

## Pulp - Story #7696

As a plugin developer, I have the Artifacts checked at pre-save time against the **ALLOWED\_CONTENT\_CHECKSUMS** instead of at `__init__`

10/12/2020 09:05 PM - bmbouter

<b>Status:</b>	CLOSED - CURRENTRELEASE	<b>Start date:</b>	
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	ggainey	<b>% Done:</b>	100%
<b>Category:</b>		<b>Estimated time:</b>	0:00 hour
<b>Sprint/Milestone:</b>	3.9.0	<b>Tags:</b>	
<b>Platform Release:</b>		<b>Sprint:</b>	Sprint 86
<b>Groomed:</b>	Yes	<b>Quarter:</b>	
<b>Sprint Candidate:</b>	Yes		

### Description

## Background

Right now the `Artifact.init` checks to see if checksums that are not allowed are being used and raises an error. That happens [here](#).

This has a problem when plugin code (like in `pulp_rpm` for example) does this:

```
foo = Artifact()
foo.md5 = 'some_md5_string'
foo.save()
```

## Improvement

Move the checking currently in `Artifact.__init__` to a `Artifact.pre_save` hook using `django-lifecycle`.

## Alternative

Add a property to each of the checksums on `Artifact` that checks if it was set post construction by `init`. This keeps the check in two places and it's more code to carry/maintain. These downsides make it the alternative.

### Related issues:

Related to Pulp - Story #5216: As a user, I can configure which checksum type...

**CLOSED - CURRENTRELEASE**

Related to Container Support - Issue #7774: `podman push` leads to missing ch...

**CLOSED - CURRENTRELEASE**

Related to Pulp - Story #7561: As a user, I can add checksums to `ALLOWED_CONT...`

**CLOSED - CURRENTRELEASE**

### Associated revisions

#### Revision 1f003206 - 12/01/2020 09:15 PM - ggainey

Added `pre_save` hook to `Artifact` to enforce checksum rules.

Closes #7696

### History

#### #1 - 10/12/2020 09:07 PM - daviddavis

- Related to Story #5216: As a user, I can configure which checksum types I want to use in Pulp added

#### #2 - 10/12/2020 09:08 PM - daviddavis

- Groomed changed from No to Yes

- Sprint Candidate changed from No to Yes

#### #3 - 11/02/2020 07:19 PM - daviddavis

- Related to Issue #7774: `podman push` leads to missing checksums on the Artifacts added

**#4 - 11/02/2020 07:28 PM - ggaaney**

We need to do more than the current `init` behavior of disallowing-FORBIDDEN. We need the presave to make sure that the Artifact has provided values for everything in `ALLOWED_` as well, to fix the issue in [#7774](#)

**#5 - 11/02/2020 07:34 PM - ipanova@redhat.com**

ggaaney wrote:

We need to do more than the current `init` behavior of disallowing-FORBIDDEN. We need the presave to make sure that the Artifact has provided values for everything in `ALLOWED_` as well, to fix the issue in [#7774](#)

In addition to this, a migration should be provided which would check for the artifacts that miss checksums specified in the `ALLOWED_` and populate missing data.

**#6 - 11/02/2020 07:51 PM - bmbouter**

Agreed with Comment 4, +1 to the pre-save approach.

For Comment 5, I agree users need some way to populate this data when it's already on-disk, but if we ship a migration they can only adjust it once before the migration runs. We likely need a django command that will "recompute missing checksums" or something like that. Something an administrator can run when Pulp refuses to start because their settings and their db's checksums do not match.

**#7 - 11/02/2020 08:28 PM - daviddavis**

For Comment 5, I agree users need some way to populate this data when it's already on-disk, but if we ship a migration they can only adjust it once before the migration runs. We likely need a django command that will "recompute missing checksums" or something like that. Something an administrator can run when Pulp refuses to start because their settings and their db's checksums do not match.

I think this is a separate use case. Yes, we need a way for users to populate this data when they make changes to `ALLOWED_CONTENT_CHECKSUMS`.

However, currently in 3.7 and 3.8, plugins can write null values to artifact checksum columns. Then when pulp restarts, it encounters the exception in [#7487](#). This is what is happening in [#7774](#). I think we need a data migration to fix this use case. The question though is whether pulpcore or the plugin is responsible for this.

**#8 - 11/03/2020 01:20 PM - ipanova@redhat.com**

daviddavis wrote:

For Comment 5, I agree users need some way to populate this data when it's already on-disk, but if we ship a migration they can only adjust it once before the migration runs. We likely need a django command that will "recompute missing checksums" or something like that. Something an administrator can run when Pulp refuses to start because their settings and their db's checksums do not match.

I think this is a separate use case. Yes, we need a way for users to populate this data when they make changes to `ALLOWED_CONTENT_CHECKSUMS`.

In my understanding once you have pulp running you cannot make additive changes but only remove checksums.

However, currently in 3.7 and 3.8, plugins can write null values to artifact checksum columns. Then when pulp restarts, it encounters the exception in [#7487](#). This is what is happening in [#7774](#). I think we need a data migration to fix this use case. The question though is whether pulpcore or the plugin is responsible for this.

It will probably be easier to have this migration in core since it can filter through the all artifacts only those that miss checksums. If a plugin would do that he would need to figure out through all the content it has, the artifacts that belong to it and then look for those that miss checksums.

