

Arvados - Story #9277

[Crunch2] System-owned container outputs should be garbage-collected

05/24/2016 06:55 PM - Peter Amstutz

Status:	Resolved	Start date:	02/16/2017
Priority:	Normal	Due date:	
Assigned To:	Peter Amstutz	% Done:	100%
Category:	API	Estimated time:	0.00 hour
Target version:	2017-03-01 sprint		
Description			
Background			
When a container finishes, crunch-run creates an output collection and records its PDH in the container record.			
On the API server, the container update triggers a hook that creates one copy of the output collection for each of N container requests that have been waiting on this container.			
The original output collection is owned by root and there is no cleanup process has trash_at set to now+defaultTrashLifetime.			
Proposed fix			
In crunch-run, create the output collection with trash_at=now.			
In the API server hook that creates a collection for each relevant container request, when looking up the container output manifest, make sure to include trashed collections in the search.			
Subtasks:			
Task # 11107: Review 9277-trash-container-outputs			Resolved
Task # 11164: Review 9277-container-output			Resolved
Related issues:			
Related to Arvados - Story #9278: [Crunch2] Document/fix handling of collecti...		In Progress	06/01/2016
Related to Arvados - Feature #11100: [CWL] Intermediary collection handling c...		Resolved	03/30/2017
Has duplicate Arvados - Feature #11064: [Crunch2] crunch-run expires output c...		Duplicate	

Associated revisions

Revision 21598295 - 02/16/2017 08:13 PM - Peter Amstutz

Merge branch '9277-trash-container-outputs' closes #9277

Revision cca78a11 - 02/24/2017 07:21 PM - Peter Amstutz

Merge branch '9277-container-output' closes #9277

History

#1 - 05/24/2016 06:59 PM - Peter Amstutz

- Description updated

#2 - 11/23/2016 08:01 PM - Peter Amstutz

- Subject changed from [Crunch2] Behavior for output collections to [Crunch2] Intermediate output collections have expiry set

- Description updated

#3 - 02/07/2017 07:08 PM - Tom Clegg

- Description updated

- Category set to API

#4 - 02/07/2017 07:09 PM - Tom Clegg

- Subject changed from [Crunch2] Intermediate output collections have expiry set to [Crunch2] System-owned container outputs should be

garbage-collected

#5 - 02/14/2017 07:35 PM - Tom Morris

- Target version set to 2017-03-01 sprint
- Story points set to 0.5

#6 - 02/15/2017 08:21 PM - Peter Amstutz

- Assigned To set to Peter Amstutz

#7 - 02/16/2017 04:57 PM - Tom Clegg

- Description updated

#8 - 02/16/2017 06:15 PM - Peter Amstutz

Even better than trash_at=now, it looks like we can just set is_trashed: true

9277-trash-container-outputs @ [72873affa7f249faa16d5d21200e935d27aea911](https://github.com/72873affa7f249faa16d5d21200e935d27aea911)

#9 - 02/16/2017 07:16 PM - Peter Amstutz

- Status changed from New to In Progress

#10 - 02/16/2017 08:02 PM - Radhika Chippada

LGTM

#11 - 02/16/2017 08:15 PM - Peter Amstutz

- Status changed from In Progress to Resolved
- % Done changed from 0 to 100

Applied in changeset arvados|commit:21598295f38998d8028aaa117f192de6b5758808.

#12 - 02/23/2017 06:18 PM - Peter Amstutz

- Status changed from Resolved to Feedback

```
def validate_output
  # Output must exist and be readable by the current user. This is so
  # that a container cannot "claim" a collection that it doesn't otherwise
  # have access to just by setting the output field to the collection PDH.
  if output_changed?
    c = Collection.
      readable_by(current_user).
      where(portable_data_hash: self.output).
      first
    if !c
      errors.add :output, "collection must exist and be readable by current user."
    end
  end
end
```

This also needs to be unscoped.

#13 - 02/24/2017 04:12 PM - Peter Amstutz

Additional fixes on 9277-container-output

Tests:

<https://ci.curoverse.com/job/developer-run-tests/178/>

#14 - 02/24/2017 06:30 PM - Radhika Chippada

It would be nice (if it doesn't already exist) to add a test such as "error setting output on running container if trashed collection is not readable". Thanks.

LGTM

#15 - 02/24/2017 07:25 PM - Peter Amstutz

- *Status changed from Feedback to Resolved*

- *% Done changed from 50 to 100*

Applied in changeset arvados|commit:cca78a112d6a2aa4f76c1956cfe8ec2d43a68759.