1. Testing Environment

1.1 Software

Kernel: sles12-sp1-rc1 (3.12.48-1-default)
HA: sles12-sp1-rc1

1.2   Hardware

Three physical machines

one for iscsi server(installed SLES12 GM),
IP: 147.2.207.200

the others for cluster(all installed SLES12 SP1 RC1).
IP: 147.2.207.235
IP: 147.2.207.233

2. Functional test (9 pass,1 fail)

 add and re-add doesn't work stable, refer to 10th as follow.

Note: function testing is based on VMs.

Node1:
zl-rc1-144,   147.2.208.144
Node2:
zl-rc1-166,   147.2.208.166
Node3:
zl-rc1-169,   147.2.208.169

2.1   create   --PASS

create local raid1 and cluster-md all work normally.

2.2   assemble   --PASS

all nodes of cluster can assemble md0 normally. Local raid1 works well.

2.3 check mdstat   --PASS (include resync checking)

/proc/mdstat   showed the correct infos.   --PASS
zl-rc1-144:~ # cat /proc/mdstat
Personalities : [raid1]
md0 : active raid1 sda[0] sdb[1]
     419136 blocks super 1.2 [2/2] [UU]
     bitmap: 0/1 pages [0KB], 65536KB chunk

unused devices: <none>

2.4 format FS --PASS (such as ocfs2)

format ocfs2 FS on md0 normally, mount and umount work well.
zl-rc1-144:~ # mkfs.ocfs2 /dev/md0
mkfs.ocfs2 1.8.2
Cluster stack: pcmk
Cluster name: zl-ha
Stack Flags: 0x0
NOTE: Feature extended slot map may be enabled
Label:
Features: sparse extended-slotmap backup-super unwritten inline-data strict-journal-super xattr indexed-dirs refcount discontig-bg
Block size: 1024 (10 bits)
Cluster size: 4096 (12 bits)
Volume size: 429195264 (104784 clusters) (419136 blocks)
Cluster groups: 14 (tail covers 4944 clusters, rest cover 7680 clusters)
Extent allocator size: 2097152 (1 groups)
Journal size: 16777216
Node slots: 2
Creating bitmaps: done
Initializing superblock: done
Writing system files: done
Writing superblock: done
Writing backup superblock: 0 block(s)
Formatting Journals: done
Growing extent allocator: done
Formatting slot map: done
Formatting quota files: done
Writing lost+found: done
mkfs.ocfs2 successful

2.5 rd/wr operations   --PASS

zl-rc1-144:~ # dd if=/dev/zero of=/dev/md0 bs=1M count=200
200+0 records in
200+0 records out
209715200 bytes (210 MB) copied, 13.0319 s, 16.1 MB/s
zl-rc1-144:~ # dd if=/dev/md0 of=md-io-test bs=1M count=200
200+0 records in
200+0 records out
209715200 bytes (210 MB) copied, 1.57526 s, 133 MB/s
zl-rc1-144:~ # du md-io-test -lh
200M   md-io-test
zl-rc1-144:~ # mount /dev/md0 /mnt/
zl-rc1-144:~ # cd /mnt/
zl-rc1-144:/mnt # ls
lost+found
zl-rc1-144:/mnt # touch 1 2 3
zl-rc1-144:/mnt # ls
1   2   3   lost+found
zl-rc1-144:/mnt # echo 123 > 1; echo 456 > 2; echo 789 > 3
zl-rc1-144:/mnt # cat 1 2 3
123
456
789
zl-rc1-144:/mnt # cd
zl-rc1-144:~ # umount /mnt
zl-rc1-144:~ # mount /dev/md0 /mnt/
zl-rc1-144:~ # rm /mnt/
1   2   3   lost+found/
zl-rc1-144:~ # rm /mnt/1 /mnt/2 /mnt/3
zl-rc1-144:~ # umount /mnt
zl-rc1-144:~ # mount /dev/md0 /mnt/
zl-rc1-144:~ # ls /mnt/
lost+found
zl-rc1-144:~ # umount /mnt

2.6 nodes update   --PASS

#mdadm --assemble md0 --nodes=3 --update=nodes /dev/sda /dev/sdb
zl-rc1-144:~ # mdadm -X /dev/sda
        Filename : /dev/sda
          Magic : 6d746962
         Version : 5

UUID : 5c040480:d452377c:ca294357:669953bc
            Chunksize : 64 MB
               Daemon : 5s flush period
          Write Mode : Normal
            Sync Size : 419136 (409.38 MiB 429.20 MB)
        <mark>Cluster nodes : 3</mark>
         Cluster name : zl-ha
            Node Slot : 0
               Events : 0
       Events Cleared : 0
                State : OK
               Bitmap : 7 bits (chunks), 0 dirty (0.0%)
            Node Slot : 1
               Events : 0
       Events Cleared : 0
                State : OK
               Bitmap : 7 bits (chunks), 0 dirty (0.0%)
            Node Slot : 2
               Events : 0
       Events Cleared : 0
                State : OK
               Bitmap : 7 bits (chunks), 0 dirty (0.0%)

2.7   home-cluster update   --PASS

 # mdadm --assemble md0 --update=home-cluster --home-cluster=zl-ha /dev/sda
/dev/sdb
zl-rc1-144:~ # mdadm -D /dev/md0
/dev/md0:
            Version : 1.2
      Creation Time : Mon Sep 28 17:32:18 2015
         Raid Level : raid1
         Array Size : 419136 (409.38 MiB 429.20 MB)
      Used Dev Size : 419136 (409.38 MiB 429.20 MB)
       Raid Devices : 2
      Total Devices : 2
        Persistence : Superblock is persistent

      Intent Bitmap : Internal

        Update Time : Mon Sep 28 17:32:18 2015
              State : clean
     Active Devices : 2
    Working Devices : 2

Failed Devices : 0
        Spare Devices : 0

              Name : zl-rc1-144:md0   (local to host zl-rc1-144)
    ==Cluster Name : zl-ha==
              UUID : 5c040480:d452377c:ca294357:669953bc
            Events : 0

     Number   Major   Minor   RaidDevice State
        0   8   0   0   active sync   /dev/sda
        1   8   16   1   active sync   /dev/sdb

2.8   examine array   --PASS

zl-rc1-144:~ # mdadm --examine /dev/sda
/dev/sda:
          Magic : a92b4efc
        Version : 1.2
     Feature Map : 0x101
      Array UUID : 5c040480:d452377c:ca294357:669953bc
           Name : zl-rc1-144:md0   (local to host zl-rc1-144)
    Cluster Name : zl-ha
   Creation Time : Mon Sep 28 17:32:18 2015
      Raid Level : raid1
    Raid Devices : 2

  Avail Dev Size : 838318 (409.40 MiB 429.22 MB)
      Array Size : 419136 (409.38 MiB 429.20 MB)
   Used Dev Size : 838272 (409.38 MiB 429.20 MB)
     Data Offset : 544 sectors
    Super Offset : 8 sectors
    Unused Space : before=464 sectors, after=46 sectors
           State : clean
     Device UUID : 95ccf63d:dba3140b:c462583e:2c9e551f

 Internal Bitmap : 8 sectors from superblock
     Update Time : Mon Sep 28 17:32:18 2015
        Checksum : 1c78885d - correct
          Events : 0

     Device Role : Active device 0

Array State : AA ('A' == active, '.' == missing, 'R' == replacing)

2.9 grow and switch bitmap --PASS (between internal and clustered)

zl-rc1-169:~ # mdadm -D /dev/md0
/dev/md0:
        Version : 1.2
  Creation Time : Mon Sep 28 14:44:42 2015
     Raid Level : raid1
     Array Size : 419136 (409.38 MiB 429.20 MB)
  Used Dev Size : 419136 (409.38 MiB 429.20 MB)
   Raid Devices : 3
  Total Devices : 3
    Persistence : Superblock is persistent

   Intent Bitmap : Internal

    Update Time : Mon Sep 28 14:48:21 2015
          State : clean
 Active Devices : 3
Working Devices : 3
 Failed Devices : 0
  Spare Devices : 0

           Name : zl-rc1-169:md0   (local to host zl-rc1-169)
   Cluster Name : cluster
           UUID : 8aa7d880:86ccaba4:e5cdbbc0:d35e389a
         Events : 64

    Number   Major   Minor   RaidDevice State
       0   8   0   0   active sync   /dev/sda
       3   8   16   1   active sync   /dev/sdb
       2   8   32   2   active sync   /dev/sdc
zl-rc1-169:~ # mdadm --grow /dev/md0 --bitmap=none
zl-rc1-169:~ # mdadm -D /dev/md0
/dev/md0:
        Version : 1.2
  Creation Time : Mon Sep 28 14:44:42 2015

```
        Raid Level : raid1
        Array Size : 419136 (409.38 MiB 429.20 MB)
     Used Dev Size : 419136 (409.38 MiB 429.20 MB)
      Raid Devices : 3
     Total Devices : 3
       Persistence : Superblock is persistent


       Update Time : Mon Sep 28 14:56:57 2015
             State : clean
    Active Devices : 3
   Working Devices : 3
    Failed Devices : 0
     Spare Devices : 0


              Name : zl-rc1-169:md0   (local to host zl-rc1-169)
              UUID : 8aa7d880:86ccaba4:e5cdbbc0:d35e389a
            Events : 65


    Number   Major   Minor   RaidDevice State
       0       8       0        0      active sync   /dev/sda
       3       8       16       1      active sync   /dev/sdb
       2       8       32       2      active sync   /dev/sdc
zl-rc1-169:~ # mdadm --grow /dev/md0 --bitmap=internal
zl-rc1-169:~ # mdadm -D /dev/md0
/dev/md0:
           Version : 1.2
     Creation Time : Mon Sep 28 14:44:42 2015
        Raid Level : raid1
        Array Size : 419136 (409.38 MiB 429.20 MB)
     Used Dev Size : 419136 (409.38 MiB 429.20 MB)
      Raid Devices : 3
     Total Devices : 3
       Persistence : Superblock is persistent


      Intent Bitmap : Internal


       Update Time : Mon Sep 28 14:57:12 2015
             State : clean
    Active Devices : 3
```

Working Devices : 3
 Failed Devices : 0
  Spare Devices : 0

         Name : zl-rc1-169:md0   (local to host zl-rc1-169)
         UUID : 8aa7d880:86ccaba4:e5cdbbc0:d35e389a
        Events : 66

    Number   Major   Minor   RaidDevice State
       0   8   0   0   active sync   /dev/sda
       3   8   16   1   active sync   /dev/sdb
       2   8   32   2   active sync   /dev/sdc
zl-rc1-169:~ # mdadm --grow /dev/md0 --bitmap=none
zl-rc1-169:~ # mdadm -D /dev/md0
/dev/md0:
        Version : 1.2
  Creation Time : Mon Sep 28 14:44:42 2015
     Raid Level : raid1
     Array Size : 419136 (409.38 MiB 429.20 MB)
  Used Dev Size : 419136 (409.38 MiB 429.20 MB)
   Raid Devices : 3
  Total Devices : 3
    Persistence : Superblock is persistent

