

Ошибка запуска сервиса `imgbase-setup.service` "Failed to start Configure Image Layers"

1. Вопрос

После установки **zVirt Node** ошибка запуска сервиса **`imgbase-setup.service`**.

В лог-файле `/var/log/messages` :

```
Failed to start Configure Image Layers.
imgbase-setup.service: Failed with result 'exit-code'.
imgbase-setup.service: Main process exited, code=exited, status=1/FAILURE
RuntimeError: Failed to parse NVR: root
raise RuntimeError("Failed to parse NVR: %s" % nvr)
File "/usr/lib/python3.6/site-packages/imgbased/naming.py", line 280, in parse
self.nvr = NVR.parse(nvr) # For convenience: Parse if necessary
File "/usr/lib/python3.6/site-packages/imgbased/naming.py", line 402, in
__init__
File "/usr/lib/python3.6/site-packages/imgbased/naming.py", line 335, in
from_nvr
```

Если продолжить установку в режиме **Hosted Engine**, то появляются ошибки:

```
[ INFO ] TASK [ovirt.hosted_engine_setup : Wait for the host to be up]
Failed to start Open vSwitch Forwarding Unit.
ovs-vswitchd.service: Failed with result 'exit-code'.
ovs-vswitchd.service: Start request repeated too quickly.
Stopped Open vSwitch Forwarding Unit.
ovs-vswitchd.service: Scheduled restart job, restart counter is at 5.
ovs-vswitchd.service: Service RestartSec=100ms expired, scheduling restart.
Failed to start Open vSwitch Forwarding Unit.
ovs-vswitchd.service: Failed with result 'exit-code'.
ovs-vswitchd.service: Control process exited, code=exited status=1
/usr/share/openvswitch/scripts/ovs-lib: line 515:
/usr/share/openvswitch/scripts/ovs-kmod-ctl: Permission denied**
```

2. Решение

- Выполнить проверку сбойных сервисов командой:

```
systemctl list-units --failed --all
```

- На Node запустить проверку **imgbase check**. Корректный результат:

```
imgbase check
Status: OK
Bootloader ... OK
  Layer boot entries ... OK
  Valid boot entries ... OK
Mount points ... OK
  Separate /var ... OK
  Discard is used ... OK
Basic storage ... OK
  Initialized VG ... OK
  Initialized Thin Pool ... OK
  Initialized LVs ... OK
Thin storage ... OK
  Checking available space in thinpool ... OK
  Checking thinpool auto-extend ... OK
```

- На хосте запустить проверку правильность разметки дисков. Корректный результат:

```
df -hT
```

Filesystem	Type	Size
Used Avail Use% Mounted on		
devtmpfs	devtmpfs	32G
0 32G 0% /dev		
tmpfs	tmpfs	32G
4.0K 32G 1% /dev/shm		
tmpfs	tmpfs	32G
610M 31G 2% /run		
tmpfs	tmpfs	32G
0 32G 0% /sys/fs/cgroup		
/dev/mapper/znn-zvirt--node--ng--3.0--0.20220410.0+1	xfs	1.6T
81G 1.5T 5% /		
/dev/sda1	xfs	1014M
350M 665M 35% /boot		
/dev/mapper/znn-tmp	xfs	1014M
40M 975M 4% /tmp		
/dev/mapper/znn-home	xfs	1014M
40M 975M 4% /home		
/dev/mapper/znn-var	xfs	15G
1.5G 14G 10% /var		
/dev/mapper/znn-var_log	xfs	8.0G
396M 7.7G 5% /var/log		
/dev/mapper/znn-var_crash	xfs	10G
105M 9.9G 2% /var/crash		
/dev/mapper/znn-var_log_audit	xfs	2.0G

84M	2.0G	5%	/var/log/audit		
/dev/sdb2				xf	1.0T
393G	632G	39%	/nfs/sdb2		
/dev/sdb1				xf	1.0T
459G	565G	45%	/nfs/sdb1		
tmpfs				tmpfs	6.3G
0	6.3G	0%	/run/user/0		

- Если для создания загрузочных USB-накопителей использовался **Ventoy**, его необходимо заменить на другое ПО, например **Rufus**.

Решение проблем zVirt

▶ Аутентификация и авторизация

▶ Виртуальные машины

▶ Гипервизоры

▶ GlusterFS

▶ Менеджер управления

▶ Миграция

▶ Обновление

▶ Работа с дисками

▶ Резервное копирование

▶ Сертификаты

▶ Сеть

▶ SDN

▶ Службы

► Установка

► Хранилище

► Keycloak

► Прочее

Создание локального хранилища

1. Подготовка каталога

- На хосте создайте каталог, который будет использоваться для локального хранилища.

```
mkdir -p /data/images
```

- Убедитесь, что каталог имеет разрешения, разрешающие доступ на чтение/запись для пользователя **vdsbm** (UID 36) и группы **kvm** (GID 36).

```
chown 36:36 /data /data/images  
chmod 0755 /data /data/images
```

2. Добавление хранилища

- Создайте новый центр данных с локальным хранилищем:

- Перейдите в **Ресурсы > Центры данных** и нажмите [**Новый**].

- В окне создания Центра данных:

- Введите желаемое имя
- В выпадающем списке **Тип хранилища** выберите **Локальный**.
- Убедитесь, что выбрана последняя доступная версия совместимости.
- Выберите желаемый уровень квотирования.
- Нажмите [**ОК**].

- В открывшемся помощнике нажмите [**настроить кластер**].

- В открывшемся окне создания кластера укажите необходимые настройки и нажмите [**ОК**].

- Добавьте в созданный кластер новый хост.

