

RPM Support - Story #2788

As a user i can configure removal of old published repodata

06/02/2017 05:58 AM - rmcgover

Status:	CLOSED - CURRENTRELEASE	Start date:	
Priority:	Urgent	Due date:	
Assignee:	jluza	% Done:	0%
Category:		Estimated time:	0:00 hour
Sprint/Milestone:		Tags:	Pulp 2
Platform Release:	2.15.0	Sprint:	
Groomed:	No	Quarter:	
Sprint Candidate:	No		
Description			
<p>When a yum repo is published, and sqlite generation is enabled (generate_sqlite=true), and incremental publish is activated (i.e. force_full is not true and units were not deleted since last publish), new sqlite files will be created and old sqlite files will be retained.</p> <p>Since the old sqlite files are never cleaned, this means that the published repodata of a repo will keep all old sqlite files since the last non-incremental publish, which wastes space and makes publishes slower.</p> <p>Steps to reproduce:</p> <ul style="list-style-type: none">• Create repo, ensure it has a yum_distributor with generate_sqlite: true• Add RPM to repo• Publish repo• Note count of *-primary.sqlite.bz2 files in published repo• Add RPM to repo• Publish repo• Note count of *-primary.sqlite.bz2 files in published repo <p>Actual result:</p> <ul style="list-style-type: none">• After first publish, 1 primary sqlite file exists• After second publish, 2 primary sqlite files exist <p>Expected result:</p> <ul style="list-style-type: none">• After first publish, 1 primary sqlite file exists• After second publish:<ul style="list-style-type: none">◦ If https://pulp.plan.io/issues/1684 is implemented, sqlite files should be subject to the configured retention options. Thus it depends on the age of the existing sqlite file.◦ Otherwise, to be consistent with XML handling, the old sqlite files should probably be deleted, so only 1 primary sqlite file exists.			
Related issues:			
Related to RPM Support - Issue #3551: RemoveOldRepodataStep for yum publisher...		CLOSED - CURRENTRELEASE	
Related to Pulp - Issue #5573: Publish won't create multiple checkchecksummed c...		CLOSED - CURRENTRELEASE	

Associated revisions

Revision 7457f4ab - 08/11/2017 02:00 PM - jluza

RemoveOldRepodataStep for yum publisher

New publish step removes repodata older than threshold which is by default 14 days.

closes #2788 <https://pulp.plan.io/issues/2788>

Revision 15bf5854 - 08/17/2017 11:42 AM - ttereshc

Fix release notes for 2.14 and 2.15 releases

re #2788 <https://pulp.plan.io/issues/2788>

Revision d1fc3eee - 12/15/2017 03:25 PM - werwty

Add remote_old_repodata and remove_old_repodata_threshold to optional config.

re #2788 <https://pulp.plan.io/issues/2788>

Revision 5ac5f94f - 12/18/2017 04:48 PM - werwty

Add remote_old_repodata and remove_old_repodata_threshold to optional config.

re #2788 <https://pulp.plan.io/issues/2788>

(cherry picked from commit d1fc3eeea10792e1bff8c4d3b84751f5a8c13e8f)

History

#1 - 06/02/2017 05:58 AM - rmcgover

- Description updated

#2 - 06/02/2017 02:39 PM - ipanova@redhat.com

The publishes that will get slower, were meant to be done with rsync distributor.

#3 - 06/02/2017 02:59 PM - ipanova@redhat.com

i had a chat with [jluza](#), we believe that If we fix this issue <https://pulp.plan.io/issues/2783>, the current one will not have slowdown in the rsync publish, because the file would have preserved the mtime and would not get rsynced

#4 - 06/02/2017 04:44 PM - ttereshc

- Triaged changed from No to Yes

#5 - 06/05/2017 01:44 AM - rmcgover

Both yum and rsync publishes will become slower due to this (though rsync is worse affected).

For example, we had a repo with a repodata directory containing about 28GB worth of sqlite files and that's enough to noticeably slow down the "copy files" step in yum publishes.

#6 - 06/12/2017 02:33 PM - ipanova@redhat.com

- Status changed from NEW to ASSIGNED

- Assignee set to [jluza](#)

#8 - 06/16/2017 04:33 PM - jluza

- Status changed from ASSIGNED to POST

PR https://github.com/pulp/pulp_rpm/pull/1056

#9 - 08/15/2017 05:33 PM - jluza

- Status changed from POST to MODIFIED

Applied in changeset [7457f4ab3fae861ac590ad584b64d655d8c91074](#).

