

## Migration Plugin - Issue #7490

### Migrating a large database with (300K rpms) uses too much memory

09/10/2020 07:28 PM - jsherril@redhat.com

<b>Status:</b>	CLOSED - CURRENTRELEASE	<b>Start date:</b>	
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	ttereshc	<b>Estimated time:</b>	0:00 hour
<b>Category:</b>		<b>Sprint Candidate:</b>	No
<b>Sprint/Milestone:</b>	0.4.0	<b>Tags:</b>	Katello
<b>Severity:</b>	2. Medium	<b>Sprint:</b>	Sprint 83
<b>Platform Release:</b>		<b>Quarter:</b>	
<b>OS:</b>			
<b>Triaged:</b>	Yes		
<b>Groomed:</b>	No		

#### Description

on a system with ~5GB free, i tried migrating a user database with 300K rpms on it. When the migration got to the 'Pre-migrating Pulp 2 RPM content (detail info)' step it got through about 200K and then the system ran out of memory:

- message: Pre-migrating Pulp 2 RPM content (detail info) code: premigrating.content.detail state: running total: 275349 done: 206000

```
Traceback: |2
  File "/usr/lib/python3.6/site-packages/rq/worker.py", line 883, in perform_job
    rv = job.perform()
  File "/usr/lib/python3.6/site-packages/rq/job.py", line 657, in perform
    self._result = self._execute()
  File "/usr/lib/python3.6/site-packages/rq/job.py", line 663, in _execute
    return self.func(*self.args, **self.kwargs)
  File "/usr/lib/python3.6/site-packages/pulp_2to3_migration/app/tasks/migrate.py", line 136
, in migrate_from_pulp2
    pre_migrate_all_content(plan)
  File "/usr/lib/python3.6/site-packages/pulp_2to3_migration/app/pre_migration.py", line 66,
in pre_migrate_all_content
    pre_migrate_content(content_model, mutable_type, lazy_type, premigrate_hook)
  File "/usr/lib/python3.6/site-packages/pulp_2to3_migration/app/pre_migration.py", line 203
, in pre_migrate_content
    content_model.pulp_2to3_detail.pre_migrate_content_detail(pulp2content_batch)
  File "/usr/lib/python3.6/site-packages/pulp_2to3_migration/app/plugin/rpm/pulp_2to3_models
.py", line 143, in pre_migrate_content_detail
    cls.objects.bulk_create(pulp2rpm_to_save, ignore_conflicts=True)
  File "/usr/lib/python3.6/site-packages/django/db/models/manager.py", line 82, in manager_m
ethod
    return getattr(self.get_queryset(), name)(*args, **kwargs)
  File "/usr/lib/python3.6/site-packages/django/db/models/query.py", line 468, in bulk_creat
e
    self._batched_insert(objs_with_pk, fields, batch_size, ignore_conflicts=ignore_conflicts
)
  File "/usr/lib/python3.6/site-packages/django/db/models/query.py", line 1211, in _batched_
insert
    self._insert(item, fields=fields, using=self.db, ignore_conflicts=ignore_conflicts)
  File "/usr/lib/python3.6/site-packages/django/db/models/query.py", line 1186, in _insert
    return query.get_compiler(using=using).execute_sql(return_id)
  File "/usr/lib/python3.6/site-packages/django/db/models/sql/compiler.py", line 1377, in ex
ecute_sql
    cursor.execute(sql, params)
  File "/usr/lib/python3.6/site-packages/django/db/backends/utils.py", line 67, in execute
    return self._execute_with_wrappers(sql, params, many=False, executor=self._execute)
  File "/usr/lib/python3.6/site-packages/django/db/backends/utils.py", line 76, in _execute_
with_wrappers
```

```
    return executor(sql, params, many, context)
File "/usr/lib/python3.6/site-packages/django/db/backends/utils.py", line 84, in _execute
    return self.cursor.execute(sql, params)
File "/usr/lib/python3.6/site-packages/django/db/utils.py", line 89, in __exit__
    raise dj_exc_value.with_traceback(traceback) from exc_value
File "/usr/lib/python3.6/site-packages/django/db/backends/utils.py", line 84, in _execute
    return self.cursor.execute(sql, params)
description: |
  out of memory
DETAIL:  Failed on request of size 1073741823 in memory context "ErrorContext".
```

I would expect that even the biggest migration should be doable on ~5 GB, so it seems like something is holding onto too much data somewhere.

---

## Associated revisions

### Revision 846bbc5b - 10/05/2020 08:58 PM - ttereshc

Pre-migrate compressed repodata for RPMs to save some memory.

Added batch\_size to all the bulk\_create calls. Decrease the size of the content batch as well.

closes #7490 <https://pulp.plan.io/issues/7490>

### Revision 846bbc5b - 10/05/2020 08:58 PM - ttereshc

Pre-migrate compressed repodata for RPMs to save some memory.

Added batch\_size to all the bulk\_create calls. Decrease the size of the content batch as well.

closes #7490 <https://pulp.plan.io/issues/7490>

### Revision 846bbc5b - 10/05/2020 08:58 PM - ttereshc

Pre-migrate compressed repodata for RPMs to save some memory.

Added batch\_size to all the bulk\_create calls. Decrease the size of the content batch as well.

closes #7490 <https://pulp.plan.io/issues/7490>

---

## History

### #1 - 09/15/2020 01:56 PM - ttereshc

- Status changed from NEW to ASSIGNED
- Assignee set to ttereshc
- Triaged changed from No to Yes
- Sprint set to Sprint 81

### #2 - 09/18/2020 09:29 PM - rchan

- Sprint changed from Sprint 81 to Sprint 82

### #3 - 10/02/2020 06:03 PM - rchan

- Sprint changed from Sprint 82 to Sprint 83

### #4 - 10/06/2020 12:02 PM - ttereshc

- Status changed from ASSIGNED to MODIFIED

Applied in changeset [pulp:pulp-2to3-migration|846bbc5bbe89dac01183b118af36fb73641151f5](https://pulp.plan.io/changesets/details/846bbc5bbe89dac01183b118af36fb73641151f5).

### #5 - 10/07/2020 09:59 AM - ttereshc

- Sprint/Milestone set to 0.4.0

### #6 - 10/07/2020 01:47 PM - pulpbot

- Status changed from MODIFIED to CLOSED - CURRENTRELEASE