# Тестирование различных кейсов поведения ZFS ZVOL

## Подготовка теста

Итак, создаем mirrored pool ~220Gb на двух дисках:

```
zpool create -f vzpool mirror /dev/sdc /dev/sdd
```

Проверяем размер:

```
zpool list
NAME    SIZE  ALLOC   FREE   CAP  DEDUP  HEALTH  ALTROOT
vzpool   222G  1,09M   222G    0%  1.00x  ONLINE  -
```

## Как работает флаг --sparse? Можно ли без него выделить диск больше, чем есть на хранилище?

Попробуем создать раздел на 300 гб (больше размера пула) без флага -s:

```
zfs create -V 300G vzpool/ct101
cannot create 'vzpool/ct101': out of space
```

С флагом -s все срабатывает отлично:

```
zfs create -V 300G -s vzpool/ct101
```

## Выделяется ли место полностью при форматировании файловой системы?

Ок, теперь создаем диск на 200 гб и форматируем его в ext4:

```
zfs create -V 200G vzpool/ct101
parted -s /dev/vzpool/ct101 mklabel gpt
parted /dev/vzpool/ct101 "mkpart primary 1 -1"
mkfs.ext4 /dev/vzpool/ct101-part1
```

Смотрим состояние пула и zvol:

```
zpool list
NAME    SIZE  ALLOC   FREE   CAP  DEDUP  HEALTH  ALTROOT
vzpool   222G  3,31G   219G    1%  1.00x  ONLINE  -

zfs list
NAME         USED  AVAIL  REFER  MOUNTPOINT
vzpool       206G  12,2G   272K  /vzpool
vzpool/ct101   206G   215G  3,31G  -
```

То есть, как можно заметить реально была выделена только область метаданных ext4, все место явно выделено не было.

## Будут ли освобождены данные при удалении файлов внутри контейнера?

Теперь попробуем смонтировать фс и залит ьв нее 50 гб мусора:

```
mount /dev/vzpool/ct101-part1 /mnt
```

Ставим dd_rescue:

```
yum install -y dd_rescue pv
```

Выделяем 50 гб файл:

```
openssl enc -aes-256-ctr -pass pass:"$(dd if=/dev/urandom bs=128 count=1 2>/dev/null | base64)"
-nosalt < /dev/zero | pv -b > /mnt/temp_file_50_Gb
```

Использовать /dev/zero в данном случае нельзя, а /dev/urandom недостаточно быстр для быстрого заполнения.

Потом делаем sync и проверяем размер файла:

```
sync
ls -al /mnt/temp_file_50_Gb
-rw-r--r-- 1 root root 51G Июл  2 14:58 /mnt/temp_file_50_Gb
```

Обращаю внимание, что truncate для этого теста совершенно не подходит, нужно забивать файл реальными данными.

zvol устройство сразу сообщит о том, что занято более 50+ Gb:

```
zfs list
NAME        USED  AVAIL  REFER  MOUNTPOINT
vzpool      206G  12,2G   272K  /vzpool
vzpool/ct101   206G   155G  63,7G  -
```

zpool также сообщит об этом:

```
zpool list
NAME    SIZE  ALLOC   FREE   CAP  DEDUP  HEALTH  ALTROOT
vzpool   222G  63,7G   158G   28%  1.00x  ONLINE  -
```

Сносим 50Gb файл:

```
zfs list
NAME        USED  AVAIL  REFER  MOUNTPOINT
vzpool      206G  12,2G   272K  /vzpool
vzpool/ct101   206G   155G  63,7G  -
zpool list
NAME    SIZE  ALLOC   FREE   CAP  DEDUP  HEALTH  ALTROOT
vzpool   222G  63,7G   158G   28%  1.00x  ONLINE  -
```

Итог - место НЕ ОСВОБОДИЛОСЬ НА ПУЛЕ!!!!!!!!!! Оно оказалось безвозвратно потеряно.