#10 - 08/15/2017 05:49 PM - ipanova@redhat.com

- Tracker changed from Issue to Story

- Subject changed from *sqlite files are retained indefinitely* to *As a user i can configure removal of old published repodata*

- % Done set to 0

#12 - 11/29/2017 03:22 AM - pcreech

- Platform Release set to 2.15.0

#13 - 11/30/2017 07:29 PM - pcreech

- Status changed from MODIFIED to 5

#14 - 12/12/2017 11:25 PM - Ichimonji10

- Status changed from 5 to ASSIGNED

Tested against Fedora 25 and Fedora 26 with Pulp 2.15 beta 2 installed. Here's the relevant packages on F25:

```
[root@fedora-25-pulp-2-15-beta ~]# rpm -qa | grep pulp | sort
pulp-admin-client-2.15.0-0.2.beta.fc25.noarch
pulp-deb-admin-extensions-1.6.0-0.2.beta.fc25.noarch
pulp-deb-plugins-1.6.0-0.2.beta.fc25.noarch
pulp-docker-admin-extensions-3.1.0-0.3.beta.fc25.noarch
pulp-docker-plugins-3.1.0-0.3.beta.fc25.noarch
pulp-ostree-admin-extensions-1.3.0-1.fc25.noarch
pulp-ostree-plugins-1.3.0-1.fc25.noarch
pulp-puppet-admin-extensions-2.15.0-0.2.beta.fc25.noarch
pulp-puppet-plugins-2.15.0-0.2.beta.fc25.noarch
pulp-puppet-tools-2.15.0-0.2.beta.fc25.noarch
pulp-python-admin-extensions-2.0.2-1.fc25.noarch
pulp-python-plugins-2.0.2-1.fc25.noarch
pulp-rpm-admin-extensions-2.15.0-0.2.beta.fc25.noarch
pulp-rpm-plugins-2.15.0-0.2.beta.fc25.noarch
pulp-selinux-2.15.0-0.2.beta.fc25.noarch
pulp-server-2.15.0-0.2.beta.fc25.noarch
python-kombu-3.0.33-8.pulp.fc25.noarch
python-pulp-bindings-2.15.0-0.2.beta.fc25.noarch
python-pulp-client-lib-2.15.0-0.2.beta.fc25.noarch
python-pulp-common-2.15.0-0.2.beta.fc25.noarch
python-pulp-deb-common-1.6.0-0.2.beta.fc25.noarch
python-pulp-docker-common-3.1.0-0.3.beta.fc25.noarch
python-pulp-oid_validation-2.15.0-0.2.beta.fc25.noarch
python-pulp-ostree-common-1.3.0-1.fc25.noarch
python-pulp-puppet-common-2.15.0-0.2.beta.fc25.noarch
python-pulp-python-common-2.0.2-1.fc25.noarch
python-pulp-repoauth-2.15.0-0.2.beta.fc25.noarch
python-pulp-rpm-common-2.15.0-0.2.beta.fc25.noarch
python-pulp-streamer-2.15.0-0.2.beta.fc25.noarch
```

And F26:

```
[root@fedora-26-pulp-2-15-beta ~]# rpm -qa | grep pulp | sort
pulp-admin-client-2.15.0-0.2.beta.fc26.noarch
pulp-deb-admin-extensions-1.6.0-0.2.beta.fc26.noarch
pulp-deb-plugins-1.6.0-0.2.beta.fc26.noarch
pulp-docker-admin-extensions-3.1.0-0.3.beta.fc26.noarch
pulp-docker-plugins-3.1.0-0.3.beta.fc26.noarch
pulp-ostree-admin-extensions-1.3.0-1.fc26.noarch
pulp-ostree-plugins-1.3.0-1.fc26.noarch
pulp-puppet-admin-extensions-2.15.0-0.2.beta.fc26.noarch
pulp-puppet-plugins-2.15.0-0.2.beta.fc26.noarch
pulp-puppet-tools-2.15.0-0.2.beta.fc26.noarch
pulp-python-admin-extensions-2.0.2-1.fc26.noarch
pulp-python-plugins-2.0.2-1.fc26.noarch
pulp-rpm-admin-extensions-2.15.0-0.2.beta.fc26.noarch
pulp-rpm-plugins-2.15.0-0.2.beta.fc26.noarch
pulp-selinux-2.15.0-0.2.beta.fc26.noarch
pulp-server-2.15.0-0.2.beta.fc26.noarch
python-pulp-bindings-2.15.0-0.2.beta.fc26.noarch
python-pulp-client-lib-2.15.0-0.2.beta.fc26.noarch
python-pulp-common-2.15.0-0.2.beta.fc26.noarch
python-pulp-deb-common-1.6.0-0.2.beta.fc26.noarch
python-pulp-docker-common-3.1.0-0.3.beta.fc26.noarch
python-pulp-oid_validation-2.15.0-0.2.beta.fc26.noarch
python-pulp-ostree-common-1.3.0-1.fc26.noarch
python-pulp-puppet-common-2.15.0-0.2.beta.fc26.noarch
python-pulp-python-common-2.0.2-1.fc26.noarch
python-pulp-repoauth-2.15.0-0.2.beta.fc26.noarch
python-pulp-rpm-common-2.15.0-0.2.beta.fc26.noarch
python-pulp-streamer-2.15.0-0.2.beta.fc26.noarch
```

