

Migration Plugin - Issue #8084

"NoneType' object is not subscriptable" error during migration

01/13/2021 04:20 AM - ggainey

Status:	CLOSED - CURRENTRELEASE	Start date:	
Priority:	High	Due date:	
Assignee:	ggainey	Estimated time:	0:00 hour
Category:		Sprint Candidate:	No
Sprint/Milestone:	0.7.0	Tags:	Katello
Severity:	2. Medium	Sprint:	Sprint 89
Platform Release:		Quarter:	
OS:			
Triaged:	No		
Groomed:	No		

Description

Migrating a large number of repositories almost succeeded, until failing with a "description": "'NoneType' object is not subscriptable" error at the migrate-distributions step. At a guess, one of the 274 repositories has something unexpected, but I have not yet tracked it down to the exact one. I do have a snapshot of the Pulp2 mongo database to use when needed.

The Giant List Of Repos in this big test can be found here :

https://github.com/ggainey/pulp_startup/blob/main/8044_create_index/all_sync_pulp2.sh

The Mongo db is 7.6GB and can be made available on request.

The task-output of the failed task:

```
(pulp) [vagrant@pulp2-nightly-pulp3-source-centos7 7781_update_ordering]$ http :/pulp/api/v3/tasks/89f2388d-6c44-4307-af8a-97afb4e940ba/
HTTP/1.1 200 OK
Access-Control-Expose-Headers: Correlation-ID
Allow: GET, PATCH, DELETE, HEAD, OPTIONS
Connection: Keep-Alive
Content-Length: 7412
Content-Type: application/json
Correlation-ID: b56675d8977a4c58b60ce861f417e6e3
Date: Sat, 09 Jan 2021 01:16:00 GMT
Keep-Alive: timeout=5, max=10000
Server: gunicorn/20.0.4
Vary: Accept, Cookie
X-Frame-Options: SAMEORIGIN

{
  "child_tasks": [],
  "created_resources": [
    "/pulp/api/v3/task-groups/22082f8a-60fd-4afd-84aa-8bc9cb43a3ed/"
  ],
  "error": {
    "description": "'NoneType' object is not subscriptable",
    "traceback": " File \"/usr/local/lib/pulp/lib64/python3.6/site-packages/rq/worker.py", line 975, in perform_job\n rv = job.perform()\n File \"/usr/local/lib/pulp/lib64/python3.6/site-packages/rq/job.py", line 696, in perform\n self._result = self._execute()\n File \"/usr/local/lib/pulp/lib64/python3.6/site-packages/rq/job.py", line 719, in _execute\n return self.func(*self.args, **self.kwargs)\n File \"/home/vagrant/devel/pulp-2to3-migration/pulp_2to3_migration/app/tasks/migrate.py", line 141, in migrate_from_pulp2\n migrate_content(plan, skip_corrupted=skip_corrupted)\n File \"/home/vagrant/devel/pulp-2to3-migration/pulp_2to3_migration/app/migration.py", line 47, in migrate_content\n plugin.migrator.migrate_content_to_pulp3(skip_corrupted=skip_corrupted)\n File \"/home/vagrant/devel/pulp-2to3-migration/pulp_2to3_migration/app/plugin/rp/migrator.py", line 145, in migrate_content_to_pulp3\n loop.run_until_complete(dm.create())\n"
```

```

File \"/usr/lib64/python3.6/asyncio/base_events.py\", line 484, in run_until_complete\n    retur
n future.result()\n File \"/home/vagrant/devel/pulp-2to3-migration/pulp_2to3_migration/app/plugin
/content.py\", line 90, in create\n    await pipeline\n File \"/home/vagrant/devel/pulpcore/pulpc
ore/plugin/stages/api.py\", line 225, in create_pipeline\n    await asyncio.gather(*futures)\n Fi
le \"/home/vagrant/devel/pulpcore/pulpcore/plugin/stages/api.py\", line 43, in __call__\n    await
self.run()\n File \"/home/vagrant/devel/pulp-2to3-migration/pulp_2to3_migration/app/plugin/conte
nt.py\", line 182, in run\n    self.migrate_to_pulp3(cmodel, ctype)\n File \"/home/vagrant/devel/
pulp-2to3-migration/pulp_2to3_migration/app/plugin/content.py\", line 277, in migrate_to_pulp3\n
    pulp3content, extra_info = pulp_2to3_detail_content.create_pulp3_content()\n File \"/home/vagra
nt/devel/pulp-2to3-migration/pulp_2to3_migration/app/plugin/rpm/pulp_2to3_models.py\", line 775, i
n create_pulp3_content\n    return (DistributionTree(**treeinfo_serialized[\"distribution_tree\"]
),\n\"
},
  \"finished_at\": \"2021-01-09T01:04:42.648930Z\",
  \"logging_cid\": \"315467a7170848f29974bcb3d5273961\",
  \"name\": \"pulp_2to3_migration.app.tasks.migrate.migrate_from_pulp2\",
  \"parent_task\": null,
  \"progress_reports\": [
    {
      \"code\": \"premigrating.content.general\",
      \"done\": 0,
      \"message\": \"Pre-migrating Pulp 2 SRPM content (general info)\",
      \"state\": \"completed\",
      \"suffix\": null,
      \"total\": 0
    },
    {
      \"code\": \"premigrating.content.detail\",
      \"done\": 0,
      \"message\": \"Pre-migrating Pulp 2 SRPM content (detail info)\",
      \"state\": \"completed\",
      \"suffix\": null,
      \"total\": 0
    },
    {
      \"code\": \"premigrating.content.general\",
      \"done\": 254,
      \"message\": \"Pre-migrating Pulp 2 MODULEMD_DEFAULTS content (general info)\",
      \"state\": \"completed\",
      \"suffix\": null,
      \"total\": 254
    },
    {
      \"code\": \"premigrating.content.general\",
      \"done\": 23,
      \"message\": \"Pre-migrating Pulp 2 DISTRIBUTION content (general info)\",
      \"state\": \"completed\",
      \"suffix\": null,
      \"total\": 23
    },
    {
      \"code\": \"premigrating.content.detail\",
      \"done\": 23,
      \"message\": \"Pre-migrating Pulp 2 DISTRIBUTION content (detail info)\",
      \"state\": \"completed\",
      \"suffix\": null,
      \"total\": 23
    },
    {
      \"code\": \"processing.repositories\",
      \"done\": 127,
      \"message\": \"Processing Pulp 2 repositories, importers, distributors\",
      \"state\": \"completed\",
      \"suffix\": null,
      \"total\": 127
    },
  ],
}

