CephFS - Bug #38835
MDSTableServer.cc: 83: FAILED assert(version == tid)

03/21/2019 09:42 AM - Dan van der Ster

Status: Resolved
Priority: High
Assignee: Zheng Yan
Category:
Target version: v15.0.0
Source: Community (user)
Tags: ceph-qa-suite
Backport: nautilus,mimic
Regression: No
Severity: 3 - minor
Reviewed:
Affected Versions: v13.2.5

Description
We just hit this on a v13.2.5 cluster with 1 active MDS:

```
-5> 2019-03-21 10:15:33.943 7fd43241a700 10 monclient: _send_mon_message to mon.cephkelly-mon-39bee08afe at 137.138.64.13:6789/0
-4> 2019-03-21 10:15:35.544 7fd43441e700 10 monclient: tick
-3> 2019-03-21 10:15:35.544 7fd43441e700 10 monclient: _check_auth_rotating have updtodate secrets (they expire after 2019-03-21 10:15:05.544674)
-2> 2019-03-21 10:15:35.856 7fd432c1b700 2 mds.0.cache check_memory_usage total 455120, rss 28648, heap 323804, baseline 32804, buffers 0, 16 / 28 inodes have caps, 35 caps, 1.25 caps per inode
-1> 2019-03-21 10:15:37.943 7fd43241a700 10 monclient: _send_mon_message to mon.cephkelly-mon-39bee08afe at 137.138.64.13:6789/0
  0> 2019-03-21 10:15:40.087 7fd42f414700 -1 /home/jenkins-build/build/workspace/ceph-build/ARCH/x86_64/AVAILABLE_ARCH/x86_64/AVAILABLE_DIST/centos7/DIST/centos7/MACHINE_SIZE/huge/release/13.2.5/rpm/el7/BUILD/ceph-13.2.5/src/mds/MDSTableServer.cc: In function 'void MDSTableServer::_prepare_logged(MMDSTableRequest*, version_t)' thread 7fd42f414700 time 2019-03-21 10:15:40.085937
/home/jenkins-build/build/workspace/ceph-build/ARCH/x86_64/AVAILABLE_ARCH/x86_64/AVAILABLE_DIST/centos7/DIST/centos7/MACHINE_SIZE/huge/release/13.2.5/rpm/el7/BUILD/ceph-13.2.5/src/mds/MDSTableServer.cc: 83: FAILED assert(version == tid)
```

ceph version 13.2.5 (cbff874f9007f1869bf3821b7e33b2a6ff4988) mimic (stable)
1: (ceph::ceph_assert_fail(char const*, char const*, int, char const*)+0xff) [0x7fd43ecddf8f]
2: (0x226d187) [0x7fd43ecde187]
3: (MDSTableServer::_prepare_logged(MMDSTableRequest*, unsigned long)+0x5d1) [0x5573985bf791]
4: (MDSIOContextBase::complete(int)+0x119) [0x5573985ec3d9]
5: (MDSLogContextBase::complete(int)+0x40) [0x5573985ec560]
6: (Finisher::finisher_thread_entry()+0x12e) [0x7fd43ecdc53e]
7: (0x7fd43c91edd5) [0x7fd43c91edd5]
8: (clone()+0x6f) [0x7fd43b9f9bea]

NOTE: a copy of the executable, or `objdump -rdS <executable>` is needed to interpret this.

The coredump is available at ceph-post-file: aea2e0f9-ed81-4b45-ab90-e084e0ca184e

Related issues:
History

#1 - 03/21/2019 07:02 PM - Patrick Donnelly
- Priority changed from Normal to High
- Target version set to v15.0.0
- Start date deleted (03/21/2019)
- Backport set to nautilus,mimic,luminous
- Labels (FS) crash added

#2 - 03/26/2019 03:07 AM - Zheng Yan
I have trouble to check the coredump file. please use gdb to print 'tid' and 'version' if you can.

#3 - 03/26/2019 11:42 AM - Dan van der Ster
Zheng Yan wrote:

> I have trouble to check the coredump file. please use gdb to print 'tid' and 'version' if you can.

```
(gdb) p tid
$1 = <optimized out>
(gdb) p version
$2 = 2
```

Another way to get the tid?

#4 - 03/27/2019 08:25 AM - Zheng Yan
goto frame 4 and try casting MDSIOContextBase into C_Prepare. tid is stored in C_Prepare

was the cluster upgraded from luminous? was it the first time that mimic version mds ran?

If both are yes, following change can explain the asserion

```
diff --git a/src/mds/SnapServer.h b/src/mds/SnapServer.h
index 0fee9db9a..a3b430f1d3 100644
--- a/src/mds/SnapServer.h
+++ b/src/mds/SnapServer.h
@@ -105,6 +105,7 @@ public:
 if (get_version() == 0) {
  // version 0 confuses snapclient code
  reset_state();
+  projected_version = version;
 upgraded = true;
 }
```
if (snaprealm_v2_since == CEPH_NOSNAP) {

#5 - 03/27/2019 10:23 AM - Dan van der Ster
Yes this was upgraded from luminous and yes that was the first time the MDS was running in mimic.

(gdb) up
#9 0x00005573985ec3d9 in complete (r=0, this=0x55739a114b40) at /usr/src/debug/ceph-13.2.5/src/include/Context.h:77
77    finish(r);
(gdb) list
72    public:
73    Context() {}
74    virtual ~Context() {} // we want a virtual destructor!!
75    virtual void complete(int r) {
76        finish(r);
77        delete this;
78    }
79    virtual bool sync_complete(int r) {
80        if (sync_finish(r)) {
(gdb) p ((C_Prepare *)this)->tid
$3 = 1

#6 - 03/27/2019 01:09 PM - Zheng Yan
after restarting the mds, the crash will not happen again.

#7 - 03/28/2019 10:06 AM - Zheng Yan
- Backport changed from nautilus,mimic,luminous to nautilus,mimic

https://github.com/ceph/ceph/pull/27238

#8 - 03/28/2019 12:46 PM - Nathan Cutler
- Status changed from New to Fix Under Review

#9 - 03/28/2019 12:46 PM - Nathan Cutler
- Pull request ID set to 27238

#10 - 04/04/2019 07:40 PM - Patrick Donnelly
- Status changed from Fix Under Review to Pending Backport
- Assignee set to Zheng Yan
While running with --resolve-parent, the script "backport-create-issue" noticed that all backports of this issue are in status "Resolved" or "Rejected".