Чтобы от этого избавится, нужно воспользоваться стандартной тулзой fstrim:

```
fstrim  /mnt
sync
```

После этого нужно подождать сброса всех буферов (видно по большому числу процессов ) место вернется:

```
zfs list
NAME         USED  AVAIL  REFER  MOUNTPOINT
vzpool       206G  12,2G   272K  /vzpool
vzpool/ct101   206G   215G  3,31G  -
```

```
zpool list
NAME    SIZE  ALLOC   FREE   CAP  DEDUP  HEALTH  ALTROOT
vzpool   222G  3,31G   219G    1%  1.00x  ONLINE  -
```

## Поведение при исчерпании пространства в пуле

Гайд: https://groups.google.com/a/zfsonlinux.org/forum/#!topic/zfs-discuss/RX9AFFOwTsM

Итак, при пуле в 220+ Gb создаем zvol в 300 Gb, без опции -s (sparsed) нам этого не сделать, так что сообщаем ее:

```
zfs create -s -V 300G vzpool/ct101
```

Форматируем:

```
parted -s /dev/vzpool/ct101 mklabel gpt
parted /dev/vzpool/ct101 "mkpart primary 1 -1"
mkfs.ext4 /dev/vzpool/ct101-part1
```

Монтируем:

```
mount /dev/vzpool/ct101-part1 /mnt
```

Смотрим размер:

```
df -h|grep '/mnt'
/dev/zd0p1     296G  191M  281G   1% /mnt
```

В то время как размер доступного дискового пространства вот такой:

```
vzpool        214G  256K  214G   1% /vzpool
```

Чистейший оверселл.

Начинаем забивать данными:

```
screen
openssl enc -aes-256-ctr -pass pass:"$(dd if=/dev/urandom bs=128 count=1 2>/dev/null | base64)"
-nosalt < /dev/zero | pv -b > /mnt/temp_file_unlimited_Gb
ctrl+a d
```

После этого ждем результатов, когда счетчик дойдет хотя бы до 214 Gb. Результаты в первую очередь в dmesg.

Вот, результат себя не заставил ждать долго:

```
Jul  2 17:48:26 zfs kernel: [20935.912818] EXT4-fs warning (device zd0p1): ext4_end_bio:
Jul  2 17:48:26 zfs kernel: [20935.912822] EXT4-fs warning (device zd0p1): ext4_end_bio:
Jul  2 17:48:26 zfs kernel: [20935.912826] EXT4-fs warning (device zd0p1): ext4_end_bio:
Jul  2 17:48:26 zfs kernel: [20935.912830] EXT4-fs warning (device zd0p1): ext4_end_bio:
Jul  2 17:48:26 zfs kernel: [20935.912835] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error
writing to inode 12 (size 0 starting block 58907648)
Jul  2 17:48:26 zfs kernel: [20935.912843] EXT4-fs warning (device zd0p1): ext4_end_bio:
Jul  2 17:48:26 zfs kernel: [20935.912848] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error
writing to inode 12 (size 0 starting block 58905600)I/O error writing to inode 12 (size 0 starting block
58906496)I/O error writing to inode 12 (size 0 starting block 58905472)
Jul  2 17:48:26 zfs kernel: [20935.912861] I/O error writing to inode 12 (size 0 starting block
58904960)I/O error writing to inode 12 (size 0 starting block 58905344)
Jul  2 17:48:26 zfs kernel: [20935.912868]
Jul  2 17:48:26 zfs kernel: [20935.912869]
Jul  2 17:48:26 zfs kernel: [20935.912872] Buffer I/O error on device zd0p1, logical block 58907264
Jul  2 17:48:26 zfs kernel: [20935.912876]
Jul  2 17:48:26 zfs kernel: [20935.912877]
Jul  2 17:48:26 zfs kernel: [20935.912879] Buffer I/O error on device zd0p1, logical block 58905216
Jul  2 17:48:26 zfs kernel: [20935.912882] Buffer I/O error on device zd0p1, logical block 58906112
Jul  2 17:48:26 zfs kernel: [20935.912886] Buffer I/O error on device zd0p1, logical block 58905088
Jul  2 17:48:26 zfs kernel: [20935.912897] lost page write due to I/O error on zd0p1
Jul  2 17:48:26 zfs kernel: [20935.912899] Buffer I/O error on device zd0p1, logical block 58904576
Jul  2 17:48:26 zfs kernel: [20935.912901] Buffer I/O error on device zd0p1, logical block 58904960
Jul  2 17:48:26 zfs kernel: [20935.912904] lost page write due to I/O error on zd0p1
Jul  2 17:48:26 zfs kernel: [20935.912906] lost page write due to I/O error on zd0p1
Jul  2 17:48:26 zfs kernel: [20935.912908] lost page write due to I/O error on zd0p1
Jul  2 17:48:26 zfs kernel: [20935.912910] Buffer I/O error on device zd0p1, logical block 58907265
Jul  2 17:48:26 zfs kernel: [20935.912917] EXT4-fs warning (device zd0p1): ext4_end_bio:
Jul  2 17:48:26 zfs kernel: [20935.913176] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error
writing to inode 12 (size 0 starting block 58905088)
Jul  2 17:48:26 zfs kernel: [20935.913366] end_request: I/O error, dev zd0, sector 471244800
Jul  2 17:48:26 zfs kernel: [20935.913369] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error
writing to inode 12 (size 0 starting block 58905728)
Jul  2 17:48:26 zfs kernel: [20935.913457] end_request: I/O error, dev zd0, sector 471268352
Jul  2 17:48:26 zfs kernel: [20935.913459] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error
writing to inode 12 (size 0 starting block 58908672)
Jul  2 17:48:26 zfs kernel: [20935.913464] end_request: I/O error, dev zd0, sector 471267328
Jul  2 17:48:26 zfs kernel: [20935.913466] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error
writing to inode 12 (size 0 starting block 58908544)
Jul  2 17:48:26 zfs kernel: [20935.913602] end_request: I/O error, dev zd0, sector 471273472
Jul  2 17:48:26 zfs kernel: [20935.913604] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error
writing to inode 12 (size 0 starting block 58909312)
Jul  2 17:48:26 zfs kernel: [20935.913622] end_request: I/O error, dev zd0, sector 471274496
Jul  2 17:48:26 zfs kernel: [20935.913624] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error
writing to inode 12 (size 0 starting block 58909440)
Jul  2 17:48:26 zfs kernel: [20935.913675] end_request: I/O error, dev zd0, sector 471196672
Jul  2 17:48:26 zfs kernel: [20935.913678] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error
writing to inode 12 (size 0 starting block 58899712)
```