```

```
    "code": "premigrating.content.detail",
    "done": 254,
    "message": "Pre-migrating Pulp 2 MODULEMD_DEFAULTS content (detail info)",
    "state": "completed",
    "suffix": null,
    "total": 254
  },
  {
    "code": "premigrating.content.general",
    "done": 75,
    "message": "Pre-migrating Pulp 2 YUM_REPO_METADATA_FILE content (general info)",
    "state": "completed",
    "suffix": null,
    "total": 75
  },
  {
    "code": "premigrating.content.detail",
    "done": 75,
    "message": "Pre-migrating Pulp 2 YUM_REPO_METADATA_FILE content (detail info)",
    "state": "completed",
    "suffix": null,
    "total": 75
  },
  {
    "code": "premigrating.content.detail",
    "done": 112495,
    "message": "Pre-migrating Pulp 2 ERRATUM content (detail info)",
    "state": "completed",
    "suffix": null,
    "total": 112495
  },
  {
    "code": "premigrating.content.general",
    "done": 371,
    "message": "Pre-migrating Pulp 2 MODULEMD content (general info)",
    "state": "completed",
    "suffix": null,
    "total": 371
  },
  {
    "code": "premigrating.content.detail",
    "done": 371,
    "message": "Pre-migrating Pulp 2 MODULEMD content (detail info)",
    "state": "completed",
    "suffix": null,
    "total": 371
  },
  {
    "code": "premigrating.content.general",
    "done": 4745,
    "message": "Pre-migrating Pulp 2 PACKAGE_GROUP content (general info)",
    "state": "completed",
    "suffix": null,
    "total": 4745
  },
  {
    "code": "premigrating.content.detail",
    "done": 4745,
    "message": "Pre-migrating Pulp 2 PACKAGE_GROUP content (detail info)",
    "state": "completed",
    "suffix": null,
    "total": 4745
  },
  {
    "code": "premigrating.content.general",
    "done": 112495,
    "message": "Pre-migrating Pulp 2 ERRATUM content (general info)",
```

```

    "state": "completed",
    "suffix": null,
    "total": 112495
  },
  {
    "code": "premigrating.content.general",
    "done": 34,
    "message": "Pre-migrating Pulp 2 PACKAGE_LANGPACKS content (general info)",
    "state": "completed",
    "suffix": null,
    "total": 34
  },
  {
    "code": "premigrating.content.detail",
    "done": 34,
    "message": "Pre-migrating Pulp 2 PACKAGE_LANGPACKS content (detail info)",
    "state": "completed",
    "suffix": null,
    "total": 34
  },
  {
    "code": "premigrating.content.general",
    "done": 396,
    "message": "Pre-migrating Pulp 2 PACKAGE_CATEGORY content (general info)",
    "state": "completed",
    "suffix": null,
    "total": 396
  },
  {
    "code": "premigrating.content.detail",
    "done": 396,
    "message": "Pre-migrating Pulp 2 PACKAGE_CATEGORY content (detail info)",
    "state": "completed",
    "suffix": null,
    "total": 396
  },
  {
    "code": "premigrating.content.general",
    "done": 180,
    "message": "Pre-migrating Pulp 2 PACKAGE_ENVIRONMENT content (general info)",
    "state": "completed",
    "suffix": null,
    "total": 180
  },
  {
    "code": "premigrating.content.detail",
    "done": 180,
    "message": "Pre-migrating Pulp 2 PACKAGE_ENVIRONMENT content (detail info)",
    "state": "completed",
    "suffix": null,
    "total": 180
  },
  {
    "code": "creating.repositories",
    "done": 127,
    "message": "Creating repositories in Pulp 3",
    "state": "completed",
    "suffix": null,
    "total": 127
  },
  {
    "code": "migrating.importers",
    "done": 127,
    "message": "Migrating importers to Pulp 3",
    "state": "completed",
    "suffix": null,
    "total": 127
  }