    Update Time : Mon Sep 28 14:57:23 2015
          State : clean
 Active Devices : 3
Working Devices : 3
 Failed Devices : 0
  Spare Devices : 0

         Name : zl-rc1-169:md0   (local to host zl-rc1-169)
         UUID : 8aa7d880:86ccaba4:e5cdbbc0:d35e389a
        Events : 67

    Number   Major   Minor   RaidDevice State
       0   8   0   0   active sync   /dev/sda
       3   8   16   1   active sync   /dev/sdb
       2   8   32   2   active sync   /dev/sdc
zl-rc1-169:~ # mdadm --grow /dev/md0 --bitmap=clustered
zl-rc1-169:~ # mdadm -D /dev/md0
/dev/md0:

```
      Version : 1.2
Creation Time : Mon Sep 28 14:44:42 2015
   Raid Level : raid1
   Array Size : 419136 (409.38 MiB 429.20 MB)
Used Dev Size : 419136 (409.38 MiB 429.20 MB)
 Raid Devices : 3
Total Devices : 3
  Persistence : Superblock is persistent

Intent Bitmap : Internal

  Update Time : Mon Sep 28 14:57:37 2015
        State : clean
Active Devices : 3
Working Devices : 3
Failed Devices : 0
 Spare Devices : 0

         Name : zl-rc1-169:md0   (local to host zl-rc1-169)
 Cluster Name : cluster
         UUID : 8aa7d880:86ccaba4:e5cdbbc0:d35e389a
       Events : 68

  Number   Major   Minor   RaidDevice State
     0       8       0        0      active sync   /dev/sda
     3       8      16        1      active sync   /dev/sdb
     2       8      32        2      active sync   /dev/sdc
```

2.10 add and re-add failure device --FAIL

Summarize

add and re-add can work well on local raid1,
also work well on the follow situation; cluster-md only start on one node of cluster, the other nodes of cluster must stop md0.

a. test under pure rc1 environment

   Totally pure SLES12 SP1 RC1 kernel and RC1 HA environment.
"add" only can pass at first time, always got failure once repeatly;   re-add cannot work.

b. verify by following packages based on rc1.

package (download them from Guoqing.):
cluster-md-kmp-default-0.0.1_k3.12.48_1-6.1.x86_64.rpm
mdadm-3.3.1-7.1.x86_64.rpm

1). assemble MD on all nodes of cluster, "add" and "re-add" operation got failure after "set fail" and "remove" device.

Create md:
mdadm --create md0 --bitmap=clustered --raid-devices=3 --level=mirror /dev/sda /dev/sdb /dev/sdc
... ...   sync the /etc/mdadm.conf   to all nodes.   ... ...

cat /etc/mdadm.conf
DEVICE /dev/sda /dev/sdb /dev/sdc
ARRAY /dev/md/md0 metadata=1.2 UUID=00fd4a8c:af0db6c9:f5088e20:0900048c

Node1:
a)   assemble and start cluster-md
zl-rc1-144:~ # mdadm -As
mdadm: /dev/md/md0 has been started with 3 drives.
zl-rc1-144:~ # cat /proc/mdstat
Personalities : [raid1]
md0 : active raid1 sda[0] sdc[2] sdb[1]
     419136 blocks super 1.2 [3/3] [UUU]
     bitmap: 0/1 pages [0KB], 65536KB chunk

unused devices: <none>

b) when "set fail" occurred
zl-rc1-144:~ # cat /proc/mdstat
Personalities : [raid1]
md0 : active raid1 sda[0] sdc[2] sdb[1](F)
     419136 blocks super 1.2 [3/2] [U_U]
     bitmap: 0/1 pages [0KB], 65536KB chunk

unused devices: <none>

c)   when "remove" operation executed on other node.
zl-rc1-144:~ # cat /proc/mdstat

Personalities : [raid1]
md0 : active raid1 sda[0] sdc[2]
    419136 blocks super 1.2 [3/2] [U_U]
    bitmap: 0/1 pages [0KB], 65536KB chunk

unused devices: <none>

d)   it's the status when "add" or "re-add"
zl-rc1-144:~ # cat /proc/mdstat
Personalities : [raid1]
md0 : active raid1 sda[0] sdc[2]
    419136 blocks super 1.2 [3/2] [U_U]
    bitmap: 0/1 pages [0KB], 65536KB chunk

unused devices: <none>

mdadm --manage /dev/md0 --add /dev/sdb
mdadm: add new device failed for /dev/sdb as 3: Invalid argument

mdadm --manage /dev/md0 --re-add /dev/sdb
mdadm: --re-add for /dev/sdb to /dev/md0 is not possible

2). only assemble md array on one node of cluster, such as "zl-rc1-169" node started md and "zl-rc1-144, zl-rc1-166" nodes all stoped; then "add" and "re-add" all works well.

node1 :
mdadm --create md0 --bitmap=clustered --raid-devices=3 --level=mirror /dev/sda /dev/sdb /dev/sdc

mdadm -Ds
ARRAY   /dev/md/md0   metadata=1.2   name=zl-rc1-169:md0
UUID=00fd4a8c:af0db6c9:f5088e20:0900048c

cat /etc/mdadm.conf
DEVICE /dev/sda /dev/sdb /dev/sdc
ARRAY /dev/md/md0 metadata=1.2 UUID=00fd4a8c:af0db6c9:f5088e20:0900048c

mdadm -S md0

```
zl-rc1-166:~ # cat /proc/mdstat
Personalities : [raid1]
unused devices: <none>


node2 :

zl-rc1-144:~ # cat /proc/mdstat
Personalities : [raid1]
unused devices: <none>


node3 :
zl-rc1-169:~ # mdadm -As
mdadm: /dev/md/md0 has been started with 3 drives.
zl-rc1-169:~ # cat /proc/mdstat
Personalities : [raid1]
md0 : active raid1 sda[0] sdc[2] sdb[1]
      419136 blocks super 1.2 [3/3] [UUU]
      bitmap: 0/1 pages [0KB], 65536KB chunk


unused devices: <none>
zl-rc1-169:~ # mdadm --manage /dev/md0 --fail /dev/sda
mdadm: set /dev/sda faulty in /dev/md0
zl-rc1-169:~ # cat /proc/mdstat
Personalities : [raid1]
md0 : active raid1 sda[0](F) sdc[2] sdb[1]
      419136 blocks super 1.2 [3/2] [_UU]
      bitmap: 0/1 pages [0KB], 65536KB chunk

unused devices: <none>
zl-rc1-169:~ # mdadm --manage /dev/md0 --remove /dev/sda
mdadm: hot removed /dev/sda from /dev/md0
zl-rc1-169:~ # cat /proc/mdstat
Personalities : [raid1]
md0 : active raid1 sdc[2] sdb[1]
      419136 blocks super 1.2 [3/2] [_UU]
      bitmap: 0/1 pages [0KB], 65536KB chunk

unused devices: <none>
zl-rc1-169:~ # mdadm --manage /dev/md0 --re-add /dev/sda
mdadm: re-added /dev/sda
zl-rc1-169:~ # cat /proc/mdstat
Personalities : [raid1]
md0 : active raid1 sda[0] sdc[2] sdb[1]
```

```
        419136 blocks super 1.2 [3/3] [UUU]
        bitmap: 0/1 pages [0KB], 65536KB chunk

unused devices: <none>
zl-rc1-169:~ # mdadm --manage /dev/md0 --fail /dev/sdb
mdadm: set /dev/sdb faulty in /dev/md0
zl-rc1-169:~ # cat /proc/mdstat
Personalities : [raid1]
md0 : active raid1 sda[0] sdc[2] sdb[1](F)
        419136 blocks super 1.2 [3/2] [U_U]
        bitmap: 0/1 pages [0KB], 65536KB chunk

unused devices: <none>
zl-rc1-169:~ # mdadm --manage /dev/md0 --remove /dev/sdb
mdadm: hot removed /dev/sdb from /dev/md0
You have new mail in /var/mail/root
zl-rc1-169:~ # cat /proc/mdstat
Personalities : [raid1]
md0 : active raid1 sda[0] sdc[2]
        419136 blocks super 1.2 [3/2] [U_U]
        bitmap: 0/1 pages [0KB], 65536KB chunk

unused devices: <none>
zl-rc1-169:~ # mdadm --manage /dev/md0 --add /dev/sdb
mdadm: re-added /dev/sdb
zl-rc1-169:~ # cat /proc/mdstat
Personalities : [raid1]
md0 : active raid1 sdb[1] sda[0] sdc[2]
        419136 blocks super 1.2 [3/3] [UUU]
        bitmap: 0/1 pages [0KB], 65536KB chunk

unused devices: <none>
```

3).   local RAID1 all pass with "add" and "re-add".

3.Stress testing

testsuit named "ocfs2-test", format ocfs2 FS on md device.

Single node test -> PASS

the following cases are listed.

```
# vim ./ocfs2-test/bin/single_run-WIP.sh
...
run_aiostress ${LOGDIR} ${DEVICE} ${MOUNTPOINT}
run_mkfs ${LOGDIR} ${DEVICE} ${MOUNTPOINT} ${CLUSTER_STACK}
${CLUSTER_NAME}
run_tunefs ${LOGDIR} ${DEVICE} ${MOUNTPOINT} ${CLUSTER_STACK}
${CLUSTER_NAME}
run_backup_super        ${LOGDIR}        ${DEVICE}        ${MOUNTPOINT}
${CLUSTER_STACK} ${CLUSTER_NAME}
run_inline_data         ${LOGDIR}        ${DEVICE}        ${MOUNTPOINT}
${CLUSTER_STACK} ${CLUSTER_NAME}
run_xattr_test ${LOGDIR} ${DEVICE} ${MOUNTPOINT} ${CLUSTER_STACK}
${CLUSTER_NAME}
run_reflink_test        ${LOGDIR}        ${DEVICE}        ${MOUNTPOINT}
${CLUSTER_STACK} ${CLUSTER_NAME}
...
```

4.Performance compare

Three disks exposed by iscsi-target(ISCSI-server IP: 147.2.207.200) are used as raid1.

4.1   dd compared between local raid1 and md.

