

Container Support - Story #4456

As a user I can use S3 as alternative storage

02/25/2019 01:33 PM - daviddavis

Status:	CLOSED - CURRENTRELEASE	Start date:	
Priority:	Normal	Due date:	
Assignee:	mdellweg	% Done:	100%
Category:		Estimated time:	0:00 hour
Sprint/Milestone:	1.4.0	Tags:	
Platform Release:		Sprint:	Sprint 73
Groomed:	No	Quarter:	
Sprint Candidate:	No		
Description			
Make sure that the pulp docker plugin works when using Pulp with S3. I see no reason why it shouldn't but it's worth testing to confirm.			
Related issues:			
Related to Pulp - Story #3900: As a user, I can use Pulp3 on S3		CLOSED - CURRENTRELEASE	
Blocks Container Support - Task #6733: Enable S3 tests on CI		CLOSED - CURRENTRELEASE	

Associated revisions

Revision 32a2d589 - 05/25/2020 06:45 PM - ipanova@redhat.com

Enable S3 storage.

closes #4456 <https://pulp.plan.io/issues/4456>

Revision 32a2d589 - 05/25/2020 06:45 PM - ipanova@redhat.com

Enable S3 storage.

closes #4456 <https://pulp.plan.io/issues/4456>

History

#1 - 02/25/2019 01:34 PM - daviddavis

- Related to Story #3900: As a user, I can use Pulp3 on S3 added

#2 - 04/26/2019 10:31 PM - bmbouter

- Tags deleted (Pulp 3)

#3 - 10/04/2019 11:30 AM - ipanova@redhat.com

- Tags Pulp 3 docker blocker added

#4 - 10/14/2019 09:58 PM - fao89

- Status changed from NEW to ASSIGNED

- Assignee set to fao89

#5 - 10/15/2019 01:33 PM - ipanova@redhat.com

- Sprint set to Sprint 60

#6 - 10/15/2019 09:51 PM - fao89

Followed the docs till reach here: <https://pulp-docker.readthedocs.io/en/latest/workflows/host.html#pull-and-run-an-image-from-pulp>
for docker:

```
(pulp) [vagrant@pulp3-source-fedora30 pulp_docker]$ sudo docker pull localhost:24816/test
Using default tag: latest
Trying to pull repository localhost:24816/test ...
error parsing HTTP 404 response body: invalid character ':' after top-level value: "404: Not Found"
```

On the server:

```
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: [2019-10-15 19:37:12 +0000] [12574] [ERROR] Error handling request
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: Traceback (most recent call last)
:
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: File "/usr/local/lib/pulp/lib64/python3.7/site-packages/aiohttp/web_protocol.py", line 275, in data_received
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: messages, upgraded, tail = self._request_parser.feed_data(data)
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: File "aiohttp/_http_parser.pyx", line 523, in aiohttp._http_parser.HttpParser.feed_data
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: aiohttp.http_exceptions.BadStatusLine: invalid HTTP method
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: 127.0.0.1 [15/Oct/2019:19:37:12 +0000] "GET /v2/ HTTP/1.1" 200 224 "-" "docker/1.13.1 go/go1.12.7 kernel/5.0.9-301.fc30.x86_64 os/linux arch/amd64 UpstreamClient(Docker-Client/1.13.1 \(linux\))"
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: [2019-10-15 19:37:12 +0000] [12574] [ERROR] Error handling request
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: Traceback (most recent call last)
:
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: File "/usr/local/lib/pulp/lib64/python3.7/site-packages/aiohttp/web_protocol.py", line 275, in data_received
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: messages, upgraded, tail = self._request_parser.feed_data(data)
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: File "aiohttp/_http_parser.pyx", line 523, in aiohttp._http_parser.HttpParser.feed_data
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: aiohttp.http_exceptions.BadStatusLine: invalid HTTP method
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: 127.0.0.1 [15/Oct/2019:19:37:12 +0000] "GET /v2/ HTTP/1.1" 200 224 "-" "docker/1.13.1 go/go1.12.7 kernel/5.0.9-301.fc30.x86_64 os/linux arch/amd64 UpstreamClient(Docker-Client/1.13.1 \(linux\))"
Oct 15 19:37:12 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: 127.0.0.1 [15/Oct/2019:19:37:12 +0000] "GET /v2/test/manifests/latest HTTP/1.1" 404 191 "-" "docker/1.13.1 go/go1.12.7 kernel/5.0.9-301.fc30.x86_64 os/linux arch/amd64 UpstreamClient(Docker-Client/1.13.1 \(linux\))"
```

