

RPM Support - Issue #4477

Duplicated empty collections are published with subsequent sync/publish of modular errata

02/26/2019 07:14 PM - ragbalak

Status:	CLOSED - CURRENTRELEASE	Start date:	
Priority:	High	Due date:	
Assignee:	ttereshc	Estimated time:	0:00 hour
Category:		Groomed:	No
Sprint/Milestone:	2.19.0	Sprint Candidate:	No
Severity:	2. Medium	Tags:	Pulp 2
Version:	2.18.1	Sprint:	Sprint 50
Platform Release:	2.19.0	Quarter:	
OS:			
Triaged:	Yes		

Description

Problem

When a repository with feed "<https://repos.fedorapeople.org/pulp/pulp/fixtures/rpm-with-modules/>" is created, synced and published, the updateinfo.xml of the published repo contains an inconsistent number of modules.

Steps to reproduce :

Create and sync a repo pointing to the rpm-with-modules fixture.

```
pulp-admin rpm repo create --feed https://repos.fedorapeople.org/pulp/pulp/fixtures/rpm-with-modules/ --repo-id test-repo
```

```
pulp-admin rpm repo sync run --repo-id test-repo
```

After the repo is synced, get the number of module tags present in the repo

```
# note the repo id and the timestamp directory after that can change
zcat 8147252315a121c89a1be2058fe3367a52bf81dac61788e53358c271421cd9d6-updateinfo.xml.gz |grep "module" | wc -l
33 # current output of the number of modules in the update_info.xml
```

If the above repository is deleted, and then again the same steps are followed for a new repo, the total number of modules in the new repo are seen to be increased

```
pulp-admin rpm repo delete --repo-id test-repo
pulp-admin rpm repo create --feed https://repos.fedorapeople.org/pulp/pulp/fixtures/rpm-with-modules/ --repo-id test-repo-new
```

```
pulp-admin rpm repo sync run --repo-id test-repo-new
```

```
zcat ef13ff9115725082dd55fef6251880d28d66c1b0acf28abf89ba7fd2b1ade96f-updateinfo.xml.gz |grep module |wc -l
36 # current output of the number of modules in the update_info.xml
```

As seen the number of modules have increased for the same feed url with different repo names.

Also the update-info.xml of the published repo is different comparing with the actual update-info.xml of the fixtures repo.

Have attached the update-info.xml of both the syncs and the fixtures repo.

Environment details :

```
# cat /etc/os-release
NAME="Red Hat Enterprise Linux Server"
```

```
VERSION="7.6 (Maipo) "
ID="rhel"
ID_LIKE="fedora"
VARIANT="Server"
VARIANT_ID="server"
VERSION_ID="7.6"
PRETTY_NAME="Employee SKU"
ANSI_COLOR="0;31"
CPE_NAME="cpe:/o:redhat:enterprise_linux:7.6:GA:server"
HOME_URL="https://www.redhat.com/"
BUG_REPORT_URL="https://bugzilla.redhat.com/"
```

```
REDHAT_BUGZILLA_PRODUCT="Red Hat Enterprise Linux 7"
REDHAT_BUGZILLA_PRODUCT_VERSION=7.6
REDHAT_SUPPORT_PRODUCT="Red Hat Enterprise Linux"
REDHAT_SUPPORT_PRODUCT_VERSION="7.6"
```

```
# rpm -qa|grep pulp|sort
libsolv-0.6.34-2.pulp.el7.x86_64
pulp-admin-client-2.19.0-0.1.alpha.201812120516git129f7b2.el7.noarch
pulp-deb-admin-extensions-1.9.0-0.1.alpha.201812120524giteefc050.el7.noarch
pulp-deb-plugins-1.9.0-0.1.alpha.201812120524giteefc050.el7.noarch
pulp-docker-admin-extensions-3.4.0-0.1.alpha.201812120503gitae36e5.el7.noarch
pulp-docker-plugins-3.4.0-0.1.alpha.201812120503gitae36e5.el7.noarch
pulp-ostree-admin-extensions-1.4.0-0.1.alpha.201812120505git6c63638.el7.noarch
pulp-ostree-plugins-1.4.0-0.1.alpha.201812120505git6c63638.el7.noarch
pulp-puppet-admin-extensions-2.19.0-0.1.alpha.201812120507gitc8c9a9a.el7.noarch
pulp-puppet-plugins-2.19.0-0.1.alpha.201812120507gitc8c9a9a.el7.noarch
pulp-puppet-tools-2.19.0-0.1.alpha.201812120507gitc8c9a9a.el7.noarch
pulp-python-admin-extensions-2.1.0-0.1.alpha.201812120501git5e2aa35.el7.noarch
pulp-python-plugins-2.1.0-0.1.alpha.201812120501git5e2aa35.el7.noarch
pulp-rpm-admin-extensions-2.19.0-0.1.alpha.201812120511git9bc9e73.el7.noarch
pulp-rpm-plugins-2.19.0-0.1.alpha.201812120511git9bc9e73.el7.noarch
pulp-selinux-2.19.0-0.1.alpha.201812120516git129f7b2.el7.noarch
pulp-server-2.19.0-0.1.alpha.201812120516git129f7b2.el7.noarch
python2-solv-0.6.34-2.pulp.el7.x86_64
python-isodate-0.5.0-4.pulp.el7.noarch
python-pulp-bindings-2.19.0-0.1.alpha.201812120516git129f7b2.el7.noarch
python-pulp-client-lib-2.19.0-0.1.alpha.201812120516git129f7b2.el7.noarch
python-pulp-common-2.19.0-0.1.alpha.201812120516git129f7b2.el7.noarch
python-pulp-deb-common-1.9.0-0.1.alpha.201812120524giteefc050.el7.noarch
python-pulp-docker-common-3.4.0-0.1.alpha.201812120503gitae36e5.el7.noarch
python-pulp-oid_validation-2.19.0-0.1.alpha.201812120516git129f7b2.el7.noarch
python-pulp-ostree-common-1.4.0-0.1.alpha.201812120505git6c63638.el7.noarch
python-pulp-puppet-common-2.19.0-0.1.alpha.201812120507gitc8c9a9a.el7.noarch
python-pulp-python-common-2.1.0-0.1.alpha.201812120501git5e2aa35.el7.noarch
python-pulp-repoauth-2.19.0-0.1.alpha.201812120516git129f7b2.el7.noarch
python-pulp-rpm-common-2.19.0-0.1.alpha.201812120511git9bc9e73.el7.noarch
python-pulp-streamer-2.19.0-0.1.alpha.201812120516git129f7b2.el7.noarch
```