4.1.1 dd read compare

local raid1:
```
# dd iflag=direct,nonblock   if=/dev/md0 of=/dev/null bs=1M count=40960
40960+0 records in
40960+0 records out
42949672960 bytes (43 GB) copied, 480.757 s, 89.3 MB/s
```
Cluster-md:
```
# dd iflag=direct,nonblock   if=/dev/md0 of=/dev/null bs=1M count=40960
40960+0 records in
40960+0 records out
42949672960 bytes (43 GB) copied, 415.498 s, 103 MB/s
```

4.1.2 dd write compare

local raid1:
dd oflag=direct,nonblock if=/dev/zero of=/dev/md0 bs=1M count=40960
40960+0 records in
40960+0 records out
42949672960 bytes (43 GB) copied, 1433.75 s, 30.0 MB/s

Cluster-md:
dd oflag=direct,nonblock if=/dev/zero of=/dev/md0 bs=1M count=40960
40960+0 records in
40960+0 records out
42949672960 bytes (43 GB) copied, 1502.29 s, 28.6 MB/s

4.1.3 dd rd/wr compare

local raid1:
# dd iflag=direct,nonblock oflag=direct,nonblock if=/dev/md0 of=/home/test/iotest
bs=1M count=40960
40960+0 records in
40960+0 records out
42949672960 bytes (43 GB) copied, 641.374 s, 67.0 MB/s

Cluster-md:
dd iflag=direct,nonblock oflag=direct,nonblock if=/dev/md0 of=/home/test/iotest
bs=1M count=40960
40960+0 records in
40960+0 records out
42949672960 bytes (43 GB) copied, 653.715 s, 65.7 MB/s

4.2   FIO test compare

4.2.1 10G + 24 threads -> Read

Local raid1:

read : io=245760MB, bw=114062KB/s, iops=1782, runt=2206338msec

Cluster-md:

read : io=245760MB, bw=111973KB/s, iops=1749, runt=2247486msec

4.2.2 10G + 24 threads -> write

Local raid1:

write: io=245760MB, bw=54994KB/s, iops=859, runt=4576126msec

Cluster-md:

write: io=245760MB, bw=55056KB/s, iops=860, runt=4570953msec

4.2.3 10G + 24 threads -> rand read

Local raid1:

read : io=245760MB, bw=114797KB/s, iops=1793, runt=2192204msec

Cluster-md:

read : io=245760MB, bw=114799KB/s, iops=1793, runt=2192170msec

4.2.4 10G + 24 threads -> rand write

Local raid1:

write: io=245760MB, bw=32192KB/s, iops=502, runt=7817456msec

Cluster-md:

write: io=245760MB, bw=32431KB/s, iops=506, runt=7759705msec

4.2.5 10G + 24 threads -> rw mixed rd

Local raid1:

read : io=171929MB, bw=71145KB/s, iops=1111, runt=2474596msec

… … …
write: io=73832MB, bw=30552KB/s, iops=477, runt=2474596msec

Cluster-md:

read : io=171929MB, bw=66869KB/s, iops=1044, runt=2632840msec
… … …
write: io=73832MB, bw=28716KB/s, iops=448, runt=2632840msec

4.3 IOzone compared with local raid1, CLVM and cluster-md

Please refer to the attached logs.

5.Performance details

5.1 Local RAID1 performance

5.1.1 dd include of read, write and rd/wr.

Read

time dd iflag=direct,nonblock   if=/dev/md0 of=/dev/null bs=1M count=40960
40960+0 records in
40960+0 records out
42949672960 bytes (43 GB) copied, 480.757 s, 89.3 MB/s

real   8m0.759s
user   0m0.072s
sys   0m4.468s

Write

time dd oflag=direct,nonblock if=/dev/zero of=/dev/md0 bs=1M count=40960
40960+0 records in
40960+0 records out

42949672960 bytes (43 GB) copied, 1433.75 s, 30.0 MB/s

real   23m53.747s
user   0m0.072s
sys   0m13.784s

Read/write

time dd iflag=direct,nonblock oflag=direct,nonblock if=/dev/md0 of=/home/test/iotest bs=1M count=40960
40960+0 records in
40960+0 records out
42949672960 bytes (43 GB) copied, 641.374 s, 67.0 MB/s

real   10m41.395s
user   0m0.128s
sys   0m10.672s

5.1.2 FIO testing

Download source code from github.
Operation contains **"read", "write", "randread", "randwrite" and "rw-mix-read"**, threads=(num of processor) x 2, size= 10G, 20G

1)   10G   + 24 threads

Read

test-read-10G: (g=0): rw=read, bs=64K-64K/64K-64K/64K-64K, ioengine=libaio, iodepth=64
...
fio-2.1.10
Starting 24 processes

test-read-10G: (groupid=0, jobs=24): err= 0: pid=16444: Mon Sep 28 02:37:30 2015
  read : io=245760MB, bw=114062KB/s, iops=1782, runt=2206338msec
    slat (usec): min=3, max=69367, avg=14.44, stdev=85.75
    clat (msec): min=11, max=54063, avg=776.60, stdev=765.73
     lat (msec): min=11, max=54063, avg=776.61, stdev=765.73

clat percentiles (msec):
| 1.00th=[322], 5.00th=[392], 10.00th=[429], 20.00th=[537],
| 30.00th=[537], 40.00th=[553], 50.00th=[685], 60.00th=[848],
| 70.00th=[1074], 80.00th=[1074], 90.00th=[1106], 95.00th=[1139],
| 99.00th=[1352], 99.50th=[1385], 99.90th=[4113], 99.95th=[16712],
| 99.99th=[16712]
  bw (KB /s): min=   36, max=62976, per=5.18%, avg=5904.14, stdev=2455.57
  lat (msec) : 20=0.01%, 50=0.01%, 100=0.27%, 250=0.41%, 500=16.78%
  lat (msec) : 750=39.19%, 1000=11.09%, 2000=32.08%, >=2000=0.16%
 cpu   : usr=0.03%, sys=0.16%, ctx=297422, majf=0, minf=13063
   IO   depths   :   1=0.1%,   2=0.1%,   4=0.1%,   8=0.1%,   16=0.1%,
32=0.1%, >=64=100.0%
     submit   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.0%, >=64=0.0%
     complete   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.1%, >=64=0.0%
   issued   : total=r=3932160/w=0/d=0, short=r=0/w=0/d=0
   latency   : target=0, window=0, percentile=100.00%, depth=64

Run status group 0 (all jobs):
       READ: io=245760MB,   aggrb=114061KB/s,   minb=114061KB/s,
maxb=114061KB/s, mint=2206338msec, maxt=2206338msec

Disk stats (read/write):
    md0:   ios=3932000/0,   merge=0/0,   ticks=0/0,   in_queue=0,   util=0.00%,
aggrios=97362/0,   aggrmerge=1212938/0,   aggrticks=70122342/0,
aggrin_queue=70122761, aggrutil=100.00%
  sdc: ios=159405/0, merge=2334393/0, ticks=139475888/0, in_queue=139477292,
util=100.00%
  sdd: ios=132682/0, merge=1304423/0, ticks=70891140/0, in_queue=70890992,
util=100.00%
  sde: ios=0/0, merge=0/0, ticks=0/0, in_queue=0, util=0.00%

Write

bj-ha-5:/opt/zhilong # cat log-write-10G.txt
test-write-10G: (g=0): rw=write, bs=64K-64K/64K-64K/64K-64K, ioengine=libaio,
iodepth=64
...
fio-2.1.10
Starting 24 processes

test-write-10G: (groupid=0, jobs=24): err= 0: pid=17383: Mon Sep 28 03:55:26 2015
  write: io=245760MB, bw=54994KB/s, iops=859, runt=4576126msec
    slat (usec): min=5, max=2629.1K, avg=432.61, stdev=24711.00
    clat (msec): min=4, max=8552, avg=1784.59, stdev=384.44
     lat (msec): min=5, max=8552, avg=1785.03, stdev=384.98
    clat percentiles (msec):
     | 1.00th=[660], 5.00th=[1303], 10.00th=[1565], 20.00th=[1713],
     | 30.00th=[1713], 40.00th=[1713], 50.00th=[1713], 60.00th=[1713],
     | 70.00th=[1729], 80.00th=[1762], 90.00th=[2180], 95.00th=[2573],
     | 99.00th=[3261], 99.50th=[3490], 99.90th=[3916], 99.95th=[4146],
     | 99.99th=[5014]
    bw (KB /s): min=   24, max=26875, per=4.41%, avg=2427.82, stdev=936.93
    lat (msec) : 10=0.01%, 20=0.01%, 50=0.01%, 100=0.01%, 250=0.03%
        lat   (msec)   :   500=0.21%,   750=0.87%,   1000=0.95%,
2000=83.68%, >=2000=14.24%
  cpu   : usr=0.03%, sys=0.08%, ctx=246358, majf=0, minf=737
    IO   depths   :   1=0.1%,   2=0.1%,   4=0.1%,   8=0.1%,   16=0.1%,
32=0.1%, >=64=100.0%
      submit   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.0%, >=64=0.0%
      complete   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.1%, >=64=0.0%
      issued   : total=r=0/w=3932160/d=0, short=r=0/w=0/d=0
      latency   : target=0, window=0, percentile=100.00%, depth=64

Run status group 0 (all jobs):
  WRITE: io=245760MB, aggrb=54993KB/s, minb=54993KB/s, maxb=54993KB/s,
mint=4576126msec, maxt=4576126msec

Disk stats (read/write):
    md0: ios=228/3932130, merge=0/0, ticks=0/0, in_queue=0, util=0.00%,
aggrios=63/167504,   aggrmerge=55/2451456,   aggrticks=5160/231404701,
aggrin_queue=231428037, aggrutil=100.00%
    sdc: ios=149/250018,   merge=106/3678377,   ticks=11136/356874316,
in_queue=356919632, util=100.00%
    sdd: ios=40/252495,   merge=59/3675993,   ticks=4344/337339788,
in_queue=337364480, util=100.00%
  sde: ios=0/0, merge=0/0, ticks=0/0, in_queue=0, util=0.00%

Rand read

bj-ha-5:/opt/zhilong # cat log-randread-10G.txt
test-randread-10G: (g=0): rw=randread, bs=64K-64K/64K-64K/64K-64K, ioengine=libaio, iodepth=64
...
fio-2.1.10
Starting 24 processes

test-randread-10G: (groupid=0, jobs=24): err= 0: pid=18927: Mon Sep 28 04:33:39 2015
  read : io=245760MB, bw=114797KB/s, iops=1793, runt=2192204msec
    slat (usec): min=3, max=1043.2K, avg=13321.16, stdev=75910.29
    clat (msec): min=1, max=3089, avg=841.14, stdev=173.65
     lat (msec): min=1, max=3089, avg=854.46, stdev=192.02
    clat percentiles (msec):
     | 1.00th=[404], 5.00th=[445], 10.00th=[461], 20.00th=[865],
     | 30.00th=[881], 40.00th=[881], 50.00th=[889], 60.00th=[889],
     | 70.00th=[898], 80.00th=[906], 90.00th=[922], 95.00th=[971],
     | 99.00th=[1270], 99.50th=[1352], 99.90th=[1582], 99.95th=[1729],
     | 99.99th=[1991]
    bw (KB /s): min=   73, max=36096, per=4.19%, avg=4807.12, stdev=886.65
    lat (msec) : 2=0.01%, 4=0.01%, 10=0.02%, 20=0.03%, 50=0.08%
    lat (msec) : 100=0.09%, 250=0.24%, 500=11.41%, 750=2.65%, 1000=81.50%
    lat (msec) : 2000=3.99%, >=2000=0.01%
  cpu   : usr=0.03%, sys=0.19%, ctx=127342, majf=0, minf=13035
  IO depths   : 1=0.1%, 2=0.1%, 4=0.1%, 8=0.1%, 16=0.1%, 32=0.1%, >=64=100.0%
     submit   : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%, >=64=0.0%
     complete : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.1%, >=64=0.0%
     issued   : total=r=3932160/w=0/d=0, short=r=0/w=0/d=0
     latency   : target=0, window=0, percentile=100.00%, depth=64

