Docker data transfer test at TRIUMF

Oct. 03 2016, by Xinli(Simon) Liu

Test environment:

We used two work nodes for this testing. One from IBM chassis, the other from SUN chassis, not meant to compare them, but to have two sets of testing beds.

wn393(node from IBM chassis):

```
Hardware: 2 sockets, 6 cores/socket, HT disabled, 48GB memory, 2*10k SAS drives,
LVM RAID0.
OS: SL7.2, 3.10.0-327.28.3.el7.x86_64
Docker Source RPM  : docker-1.10.3-44.el7.centos.src.rpm
```

wn242(node from SUN chassis):

```
Hardware: 2 sockets, 4 cores/socket, HT disabled, 24GB memory, 2*10k SAS drives,
LVM RAID0.
OS: SL7.2, 3.10.0-327.28.3.el7.x86_64
Docker Source RPM  : docker-1.10.3-44.el7.centos.src.rpm
```

Test method:

The test was meant to measure the data transfer rate from SE to host node and docker node via different protocols.

The following popular protocols were tested, the second column are the tools used for testing

```
gsiftp    gfal-copy
dcap      gfal-copy and dccp
http      gfal-copy and curl
root      gfal-copy and xrdcp
```

No tuning options set to gfal-copy, curl and xrdcp. Dccp uses tuned transfer buffers(TRIUMF production env tuning, however this is not the case on wn393 and its docker node)

Tested on 2 local nodes, 2 docker nodes for 4 protocols to fetch a 1.24GB file from production SE:

1. Two local nodes run the test in parallel, read the same file via 4 protocols sequentially and repeat for 100 times.
2. start two docker nodes and run the same test

Note:

1. The test file is a compressed file used to be production tape file, pinned to 5 different dcache pool nodes(10G net) and on different storage units.
2. The test ran along with production activities, so there could be some impact from production. However, with 100 times 4 protocols sequential copy, the impact to different nodes/protocols should be fairly same. Plus production SE activities were quiet during the testing period.
3. To minimize overhead, all protocols directly contact door nodes for data transfer, though 3 protocols still need authentication check. Single stream transfer.
4. 3 round testings were done, no major different found between client tools(except dcap), so in the following report, only protocol and nodes are addressed.

Testing surl files

```
gurl="gsiftp://dpool15.triumf.ca:2811//atlas/atlasscratchdisk/simon1_24_GiBtestfil
e"
durl="dcap://dpool15.triumf.ca:22125/pnfs/triumf.ca/data/atlas/atlasscratchdisk/si
mon1_24_GiBtestfile"
```

```
xurl="root://dpool15.triumf.ca:1094/atlas/atlasscratchdisk/simon1_24_GiBtestfile"
hurl="https://dpool15.triumf.ca:2880/atlas/atlasscratchdisk/simon1_24_GiBtestfile"
```

Command examples:

```
gfal-copy $gurl file://$filename

curl --cacert /tmp/x509up_u501 --capath /etc/grid-security/certificates/ --cert
/tmp/x509up_u501 -o $filename -L $hurl

/bin/dccp -X -io-queue=regular -a -b 10485760 -B 1048576 -r 1048576 -s 1048576
$durl $filename

xrdcp -f $xurl $filename
```

Testing matrix

Here is a table matrix shows data transfer rate on different nodes, using different protocols in MiB/sec

	Gsiftp	Dcap	Http	root
wn393	122.23	107.13	122.29	122.72
vn31073(docker node on wn393)	123.00	107.61	122.42	122.38
wn242	50.92	104.96 45.85	60.04	45.15
vn2033(docker node on wn242)	60.78	105.51 48.07	49.45	51.03

Red: using gfal-copy
blue: using dccp with tuned parameters
Green: individual test in any testing set shows vary large range of result, For example: on wn242, observed xrootd transfer rate ranged 32Mib/sec to 97Mib/sec. Average speed of each set of test result also not consistent, range in 10%-30% different.

	dccp with tuned parameters	dccp
Wn242	104.96	39.04
vn2033(docker node on wn242)	105.51	41.48

Summary

1. On wn393, no transfer efficiency different observed between host node and docker node on different protocols tested.

2. On wn242, both host and docker nodes show much less performance than wn393. It has less memory (24GB), but this should not be the performance bottle neck. No consistent result found in both individual and set test by average(100 times). Further looking on chassis/system/networking tuning is suggested.

3. No data transfer penalty observed on docker node.

Followup testing

To understand why the data transfer performance is lower on wn242 and its docker node. Di and I did

some testings. No docker involved.

1. Did some testing on two other work nodes from old IBM and SUN chassis, we observed the same pool performance as wn242's

2. Added 16GB memory to wn242, made it to 40GB, we see the same pool performance as wn242's.

Note: the work node from old IBM chassis is the same type of blade that wn393 is. Just seating in the older generation chassis.

Overall, poor data transfer performance seems not related with individual nodes, kernel, memory etc.. We suspect it's related with old chassis architecture(or other aspects we haven't touched yet). Further testing may needed but it's definitely out of scope of this docker testing.