Here's the test script:

```
#!/usr/bin/bash
set -eou pipefail

readonly rpms=(
  'https://repos.fedorapeople.org/pulp/pulp/fixtures/rpm-unsigned/bear-4.1-1.noarch.rpm'
  'https://repos.fedorapeople.org/pulp/pulp/fixtures/rpm-unsigned/camel-0.1-1.noarch.rpm'
```

```
'https://repos.fedorapeople.org/pulp/pulp/fixtures/rpm-unsigned/cat-1.0-1.noarch.rpm'
)
counts=()

pulp-admin login -u admin -p admin
pulp-admin rpm repo create --repo-id foo --generate-sqlite true
for rpm in "${rpms[@]"; do
    wget "$rpm"
    pulp-admin rpm repo uploads rpm --repo-id foo --file "$(basename "$rpm")"
    pulp-admin rpm repo publish run --repo-id foo
    find /var/lib/pulp/published/yum/master -type f -name '*-primary.sqlite.bz2'
    counts+=("${find /var/lib/pulp/published/yum/master -type f -name '*-primary.sqlite.bz2' | wc --lines}")
done
pulp-admin rpm repo delete --repo-id foo
echo "${counts[*]}"
```

Here's the final output on F25:

```
1 2 3
```

And on F26:

```
1 2 3
```

#15 - 12/12/2017 11:26 PM - Ichimonji10

- File *fedora-26-pulp-2-15-beta.txt* added

#16 - 12/13/2017 03:51 PM - ipanova@redhat.com

@jeremy did you configure and notice the new options for the distributor?

https://github.com/pulp/pulp_rpm/pull/1056/files#diff-7cb9179c56cd2d7232d3e801a56ffceaR631

You did not mention any of those in your testing.

#17 - 12/13/2017 03:56 PM - Ichimonji10

No, I didn't use any of those new options. I ran the script listed in my previous comment. I'll re-test.

#18 - 12/13/2017 03:58 PM - Ichimonji10

I performed testing in the way that I did because of what the issue states:

Expected result:

After first publish, 1 primary sqlite file exists. After second publish:

- If <https://pulp.plan.io/issues/1684> is implemented, sqlite files should be subject to the configured retention options. Thus it depends on the age of the existing sqlite file.
- Otherwise, to be consistent with XML handling, the old sqlite files should probably be deleted, so only 1 primary sqlite file exists.

Is it fair to say that Pulp 2.15 doesn't act according to this description? If so, that's OK. I just want to have that explicit confirmation.

#19 - 12/13/2017 10:42 PM - Ichimonji10

Fail, again. When one attempts to configure a distributor with the `remove_old_repodata` or `remove_old_repodata_threshold` options, an error is returned. Peeking in the logs shows errors like:

```
pulp.plugins.pulp_rpm.plugins.distributors.yum.configuration:ERROR: Configuration key [remove_old_repodata] is
not supported
pulp.server.controllers.repository:ERROR: (2257-64544) Exception adding distributor to repo [06acfb4-96b4-4e6
a-a418-32bed2f55bf4]; the repo will be deleted
pulp.server.controllers.repository:ERROR: (2257-64544) Traceback (most recent call last):
pulp.server.controllers.repository:ERROR: (2257-64544)   File "/usr/lib/python2.7/site-packages/pulp/server/co
ntrollers/repository.py", line 433, in create_repo
pulp.server.controllers.repository:ERROR: (2257-64544)     dist_controller.add_distributor(repo_id, type_id, p
lugin_config, auto_publish, dist_id)
pulp.server.controllers.repository:ERROR: (2257-64544)   File "/usr/lib/python2.7/site-packages/pulp/server/co
ntrollers/distributor.py", line 77, in add_distributor
pulp.server.controllers.repository:ERROR: (2257-64544)     raise exceptions.PulpDataException(message)
pulp.server.controllers.repository:ERROR: (2257-64544) PulpDataException: Configuration key [remove_old_repod
ata] is not supported
pulp.server.webservices.middleware.exception:INFO: Configuration key [remove_old_repodata] is not supported
```

