

Arvados - Bug #17962

[arv-copy] blows up when copying a project with a registered workflow in it

08/02/2021 03:26 PM - Ward Vandewege

Status:	Resolved	Start date:	11/17/2021
Priority:	Normal	Due date:	
Assigned To:	Peter Amstutz	% Done:	100%
Category:		Estimated time:	0.00 hour
Target version:	2021-11-24 sprint		

Description

Trying to copy a test project from pirca to tordo.

First, the project contained only collections and container requests. The collections were copied, container requests ignored (as expected).

```
# on shell.tordo
$ arv-copy pirca-j7d0g-bbu5p6jsjffokta
pirca-4zz18-ccifrm41a3z84: 0M / 0M 100.0%
pirca-4zz18-aneymw4b20pwug0: 0M / 0M 100.0%
pirca-4zz18-smataa81t834uug: 0M / 0M 100.0%
pirca-4zz18-k03bijojmx1m3n1: 0M / 0M 100.0%
pirca-4zz18-ldybp10d3adj60u: 0M / 0M 100.0%
pirca-4zz18-nuz3ntrreffkanr: 0M / 0M 100.0%
pirca-4zz18-3xsu89n5mbiy2l6: 0M / 0M 100.0%
pirca-4zz18-ghfqhzb8z9bghc: 0M / 0M 100.0%
pirca-4zz18-uvwqkky5kjtthem: 0M / 0M 100.0%
2021-08-02 15:17:03 arvados.arv-copy[26659] INFO:
2021-08-02 15:17:03 arvados.arv-copy[26659] INFO: Success: created copy with uuid
tordo-j7d0g-2s0oya2vzqps25h
```

Then, I copied a registered workflow into the project, and tried to copy it again:

```
$ arv-copy pirca-j7d0g-bbu5p6jsjffokta
Traceback (most recent call last):
  File "/usr/share/python3/dist/python3-arvados-python-client/lib/python3.7/site-packages/ruamel/yaml/reader.py", line 105, in stream
    return self._stream
AttributeError: 'SafeLoader' object has no attribute '_stream'
```

During handling of the above exception, another exception occurred:

```
Traceback (most recent call last):
  File "/usr/bin/arv-copy", line 7, in <module>
    main()
  File "/usr/share/python3/dist/python3-arvados-python-client/lib/python3.7/site-packages/arvados/commands/arv_copy.py", line 144, in main
    result = copy_project(args.object_uuid, src_arv, dst_arv, args.project_uuid, args)
  File "/usr/share/python3/dist/python3-arvados-python-client/lib/python3.7/site-packages/arvados/commands/arv_copy.py", line 683, in copy_project
    copy_workflow(w["uuid"], src, dst, args)
  File "/usr/share/python3/dist/python3-arvados-python-client/lib/python3.7/site-packages/arvados/commands/arv_copy.py", line 291, in copy_workflow
    wf_def = yaml.safe_load(wf["definition"])
  File "/usr/share/python3/dist/python3-arvados-python-client/lib/python3.7/site-packages/ruamel/yaml/main.py", line 994, in safe_load
    return load(stream, SafeLoader, version)
  File "/usr/share/python3/dist/python3-arvados-python-client/lib/python3.7/site-packages/ruamel/yaml/main.py", line 947, in load
    loader = Loader(stream, version, preserve_quotes=preserve_quotes)
  File "/usr/share/python3/dist/python3-arvados-python-client/lib/python3.7/site-packages/ruamel/yaml/loader.py", line 40, in __init__
    Scanner.__init__(self, loader=self)
```

```

File "/usr/share/python3/dist/python3-arvados-python-client/lib/python3.7/site-packages/ruamel/yaml/scanner.py", line 82, in __init__
    self.reset_scanner()
File "/usr/share/python3/dist/python3-arvados-python-client/lib/python3.7/site-packages/ruamel/yaml/scanner.py", line 104, in reset_scanner
    self.fetch_stream_start()
File "/usr/share/python3/dist/python3-arvados-python-client/lib/python3.7/site-packages/ruamel/yaml/scanner.py", line 446, in fetch_stream_start
    mark = self.reader.get_mark()
File "/usr/share/python3/dist/python3-arvados-python-client/lib/python3.7/site-packages/ruamel/yaml/reader.py", line 180, in get_mark
    if self.stream is None:
File "/usr/share/python3/dist/python3-arvados-python-client/lib/python3.7/site-packages/ruamel/yaml/reader.py", line 107, in stream
    raise YAMLStreamError('input stream needs to specified')
ruamel.yaml.error.YAMLStreamError: input stream needs to specified

```

Interestingly, this doesn't seem to happen all the time:

```

ward@shell:~$ dpkg -l |grep arv
ii arvados-client                2.3.0~dev20210803154620-1      amd64
   Arvados command line tool (beta)
ii arvados-src                   2.3.0~dev20210803154620-1      all
   The Arvados source code
ii libpam-arvados-go             2.3.0~dev20210803154620-1      amd64
   Arvados PAM authentication module
ii python3-arvados-cwl-runner    2.3.0~dev20210729201354-1      amd64
   Arvados Common Workflow Language runner
ii python3-arvados-fuse          2.3.0~dev20210726143205-1      amd64
   Arvados FUSE driver
ii python3-arvados-python-client 2.3.0~dev20210726143205-1      amd64
   Arvados client library