for podman:

```
(pulp) [vagrant@pulp3-source-fedora30 pulp_docker]$ podman pull localhost:24816/test
Trying to pull localhost:24816/test...
error parsing HTTP 404 response body: invalid character ':' after top-level value: "404: Not Found"
Error: error pulling image "localhost:24816/test": unable to pull localhost:24816/test: unable to pull image: Error initializing source docker://localhost:24816/test:latest: Error reading manifest latest in localhost:24816/test: error parsing HTTP 404 response body: invalid character ':' after top-level value: "404: Not Found"
```

On the server:

```
Oct 15 19:36:37 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: [2019-10-15 19:36:37 +0000] [12575] [ERROR] Error handling request
Oct 15 19:36:37 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: Traceback (most recent call last)
:
Oct 15 19:36:37 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: File "/usr/local/lib/pulp/lib64/python3.7/site-packages/aiohttp/web_protocol.py", line 275, in data_received
Oct 15 19:36:37 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: messages, upgraded, tail = self._request_parser.feed_data(data)
Oct 15 19:36:37 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: File "aiohttp/_http_parser.pyx", line 523, in aiohttp._http_parser.HttpParser.feed_data
Oct 15 19:36:37 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: aiohttp.http_exceptions.BadStatusLine: invalid HTTP method
Oct 15 19:36:37 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: 127.0.0.1 [15/Oct/2019:19:36:37 +0000] "GET /v2/ HTTP/1.1" 200 224 "-" "libpod/1.6.1"
Oct 15 19:36:37 pulp3-source-fedora30.localhost.example.com gunicorn[12571]: 127.0.0.1 [15/Oct/2019:19:36:37 +0000] "GET /v2/test/manifests/latest HTTP/1.1" 404 191 "-" "libpod/1.6.1"
```

#7 - 10/15/2019 10:19 PM - fao89

I destroyed vagrant and started again, but still having problems:

```
(pulp) [vagrant@pulp3-source-fedora30 pulp_docker]$ podman pull localhost:24816/test
Trying to pull localhost:24816/test...
Get https://localhost:24816/v2/: http: server gave HTTP response to HTTPS client
Error: error pulling image "localhost:24816/test": unable to pull localhost:24816/test: unable to pull image: Error initializing source docker://localhost:24816/test:latest: pinging docker registry returned: Get https://localhost:24816/v2/: http: server gave HTTP response to HTTPS client
```

following the docs I edited: /etc/containers/registries.conf
and then:

```
(pulp) [vagrant@pulp3-source-fedora30 pulp_docker]$ podman pull localhost:24816/test
Trying to pull localhost:24816/test...
  error parsing HTTP 404 response body: invalid character ':' after top-level value: "404: Not Found"
Error: error pulling image "localhost:24816/test": unable to pull localhost:24816/test: unable to pull image:
Error initializing source docker://localhost:24816/test:latest: Error reading ma
nifest latest in localhost:24816/test: error parsing HTTP 404 response body: invalid character ':' after top-l
evel value: "404: Not Found"
```

#8 - 10/15/2019 10:39 PM - fao89

PS: I've been pinning host when self.context['request'] is None:

```
class RegistryPathField(serializers.CharField):
    """
    Serializer Field for the registry_path field of the DockerDistribution.
    """

    def to_representation(self, value):
        """
        Converts a base_path into a registry path.
        """
        if settings.CONTENT_HOST:
            host = settings.CONTENT_HOST
        else:
            try:
                host = self.context['request'].get_host()
            except:
                host = "http://localhost:24817"
        return ''.join([host, '/', value])
```

