Subject: Re: NFS-write blocks all processes
Posted by Bernd_K on Mon, 20 Oct 2014 07:11:38 GMT

View Forum Message <> Reply to Message

Hi Thorsten.

Thank you very much for your reply.

I am user only, not a sysman, so some of the answers may not be exactly what you are expecting, but I will do my very best

----

- more Info about this setup? - kernel version?

It is a seismic processing environment, 13 nodes dual Xeon E5-2667 v2, Linux version 2.6.18-371.el5 .

- any hardware problems?

No hardware problems as such, just a constant fight against runtimes, I/O bottlenecks...
Everything is too slow and too small.
This is why the gaps in NFS-read (during NFS-write) really hurt.

My hope was that tweaking some parameters would enable the system to simultaneously read and write from/to disk.

- which kind of RAID Controller and Level ist used? - how many Disk / Spindles are used?

22 Disks, 40TB total, RAID controller LS8271-8I, Level 5 and 6 (says sysman)
The above is the "bulk-data" storage, the one which has critical impact on runtime if it does not constantly read/write as much as possible.

There is another RAID of similar size, holding "database" and other stuff, which normally is not constantly active and has no or small throughput during runtime.

Both RAIDs have separate IP-addresses.

- post the outpot from nfsstat?

from RAID:

-sh-4.1$ nfsstat -o all
Server packet stats:
packets   udp   tcp   tcpconn
3374939671 494   3374660041 15960

Server rpc stats:
calls   badcalls badclnt   badauth   xdrcall
3375382591 0   0   0   0

Server reply cache:
hits   misses   nocache
0   91841868 3282784906

Server file handle cache:
lookup   anon   ncachedir ncachedir stale
0   0   0   0   417

Server nfs v2:
null   getattr   setattr   root   lookup   readlink
55   11% 178   36% 0   0% 0   0% 15   3% 0   0%
read   wrcache   write   create   remove   rename
0   0%   0   0% 0   0% 0   0% 0   0% 0   0%
link   symlink   mkdir   rmdir   readdir   fsstat
0   0%   0   0% 0   0% 0   0% 39   7% 207   41%

Server nfs v3:
null   getattr   setattr   lookup   access   readlink
205   0%   148779700 4% 128541   0% 120366   0% 1448436 0% 0   0%
read   write   create   mkdir   symlink   mknod
3082738953 92% 63675228 1% 29584 0% 4483   0% 0   0% 0   0%
remove   rmdir   rename   link   readdir   readdirplus
24735   0% 3742   0% 166   0% 0   0% 259   0% 238980 0%
fsstat   fsinfo   pathconf   commit
36202899 1% 192   0% 2   0% 10059450 0%

Server nfs v4:
null   compound
66   0%   31900709 99%

Server nfs v4 operations:
op0-unused op1-unused op2-future access   close   commit

```
0       0% 0       0% 0       0% 325021  0% 257650  0% 321645  0%
create     delegpurge  delegreturn getattr     getfh       link
58040   0% 0       0% 3622    0% 29849717 32% 317903  0% 0       0%
lock       lockt       locku       lookup      lookup_root nverify
0       0% 0       0% 0       0% 5092    0% 0       0% 0       0%
open       openattr    open_conf   open_dgrd   putfh       putpubfh
258028  0% 0       0% 74      0% 1       0% 31287868 33% 0       0%
putrootfh  read        readdir     readlink    remove      rename
128     0% 786775  0% 101225  0% 0       0% 194679  0% 0       0%
renew      restorefh   savefh      secinfo     setattr     setcltid
612641  0% 0       0% 0       0% 2       0% 967416  1% 65      0%
setcltidconf verify    write       rellockowner bc_ctl     bind_conn
65      0% 0       0% 26753894 29% 0       0% 0       0% 0       0%
exchange_id create_ses  destroy_ses free_stateid getdirdeleg getdevinfo
0       0% 0       0% 0       0% 0       0% 0       0% 0       0%
getdevlist layoutcommit layoutget  layoutreturn secinfononam sequence
0       0% 0       0% 0       0% 0       0% 0       0% 0       0%
set_ssv    test_stateid want_deleg  destroy_clid reclaim_comp
0       0% 0       0% 0       0% 0       0% 0       0%
```

Client packet stats:
```
packets    udp        tcp        tcpconn
0          0          0          0
```

Client rpc stats:
```
calls      retrans    authrefrsh
565223     89         565223
```

Client nfs v4:
```
null       read        write       commit      open        open_conf
0       0% 953     0% 1       0% 0       0% 2       0% 0       0%
open_noat  open_dgrd   close       setattr     fsinfo      renew
0       0% 0       0% 2       0% 1       0% 6       0% 564196  99%
setclntid  confirm     lock        lockt       locku       access
1       0% 1       0% 0       0% 0       0% 0       0% 11      0%
getattr    lookup      lookup_root remove      rename      link
12      0% 9       0% 2       0% 0       0% 0       0% 0       0%
symlink    create      pathconf    statfs      readlink    readdir
0       0% 0       0% 4       0% 8       0% 0       0% 3       0%
server_caps delegreturn getacl     setacl      fs_locations rel_lkowner
10      0% 0       0% 0       0% 0       0% 0       0% 0       0%
secinfo    exchange_id create_ses  destroy_ses sequence    get_lease_t
0       0% 0       0% 0       0% 0       0% 0       0% 0       0%
reclaim_comp layoutget getdevinfo  layoutcommit layoutreturn getdevlist
0       0% 0       0% 0       0% 0       0% 0       0% 0       0%
(null)
0       0%
```

From one of the nodes:

-bash-3.2$ nfsstat nfs -o all
Server packet stats:
packets   udp   tcp   tcpconn
0   0   0   0

Server rpc stats:
calls   badcalls badclnt   badauth   xdrcall
0   0   0   0   0

Server reply cache:
hits   misses   nocache
0   0   0

Server file handle cache:
lookup   anon   ncachedir ncachedir stale
0   0   0   0   0

Client packet stats:
packets   udp   tcp   tcpconn
0   0   0   0

Client rpc stats:
calls   retrans   authrefrsh
1495170660 33   0

Client nfs v3:
null   getattr   setattr   lookup   access   readlink
0   0% 228264410 15% 511297   0% 493031   0% 5084816 0% 11463   0%
read   write   create   mkdir   symlink   mknod
804422516 53% 333380788 22% 81229   0% 8923   0% 260   0% 0   0%
remove   rmdir   rename   link   readdir   readdirplus
83241   0% 9101   0% 21029   0% 121   0% 3809   0% 313854   0%
fsstat   fsinfo   pathconf   commit
21495660 1% 10   0% 0   0% 100985097 6%

Thanks and cheers,
Bernd