

Arvados - Support #16421

[doc] document deletion lifecycle of collections, and steps to undelete collections

05/08/2020 02:14 PM - Ward Vandewege

Status:	Resolved	Start date:	09/02/2020
Priority:	Normal	Due date:	
Assigned To:	Ward Vandewege	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	2020-09-09 Sprint		
Description			
<p>We have some 'Keep collection lifecycle' documentation at https://doc.arvados.org/v2.0/user/tutorials/tutorial-keep-collection-lifecycle.html, which is intended for Arvados users.</p> <p>We need something similar but more in-depth geared at administrators. It should cover:</p> <ol style="list-style-type: none">the various phases of deletionwhat is necessary to enable deletion (keep-balance and keepstore configuration)what are the flags to tweak the duration of each of the phases of deletionsteps to recover collections in each of the phases of deletion (where possible) <p>For d) we may want to write some code to automate the steps. See also #16427.</p> <p>Please also review https://doc.arvados.org/v2.0/admin/keep-balance.html which covers some of this ground.</p> <p>And also https://dev.arvados.org/projects/arvados/wiki/Recovering_lost_data as well as https://doc.arvados.org/2.0/admin/recovering-deleted-collections.html</p> <p>Let's consolidate all this stuff in one place. Note that that last URL is referenced from the 1.3.3 release notes at https://lists.arvados.org/pipermail/arvados/2019-May/000210.html so it should be replaced with a link to the new canonical source, not just deleted.</p>			
Subtasks:			
Task # 16767: Review 16421-document-collection-deletion-lifecycle			Resolved
Related issues:			
Related to Arvados - Story #16427: "undelete" command to recover trashed bloc...		Resolved	06/01/2020
Blocks Arvados Epics - Story #16514: Actionable insight into keep usage		New	09/01/2021 12/31/2021

Associated revisions

Revision 9314e5db - 09/08/2020 03:41 PM - Ward Vandewege

Merge branch '16421-document-collection-deletion-lifecycle'

closes #16421

Arvados-DCO-1.1-Signed-off-by: Ward Vandewege <ward@curii.com>

History

#1 - 05/08/2020 02:17 PM - Ward Vandewege

- Description updated

#2 - 05/08/2020 02:51 PM - Ward Vandewege

- Description updated

#3 - 05/08/2020 08:19 PM - Ward Vandewege

- Description updated

#4 - 05/22/2020 09:30 PM - Ward Vandewege

- Related to Story #16427: "undelete" command to recover trashed blocks and restore a deleted collection added

#5 - 05/22/2020 09:30 PM - Ward Vandewege

- Description updated

#6 - 06/10/2020 03:38 PM - Ward Vandewege

- Blocks Story #16514: Actionable insight into keep usage added

#7 - 08/05/2020 02:57 PM - Peter Amstutz

- Release set to 25

#8 - 08/19/2020 02:47 PM - Peter Amstutz

- Target version changed from To Be Groomed to 2020-09-09 Sprint

#9 - 08/26/2020 04:12 PM - Peter Amstutz

- Assigned To set to Ward Vandewege

#10 - 09/01/2020 02:16 PM - Ward Vandewege

- Description updated

#11 - 09/02/2020 02:29 PM - Ward Vandewege

- Status changed from New to In Progress

#12 - 09/02/2020 09:38 PM - Ward Vandewege

Ready for first look in [f7f5dfb456c1ddb6db63cf6e6a79e5d468d20a20](https://github.com/16421-document-collection-deletion-lifecycle/commit/f7f5dfb456c1ddb6db63cf6e6a79e5d468d20a20) on 16421-document-collection-deletion-lifecycle

Pushed missing files in [8a00acc17403e7836b88b1e9e66b4ff47d5505f2](https://github.com/16421-document-collection-deletion-lifecycle/commit/8a00acc17403e7836b88b1e9e66b4ff47d5505f2)

#13 - 09/03/2020 02:54 PM - Peter Amstutz

Ward Vandewege wrote:

Ready for first look in [f7f5dfb456c1ddb6db63cf6e6a79e5d468d20a20](https://github.com/16421-document-collection-deletion-lifecycle/commit/f7f5dfb456c1ddb6db63cf6e6a79e5d468d20a20) on 16421-document-collection-deletion-lifecycle

Recovering data

I usually try to have a brief introduction at the top, like, "Arvados has several features to prevent accidental loss/deletion of data, but accidents happen..."

