UNDERSTANDING MDS CACHE SIZE LIMITS

This section describes ways to limit MDS cache size.

You can limit the size of the Metadata Server (MDS) cache by:

- *A memory limit*: A new behavior introduced in the Luminous release. Use the *mds_cache_memory_limit* parameters. We recommend to use memory limits instead of inode count limits.
- *Inode count*: Use the *mds_cache_size* parameter. By default, limiting the MDS cache by inode count is disabled.

In addition, you can specify a cache reservation by using the *mds_cache_reservation* parameter for MDS operations. The cache reservation is limited as a percentage of the memory or inode limit and is set to 5% by default. The intent of this parameter is to have the MDS maintain an extra reserve of memory for its cache for new metadata operations to use. As a consequence, the MDS should in general operate below its memory limit because it will recall old state from clients in order to drop unused metadata in its cache.

The *mds_cache_reservation* parameter replaces the *mds_health_cache_threshold* in all situations except when MDS nodes sends a health alert to the Monitors indicating the cache is too large. By default, *mds_health_cache_threshold* is 150% of the maximum cache size.

Be aware that the cache limit is not a hard limit. Potential bugs in the CephFS client or MDS or misbehaving applications might cause the MDS to exceed its cache size. The *mds_health_cache_threshold* configures the cluster health warning message so that operators can investigate why the MDS cannot shrink its cache.

> **Warning:** The memory tracking used is currently imprecise by a constant factor. This will be addressed in http://tracker.ceph.com/issues/22599. MDS deployments with large *mds_cache_memory_limit* (64GB+) should underallocate RAM to accomodate.