Pulp - Story #23

As a user, I can rest easy in the knowledge that Pulp's Celery tasks are not serialized dangerously

12/18/2014 05:12 PM - Anonymous

Status:	CLOSED - CURRENTRELEASE	**Start date:**	
Priority:	Normal	**Due date:**	
Assignee:	ipanova@redhat.com	**% Done:**	100%
Category:		**Estimated time:**	0:00 hour
Sprint/Milestone:			
Platform Release:	2.8.0	**Tags:**	Pulp 2
Groomed:	Yes	**Sprint:**	
Sprint Candidate:	Yes	**Quarter:**	

Description

Our tasks currently need to use pickle for serialization, because we pass objects as arguments to some of them. We likely also return objects from some tasks.We should either make our objects json serializable, or change our tasks to accept simple types that json can handle.Once this is complete, we should configure Celery to use json, and configure the workers not to accept pickle.

Related issues:

Has duplicate Pulp - Issue #575: Warning in the log when starting celery serv...	**CLOSED - DUPLICATE**

Associated revisions

Revision 2ed8b7f5 - 12/15/2015 10:43 AM - Shubham Bhawsinka

Progress for story 23

- Recursive serialization and deserialization
- Force all Pulp Celery tasks to inherit from PulpTask

Revision 4ac8b815 - 12/15/2015 10:43 AM - bmbouter

Removes explicit type validation, and cleans up docstrings some.

Revision ed367081 - 12/15/2015 11:03 AM - ipanova@redhat.com

Introducing from_dict() and to_dict()

closes #23 https://pulp.plan.io/issues/23

UnitAssociationCriteria arrives as dict in the task code, when it needs to be UnitAssociationCriteria object. to_dict() method was introduced to serialize UnitAssociationCriteria to a dictionary prior to it being sent as an argument to apply_sync(). the from_dict() method was introduced to return UnitAssociationCriteria object from a dictionary that was produced by to_dict()

This commit also introduces a testcase that checks that all Celery tasks defined in Pulp inhherit from PulpTask and not from Celery directly.

There have been done also changes to criteria. Due to the introduction of de/sereliazions methods, criteria's value cannot be None. Code and tests were updated accordigly.

Revision ed367081 - 12/15/2015 11:03 AM - ipanova@redhat.com

Introducing from_dict() and to_dict()

closes #23 https://pulp.plan.io/issues/23

UnitAssociationCriteria arrives as dict in the task code, when it needs to be UnitAssociationCriteria object. to_dict() method was introduced to serialize UnitAssociationCriteria to a dictionary prior to it being sent as an argument to apply_sync(). the from_dict() method was introduced to return UnitAssociationCriteria object from a dictionary that was produced by to_dict()

This commit also introduces a testcase that checks that all Celery tasks defined in Pulp inhherit from PulpTask and not from Celery directly.

There have been done also changes to criteria. Due to the introduction of de/sereliazions methods, criteria's value cannot be None. Code and tests were updated accordigly.

Revision 478443a7 - 01/22/2016 06:50 PM - bmbouter

Disables all other celery serializers except JSON

This removes a scary, multiline warning from celery workers when they start up

https://pulp.plan.io/issues/23 re #23

Revision 478443a7 - 01/22/2016 06:50 PM - bmbouter

Disables all other celery serializers except JSON

This removes a scary, multiline warning from celery workers when they start up

https://pulp.plan.io/issues/23 re #23

History

#1 - 12/18/2014 10:17 PM - rbarlow

- Project changed from 22 to Pulp

#2 - 09/11/2015 06:15 PM - bmbouter

- Groomed set to No

- Sprint Candidate set to No

#3 - 09/14/2015 06:14 PM - bmbouter

- Groomed changed from No to Yes

#4 - 09/14/2015 07:28 PM - bmbouter

- Sprint Candidate changed from No to Yes

#5 - 12/02/2015 06:57 PM - bmbouter

The branch with the current work is here: https://github.com/bmbouter/pulp/tree/23-farmers-ale

It's currently branched from 2.6, and has 2 additional commits (50d785d and 8bf9633).

Here is what needs to be done:
1) First, rebase or cherry pick against master. The two existing commits will have their hashes changed in the process, but please leave them and their commit messages in place (for attribution). Once rebased into a branch that branches from master, make your commits on top of the existing 2.