```

```

    },
    {
      "code": "migrating.rpm.content",
      "done": 259738,
      "message": "Migrating rpm content to Pulp 3 rpm",
      "state": "completed",
      "suffix": null,
      "total": 259738
    },
    {
      "code": "migrating.content",
      "done": 0,
      "message": "Migrating content to Pulp 3",
      "state": "failed",
      "suffix": null,
      "total": 0
    },
    {
      "code": "premigrating.content.general",
      "done": 259738,
      "message": "Pre-migrating Pulp 2 RPM content (general info)",
      "state": "completed",
      "suffix": null,
      "total": 259738
    },
    {
      "code": "premigrating.content.detail",
      "done": 259738,
      "message": "Pre-migrating Pulp 2 RPM content (detail info)",
      "state": "completed",
      "suffix": null,
      "total": 259738
    },
    {
      "code": "migrating.rpm.content",
      "done": 0,
      "message": "Migrating rpm content to Pulp 3 srpm",
      "state": "completed",
      "suffix": null,
      "total": 0
    },
    {
      "code": "migrating.rpm.content",
      "done": 0,
      "message": "Migrating rpm content to Pulp 3 distribution",
      "state": "failed",
      "suffix": null,
      "total": 23
    }
  ],
  "pulp_created": "2021-01-08T20:40:04.242552Z",
  "pulp_href": "/pulp/api/v3/tasks/89f2388d-6c44-4307-af8a-97afb4e940ba/",
  "reserved_resources_record": [
    "pulp_2to3_migration"
  ],
  "started_at": "2021-01-08T20:40:04.441480Z",
  "state": "failed",
  "task_group": "/pulp/api/v3/task-groups/22082f8a-60fd-4afd-84aa-8bc9cb43a3ed/",
  "worker": "/pulp/api/v3/workers/2d6a7e75-85d7-471b-9ce0-aa850599656d/"
}

```

Associated revisions

Revision c66c3cba - 02/03/2021 06:19 PM - ggainey

Taught rpm-migration to be more stable in the face of missing treeinfos.

fixes #8084

Revision c66c3cba - 02/03/2021 06:19 PM - ggaaney

Taught rpm-migration to be more stable in the face of missing treeinfos.

fixes #8084

Revision c66c3cba - 02/03/2021 06:19 PM - ggaaney

Taught rpm-migration to be more stable in the face of missing treeinfos.

