Arvados - Bug #13609

[R SDK] Copying files using the R SDK results in 0 byte files

06/11/2018 07:43 PM - Bryan Cosca

Status:	Resolved		**Start date:**	07/19/2018
Priority:	Normal		**Due date:**	
Assigned To:	Fuad Muhic		**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
Target version:	2018-08-15 Sprint			

Description

Using ArvadosR package 0.0.4, This is the script I'm using to reproduce:

```
library(ArvadosR)
packageVersion("ArvadosR")
arv <- Arvados$new("api-token-redacted", "e51c5.arvadosapi.com", numRetries = 3)
previous_collection <- Collection$new(arv, "e51c5-4zz18-ko243dg1k5c6hid")
message("accessed previous collection")
file_to_be_transferred <- previous_collection$get("foo")
message("created ArvadosFile object to transfer")
newcoll <- arv$createCollection(list(name = "test-copy-foo", description = "foo"))
message("created new collection")
new_collection <- Collection$new(arv, newcoll$uuid)
message("created new collection object")
new_collection$add(file_to_be_transferred)
message("transferred file")
new_collection$uuid
```

It resulted in this collection: https://workbench.e51c5.arvadosapi.com/collections/e51c5-4zz18-19p0djjs1yt7xq0

```
$ ls -lah ~/keep-e51c5/e51c5-4zz18-19p0djjs1yt7xq0
total 1.5K
drwxrwxrwx 1 bcosc bcosc 0 Jun 11 19:40 .
dr-xr-xr-x 1 bcosc bcosc 0 May 17 18:25 ..
-rwxrwxrwx 1 bcosc bcosc 0 Jun 11 19:40 foo
```

Subtasks:

Task # 13845: Review   **Resolved**

Associated revisions

Revision 336638f4 - 08/02/2018 06:00 PM - Peter Amstutz

Merge branch '13609-r-sdk-copy' refs #13609

Arvados-DCO-1.1-Signed-off-by: Peter Amstutz <pamstutz@veritasgenetics.com>

History

#1 - 06/11/2018 07:47 PM - Bryan Cosca

- Description updated

#2 - 06/12/2018 06:37 PM - Peter Amstutz

So this seems to be a bug or oversight in the R SDK, it doesn't handle this case at all.

However, I think using the API this way is probably an error. It isn't clear what it should mean to "add" a file that already belongs to another collection. The intention here is was for it to be copied, but it would make just as much sense for it to be moved. WebDAV has MOVE and COPY operations, but they are not implemented for all cases (in particular MOVE between collections isn't supported by WebDAV).

Suggested fix:

- Using Collection$add() with an ArvadosFile object that already belongs to a collection should be an error
- Add a new Collection$copy() method which accepts a source ArvadosFile object and returns a new ArvadosFile representing the file that has been copied into the collection.

#3 - 06/12/2018 06:41 PM - Peter Amstutz

In the meantime, the workaround is something like:

```
library(ArvadosR)
packageVersion("ArvadosR")
arv <- Arvados$new("api-token-redacted", "e51c5.arvadosapi.com", numRetries = 3)
previous_collection <- Collection$new(arv, "e51c5-4zz18-ko243dg1k5c6hid")
message("accessed previous collection")
file_to_be_transferred <- previous_collection$get("foo")
message("created ArvadosFile object to transfer")
newcoll <- arv$createCollection(list(name = "test-copy-foo", description = "foo"))
message("created new collection")
new_collection <- Collection$new(arv, newcoll$uuid)
message("created new collection object")
dest_file <- new_collection$create(file_to_be_transferred$getName())
dest_file$write(file_to_be_transferred$read())
message("transferred file")
new_collection$uuid
```

#4 - 06/13/2018 02:12 PM - Tom Morris

- Target version set to Arvados Future Sprints

#5 - 06/13/2018 02:14 PM - Fuad Muhic

- Assigned To set to Fuad Muhic

#6 - 07/03/2018 04:04 PM - Tom Morris

- Target version changed from Arvados Future Sprints to 2018-07-18 Sprint

- Story points set to 0.5

#7 - 07/03/2018 04:05 PM - Tom Morris

- Subject changed from Copying files using the R SDK results in 0 byte files to [R SDK] Copying files using the R SDK results in 0 byte files

