Docker Support - Test #5181

Removing docker manifests from a docker repository takes a long time

07/24/2019 02:58 PM - bherring

Status:	CLOSED - COMPLETE		
Priority:	Normal		
Assignee:	bherring		
Category:			
Sprint/Milestone:			
Platform Release:		**Sprint:**	
Tags:	Pulp 2	**Quarter:**	

Description

Original Dev Problem

Removing all docker manifests from a large docker repo seems to take a long time:

~300 manifests takes ~2 minutes
~2000 manifests takes ~30-40 minutes

Reproducer:

1. Create and sync a docker repo such as: https://quay.io datawire/ambassador
2. Remove all docker manifests from the repo: pulp-admin docker repo remove manifest
--repo-id=1-docker-dev-7915f7d0-7a98-4131-9c41-1be7b578d442 --not id=foo

Discussion with asmacado

- A lot of the delay were caused by the sequential removals
- If there was an outage while the task was on-going, the removal was left in a non-deterministic state
- The timing model for removal for general use cases grew incredibly fast

Solutions with the new refactor:

- All BLOBS and MANIFESTS are all removed at once
- Higher level units are all removed second
- Timing shrank from 300 ~=1hr to 300 ~=a few seconds

Concerns:

- This refactor would affect removals and syncs
- The removal, if interrupted, would leave Orphans which could affect upstream verification

Actions:

Upstream

- Need an upstream verification that ensures there is no change in functionality with remove units and sync units

Downstream

- Need to verify downstream verification are not affected by this change/patch
- BZ verification will just be the removal of a large number of repos takes a short amount of time

Test Permutations

Remove

Variables:

- (Permutations) Shared or Not Shared [manifest_lists, manifests]

- type_ids": ['docker_tag', 'docker_manifest_list', 'docker_manifest', 'docker_blob']

Verification Scenarios:

- For Not Shared, remove $REMOVAL_TYPES and verify all units have been removed
- For Shared, remove $REMOVAL_TYPES and verify only non-shared units have been removed. Shared units should remain.

Note:

- Removal Criteria short-cut for relevant type_id being removed.

```
criteria = {
    "type_ids": ['docker_tag', 'docker_manifest_list', 'docker_manifest', 'docker_blob'],
    "filters": {"unit": {"_id": {"$in": all_units}}},
}
```

- There are already tests covering tag removal

Related issues:	
Copied from Docker Support - Issue #4549: Removing docker manifests from a do...	**CLOSED - CURRENTRELEASE**

Associated revisions

Revision 086b660a - 08/01/2019 06:51 PM - bherring

Adding tests to docker plugin for removing and syncing of docker units

Issue #4549 refactors the handling of removing tags, maniftest_lists, manifests, and blobs within the docker plugin.

Adding test permutations to ensure the behavior of removing all, shared, and non-shared units remains the same.

This test will not address the large performance gain by the refactor.

Refs #5181

Revision ecba1db4 - 08/08/2019 07:25 PM - bherring

Refactor of test_remove.py with changes from 4549

```
With the refactor of the docker importer's remove function to
increase performance, content removal needs to be functional verified.

The cases covered with content post-count verification for all units:

1. Remove all manifest_lists sequentially.
2. Remove all manifests sequentially.
3. Remove all blobs sequentially.
4. Remove all manifest_lists batch.
5. Remove all manifests batch.
6. Remove all blobs batch.
7. Remove some non-shared manifest lists.
8. Remove some non-shared manifest.
9. Remove some shared manifests lists and verify shared units are not
   recursively removed.
10. Remove some shared manifests and verify shared units are not
    recursively removed.

The fixture includes:

* 2 relatively independent manifest lists (no shared manifests,
  no shared blobs between them)
* 2 manifest lists that share some (but not all) manifests, and those
  manifest share some (but not all) blobs. This only requires the creation
  of 1 manifest list that shares some content with one of the first
  "independent manifest lists".
* 2 relatively independent manifests
* 2 manifests that share (some but not all) blobs

In order to sync the content, each content unit must be recursively related
to at least 1 tag.
```

refs #4549 #5161

closes #5181

History

#1 - 07/24/2019 02:58 PM - bherring

- Copied from Issue #4549: Removing docker manifests from a docker repository takes a long time added

#2 - 07/24/2019 03:01 PM - bherring

- Description updated

#3 - 07/25/2019 02:48 PM - amacdona@redhat.com

- File add_remove_utils.py added

- File test_remove.py added

#4 - 07/30/2019 02:12 PM - bherring

- Description updated

- Status changed from NEW to ASSIGNED

#5 - 08/09/2019 07:49 PM - bherring

- Status changed from ASSIGNED to CLOSED - COMPLETE

Update

All testing complete.

All information aggregated in PR via https://projects.engineering.redhat.com/browse/SATQE-5197

Files

almost-all-man-lists-busybox	210 KB	07/09/2019	amacdona@redhat.com
add_remove_utils.py	5.76 KB	07/25/2019	amacdona@redhat.com
test_remove.py	15 KB	07/25/2019	amacdona@redhat.com