fixes #8084

History

#1 - 01/14/2021 05:02 PM - jsherril@redhat.com

- Priority changed from Normal to High

- Tags Katello added

I hit this as well on a user database

#3 - 01/19/2021 02:32 AM - dalley

```
def get_treeinfo_serialized(self):
    """
    Create a Pulp 3 Distribution content for saving later in a bulk operation.

    Returns:
        dict: a serialized treeinfo data

    """
    namespaces = [".treeinfo", "treeinfo"]
    for namespace in namespaces:
        treeinfo = PulpTreeInfo()
        try:
            treeinfo.load(f=os.path.join(self.pulp2content.pulp2_storage_path, namespace))
        except FileNotFoundError:
            continue
        self.filename = namespace
        treeinfo_parsed = treeinfo.parsed_sections()
        treeinfo_serialized = TreeinfoData(treeinfo_parsed).to_dict(filename=namespace)
        # Pulp 2 only knows about the top level kickstart repository
        treeinfo_serialized["repositories"] = {'.': None}
        # Pulp 2 did not support addon repositories, so we should not list them here either
        treeinfo_serialized['addons'] = {}
        # Pulp 2 only supported variants that are in the root of the repository
        variants = {}
        for name, variant in treeinfo_serialized['variants'].items():
            if variant['repository'] == '.':
                variants[name] = variant
        treeinfo_serialized['variants'] = variants
        # Reset build_timestamp so Pulp will fetch all the addons during the next sync
        treeinfo_serialized['distribution_tree']['build_timestamp'] = 0
    return treeinfo_serialized
```

```
def create_pulp3_content(self):
    """
    Create a Pulp 3 Distribution content for saving later in a bulk operation.

    Returns:
        (DistributionTree, dict): an in-memory DistributionTree content and serialized
        treeinfo data

    """
    treeinfo_serialized = self.get_treeinfo_serialized()
    return (DistributionTree(**treeinfo_serialized["distribution_tree"]),
            treeinfo_serialized)
```

It looks like "treeinfo_serilized" is None and therefore the subscript operation fails. Either .get_treeinfo_serialized() is falsely returning None, or this chunk of code is being called in a situation where it shouldn't be.

.get_treeinfo_serialized() should probably throw an assertion failure if it can't find what it's looking for, instead of silent None regardless of where the problem is.

Tanya do you have any ideas?

#4 - 01/19/2021 07:47 PM - dalley

- Sprint set to Sprint 88

#5 - 01/19/2021 08:16 PM - ttereshc

- Sprint/Milestone set to 0.9.0

#6 - 01/22/2021 09:32 PM - rchan

- Sprint changed from Sprint 88 to Sprint 89

#7 - 01/27/2021 04:06 PM - ggainey

- Status changed from NEW to ASSIGNED

- Assignee set to ggainey

#8 - 01/28/2021 02:03 PM - ggainey

debugging notes:

```
Jan 28 03:18:22 pulp2-nightly-pulp3-source-centos7.padre-fedora.example.com rq[6382]:
pulp [66acc5ec9aad4796b3cea9ecf2321620]: pulp_2to3_migration.app.plugin.rpm.pulp_2to3_models:
ERROR: >>> NO TREEINFO FOUND for
/var/lib/pulp/content/units/distribution/82/8a832a0d16c7d561db0fa2ac129112501b3003ce1fc677917bc548a430f9bc
```

```
pulp=> select * from pulp_2to3_migration_pulp2content where pulp2_storage_path = '/var/lib/pulp/content/units/distribution/82/8a832a0d16c7d561db0fa2ac129112501b3003ce1fc677917bc548a430f9bc'
```

```
pulp-> ;
```

pulp2_id	pulp_id	pulp_created	pulp_last_updated	pulp2_storage_path
94	1610014880	2021-01-27 23:41:19.161583+00	2021-01-27 23:41:19.161616+00	/var/lib/pulp/content/units/distribution/82/8a832a0d16c7d561db0fa2ac129112501b3003ce1fc677917bc548a430f9bc

```
pulp=> select * from pulp_2to3_migration_pulp2distribution where pulp_2to3_migration_pulp2distribution.pulp2content_id = '63ca3324-c132-4c8e-bb6f-cef1d64c6913';
```

distribution_id	pulp_id	pulp_created	pulp_last_updated	pulp2content_id
ks-Red Hat Enterprise Linux-BaseOS-8.3-x86_64	1610014880	2021-01-27 23:41:19.850627+00	2021-01-27 23:41:19.850635+00	63ca3324-c132-4c8e-bb6f-cef1d64c6913