Jul  2 17:48:26 zfs kernel: [20935.913946] end_request: I/O error, dev zd0, sector 471269376
Jul  2 17:48:26 zfs kernel: [20935.913948] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error writing to inode 12 (size 0 starting block 58908800)
Jul  2 17:48:26 zfs kernel: [20935.913987] end_request: I/O error, dev zd0, sector 471278592
Jul  2 17:48:26 zfs kernel: [20935.913989] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error writing to inode 12 (size 0 starting block 58909952)
Jul  2 17:48:26 zfs kernel: [20935.914048] end_request: I/O error, dev zd0, sector 471279616
Jul  2 17:48:26 zfs kernel: [20935.914051] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error writing to inode 12 (size 0 starting block 58910080)
Jul  2 17:48:26 zfs kernel: [20935.914063] end_request: I/O error, dev zd0, sector 471280640
Jul  2 17:48:26 zfs kernel: [20935.914066] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error writing to inode 12 (size 0 starting block 58910208)
Jul  2 17:48:26 zfs kernel: [20935.918637] I/O error writing to inode 12 (size 0 starting block 58906880)
Jul  2 17:48:26 zfs kernel: [20935.939194] end_request: I/O error, dev zd0, sector 471266304
Jul  2 17:48:26 zfs kernel: [20935.939197] end_request: I/O error, dev zd0, sector 471252992
Jul  2 17:48:26 zfs kernel: [20935.939199] end_request: I/O error, dev zd0, sector 471224320
Jul  2 17:48:26 zfs kernel: [20935.939203] end_request: I/O error, dev zd0, sector 471270400
Jul  2 17:48:26 zfs kernel: [20935.939206] end_request: I/O error, dev zd0, sector 471276544
Jul  2 17:48:26 zfs kernel: [20935.939209] end_request: I/O error, dev zd0, sector 471316480
Jul  2 17:48:26 zfs kernel: [20935.939212] end_request: I/O error, dev zd0, sector 471285760
Jul  2 17:48:26 zfs kernel: [20935.939214] EXT4-fs warning (device zd0p1): ext4_end_bio:
Jul  2 17:48:26 zfs kernel: [20935.939217] EXT4-fs warning (device zd0p1): ext4_end_bio:
Jul  2 17:48:26 zfs kernel: [20935.939219] EXT4-fs warning (device zd0p1): ext4_end_bio:
Jul  2 17:48:26 zfs kernel: [20935.939221] EXT4-fs warning (device zd0p1): ext4_end_bio:
Jul  2 17:48:26 zfs kernel: [20935.939223] EXT4-fs warning (device zd0p1): ext4_end_bio:
Jul  2 17:48:26 zfs kernel: [20935.939226] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error writing to inode 12 (size 0 starting block 58906752)I/O error writing to inode 12 (size 0 starting block 58903168)I/O error writing to inode 12 (size 0 starting block 58908928)I/O error writing to inode 12 (size 0 starting block 58909696)I/O error writing to inode 12 (size 0 starting block 58910848)
Jul  2 17:48:26 zfs kernel: [20935.939238] I/O error writing to inode 12 (size 0 starting block 58914688)

