

Arvados - Bug #5426

[Workbench] Large downloads through workbench fail

03/10/2015 01:23 PM - Peter Amstutz

Status:	Resolved	Start date:	03/18/2015
Priority:	Normal	Due date:	
Assigned To:		% Done:	100%
Category:	Workbench	Estimated time:	0.00 hour
Target version:	Deferred		
Description			
Right around 1 GiB, this download fails (notice that it fails at two different positions, but the same position the last 2 times...)			
<pre>--2015-03-10 09:03:52-- https://workbench.qr1hi.arvadosapi.com/collections/download/qr1hi-4zz18-b1uuzkf11kg3huv/3yfrrbhsh4t1qyr8catlfa5q8uy2m7wscuvdr4d485hqgy9u/lobstr_v3.0.2_hg19_ref/lobSTR_ref.fasta Reusing existing connection to workbench.qr1hi.arvadosapi.com:443. HTTP request sent, awaiting response... 200 OK Length: unspecified [application/octet-stream] Saving to: 'lobstr_v3.0.2_hg19_ref/lobSTR_ref.fasta' lobstr_v3.0.2_hg19_ [<=>] 1.01G 2.98MB/s in 6m 17s Last-modified header missing -- time-stamps turned off. 2015-03-10 09:10:10 (2.74 MB/s) - Read error at byte 1083196276 (The request is invalid.).Retrying . --2015-03-10 09:10:11-- (try: 2) https://workbench.qr1hi.arvadosapi.com/collections/download/qr1hi-4zz18-b1uuzkf11kg3huv/3yfrrbhsh4t1qyr8catlfa5q8uy2m7wscuvdr4d485hqgy9u/lobstr_v3.0.2_hg19_ref/lobSTR_ref.fasta Connecting to workbench.qr1hi.arvadosapi.com (workbench.qr1hi.arvadosapi.com) 54.88.31.97 :443... connected. HTTP request sent, awaiting response... 200 OK Length: unspecified [application/octet-stream] Saving to: 'lobstr_v3.0.2_hg19_ref/lobSTR_ref.fasta' lobstr_v3.0.2_hg19_ [<=>] 1.01G 3.05MB/s in 6m 34s 2015-03-10 09:16:47 (2.63 MB/s) - Read error at byte 1084238716 (The request is invalid.).Retrying . --2015-03-10 09:16:49-- (try: 3) https://workbench.qr1hi.arvadosapi.com/collections/download/qr1hi-4zz18-b1uuzkf11kg3huv/3yfrrbhsh4t1qyr8catlfa5q8uy2m7wscuvdr4d485hqgy9u/lobstr_v3.0.2_hg19_ref/lobSTR_ref.fasta Connecting to workbench.qr1hi.arvadosapi.com (workbench.qr1hi.arvadosapi.com) 54.88.31.97 :443... connected. HTTP request sent, awaiting response... 200 OK Length: unspecified [application/octet-stream] Saving to: 'lobstr_v3.0.2_hg19_ref/lobSTR_ref.fasta' lobstr_v3.0.2_hg19_ [<=>] 1.01G 1.98MB/s in 7m 59s 2015-03-10 09:24:50 (2.16 MB/s) - Read error at byte 1084238716 (The request is invalid.).Retrying .</pre>			
Subtasks:			
Task # 5503: Deploy "proxy_buffer off;" on /collections/download			Resolved
Related issues:			
Related to Arvados - Story #5824: [Workbench] [Keep] collection browse/downlo...			Resolved 05/21/2015

History

#1 - 03/10/2015 01:23 PM - Peter Amstutz

- Description updated

#2 - 03/10/2015 01:29 PM - Peter Amstutz

- Description updated

#3 - 03/11/2015 02:47 PM - Tom Clegg

- Subject changed from [Keep] Large downloads through workbench fail to [Workbench] Large downloads through workbench fail

- Category set to Workbench

Thoughts

- Could this be a proxy issue? (Try bypassing nginx and downloading from Workbench directly, from inside the firewall?)
- Anything in Workbench logs?
- Anything in nginx logs?
- Confirmed there's no problem retrieving the entire file with other tools?

