TROUBLESHOOTING

SLOW/STUCK OPERATIONS

If you are experiencing apparent hung operations, the first task is to identify where the problem is occurring: in the client, the MDS, or the network connecting them. Start by looking to see if either side has stuck operations (Slow requests (MDS), below), and narrow it down from there.

RADOS HEALTH

If part of the CephFS metadata or data pools is unavaible and CephFS is not responding, it is probably because RADOS itself is unhealthy. Resolve those problems first (Troubleshooting).

THE MDS

If an operation is hung inside the MDS, it will eventually show up in ceph `health`, identifying “slow requests are blocked”. It may also identify clients as “failing to respond” or misbehaving in other ways. If the MDS identifies specific clients as misbehaving, you should investigate why they are doing so. Generally it will be the result of 1) overloading the system (if you have extra RAM, increase the “mds cache size” config from its default 100000; having a larger active file set than your MDS cache is the #1 cause of this!) 2) running an older (misbehaving) client, or 3) underlying RADOS issues.

Otherwise, you have probably discovered a new bug and should report it to the developers!

SLOW REQUESTS (MDS)

You can list current operations via the admin socket by running:

```
ceph daemon mds.<name> dump_ops_in_flight
```

from the MDS host. Identify the stuck commands and examine why they are stuck. Usually the last “event” will have been an attempt to gather locks, or sending the operation off to the MDS log. If it is waiting on the OSDs, fix them. If operations are stuck on a specific inode, you probably have a client holding caps which prevent others from using it, either because the client is trying to flush out dirty data or because you have encountered a bug in CephFS’ distributed file lock code (the file “capabilities” [“caps”] system).

If it’s a result of a bug in the capabilities code, restarting the MDS is likely to resolve the problem.

If there are no slow requests reported on the MDS, and it is not reporting that clients are misbehaving, either the client has a problem or its requests are not reaching the MDS.

CEPH-FUSE DEBUGGING

ceph-fuse also supports dump_ops_in_flight. See if it has any and where they are stuck.

DEBUG OUTPUT

To get more debugging information from ceph-fuse, try running in the foreground with logging to the console (`-d`) and enabling client debug (`--debug-client=20`), enabling prints for each message sent (`--debug-ms=1`).

If you suspect a potential monitor issue, enable monitor debugging as well (`--debug-monc=20`).

KERNEL MOUNT DEBUGGING

SLOW REQUESTS

Unfortunately the kernel client does not support the admin socket, but it has similar (if limited) interfaces if your kernel has debugfs enabled. There will be a folder in `sys/kernel/debug/ceph/`, and that folder (whose name will look something like `28f7427e-5558-4ffd-ae1a-51ec3042759a.client25386880`) will contain a variety of files that output interesting output when you `cat` them. These files are described below; the most interesting when debugging slow requests are probably the `mdsc` and `osdc` files.

- bdi: BDI info about the Ceph system (blocks dirtied, written, etc)
- caps: counts of file "caps" structures in-memory and used
- client_options: dumps the options provided to the CephFS mount
- dentry_lru: Dumps the CephFS dentries currently in-memory
- mdsc: Dumps current requests to the MDS
- mdsmap: Dumps the current MDSMap epoch and MDSes
- mds_sessions: Dumps the current sessions to MDSes
- monc: Dumps the current maps from the monitor, and any "subscriptions" held
- monmap: Dumps the current monitor map epoch and monitors
- osdc: Dumps the current ops in-flight to OSDs (ie, file data IO)
- osdmap: Dumps the current OSDMap epoch, pools, and OSDs

If there are no stuck requests but you have file IO which is not progressing, you might have a...

DISCONNECTED+REMOUNTED FS

Because CephFS has a "consistent cache", if your network connection is disrupted for a long enough time, the client will be forcibly disconnected from the system. At this point, the kernel client is in a bind: it cannot safely write back dirty data, and many applications do not handle IO errors correctly on close(). At the moment, the kernel client will remount the FS, but outstanding filesystem IO may or may not be satisfied. In these cases, you may need to reboot your client system.

You can identify you are in this situation if dmesg/kern.log report something like:

```
Jul 20 08:14:38 teuthology kernel: [3677601.123718] ceph: mds0 closed our session
Jul 20 08:14:38 teuthology kernel: [3677601.128019] ceph: mds0 reconnect start
Jul 20 08:14:39 teuthology kernel: [3677602.093378] ceph: mds0 reconnect denied
Jul 20 08:14:39 teuthology kernel: [3677602.098525] ceph:  dropping dirty+flushing Fw state f
Jul 20 08:14:39 teuthology kernel: [3677602.107145] ceph:  dropping dirty+flushing Fw state f
Jul 20 08:14:39 teuthology kernel: [3677602.196747] libceph: mds0 172.21.5.114:6812 socket cl
Jul 20 08:14:40 teuthology kernel: [3677603.126214] libceph: mds0 172.21.5.114:6812 connectio
Jul 20 08:14:40 teuthology kernel: [3677603.132176] libceph: reset on mds0
```

This is an area of ongoing work to improve the behavior. Kernels will soon be reliably issuing error codes to in-progress IO, although your application(s) may not deal with them well. In the longer-term, we hope to allow reconnect and reclaim of data in cases where it won't violate POSIX semantics (generally, data which hasn't been accessed or modified by other clients).

MOUNTING

MOUNT 5 ERROR

A mount 5 error typically occurs if a MDS server is laggy or if it crashed. Ensure at least one MDS is up and running, and the cluster is `active + healthy`.

MOUNT 12 ERROR

A mount 12 error with `cannot allocate memory` usually occurs if you have a version mismatch between the Ceph Client version and the Ceph Storage Cluster version. Check the versions using:

```
ceph -v
```

If the Ceph Client is behind the Ceph cluster, try to upgrade it:

```
sudo apt-get update && sudo apt-get install ceph-common
```

You may need to uninstall, autoclean and autoremove ceph-common and then reinstall it so that you have the latest version.