Jul  2 17:50:33 zfs kernel: [21062.997747] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error writing to inode 12 (size 0 starting block 32683)
Jul  2 17:50:33 zfs kernel: [21062.997794] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error writing to inode 12 (size 0 starting block 32804)
Jul  2 17:50:33 zfs kernel: [21062.997829] EXT4-fs warning (device zd0p1): ext4_end_bio: I/O error writing to inode 12 (size 0 starting block 32427)
Jul  2 17:50:39 zfs kernel: [21068.632693] ------------[ cut here ]------------
Jul  2 17:50:39 zfs kernel: [21068.632742] WARNING: at fs/buffer.c:1182 mark_buffer_dirty+0x82/0xa0() (Tainted: P           ---------------  )
Jul  2 17:50:39 zfs kernel: [21068.632821] Hardware name: System Product Name
Jul  2 17:50:39 zfs kernel: [21068.632859] Modules linked in: vzethdev pio_nfs pio_direct pfmt_raw pfmt_ploop1 ploop simfs vzrst nf_nat nf_conntrack_ipv4 nf_defrag_ipv4 vzcpt nfs lockd fscache auth_rpcgss nfs_acl sunrpc nf_conntrack vziolimit vzdquota ip6t_REJECT ip6table_mangle ip6table_filter ip6_tables xt_length xt_hl xt_tcpmss xt_TCPMSS iptable_mangle iptable_filter xt_multiport xt_limit xt_dscp ipt_REJECT ip_tables vzevent vznetdev vzmon vzdev ipv6 zfs(P)(U) zcommon(P)(U) znvpair(P)(U) zavl(P)(U) zunicode(P)(U) spl(U) zlib_deflate ext3 jbd cpufreq_ondemand acpi_cpufreq freq_table mperf iTCO_wdt iTCO_vendor_support i2c_i801 i2c_core sg lpc_ich mfd_core shpchp e1000e ptp pps_core ext4 jbd2 mbcache sd_mod crc_t10dif xhci_hcd video output ahci dm_mirror dm_region_hash dm_log dm_mod [last unloaded: scsi_wait_scan]
Jul  2 17:50:39 zfs kernel: [21068.633409] Pid: 4125, comm: jbd2/zd0p1-8 veid: 0 Tainted: P           ---------------  2.6.32-042stab090.5 #1
Jul  2 17:50:39 zfs kernel: [21068.633488] Call Trace:
Jul  2 17:50:39 zfs kernel: [21068.633528]  [<ffffffff81074f97>] ? warn_slowpath_common+0x87/0xc0
Jul  2 17:50:39 zfs kernel: [21068.633573]  [<ffffffff81074fea>] ? warn_slowpath_null+0x1a/0x20
Jul  2 17:50:39 zfs kernel: [21068.633617]  [<ffffffff811e5092>] ? mark_buffer_dirty+0x82/0xa0

```
Jul  2 17:50:39 zfs kernel: [21068.633666]  [<ffffffffa0090315>] ?
__jbd2_journal_temp_unlink_buffer+0xa5/0x140 [jbd2]
Jul  2 17:50:39 zfs kernel: [21068.633745]  [<ffffffffa00903c6>] ?
__jbd2_journal_unfile_buffer+0x16/0x30 [jbd2]
Jul  2 17:50:39 zfs kernel: [21068.633822]  [<ffffffffa0090728>] ?
__jbd2_journal_refile_buffer+0xc8/0xf0 [jbd2]
Jul  2 17:50:39 zfs kernel: [21068.633900]  [<ffffffffa00933a8>] ?
jbd2_journal_commit_transaction+0xd38/0x1500 [jbd2]
Jul  2 17:50:39 zfs kernel: [21068.633978]  [<ffffffff81009770>] ? __switch_to+0xd0/0x320
Jul  2 17:50:39 zfs kernel: [21068.634023]  [<ffffffff81088cdb>] ? try_to_del_timer_sync+0x7b/0xe0
Jul  2 17:50:39 zfs kernel: [21068.634073]  [<ffffffffa0098bb8>] ? kjournald2+0xb8/0x220 [jbd2]