Related issues:

Related to Pulp - Test #4112: Verify module information on errata search api

CLOSED - COMPLETE

Related to RPM Support - Issue #4056: module information missing from errata ...

CLOSED - CURRENTRELEASE

Associated revisions

Revision 641750dc - 03/18/2019 11:00 AM - ttereshc

Handle duplicated modular collections in errata at publish time

closes #4477 <https://pulp.plan.io/issues/4477>

Revision f0ec47d9 - 03/25/2019 02:27 PM - Bruno Rocha

Adding tests for Modular content module counts after republish

Ensure module count is the same after republishing 2 different repositories with the same feed.

closes #4477

History

#1 - 02/26/2019 07:53 PM - ragbalak

- File `update-info.xml.secdone` added
- File `update-info.xml.firstone` added
- File `update-info.xml.fixture` added
- Subject changed from `Updateinfo information contains` to `Updateinfo information of published repo inconsistent with the actual update info`
- Description updated

#2 - 02/28/2019 04:26 PM - kersom

- Related to `Test #4112: Verify module information on errata search api` added

#3 - 02/28/2019 04:28 PM - kersom

- Related to `Issue #4056: module information missing from errata search api` added

#4 - 03/01/2019 04:35 PM - CodeHeeler

- Project changed from `Pulp` to `RPM Support`

#5 - 03/04/2019 09:21 AM - ttereshc

[ragbalak](#), do you see the same behavior on 2.18.1? Or is it specific to 2.19?

#6 - 03/04/2019 09:22 AM - ttereshc

- Subject changed from `Updateinfo information of published repo inconsistent with the actual update info` to `Duplicated empty collections are published with subsequent sync/publish of modular errata`
- Priority changed from `Normal` to `High`

#7 - 03/05/2019 04:33 PM - CodeHeeler

- Triaged changed from `No` to `Yes`

#8 - 03/07/2019 03:41 PM - ttereshc

- Sprint set to `Sprint 49`

#9 - 03/08/2019 11:41 PM - rchan

- Sprint changed from `Sprint 49` to `Sprint 50`

#10 - 03/12/2019 03:18 PM - bherring

Additional Investigation Notes

The same behavior exists in 2.18.1.

As a note from Raga's investigation, removing orphans between runs will stop this behavior, but it seems like a non-intended design decision.

#12 - 03/14/2019 05:43 PM - ttereshc

- Version set to `2.18.1`

#13 - 03/14/2019 06:01 PM - ttereshc

- Status changed from `NEW` to `ASSIGNED`
- Assignee set to `ttereshc`

#14 - 03/15/2019 10:28 AM - ttereshc

I think I know why it is happening.
For every new repository if it synced from feed, erratum collection is created, even if it's a duplicate. That's expected.

The logic of publishing an erratum is complicated and removal of duplicates is a part of it.
Erratum collection won't be published if it's non-modular and has no packages in it.
Erratum collection will be published if it's modular and even if it has no packages.
I guess, the duplicated RPMs are removed, while modular data is still in the collection, so all collections with modular info are published.

#15 - 03/15/2019 04:49 PM - ttereshc

- Sprint/Milestone set to 2.19.0
- Platform Release set to 2.19.0

#16 - 03/18/2019 11:02 AM - ttereshc

- Status changed from ASSIGNED to POST

https://github.com/pulp/pulp_rpm/pull/1301

#17 - 03/18/2019 02:05 PM - ttereshc

- Status changed from POST to MODIFIED

Applied in changeset [641750dc0c1e6eadabed27c86ac676de90318b27](#).

#19 - 03/20/2019 10:03 AM - ttereshc

- Status changed from MODIFIED to 5

#20 - 04/02/2019 10:52 PM - ttereshc

- Status changed from 5 to CLOSED - CURRENTRELEASE

#21 - 04/15/2019 10:01 PM - bmbouter

- Tags Pulp 2 added

Files

update-info.xml.secdone	12.1 KB	02/26/2019	ragbalak
update-info.xml.firstone	11.4 KB	02/26/2019	ragbalak
update-info.xml.fixture	4.26 KB	02/26/2019	ragbalak