

D4Science Infrastructure - Incident #5638

Task # 5590 (Closed): DataMiner as Generic Worker

Check data availability through URI Resolver

Nov 03, 2016 02:45 PM - Gianpaolo Coro

Status:	Rejected	Start date:	Nov 03, 2016
Priority:	Normal	Due date:	
Assignee:	Roberto Cirillo	% Done:	100%
Category:	Data Management	Estimated time:	0.00 hour
Target version:	Data Processing		
Infrastructure:	Production		
Description			
Sequential access to this file http://goo.gl/FcnUc0 (FishBase taxonomic file) fails. Concurrent access seems to systematically fail. This prevents using the file in experiments and could be affecting the system performance.			

History

#1 - Nov 03, 2016 07:54 PM - Gianpaolo Coro

I have done some stress tests, since the link was working again this afternoon.

On the access-d.d4science.org machine, I run these benchmark processes few times on both the short Url and the long url:

```
ab -n 1000 -c 500 "http://data.d4science.org/smp?fileName=FISHBASE_taxa.taf.gz&contentType=application%2Fxml&smp-uri=smp%3A%2F%2FShare%2F89971b8f-a993-4e7b-9a95-8d774cb68a99%2FWork+Packages%2FWP+6+-+Virtual+Research+Environments+Deployment+and+Operation%2FT6.2+Resources+and+Tools%2FCOMET-Species-Matching-Engine%2FYASMEEN%2F1.2.0%2FData%2FBiOnymTAF%2FFISHBASE_taxa.taf.gz%3F5ezVfBOLqb3YESyI%2FkesN4T%2BZD0mtmc%2F4sZ0vGMrl0lgx7k85j8o2Q1vF0ezJi%2FxiGDhncO9jOkV1T8u6Db7GZ%2F4ePgMws8Jxu8ierJajHBd20bUotElN0kyA%2Bs3HQuMVYbva9MKgw1wahC7aUCyaItSZIQuKpu4pSjoDP8iox%2FX02bqsokgB5v1H%2FQUQgN"
```

```
ab -n 1000 -c 500 "http://goo.gl/FcnUc0"
```

and several wgets on the same files to check availability. After a few attempts (the benchmarks were always successful) the server began to be non-responding at all to the wgets. After a while, the file was available again (the server returned to be responding).

#2 - Nov 04, 2016 09:40 AM - Roberto Cirillo

- Status changed from New to Rejected

At this time we are not able to handle 1000 requests, with 500 requests running concurrently for the same file.

Is there a real case that need 500 concurrent access to the same file?

If the answer is yes (please specify the case), we need to convert our system in a sharding system with a horizontal scalability.

#3 - Nov 04, 2016 09:58 AM - Gianpaolo Coro

- Status changed from Rejected to In Progress

ab simulates high traffic but does not download the file. You can test also with less calls and concurrency. I wanted to demonstrate that under certain conditions the uri resolver does not respond.

The concrete case is that concurrent calls from the DataMiners often fail. We need to understand which is this maximum degree of concurrency we support.

If we do not support medium concurrency, then people are justified to use their own services to store and publish data.

Currently, a bug in the ic-client prevents to run tests with DataMiner, but I think the issue reported in this ticket is crucial and we cannot ignore it. We cannot use DataMiner to test the other services in the e-Infrastructure.

#4 - Nov 04, 2016 10:35 AM - Gianpaolo Coro

- Status changed from In Progress to Rejected

Since this issue requires further investigation and is possibly related to other issues than concurrency, I close it and a new one will be opened.

#5 - Nov 04, 2016 11:13 AM - Pasquale Pagano

I made same tests with three types of files.

The first file is 15 Mb and I made 1000 attempts with concurrency set to 50.

```
mb-pagano:~ pasqualepagano$ ab -n 1000 -c 50 http://data.d4science.org/dk9oekp4b1ZFajc5Z1ZXyXlUOUtMaUgzYUJOWXE5eDdHbWJQNStIS0N6Yz0
This is ApacheBench, Version 2.3 <$Revision: 1748469 $>
Copyright 1996 Adam Twiss, Zeus Technology Ltd, http://www.zeustech.net/
Licensed to The Apache Software Foundation, http://www.apache.org/
```

Benchmarking data.d4science.org (be patient)

Finished 1000 requests

```
Server Software:      Apache-Coyote/1.1
Server Hostname:      data.d4science.org
Server Port:          80

Document Path:        /dk9oekp4b1ZFajc5Z1ZXyXlUOUtMaUgzYUJOWXE5eDdHbWJQNStIS0N6Yz0
Document Length:      14229041 bytes

Concurrency Level:     50
Time taken for tests:   226.883 seconds
Complete requests:     1000
Failed requests:        0
Total transferred:     14229343000 bytes
HTML transferred:      14229041000 bytes
Requests per second:   4.41 [#/sec] (mean)
Time per request:      11344.144 [ms] (mean)
Time per request:      226.883 [ms] (mean, across all concurrent requests)
Transfer rate:         61246.77 [Kbytes/sec] received
```

```
Connection Times (ms)
              min  mean[+/-sd] median   max
Connect:        0    2    6.4      2    142
Processing:  4993 10770 2789.5   10396  27958
Waiting:        1354 3824 2176.4    3244  21097
Total:          4996 10772 2789.5   10397  27960
```

