

RPM Support - Issue #1476

The RPM models are indexing a lot of fields that don't make sense to have indices

01/05/2016 09:54 PM - rbarlow

Status: CLOSED - WONTFIX	Start date:
Priority: High	Due date:
Assignee:	Estimated time: 0:00 hour
Category:	
Sprint/Milestone:	
Severity: 1. Low	Groomed: No
Version: 2.4.0	Sprint Candidate: No
Platform Release:	Tags: Pulp 2
OS:	Sprint:
Triaged: Yes	Quarter:
Description	
The RPM models are setting up these indices :	
<code>"name", "epoch", "version", "release", "arch", "filename", "checksum", "checksumtype", "version_sort_index"</code>	
Almost all of these do not make sense, and they each will significantly increase the RAM required of the MongoDB server and reduce write performance. I'll go one by one and explain each:	
name: The name is also indexed by the first field of the unit key uniqueness constraint that is set up on line 369 below, so this index is redundant.	
epoch: Why would users search by epoch?	
release: Why would users search by release?	
arch: Why would users search by arch?	
filename: I could see a case for this one, especially with a contains, but I'm still a little skeptical. I'm a maybe on it.	
checksum: This can be a useful way to look up units, so I think it's OK.	
checksumtype: Why would users search with this?	
version_sort_index: This is also the first field of the compound index of version_sort_index and release_sort_index, so it is redundant like the name index.	
I would mark this easy fix or even just make a pull request, but if we've made a release with this we'll need a migration to drop the indices. I haven't done the research to determine if we've released with these indices. If we haven't released, I'd consider this a 2.8 blocker.	
Related issues:	
Related to Pulp - Issue #1477: With mongoengine models, developers have to ma...	CLOSED - CURRENTRELEASE
Has duplicate RPM Support - Issue #630: RPM unit types have redundant search ...	CLOSED - DUPLICATE

History

#1 - 01/05/2016 10:22 PM - bmbouter

+1 to getting this right.

The port to mongoengine was designed to be as similar as possible to what was there before. I took a quick look at an rpm based 2.7.0 installation I have, and it shows all of those fields are indexed. I'm open to changing these in a major release since the memory improvement is worth the losses of indexes that aren't beneficial.

Given that, should this be a story or a bug? What about making it a 2.8.0 release blocker?

The release that fixes this should have release notes for it. I've written some like these for the 2.8.0 upgrade already[0]. As an example, here is the corresponding migration[1] with those changes.

[0]:
https://github.com/pulp/pulp_rpm/blob/30b8ea06ea81a410c7ab83941c706e7b217b9993/docs/user-guide/release-notes/2.8.x.rst#database-index-changes

[1]:
https://github.com/pulp/pulp_rpm/blob/bbadd8f7c0b65914c2c01327f4b5f9370228fcf/plugins/pulp_rpm/plugins/migrations/0022_rename_unit_id_field_s.py

#2 - 01/05/2016 10:44 PM - bmbouter

- Related to Issue #1477: With mongoengine models, developers have to manually configure uniqueness constraints for model unit keys added

#3 - 01/06/2016 08:21 PM - rbarlow

bmbouter wrote:

Given that, should this be a story or a bug? What about making it a 2.8.0 release blocker?

I think it should be a bug or a refactor, but it's definitely not a feature. That's really the crux of it, we are wasting RAM here and the RAM isn't supporting a user story. Since you have confirmed that it is not a new problem, I don't think it should be a 2.8 blocker. Since 2.8.0 is already crushing us, let's save it for later.

#4 - 01/08/2016 05:01 PM - mhrivnak

- Priority changed from Normal to High

- Severity changed from 2. Medium to 1. Low

- Triaged changed from No to Yes

#5 - 02/23/2016 10:21 PM - rbarlow

- Has duplicate Issue #630: RPM unit types have redundant search indices specified added

#6 - 02/23/2016 10:23 PM - rbarlow

- Version set to 2.4.0

#7 - 02/24/2016 07:28 PM - rbarlow

- Triaged changed from Yes to No

We should rediscuss this in light of https://github.com/pulp/pulp_rpm/pull/812.

#8 - 02/26/2016 04:45 PM - mhrivnak

- Triaged changed from No to Yes

#9 - 08/10/2016 11:25 PM - mhrivnak

- Related to Refactor #2159: Convert unit models to django added

#10 - 10/18/2017 04:05 PM - amadonna@redhat.com

- Related to deleted (Refactor #2159: Convert unit models to django)

#11 - 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](#).

#12 - 04/15/2019 10:39 PM - bmbouter

- Tags Pulp 2 added