#4 - 03/11/2015 08:22 PM - Ward Vandewege

- Category deleted (Workbench)

The bug appears to be in our code. Workbench does a fork (IO.popen) to call arv-get and streams the files. Nginx says in the logs:

```
2015/03/10 13:21:07 [error] 5544#0: *395704 upstream prematurely closed connection while reading upstream, client: 74.118.24.162, server: workbench.qr1hi.arvadosapi.com, request: "GET /collections/download/qr1hi-4zz18-bl1uuzkf11kg3huv/3yfrbhnsh4t1qyr8cat1fa5q8uy2m7wscuvdrm4d485hqqy9u/lobstr_v3.0.2_hg19_ref/lobSTR_ref.fasta HTTP/1.1", upstream: "http://127.0.0.1:9000/collections/download/qr1hi-4zz18-bl1uuzkf11kg3huv/3yfrbhnsh4t1qyr8cat1fa5q8uy2m7wscuvdrm4d485hqqy9u/lobstr_v3.0.2_hg19_ref/lobSTR_ref.fasta", host: "workbench.qr1hi.arvadosapi.com", referer: "https://workbench.qr1hi.arvadosapi.com/collections/download/qr1hi-4zz18-bl1uuzkf11kg3huv/3yfrbhnsh4t1qyr8cat1fa5q8uy2m7wscuvdrm4d485hqqy9u/"
```

There is nothing in the nginx error log for the process running on port 9000.

So, it looks like the IO.popen dies or arv-get dies, without logging anything in the webserver logs. This happens reliably at sizes just over 1 GiB.

#5 - 03/11/2015 08:22 PM - Ward Vandewege

- Category set to Workbench

#6 - 03/11/2015 08:44 PM - Peter Amstutz

- Target version changed from Bug Triage to 2015-04-01 sprint

#7 - 03/18/2015 05:52 PM - Peter Amstutz

- Assigned To set to Peter Amstutz

#8 - 03/18/2015 05:52 PM - Peter Amstutz

- Status changed from New to In Progress

#9 - 04/01/2015 07:04 PM - Ward Vandewege

- Target version changed from 2015-04-01 sprint to 2015-04-29 sprint

#10 - 04/01/2015 07:04 PM - Ward Vandewege

- Story points set to 0.5

#11 - 04/02/2015 05:01 PM - Ward Vandewege

- Status changed from In Progress to Resolved

#12 - 04/18/2015 12:46 AM - Ward Vandewege

- Status changed from Resolved to In Progress

- Target version changed from 2015-04-29 sprint to Bug Triage

I'm re-opening this bug.

The collection mentioned above downloads fine, now that we have proxy_buffering disabled. That's roughly 1.8 GiB.

However - this collection ([qr1hi-4zz18-w0t3gbd4u8n5o9h](#)) has a 16.4 fasta file in it, and.... it terminates download after roughly 1 GiB when downloaded through the browser. With arv keep get on a shell node, we get all 16.4 GiB without issues.

#13 - 04/20/2015 01:18 PM - Radhika Chippada

Actually, it appears that the original download listed in the ticket also does not download completely. It terminates after **1.08GB** and not all of the 1.8GB is downloaded.

https://workbench.qr1hi.arvadosapi.com/collections/d341a6f1db391a780d694e240e95e475+3805/lobstr_v3.0.2_hg19_ref/lobSTR_ref.fasta?disposition=attachment&size=1885053904

#14 - 08/11/2015 02:08 PM - Brett Smith

- *Assigned To deleted (Peter Amstutz)*

- *Target version changed from Bug Triage to Deferred*

I think we're very likely to deal with this via [#5824](#).

#15 - 05/18/2017 03:57 PM - Peter Amstutz

- *Status changed from In Progress to Resolved*

Now using keep-web and planning to remove workbench download entirely.