

Ansible Plugin - Issue #7564

Excess number of suboptimal queries when querying collections list

09/24/2020 02:55 PM - osapryki

Status:	NEW	Start date:	
Priority:	Normal	Due date:	
Assignee:		Estimated time:	0:00 hour
Category:			
Sprint/Milestone:			
Severity:	2. Medium	Sprint Candidate:	No
Platform Release:		Tags:	
OS:		Sprint:	
Triaged:	No	Quarter:	
Groomed:	No		

Description

Request to `/pulp/api/v3/content/ansible/collection_versions` (see Request) produces **6 + 2N** SQL queries, where N is a number of collection version in response.

This is highly inefficient and may result in a high request latency in production environments. For Automation Hub it results in **1.8-4.2** seconds per request.

Queries list includes 2 queries that runs SELECT DISTINCT over all fields and includes explicit list of UUIDs of all certified collections in the IN clause.

Request:

GET

`/pulp/api/v3/content/ansible/collection_versions/?is_highest=True&offset=0&limit=12&exclude_fields=docs_blob&certification=certified&deprecated=false`

Queries:

1. Queries list:

<https://gist.github.com/cutwater/70f6adbe68d88cec0df14141d508f5cf>

1. Explain analyze of query executed against the production database:

To be added later in comments.

Versions tested:

1. pulpcore==3.5.0 + pulp-ansible==0.2.0b15
2. pulpcore==3.7.0 + pulp-ansible==0.4.0

Actions required:

1. Eliminate 2N queries.
2. Optimize SELECT DISTINCT or remove.
3. Remove explicit list of all certified collections in the IN clause. Replace with a subquery if needed.

History

#1 - 09/25/2020 05:12 PM - osapryki

- Priority changed from High to Normal

#2 - 09/28/2020 07:56 PM - fao89

the distinct comes from MultipleChoiceFilter: <https://django-filter.readthedocs.io/en/stable/ref/filters.html?highlight=distinct#multiplechoicefilter>

#3 - 10/08/2020 02:57 PM - osapryki

- File Screenshot_20201008_143356.png added

This is a partial dump of SQL query generated on CI environment which has 5290 collection version records in the database.

#4 - 10/08/2020 02:58 PM - osapryki

The previous dump is generated on API call to /api/automation-hub/v3/collections/ which calls pulp_ansible.app.galaxy.v3.CollectionViewSet

#5 - 10/08/2020 06:07 PM - osapryki

Full query log: <https://gist.github.com/cutwater/8ad7eb7aa9a86232d2169c61b18e61cf>

Files

Screenshot_20201008_143356.png	1.07 MB	10/08/2020	osapryki
--------------------------------	---------	------------	----------