

RPM Support - Issue #2094

Most DRPMS are missing after a sync

07/20/2016 04:35 PM - jcline@redhat.com

Status:	CLOSED - WONTFIX	Start date:	
Priority:	High	Due date:	
Assignee:		Estimated time:	0:00 hour
Category:			
Sprint/Milestone:			
Severity:	2. Medium	Groomed:	No
Version:	2.8.5	Sprint Candidate:	No
Platform Release:		Tags:	Pulp 2
OS:		Sprint:	
Triaged:	Yes	Quarter:	
Description			
When syncing http://mirror.centos.org/centos/6/updates/x86_64/ , a user noticed that even though there are ~1000 DRPMS, only ~220 get synced. I was able to reproduce this on 2.8.6 (there's no option for 2.8.6 in the "Version" field).			
To reproduce:			
#. pulp-admin rpm repo create --repo-id el6-updates --download-policy on_demand --feed http://mirror.centos.org/centos/6/updates/x86_64/			
#. pulp-admin rpm repo sync run --repo-id el6-updates			
#. Note the large number of missing DRPMS			
There is nothing in the logs that indicates something went wrong.			

History

#1 - 07/21/2016 03:13 PM - mhrivnak

- Status changed from NEW to CLOSED - NOTABUG

Most of the drpms listed in the HTML directory index for that repo are not present in the XML metadata. The crude parsing method below shows about 234 drpms in the XML metadata.

```
$ curl -s http://ftp.cvut.cz/centos/6/updates/x86_64/repodata/245eccdf7402d7d6cfb7d87245a387ee06d679760b9414cb5ba2938ab074ca18-prestodelta.xml.gz | zgrep "<newpackage"| wc -l
234
```

For any DRPMS not listed in the prestodelta file, pulp (or any other client) does not have a way to find them.

So it appears that pulp is working correctly, and the creator of that repo chose to not keep older drpms in the repo's metadata.

#2 - 07/21/2016 03:20 PM - jcline@redhat.com

- Status changed from CLOSED - NOTABUG to NEW

That is *not* how prestodelta.xml works and Pulp is *not* working correctly. See

http://docs.pulpproject.org/plugins/pulp_rpm/tech-reference/rpm.html#prestodelta-xml

```
curl -s http://ftp.cvut.cz/centos/6/updates/x86_64/repodata/245eccdf7402d7d6cfb7d87245a387ee06d679760b9414cb5ba2938ab074ca18-prestodelta.xml.gz | zgrep "</delta>"| wc -l
1079
```

#3 - 07/22/2016 06:13 AM - mhrivnak

Ah, very interesting. So pulp has been parsing these incorrectly for a very long time. I'm surprised it took this long to notice. Looking at the the parsing logic, it assumes only one file per "newpackage" entry, and uses whatever the first "delta" match is returned by ElementTree.

Correcting that could be a pretty big change. The filename is part of the unit key, so at least the modeling can probably stay the same. But parsing in the importer will have to be improved to iterate over all the delta entries, and then the distributor will need to sort all the units by "newpackage" (which is not in the unit key, but does get saved on the model) so it can publish the deltas together.

#4 - 07/22/2016 04:50 PM - amacdona@redhat.com

- Priority changed from Normal to High
- Triaged changed from No to Yes
- Tags Pulp 3 Prefactor added

There is a workaround which is that the yum user download and install a whole package instead of the delta.

#5 - 07/31/2017 06:43 PM - bmbouter

- Tags deleted (Pulp 3 Prefactor)

#6 - 04/12/2019 10:18 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](#).

#7 - 04/15/2019 10:27 PM - bmbouter

- Tags Pulp 2 added