Arvados - Bug #17267
Can't run when $schema links are broken
01/18/2021 05:55 PM - Peter Amstutz

Description
Arvados uploads URLs in $schemas and converts them into keep references.

When it goes through and downloads URLs, a broken link is fatal error. This is a problem because these $schemas URLs are nonessential to running the workflow, and schema_salad can ignore broken URLs.

Arvados-cwl-runner needs to be able to ignore broken URLs in $schemas.

Subtasks:
Task #17268: Review 17267-broken-schema-links Resolved

Related issues:
Related to Arvados - Bug #16477: arvados-cwl-runner is missing --skip-schemas Resolved
Related to Arvados - Bug #11257: [CWL] Fails with $schemas referencing remote... Resolved

Associated revisions
Revision 56816829 - 01/19/2021 10:01 PM - Peter Amstutz
Merge branch '17267-broken-schema-links' refs #17267
Arvados-DCO-1.1-Signed-off-by: Peter Amstutz <peter.amstutz@curii.com>

History
#1 - 01/18/2021 05:59 PM - Peter Amstutz
- Description updated

#2 - 01/18/2021 06:57 PM - Peter Amstutz
- Related to Bug #16477: arvados-cwl-runner is missing --skip-schemas added

#3 - 01/18/2021 06:57 PM - Peter Amstutz
- Related to Bug #11257: [CWL] Fails with $schemas referencing remote resources. added

#4 - 01/18/2021 09:05 PM - Peter Amstutz
- Status changed from New to In Progress

#5 - 01/18/2021 09:05 PM - Peter Amstutz
- Assigned To set to Peter Amstutz

#6 - 01/18/2021 09:06 PM - Peter Amstutz
17267-broken-schema-links @ 29a8298caf922a2651fd51baf7b97b58ae25840

#7 - 01/19/2021 02:18 PM - Lucas Di Pentima
- Given that this issue is IMO pretty critical, I think it would be convenient to have it tested.
- The changes on pathmapper.py's visit() seems to be applicable for other callers than the one processing the $schema, maybe it would need an additional argument so that the caller can ask to ignore broken links on certain cases?
- Related Q: Would a broken schema link make validation impossible, or the schemas have other uses?
Lucas Di Pentima wrote:

- Given that this issue is IMO pretty critical, I think it would be convenient to have it tested.

Yes, I wasn't quite sure if a unit test would make sense, but I could add an integration test.

- The changes on pathmapper.py's visit() seems to be applicable for other callers than the one processing the $schema, maybe it would need an additional argument so that the caller can ask to ignore broken links on certain cases?

So, this only applies to things that are referenced in the CWL document with http links, that need to be downloaded on the fly into Arvados.

The prior behavior was that, if a http download failed, it would log an error, throw a fatal exception.

The new behavior is that if a http download fails, it logs an error, does not add an entry in the "mapper". If something tries to look the mapping without checking to see if it exists first, it will throw a fatal exception.

So I think the behavior is functionally the same.

- Related Q: Would a broken schema link make validation impossible, or the schemas have other uses?

No, they are entirely optional. The only thing they are really used for is to check if your annotations appear in the vocabulary of the extension schema, and warn if not found. It is more to assist with spellchecking than anything.

17267-broken-schema-links @ b01ff2414daaf5fd8ff7f0e78ed49e63d431ccd3

Added integration test.

This LGTM, thanks!

- Status changed from In Progress to Resolved

- Release set to 37