Run status group 0 (all jobs):
   READ: io=245760MB, aggrb=114796KB/s, minb=114796KB/s, maxb=114796KB/s, mint=2192204msec, maxt=2192204msec

Disk stats (read/write):

md0: ios=3932069/0, merge=0/0, ticks=0/0, in_queue=0, util=0.00%, aggrios=1299893/0, aggrmerge=9691/0, aggrticks=117282861/0, aggrin_queue=117283254, aggrutil=90.03%

  sdc: ios=1948709/0, merge=13481/0, ticks=170119768/0, in_queue=170120268, util=89.08%

  sdd: ios=1950970/0, merge=15593/0, ticks=181728816/0, in_queue=181729496, util=90.03%

  sde: ios=0/0, merge=0/0, ticks=0/0, in_queue=0, util=0.00%

Randwrite

bj-ha-5:/opt/zhilong # cat log-randwrite-10G.txt
test-randwrite-10G: (g=0): rw=randwrite, bs=64K-64K/64K-64K/64K-64K, ioengine=libaio, iodepth=64
...
fio-2.1.10
Starting 24 processes

test-randwrite-10G: (groupid=0, jobs=24): err= 0: pid=19822: Mon Sep 28 06:45:38 2015
  write: io=245760MB, bw=32192KB/s, iops=502, runt=7817456msec
    slat (usec): min=9, max=55867K, avg=47656.99, stdev=475909.49
    clat (msec): min=15, max=59733, avg=2998.55, stdev=3565.29
    lat (msec): min=15, max=59733, avg=3046.21, stdev=3592.74
    clat percentiles (msec):
    | 1.00th=[1287], 5.00th=[1713], 10.00th=[1713], 20.00th=[1729],
    | 30.00th=[1942], 40.00th=[2278], 50.00th=[2507], 60.00th=[2737],
    | 70.00th=[3130], 80.00th=[3556], 90.00th=[4015], 95.00th=[4490],
    | 99.00th=[12125], 99.50th=[16712], 99.90th=[16712], 99.95th=[16712],
    | 99.99th=[16712]
    bw (KB /s): min= 1, max=16220, per=4.85%, avg=1561.42, stdev=532.97
    lat (msec) : 20=0.01%, 50=0.01%, 100=0.01%, 250=0.01%, 500=0.03%
    lat (msec) : 750=0.02%, 1000=0.03%, 2000=31.07%, >=2000=68.84%
  cpu : usr=0.02%, sys=0.08%, ctx=251489, majf=0, minf=708
    IO depths : 1=0.1%, 2=0.1%, 4=0.1%, 8=0.1%, 16=0.1%, 32=0.1%, >=64=100.0%
    submit : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%, >=64=0.0%
    complete : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.1%, >=64=0.0%
    issued : total=r=0/w=3932160/d=0, short=r=0/w=0/d=0

```
        latency    : target=0, window=0, percentile=100.00%, depth=64
```

Run status group 0 (all jobs):
  WRITE: io=245760MB, aggrb=32191KB/s, minb=32191KB/s, maxb=32191KB/s, mint=7817456msec, maxt=7817456msec

Disk stats (read/write):
    md0: ios=342/3932072, merge=0/0, ticks=0/0, in_queue=0, util=0.00%, aggrios=79/2610013,   aggrmerge=55/9895,   aggrticks=16420/424011650, aggrin_queue=424034412, aggrutil=100.00%
      sdc: ios=231/3930295,   merge=155/1947,   ticks=12036/118635684, in_queue=118649840, util=74.51%
      sdd: ios=6/3899744,   merge=10/27740,   ticks=37224/1153399268, in_queue=1153453396, util=100.00%
  sde: ios=0/0, merge=0/0, ticks=0/0, in_queue=0, util=0.00%

Rw-mix-read

bj-ha-5:/opt/zhilong # cat log-randrw\ -rwmixread\=70-10G.txt
test-randrw: (g=0): rw=randrw, bs=64K-64K/64K-64K/64K-64K, ioengine=libaio, iodepth=64
...
fio-2.1.10
Starting 24 processes