**#9 - 11/03/2020 01:39 PM - daviddavis**

[ipanova@redhat.com](mailto:ipanova@redhat.com) wrote:

daviddavis wrote:

For Comment 5, I agree users need some way to populate this data when it's already on-disk, but if we ship a migration they can only adjust it once before the migration runs. We likely need a django command that will "recompute missing checksums" or something like that. Something an administrator can run when Pulp refuses to start because their settings and their db's checksums do not match.

I think this is a separate use case. Yes, we need a way for users to populate this data when they make changes to `ALLOWED_CONTENT_CHECKSUMS`.

In my understanding once you have pulp running you cannot make additive changes but only remove checksums.

That's the current state today but see <https://pulp.plan.io/issues/7561>. Allowing users to populate missing checksums would have to be part of [#7561](#).

However, currently in 3.7 and 3.8, plugins can write null values to artifact checksum columns. Then when pulp restarts, it encounters the exception in [#7487](#). This is what is happening in [#7774](#). I think we need a data migration to fix this use case. The question though is whether pulpcore or the plugin is responsible for this.

It will probably be easier to have this migration in core since it can filter through the all artifacts only those that miss checksums. If a plugin would do that he would need to figure out through all the content it has, the artifacts that belong to it and then look for those that miss checksums.

My only misgiving here is that users would only run this migration once and a plugin might have a bug where it continues to insert blank/null checksums (even if we add this pre-save check as things like bulk\_create wouldn't hit it). If a plugin fixes such a bug, it should probably fix the data too.

That said, if we fix <https://pulp.plan.io/issues/7561>, plugins can just ask users to run whatever solution we come up with to fix their missing data so maybe we should address [#7561](#) with this story.

**#10 - 11/03/2020 01:50 PM - ipanova@redhat.com**

daviddavis wrote:

[ipanova@redhat.com](mailto:ipanova@redhat.com) wrote:

daviddavis wrote:

For Comment 5, I agree users need some way to populate this data when it's already on-disk, but if we ship a migration they can only adjust it once before the migration runs. We likely need a django command that will "recompute missing checksums" or something like that. Something an administrator can run when Pulp refuses to start because their settings and their db's checksums do not match.

I think this is a separate use case. Yes, we need a way for users to populate this data when they make changes to ALLOWED\_CONTENT\_CHECKSUMS.

In my understanding once you have pulp running you cannot make additive changes but only remove checksums.

That's the current state today but see <https://pulp.plan.io/issues/7561>. Allowing users to populate missing checksums would have to be part of [#7561](#).

However, currently in 3.7 and 3.8, plugins can write null values to artifact checksum columns. Then when pulp restarts, it encounters the exception in [#7487](#). This is what is happening in [#7774](#). I think we need a data migration to fix this use case. The question though is whether pulpcore or the plugin is responsible for this.

It will probably be easier to have this migration in core since it can filter through the all artifacts only those that miss checksums. If a plugin would do that he would need to figure out through all the content it has, the artifacts that belong to it and then look for those that miss checksums.

My only misgiving here is that users would only run this migration once and a plugin might have a bug where it continues to insert blank/null checksums (even if we add this pre-save check as things like bulk\_create wouldn't hit it). If a plugin fixes such a bug, it should probably fix the data too.

if we deliver migration together with [#7696](#) and [#7561](#) it will 1) fix/populate all existing artifacts with missing checksums and 2) ~~make it impossible to create further artifacts with missing checksums~~, so even if the plugin has the bug, he would need to first fix it in the plugin and only after that he would be able to create artifact

Forgot about bulk\_create(). At this point it makes sense that plugins would provide the migration, because as you have stated even if we provide the core migration, plugin would need to fix the issue eventually and then again provide another migration.

That said, if we fix <https://pulp.plan.io/issues/7561>, plugins can just ask users to run whatever solution we come up with to fix their missing data so maybe we should address [#7561](#) with this story.

**#11 - 11/03/2020 02:01 PM - ipanova@redhat.com**

- Related to Story #7561: As a user, I can add checksums to ALLOWED\_CONTENT\_CHECKSUMS added

**#12 - 11/04/2020 07:36 PM - ipanova@redhat.com**

- Checklist item [ ] add django command that will recompute missing checksums added

**#13 - 11/04/2020 07:45 PM - ipanova@redhat.com**

- Sprint/Milestone set to 3.9.0

- Sprint set to Sprint 85

**#14 - 11/04/2020 07:46 PM - ipanova@redhat.com**

We should provide django command that will recompute and add missing checksums, data migration whether written in core or plugin is not a viable solution

**#15 - 11/04/2020 11:10 PM - ipanova@redhat.com**

- Checklist item deleted (add django command that will recompute missing checksums)

**#16 - 11/13/2020 09:53 PM - rchan**

- Sprint changed from Sprint 85 to Sprint 86

**#17 - 11/14/2020 10:18 PM - pulpbot**

- Status changed from NEW to POST

PR: <https://github.com/pulp/pulpcore/pull/1027>

**#18 - 11/19/2020 09:00 PM - daviddavis**

- Assignee set to ggainey

**#19 - 12/01/2020 09:15 PM - ggainey**

- Status changed from POST to MODIFIED

- % Done changed from 0 to 100

Applied in changeset [pulpcore|1f003206393c33bc451d9ab64601105099bbd087](#).

**#20 - 12/07/2020 08:10 PM - pulpbot**

- Status changed from MODIFIED to CLOSED - CURRENTRELEASE