and:

```
pulp.plugins.pulp_rpm.plugins.distributors.yum.configuration:ERROR: Configuration key [remove_old_repodata_threshold] is not supported
pulp.server.controllers.repository:ERROR: (2256-33152) Exception adding distributor to repo [9d542ad1-7ca6-4a90-9906-7ee23a775932]; the repo will be deleted
pulp.server.controllers.repository:ERROR: (2256-33152) Traceback (most recent call last):
pulp.server.controllers.repository:ERROR: (2256-33152)   File "/usr/lib/python2.7/site-packages/pulp/server/controllers/repository.py", line 433, in create_repo
pulp.server.controllers.repository:ERROR: (2256-33152)     dist_controller.add_distributor(repo_id, type_id, plugin_config, auto_publish, dist_id)
pulp.server.controllers.repository:ERROR: (2256-33152)   File "/usr/lib/python2.7/site-packages/pulp/server/controllers/distributor.py", line 77, in add_distributor
pulp.server.controllers.repository:ERROR: (2256-33152)     raise exceptions.PulpDataException(message)
pulp.server.controllers.repository:ERROR: (2256-33152) PulpDataException: Configuration key [remove_old_repodata_threshold] is not supported
pulp.server.webservices.middleware.exception:INFO: Configuration key [remove_old_repodata_threshold] is not supported
```

Have a look at [Pulp Smash 826](#), and let me know if something is wrong with the test case.

#20 - 12/14/2017 01:57 PM - ipanova@redhat.com

new distributor parameters were not added to the validation config

https://github.com/pulp/pulp_rpm/blob/master/plugins/pulp_rpm/plugins/distributors/yum/configuration.py#L58

#21 - 12/15/2017 03:23 PM - pcreech

- Priority changed from Normal to Urgent

With the failing of verification, this issue is being considered a blocker for 2.15.0 until further notice

#22 - 12/18/2017 06:20 PM - Ichimonji10

This issue has been fixed in 2.15 beta 3. See:

- <https://github.com/PulpQE/pulp-smash/pull/826>
- <https://github.com/PulpQE/pulp-smash/issues/681>

Here's a sample of the RPMs installed when doing testing:

```
[root@fedora-26-pulp-2-15-beta ~]# rpm -qa | grep pulp | sort
pulp-admin-client-2.15.0-0.2.beta.fc26.noarch
pulp-deb-admin-extensions-1.6.0-0.2.beta.fc26.noarch
pulp-deb-plugins-1.6.0-0.2.beta.fc26.noarch
pulp-docker-admin-extensions-3.1.0-0.3.beta.fc26.noarch
pulp-docker-plugins-3.1.0-0.3.beta.fc26.noarch
pulp-ostree-admin-extensions-1.3.0-1.fc26.noarch
pulp-ostree-plugins-1.3.0-1.fc26.noarch
pulp-puppet-admin-extensions-2.15.0-0.2.beta.fc26.noarch
pulp-puppet-plugins-2.15.0-0.2.beta.fc26.noarch
pulp-puppet-tools-2.15.0-0.2.beta.fc26.noarch
pulp-python-admin-extensions-2.0.2-1.fc26.noarch
pulp-python-plugins-2.0.2-1.fc26.noarch
pulp-rpm-admin-extensions-2.15.0-0.3.beta.fc26.noarch
pulp-rpm-plugins-2.15.0-0.3.beta.fc26.noarch
pulp-selinux-2.15.0-0.2.beta.fc26.noarch
pulp-server-2.15.0-0.2.beta.fc26.noarch
python-pulp-bindings-2.15.0-0.2.beta.fc26.noarch
python-pulp-client-lib-2.15.0-0.2.beta.fc26.noarch
python-pulp-common-2.15.0-0.2.beta.fc26.noarch
python-pulp-deb-common-1.6.0-0.2.beta.fc26.noarch
python-pulp-docker-common-3.1.0-0.3.beta.fc26.noarch
python-pulp-oid_validation-2.15.0-0.2.beta.fc26.noarch
python-pulp-ostree-common-1.3.0-1.fc26.noarch
python-pulp-puppet-common-2.15.0-0.2.beta.fc26.noarch
python-pulp-python-common-2.0.2-1.fc26.noarch
python-pulp-repoauth-2.15.0-0.2.beta.fc26.noarch
python-pulp-rpm-common-2.15.0-0.3.beta.fc26.noarch
python-pulp-streamer-2.15.0-0.2.beta.fc26.noarch
```

Notice that the RPM-related RPMs have had their version numbers bumped.

#23 - 01/17/2018 10:08 PM - pcreech

- Status changed from ASSIGNED to CLOSED - CURRENTRELEASE

#24 - 06/04/2018 12:27 AM - ipanova@redhat.com

- Related to Issue #3551: RemoveOldRepodataStep for yum publisher not checking repomd.xml to remove old files added

#25 - 04/15/2019 10:17 PM - bmbouter

- Tags Pulp 2 added

#26 - 10/16/2019 07:32 PM - mihai.ibanescu@gmail.com

- Related to Issue #5573: Publish won't create multiple checkchecksummed copies of primary.xml, fileliststs.xml etc even when in fast-forward mode added

Files

fedora-26-pulp-2-15-beta.txt	8.62 KB	12/12/2017	Ichimonji10
------------------------------	---------	------------	-------------