

RPM Support - Issue #1088

Incremental publish could use the wrong primary.xml during fast-forward publish

06/24/2015 10:53 PM - bcourt

Status: CLOSED - WONTFIX	Start date:
Priority: Low	Due date:
Assignee:	Estimated time: 0:00 hour
Category:	
Sprint/Milestone:	
Severity: 1. Low	Groomed: Yes
Version:	Sprint Candidate: No
Platform Release:	Tags: Pulp 2
OS:	Sprint:
Triaged: Yes	Quarter:
Description <p>The incremental publish searches for the most recently modified file with a filename that matches a regex. It would be more robust to identify the file from the old repomd.xml instead of searching for one on the filesystem. For example, if there are two primary.xml files in the repodata directory, the incremental publish should use the one from the repodata.xml and not the first one it finds.</p> <p>This regex is performed here[0].</p> <p>[0]: https://github.com/pulp/pulp/blob/60d8aa1406703bb0f20d3632d0a8e2f5f48b66a6/server/pulp/plugins/util/metadata_writer.py#L325-L334</p>	
Related issues:	
Related to RPM Support - Issue #1087: publishing of distribution units should...	CLOSED - CURRENTRELEASE
Related to RPM Support - Issue #1086: pulp_distribution.xml sync should ignor...	CLOSED - CURRENTRELEASE

History

#1 - 06/26/2015 04:59 PM - bcourt

- Triaged changed from No to Yes

#2 - 01/08/2016 10:33 PM - bmbouter

- Related to Issue #1087: publishing of distribution units should ignore files in repodata directory added

#3 - 01/08/2016 11:02 PM - bmbouter

- Related to Issue #1086: pulp_distribution.xml sync should ignore repodata/* added

#4 - 01/11/2016 11:28 PM - bmbouter

- Tracker changed from Issue to Refactor

- Description updated

- Groomed set to No

- Sprint Candidate set to Yes

#5 - 05/04/2016 03:20 PM - mhrivnak

- Sprint Candidate changed from Yes to No

#6 - 05/18/2016 01:12 AM - dgregor@redhat.com

If I understand the description correctly, pulp could use the wrong metadata files for a fast-forward publish, which would result in incorrect metadata being published. Please change this to a bug and raise the priority.

#7 - 05/18/2016 05:07 PM - bmbouter

- *Tracker changed from Refactor to Issue*
- *Subject changed from incremental publish should use the repomd.xml to determine files to fast-forward to Incremental publish could use the wrong primary.xml during fast-forward publish*
- *Severity set to 2. Medium*
- *Triaged set to No*

Switching from a refactor to a issue because this could produce incorrect behavior.

#8 - 05/18/2016 05:08 PM - bmbouter

- *Sprint Candidate changed from No to Yes*

#9 - 05/18/2016 06:34 PM - bmbouter

- *Groomed changed from No to Yes*

After discussion in #pulp-dev produced a +1 to grooming this issue, I'm marking it as groomed.

#10 - 05/19/2016 04:51 PM - bmbouter

- *Sprint/Milestone set to 21*

#11 - 05/20/2016 05:38 PM - mhrivnak

- *Severity changed from 2. Medium to 1. Low*
- *Triaged changed from No to Yes*

#12 - 05/24/2016 07:57 PM - jortel@redhat.com

- *Status changed from NEW to ASSIGNED*
- *Assignee set to jortel@redhat.com*

#13 - 05/24/2016 08:17 PM - jortel@redhat.com

- *Description updated*

#14 - 06/13/2016 03:56 PM - mhrivnak

- *Sprint/Milestone changed from 21 to 22*

#15 - 06/13/2016 05:53 PM - jortel@redhat.com

- *File repomd.py added*
- *Sprint/Milestone changed from 22 to 21*

After digging into this, I found several things that need to be addressed.

1. The *fast forward* context in the platform is biased toward RPM and is trying to do too much. I suggest the platform classes should only provide the capability of *editing* an existing XML file that can be used at the discretion of the plugin. The plugin can determine if the file exists and have the writer open it (for edit) instead of creating a new file. I don't think much of the existing logic should be reused. It's not very robust. It makes assumptions about the complexity and structure of the XML document and does not account for namespaces. This is part of the RPM bias.
2. The distributor needs to read the existing repomd.xml before the new repomd.xml context is created. This way the information can be used to safely locate existing files. See attached example object for reading the repomd.xml.
3. The distributor changed to use the repomd information to locate existing files for fast forward.

