

Pulp - Refactor #888

Refactor # 765 (CLOSED - WONTFIX): Convert Pulp to use MongoEngine

pulp_manage_db needs to run .ensure_indexes() on MongoEngine platform models

04/14/2015 08:57 PM - bcourt

Status:	CLOSED - CURRENTRELEASE	Start date:	
Priority:	Normal	Due date:	
Assignee:	bcourt	% Done:	100%
Category:		Estimated time:	0:00 hour
Sprint/Milestone:		Tags:	Pulp 2
Platform Release:	2.7.0	Sprint:	April 2015
Groomed:	Yes	Quarter:	
Sprint Candidate:	No		
Description			
After pulp-manage-db performs all the migrations it should load all the models for the core collections (repos, tasks, repo_content_units, etc.) and run the ensure_indexes() method on the models to make sure that the Pulp required indexes are maintained.			
Deliverables:			
<ul style="list-style-type: none">• Create a list that stores the class paths to provide the loading of all models for core collections• Platform models that have already been converted are included in the list of models above• Updates to pulp-manage-db calls ensure_indexes() on each loaded class in the list above• Mongoengine conversion guide is updated to include adding the model to this collection of calls.			

History

#1 - 04/14/2015 09:02 PM - bcourt

- Assignee set to bcourt
- Tags Groomed added

#2 - 04/14/2015 09:07 PM - bmbouter

- Description updated

#3 - 04/14/2015 09:08 PM - bmbouter

- Description updated

#4 - 04/14/2015 09:10 PM - bcourt

- Status changed from NEW to ASSIGNED
- Platform Release set to 2.7.0

#5 - 04/15/2015 04:14 PM - rbarlow

On 04/14/2015 02:57 PM, Pulp wrote:

After pulp-manage-db performs all the migrations it should load all the models for the core collections (repos, tasks, repo_content_units, etc.) and run the ensure_indexes() method on the models to make sure that the Pulp required indexes are maintained.

Is this true? When I used Mongoengine before, I believe it ensured all the indices for me upon connecting to the DB. However, it has been ~3.5 years since I last used Mongoengine.

--
Randy Barlow

#6 - 04/15/2015 08:40 PM - bcourt

- Status changed from ASSIGNED to CLOSED - CURRENTRELEASE

#7 - 04/15/2015 08:41 PM - bcourt

- Sprint/Milestone set to 15

- % Done changed from 0 to 100

#8 - 04/16/2015 03:39 PM - bmbouter

I just tested this, and I think that mongoengine calls ensureindexes when accessing the db automatically. I am using mongoengine==0.7.10

Adapted from the mongoengine gettings started guide, I create a simple model in foo.py:

```
from mongoengine import *
connect('tumblelog')
class User(Document):
    email = StringField(required=True)
    first_name = StringField(max_length=50)
    last_name = StringField(max_length=50)

User(email='ross@example.com', first_name='Ross', last_name='Lawley').save()
```

I run foo.py. I then list the indexes using the following command and I see 2. One for the primary key and one for some other index that I don't understand.

```
[bmbouter@server ~]$ mongo tumblelog --eval 'db.user.getIndexes()'
MongoDB shell version: 2.4.6
connecting to: tumblelog
[object Object],[object Object]
```

Then I add some indexes to foo.py and run it:

```
from mongoengine import *
connect('tumblelog')
class User(Document):
    meta = {'indexes': ['first_name', 'last_name']}

    email = StringField(required=True)
    first_name = StringField(max_length=50)
    last_name = StringField(max_length=50)

User(email='another@example.com', first_name='Some', last_name='User').save()
```

Then when I list the indexes I see the two additional ones on the collection.

```
[bmbouter@server ~]$ mongo tumblelog --eval 'db.user.getIndexes()'
MongoDB shell version: 2.4.6
connecting to: tumblelog
[object Object],[object Object],[object Object],[object Object]
```

I suspect that the connect() gratuitously calls createIndex() or ensureIndex(). ensureIndex() was deprecated by MongoDB in favor of createIndex later. Having it check with each connection is a mongoengine behavior which can be disabled with the auto_create_index set to False in the meta of the model definition. I think we should leave it to True, and remove the pulp-db and revert this story.

This would have been much more useful two days ago, but I assumed mongoengine didn't do this correctly then.

#9 - 04/16/2015 04:45 PM - bmbouter

Disabling auto_create_index could have performance gains. Here is a quick analysis of the overhead of auto_create_index. <http://fpaste.org/211905/>

#10 - 04/16/2015 05:34 PM - bmbouter

After some IRC discussion, we are going to leave in place the PR [0] that runs an ensure_index() on all mongoengine models at the end of pulp-manage-db. This will push any index creations on big collections to pulp-manage-db time instead of runtime. We are also choosing to leave auto_index_create at its default (enabled) so that if Pulp developers fail to add a model to the pulp-manage-db codepath, mongoengine will still autcreate the missing indexes at runtime.

[0]: <https://github.com/pulp/pulp/pull/1784/files>

#11 - 05/04/2015 09:45 PM - bmbouter

- Groomed set to Yes

- Sprint Candidate set to No

- Tags deleted (Groomed)

#12 - 03/08/2018 07:14 PM - bmbouter

- *Sprint set to April 2015*

#13 - 03/08/2018 07:15 PM - bmbouter

- *Sprint/Milestone deleted (15)*

#14 - 04/15/2019 10:50 PM - bmbouter

- *Tags Pulp 2 added*