

Pulp - Issue #3412

Trailing slash in distribution base_path causes breakage (Pulp 3)

02/28/2018 09:07 PM - Ichimonji10

Status:	CLOSED - CURRENTRELEASE	Start date:	
Priority:	Normal	Due date:	
Assignee:	daviddavis	Estimated time:	0:00 hour
Category:		Groomed:	No
Sprint/Milestone:	3.0.0	Sprint Candidate:	No
Severity:	2. Medium	Tags:	
Version:		Sprint:	Sprint 34
Platform Release:		Quarter:	
OS:			
Triaged:	Yes		

Description

Let's say I create a distribution by making an HTTP POST request to `/api/v3/distributions/`, with the following JSON body:

```
{
  "base_path": "my-base-path",
  "http": true,
  "https": true,
  "name": "distribution-name",
  "publication": "http://pulp-3.example.com:8000/api/v3/publications/publication-id/",
}
```

This will return a distribution:

```
{
  "_href": "http://pulp-3.example.com:8000/api/v3/distributions/distribution-name/",
  "base_path": "my-base-path",
  "base_url": "pulp-3.example.com:8000/content/my-base-path",
  "http": true,
  "https": true,
  "name": "distribution-name",
  "publication": "http://pulp-3.example.com:8000/api/v3/publications/publication-id/",
  "publisher": null,
  "repository": null
}
```

Furthermore, let's say that the publication at `.../api/v3/publications/publication-id/` contains a file named `1.iso`. At which URLs are this file available? The following:

- `http://pulp-3.example.com:8000/content/my-base-path/1.iso`
- `https://pulp-3.example.com:8000/content/my-base-path/1.iso`

Before going any further, note that the process of generating these paths is awkward, because the data in the distributor is awkward. Let's say that I write some Python code to come up with the path to `1.iso`. A naive attempt will totally fail:

```
from urllib.parse import urljoin

distribution = {
  '_href': 'http://pulp-3.example.com:8000/api/v3/distributions/distribution-name/',
  'base_path': 'my-base-path',
  'base_url': 'pulp-3.example.com:8000/content/my-base-path',
  'http': True,
  'https': True,
  'name': 'distribution-name',
  'publication': 'http://pulp-3.example.com:8000/api/v3/publications/publication-id/',
  'publisher': None,
```

```

    'repository': None
}
basename = '1.iso'
# ['1.iso', '1.iso']
urls = [
    urljoin(distribution['base_url'], basename)
    for scheme in ('http', 'https')
    if distribution[scheme]
]

# ['http://pulp-3.example.com:8000/content/1.iso', 'https://pulp-3.example.com:8000/content/1.iso']
]
urls = [
    urljoin(scheme + '://' + distribution['base_url'], basename)
    for scheme in ('http', 'https')
    if distribution[scheme]
]

# ['http://pulp-3.example.com:8000/content/my-base-path/1.iso', 'https://pulp-3.example.com:8000/c
ontent/my-base-path/1.iso']
urls = [
    urljoin(scheme + '://' + distribution['base_url'] + '/', basename)
    for scheme in ('http', 'https')
    if distribution[scheme]
]

```

This user experience sucks. But that's not what this bug is about. This bug is **really** about the fact Pulp accepts `base_paths` that it subsequently barfs on. Let's say that I create the exact same distribution as before, but with a trailing slash on the `base_path`:

```

{
  "base_path": "my-base-path/",
  "http": true,
  "https": true,
  "name": "distribution-name",
  "publication": "http://pulp-3.example.com:8000/api/v3/publications/publication-id/",
}

```

Pulp will accept the trailing slash:

```

{
  "_href": "http://pulp-3.example.com:8000/api/v3/distributions/distribution-name/",
  "base_path": "my-base-path/",
  "base_url": "pulp-3.example.com:8000/content/my-base-path/",
  "http": true,
  "https": true,
  "name": "distribution-name",
  "publication": "http://pulp-3.example.com:8000/api/v3/publications/publication-id/",
  "publisher": null,
  "repository": null
}

```

The file named `1.iso` **should** be available at the following URLs:

- `http://pulp-3.example.com:8000/content/my-base-path/1.iso`
- `https://pulp-3.example.com:8000/content/my-base-path/1.iso`

Unfortunately, HTTP requests to these URLs will produce HTTP 404s.

Please do one of the following:

- Reject `base_urls` that have a trailing slash.
- Accept `base_urls` that have a trailing slash, and handle them properly.

When deciding upon a solution, please consider the user story outlined above.

Associated revisions

Revision 4c92978e - 03/15/2018 06:32 PM - daviddavis

Validate distribution base paths

Validate that distribution base paths contain valid urls and don't end or begin with slashes

fixes #3412 <https://pulp.plan.io/issues/3412>

Revision 4c92978e - 03/15/2018 06:32 PM - daviddavis

Validate distribution base paths

Validate that distribution base paths contain valid urls and don't end or begin with slashes

fixes #3412 <https://pulp.plan.io/issues/3412>

History

#1 - 02/28/2018 09:14 PM - Ichimonji10

- Description updated

#2 - 03/01/2018 05:01 PM - Ichimonji10

- Subject changed from *Trailing slash in distribution base_url causes breakage (Pulp 3)* to *Trailing slash in distribution base_path causes breakage (Pulp 3)*

#3 - 03/01/2018 05:19 PM - daviddavis

- Project changed from *File Support* to *Pulp*

#4 - 03/02/2018 04:49 PM - dalley

- Sprint/Milestone set to 56

- Triaged changed from *No* to *Yes*

#5 - 03/09/2018 12:35 AM - bmbouter

- Sprint set to *Sprint 33*

#6 - 03/09/2018 12:36 AM - bmbouter

- Sprint/Milestone deleted (56)

#7 - 03/12/2018 02:37 PM - jortel@redhat.com

- Sprint changed from *Sprint 33* to *Sprint 34*

#8 - 03/13/2018 02:13 AM - daviddavis

- Status changed from *NEW* to *POST*

- Assignee set to *daviddavis*

<https://github.com/pulp/pulp/pull/3371>

#10 - 03/15/2018 06:32 PM - daviddavis

- Status changed from *POST* to *MODIFIED*

Applied in changeset <pulp|4c92978e13566e3a3fbd84582f0b29198cbf392e>.

#11 - 04/25/2019 06:46 PM - daviddavis

- Sprint/Milestone set to *3.0.0*

#12 - 04/26/2019 10:36 PM - bmbouter

- Tags deleted (*Pulp 3*)

#13 - 12/13/2019 06:30 PM - bmbouter

- Status changed from *MODIFIED* to *CLOSED - CURRENTRELEASE*