

RPM Support - Issue #3551

RemoveOldRepodataStep for yum publisher not checking repomd.xml to remove old files

04/05/2018 11:51 AM - mtahir

Status:	CLOSED - CURRENTRELEASE	Start date:	
Priority:	Normal	Due date:	
Assignee:		Estimated time:	0:00 hour
Category:			
Sprint/Milestone:			
Severity:	3. High	Groomed:	No
Version:	Master	Sprint Candidate:	No
Platform Release:	2.17.0	Tags:	Pulp 2
OS:		Sprint:	
Triaged:	Yes	Quarter:	
Description			
Current code in yum distributor publish doesn't actually check the files within the repodata.xml to compare the files that should be kept and files that should be deleted. Instead it makes the assumption that the latest one of each file type (to be removed) is present and removes the files older then the threshold (default 14 days). Given that mtime could be changed by someone touching a file or even moving it on the file system, it may lead to deletion of incorrect files.			
related code:			
<pre>for key, val in to_remove.iteritems(): # preserve at least one file of each kind - pop out latest if not set(grouped[key]) - set(val): val.pop(0) for f in val: self.remove_repodata_file(f[0])</pre>			
Related issues:			
Related to RPM Support - Story #2788: As a user i can configure removal of ol...		CLOSED - CURRENTRELEASE	
Related to Pulp - Issue #5573: Publish won't create multiple checkchecksummed c...		CLOSED - CURRENTRELEASE	

Associated revisions

Revision 3774e387 - 06/21/2018 06:13 AM - mtahir

Fix RemoveOldRepodataStep for yum publisher

Previously RemoveOldRepodataStep step didn't actually check the files within the repodata.xml to find the files that should be deleted. Instead it made the assumption that the latest one of each file type (to be removed) was present, and removed the files older than the threshold

fixes #3551 <https://pulp.plan.io/issues/3551>

Co-authored-by: Blake Mclvor bmcivor@redhat.com

History

#1 - 04/06/2018 04:56 PM - bmbouter

- Triaged changed from No to Yes

I think in order to do this more accurately we would need to store a record of the published data. Where would we store this type of data?

The rpm devs are really Pulp3 focused right now and Pulp2 is almost near it's maintenance phase. For Pulp2 we could work with you to help prepare a fix. Is there any possibility you would be able to contribute an improvement in this area?

#2 - 04/10/2018 08:25 PM - mtahir

I have created a fix[1] and pushed it for review. With regards to storing the record of published data, I created the fix with the assumption that any file that is not in repomd.xml, and older than the threshold should be removed.

[1] https://github.com/mztahir/pulp_rpm/commit/0b1c9edaa853b28bdead8af6e5ed36686aa6caa6

#3 - 04/13/2018 05:56 PM - ipanova@redhat.com

@mtahir, can you actually submit this as a PR to upstream pulp_rpm? thank you.

#4 - 05/17/2018 08:36 AM - bmcivor

@ipanova, PR has been created on github for this issue. If there's any more information you need, feel free to ask.

#5 - 06/04/2018 12:11 AM - ipanova@redhat.com

- Status changed from NEW to POST

https://github.com/pulp/pulp_rpm/pull/1110

#6 - 06/04/2018 12:27 AM - ipanova@redhat.com

- Related to Story #2788: As a user i can configure removal of old published repodata added

#8 - 06/27/2018 07:35 PM - mtahir

- Status changed from POST to MODIFIED

Applied in changeset [3774e3870e2ec1b485ee5cd1f49e98d7c696399a](#).

#9 - 06/28/2018 05:41 PM - ipanova@redhat.com

for QE https://github.com/pulp/pulp_rpm/pull/1110#pullrequestreview-132534392

#10 - 09/18/2018 07:50 PM - dkliban@redhat.com

- Status changed from MODIFIED to CLOSED - CURRENTRELEASE

- Platform Release set to 2.17.0

#11 - 04/15/2019 10:12 PM - bmbouter

- Tags Pulp 2 added

#12 - 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