Arvados - Feature #12320

Access Arvados projects under the "by_id" directory in keep mount

09/25/2017 04:34 PM - Abram Connelly

Status:	Resolved		**Start date:**	09/25/2017
Priority:	Normal		**Due date:**	
Assigned To:	Fuad Muhic		**% Done:**	100%
Category:			**Estimated time:**	0.00 hour
Target version:	2018-04-25 Sprint			

Description

One can access collections either by their portable data hash or UUID from the "by_id" directory in the keep mount. Trying to do the same for an Arvados project will fail with a "no such file or directory".

It would be nice to be able to access Arvados projects by their UUID under the keep "by_id" directory.

Implementation details:

1. The code is in arvados/services/fuse
2. The mount point configuration is initialized in arvados_fuse/command.py
3. The "by_id" directory is implemented by the MagicDirectory class in arvados_fuse/fusedir.py
4. The MagicDirectory is special, subdirectories only appear in the listing after they are first accessed
5. The MagicDirectory.__contains__ method is called to determine if a subdirectory exists (named by portable data hash or uuid) and creates on demand it if necessary
6. Currently it creates CollectionDirectory object for collections.
7. The goal of this task is to add support to MagicDirectory for accessing projects (ProjectDirectory) by uuid.

Tests are in arvados/services/fuse/tests

The simplest way to run them is using [arvbox](#) on Linux (note, won't work on Windows/MacOS so in that case you will need to use a VM).

```
$ arvbox start test --only services/fuse
```

Alternately you can set up your workstation or a VM following these instructions:
https://dev.arvados.org/projects/arvados/wiki/Hacking_prerequisites (but personally I use arvbox for all my development).

The test framework starts up the API server in testing mode with a known database, so some tests rely on the API server returning known values. These test values are defined in arvados/services/api/test/fixtures. You can probably use an existing record, such as "aproject" (uuid [zzzzz-j7d0g-v955i6s2oi1cbso](#)) defined groups.yml.

As a starting point, try looking at FuseMagicTest in arvados/services/fuse/tests/test_mount.py

To run a single test:

```
$ arvbox start test --only services/fuse services/fuse_test=--test-suite=tests.test_mount.FuseMagi
cTest
```

Subtasks:

Task # 13226: Review 12320-access-projects-from-by_id	**Resolved**

Associated revisions

Revision ec0d70da - 04/13/2018 07:59 AM - Fuad Muhic

Merge branch '12320-access-projects-from-by_id'

closes #12320

Arvados-DCO-1.1-Signed-off-by: Fuad Muhic <fmuhic@capeannenterprises.com>

History

#1 - 09/25/2017 06:37 PM - Tom Morris

- Target version set to Arvados Future Sprints

#2 - 09/26/2017 06:52 PM - Tom Morris

- Story points set to 0.5

#3 - 03/14/2018 07:39 PM - Tom Morris

- Assigned To set to Fuad Muhic

- Target version changed from Arvados Future Sprints to 2018-03-28 Sprint

#4 - 03/16/2018 08:45 PM - Peter Amstutz

- Description updated

#5 - 03/28/2018 03:26 PM - Tom Morris

- Status changed from New to In Progress

- Target version changed from 2018-03-28 Sprint to 2018-04-11 Sprint

#6 - 04/10/2018 03:12 PM - Lucas Di Pentima

Reviewing branch 12320-access-projects-from-by_id at [1f15c5b6060b0d82c9b17fb724b7128374f65747](1f15c5b6060b0d82c9b17fb724b7128374f65747)

Just one question:

Should all types of groups be accessible on fuse mount? Maybe we're only looking for the ones with ['group_class', '=', 'project'] (just in case, double check with Tom or Peter)

#7 - 04/10/2018 04:33 PM - Fuad Muhic

Lucas Di Pentima wrote:

> Reviewing branch 12320-access-projects-from-by_id at [1f15c5b6060b0d82c9b17fb724b7128374f65747](1f15c5b6060b0d82c9b17fb724b7128374f65747)
>
> Just one question:
>
> Should all types of groups be accessible on fuse mount? Maybe we're only looking for the ones with ['group_class', '=', 'project'] (just in case, double check with Tom or Peter)

You are right, I fixed it.

#8 - 04/11/2018 01:05 PM - Lucas Di Pentima

Reviewing fcdea9d131e4ba823ac8b69224d90b8eb5f4ae2e

Do you think wouldn't be better to add a filter to the api call so that it isn't retrieved in the first place?

#9 - 04/11/2018 01:11 PM - Fuad Muhic

Lucas Di Pentima wrote:

> Reviewing fcdea9d131e4ba823ac8b69224d90b8eb5f4ae2e
>
> Do you think wouldn't be better to add a filter to the api call so that it isn't retrieved in the first place?

Do you mean to do a api group list with filter ['group_class', '=', 'project'] and ['uuid', '=', k]?

#10 - 04/11/2018 03:26 PM - Tom Morris

- Target version changed from 2018-04-11 Sprint to 2018-04-25 Sprint

#11 - 04/12/2018 03:23 PM - Lucas Di Pentima

Fuad Muhic wrote:

>> Do you think wouldn't be better to add a filter to the api call so that it isn't retrieved in the first place?
>
> Do you mean to do a api group list with filter ['group_class', '=', 'project'] and ['uuid', '=', k]?

Yes, I believe it would be better in terms of bandwidth usage, for example if some user tries to access lots of group uuids.

#12 - 04/12/2018 03:45 PM - Lucas Di Pentima

[a840e658554cccdf5cda70c71666a62b862f9f48](#) LGTM.

#13 - 04/13/2018 08:08 AM - Fuad Muhic

- Status changed from In Progress to Resolved

#14 - 07/23/2018 06:41 PM - Tom Morris

- Release set to 13