Mention data recovery can be used for both deleting collections and overwritten collection contents.

Link to "Data lifecycle" architecture page.

Tell user to look in the trash first.

Tell user collection contents may be available somewhere else and try searching by PDH.

Mention/link to collection versioning as a way to prevent accidental overwriting of collection contents.

When mentioning configuration items, include the section, eg AuditLogs.UnloggedAttributes, AuditLogs.MaxAge, Collections.BlobMissingReport

"Obviously this could be improved ..." if it is obvious you don't need to say it, but then raises the question of why the example doesn't do it.

Architecture

I would tweak the section labels to avoid parenthesis, something like "Storage in Keep" and "Computation with Crunch"

"Keep - Content-Addressable Storage" the leading "Keep" in the title is redundant.

I would reorder the pages:

1. Content-Addressable Storage
2. Keep clients
3. Manifest format
4. Data lifecycle

User guide

On "Keep collection lifecycle" there's a section "Collection lifecycle attributes" which is mostly a duplicate of the "Data lifecycle" page, it should just link to that page instead.

#14 - 09/04/2020 03:27 PM - Ward Vandewege

Peter Amstutz wrote:

Ward Vandewege wrote:

Ready for first look in [f7f5dfb456c1ddb6db63cf6e6a79e5d468d20a20](#) on 16421-document-collection-deletion-lifecycle

Recovering data

I usually try to have a brief introduction at the top, like, "Arvados has several features to prevent accidental loss/deletion of data, but accidents happen..."

Mention data recovery can be used for both deleting collections and overwritten collection contents.

Link to "Data lifecycle" architecture page.

Tell user to look in the trash first.

Tell user collection contents may be available somewhere else and try searching by PDH.

Mention/link to collection versioning as a way to prevent accidental overwriting of collection contents.

OK, I've made a bunch of changes along those lines. I've also reduced duplication in the user guide, by making the "Keep lifecycle" page there just about "Trashing and untrashing data", linking to the main "Collection lifecycle" page in the Architecture section.

When mentioning configuration items, include the section, eg AuditLogs.UnloggedAttributes, AuditLogs.MaxAge, Collections.BlobMissingReport

Fixed.

"Obviously this could be improved ..." if it is obvious you don't need to say it, but then raises the question of why the example doesn't do it.

I dropped that line. That code is mostly redundant, the new recover-collections tool is the better way to get out of that situation. But it's helpful to show the 'manual way' so that the admin can see that there **is** a manual way that is not complicated. Increases confidence.

Architecture

I would tweak the section labels to avoid parenthesis, something like "Storage in Keep" and "Computation with Crunch"

OK done.

"Keep - Content-Addressable Storage" the leading "Keep" in the title is redundant.

The title in the menu is also the title of the page. So I want to keep the word 'Keep' in there. I've renamed it to 'Introduction to Keep'.

I would reorder the pages:

1. Content-Addressable Storage
2. Keep clients
3. Manifest format
4. Data lifecycle

Done, though I swapped the lifecycle and manifest format pages, the latter being lower level.

User guide

On "Keep collection lifecycle" there's a section "Collection lifecycle attributes" which is mostly a duplicate of the "Data lifecycle" page, it should just link to that page instead.

Ah, yes, I forgot to do that. Done now.

changes ready for another look at [dafd68e639a416fcae216c639c6e38beab53b214](#) on branch 16421-document-collection-deletion-lifecycle

#15 - 09/08/2020 03:38 PM - Peter Amstutz

LGTM

#16 - 09/08/2020 03:42 PM - Ward Vandewege

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

Applied in changeset [arvados|9314e5dbcf51231eba0deb7e14522a281da22925](#).