

Ceph - Tasks #13618

objectstore_bench did not support newstore

10/27/2015 06:31 PM - James Liu

Status: New	% Done: 0%
Priority: Normal	Spent time: 0.00 hour
Assignee:	
Category:	
Target version:	
Tags:	Affected Versions:
Reviewed:	Pull request ID:
Description ObjectStore_bench did not support newstore	

History

#1 - 10/27/2015 08:02 PM - Nathan Cutler

- Project changed from Stable releases to Ceph

#2 - 11/20/2015 05:41 PM - Igor Fedotov

I am able to use newstore in objectstore_bench using following config file (provided with -c command option):

; generated by vstart.sh on Thu Nov 19 15:42:05 MSK 2015

[global]

filestore fd cache size = 32

osd objectstore = newstore

newstore backend = leveldb

enable experimental unrecoverable data corrupting features = *

[osd]

osd data = /var/lib/ceph/osd/ceph-admin

osd journal = /var/lib/ceph2/osdd/ceph-admin2

osd journal size = 100

#osd class tmp = out

#osd class dir = .libs

#osd scrub load threshold = 5.0

#osd debug op order = true

filestore wbthrottle xfs ios start flusher = 10

filestore wbthrottle xfs ios hard limit = 20

filestore wbthrottle xfs inodes hard limit = 30

filestore wbthrottle btrfs ios start flusher = 10

filestore wbthrottle btrfs ios hard limit = 20

filestore wbthrottle btrfs inodes hard limit = 30

#3 - 11/25/2015 10:46 PM - James Liu

thanks for providing the configure file. I am able to run objectstore_bench for newstore too over here. However, the cmake did not support rocksdb yet. we have no way to test newstore with rocksdb. any ideas?

#4 - 11/25/2015 10:53 PM - James Liu

```
sudo ./src/test/ceph_objectstore_bench --conf=./src/test/ceph.conf --size=2000M --block-size=16K --threads=1
```

Whenever I tried to run smaller block size like 16K, I got below errors, any clues?

```
-5> 2015-11-25 14:51:51.194562 7f93bae86700 2 newstore(/mnt/nvmedevice) _do_write prepared aio 0x4b2f210
-4> 2015-11-25 14:51:51.194579 7f93bae86700 -1 newstore(/mnt/nvmedevice) _open_fid on 1/1: (24)
-3> 2015-11-25 14:51:51.194589 7f93bae86700 0 newstore(/mnt/nvmedevice) error (24) not handled on operation 10 (op 0, counting from 0)
-2> 2015-11-25 14:51:51.194592 7f93bae86700 0 newstore(/mnt/nvmedevice) unexpected error code
-1> 2015-11-25 14:51:51.194593 7f93bae86700 0 newstore(/mnt/nvmedevice) transaction dump:
```

```
{
"ops": [
{
"op_num": 0,
"op_name": "write",
"collection": "0.0_head",
"oid": "-2V00000000Vosbench/head",
"length": 1048576,
"offset": 1061158912,
"bufferlist length": 1048576
}
]
}
```

```
0> 2015-11-25 14:51:51.197631 7f93bae86700 -1 /home/jamesliu/WorkSpace/ceph_casey/src/os/newstore/NewStore.cc: In function '(null)' thread 7f93bae86700 time 2015-11-25 14:51:51.194646
```

```
/home/jamesliu/WorkSpace/ceph_casey/src/os/newstore/NewStore.cc: 3151: FAILED assert(0 == "unexpected error")
```

```
ceph version v9.0.3-1491-gc229310 (c2293104d26bd82731a5ff28c699254ce8634e7d)
```

```
1: (ceph::__ceph_assert_fail(char const*, char const*, int, char const*)+0x95) [0xcc0717]
2: (NewStore::_txc_add_transaction(NewStore::TransContext*, ObjectStore::Transaction*)+0x1b38) [0xbc0154]
3: (NewStore::queue_transactions(ObjectStore::Sequencer*, std::list<ObjectStore::Transaction*, std::allocator<ObjectStore::Transaction*> >&, std::shared_ptr<TrackedOp>, ThreadPool::TPHandle*)+0x4f0) [0xbbd904]
4: (ObjectStore::queue_transactions(ObjectStore::Sequencer*, std::list<ObjectStore::Transaction*, std::allocator<ObjectStore::Transaction*> >&, Context*, Context*, Context*, std::shared_ptr<TrackedOp>, ThreadPool::TPHandle*)+0xe2) [0xb152a0]
5: (osbench_worker(ObjectStore*, Config const&, coll_t, ghobject_t, unsigned long)+0x50e) [0xb10a0a]
6: (void std::_Bind_simple<void ((ObjectStore, std::reference_wrapper<Config>, coll_t, ghobject_t, unsigned long))(ObjectStore*, Config const&, coll_t, ghobject_t, unsigned long)>::_M_invoke<0ul, 1ul, 2ul, 3ul, 4ul>(std::_Index_tuple<0ul, 1ul, 2ul, 3ul, 4ul>)+0xf2) [0xb1d9ce]
7: (std::_Bind_simple<void ((ObjectStore, std::reference_wrapper<Config>, coll_t, ghobject_t, unsigned long))(ObjectStore*, Config const&, coll_t, ghobject_t, unsigned long)>::operator()()+0x1b) [0xb1d757]
8: (std::thread::_Impl<std::_Bind_simple<void ((ObjectStore, std::reference_wrapper<Config>, coll_t, ghobject_t, unsigned long))(ObjectStore*, Config const&, coll_t, ghobject_t, unsigned long)> >::_M_run()+0x1c) [0xb1d6f0]
9: ((()+0xb1a40) [0x7f93c8a7fa40]
10: ((()+0x8182) [0x7f93c9c22182]
11: (clone()+0x6d) [0x7f93c81e730d]
```

