

## Arvados - Feature #8993

### arv-put: options for 3 modes of "resumption"

04/14/2016 01:40 PM - Peter Grandi

<b>Status:</b> Closed	<b>Start date:</b> 04/14/2016
<b>Priority:</b> Normal	<b>Due date:</b>
<b>Assigned To:</b>	<b>% Done:</b> 0%
<b>Category:</b>	<b>Estimated time:</b> 0.00 hour
<b>Target version:</b>	
<b>Description</b>	
Because of <a href="https://dev.arvados.org/issues/8878">https://dev.arvados.org/issues/8878</a> it seems to be that arv-put has 2 modes of operation when uploading hash+block:	
<ul style="list-style-type: none"><li>• Without --no-resume, and if the hash+block is listed in the resume list, and the permissions token for it is not expired, neither hash nor block are uploaded, they are presume to be present in Keep, and simply added to the upload manifest.</li><li>• Otherwise, both hash and block are uploaded and it is Keep's job to avoid unnecessarily duplicating them.</li></ul>	
The problem with with the first case is that is the resume list is assumed to be a valid cache (until permission token expiry), while instead <b>arguably</b> it should be treated as a hint and verified before use.	
The problem with the second case is that the whole block is uploaded, consuming resources, even if Keep then determines it is already present.	
This request is for a 3rd case and a different default, for example in the following form, with 3 values for a new option --upload-again:	
<ul style="list-style-type: none"><li>• yes with the same meaning as current --no-resume, that is unconditionally upload all hashes and their blocks.</li><li>• no with a similar (or even identical) meaning as current --resume, that is upload hashes and blocks only if they are not mentioned in whichever already-uploaded hash list is available.</li><li>• check with a new meaning, to send to all Keepstore daemons the list of hashes to be uploaded (possibly in subsets), which then return a list of those that are found present, with an <i>absolute</i> expiry time, and then to upload all other hashes and blocks, and at the end upload the hashes and blocks in the returned list only if their lifetime has expired, and then write the manifest. Or some obvious variant.</li></ul>	
The default would be --upload-again=yes for safety, with check recommended and no suggested only for "optimistic" cases.	
The main flaw of the current option is that it hold outside Keep block state that is persistent and does not get verified, even if the permission expiry time is a block lifetime time only advisoirly.	
<b>Related issues:</b>	
Related to Arvados - Story #8937: [SDKs] Write integration test for when arv-...	<b>Resolved</b> 04/13/2016
Related to Arvados - Bug #8878: Keep: sudden appearance of "missing" blocks	<b>Closed</b> 04/04/2016

## History

### #1 - 04/14/2016 10:43 PM - Brett Smith

Peter Grandi wrote:

The problem with with the first case is that is the resume list is assumed to be a valid cache (until permission token expiry), while instead **arguably** it should be treated as a hint and verified before use.

We don't think it's arguable—you're right that arv keep put should check its own cache before relying on it. Our plan for this is [#8937](#). In short, this would have arv keep put verify that the first block locator in its resume cache is still usable before relying on it. This means that uploads would start fresh in any of these cases:

- The block has since been removed from Keep (the case that we believe affected you in [#8878](#))
- The block's access token has since expired
- The block's access token is now invalid because it was generated with a signing key no longer in use

Put another way: we're planning on implementing something functionally identical to your check behavior, and we'll make that what you get in the --resume case. arv keep put already takes several steps to verify that its previously cached results are still usable, by making sure the underlying files haven't changed. We want to extend that to verifying things on the Keep end as well, and we believe [#8937](#) will do that.

That work is currently slated to happen this sprint, so it would be done in the next couple of weeks.

**#2 - 04/15/2016 08:28 AM - Peter Grandi**

Added a slightly different argument in <https://dev.arvados.org/issues/8997> "just-in-case".

**#3 - 05/30/2019 05:54 PM - Tom Morris**

- *Related to Bug #8878: Keep: sudden appearance of "missing" blocks added*

**#4 - 01/18/2020 01:13 AM - Peter Amstutz**

- *Status changed from New to Closed*