CephFS - Bug #19306

fs: mount NFS to cephfs, and then ls a directory containing a large number of files, resulting in ls hang.

03/19/2017 08:25 AM - geng jichao

Status:	Resolved	**% Done:**	0%
Priority:	Normal		
Assignee:	Zheng Yan		
Category:			
Target version:			
Source:		**ceph-qa-suite:**	
Tags:		**Component(FS):**	
Backport:		**Labels (FS):**	
Regression:	No	**Pull request ID:**	
Severity:	3 - minor	**Crash signature (v1):**	
Reviewed:		**Crash signature (v2):**	
Affected Versions:			

Description

The ceph_readdir function save lot of date in the file->private_date, include the last_name which uses as offset.However, in the nfs or cifs system file, when read a directory, they will open and close the directory many times, because the contents of the directory cannot be read once, this lead to last_name be null, and start reading from the beginning every time.Finaly, the time complexity of readdir is O(n^2), the n is file nums/max_readdir.

the nfs readdir code is at fs/nfsd/vfs.c/nfsd_readdir.

the kernel version is 4.4.0-46.

History

#1 - 03/19/2017 09:07 AM - Nathan Cutler

- Tracker changed from Tasks to Support

- Project changed from Stable releases to CephFS

- Target version deleted (v10.2.7)

- Release set to jewel

- Affected Versions v10.2.6 added

#2 - 03/21/2017 03:18 PM - John Spray

- Tracker changed from Support to Bug

- Project changed from CephFS to Linux kernel client

- Subject changed from mount NFS to cephfs, and then ls a directory containing a large number of files, resulting in ls hang. to kcephfs: mount NFS to cephfs, and then ls a directory containing a large number of files, resulting in ls hang.

- Regression set to No

- Severity set to 3 - minor

- Release deleted (jewel)

- Affected Versions deleted (v10.2.6)

#3 - 03/22/2017 03:38 AM - Zheng Yan

- Project changed from Linux kernel client to CephFS

- Subject changed from kcephfs: mount NFS to cephfs, and then ls a directory containing a large number of files, resulting in ls hang. to fs: mount NFS to cephfs, and then ls a directory containing a large number of files, resulting in ls hang.

This bug is not specific to kernel client. Enabling directory fragments can help. The complete fix is make client encode hash of last dentry in readdir request.

#4 - 03/22/2017 06:31 AM - geng jichao

I have used the offset parameter of the ceph_dir_llseek function, and it will be passed to mds in readdir request,if the last_name is null, I will use the offset as offset_hash in the handle_client_readdir.it can avoid unnecessary requests to mds, the performance has been greatly improved, but I do not know how to fill in the cache, the ceph_readdir_cache_control.index may be reset to zero, which will cause the contents of the cache error.

#5 - 03/27/2017 01:40 PM - John Spray

- Assignee set to Jeff Layton

#6 - 04/03/2017 12:53 PM - Zheng Yan

- Status changed from New to In Progress

- Assignee changed from Jeff Layton to Zheng Yan

#7 - 04/04/2017 01:50 PM - Zheng Yan

- Status changed from In Progress to Fix Under Review

https://github.com/ceph/ceph/pull/14317

#8 - 04/05/2017 01:32 AM - Zheng Yan

kernel patch https://github.com/ceph/ceph-client/commit/b7e2eee12aa174bc91279a7cee85e9ea73092bad

#9 - 04/05/2017 05:52 AM - geng jichao

I have a question, if the file struct is destroyed□how to ensure that cache_ctl.index is correct□
In other words□req->r_readdir_cache_idx = fi.readdir_cache_idx, then cache_ctl.index = req->r_readdir_cache_idx, but when the file struct is destroyed, the fi.readdir_cache_idx is reset to zero□this causes the cache error□

#10 - 04/05/2017 09:20 AM - Zheng Yan

geng jichao wrote:

> I have a question, if the file struct is destroyed□how to ensure that cache_ctl.index is correct□
> In other words□req->r_readdir_cache_idx = fi.readdir_cache_idx, then cache_ctl.index = req->r_readdir_cache_idx, but when the file struct is destroyed, the fi.readdir_cache_idx is reset to zero□this causes the cache error□

req->r_readdir_cache_idx is -1 by default. cache is disabled unless ceph_readdir_prepopulate() set it to 0

#11 - 04/24/2017 09:12 PM - John Spray

The userspace piece (https://github.com/ceph/ceph/pull/14317) has merged.

Zheng: please resolve the ticket when the kernel part has gone upstream

#12 - 09/28/2017 10:04 AM - Zheng Yan

- Status changed from Fix Under Review to Resolved