

## RPM Support - Refactor #1013

### Reduce sync time spent processing metadata up-front

05/29/2015 08:44 PM - mhrivnak

<b>Status:</b> CLOSED - WONTFIX	<b>Start date:</b>
<b>Priority:</b> High	<b>Due date:</b>
<b>Assignee:</b>	<b>% Done:</b> 0%
<b>Category:</b>	<b>Estimated time:</b> 0:00 hour
<b>Sprint/Milestone:</b>	<b>Tags:</b> Performance, Pulp 2
<b>Platform Release:</b>	<b>Sprint:</b>
<b>Groomed:</b> No	<b>Quarter:</b>
<b>Sprint Candidate:</b> No	

**Description**

The rpm sync workflow spends a lot of time early on parsing metadata (filelist.xml and other.xml) and writing it to disk in an indexed data structure. Actual downloading of files doesn't happen until afterward. This can take multiple minutes.

There are several options for improving this workflow.

- start downloading before indexing is complete. This would require a new queue of completed downloads, and an additional thread that finishes processing and saving them once the indexing is complete.
- Only index the rpms that will actually be downloaded. This would not help on the first sync (except in rare cases).
- use the sqlite file (if available) to get this metadata instead of indexing the data ourselves. The downside here is we'd have to convert it to XML.
- there may be other options

This refactor is to employ whatever improvements are reasonable to reduce the amount of time spent chewing on metadata before downloading starts.

**Related issues:**

Related to RPM Support - Story #2567: As a user, I can quickly propogate one ...	<b>CLOSED - WONTFIX</b>
Blocked by RPM Support - Refactor #874: Convert pulp_rpm to use MongoEngine M...	<b>CLOSED - CURRENTRELEASE</b>

#### History

**#1 - 07/13/2015 04:31 AM - mhrivnak**

- Blocked by Refactor #874: Convert pulp\_rpm to use MongoEngine Models added

**#2 - 10/19/2015 10:51 PM - mhrivnak**

- Platform Release set to 2.8.0

**#3 - 01/29/2016 05:32 PM - jortel@redhat.com**

- Priority changed from Normal to High

- Platform Release deleted (2.8.0)

**#4 - 05/04/2016 03:20 PM - mhrivnak**

- Sprint Candidate changed from Yes to No

**#6 - 02/27/2017 02:28 PM - dkliban@redhat.com**

- Related to Story #2567: As a user, I can quickly propogate one package change from one Pulp to another Pulp added

**#7 - 02/27/2017 02:49 PM - dkliban@redhat.com**

- Tags Sync Performance added

**#8 - 03/12/2019 07:20 PM - dkliban@redhat.com**

- Status changed from NEW to CLOSED - WONTFIX

This is going to be addressed by the stream processing design of Pulp 3.

**#9 - 04/15/2019 10:48 PM - bmbouter**

- *Tags Pulp 2 added*

**#10 - 06/16/2020 10:59 PM - bmbouter**

- *Tags Performance added*

- *Tags deleted (Sync Performance)*