Percentage of the requests served within a certain time (ms)

```
 50%   10397
 66%   11657
 75%   12407
 80%   12814
 90%   14275
 95%   15271
 98%   17536
 99%   19947
100%   27960 (longest request)
```

Note that the amount of data transferred is 14229343000 bytes that means 13 GB; 100% success; 4.41 [#/sec](#)

The second test I did is with a smaller file, 500 kb.
I made 1000 attempts with concurrency set to 50.

```
mb-pagano:~ pasqualepagano$ ab -n 1000 -c 50 http://data.d4science.org/eWlXRlgvM05iZFRWUWhEWktTeVNVdWdTWGF0VTRICvJHbWJQNStIS0N6Yz0
This is ApacheBench, Version 2.3 <$Revision: 1748469 $>
Copyright 1996 Adam Twiss, Zeus Technology Ltd, http://www.zeustech.net/
Licensed to The Apache Software Foundation, http://www.apache.org/
```

Benchmarking data.d4science.org (be patient)

Finished 1000 requests

```
Server Software:      Apache-Coyote/1.1
Server Hostname:      data.d4science.org
Server Port:          80

Document Path:        /eWlXRlgvM05iZFRWUWhEWktTeVNVdWdTWGF0VTRICvJHbWJQNStIS0N6Yz0
Document Length:      574815 bytes

Concurrency Level:     50
Time taken for tests:   119.583 seconds
Complete requests:     1000
Failed requests:        0
Total transferred:     575046000 bytes
HTML transferred:      574815000 bytes
```

```
Requests per second:      8.36 [#/sec] (mean)
Time per request:         5979.173 [ms] (mean)
Time per request:         119.583 [ms] (mean, across all concurrent requests)
Transfer rate:            4696.04 [Kbytes/sec] received
```

```
Connection Times (ms)
              min    mean[+/-sd] median    max
Connect:        0      1    5.7        0    151
Processing:    2401  5863 1456.0    5708  10291
Waiting:       2382  5777 1437.1    5606  10279
Total:         2401  5865 1456.0    5709  10292
```

```
Percentage of the requests served within a certain time (ms)
 50%    5709
 66%    6329
 75%    6783
 80%    7062
 90%    7937
 95%    8677
 98%    9192
 99%    9405
100%   10292 (longest request)
```

Even in this case the percentage of success is 100%. The total transfer is 0.5 GB; the medium number of request per second are 8.36 [#/sec].

Finally, the third attempt was done with an even smaller file, 150 Kb, but with 100 concurrency accesses.

```
mb-pagano:~ pasqualepagano$ ab -n 1000 -c 100 http://data.d4science.org/LzhNd1h4c2VuVEo4YW5oVVRHbTBpcWhkeDhTUWRDeWxHbWJQNStIS0N6Yz0
This is ApacheBench, Version 2.3 <$Revision: 1748469 $>
Copyright 1996 Adam Twiss, Zeus Technology Ltd, http://www.zeustech.net/
Licensed to The Apache Software Foundation, http://www.apache.org/
```

```
Benchmarking data.d4science.org (be patient)
Finished 1000 requests
```

```
Server Software:      Apache-Coyote/1.1
Server Hostname:      data.d4science.org
Server Port:          80
```

```
Document Path:        /LzhNd1h4c2VuVEo4YW5oVVRHbTBpcWhkeDhTUWRDeWxHbWJQNStIS0N6Yz0
Document Length:      147082 bytes
```

```
Concurrency Level:     100
Time taken for tests:   116.561 seconds
Complete requests:     1000
Failed requests:        0
Total transferred:     147360000 bytes
HTML transferred:      147082000 bytes
Requests per second:   8.58 [#/sec] (mean)
Time per request:      11656.074 [ms] (mean)
Time per request:      116.561 [ms] (mean, across all concurrent requests)
Transfer rate:         1234.60 [Kbytes/sec] received
```

```
Connection Times (ms)
              min    mean[+/-sd] median    max
Connect:        0      2    9.1        0    147
Processing:    4516  11298 2868.0   11350  18630
Waiting:       4513  11168 2828.1   11204  18626
Total:         4517  11301 2867.6   11352  18633
```

```
Percentage of the requests served within a certain time (ms)
 50%    11352
 66%    12708
 75%    13399
 80%    13864
 90%    14958
 95%    16149
 98%    16927
 99%    17364
100%   18633 (longest request)
```

Again 100% success with 8.58 requests served per second.

It seems to me that overall the system works well. We need more details on the average volume to transfer and an average number of reasonable concurrent accesses to serve.

#6 - Nov 04, 2016 11:32 AM - Roberto Cirillo

I've created another ticket for the specific problem with the old smp uri reported by @gianpaolo.coro@isti.cnr.it : [#5646](#)