

Docker Support - Issue #2130

Old manifests with old tags are not purged during syncing

08/03/2016 10:59 PM - jsherril@redhat.com

Status: CLOSED - NOTABUG	Start date:
Priority: Normal	Due date:
Assignee:	Estimated time: 0:00 hour
Category:	
Sprint/Milestone:	
Severity: 2. Medium	Groomed: No
Version - Docker:	Sprint Candidate: No
Platform Release:	Tags: Pulp 2
Target Release - Docker:	Sprint:
OS:	Quarter:
Triaged: No	

Description

When re-syncing a docker repo, if any tag has been updated with a new manifest (say a new build of 'latest'), pulp correctly updates the 'tag' in units_docker_tags, but fails to delete the manifest from units_docker_manifests.

Steps to reproduce:

1. Sync a docker repo with some tag 'latest'
2. Using the docker command, Create a new image, tag it to latest, and push the image to the docker repo
3. Resync the docker repo

If you list out manifests, you'll notice that there are still two manifests in the 'units_docker_tags' collection with the 'latest' tag. I believe I would expect the old manifest to be deleted.

At a minimum, its 'tag' attribute should be null or empty

History

#1 - 08/03/2016 11:02 PM - jsherril@redhat.com

- Description updated

#2 - 08/04/2016 07:47 PM - mhrivnak

Since clients can access manifests by their digest value, many docker users want to keep old manifests and blobs in their repository even when there is no longer an active tag referencing them. Deploying a manifest by its digest ID is a best practice as it allows the deployer to know with confidence exactly what is deployed.

Note that the "tag" attribute of a Manifest is a piece of data baked into the manifest itself. It is completely valid for there to be many manifests in a docker repository that all have the same "tag" value in their json data. Of course only one of them will be returned by the registry API's "tags" endpoint, and presumably that would be the newest one.

<https://docs.docker.com/registry/spec/api/#/listing-image-tags>

In order to allow pulp users to control which manifest is the "active" one returned by the registry API for a given tag name, pulp has the Tag unit type. During a sync, if the remote API has a new manifest referenced from a given tag name, pulp will update the corresponding Tag unit. This allows a point-in-time snapshot of a repository to be promoted by copying a Tag, Manifest, and associated Blob units to a new repo. In the future, if another copy is made and the source repo's Tag unit references a different manifest, that will overwrite the one in the destination repo.

Given all of that, I do not think this issue describes a bug. There are at least two opportunities for improvement though:

- Better document how Tags are intended to be used. I was surprised to see that we don't have much currently.
- An enhancement would be to add an importer option to purge non-tagged content during sync, possibly keeping some number of the newest non-tagged manifests. But this should be an optional behavior that is not the default; when uncertain about whether a user wants pulp to keep a piece of content, pulp should default to keeping it.

#3 - 08/09/2016 04:21 PM - jsherril@redhat.com

- Status changed from NEW to CLOSED - NOTABUG

Based on discussion seems this is not a bug

#4 - 04/15/2019 10:26 PM - bmbouter

- Tags Pulp 2 added