

## RPM Support - Test #4566

### Improve performance of rpm duplicate nevra check

03/25/2019 01:53 PM - bherring

<b>Status:</b> CLOSED - COMPLETE	
<b>Priority:</b> High	
<b>Assignee:</b> bherring	
<b>Category:</b>	
<b>Sprint/Milestone:</b>	
<b>Version:</b>	<b>Sprint:</b>
<b>Platform Release:</b>	<b>Quarter:</b>
<b>Tags:</b> Pulp 2	
<b>Description</b>	
In current versions of Pulp 2.x, uploading an RPM to a repo will remove other RPMs with the same NEVRA.	
Currently, we are upgrading from an old version of Pulp 2.7, and I've found that performance of import_uploaded_unit tasks for RPMs has regressed significantly. In Pulp 2.7, imports would usually take around 0.5s. In Pulp 2-master, imports to the same repos have taken from 8 to 130 seconds, depending on the size of the repo.	
By debugging I've found most of the time is spent in this duplicate check (remove_unit_duplicate_nevra).	
This issue is for improving the performance of remove_unit_duplicate_nevra to reduce the severity of the performance regression.	
<b>Related issues:</b>	
Copied from RPM Support - Story #4527: Improve performance of rpm duplicate n...	<b>CLOSED - CURRENTRELEASE</b>

#### History

##### #1 - 03/25/2019 01:53 PM - bherring

- Copied from Story #4527: Improve performance of rpm duplicate nevra check added

##### #2 - 03/25/2019 02:22 PM - bherring

- Priority changed from Normal to High

##### #3 - 03/25/2019 03:01 PM - bherring

- Assignee set to bherring

#### Notes

Verification will include:

- No Regression in automation testing for 2.19RC
- Manually testing comparison just to note and agree with previous notations, assumptions and findings pre-build (verify there is an improvement).

##### #4 - 03/25/2019 07:56 PM - bherring

- Status changed from NEW to ASSIGNED

##### #5 - 03/27/2019 06:21 PM - bherring

#### Notes

Manual verification of uploaded units proved the single order of magnitude increase on upload to a pre-populated repo without duplicating RPMs via the python script and more functionally through the CLI

#### CLI-based script

```
## Test Scenario:
## Resources:
#   Box 1: 2.17.1
#   Box 2: 2-master
```

```
## Tests:
# 1. Dev provided Python based test
# 2. CLI based test below

# Login
pulp-admin login -u admin -p admin

# Create Repo A and B
pulp-admin rpm repo create --repo-id a --feed https://fedora.mirror.constant.com/fedora/linux/releases/29/Everything/x86_64/os/ --download-policy=on_demand

pulp-admin rpm repo create --repo-id b --feed https://repos.fedorapeople.org/pulp/pulp/fixtures/rpm-with-modules/

# Sync Up
pulp-admin rpm repo sync run --repo-id b
pulp-admin rpm repo sync run --repo-id a

# Test Upload of content to A (B is no longer used)
yum install wget -y
wget --recursive --no-parent https://repos.fedorapeople.org/pulp/pulp/fixtures/rpm-with-modules/

# Timing Copy of unit from B to A
## declare an array
declare -a elementArray=("bear" "camel" "cat" "cheetah" "chimpanzee" "cockateel" "cow" "crow" "dog" "duck" "elephant" "fox" "frog")

## loop through all elements
for element in "${elementArray[@]}"
do
    echo -e "$element\n"
    time pulp-admin rpm repo uploads rpm --repo-id a --file $element* --recursive
done
```

## Python Test

The python testing locally on each box using `python test.py` returned the same result as the gitgist provided in the original PR for verification.

### #6 - 03/27/2019 06:21 PM - bherring

- Status changed from ASSIGNED to MODIFIED

### #7 - 04/15/2019 09:59 PM - bmbouter

- Tags Pulp 2 added

### #8 - 05/21/2019 09:34 PM - bherring

- Status changed from MODIFIED to CLOSED - COMPLETE