test-randrw: (groupid=0, jobs=24): err= 0: pid=22737: Mon Sep 28 07:28:33 2015
  read : io=171929MB, bw=71145KB/s, iops=1111, runt=2474596msec
    slat (usec): min=4, max=50493K, avg=9637.71, stdev=136664.52
    clat (msec): min=2, max=107322, avg=901.72, stdev=1903.76
     lat (msec): min=2, max=107322, avg=911.36, stdev=1912.56
    clat percentiles (msec):
     | 1.00th=[330], 5.00th=[545], 10.00th=[562], 20.00th=[586],
     | 30.00th=[611], 40.00th=[717], 50.00th=[832], 60.00th=[873],
     | 70.00th=[906], 80.00th=[1004], 90.00th=[1221], 95.00th=[1450],
     | 99.00th=[1745], 99.50th=[2311], 99.90th=[15139], 99.95th=[16712],
     | 99.99th=[16712]
    bw (KB /s): min=   1, max=15168, per=4.63%, avg=3293.50, stdev=957.45
  write: io=73832MB, bw=30552KB/s, iops=477, runt=2474596msec
    slat (usec): min=7, max=50537K, avg=27443.20, stdev=302626.50
```

clat (msec): min=6, max=107322, avg=1038.06, stdev=2125.45
     lat (msec): min=6, max=107322, avg=1065.50, stdev=2163.58
    clat percentiles (msec):
     |  1.00th=[  519],  5.00th=[  570], 10.00th=[  594], 20.00th=[  791],
     | 30.00th=[  840], 40.00th=[  865], 50.00th=[  889], 60.00th=[  914],
     | 70.00th=[  996], 80.00th=[ 1188], 90.00th=[ 1434], 95.00th=[ 1565],
     | 99.00th=[ 1844], 99.50th=[ 3425], 99.90th=[15533], 99.95th=[16712],
     | 99.99th=[16712]
    bw (KB  /s): min=    1, max= 7746, per=4.64%, avg=1416.33, stdev=518.28
    lat (msec) : 4=0.01%, 10=0.01%, 20=0.02%, 50=0.05%, 100=0.05%
    lat (msec) : 250=0.11%, 500=1.27%, 750=33.17%, 1000=42.25%, 2000=22.46%
    lat (msec) : >=2000=0.62%
  cpu        : usr=0.03%, sys=0.19%, ctx=166504, majf=0, minf=731
    IO    depths    :   1=0.1%,    2=0.1%,    4=0.1%,    8=0.1%,    16=0.1%,
32=0.1%, >=64=100.0%
      submit    :   0=0.0%,    4=100.0%,    8=0.0%,    16=0.0%,    32=0.0%,
64=0.0%, >=64=0.0%
      complete   :   0=0.0%,    4=100.0%,    8=0.0%,    16=0.0%,    32=0.0%,
64=0.1%, >=64=0.0%
    issued    : total=r=2750856/w=1181304/d=0, short=r=0/w=0/d=0
    latency   : target=0, window=0, percentile=100.00%, depth=64


Run status group 0 (all jobs):
   READ: io=171929MB, aggrb=71144KB/s, minb=71144KB/s, maxb=71144KB/s,
mint=2474596msec, maxt=2474596msec
   WRITE: io=73832MB, aggrb=30551KB/s, minb=30551KB/s, maxb=30551KB/s,
mint=2474596msec, maxt=2474596msec


Disk stats (read/write):
    md0: ios=2750989/1181257, merge=0/0, ticks=0/0, in_queue=0, util=0.00%,
aggrios=913305/786114,    aggrmerge=3428/1312,    aggrticks=73819260/63246560,
aggrin_queue=137071953, aggrutil=91.11%
    sdc:  ios=1384370/1178211,  merge=7328/2854,  ticks=127597484/110271136,
in_queue=237877756, util=91.11%
    sdd:  ios=1355546/1180133,  merge=2958/1084,  ticks=93860296/79468544,
in_queue=173338104, util=86.47%
  sde: ios=0/0, merge=0/0, ticks=0/0, in_queue=0, util=0.00%

## 2) 20G + 24 threads

## Read

bj-ha-5:/opt/zhilong # cat log-read-20G*
test-read-20G: (g=0): rw=read, bs=64K-64K/64K-64K/64K-64K, ioengine=libaio, iodepth=64
...
fio-2.1.10
Starting 24 processes


test-read-20G: (groupid=0, jobs=24): err= 0: pid=23591: Mon Sep 28 08:45:21 2015
  read : io=491520MB, bw=114203KB/s, iops=1784, runt=4407223msec
    slat (usec): min=3, max=12909, avg=14.64, stdev=11.06
    clat (msec): min=11, max=7482, avg=806.57, stdev=205.17
     lat (msec): min=11, max=7482, avg=806.59, stdev=205.17
    clat percentiles (msec):
     |  1.00th=[  322],  5.00th=[  392], 10.00th=[  392], 20.00th=[  685],
     | 30.00th=[  758], 40.00th=[  832], 50.00th=[  865], 60.00th=[  889],
     | 70.00th=[  938], 80.00th=[  971], 90.00th=[  996], 95.00th=[ 1012],
     | 99.00th=[ 1057], 99.50th=[ 1205], 99.90th=[ 1713], 99.95th=[ 1975],
     | 99.99th=[ 2311]
    bw (KB  /s): min=   78, max=82944, per=4.55%, avg=5196.46, stdev=1559.70
    lat (msec) : 20=0.01%, 50=0.02%, 100=0.01%, 250=0.09%, 500=12.71%
    lat (msec) : 750=16.69%, 1000=63.48%, 2000=6.98%, >=2000=0.02%
  cpu        : usr=0.03%, sys=0.16%, ctx=655094, majf=0, minf=13060
    IO    depths    :  1=0.1%,    2=0.1%,    4=0.1%,    8=0.1%,    16=0.1%,
32=0.1%, >=64=100.0%
     submit     :  0=0.0%,    4=100.0%,    8=0.0%,    16=0.0%,    32=0.0%,
64=0.0%, >=64=0.0%
     complete   :  0=0.0%,    4=100.0%,    8=0.0%,    16=0.0%,    32=0.0%,
64=0.1%, >=64=0.0%
     issued     : total=r=7864320/w=0/d=0, short=r=0/w=0/d=0
     latency    : target=0, window=0, percentile=100.00%, depth=64


Run status group 0 (all jobs):
        READ:      io=491520MB,      aggrb=114202KB/s,      minb=114202KB/s,
maxb=114202KB/s, mint=4407223msec, maxt=4407223msec

Disk stats (read/write):

    md0:   ios=7864320/0,   merge=0/0,   ticks=0/0,   in_queue=0,   util=0.00%,
aggrios=216612/0,           aggrmerge=2403853/0,           aggrticks=153367650/0,
aggrin_queue=153368130, aggrutil=100.00%

  sdc: ios=382104/0, merge=5490368/0, ticks=331003688/0, in_queue=331004856,
util=100.00%

  sdd: ios=267733/0, merge=1721192/0, ticks=129099264/0, in_queue=129099536,
util=100.00%

  sde: ios=0/0, merge=0/0, ticks=0/0, in_queue=0, util=0.00%
## start read 20G testing ##
        Mon Sep 28 07:31:53 EDT 2015
## finish read 20G testing ##
        Mon Sep 28 08:45:21 EDT 2015


## Write


bj-ha-5:/opt/zhilong # cat log-write-20G.txt
test-write-20G: (g=0): rw=write, bs=64K-64K/64K-64K/64K-64K, ioengine=libaio,
iodepth=64
...
fio-2.1.10
Starting 24 processes


test-write-20G: (groupid=0, jobs=24): err= 0: pid=25288: Mon Sep 28 11:25:09 2015
  write: io=491520MB, bw=53052KB/s, iops=828, runt=9487299msec
    slat (usec): min=5, max=2990.6K, avg=447.26, stdev=25982.45
    clat (msec): min=4, max=11870, avg=1850.93, stdev=449.22
     lat (msec): min=7, max=11870, avg=1851.37, stdev=449.73
    clat percentiles (msec):
     |  1.00th=[  644],  5.00th=[ 1188], 10.00th=[ 1582], 20.00th=[ 1713],
     | 30.00th=[ 1713], 40.00th=[ 1713], 50.00th=[ 1729], 60.00th=[ 1762],
     | 70.00th=[ 1844], 80.00th=[ 2008], 90.00th=[ 2343], 95.00th=[ 2802],
     | 99.00th=[ 3490], 99.50th=[ 3687], 99.90th=[ 4178], 99.95th=[ 4293],
     | 99.99th=[ 4817]
    bw (KB  /s): min=   21, max=28228, per=4.49%, avg=2384.04, stdev=977.75
    lat (msec) : 10=0.01%, 20=0.01%, 50=0.01%, 100=0.01%, 250=0.03%
        lat      (msec)     :      500=0.19%,      750=1.02%,      1000=1.40%,
2000=77.14%, >=2000=20.22%
  cpu        : usr=0.03%, sys=0.07%, ctx=488006, majf=0, minf=738
    IO    depths    :   1=0.1%,   2=0.1%,   4=0.1%,   8=0.1%,   16=0.1%,
32=0.1%, >=64=100.0%
       submit      :    0=0.0%,    4=100.0%,    8=0.0%,    16=0.0%,    32=0.0%,

64=0.0%, >=64=0.0%
         complete   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.1%, >=64=0.0%
    issued    : total=r=0/w=7864320/d=0, short=r=0/w=0/d=0
    latency   : target=0, window=0, percentile=100.00%, depth=64


Run status group 0 (all jobs):
  WRITE: io=491520MB, aggrb=53051KB/s, minb=53051KB/s, maxb=53051KB/s,
mint=9487299msec, maxt=9487299msec


Disk stats (read/write):
    md0:  ios=228/7864309,  merge=0/0,  ticks=0/0,  in_queue=0,  util=0.00%,
aggrios=46/334603,        aggrmerge=36/4904912,        aggrticks=12341/475926781,
aggrin_queue=475964705, aggrutil=100.00%
      sdc:    ios=120/501525,    merge=108/7357593,    ticks=22928/688069940,
in_queue=688142084, util=100.00%
      sdd:    ios=18/502285,    merge=2/7357144,    ticks=14096/739710404,
in_queue=739752032, util=100.00%
  sde: ios=0/0, merge=0/0, ticks=0/0, in_queue=0, util=0.00%
bj-ha-5:/opt/zhilong # cat time-write-20G.txt
## start write 20G testing ##
        Mon Sep 28 08:47:01 EDT 2015
## finish write 20G testing ##
        Mon Sep 28 11:25:09 EDT 2015


**Rand read**


bj-ha-5:/opt/zhilong # cat log-randread-20G.txt
test-randread-20G:    (g=0):    rw=randread,    bs=64K-64K/64K-64K/64K-64K,
ioengine=libaio, iodepth=64
...
fio-2.1.10
Starting 24 processes


test-randread-20G: (groupid=0, jobs=24): err= 0: pid=32299: Mon Sep 28 16:52:01
2015
  read : io=491520MB, bw=25798KB/s, iops=403, runt=19510211msec
    slat (usec): min=4, max=5687.2K, avg=59470.63, stdev=352918.06
    clat (msec): min=2, max=13149, avg=3749.69, stdev=1287.78

lat (msec): min=2, max=14024, avg=3809.16, stdev=1355.33
    clat percentiles (msec):
    |  1.00th=[  668],  5.00th=[ 1336], 10.00th=[ 1795], 20.00th=[ 2409],
    | 30.00th=[ 3425], 40.00th=[ 3818], 50.00th=[ 4015], 60.00th=[ 4228],
    | 70.00th=[ 4424], 80.00th=[ 4686], 90.00th=[ 5080], 95.00th=[ 5538],
    | 99.00th=[ 6456], 99.50th=[ 6783], 99.90th=[ 7767], 99.95th=[ 8291],
    | 99.99th=[ 9372]
    bw (KB  /s): min=    9, max=52254, per=4.85%, avg=1251.88, stdev=780.12
    lat (msec) : 4=0.01%, 10=0.01%, 20=0.01%, 50=0.02%, 100=0.02%
    lat (msec) : 250=0.06%, 500=0.15%, 750=1.10%, 1000=0.72%, 2000=11.23%
    lat (msec) : >=2000=86.67%
  cpu         : usr=0.01%, sys=0.05%, ctx=245542, majf=0, minf=13038
    IO    depths    :  1=0.1%,   2=0.1%,   4=0.1%,   8=0.1%,   16=0.1%,
32=0.1%, >=64=100.0%
      submit    :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.0%, >=64=0.0%
      complete    :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.1%, >=64=0.0%
    issued    : total=r=7864320/w=0/d=0, short=r=0/w=0/d=0
    latency    : target=0, window=0, percentile=100.00%, depth=64


Run status group 0 (all jobs):
   READ: io=491520MB, aggrb=25797KB/s,  minb=25797KB/s,  maxb=25797KB/s,
mint=19510211msec, maxt=19510211msec


Disk stats (read/write):
    md0:   ios=7864146/0,   merge=0/0,   ticks=0/0,   in_queue=0,   util=0.00%,
aggrios=2607704/0,            aggrmerge=12591/0,            aggrticks=1034380844/0,
aggrin_queue=1034381173, aggrutil=91.65%
  sdc: ios=3912683/0, merge=13330/0, ticks=1366947264/0, in_queue=1366945836,
util=89.07%
  sdd: ios=3910429/0, merge=24444/0, ticks=1736195268/0, in_queue=1736197684,
util=91.65%
  sde: ios=0/0, merge=0/0, ticks=0/0, in_queue=0, util=0.00%
bj-ha-5:/opt/zhilong # cat time-randread-20G.txt
## start randread 20G testing ##
        Mon Sep 28 11:26:49 EDT 2015
## finish randread 20G testing ##
        Mon Sep 28 16:52:01 EDT 2015

## Rand write

bj-ha-5:/opt/zhilong # cat log-randwrite-20G.txt
test-randwrite-20G: (g=0): rw=randwrite, bs=64K-64K/64K-64K/64K-64K, ioengine=libaio, iodepth=64
...
fio-2.1.10
Starting 24 processes


test-randwrite-20G: (groupid=0, jobs=24): err= 0: pid=6752: Tue Sep 29 00:26:48 2015
  write: io=491520MB, bw=18515KB/s, iops=289, runt=27184473msec
    slat (usec): min=9, max=93593K, avg=82815.37, stdev=448274.00
    clat (msec): min=13, max=143664, avg=5214.35, stdev=2980.79
     lat (msec): min=13, max=143664, avg=5297.17, stdev=3001.45
    clat percentiles (msec):
     |  1.00th=[ 1713],  5.00th=[ 2311], 10.00th=[ 3130], 20.00th=[ 4080],
     | 30.00th=[ 4490], 40.00th=[ 4752], 50.00th=[ 5014], 60.00th=[ 5276],
     | 70.00th=[ 5669], 80.00th=[ 6128], 90.00th=[ 6849], 95.00th=[ 7767],
     | 99.00th=[10421], 99.50th=[14091], 99.90th=[16712], 99.95th=[16712],
     | 99.99th=[16712]
    bw (KB  /s): min=    1, max=19401, per=4.66%, avg=863.10, stdev=352.01
    lat (msec) : 20=0.01%, 50=0.01%, 100=0.01%, 250=0.01%, 500=0.01%
    lat (msec) : 750=0.01%, 1000=0.01%, 2000=3.44%, >=2000=96.54%
  cpu          : usr=0.01%, sys=0.04%, ctx=514615, majf=0, minf=693
    IO    depths    : 1=0.1%, 2=0.1%, 4=0.1%, 8=0.1%, 16=0.1%, 32=0.1%, >=64=100.0%
       submit    : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%, >=64=0.0%
       complete  : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.1%, >=64=0.0%
    issued    : total=r=0/w=7864320/d=0, short=r=0/w=0/d=0
    latency   : target=0, window=0, percentile=100.00%, depth=64


Run status group 0 (all jobs):
  WRITE: io=491520MB, aggrb=18514KB/s, minb=18514KB/s, maxb=18514KB/s, mint=27184473msec, maxt=27184473msec


Disk stats (read/write):
    md0:  ios=228/7864196,  merge=0/0,  ticks=0/0,  in_queue=0,  util=0.00%,

aggrios=52/5227820, aggrmerge=36/13762, aggrticks=62312/70116317, aggrin_queue=70239010, aggrutil=100.00%

sdc: ios=140/7821461, merge=108/38538, ticks=92076/4075585868, in_queue=4075850040, util=100.00%

sdd: ios=16/7862000, merge=2/2750, ticks=94860/429730380, in_queue=429834288, util=70.85%

sde: ios=0/0, merge=0/0, ticks=0/0, in_queue=0, util=0.00%

bj-ha-5:/opt/zhilong # cat time-randwrite-20G.txt

## start randwrite 20G testing ##

Mon Sep 28 16:53:41 EDT 2015

## finish randwrite 20G testing ##

Tue Sep 29 00:26:48 EDT 2015


## rw-mix-read


downgrade the priority, and kill it due to not enough time,


# 5.1.3 IOzone, ext4 format on local raid1


```
# ./iozone -i 0 -i 1 -i 2 -i 3 -i 4 -i 5 -i 6 -i 7 -i 8 -f /home/test/iozone.tmpfile -Rab /test-iozone.xls -g 8G -n 4G -C
        Iozone: Performance Test of File I/O
                Version $Revision: 3.430 $
                Compiled for 64 bit mode.
                Build: linux

        Contributors:William Norcott, Don Capps, Isom Crawford, Kirby Collins
                Al Slater, Scott Rhine, Mike Wisner, Ken Goss
                Steve Landherr, Brad Smith, Mark Kelly, Dr. Alain CYR,
                Randy Dunlap, Mark Montague, Dan Million, Gavin Brebner,
                Jean-Marc Zucconi, Jeff Blomberg, Benny Halevy, Dave Boone,
                Erik Habbinga, Kris Strecker, Walter Wong, Joshua Root,
                Fabrice Bacchella, Zhenghua Xue, Qin Li, Darren Sawyer,
                Vangel Bojaxhi, Ben England, Vikentsi Lapa.

