

Infrastructure - Bug #8812

Resolve service returns 500 for HTTP HEAD request

2019-05-22 16:59 - Peter Slaughter

| | | | |
|-------------------------|---------------|------------------------|------------|
| Status: | Closed | Start date: | 2019-05-22 |
| Priority: | High | Due date: | |
| Assignee: | Rob Nahf | % Done: | 100% |
| Category: | d1_cn_service | Estimated time: | 0.00 hour |
| Target version: | | Story Points: | |
| Milestone: | None | | |
| Product Version: | * | | |

Description

The CN resolve service doesn't currently support an HTTP HEAD request but probably should, as this can be used by clients to check if a URL is valid and to check the "Content-length" for potentially large objects. Currently a HEAD request for the resolve service returns an HTTP 500 status:

```
curl -v -I https://cn.dataone.org/cn/v2/resolve/urn%3Auuid%3A7098ba54-ca6f-4e35-beb3-718bd0fe58a8
```

I'm marking this issue as 'high' because this functionality is required by the DataONE FAIR suite.

Related issues:

Related to Infrastructure - Feature #8766: support server-side link checking ...

In Progress **2019-02-15**

History

#1 - 2019-05-22 19:59 - Rob Nahf

- Related to Feature #8766: support server-side link checking for the 303 redirect url in the resolve call added

#2 - 2019-05-22 20:03 - Rob Nahf

Update: The better implementation is to use 307 instead of 303 or 302. see <https://stackoverflow.com/a/2068485>

(Because of READ logging, we need to avoid creating GETs.)

The most straightforward implementation to support HEAD against the redirect is to return a 302 response instead of 303. (I'm not sure why we are using 303) This would let the user agent use the same method (HEAD in this case) in the request to the Location URL.

Here's a good explanation: <https://serverfault.com/questions/391181/examples-of-302-vs-303>

The related task is a bit tangential, but still related :-)

#3 - 2019-05-23 17:05 - Rob Nahf

When we add a handler for HEAD resolve, a 303 converts the HEAD into a GET, which will create a READ event in the logs of the MN or CN receiving the GET /object call.

For counter compliance, we need to avoid artificially increasing the number of READ events (that's also why we have getReplica).

Many client implementations convert 302s into 303s, so we should return a 307 instead, when we handle the HEAD resolve call. 307 is the 1.1 equivalent of 1.0 302 - in that it disallows the user agent from converting the HEAD to another method.

#4 - 2019-05-30 22:56 - Rob Nahf

testing a solution in DEV. For the resolve proxy, I needed to convert the HEAD request into a GET, so we could still get the systemMetadata from Metacat.

Confirming output:

```
EPSCoR-MBP-7BAC:d1_cn_index_processor rnahf$ curl -IL https://cn-dev-ucsb-1.test.dataone.org/cn/v2/resolve/urn
:uuid:5cd8fcba-2841-4688-971d-189ee56fb59b
HTTP/1.1 307 Temporary Redirect
Date: Thu, 30 May 2019 22:52:54 GMT
Server: Apache/2.4.7 (Ubuntu)
Vary: User-Agent
Set-Cookie: JSESSIONID=973892A50C7A73ECEDA3755573D38B05; Path=/cn/; Secure; HttpOnly
Location: https://test.arcticdata.io/metacat/d1/mn/v2/object/urn:uuid:5cd8fcba-2841-4688-971d-189ee56fb59b
Access-Control-Allow-Origin:
Access-Control-Allow-Credentials: true
Access-Control-Allow-Headers: Authorization, Content-Type, Location, Content-Length, x-annotator-auth-token
Access-Control-Expose-Headers: Content-Length, Content-Type, Location
Access-Control-Allow-Methods: POST, GET, OPTIONS, PUT, DELETE
Content-Type: text/xml;charset=UTF-8
```

```
HTTP/1.1 200 200
Date: Thu, 30 May 2019 22:52:55 GMT
Server: Apache/2.4.29 (Ubuntu)
X-Frame-Options: SAMEORIGIN
Vary: User-Agent
Set-Cookie: JSESSIONID=374C652FC1DBA5F93AE2AFEAB981C178; Path=/metacat; Secure
DataONE-Checksum: SHA-1,7181723277085f1dfd25a895a07ead830ac4e605
Last-Modified: Thu, 01 Jan 1970 00:00:00 GMT
DataONE-ObjectFormat: eml://ecoinformatics.org/eml-2.1.0
DataONE-SerialVersion: 0
Content-Length: 137097
X-Frame-Options: sameorigin
Access-Control-Allow-Origin:
Access-Control-Allow-Headers: Authorization
Access-Control-Allow-Credentials: true
Access-Control-Allow-Methods: GET, OPTIONS,PUT, POST
Content-Type: text/xml
```

#5 - 2019-05-31 15:18 - Rob Nahf

- % Done changed from 0 to 30
- Assignee set to Rob Nahf
- Status changed from New to In Progress
- Category set to d1_cn_service

#6 - 2019-06-11 22:13 - Rob Nahf

branching the solution in trunk cn/d1_cn_rest as D1_CN_REST_v2.4.

In trunk, these calls demonstrate the solution (v1 and v2 APIs):

```
curl -IL https://cn-dev-ucsb-1.test.dataone.org/cn/v1/resolve/urn:uuid:5cd8fcba-2841-4688-971d-189ee56fb59b
curl -IL https://cn-dev-ucsb-1.test.dataone.org/cn/v2/resolve/urn:uuid:5cd8fcba-2841-4688-971d-189ee56fb59b
```

#7 - 2019-06-13 18:15 - Rob Nahf

Tested D1_CN_REST_v2.4 in sandbox. Tests pass. curl seems to be slower than the Chrome browser. /v1, /v2 endpoints are similar in timing. resolve is much faster on ORC than UCSB (where the new solution is deployed)

#8 - 2019-06-18 05:52 - Rob Nahf

the new HEAD /resolve functionality has been ruled out as causing the reponse deley. it has been isolated to the upgrade of the openssl debian package from 1.1.0g to 1.1.1.

#9 - 2019-09-12 20:06 - Rob Nahf

- % Done changed from 30 to 100
- Status changed from In Progress to Closed

ready to deploy