Arvados - Bug #16138
arv-copy can't copy collections
02/06/2020 09:35 PM - Peter Amstutz

Status: Resolved
Priority: Normal
Assigned To: Peter Amstutz
Category: 
Target version: 2020-04-08 Sprint

Description

Subtasks:
Task # 16233: Review 16138-fix-arv-copy

Associated revisions
Revision 28bf445f - 03/25/2020 06:35 PM - Peter Amstutz
Merge branch '16138-fix-arv-copy' refs #16138
Arvados-DCO-1.1-Signed-off-by: Peter Amstutz <peter.amstutz@curii.com>

History
#1 - 02/06/2020 09:35 PM - Peter Amstutz
- Description updated

#2 - 02/12/2020 04:40 PM - Peter Amstutz
- Target version changed from 2020-02-26 Sprint to 2020-03-11 Sprint

#3 - 02/26/2020 05:05 PM - Peter Amstutz
- Target version changed from 2020-03-11 Sprint to 2020-03-25 Sprint

#4 - 03/11/2020 03:44 PM - Peter Amstutz
- Assigned To set to Peter Amstutz

#5 - 03/20/2020 07:53 PM - Peter Amstutz
- Status changed from New to In Progress

#6 - 03/23/2020 01:12 PM - Lucas Di Pentima
Some comments & questions:
- Do we need more testing on this command?
- There're more references to "pipeline" on argument descriptions, and code comments.
- It seems there's more pipeline related code:
  - Line 273: migrate_components_filters()
  - Line 410: migrate_jobspec()
  - Line 452: copy.git_repos()
  - Line 754: copy_docker_images() (it looks for a components key, so it seems pipeline-specific)
- Is the whilelisted portable_data_hash field necessary? Isn't the PDH computed from others?

#7 - 03/23/2020 04:20 PM - Peter Amstutz
- Target version deleted (2020-03-25 Sprint)
- Category deleted (SDKs)

Lucas Di Pentima wrote:
Some comments & questions:

- Do we need more testing on this command?

There's no real testing of arv-copy right now. It would require either setting up an environment with two clusters (which used to be impossible without arvbox, with arvados-server boot maybe easier now) or using mocks for everything (which probably wouldn’t have detected the original bug, thus would be low-quality tests).

So I was hoping to just leave it the way it is. Eventually arv-copy will be rewritten in Go and have real testing.

- There’re more references to “pipeline” on argument descriptions, and code comments.

Removed most of them I think.

- It seems there’s more pipeline related code:
  - Line 273: migrate_components_filters()
  - Line 410: migrate_jobspec()
  - Line 452: copy_git_repos()
  - Line 754: copy_docker_images() (it looks for a components key, so it seems pipeline-specific)

Done.

- Is the whilelisted portable_data_hash field necessary? Isn’t the PDH computed from others?

Copying it through ensures that manifest text wasn’t corrupted when it was rewritten.

#8 - 03/23/2020 04:20 PM - Peter Amstutz
16138-fix-arv-copy @ 20e09ab0726d7f0edb4333277c3403f995b960ce

#9 - 03/23/2020 04:58 PM - Lucas Di Pentima
I've tested it manually with 2 arvbox instances, it worked OK.
There're some more pipeline references that may need cleaning, other than that, it LGTM.

$ grep -in pipeline ./sdk/python/build/lib/arvados/commands/arv_copy.py
103: help='The name of the destination git repository. Required when copying a pipeline recursively.'
106: help='The UUID of the project at the destination to which the pipeline should be copied.'
121: description='Copy a pipeline instance, template, workflow, or collection from one Arvados instance to another.',
471: # as the root item (usually a pipeline) being copied.
710: # It returns a string such as 'Collection', 'PipelineInstance', etc.

#10 - 03/25/2020 06:29 PM - Peter Amstutz
- Target version set to 2020-04-08 Sprint

#11 - 03/25/2020 06:58 PM - Peter Amstutz
- Status changed from In Progress to Resolved

#12 - 03/27/2020 02:40 PM - Peter Amstutz
- Release set to 30