#9 - 10/15/2019 11:19 PM - fao89

went to following script on the docs:

```
#!/usr/bin/env bash

DOCKER_TAG='manifest_a'

echo "Setting REGISTRY_PATH, which can be used directly with the Docker Client."
export REGISTRY_PATH=$(http $BASE_ADDR$DISTRIBUTION_HREF | jq -r '.registry_path')

echo "Next we pull the image from pulp and run it."
echo "$REGISTRY_PATH:$DOCKER_TAG"
sudo docker run $REGISTRY_PATH:$DOCKER_TAG
```

but I ran on podman:

```
(pulp) [vagrant@pulp3-source-fedora30 pulp_docker]$ echo "$REGISTRY_PATH:$DOCKER_TAG"
localhost:24817/test:manifest_a
(pulp) [vagrant@pulp3-source-fedora30 pulp_docker]$ podman run $REGISTRY_PATH:$DOCKER_TAG
Trying to pull localhost:24817/test:manifest_a...
  Get https://localhost:24817/v2/: net/http: TLS handshake timeout
Error: unable to pull localhost:24817/test:manifest_a: unable to pull image: Error initializing source docker:
//localhost:24817/test:manifest_a: pinging docker registry returned: Get https:/
localhost:24817/v2/: net/http: TLS handshake timeout
```

changed the port:

```
(pulp) [vagrant@pulp3-source-fedora30 pulp_docker]$ podman run localhost:24816/test:manifest_a

Trying to pull localhost:24816/test:manifest_a...
  received unexpected HTTP status: 500 Internal Server Error
Error: unable to pull localhost:24816/test:manifest_a: unable to pull image: Error initializing source docker:
//localhost:24816/test:manifest_a: Error reading manifest manifest_a in localhos
t:24816/test: received unexpected HTTP status: 500 Internal Server Error
```

On the server:

```
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: [2019-10-15 21:15:12 +0000] [1883]
[ERROR] Error handling request
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: Traceback (most recent call last):
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]:   File "/usr/local/lib/pulp/lib64/
```

```
python3.7/site-packages/aiohttp/web_protocol.py", line 275, in data_received
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: messages, upgraded, tail = self._request_parser.feed_data(data)
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: File "aiohttp/_http_parser.pyx", line 523, in aiohttp._http_parser.HttpParser.feed_data
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: aiohttp.http_exceptions.BadStatusLine: invalid HTTP method
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: 127.0.0.1 [15/Oct/2019:21:15:12 +0000] "GET /v2/ HTTP/1.1" 200 224 "-" "libpod/1.6.1"
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: [2019-10-15 21:15:12 +0000] [1883] [ERROR] Error handling request
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: Traceback (most recent call last):
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: File "/usr/local/lib/pulp/lib64/python3.7/site-packages/aiohttp/web_protocol.py", line 418, in start
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: resp = await task
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: File "/usr/local/lib/pulp/lib64/python3.7/site-packages/aiohttp/web_app.py", line 458, in _handle
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: resp = await handler(request)
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: File "/home/vagrant/devel/pulp_docker/pulp_docker/app/registry.py", line 168, in get_tag
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: return await Registry.dispatch_tag(tag, response_headers)
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: File "/home/vagrant/devel/pulp_docker/pulp_docker/app/registry.py", line 191, in dispatch_tag
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: response_headers)
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: File "/home/vagrant/devel/pulp_docker/pulp_docker/app/registry.py", line 89, in _dispatch
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: full_headers['Content-Length'] = os.path.getsize(path)
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: File "/usr/lib64/python3.7/genericpath.py", line 50, in getsize
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: return os.stat(filename).st_size
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: FileNotFoundError: [Errno 2] No such file or directory: 'artifact/21/e3caae28758329318c8a868a80daa37ad8851705155fc28767852c73d36af5'
Oct 15 21:15:12 pulp3-source-fedora30.localhost.example.com gunicorn[1870]: 127.0.0.1 [15/Oct/2019:21:15:12 +0000] "GET /v2/test/manifests/manifest_a HTTP/1.1" 500 244 "-" "libpod/1.6.1"
```