2) Reproduce the known issue with this code by having a working pulp installation and running:

```
pulp-admin rpm repo create --repo-id bar
pulp-admin rpm repo create --repo-id foo
pulp-admin rpm repo copy rpm --from-repo-id=bar --to-repo-id=foo
```

3) The issue is that the UnitAssociationCriteria is arriving as a dict in the task code, when it needs to be a UnitAssociationCriteria (which inherits from dict). To resolve this, the UnitAssociationCriteria object needs to be explicitly serialized to a dictionary prior to it being sent as an argument of apply_async(). Add a to_dict() method on UnitAssociationCriteria() to return the dict. Find the place where the UnitAssociationCriteria is passed as an argument by looking at what API endpoint is being used by the repo copy rpm command.

4) Add a classmethod to UnitAssociationCriteria which can instantiate and return a UnitAssociationCriteria object from a dictionary that was produced by to_dict().

5) Use the classmethod above in the task code to produce the UnitAssociationCriteria from the dict passed in as an argument

6) Do some other general testing if you like (not required)

7) Write test code for everything

8) Write one specific test which looks through the celery registry and asserts that each entry in it is a subclass of PulpTask.

#6 - 12/04/2015 04:40 PM - bmbouter

Here is what I put together as a start for item #8 above.

```
from pulp.server.async import app
from pulp.server import tasks # you have to import this to discover all pulp tasks
from pulp.server.async.tasks import PulpTask

for task_name, task in app.celery.tasks.iteritems():
    if task.startswith('pulp'):
        if not isinstance(task, PulpTask):
            self.fail('task named %s must inherit from %s) % (task_name, PulpTask)
```

#7 - 12/09/2015 04:16 PM - ipanova@redhat.com

- Status changed from NEW to ASSIGNED

- Assignee set to ipanova@redhat.com

#8 - 12/11/2015 05:22 PM - ipanova@redhat.com

- Status changed from ASSIGNED to MODIFIED

https://github.com/pulp/pulp/pull/2241

#9 - 12/11/2015 06:04 PM - bmbouter

It's not merged yet, so moving back to POST.

#10 - 12/15/2015 03:36 PM - ipanova@redhat.com

- % Done changed from 0 to 100

Applied in changeset pulp|ed36708132e18bac4ab8b58d0b56cdb2ffe95685.

#11 - 12/17/2015 04:55 PM - amacdona@redhat.com

- Platform Release set to 2.8.0

#12 - 01/22/2016 06:43 PM - bmbouter

This merged change does switch our usage to use the JSON serializer, but we haven't actually disabled the other serializers. As such the following banner is still visible:

```
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000) /usr/lib/python2.7/site-packages/celery/ap
ps/worker.py:161: CDeprecationWarning:
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000) Starting from version 3.2 Celery will refu
se to accept pickle by default.
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000)
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000) The pickle serializer is a security concer
n as it may give attackers
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000) the ability to execute any command.  It's
important to secure
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000) your broker from unauthorized access when
using pickle, so we think
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000) that enabling pickle should require a deli
berate action and not be
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000) the default choice.
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000)
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000) If you depend on pickle then you should se
t a setting to disable this
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000) warning and to be sure that everything wil
l continue working
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000) when you upgrade to Celery 3.2::
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000)
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000)     CELERY_ACCEPT_CONTENT = ['pickle', 'js
on', 'msgpack', 'yaml']
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000)
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000) You must only enable the serializers that
you will actually use.
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000)
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000)
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000)   warnings.warn(CDeprecationWarning(W_PICK
LE_DEPRECATED))
Jan 22 01:35:01 dev pulp[22849]: py.warnings:WARNING: (22849-36000)
```

I am fixing this now with a small PR to disable the other serializers.

#13 - 01/22/2016 06:53 PM - bmbouter

PR available at: https://github.com/pulp/pulp/pull/2346

#14 - 02/18/2016 03:28 AM - bmbouter

- Has duplicate Issue #575: Warning in the log when starting celery services added

#15 - 02/23/2016 09:55 PM - dkliban@redhat.com

- Status changed from MODIFIED to 5

#16 - 03/23/2016 07:14 PM - dkliban@redhat.com

- Status changed from 5 to CLOSED - CURRENTRELEASE

#18 - 04/15/2019 11:22 PM - bmbouter

- Tags Pulp 2 added