RPM Support - Story #5202

Story # 5110 (CLOSED - DUPLICATE): [Epic] As a user, I can manage Kickstart repositories

As a user, I can sync distribution trees

07/31/2019 09:24 PM - daviddavis

Status:	CLOSED - CURRENTRELEASE		**Start date:**	
Priority:	Normal		**Due date:**	
Assignee:	fao89		**% Done:**	100%
Category:			**Estimated time:**	0:00 hour
Sprint/Milestone:	Pulp 3.x RPM (Katello 3.16)			
Platform Release:			**Tags:**	
Groomed:	Yes		**Sprint:**	Sprint 58
Sprint Candidate:	No		**Quarter:**	

Description

For now, mirror all images and a main yum repo from a kickstart repo.

Design

Have the FirstStage of pulp_rpm try to download and parse the tree info file. From this file create the declarative content unit of the DistributionTree model (the actual content unit). Have the Stages API process this unit like any other.

When actually saving the DistributionTree content unit in the ContentUnitSaver stage, have any additional model creation happen then.

Syncing Sub Repos

Some entries in the kickstart tree identify a "sub repo". that's not an official term, but it's intended to mean another valid repodata folder that identifies another set of RPMs that is outside of the main /repodata/ folder. One main challenge is efficiently syncing these "sub repos".

One easy solution we came up with is to instantiate multiple RPMDeclarativeVersion instances, one for each sub-repo and one for the main repo. This allows each one to work out of a different repository. The trick is to not instantiate more than 1 at once that way rate limiting, etc won't be an issue.

What about remotes for sub Repos?

We should create an in-memory Remote for the sub-repos. It shouldn't modify the real Remote because we don't want to do that. Also it needs all the settings from the original remote, e.g. proxy settings, ssl, etc so all those should be copied. The one difference needs to be the url, it needs to be appended with the sub-path url of the sub-repo from the treeinfo file.

Associated revisions

Revision 0057cab2 - 09/04/2019 06:44 PM - Fabricio Aguiar

kickstart syncing

closes #5202 https://pulp.plan.io/issues/5202 Required PR: https://github.com/pulp/pulp_rpm/pull/1418

History

#1 - 08/16/2019 05:13 PM - bmbouter

- Groomed changed from No to Yes

- Sprint set to Sprint 57

This makes sense to me if the models were to exist.

#2 - 08/21/2019 07:03 PM - bmbouter

- Description updated

#3 - 08/23/2019 12:01 AM - bmbouter

- Description updated

#4 - 08/23/2019 12:04 AM - bmbouter

If we don't want to extend the Stages API immediately with a "multi-sync" capability, we could instantiate N copies of the RpmDeclarativeVersion pipeline and run N of them on the asyncio scheduler concurrently.

The downside of this approach is that it interferes with the various limiting features the Stages API itself which is ok in the short term but not the long. Specifically download concurrency would be N times what the user expects, also the memory restriction features would be N times larger.

#5 - 08/23/2019 09:51 PM - bmbouter

- Description updated

#6 - 08/26/2019 04:29 PM - rchan

- Sprint changed from Sprint 57 to Sprint 58

#7 - 08/26/2019 06:58 PM - fao89

- Status changed from NEW to POST

#8 - 08/26/2019 09:43 PM - fao89

- Assignee set to fao89

#9 - 08/27/2019 09:16 PM - fao89

- Assignee deleted (fao89)

https://github.com/pulp/pulp_rpm/pull/1427

#10 - 08/28/2019 12:49 PM - ipanova@redhat.com

- Assignee set to fao89

#11 - 09/04/2019 11:06 PM - Anonymous

- Status changed from POST to MODIFIED

- % Done changed from 0 to 100

Applied in changeset 0057cab262fa5afa0ee9164f74b1ce60c8fdcc87.

#12 - 12/13/2019 08:29 AM - ttereshc

- Status changed from MODIFIED to CLOSED - CURRENTRELEASE

#13 - 02/26/2020 05:37 PM - ttereshc

- Sprint/Milestone set to Pulp 3.x RPM (Katello 3.16)