

Pulp - Issue #6362

Check for duplicated content happens without plugin input

03/18/2020 03:12 PM - ttereshc

Status: CLOSED - CURRENTRELEASE	Start date:
Priority: Normal	Due date:
Assignee: daviddavis	Estimated time: 0:00 hour
Category:	
Sprint/Milestone: 3.3.0	
Severity: 2. Medium	Groomed: No
Version:	Sprint Candidate: No
Platform Release:	Tags:
OS:	Sprint: Sprint 69
Triaged: Yes	Quarter:
Description Currently, due to #5567 , a check is performed in add_content and the error is raised , plugins might want to deal with duplicates themselves at the finalization stage of the repo version creation. For the cases when a user can't change much (e.g. sync from a remote, copy different versions of the same advisory from multiple repos), Pulp should try its best to solve issues itself without producing an error. The check can be performed after finalize_repo_version is happened but before a repo version is marked as complete.	
Related issues: Related to Pulp - Issue #5567: Content with duplicate repo_key_fields can be ... CLOSED - CURRENTRELEASE Related to Pulp - Issue #6217: First repo_version can contain duplicates CLOSED - CURRENTRELEASE	

Associated revisions

Revision db31633d - 03/20/2020 12:50 PM - daviddavis

Checking for dupe content in finalize

fixes #6362 <https://pulp.plan.io/issues/6362>

Required PR: https://github.com/pulp/pulp_file/pull/368

Revision dc6acb9d - 03/31/2020 12:01 AM - daviddavis

Moving duplicate content check to finalize

fixes #6362 <https://pulp.plan.io/issues/6362>

Required PR: <https://github.com/pulp/pulpcore/pull/599>

Revision 7c032a16 - 03/31/2020 06:30 PM - daviddavis

Moving duplicate content check to finalize

fixes #6362 <https://pulp.plan.io/issues/6362>

Required PR: <https://github.com/pulp/pulpcore/pull/599>

(cherry picked from commit dc6acb9d60a1d3aa61ff4def99cc12ddd9981076)

History

#1 - 03/18/2020 03:46 PM - ttereshc

- Description updated

#2 - 03/18/2020 03:46 PM - ttereshc

- Related to Issue #5567: Content with duplicate repo_key_fields can be added to a repo version added

#3 - 03/18/2020 03:46 PM - ttereshc

- Related to Issue #6217: First repo_version can contain duplicates added

#4 - 03/18/2020 04:01 PM - daviddavis

We should either move this check to `remove_duplicates[0]` which is called from `finalize_new_version` or roll a new repo util method that plugins can call during `finalize` to check for duplicates. The latter gives the plugin writer greater flexibility. However, I can't think of a case where a plugin might want to actually have that flexibility.

[0] https://github.com/pulp/pulpcore/blob/747504c019e618dbb5949ead75144f7d1e731cff/pulpcore/plugin/repo_version_utils.py#L16

#5 - 03/18/2020 04:19 PM - daviddavis

Looks like plugins are calling `remove_duplicates` before resolving `content[0]`. If we go with updating `remove_duplicates`, we'll need to move it to the end.

[0] https://github.com/pulp/pulp_rpm/blob/0464c62ff2b157124a1b86e6fab5a8b562d062fe/pulp_rpm/app/models/repository.py#L118

#6 - 03/19/2020 05:23 PM - daviddavis

- Status changed from NEW to POST

- Assignee set to daviddavis

<https://github.com/pulp/pulpcore/pull/599> https://github.com/pulp/pulp_file/pull/368

#7 - 03/20/2020 03:37 PM - fao89

- Triaged changed from No to Yes

- Sprint set to Sprint 69

#8 - 03/30/2020 09:57 PM - pulpbot

PR: <https://github.com/pulp/pulpcore/pull/599>

#9 - 03/30/2020 11:59 PM - pulpbot

PR: https://github.com/pulp/pulp_file/pull/368

#10 - 03/31/2020 12:01 AM - daviddavis

- Status changed from POST to MODIFIED

Applied in changeset [pulpcore|db31633d5473d69ed8dbf76adeb4377ac2b30e8a](#).

#11 - 03/31/2020 12:01 AM - daviddavis

Applied in changeset [pulp_file:dc6acb9d60a1d3aa61ff4def99cc12ddd9981076](#).

#12 - 03/31/2020 06:30 PM - daviddavis

Applied in changeset [pulp_file:7c032a16021c1fe6e091273d6b8e3d04b3d16b43](#).

#13 - 04/15/2020 09:55 PM - ttereshc

- Status changed from *MODIFIED* to *CLOSED - CURRENTRELEASE*

- *Sprint/Milestone* set to 3.3.0