Pulp - Story #5008

As a user, I can enforce content specific uniqueness constaints in a repo version

06/21/2019 04:30 PM - ttereshc

Status:	CLOSED - CURRENTRELEASE		**Start date:**	
Priority:	Normal		**Due date:**	
Assignee:	bmbouter		**% Done:**	100%
Category:			**Estimated time:**	0:00 hour
Sprint/Milestone:	3.0.0			
Platform Release:			**Tags:**	
Groomed:	Yes		**Sprint:**	Sprint 60
Sprint Candidate:	Yes		**Quarter:**	

Description

RemoveDuplicates stage provides the opportunity to enforce uniqueness constraints for content at the sync time.
The duplicate problem can be encountered at any time content is added to a repository, not only at sync time.
E.g. Content was uploaded, or content was synced as a part of other repo, and now it's added/copied to a new repository.

It would be good if some check/validation happened for any added content, e.g. at this stage

Examples for where uniqueness might be needed in a repo version:
- only one content unit with a certain characteristics should be present in a repo (only one advisory with the same id, only one module_defaults for a module, etc)

This issue was discussed on this thread: https://www.redhat.com/archives/pulp-dev/2019-May/msg00061.html

Solution

Plugins:
On the plugin content model define a repo_key - one or more fields which must be unique within a repo version.

Pulpcore:
Check uniqueness of the repo_key for each added content type at the repository version creation time
https://github.com/pulp/pulpcore/blob/aef490e201f89fc005ba3239fda3a79c05e28fd7/pulpcore/app/models/repository.py#L343.
Whether it's sync, copy, or upload, repo_key uniqueness will be ensured if core/plugin devs use with RepositoryVersion.create(...) context manager.

Related issues:

Related to File Support - Task #4028: Prevent duplicate files in repositories	**CLOSED - CURRENTRELEASE**
Related to RPM Support - Story #4898: As a user, no "duplicated" content is p...	**CLOSED - CURRENTRELEASE**

Associated revisions

Revision b41a0923 - 10/08/2019 08:55 PM - bmbouter

Add repo_key support for add_content

The Content model now supports a repo_key attribute which defaults to (). Plugin writers can specify this, and the tuple field names will automatically replace another unit of the same type with these field names.

Required PR: https://github.com/pulp/pulp_file/pull/290

https://pulp.plan.io/issues/5008 closes #5008

History

#1 - 06/21/2019 04:32 PM - ttereshc

- Description updated

#2 - 06/21/2019 04:34 PM - ttereshc

- Related to Task #4028: Prevent duplicate files in repositories added

#3 - 06/21/2019 04:34 PM - ttereshc

- *Related to Story #4898: As a user, no "duplicated" content is present in a repo version added*

#4 - 06/25/2019 04:48 PM - amacdona@redhat.com

- *Triaged changed from No to Yes*

#5 - 07/30/2019 03:36 PM - ttereshc

- *Description updated*

#6 - 07/30/2019 03:38 PM - ttereshc

- *Description updated*

#7 - 07/30/2019 04:50 PM - amacdona@redhat.com

- *Sprint set to Sprint 56*

#8 - 07/30/2019 06:31 PM - bmbouter

This all looks good. Should we convert this to a story? To me in the changelog this would make the most sense as a new feature in the pulpcore-plugin package. What do you think?

#9 - 07/31/2019 03:40 PM - ttereshc

- *Tracker changed from Issue to Story*

- *Subject changed from No way to enforce content specific uniqueness constaints in a repo version to As a user, I can enforce content specific uniqueness constaints in a repo version*

- *% Done set to 0*

#10 - 07/31/2019 04:07 PM - bmbouter

- *Groomed changed from No to Yes*

- *Sprint Candidate changed from No to Yes*

This looks good to me. Thank you for recapping ttereshc!

#11 - 08/02/2019 03:14 PM - rchan

- *Sprint changed from Sprint 56 to Sprint 57*

#12 - 08/26/2019 04:31 PM - rchan

- *Sprint changed from Sprint 57 to Sprint 58*

#13 - 09/10/2019 08:53 PM - bmbouter

- *Status changed from NEW to ASSIGNED*

- *Assignee set to bmbouter*

#14 - 09/13/2019 04:59 PM - rchan

- *Sprint changed from Sprint 58 to Sprint 59*

#15 - 09/27/2019 12:55 PM - daviddavis

- *Sprint/Milestone set to 3.0.0*

#16 - 09/27/2019 03:46 PM - rchan

- *Sprint changed from Sprint 59 to Sprint 60*

#17 - 10/08/2019 12:06 AM - bmbouter

- *Status changed from ASSIGNED to POST*

PR available at: https://github.com/pulp/pulpcore/pull/331

#18 - 10/08/2019 09:52 PM - bmbouter

- Status changed from POST to MODIFIED

- % Done changed from 0 to 100

Applied in changeset [pulpcore|b41a0923183c4316401349fb4f823b76433e9ed9](#).

#19 - 12/13/2019 06:28 PM - bmbouter

- Status changed from MODIFIED to CLOSED - CURRENTRELEASE