        Run began: Tue Sep 29 04:21:08 2015

        Excel chart generation enabled
        Auto Mode
        Using maximum file size of 8388608 kilobytes.
        Using minimum file size of 4194304 kilobytes.
        Command line used: ./iozone -i 0 -i 1 -i 2 -i 3 -i 4 -i 5 -i 6 -i 7 -i 8 -f /home/test/iozone.tmpfile -Rab /test-iozone.xls -g 8G -n 4G -C
        Output is in kBytes/sec
        Time Resolution = 0.000001 seconds.
        Processor cache size set to 1024 kBytes.
        Processor cache line size set to 32 bytes.
        File stride size set to 17 * record size.
                                                        random   random    bkwd    record   stride
              kB  reclen    write  rewrite     read    reread    read    write    read   rewrite     read    fwrite frewrite    fread   freread
         4194304      64  2631700  4063515  8130449  8055329  7850882  4602944  7931318  8541605  8095549  3888500  3774905  8104059  6912089
         4194304     128  2599403  3983341  7962427  8390085  7847139  4813498  7909036  9504687  8888777  4121449  3910867  7963681  8353658
         4194304     256  2605620  4102546  7351191  8123519  7495679  4713755  7353034  8286729  7448225  3799405  3738858  7343032  7366144
         4194304     512  2659410  4127037  7386768  8256734  7421918  4754325  7410444  8013864  7566070  3929414  3749239  7445631  7517212
         4194304    1024  2686711  4151875  7475959  7484645  7646931  4793649  7730169  8181423  7587571  3365715  3367369  7519719  8367040
         4194304    2048  2690025  4192727  7549945  8420691  7637589  4761044  7534863  8244623  7608149  3082342  3043410  7464504  7567163
         4194304    4096  2684222  4139823  7614861  7932115  7631268  4532353  7565565  7578331  7626808  2935627  2892248  7585170  7608110
         4194304    8192  2495810  3757091  5840486  5987021  6329800  4209509  5914576  5280476  5904510  2869334  2847507  5909734  5993079
         4194304   16384  2428854  3617391  4993236  5005637  4992803  4042099  4972265  4326531  4978460  2857569  2823354  4956774  5123777
         8388608      64   361560   367596  8472872  8997589  8290048   190662  8387248  9298451  8768118   340854   364664  8492152  8464022
         8388608     128   362992   368763  8373352  8578784  8438819   235285  8395073  9707632  8885737   363633   370772  8421569  8425114
         8388608     256   354848   367809  7710017  7899739  7697975   261614  7721422  8099792  7671850   366827   368049  7713229  7766530
         8388608     512   359943   366954  7796907  8237903  7924733   280900  7771864  7977031  7915086   366577   365622  7784651  7820073
         8388608    1024   363003   368856  7880713  8097721  8026892   301096  7927872  7966131  8039415   366661   365003  7925700  7928703
         8388608    2048   361908   372107  7988906  8157085  8029667   318899  7953864  8147442  8135094   365680   366405  7961429  7980607
         8388608    4096   362037   366475  7988108  8128946  8018819   338158  7969688  8254024  8053671   361749   362611  7890943  8261926
         8388608    8192   363402   367754  6039793  6311448  6099541   341651  6086807  5166633  6087228   364657   365334  6056065  6086605
         8388608   16384   359135   366503  5064627  5171388  5088690   354730  5133557  4302569  5102665  361872   366315  5105416  5095137
iozone test complete.
Excel output is below:

"Writer report"
           "4"   "8"   "16"   "32"   "64"   "128"   "256"   "512"   "1024"   "2048"   "4096"   "8192"   "16384"
"4194304"   0   0   0   0  2631700  2599403  2605620  2659410  2686711  2690025  2684222  2495810  2428854
"8388608"   0   0   0   0  361560  362992  354848  359943  363003  361908  362037  363402  359135

"Re-writer report"
           "4"   "8"   "16"   "32"   "64"   "128"   "256"   "512"   "1024"   "2048"   "4096"   "8192"   "16384"
```

Please refer to the detail log named "local-raid1-iozone.txt"

5.2 cluster-md performance

5.2.1 dd include of read, write and rd/wr.

READ

```
#  time dd iflag=direct,nonblock  if=/dev/md0 of=/dev/null bs=1M count=40960
40960+0 records in
40960+0 records out
42949672960 bytes (43 GB) copied, 415.498 s, 103 MB/s

real    6m55.526s
user    0m0.060s
sys     0m3.776s
```

Write

```
#  time dd oflag=direct,nonblock if=/dev/zero of=/dev/md0 bs=1M count=40960
40960+0 records in
40960+0 records out
42949672960 bytes (43 GB) copied, 1502.29 s, 28.6 MB/s


real    25m2.294s
user    0m0.044s
sys     0m10.068s
```

Read/Write

```
#    time    dd    iflag=direct,nonblock    oflag=direct,nonblock    if=/dev/md0
of=/home/test/iotest bs=1M count=40960
40960+0 records in
40960+0 records out
42949672960 bytes (43 GB) copied, 653.715 s, 65.7 MB/s