```

```

ward@shell:~$ arv-copy pirca-j7d0g-qv5r6tyawa0fri3 --project-uuid ce8i5-j7d0g-3b990qlglr7cmse
pirca-4zz18-3ik8491082fqjcz: 0M / 0M 100.0%
pirca-4zz18-xa0i7qjide8td5d: 156M / 156M 100.0%
pirca-4zz18-n9os5bjq4ipwe7d: 23M / 23M 100.0%
pirca-4zz18-xos3zqns62comjb: 94M / 94M 100.0%
pirca-4zz18-ckim3zvdmbf3i05: 0M / 0M 100.0%
pirca-4zz18-3n2p2xu20q42it0: 259M / 259M 100.0%
2021-08-03 17:26:39 arvados.arv-copy[28298] INFO:
2021-08-03 17:26:39 arvados.arv-copy[28298] INFO: Success: created copy with uuid
ce8i5-j7d0g-fqjgh6cs9nezwey

```

Subtasks:

Task # 18038: Review 17962-check-wf-definition

Resolved

Associated revisions

Revision 125fbfaa - 11/17/2021 10:03 PM - Peter Amstutz

Merge branch '17962-check-wf-definition' refs #17962

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

Revision cd034042 - 11/17/2021 10:06 PM - Peter Amstutz

Merge branch '17962-check-wf-definition' refs #17962

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

History

#1 - 08/02/2021 03:28 PM - Ward Vandewege

- Description updated

#2 - 08/02/2021 03:55 PM - Peter Amstutz

That's really weird. Seems like some kind of ruamel.yaml failure. Bad version maybe?

#3 - 08/03/2021 03:10 PM - Peter Amstutz

- Target version changed from 2021-08-18 sprint to 2021-09-01 sprint

#4 - 08/03/2021 05:44 PM - Ward Vandewege

- Description updated

#5 - 08/03/2021 05:46 PM - Ward Vandewege

- Description updated

#6 - 08/18/2021 03:18 PM - Peter Amstutz

- Assigned To set to Peter Amstutz

#7 - 09/01/2021 01:47 PM - Peter Amstutz

- Target version changed from 2021-09-01 sprint to 2021-09-15 sprint

#8 - 09/01/2021 01:49 PM - Peter Amstutz

- Assigned To deleted (Peter Amstutz)

#9 - 09/01/2021 03:24 PM - Peter Amstutz

- Target version changed from 2021-09-15 sprint to 2021-09-29 sprint

#10 - 09/15/2021 03:08 PM - Peter Amstutz

- Assigned To set to Peter Amstutz

#11 - 09/29/2021 01:58 PM - Peter Amstutz

- Target version changed from 2021-09-29 sprint to 2021-10-13 sprint

#12 - 10/13/2021 03:25 PM - Peter Amstutz

- Target version changed from 2021-10-13 sprint to 2021-10-27 sprint

#13 - 10/27/2021 02:54 PM - Peter Amstutz

- Target version changed from 2021-10-27 sprint to 2021-11-10 sprint

#14 - 11/10/2021 04:06 PM - Peter Amstutz

- Target version changed from 2021-11-10 sprint to 2021-11-24 sprint

#15 - 11/16/2021 04:28 PM - Peter Amstutz

- Release set to 42

#16 - 11/16/2021 04:29 PM - Peter Amstutz

- Release deleted (42)

#17 - 11/17/2021 09:24 PM - Peter Amstutz

17962-check-wf-definition @ [972272c42f7ad8daaef29d3d981968e566f3f087](#)

[developer-run-tests: #2806](#) [icon?job=developer-run-tests&build=2806](#)

The reason this failed was because the specific workflow record in the bug report had null in the "definition" field. I don't know how that happened, but arv-copy didn't like it.

I've improved the error handling for arv-copy a little bit.

```
$ arv-copy pirca-j7d0g-bbu5p6jsjjfokta
2021-11-17 16:15:33 arvados.arv-copy[1853791] WARNING: Workflow object pirca-7fd4e-hx0a8x0sskhd1od
has an empty or null definition, it won't do anything.
2021-11-17 16:15:33 arvados.arv-copy[1853791] WARNING: Warning: created copy with uuid
x2z00-j7d0g-hqfaopz1teptkn0 but failed to copy some items:
Error while copying pirca-7fd4e-hx0a8x0sskhd1od: input stream needs to specified
```

#18 - 11/17/2021 09:25 PM - Peter Amstutz

- Release set to 45

#19 - 11/17/2021 09:43 PM - Lucas Di Pentima

- At line 164: I think that print() call was for debugging?

LGTM otherwise, thanks!

#20 - 11/17/2021 09:45 PM - Peter Amstutz

Lucas Di Pentima wrote:

- At line 164: I think that print() call was for debugging?

LGTM otherwise, thanks!

Actually it was on purpose, it makes it usable in a script (you can capture stdout to get the uuid of the copy). Although it's technically out of scope of the ticket so I can pull it out if you think I should.

#21 - 11/17/2021 09:50 PM - Lucas Di Pentima

I think it's a good idea but also guess that it should be validated with a test and maybe some documentation (ie: on the command's --help message)

#22 - 11/17/2021 10:06 PM - Peter Amstutz

- Status changed from New to Resolved