2020 09:06 PM - Tom Clegg

Looks like keep-web-urls.html.textile.liquid didn't get git-added, but the rest LGTM, thanks!

#20 - 10/19/2020 03:03 PM - Peter Amstutz

- Status changed from *In Progress* to *Resolved*

#21 - 05/13/2021 03:02 PM - Peter Amstutz

- Release set to 38