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

#5 - 11/27/2015 12:15 PM - Igor Fedotov

James Liu wrote:

thanks for providing the configure file. I am able to run objectstore_bench for newstore too over here. However, the cmake did not support rocksdb yet. we have no way to test newstore with rocksdb. any ideas?

I didn't use CMake but I managed to build objectstore_bench with rocksdb support using regular make.

For fresh ceph repo clone do the following:

```
./autogen.sh
./configure --with-librocksdb
make
make ceph_objectstore_bench
```

#and don't forget to enable rocksdb in config file

#6 - 11/27/2015 12:22 PM - Igor Fedotov

James Liu wrote:

```
sudo ./src/test/ceph_objectstore_bench --conf=./src/test/ceph.conf --size=2000M --block-size=16K --threads=1
```

Whenever I tried to run smaller block size like 16K, I got below errors, any clues?

Not sure this depends on block size. I tried similar command with --size=200M and it works fine for me (newstore+leveldb). Could you reduce the size and check?

#7 - 12/01/2015 12:08 AM - James Liu

The problem was caused by no of open file limits:

1. Please check your open file limits in the system by command :

```
ulimit -n
```

2. The default is 1024. please edit

```
/etc/security/limits.conf
```

and add below section to the file * - nofile 655360

You also need to edit /etc/pam.d/common-session* and add the following line to the end:

```
session required pam_limits.so
```

#8 - 12/02/2015 05:50 PM - James Liu

Got another error to run the ceph_objectstore_bench against key value store.

```
-11> 2015-12-02 09:46:52.093902 7fadf5a60700 5 _do_op 0x7fadfcf38e50 seq 1 osr(main 0x7fffa043aa80)/0x7fffa043aa80 start
-10> 2015-12-02 09:46:52.093906 7fadf5a60700 10 _do_transaction on 0x7fffa043add0
-9> 2015-12-02 09:46:52.093915 7fadf5a60700 15 _remove 0.0_head/-2/00000000/osbench/head
-8> 2015-12-02 09:46:52.093926 7fadf5a60700 20 remove_header removing 1 cid 0.0_head oid -2/00000000/osbench/head
-7> 2015-12-02 09:46:52.093940 7fadf5a60700 20 clear_header clearing seq 1
-6> 2015-12-02 09:46:52.093979 7fadf5a60700 10 _remove 0.0_head/-2/00000000/osbench/head = 0
-5> 2015-12-02 09:46:52.094205 7fadf5a60700 5 submit_transaction r = 0
-4> 2015-12-02 09:46:52.094222 7fadf5a60700 10 _do_op 0x7fadfcf38e50 seq 1 r = 0, finisher 0x7fadfcf0de50
0
-3> 2015-12-02 09:46:52.094227 7fadf5a60700 10 _finish_op 0x7fadfcf38e50 seq 1 osr(main 0x7fffa043aa80)/0x7fffa043aa80 lat 0.000387
-2> 2015-12-02 09:46:52.094233 7fadf5a60700 15 KeyValueStore::op_tp worker wq KeyValueStore::OpWQ done processing 0x7fadfcf38350 (0 active)
-1> 2015-12-02 09:46:52.094236 7fadf5a60700 20 KeyValueStore::op_tp worker waiting
0> 2015-12-02 09:46:52.094537 7fadf525f700 -1 common/Finisher.cc: In function 'void* Finisher::finisher_thread_entry()' thread 7fadf525f700 time 2015-12-02 09:46:52.093392
common/Finisher.cc: 74: FAILED assert(!ls_rval.empty())
```

```
ceph version 10.0.0-793-ge5c9e20 (e5c9e20a1689116e7f2fb019652a6b08eadf118e)
1: (ceph::__ceph_assert_fail(char const*, char const*, int, char const*)+0x8b) [0x7fadfaa926cb]
2: (Finisher::finisher_thread_entry()+0xae1) [0x7fadfaa71aa1]
3: ((()+0x8182) [0x7fadf8f4f182]
4: (clone()+0x6d) [0x7fadf798530d]
NOTE: a copy of the executable, or `objdump -rdS &lt;executable>` is needed to interpret this.
```