Jul  2 17:50:39 zfs kernel: [21068.634118]  [<ffffffff810a1550>] ?
autoremove_wake_function+0x0/0x40
Jul  2 17:50:39 zfs kernel: [21068.634167]  [<ffffffffa0098b00>] ? kjournald2+0x0/0x220 [jbd2]
Jul  2 17:50:39 zfs kernel: [21068.634212]  [<ffffffff810a0f36>] ? kthread+0x96/0xa0
Jul  2 17:50:39 zfs kernel: [21068.634254]  [<ffffffff8100c34a>] ? child_rip+0xa/0x20
Jul  2 17:50:39 zfs kernel: [21068.634306]  [<ffffffff810a0ea0>] ? kthread+0x0/0xa0
Jul  2 17:50:39 zfs kernel: [21068.634350]  [<ffffffff8100c340>] ? child_rip+0x0/0x20
Jul  2 17:50:39 zfs kernel: [21068.634392] ---[ end trace 520ff3534d63f82f ]---
Jul  2 17:50:39 zfs kernel: [21068.634431] Tainting kernel with flag 0x9
Jul  2 17:50:39 zfs kernel: [21068.634470] Pid: 4125, comm: jbd2/zd0p1-8 veid: 0 Tainted: P
---------------    2.6.32-042stab090.5 #1
Jul  2 17:50:39 zfs kernel: [21068.634548] Call Trace:
Jul  2 17:50:39 zfs kernel: [21068.634585]  [<ffffffff81074e21>] ? add_taint+0x71/0x80
Jul  2 17:50:39 zfs kernel: [21068.634627]  [<ffffffff81074fa4>] ?
warn_slowpath_common+0x94/0xc0
Jul  2 17:50:39 zfs kernel: [21068.634672]  [<ffffffff81074fea>] ? warn_slowpath_null+0x1a/0x20
Jul  2 17:50:39 zfs kernel: [21068.634715]  [<ffffffff811e5092>] ? mark_buffer_dirty+0x82/0xa0
Jul  2 17:50:39 zfs kernel: [21068.634762]  [<ffffffffa0090315>] ?
__jbd2_journal_temp_unlink_buffer+0xa5/0x140 [jbd2]
Jul  2 17:50:39 zfs kernel: [21068.634841]  [<ffffffffa00903c6>] ?
__jbd2_journal_unfile_buffer+0x16/0x30 [jbd2]
Jul  2 17:50:39 zfs kernel: [21068.634918]  [<ffffffffa0090728>] ?
__jbd2_journal_refile_buffer+0xc8/0xf0 [jbd2]
Jul  2 17:50:39 zfs kernel: [21068.634996]  [<ffffffffa00933a8>] ?
jbd2_journal_commit_transaction+0xd38/0x1500 [jbd2]
Jul  2 17:50:39 zfs kernel: [21068.635072]  [<ffffffff81009770>] ? __switch_to+0xd0/0x320
Jul  2 17:50:39 zfs kernel: [21068.635116]  [<ffffffff81088cdb>] ? try_to_del_timer_sync+0x7b/0xe0
Jul  2 17:50:39 zfs kernel: [21068.635164]  [<ffffffffa0098bb8>] ? kjournald2+0xb8/0x220 [jbd2]
Jul  2 17:50:39 zfs kernel: [21068.635208]  [<ffffffff810a1550>] ?
autoremove_wake_function+0x0/0x40
Jul  2 17:50:39 zfs kernel: [21068.635256]  [<ffffffffa0098b00>] ? kjournald2+0x0/0x220 [jbd2]
Jul  2 17:50:39 zfs kernel: [21068.635305]  [<ffffffff810a0f36>] ? kthread+0x96/0xa0
Jul  2 17:50:39 zfs kernel: [21068.635348]  [<ffffffff8100c34a>] ? child_rip+0xa/0x20
Jul  2 17:50:39 zfs kernel: [21068.635392]  [<ffffffff810a0ea0>] ? kthread+0x0/0xa0
Jul  2 17:50:39 zfs kernel: [21068.635433]  [<ffffffff8100c340>] ? child_rip+0x0/0x20
Jul  2 17:51:13 zfs kernel: [21102.995512] __ratelimit: 628 callbacks suppressed
```

```
Jul  2 17:51:13 zfs kernel: [21102.995558] Buffer I/O error on device zd0p1, logical block 295937
Jul  2 17:51:13 zfs kernel: [21102.995601] lost page write due to I/O error on zd0p1
```