pulp=>

#9 - 01/28/2021 02:35 PM - ggainey

```
> db.units_distribution.find({'distribution_id': 'ks-Red Hat Enterprise Linux-BaseOS-8.3-x86_64'})
{"_id": "1da47a3a-b568-4efa-be13-59854a0ead94",
"pulp_user_metadata": {},
"_last_updated": 1610014880,
"_storage_path": "/var/lib/pulp/content/units/distribution/82/8a832a0d16c7d561db0fa2ac129112501b3003ce1fc677917bc548a430f9bc",
"downloaded": false,
"distribution_id": "ks-Red Hat Enterprise Linux-BaseOS-8.3-x86_64",
"family": "Red Hat Enterprise Linux", "variant": "BaseOS", "version": "8.3", "arch": "x86_64",
"files": [
{"relativepath": "images/pxeboot/initrd.img", "checksumtype": "sha256", "checksum": "2e974f7a2e79d7f7db24"}]
```

```
beccf9af3b244f9edcd464990c9f6d7bc2e3d9b7151b" },
  { "relativepath" : "images/pxeboot/vmlinuz", "checksumtype" : "sha256", "checksum" : "98371ec7d815c710fa21a7a
b3c59a9c3aed98b7a3a2c78237c24528cd5aa3353" },
  { "relativepath" : "images/boot.iso", "checksumtype" : "sha256", "checksum" : "1b73ebfebd1f9424c806032168873b
067259d8b29f4e9d39ae0e4009ccea49b93" },
  { "relativepath" : "images/efiboot.img", "checksumtype" : "sha256", "checksum" : "cfa231ea085e32b116afcbcdcc8
7857948f0172b090427c52e45bbc80cefa2f3" },
  { "relativepath" : "images/install.img", "checksumtype" : "sha256", "checksum" : "4e00b6a55dd0dc4d560aada4b9c9
20ec1bdbf1df69c273bb4a59ba98d9e7e9ca" }
],
"timestamp" : 1602239855, "packagedir" : "Packages", "version_sort_index" : "01-8.01-3", "_ns" : "units_distrib
ution", "_content_type_id" : "distribution" }
```

#10 - 01/28/2021 02:52 PM - ggainey

Current hypothesis: All the repos in this example are on-demand in pulp2, and have no content sync'd. .treeinfo files never get lazy-content entries (I think). So this code fails, because it 'assumes' that it will always be able to find a .treeinfo file 'on disk' in pulp2.

#11 - 01/28/2021 05:30 PM - ggainey

OK< the source of the problem described here, is that both test-cases are running against a mongo-db export from a long-gone pulp2 system - *nothing* exists "on disk", only in the DB. Running with skip-corrupted should let the migration complete (after a lot of bitter complaints). However, there is a path involving missing-treeinfo that results in the python-exception that interrupts the migration.

I think I have a path forward that will let this scenario run to completion. We're testing something that is a never-happen-case for "actual" users of migration - but which is very very common for testers...

#12 - 01/29/2021 05:25 PM - ggainey

Testing:

```
# Sync one ks-repo into Pulp2
pulp-admin login -u admin -p admin
pulp-admin rpm repo create --repo-id rhel8-ks --feed https://cdn.redhat.com/content/dist/rhel8/8/x86_64/baseos
/kickstart \
  --download-policy on_demand --generate-sqlite false --repoview false \
  --feed-ca-cert /home/vagrant/devel/pulp_startup/CDN_cert/redhat-uep.pem \
  --feed-cert /home/vagrant/devel/pulp_startup/CDN_cert/cdn.crt \
  --feed-key /home/vagrant/devel/pulp_startup/CDN_cert/cdn.key
pulp-admin rpm repo sync run --repo-id rhel8-ks
# Dump the resulting mongodb
mongodump --db pulp_database --gzip
# Reset pulp2 to 'empty'
reset_pulp2
# Restore just the mongodb
mongorestore --gzip .
# Set up and migrate that data into pulp3
http POST :24817/pulp/api/v3/migration-plans/ plan='{ "plugins": [{"type": "rpm"}]}'
http POST /pulp/api/v3/migration-plans/<migration-plan-UUID>/run/ skip_corrupted=true
```

Without a fix, this will FAIL with the "'NoneType' object is not subscriptable" error. With the fix, the task will complete successfully while complaining about missing/corrupted files.

#13 - 02/01/2021 02:58 PM - ggainey

- Status changed from ASSIGNED to POST

<https://github.com/pulp/pulp-2to3-migration/pull/303>

#14 - 02/03/2021 06:15 PM - ttereshc

- Sprint/Milestone changed from 0.9.0 to 0.7.0

#15 - 02/03/2021 06:19 PM - ggainey

- Status changed from POST to MODIFIED

Applied in changeset [pulp:pulp-2to3-migration|c66c3c6a6567844fd0a5893f4577a50ed785c011](https://pulp.org/pulp-2to3-migration/c66c3c6a6567844fd0a5893f4577a50ed785c011).

#16 - 02/04/2021 12:39 PM - pulpbot

- Status changed from MODIFIED to CLOSED - CURRENTRELEASE