real    10m53.731s
user    0m0.144s
```

sys   0m10.680s

5.2.2 FIO testing,

operation   contains   "read",   "write",   "randread",   "randwrite"   and
"rw-mix-read",   threads=(num of processor) x 2, size= 10G, 20G, 30G

1)   10G + 24 threads

Read

bj-ha-5:/opt/md-test # cat log-read-10G.txt
test-read: (g=0):   rw=read,   bs=64K-64K/64K-64K/64K-64K,   ioengine=libaio,
iodepth=64
...
fio-2.1.10
Starting 24 processes

test-read: (groupid=0, jobs=24): err= 0: pid=10009: Sun Sep 27 08:05:03 2015
  read : io=245760MB, bw=111973KB/s, iops=1749, runt=2247486msec
    slat (usec): min=3, max=3318, avg=11.85, stdev= 8.72
    clat (msec): min=12, max=54856, avg=789.86, stdev=434.81
     lat (msec): min=12, max=54856, avg=789.87, stdev=434.81
    clat percentiles (msec):
     | 1.00th=[322], 5.00th=[383], 10.00th=[465], 20.00th=[482],
     | 30.00th=[537], 40.00th=[660], 50.00th=[783], 60.00th=[922],
     | 70.00th=[1020], 80.00th=[1045], 90.00th=[1188], 95.00th=[1221],
     | 99.00th=[1270], 99.50th=[1319], 99.90th=[2008], 99.95th=[2311],
     | 99.99th=[7242]
    bw (KB /s): min=   38, max=115712, per=5.10%, avg=5715.55, stdev=2993.05
    lat (msec) : 20=0.01%, 50=0.16%, 100=0.41%, 250=0.23%, 500=22.17%
    lat (msec) : 750=25.38%, 1000=18.25%, 2000=33.30%, >=2000=0.10%
  cpu   : usr=0.02%, sys=0.13%, ctx=309249, majf=0, minf=13061
    IO   depths   :   1=0.1%,   2=0.1%,   4=0.1%,   8=0.1%,   16=0.1%,
32=0.1%, >=64=100.0%
      submit   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.0%, >=64=0.0%
      complete   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.1%, >=64=0.0%
    issued   : total=r=3932160/w=0/d=0, short=r=0/w=0/d=0

latency   : target=0, window=0, percentile=100.00%, depth=64

Run status group 0 (all jobs):
      READ: io=245760MB,   aggrb=111973KB/s,   minb=111973KB/s,
maxb=111973KB/s, mint=2247486msec, maxt=2247486msec

Disk stats (read/write):
    md0:   ios=3931962/0,   merge=0/0,   ticks=0/0,   in_queue=0,   util=0.00%,
aggrios=153637/0,   aggrmerge=1811960/0,   aggrticks=109535188/0,
aggrin_queue=109535582, aggrutil=100.00%
  sdc: ios=156969/0, merge=2132120/0, ticks=133679772/0, in_queue=133679816,
util=100.00%
  sdd: ios=150305/0, merge=1491800/0, ticks=85390604/0, in_queue=85391348,
util=99.99%

bj-ha-5:/opt/md-test # cat log-read-10G-time.txt
start read 10G testing
        Sun Sep 27 07:27:35 EDT 2015
finish read testing
        Sun Sep 27 08:05:03 EDT 2015

Write

bj-ha-5:/opt/md-test # cat log-write-10G*
start write 10G testing
        Sun Sep 27 08:06:43 EDT 2015
finish write testing
        Sun Sep 27 09:22:55 EDT 2015
test-write: (g=0): rw=write, bs=64K-64K/64K-64K/64K-64K, ioengine=libaio,
iodepth=64
...
fio-2.1.10
Starting 24 processes

test-write: (groupid=0, jobs=24): err= 0: pid=11026: Sun Sep 27 09:22:55 2015
  write: io=245760MB, bw=55056KB/s, iops=860, runt=4570953msec
    slat (usec): min=5, max=2565.2K, avg=452.02, stdev=25179.68
    clat (msec): min=2, max=14247, avg=1782.70, stdev=397.56
     lat (msec): min=2, max=14247, avg=1783.15, stdev=398.12

clat percentiles (msec):
| 1.00th=[635], 5.00th=[1221], 10.00th=[1565], 20.00th=[1713],
| 30.00th=[1713], 40.00th=[1713], 50.00th=[1713], 60.00th=[1713],
| 70.00th=[1729], 80.00th=[1745], 90.00th=[2212], 95.00th=[2606],
| 99.00th=[3228], 99.50th=[3425], 99.90th=[3884], 99.95th=[3982],
| 99.99th=[7504]
bw (KB /s): min=   24, max=38740, per=4.45%, avg=2448.73, stdev=1000.74
lat (msec) : 4=0.01%, 10=0.01%, 20=0.01%, 50=0.01%, 100=0.01%
lat (msec) : 250=0.03%, 500=0.23%, 750=0.93%, 1000=1.14%, 2000=83.76%
lat (msec) : >=2000=13.88%
cpu   : usr=0.03%, sys=0.06%, ctx=245931, majf=0, minf=738
IO   depths   : 1=0.1%,   2=0.1%,   4=0.1%,   8=0.1%,   16=0.1%,
32=0.1%, >=64=100.0%
submit   : 0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.0%, >=64=0.0%
complete : 0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.1%, >=64=0.0%
issued   : total=r=0/w=3932160/d=0, short=r=0/w=0/d=0
latency   : target=0, window=0, percentile=100.00%, depth=64

Run status group 0 (all jobs):
  WRITE: io=245760MB, aggrb=55055KB/s, minb=55055KB/s, maxb=55055KB/s,
mint=4570953msec, maxt=4570953msec

Disk stats (read/write):
    md0: ios=342/3932055, merge=0/0, ticks=0/0, in_queue=0, util=0.00%,
aggrios=130/252012,   aggrmerge=110/3676263,   aggrticks=14862/346463136,
aggrin_queue=346536780, aggrutil=100.00%
    sdc: ios=222/251837,   merge=153/3676295,   ticks=16816/324810208,
in_queue=324862116, util=100.00%
    sdd: ios=38/252188,   merge=67/3676231,   ticks=12908/368116064,
in_queue=368211444, util=100.00%

rand read

bj-ha-5:/opt/md-test # cat log-randread-10G*
start randread 10G testing
        Sun Sep 27 09:24:35 EDT 2015
finish randread testing

Sun Sep 27 10:01:08 EDT 2015
test-randread: (g=0): rw=randread, bs=64K-64K/64K-64K/64K-64K, ioengine=libaio, iodepth=64
...
fio-2.1.10
Starting 24 processes

test-randread: (groupid=0, jobs=24): err= 0: pid=12837: Sun Sep 27 10:01:07 2015
  read : io=245760MB, bw=114799KB/s, iops=1793, runt=2192170msec
    slat (usec): min=4, max=1000.6K, avg=13323.72, stdev=75908.81
    clat (usec): min=855, max=2765.7K, avg=841172.22, stdev=171555.84
     lat (usec): min=881, max=2929.7K, avg=854496.35, stdev=190125.71
    clat percentiles (msec):
     | 1.00th=[408], 5.00th=[445], 10.00th=[461], 20.00th=[865],
     | 30.00th=[881], 40.00th=[881], 50.00th=[889], 60.00th=[889],
     | 70.00th=[898], 80.00th=[906], 90.00th=[922], 95.00th=[971],
     | 99.00th=[1237], 99.50th=[1336], 99.90th=[1582], 99.95th=[1680],
     | 99.99th=[1909]
    bw (KB /s): min=   74, max=66816, per=4.19%, avg=4809.40, stdev=915.76
    lat (usec) : 1000=0.01%
    lat (msec) : 4=0.01%, 10=0.01%, 20=0.03%, 50=0.07%, 100=0.08%
    lat (msec) : 250=0.27%, 500=11.25%, 750=2.61%, 1000=81.69%, 2000=3.98%
    lat (msec) : >=2000=0.01%
  cpu   : usr=0.03%, sys=0.18%, ctx=126611, majf=0, minf=13035
  IO depths   : 1=0.1%, 2=0.1%, 4=0.1%, 8=0.1%, 16=0.1%, 32=0.1%, >=64=100.0%
     submit   : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%, >=64=0.0%
     complete : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.1%, >=64=0.0%
     issued   : total=r=3932160/w=0/d=0, short=r=0/w=0/d=0
     latency   : target=0, window=0, percentile=100.00%, depth=64

Run status group 0 (all jobs):
   READ: io=245760MB, aggrb=114798KB/s, minb=114798KB/s, maxb=114798KB/s, mint=2192170msec, maxt=2192170msec

Disk stats (read/write):
  md0: ios=3932160/0, merge=0/0, ticks=0/0, in_queue=0, util=0.00%, aggrios=1949821/0, aggrmerge=14565/0, aggrticks=174918046/0, aggrin_queue=174917166, aggrutil=89.92%

sdc: ios=1951725/0, merge=16378/0, ticks=181188636/0, in_queue=181187852, util=89.92%

sdd: ios=1947918/0, merge=12752/0, ticks=168647456/0, in_queue=168646480, util=88.81%

rand write

bj-ha-5:/opt/md-test # cat log-randwrite-10G*
start randwrite 10G testing
        Sun Sep 27 10:02:48 EDT 2015
finish randwrite testing
        Sun Sep 27 12:12:08 EDT 2015
test-randwrite: (g=0):   rw=randwrite,   bs=64K-64K/64K-64K/64K-64K, ioengine=libaio, iodepth=64
...
fio-2.1.10
Starting 24 processes

test-randwrite: (groupid=0, jobs=24): err= 0: pid=13736: Sun Sep 27 12:12:08 2015
  write: io=245760MB, bw=32431KB/s, iops=506, runt=7759705msec
    slat (usec): min=7, max=65966K, avg=47273.16, stdev=350413.40
    clat (msec): min=6, max=76582, avg=2974.12, stdev=2578.05
     lat (msec): min=6, max=76582, avg=3021.40, stdev=2598.15
    clat percentiles (msec):
     | 1.00th=[1303], 5.00th=[1713], 10.00th=[1713], 20.00th=[1827],
     | 30.00th=[2180], 40.00th=[2409], 50.00th=[2638], 60.00th=[2966],
     | 70.00th=[3326], 80.00th=[3687], 90.00th=[4146], 95.00th=[4621],
     | 99.00th=[6259], 99.50th=[11731], 99.90th=[16712], 99.95th=[16712],
     | 99.99th=[16712]
    bw (KB /s): min=   0, max=26368, per=4.60%, avg=1491.43, stdev=513.74
    lat (msec) : 10=0.01%, 20=0.01%, 50=0.01%, 100=0.01%, 250=0.01%
        lat   (msec)   :   500=0.02%,   750=0.02%,   1000=0.04%, 2000=23.35%, >=2000=76.55%
  cpu   : usr=0.02%, sys=0.06%, ctx=252496, majf=0, minf=709
    IO   depths   :   1=0.1%,   2=0.1%,   4=0.1%,   8=0.1%,   16=0.1%, 32=0.1%, >=64=100.0%
        submit   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%, 64=0.0%, >=64=0.0%
        complete   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%, 64=0.1%, >=64=0.0%
      issued   : total=r=0/w=3932160/d=0, short=r=0/w=0/d=0
      latency   : target=0, window=0, percentile=100.00%, depth=64

Run status group 0 (all jobs):
  WRITE: io=245760MB, aggrb=32431KB/s, minb=32431KB/s, maxb=32431KB/s, mint=7759705msec, maxt=7759705msec

Disk stats (read/write):
    md0: ios=342/3932068, merge=0/0, ticks=0/0, in_queue=0, util=0.00%, aggrios=96/3915771,   aggrmerge=82/14168,   aggrticks=38308/633055666, aggrin_queue=633110144, aggrutil=100.00%
    sdc: ios=170/3899883,   merge=100/27732,   ticks=12832/1149929820, in_queue=1149976416, util=100.00%
    sdd: ios=23/3931660,   merge=65/604,   ticks=63784/116181512, in_queue=116243872, util=73.10%

rw-mix-read

bj-ha-5:/opt/md-test # cat log-randrw\ -rwmixread\=70-10G.txt
test-randrw: (g=0): rw=randrw, bs=64K-64K/64K-64K/64K-64K, ioengine=libaio, iodepth=64
...
fio-2.1.10
Starting 24 processes

test-randrw: (groupid=0, jobs=24): err= 0: pid=16386: Sun Sep 27 12:57:42 2015
  read : io=171929MB, bw=66869KB/s, iops=1044, runt=2632840msec
    slat (usec): min=4, max=42828K, avg=9934.49, stdev=171992.68
    clat (msec): min=1, max=44439, avg=959.18, stdev=1762.23
     lat (msec): min=1, max=44439, avg=969.11, stdev=1771.46
    clat percentiles (msec):
     | 1.00th=[314], 5.00th=[537], 10.00th=[562], 20.00th=[586],
     | 30.00th=[611], 40.00th=[742], 50.00th=[840], 60.00th=[873],
     | 70.00th=[914], 80.00th=[1020], 90.00th=[1287], 95.00th=[1467],
     | 99.00th=[2089], 99.50th=[5473], 99.90th=[16712], 99.95th=[16712],
     | 99.99th=[16712]
    bw (KB /s): min=   4, max=19456, per=4.87%, avg=3253.42, stdev=1001.66
  write: io=73832MB, bw=28716KB/s, iops=448, runt=2632840msec
    slat (usec): min=7, max=42809K, avg=29881.78, stdev=322523.42
    clat (msec): min=7, max=44439, avg=1102.48, stdev=1930.27

lat (msec): min=7, max=44764, avg=1132.36, stdev=1961.98
clat percentiles (msec):
| 1.00th=[510], 5.00th=[570], 10.00th=[594], 20.00th=[791],
| 30.00th=[840], 40.00th=[865], 50.00th=[889], 60.00th=[922],
| 70.00th=[1020], 80.00th=[1221], 90.00th=[1434], 95.00th=[1582],