Подробный лог на 46 мегабайт сохранен в: zfs_ext4_overflow.log

TODO: почему ZFS ни слова не пишет в лог об окончании места в пуле при этом?

TODO: почему бы openvz не монтировать диски в режиме onerror=ro?

Все ресурсы ZFS были забиты:

```
[root@zfs ~]# zfs  list
NAME         USED  AVAIL  REFER  MOUNTPOINT
vzpool       219G     0   272K  /vzpool
vzpool/ct101   219G     0   219G  -

[root@zfs ~]# zpool list
NAME    SIZE  ALLOC   FREE    CAP  DEDUP  HEALTH  ALTROOT
vzpool   222G   219G  3,47G    98%  1.00x  ONLINE  -
```

После этого размонтируем файловую систему:

```
umount /mnt
```

При этом в dmesg вылетит следующее:

```
[21102.995558] Buffer I/O error on device zd0p1, logical block 295937
[21102.995601] lost page write due to I/O error on zd0p1
[32373.387442] Aborting journal on device zd0p1-8.
[32373.387622] EXT4-fs error (device zd0p1): ext4_put_super: Couldn't clean up the journal
[32373.387683] EXT4-fs (zd0p1): Remounting filesystem read-only
```

Пробуем смонтировать вновь:

```
LANG=C mount /dev/vzpool/ct101-part1 /mnt
mount: wrong fs type, bad option, bad superblock on /dev/zd0p1,
missing codepage or helper program, or other error
In some cases useful info is found in syslog - try
dmesg | tail or so
```

В dmesg следующее:

```
[root@zfs ~]# dmesg
[34997.800306] __ratelimit: 520 callbacks suppressed
[34997.800351] __ratelimit: 520 callbacks suppressed
[34997.800386] Buffer I/O error on device zd0p1, logical block 58720261
[34997.800388] Buffer I/O error on device zd0p1, logical block 59244555
[34997.800391] Buffer I/O error on device zd0p1, logical block 58720264
[34997.800393] Buffer I/O error on device zd0p1, logical block 59244558
[34997.800397] Buffer I/O error on device zd0p1, logical block 295937
[34997.800399] Buffer I/O error on device zd0p1, logical block 59244551
[34997.800402] lost page write due to I/O error on zd0p1
[34997.800404] Buffer I/O error on device zd0p1, logical block 59244548
[34997.800407] lost page write due to I/O error on zd0p1
[34997.800409] lost page write due to I/O error on zd0p1
[34997.800411] lost page write due to I/O error on zd0p1
[34997.800412] lost page write due to I/O error on zd0p1
[34997.800414] lost page write due to I/O error on zd0p1
[34997.800427] Buffer I/O error on device zd0p1, logical block 58720266
[34997.800430] end_request: I/O error, dev zd0, sector 478152768
[34997.800432] Buffer I/O error on device zd0p1, logical block 59768835
[34997.800435] Buffer I/O error on device zd0p1, logical block 59768839
[34997.800437] lost page write due to I/O error on zd0p1
[34997.800439] end_request: I/O error, dev zd0, sector 478152776
[34997.800442] end_request: I/O error, dev zd0, sector 478152784
[34997.800444] lost page write due to I/O error on zd0p1
[34997.800446] end_request: I/O error, dev zd0, sector 478152808
[34997.800448] lost page write due to I/O error on zd0p1
[34997.800455] end_request: I/O error, dev zd0, sector 478152816
[34997.803636] lost page write due to I/O error on zd0p1
[34998.083999] end_request: I/O error, dev zd0, sector 624953408
[34998.084505] JBD: recovery failed
[34998.084544] EXT4-fs (zd0p1): error loading journal
```