But the artifact is there:

<https://test-pulp3.s3.us-east-2.amazonaws.com/artifact/21/e3caae28758329318c8a868a80daa37ad8851705155fc28767852c73d36af5>

I believe it is related to MEDIA_ROOT, which I set as MEDIA_ROOT = " like on the docs:

<https://docs.pulpproject.org/en/3.0/nightly/installation/storage.html#configuring-pulp>

#10 - 10/16/2019 05:31 PM - fao89

with these changes:

https://github.com/pulp/pulp_docker/pull/433

```
(pulp) [vagrant@pulp3-source-fedora30 pulp_docker]$ http GET $CONTENT_ADDR/v2/test/manifests/manifest_a "Accept:application/vnd.docker.distribution.manifest.v2+json"
HTTP/1.1 302 Found
Content-Disposition: attachment; filename=e3caae28758329318c8a868a80daa37ad8851705155fc28767852c73d36af5
Content-Length: 524
Content-Type: application/vnd.docker.distribution.manifest.v2+json; charset=utf-8
Date: Wed, 16 Oct 2019 15:20:20 GMT
Docker-Content-Digest: sha256:21e3caae28758329318c8a868a80daa37ad8851705155fc28767852c73d36af5
Docker-Distribution-API-Version: registry/2.0
Location: https://s3.us-east-2.amazonaws.com/test-pulp3/artifact/21/e3caae28758329318c8a868a80daa37ad8851705155fc28767852c73d36af5?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=*****/20191016/us-east-2/s3/aws4_request&X-Amz-Date=20191016T152020Z&X-Amz-Expires=3600&X-Amz-SignedHeaders=host&X-Amz-Signature=2ece0668eff58a64d2f8abb3bbe0be58bda69bb8bad3d7a26c0da0191d81d8
Server: Python/3.7 aiohttp/3.6.2
```

302: Found

```
(pulp) [vagrant@pulp3-source-fedora30 pulp_docker]$ sudo docker pull localhost:24816/test:manifest_a
Trying to pull repository localhost:24816/test ...
unsupported schema version 2
```

```
(pulp) [vagrant@pulp3-source-fedora30 pulp_docker]$ sudo docker run $REGISTRY_PATH:$DOCKER_TAG
```

```
Unable to find image 'localhost:24816/test:manifest_a' locally
```

```
Trying to pull repository localhost:24816/test ...
```

```
/usr/bin/docker-current: unsupported schema version 2.
```

#11 - 10/16/2019 05:46 PM - ipanova@redhat.com

We figured out that problem is that fact that the headers are not set on S3. Docker clients checks for content-type, digest, etc and other headers. otherwise it fails the pull.

There is a way how to set headers in s3 <https://docs.aws.amazon.com/AmazonS3/latest/dev/cors.html> but in our case for each served file we need to somehow figure out its content-type and digest.

#12 - 10/16/2019 06:01 PM - dkliban@redhat.com

When Pulp is using S3, it needs to set these headers on the file when the Artifact is being created. That way S3 knows right away how to serve that file. Not sure how to achieve this though.

#13 - 10/16/2019 08:12 PM - fao89

- Status changed from ASSIGNED to NEW

- Assignee deleted (fao89)

#14 - 10/16/2019 08:38 PM - daviddavis

When Pulp is using S3, it needs to set these headers on the file when the Artifact is being created. That way S3 knows right away how to serve that file. Not sure how to achieve this though.

Does docker allow artifacts to be shared between content units? If so, what will it do in that case where an artifact could have different header info?

Another option might be to have the content app stream the artifact (although this is not really optimal).