| 99.00th=[2212], 99.50th=[8094], 99.90th=[16712], 99.95th=[16712],
| 99.99th=[16712]
bw (KB /s): min=   4, max= 7787, per=4.88%, avg=1399.94, stdev=532.30
lat (msec) : 2=0.01%, 4=0.01%, 10=0.01%, 20=0.02%, 50=0.07%
lat (msec) : 100=0.05%, 250=0.14%, 500=1.50%, 750=31.91%, 1000=42.05%
lat (msec) : 2000=23.00%, >=2000=1.25%
cpu   : usr=0.03%, sys=0.15%, ctx=166994, majf=0, minf=731
   IO   depths   :   1=0.1%,   2=0.1%,   4=0.1%,   8=0.1%,   16=0.1%,
32=0.1%, >=64=100.0%
      submit   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.0%, >=64=0.0%
      complete   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.1%, >=64=0.0%
   issued   : total=r=2750856/w=1181304/d=0, short=r=0/w=0/d=0
   latency   : target=0, window=0, percentile=100.00%, depth=64

Run status group 0 (all jobs):
  READ: io=171929MB, aggrb=66868KB/s, minb=66868KB/s, maxb=66868KB/s,
mint=2632840msec, maxt=2632840msec
  WRITE: io=73832MB, aggrb=28715KB/s, minb=28715KB/s, maxb=28715KB/s,
mint=2632840msec, maxt=2632840msec

Disk stats (read/write):
      md0:   ios=2751175/1181294,   merge=0/0,   ticks=0/0,   in_queue=0,
util=0.00%,   aggrios=1369876/1179133,
aggrmerge=5280/2047, aggrticks=119884958/102992140, aggrin_queue=222885698,
aggrutil=90.77%
   sdc: ios=1393493/1180027,   merge=3615/1236,   ticks=100117684/83900596,
in_queue=184026136, util=86.82%
   sdd: ios=1346260/1178239,   merge=6945/2859,   ticks=139652232/122083684,
in_queue=261745260, util=90.77%

2) 20G + 24 threads

Read

bj-ha-5:/opt/md-test # cat log-read-20G*
start read 20G testing
        Sun Sep 27 13:01:02 EDT 2015
finish read testing
        Sun Sep 27 14:14:13 EDT 2015
test-read: (g=0): rw=read, bs=64K-64K/64K-64K/64K-64K, ioengine=libaio,
iodepth=64
...
fio-2.1.10
Starting 24 processes

test-read: (groupid=0, jobs=24): err= 0: pid=17526: Sun Sep 27 14:14:13 2015
  read : io=491520MB, bw=114623KB/s, iops=1790, runt=4391069msec
    slat (usec): min=3, max=1678, avg=11.61, stdev= 7.98
    clat (msec): min=6, max=5967, avg=773.39, stdev=265.39
     lat (msec): min=6, max=5967, avg=773.40, stdev=265.39
    clat percentiles (msec):
     | 1.00th=[359], 5.00th=[482], 10.00th=[545], 20.00th=[570],
     | 30.00th=[570], 40.00th=[570], 50.00th=[603], 60.00th=[848],
     | 70.00th=[1106], 80.00th=[1106], 90.00th=[1106], 95.00th=[1123],
     | 99.00th=[1156], 99.50th=[1188], 99.90th=[1680], 99.95th=[1844],
     | 99.99th=[2409]
    bw (KB /s): min=   50, max=115968, per=5.08%, avg=5818.71, stdev=2013.23
    lat (msec) : 10=0.01%, 20=0.01%, 50=0.05%, 100=0.01%, 250=0.04%
        lat   (msec)   :   500=7.81%,   750=50.82%,   1000=7.25%,
2000=33.99%, >=2000=0.03%
  cpu   : usr=0.02%, sys=0.13%, ctx=558859, majf=0, minf=13085
    IO   depths   :   1=0.1%,   2=0.1%,   4=0.1%,   8=0.1%,   16=0.1%,
32=0.1%, >=64=100.0%
      submit   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.0%, >=64=0.0%
      complete   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.1%, >=64=0.0%
    issued   : total=r=7864320/w=0/d=0, short=r=0/w=0/d=0
    latency   : target=0, window=0, percentile=100.00%, depth=64

Run status group 0 (all jobs):
    READ: io=491520MB,   aggrb=114622KB/s,   minb=114622KB/s,
maxb=114622KB/s, mint=4391069msec, maxt=4391069msec

Disk stats (read/write):
    md0:   ios=7864184/0,   merge=0/0,   ticks=0/0,   in_queue=0,   util=0.00%,
aggrios=275850/0,   aggrmerge=3655450/0,   aggrticks=202788660/0,
aggrin_queue=202791622, aggrutil=100.00%
  sdc: ios=301856/0, merge=4497783/0, ticks=265281564/0, in_queue=265286736,
util=100.00%
  sdd: ios=249844/0, merge=2813117/0, ticks=140295756/0, in_queue=140296508,
util=100.00%

Write

bj-ha-5:/opt/md-test # cat log-write-20G*
start write 20G testing
        Sun Sep 27 14:15:53 EDT 2015
finish write testing
        Sun Sep 27 16:53:16 EDT 2015
test-write: (g=0):   rw=write,   bs=64K-64K/64K-64K/64K-64K,   ioengine=libaio,
iodepth=64
...
fio-2.1.10
Starting 24 processes

test-write: (groupid=0, jobs=24): err= 0: pid=19034: Sun Sep 27 16:53:16 2015
  write: io=491520MB, bw=53307KB/s, iops=832, runt=9441863msec
    slat (usec): min=5, max=2951.8K, avg=420.14, stdev=24701.07
    clat (msec): min=5, max=13264, avg=1842.14, stdev=426.60
     lat (msec): min=5, max=13264, avg=1842.56, stdev=427.08
    clat percentiles (msec):
     | 1.00th=[627], 5.00th=[1237], 10.00th=[1614], 20.00th=[1713],
     | 30.00th=[1713], 40.00th=[1729], 50.00th=[1729], 60.00th=[1762],
     | 70.00th=[1844], 80.00th=[1991], 90.00th=[2311], 95.00th=[2737],
     | 99.00th=[3392], 99.50th=[3621], 99.90th=[4015], 99.95th=[4178],
     | 99.99th=[4555]
    bw (KB /s): min=   21, max=39071, per=4.49%, avg=2393.26, stdev=1001.09
    lat (msec) : 10=0.01%, 20=0.01%, 50=0.01%, 100=0.01%, 250=0.03%
        lat   (msec)   :   500=0.21%,   750=1.12%,   1000=1.18%,

2000=78.35%, >=2000=19.11%
  cpu   : usr=0.02%, sys=0.06%, ctx=488248, majf=0, minf=738
    IO   depths   :   1=0.1%,   2=0.1%,   4=0.1%,   8=0.1%,   16=0.1%,
32=0.1%, >=64=100.0%
      submit   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.0%, >=64=0.0%
      complete   :   0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.1%, >=64=0.0%
    issued   : total=r=0/w=7864320/d=0, short=r=0/w=0/d=0
    latency   : target=0, window=0, percentile=100.00%, depth=64

Run status group 0 (all jobs):
  WRITE: io=491520MB, aggrb=53306KB/s, minb=53306KB/s, maxb=53306KB/s,
mint=9441863msec, maxt=9441863msec

Disk stats (read/write):
    md0:   ios=342/7864288,   merge=0/0,   ticks=0/0,   in_queue=0,   util=0.00%,
aggrios=96/501910,   aggrmerge=82/7357419,   aggrticks=12914/715405288,
aggrin_queue=715469476, aggrutil=100.00%
    sdc: ios=155/504445,   merge=100/7354826,   ticks=7080/670239748,
in_queue=670285076, util=100.00%
    sdd: ios=38/499376,   merge=65/7360013,   ticks=18748/760570828,
in_queue=760653876, util=100.00%

randread

bj-ha-5:/opt/md-test # cat log-randread-20G*
start randread 20G testing
        Sun Sep 27 16:54:56 EDT 2015
finish randread testing
        Sun Sep 27 22:20:57 EDT 2015
test-randread: (g=0): rw=randread, bs=64K-64K/64K-64K/64K-64K, ioengine=libaio,
iodepth=64
...
fio-2.1.10
Starting 24 processes

test-randread: (groupid=0, jobs=24): err= 0: pid=22344: Sun Sep 27 22:20:56 2015
  read : io=491520MB, bw=25734KB/s, iops=402, runt=19558360msec

slat (usec): min=4, max=6715.2K, avg=59622.64, stdev=351241.15
clat (msec): min=2, max=14430, avg=3758.68, stdev=1200.36
lat (msec): min=2, max=14430, avg=3818.30, stdev=1270.15
clat percentiles (msec):
| 1.00th=[709], 5.00th=[1467], 10.00th=[1876], 20.00th=[2573],
| 30.00th=[3490], 40.00th=[3752], 50.00th=[3982], 60.00th=[4228],
| 70.00th=[4490], 80.00th=[4752], 90.00th=[5080], 95.00th=[5342],
| 99.00th=[5800], 99.50th=[6194], 99.90th=[7439], 99.95th=[7832],
| 99.99th=[8979]
bw (KB /s): min= 9, max=38730, per=4.73%, avg=1216.22, stdev=703.65
lat (msec) : 4=0.01%, 10=0.01%, 20=0.01%, 50=0.03%, 100=0.02%
lat (msec) : 250=0.06%, 500=0.15%, 750=0.91%, 1000=0.73%, 2000=10.24%
lat (msec) : >=2000=87.86%
cpu : usr=0.01%, sys=0.04%, ctx=246275, majf=0, minf=13039
IO depths : 1=0.1%, 2=0.1%, 4=0.1%, 8=0.1%, 16=0.1%, 32=0.1%, >=64=100.0%
submit : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.0%, >=64=0.0%
complete : 0=0.0%, 4=100.0%, 8=0.0%, 16=0.0%, 32=0.0%, 64=0.1%, >=64=0.0%
issued : total=r=7864320/w=0/d=0, short=r=0/w=0/d=0
latency : target=0, window=0, percentile=100.00%, depth=64

Run status group 0 (all jobs):
   READ: io=491520MB, aggrb=25734KB/s, minb=25734KB/s, maxb=25734KB/s, mint=19558360msec, maxt=19558360msec

Disk stats (read/write):
    md0: ios=7864320/0, merge=0/0, ticks=0/0, in_queue=0, util=0.00%, aggrios=3911329/0, aggrmerge=19125/0, aggrticks=1543710402/0, aggrin_queue=1543712000, aggrutil=90.27%
  sdc: ios=3909716/0, merge=22347/0, ticks=1551574116/0, in_queue=1551574980, util=90.27%
  sdd: ios=3912943/0, merge=15903/0, ticks=1535846688/0, in_queue=1535849020, util=90.20%

randwrite

bj-ha-5:/opt/md-test # cat log-randwrite-20G*

start randwrite 20G testing
        Sun Sep 27 22:22:37 EDT 2015
test-randwrite: (g=0):   rw=randwrite,   bs=64K-64K/64K-64K/64K-64K,
ioengine=libaio, iodepth=64
...
fio-2.1.10
Starting 24 processes

fio: terminating on signal 2

test-randwrite: (groupid=0, jobs=24): err= 0: pid=29005: Sun Sep 27 22:38:20 2015
  write: io=16983MB, bw=18432KB/s, iops=288, runt=943478msec
    slat (usec): min=11, max=81651K, avg=83287.29, stdev=1089145.02
    clat (msec): min=2, max=188687, avg=5139.77, stdev=13031.19
     lat (msec): min=257, max=188701, avg=5223.06, stdev=13143.76
    clat percentiles (msec):
     | 1.00th=[1713],   5.00th=[2089], 10.00th=[2376], 20.00th=[2802],
     | 30.00th=[3359], 40.00th=[3916], 50.00th=[4146], 60.00th=[4359],
     | 70.00th=[4686], 80.00th=[5014], 90.00th=[5407], 95.00th=[5997],
     | 99.00th=[15795], 99.50th=[16712], 99.90th=[16712], 99.95th=[16712],
     | 99.99th=[16712]
    bw (KB /s): min=   0, max= 3456, per=5.65%, avg=1040.82, stdev=458.72
    lat (msec) : 4=0.01%, 10=0.01%, 20=0.01%, 50=0.01%, 100=0.01%
    lat (msec) : 250=0.01%, 500=0.01%, 750=0.01%, 1000=0.01%, 2000=3.45%
    lat (msec) : >=2000=96.53%
  cpu   : usr=0.01%, sys=0.04%, ctx=25716, majf=0, minf=712
  IO depths   : 1=0.1%, 2=0.1%, 4=0.1%, 8=0.1%, 16=0.1%, 32=0.3%, >=64=99.4%
      submit   : 0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.0%, >=64=0.0%
      complete   : 0=0.0%,   4=100.0%,   8=0.0%,   16=0.0%,   32=0.0%,
64=0.1%, >=64=0.0%
      issued   : total=r=0/w=271722/d=0, short=r=0/w=0/d=0
      latency   : target=0, window=0, percentile=100.00%, depth=64

Run status group 0 (all jobs):
  WRITE: io=16983MB, aggrb=18432KB/s, minb=18432KB/s, maxb=18432KB/s,
mint=943478msec, maxt=943478msec

Disk stats (read/write):
    md0: ios=233/271722, merge=0/0, ticks=0/0, in_queue=0, util=0.00%,

aggrios=116/272170, aggrmerge=102/5, aggrticks=42380/70887820, aggrin_queue=70930136, aggrutil=100.00%

  sdc: ios=165/272221, merge=137/0, ticks=1648/14250316, in_queue=14252036, util=73.58%

  sdd: ios=68/272120, merge=67/11, ticks=83112/127525324, in_queue=127608236, util=100.00%

rw-mix-read

downgrade the priority, and kill it due to not enough time,

5.2.3 IOzone format ocfs2 FS on cluster-md

Please refer to the Excel spreadsheet named "test-md-iozone"

5.3 IOzone ocfs2 FS on CLVM

Please refer to the TXT and xls file name "test-clvm-iozone"