Если попробовать включить fsck, то в целом он пройдет успешно:

```
fsck.ext4 /dev/vzpool/ct101-part1 -n
e2fsck 1.41.12 (17-May-2010)
Warning: skipping journal recovery because doing a read-only filesystem check.
/dev/vzpool/ct101-part1 contains a file system with errors, check forced.
Pass 1: Checking inodes, blocks, and sizes
Inode 12 has an invalid extent node (blk 59344896, lblk 55537664)
Clear? no
Inode 12 has an invalid extent node (blk 71174144, lblk 66643968)
Clear? no
Inode 12 has an invalid extent node (blk 65042464, lblk 76466176)
Clear? no
Inode 12 has an invalid extent node (blk 295937, lblk 77351929)
Clear? no
Inode 12, i_blocks is 618846512, should be 444301360.  Fix? no
Pass 2: Checking directory structure
Pass 3: Checking directory connectivity
Pass 4: Checking reference counts
Pass 5: Checking group summary information
Block bitmap differences:  -(9255--32547) -(33824--34495) -(99329--100351) -(164866--165887)
-(230401--231423) -(295937--296959) -(532512--557055) -(820256--821247) -(885761--886783)
-(1056800--1081343) -(1581088--1605631) -(1606688--1606985) -(2105376--2129919)
```

-(2629664--2654207) -(2655264--2656255) -(3153952--3178495) -(3678240--3702783)
-(4097056--4098047) -(4202528--4227071) -(4726816--4751359) -(5251104--5275647)
-(5775392--5799935) -(6299680--6324223) -(6823968--6848511) -(7348256--7372799)
-(7872544--7897087) -(7963680--7964127) -(8396832--8421375) -(8921120--8945663)
-(9445408--9469951) -(9969696--9994239) -(10493984--10518527) -(11018272--11042815)
-(11240480--11240671) -(11542560--11567103) -(11896833--11898879) -(12066848--12091391)
-(12591136--12615679) -(13115424--13139967) -(13639712--13664255) -(14164000--14188543)
-(14688288--14712831) -(15212576--15237119) -(15736864--15761407) -(16261152--16285695)
-(16785440--16809983) -(17309728--17334271) -(17834016--17858559) -(18358304--18382847)
-(18882592--18907135) -(19406880--19431423) -(19931168--19955711) -(20455456--20479999)
-(20481056--20481632) -(20979744--21004287) -(21504032--21528575) -(22028320--22052863)
-(22552608--22577151) -(23076896--23101439) -(23601184--23625727) -(23758849--23760895)
-23888897 -(23888928--23889119) -(24125472--24150015) -(24649760--24674303)
-(25174048--25198591) -(25698336--25722879) -(26222624--26247167) -(26746912--26771455)
-(27271200--27295743) -(27795488--27820031) -(28319776--28344319) -(28844064--28868607)
-(29368352--29392895) -(29892640--29917183) -(30416928--30441471) -(30941216--30965759)
-(31465504--31490047) -(31989792--32014335) -(32514080--32538623) -(33038368--33062911)
-(33562656--33587199) -(34086944--34111487) -(34611232--34635775) -(35135520--35160063)
-(35588097--35590143) -(35659808--35684351) -(36184096--36208639) -(36708384--36732927)
-(37232672--37257215) -(37756960--37781503) -(38281248--38305791) -(38805536--38830079)
-(39329824--39354367) -(39854112--39878655) -(40378400--40402943) -(40902688--40927231)
-(41426976--41451519) -(41951264--41975807) -(42475552--42500095) -(42999840--43024383)
-(43524128--43548671) -(44048416--44072959) -(44572704--44597247) -(45096992--45121535)
-(45621280--45645823) -(46145568--46170111) -(46669856--46694399) -(47194144--47218687)
-(47482882--47484927) -(47718432--47742975) -(48242720--48267263) -(48767008--48791551)
-(49291296--49315839) -(49815584--49840127) -(50339872--50364415) -(50864160--50888703)
-(51388448--51412991) -(51912736--51937279) -(52437024--52461567) +(52494336--52953087)
-(52961312--52985855) +(53018624--53477375) -(53485600--53510143) +(53542912--54001663)
-(54009888--54034431) +(54067200--54525951) -(54534176--54558719) +(54591488--55050239)
-(55058464--55083007) +(55115776--55574527) -(55582752--55607295) +(55640064--56098815)
-(56107040--56131583) +(56164352--56623103) -(56631328--56655871) +(56688640--57147391)
-(57155616--57180159) +(57212928--57671679) -(57679904--57704447) +(57737216--58195967)
-(58204192--58228735) +(58261504--58720255) -(58728480--58753023) +(58785792--59244543)
-(59252768--59277311) +(59310080--59342847) -(59777056--59834367) -(60301344--60358655)
-(60825632--60882943) -(61349920--61407231) -(61874208--61931519) -(62398496--62455807)
-(62922784--62980095) -(63447072--63504383) -(63971360--64028671) -(64495648--64552959)
-(65019936--65077247) -(65544224--65601535) -(66068512--66125823) -(66592800--66650111)
-(67117088--67174399) -(67641376--67698687) -(68165664--68222975) -(68689952--68747263)
-(69214240--69271551) -(69738528--69795839) -(70262816--70320127) -(70787104--70844415)
-(71311392--71368703) -(71630848--71663615) -(71664642--71664643) -(71664672--71696383)
-(71835680--71892991) -(72359968--72417279) -(72884256--72941567) -(73408544--73465855)
-(73932832--73990143) -(74457120--74514431) -(74981408--75038719) -(75505696--75563007)
-(76029984--76087295) -(76554272--76611583) -(77078560--77135871) -(77602848--77660159)
-(78127136--78184447) -(78577664--78642687)
Fix? no
Free blocks count wrong for group #1602 (0, counted=32768).
Fix? no
Free blocks count wrong for group #1603 (0, counted=32768).
Fix? no
Free blocks count wrong for group #2397 (0, counted=32768).
Fix? no
Free blocks count wrong (77359914, counted=22810628).
Fix? no
Free inodes count wrong (19660789, counted=19660788).
Fix? no

```
/dev/vzpool/ct101-part1: ********** WARNING: Filesystem still has errors **********
/dev/vzpool/ct101-part1: 11/19660800 files (0.0% non-contiguous), 1282774/78642688 blocks
```