#15 - 10/16/2019 09:00 PM - bmbouter

Isn't there a way to have AWS hand specific headers to the client by settings them upon the redirect?

We could manipulate those headers here maybe? <https://github.com/pulp/pulpcore/blob/master/pulpcore/content/handler.py#L429>

#16 - 10/16/2019 09:16 PM - daviddavis

Looks like maybe. Googling a bit lead me to this page: https://docs.aws.amazon.com/AmazonS3/latest/API/API_GetObject.html. See the section "Overriding Response Header Values". It looks like it's only a small subset of headers that can be specified.

#17 - 10/17/2019 02:02 PM - ipanova@redhat.com

bmbouter wrote:

Isn't there a way to have AWS hand specific headers to the client by settings them upon the redirect?

We could manipulate those headers here maybe? <https://github.com/pulp/pulpcore/blob/master/pulpcore/content/handler.py#L429>

This will only ensure that the 302 redirect contains proper headers.[0]

Problem consists in the fact that when the client follows the redirect through the "location" header it gets the file, stored on S3 that has no headers set in the response. These are docker registry specific headers and without them docker/podman client will refuse to pull.

[0] https://github.com/pulp/pulp_docker/pull/433/files#diff-1f37e1bf95e24a173326983f481027a7R103

#18 - 10/17/2019 02:09 PM - ipanova@redhat.com

dkliban@redhat.com wrote:

When Pulp is using S3, it needs to set these headers on the file when the Artifact is being created. That way S3 knows right away how to serve that file. Not sure how to achieve this though.

I still don't understand how that will help S3 to know how to serve the file? once we trigger 302 redirect we loose control by directing the client to S3 to fetch the file from there. Unless we can set somehow those headers directly in the response from S3, we won't be able to make docker pull work.

#19 - 10/18/2019 03:22 PM - ipanova@redhat.com

There is a way to specify headers when uploading file to S3

via direct s3 api https://docs.aws.amazon.com/AmazonS3/latest/API/API_PutObject.html#API_PutObject_RequestSyntax

We need to take a look how to do it via boto3 or django-storages.

#20 - 10/25/2019 09:46 AM - rchan

- Sprint changed from Sprint 60 to Sprint 61

#21 - 10/25/2019 10:39 AM - ipanova@redhat.com

- Sprint deleted (Sprint 61)

- Tags deleted (Pulp 3 docker blocker)

#22 - 10/25/2019 10:40 AM - ipanova@redhat.com

We will address this post GA

https://github.com/pulp/pulp_docker/pull/433

#23 - 11/20/2019 07:17 PM - ipanova@redhat.com

- Project changed from Docker Support to Container Support

#24 - 05/15/2020 01:54 PM - ipanova@redhat.com

- Tracker changed from Task to Story

- Subject changed from Test out docker with S3 to As a user I can use S3 as alternative storage

#25 - 05/15/2020 03:41 PM - ipanova@redhat.com

- Blocks Task #6733: Enable S3 tests on CI added

#26 - 05/15/2020 03:41 PM - ipanova@redhat.com

- Sprint set to Sprint 73

#27 - 05/20/2020 10:33 AM - mdellweg

- Status changed from NEW to ASSIGNED

- Assignee set to mdellweg

#28 - 05/25/2020 05:48 PM - pulpbot

- Status changed from ASSIGNED to POST

PR: https://github.com/pulp/pulp_container/pull/80

#29 - 05/25/2020 06:59 PM - ipanova@redhat.com

- Status changed from POST to MODIFIED

- % Done changed from 0 to 100

Applied in changeset [32a2d58955158b93a394c4303b90ee329bf0cecb](https://github.com/pulp/pulp_container/pull/80).

#30 - 05/28/2020 08:25 PM - ipanova@redhat.com

- Sprint/Milestone set to 1.4.0

#31 - 05/28/2020 08:25 PM - ipanova@redhat.com

- Status changed from MODIFIED to CLOSED - CURRENTRELEASE