#16 - 06/14/2016 09:07 PM - jortel@redhat.com

- Status changed from ASSIGNED to NEW

- Assignee deleted (jortel@redhat.com)

#17 - 06/30/2016 04:07 PM - mhrivnak

- Sprint/Milestone deleted (21)

#18 - 06/30/2016 04:38 PM - mhrivnak

- Sprint Candidate changed from Yes to No

#20 - 10/20/2016 04:07 PM - mhrivnak

- Sprint Candidate changed from No to Yes

#21 - 10/20/2016 04:08 PM - mhrivnak

- Sprint/Milestone set to 28

#22 - 10/25/2016 08:32 PM - Anonymous

- Severity changed from 1. Low to 3. High

#23 - 10/28/2016 03:55 PM - jortel@redhat.com

- Priority changed from Normal to High

- Severity changed from 3. High to 1. Low

#24 - 10/31/2016 04:25 PM - mhrivnak

bmbouter, jortel, tteresch and I just talked through this problem.

Next steps: bmbouter is going to try to reproduce and will add notes here.

We all agreed that the likely culprit is the general one described previously in this bug, that timestamps are being used to choose the correct updateinfo file to use during publish, and those timestamps are likely being changed (mangled) by a preceding copy operation.

Open question: Does updateinfo get fast-forwarded at all?

Open question: At some point, the wrong updateinfo file is being chosen for use. Is that either 1) during fast-forward, when the previous publish is being inspected, or 2) at the very end when a new repomd file is being created, and the distributor just wants to know which file was created during the current publish.

We might be able to resolve the issue being seen currently, which only appears to affect updateinfo files, by doing issue [#2096](#) and stop preserving old updateinfo files from one publish to the next. That would need to be vetted as an acceptable behavioral change.

If we find that the problem could be resolved by correctly inspecting a repomd file to identify the current updateinfo file, we could use the createrepo_c python library to do much of that work.

http://rpm-software-management.github.io/createrepo_c/python/lib.html#createrepo_c.Repomd

#25 - 10/31/2016 09:00 PM - bmbouter

- File `1088_reproducer.py` added

#26 - 10/31/2016 09:05 PM - bmbouter

In addition to the reproducer, I manually verified that the `updateinfo` file that is not referenced by the `repo.md` file does contain the expected errata that should be included by the incremental publish.

#27 - 10/31/2016 09:26 PM - bmbouter

Based on some of my testing, I don't think that `updateinfo.xml` is written incrementally during an incremental publish. The link in the original ticket description links to a codepath which does not provide handling for `updateinfo.xml`. It does provide incremental publish handling for several other metadata types.

I think that the `repo.md` writing logic is selecting the wrong `updateinfo.xml`. I am looking for where that codepath lives.

#28 - 11/02/2016 10:43 PM - bmbouter

- Priority changed from High to Low

- Sprint/Milestone deleted (28)

- Sprint Candidate changed from Yes to No

This bug was not the cause of the symptoms we were investigating like we originally thought. Since it was put onto the sprint on that basis I'm removing it from the sprint.

#29 - 11/02/2016 11:02 PM - bmbouter

- File deleted (`1088_reproducer.py`)

#30 - 04/12/2019 10:17 PM - bmbouter

- Status changed from NEW to CLOSED - WONTFIX

Pulp 2 is approaching maintenance mode, and this Pulp 2 ticket is not being actively worked on. As such, it is being closed as WONTFIX. Pulp 2 is still accepting contributions though, so if you want to contribute a fix for this ticket, please reopen or comment on it. If you don't have permissions to reopen this ticket, or you want to discuss an issue, please reach out via the [developer mailing list](#).

#31 - 04/15/2019 10:46 PM - bmbouter

- Tags `Pulp 2` added

Files

<code>repomd.py</code>	960 Bytes	06/13/2016	jortel@redhat.com
------------------------	-----------	------------	--