После этого увеличиваем пул ZFS еще на два зеркальных диска:

```
zpool add vzpool mirror /dev/sdb  /dev/sda5
```

После этого место в пуле появится:

```
[root@zfs ~]# zpool list
NAME    SIZE  ALLOC   FREE    CAP  DEDUP  HEALTH  ALTROOT
vzpool  1,92T   219G  1,71T    11%  1.00x  ONLINE  -
[root@zfs ~]# zfs list
NAME         USED  AVAIL  REFER  MOUNTPOINT
vzpool       219G  1,68T   272K  /vzpool
vzpool/ct101   219G  1,68T   219G  -
```

После этого файловая система смонтируется:

```
mount /dev/vzpool/ct101-part1 /mnt/
```

С минимумом ошибок в dmesg:

```
[37464.418776] EXT4-fs warning (device zd0p1): ext4_clear_journal_err: Filesystem error recorded from previous mount: IO failure
[37464.418859] EXT4-fs warning (device zd0p1): ext4_clear_journal_err: Marking fs in need of filesystem check.
[37464.419037] EXT4-fs (zd0p1): warning: mounting fs with errors, running e2fsck is recommended
[37464.419377] EXT4-fs (zd0p1): recovery complete
[37464.419488] EXT4-fs (zd0p1): mounted filesystem with ordered data mode. Opts:
```

И файлы уже доступны:

```
ls -al /mnt
итого 296G
drwxr-xr-x  3 root root 4,0K Июл  2 17:23 .
drwxr-xr-x 26 root root 4,0K Июл  2 12:01 ..
drwx------  2 root root  16K Июл  2 17:19 lost+found
-rw-r--r--  1 root root 296G Июл  2 17:50 temp_file_unlimited_Gb
```

Но мы все равно должны запустить fsck, ну что же, запустим!

```
fsck.ext4 /dev/vzpool/ct101-part1
e2fsck 1.41.12 (17-May-2010)
/dev/vzpool/ct101-part1 contains a file system with errors, check forced.
Pass 1: Checking inodes, blocks, and sizes
Pass 2: Checking directory structure
Pass 3: Checking directory connectivity
Pass 4: Checking reference counts
Pass 5: Checking group summary information
Block bitmap differences:  +(52494336--52953087) +(53018624--53477375)
+(53542912--54001663) +(54067200--54525951) +(54591488--55050239)
+(55115776--55574527) +(55640064--56098815) +(56164352--56623103)
+(56688640--57147391) +(57212928--57671679) +(57737216--58195967)
+(58261504--58720255) +(58785792--58884095) +(59342848--59344896)
+(59346944--59375615) +(71172096--71174144) +(71176192--71204863)
Fix<y>? yes

/dev/vzpool/ct101-part1: ***** FILE SYSTEM WAS MODIFIED *****
/dev/vzpool/ct101-part1: 12/19660800 files (0.0% non-contiguous), 78638588/78642688 blocks
```

После этого файловая система пришла в нормальное состояние и продолжила работать.