#8 - 07/18/2018 03:25 PM - Tom Morris

- Target version changed from 2018-07-18 Sprint to 2018-08-01 Sprint

#9 - 07/19/2018 10:23 AM - Fuad Muhic

Currently, R SDK doesn't support copy operation at all. Move operation is supported only for moves within same collection. Add operation behaves similarly to create operation. It will always create new empty file(s). So if you grab ArvadosFile from a collection and add it to same collection on difference place, file won't be moved. Instead, new empty file will be created with same name as file we were trying to add (That is probably a mistake, it should be a warning/error or this case should be detected and treated as move). If you want to move file, you need to use move method explicitly.

If we have a composite of ArvadosFiles/Subcollections and it doesn't belong to any collection, we can freely manipulate it (this way we can freely create any tree structure and add it to existing collection, and changes will be applied recursively). However if that composite tree structure belongs to a collection, any change to it will include REST call. This way we ensure that local and remote content are in sync. Also, this interaction is based on assumption that we are always working within same collection. If we want to support any operation between collections in R SDK (add, move or copy) we need WebDAV support for it, and we need to update the way nodes are managed inside collections in-memory tree structure.

#10 - 07/19/2018 06:48 PM - Peter Amstutz

Fuad Muhic wrote:

> Currently, R SDK doesn't support copy operation at all. Move operation is supported only for moves within same collection. Add operation behaves similarly to create operation. It will always create new empty file(s). So if you grab ArvadosFile from a collection and add it to same collection on difference place, file won't be moved. Instead, new empty file will be created with same name as file we were trying to add (That is probably a mistake, it should be a warning/error or this case should be detected and treated as move). If you want to move file, you need to use move method explicitly.
>
> If we have a composite of ArvadosFiles/Subcollections and it doesn't belong to any collection, we can freely manipulate it (this way we can freely create any tree structure and add it to existing collection, and changes will be applied recursively). However if that composite tree structure belongs to a collection, any change to it will include REST call. This way we ensure that local and remote content are in sync. Also, this interaction is based on assumption that we are always working within same collection. If we want to support any operation between collections in R SDK (add, move or copy) we need WebDAV support for it, and we need to update the way nodes are managed inside collections in-memory tree structure.

I think the confusion is because the R SDK API has two ways to create files:

- collection$create("newfile")
- collection$add(ArvadosFile$new("newfile"))

I assume "add()" was really intended to be a package-internal helper method for constructing the file listing. The problem is that it is a documented public method, but doesn't work (and isn't intended to work) for all the ways people want to use it.

I recommend:

- Remove the add() method from the examples / document that it is internal and users shouldn't use it. Fail if the parent collection of "newfile" isn't null.

- Add Collection$copy() / Subcollection$copy() method:

  filecopy = subcollection$copy(source, relativeDestination = "")

Copy the "source" ArvadosFile to "relativeDestination" (a file or subcollection under the subcollection object it is called on) and return a new ArvadosFile for the copy. Should issue a WebDAV COPY request. Tom says COPY is probably supported, but you'll have to confirm that in testing.

#11 - 07/20/2018 12:12 PM - Fuad Muhic

- Status changed from New to In Progress

#12 - 07/23/2018 07:07 PM - Tom Morris

- Release set to 13

#13 - 07/26/2018 03:29 PM - Fuad Muhic

13609-r-sdk-copy @ c9bfa91e3c3ec90778795969fdd3787578a76ed4
test-run: https://ci.curoverse.com/job/developer-run-tests/827/
- implemented copy for for Collection, Subcollection and ArvadosFile (works only for files and folders inside same collection)
- small bug-fix for move in REST
- reworked create method in Collection
- add method in Collection is private now (removed from official documentation)
- updated documentation

WEBDAV move doesn't support moving content between different collection. Result of moving is "Not Found".

#14 - 08/01/2018 03:15 PM - Peter Amstutz

- Target version changed from 2018-08-01 Sprint to 2018-08-15 Sprint

#15 - 08/02/2018 06:05 PM - Peter Amstutz

- Status changed from In Progress to Resolved