Arvados - Feature #17351

[arv-put] Storage classes

02/08/2021 08:09 PM - Nico César

Status:	Resolved		**Start date:**	06/03/2021
Priority:	Normal		**Due date:**	
Assigned To:	Lucas Di Pentima		**% Done:**	100%
Category:	Keep		**Estimated time:**	0.00 hour
Target version:	2021-06-23 sprint			

Description

Goals of this ticket:

- to define the command line arguments to specify storage clases
- to have all the expected behaviour for arv-put
- to add to the documentation this behaviour
- to add necesary tests to make sure we comply with this behaviour

to discuss:

- shall we migrate arv-put to Go in this instance or is this a future work?

Current command line arguments (arvados 2.1.0) :

```
arv-put --help
usage: arv-put [-h] [--version] [--normalize | --dry-run]
               [--as-stream | --stream | --as-manifest | --in-manifest | --manifest | --as-raw | -
-raw]
               [--update-collection UUID] [--use-filename FILENAME]
               [--filename FILENAME] [--portable-data-hash] [--replication N]
               [--storage-classes STORAGE_CLASSES] [--threads N]
               [--exclude PATTERN] [--follow-links | --no-follow-links]
               [--trash-at YYYY-MM-DDTHH:MM | --trash-after DAYS]
               [--project-uuid UUID] [--name NAME]
               [--progress | --no-progress | --batch-progress] [--silent]
               [--resume | --no-resume] [--cache | --no-cache]
               [--retries RETRIES]
               [path [path ...]]
(..)
  --replication N       Set the replication level for the new collection: how
                        many different physical storage devices (e.g., disks)
                        should have a copy of each data block. Default is to
                        use the server-provided default (if any) or 2.
  --storage-classes STORAGE_CLASSES
                        Specify comma separated list of storage classes to be
                        used when saving data to Keep.
```

base casecase

```
arv-put --replication N  --storage-classes STORAGE_CLASSES directory
```

Expected behaviour: ...

updating an existing collection

```
arv-put --replication N  --storage-classes STORAGE_CLASSES directory --update-collection
zzzzz-4zz18-xxxxxxxxxxxxxxx
```

Expected behaviour: ...

giving conflicting options for resume transaction

```
arv-put --replication N  --storage-classes STORAGE_CLASSES directory
arv-put --replication M  --storage-classes DIFFERENT_STORAGE_CLASSES directory --resume
```

Expected behaviour: ...

Subtasks:		
Task # 17736: Review 17351-arvput-keepclient-storage-support		**Resolved**

Related issues:			
Related to Arvados Epics - Story #16107: Storage classes	**Resolved**	**03/01/2021**	**09/30/2021**
Related to Arvados - Story #17465: Support writing blocks to correct storage ...	**Resolved**	**06/01/2021**	

Associated revisions

Revision 523d1c2a - 06/10/2021 09:54 PM - Lucas Di Pentima

Merge branch '17351-arvput-keepclient-storage-support'
Closes #17351

Arvados-DCO-1.1-Signed-off-by: Lucas Di Pentima <lucas.dipentima@curii.com>

History

#1 - 02/08/2021 08:12 PM - Nico César

- Related to Story #16107: Storage classes added

#2 - 02/08/2021 08:13 PM - Nico César

- Target version set to To Be Groomed

- Category set to Keep

#3 - 02/15/2021 07:11 PM - Nico César

- Description updated

- Subject changed from [arv-put] [and other keep clients] Storage tiers design to [arv-put] Storage tiers design

#4 - 02/15/2021 07:12 PM - Nico César

- Description updated

#5 - 02/15/2021 07:45 PM - Nico César

- Subject changed from [arv-put] Storage tiers design to [arv-put] Storage classes revisit

#6 - 02/15/2021 07:51 PM - Nico César

- Subject changed from [arv-put] Storage classes revisit to [arv-put] Storage tiers design

#7 - 02/15/2021 07:51 PM - Nico César

- Subject changed from [arv-put] Storage tiers design to [arv-put] Storage classes revisit

#8 - 03/16/2021 05:11 PM - Lucas Di Pentima

- Target version changed from To Be Groomed to 2021-04-14 sprint

#9 - 03/31/2021 03:42 PM - Peter Amstutz

- Target version changed from 2021-04-14 sprint to 2021-05-26 sprint

#10 - 05/12/2021 04:15 PM - Peter Amstutz

- Subject changed from [arv-put] Storage classes revisit to [arv-put] Storage classes

#11 - 05/12/2021 04:16 PM - Peter Amstutz

- Assigned To set to Lucas Di Pentima

- Subject changed from [arv-put] Storage classes to [arv-put] Storage classes

#12 - 05/26/2021 02:21 PM - Lucas Di Pentima

- Target version changed from 2021-05-26 sprint to 2021-06-09 sprint

#13 - 06/01/2021 02:13 PM - Lucas Di Pentima

- Status changed from New to In Progress

#14 - 06/03/2021 12:32 AM - Lucas Di Pentima

Updates at a0fcd46cb - branch 17351-arvput-keepclient-storage-support
Test run: https://ci.arvados.org/job/developer-run-tests/2508/

- Removes limitation of no more than 1 storage classes. (couldn't find the reason of that limitation, introduced in #13430)
- Passes storage classes data at Collection instantiation time instead of passing it to the .save() or .save_new() methods. This produces that the keep client used to upload files will write to keep directly to the specified classes.

#15 - 06/03/2021 07:55 PM - Tom Clegg

- Related to Story #17465: Support writing blocks to correct storage classes in Python SDK added

#16 - 06/07/2021 04:05 PM - Lucas Di Pentima

Rebased to the latest #17465 changes at 57a26e5
Test run: https://ci.arvados.org/job/developer-run-tests/2510/

#17 - 06/09/2021 02:33 PM - Lucas Di Pentima

- Target version changed from 2021-06-09 sprint to 2021-06-23 sprint

#18 - 06/09/2021 09:33 PM - Lucas Di Pentima

While working on #17572 I realized that making arv-put to honor a previously created collection's desired_storage_classes field may produce surprising results to the user, for example:

1. The user creates an empty collection via the CLI tools, assigning a desired_storage_classes list with nonexistent classes.
2. The RailsAPI will be OK with that, so it gets created.
3. Then, the user executes arv-put without any --storage-classes argument but using --update-collection UUID with the previously created collection's UUID.
4. The user will get an error from arv-put because Keep returns 503 (I think) when a non-valid class is specified. This is because the command will honor the storage class set up on the collection record if none is specified.

If our priority is making sure that keep writes get done on the correct classes or nowhere, I think the solution would be to make RailsAPI or controller error out when a non-valid class is requested. WDYT?

#19 - 06/10/2021 04:04 PM - Tom Clegg

I think the "error because existing collection has unwritable classes" outcome is acceptable. Even if we validate classes when creating/saving a collection, this same condition can happen if all volumes with a given class become read-only, temporarily unreachable, or full.

We should probably check that the error message in such cases is not too confusing, though.

#20 - 06/10/2021 07:12 PM - Tom Clegg

57a26e595 LGTM, thanks

#21 - 06/10/2021 10:38 PM - Lucas Di Pentima

- % Done changed from 0 to 100

- Status changed from In Progress to Resolved

Applied in changeset arvados|523d1c2a9963edc25becf7958e024992ed8a6e66.

#22 - 11/16/2021 04:26 PM - Peter Amstutz

- Release set to 42