- При необходимости включите хост:

- Переходим **Ресурсы > Хосты**.
- Выбираем хост в списке **управление > включить**.

- Перейдите в **Хранилище > Домены** и нажмите [**Новый домен**].

- В окне добавления домена:

- В качестве центра данных укажите **центр данных**, созданный шагами ранее.
- В поле тип хранилища укажите **Локальный на хосте**.
- В поле **Путь** укажите путь к созданному каталогу
- В поле **Имя** укажите имя домена
- При необходимости измените дополнительные параметры.
- Нажмите [**ОК**].

Рекомендации по организации хранения ISO-образов

1. Окружение

zVirt 4.2

2. Пояснение

На данный момент в zVirt есть ограничения при работе с ISO-образами:

- После перемещения ISO-образов из одного домена хранения в другой теряется возможность загрузки с них виртуальных машин.
- После перемещения ISO-образа в домене хранения у пользователя нет возможности произвести смену его названия.
- Не рекомендуется выполнять параллельную загрузку двух и более ISO-образов.

3. Рекомендации

В связи с указанными ограничениями рекомендуется поступать следующим образом:

- Переименовывать файл ISO до загрузки в домен хранения таким образом, чтобы по названию можно было понять его назначение.
- Использовать отдельный домен хранения для размещения ISO-образов, например ISO, с типом **Данные**. Для этого:
 1. Авторизуйтесь в портале администратора,
 2. Перейдите в раздел **Хранилище > Диски**,
 3. Выберите кнопку меню [**Загрузить**] и нажмите [**Начать**],
 4. Нажмите [**Тест соединения**], чтобы провести тестирование корректности настройки системы до начала загрузки ISO-образа,
 5. Нажмите [**Выберите файл**],
 6. В диалоговом окне выберите ISO-образ для загрузки,
 7. В полях **Центр данных**, **Домен хранения**, **Профиль диска** и **Хост** выберите необходимый домен хранения, а также хост, который будет использоваться в

качестве прокси-сервера для загрузки ISO-образа.



При выборе хоста рекомендуется выбирать наименее загруженный хост.

- Для размещения ISO-образов также можно использовать API-функционал. Подробнее — [Загрузка образов дисков в хранилище с помощью утилиты "upload disk"](#).

4. Имеющиеся проблемы и способы их решения

Если после перемещения ISO-образ перестает быть загрузочным, то необходимо выполнить запрос к API с указанием ID диска:

```
USER=admin@internal ①
PASSWORD=1
FQDN="engine.ru"
PEM="engine.cer"
DISK_ID="b8a70a22-53b1-43af-ab04-0abbe53f359a"
curl -s \
    --cacert ${PEM} \
    --request POST \
    --header 'Version: 4' \
    --header 'Accept: application/xml' \
    --header "Content-Type: application/xml" \
    --user ${USER}:'${PASSWORD}' \
    --data "<action><disk><sparse>false</sparse><format>raw</format></disk>
</action>" \
    https://${FQDN}/ovirt-engine/api/disks/${DISK_ID}/convert
```

① В случае интеграции с Keycloak используйте учетную запись **admin@zvirt@internalsso**

Восстановление метаданных домена



Данная инструкция выполняется в том случае, когда бэкап был предварительно сделан, или возможности его сделать нет

1. Диагностика

Выполняется на SPM-хосте, к которому прикреплен проблемный домен:



В случае наличия файла `/etc/lvm/devices/system.devices`, его необходимо временно переименовать.

1. В лог-файле `/var/log/vdsm/vdsm.log` наблюдаем следующее:

```
Cannot process volume group 939d6cac-cc99-4216-b1dd-d34ec2fc5052\'\' '
(lvm:615)
2024-05-19 13:03:58,868+0500 INFO (jsonrpc/3) [storage.StorageDomainCache]
Looking up domain 939d6cac-cc99-4216-b1dd-d34ec2fc5052: 0.11 seconds
(utils:390)
2024-05-19 13:03:58,868+0500 INFO (jsonrpc/3) [vdsm.api] FINISH
teardownImage error=Storage domain does not exist: ('939d6cac-cc99-4216-
b1dd-d34ec2fc5052',)
```

2. В выводе команды `vgscan --config 'devices{filter=["a|.|"]}'` указанной ***Volume group** также нет
3. В выводе команды `pvscan --config 'devices{filter=["a|.|"]}'` нет ***PV**, который принадлежал бы проблемной **VG**
4. В выводе команды `lsblk` находим **mpath** устройство, которое связано с нашей **Volume Group**, и выполняем по нему вывод `parted`:

```
sdm                                     8:192 0 9.8T 0 disk
└─36d039ea0002b2950000011ae65580912 253:17 0 9.8T 0 mpath
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-ids 253:19 0 128M 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-leases 253:20 0 2G 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-metadata 253:21 0 128M 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-inbox 253:22 0 128M 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-outbox 253:23 0 128M 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-xleases 253:24 0 16 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-master 253:25 0 16 0 lvm /rhev/
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-a2912688--f646--4fc1--a7c6--e9f17a90b689 253:33 0 5G 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-d8e3cde4--eb15--4ed8--afc3--936a1dd8f8e2 253:38 0 2.6G 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-ea008834--c3e3--4df6--8a8f--f1b2e595de1f 253:40 0 4.5G 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-13cc86bb--172f--4671--b111--4fc81d9b6662 253:45 0 2G 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-87216cc0--022e--44b7--b8b9--b7c8b721f234 253:48 0 8.3G 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-b132f96e--0c70--4156--97ba--b7dedcc9971a 253:54 0 62.3G 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-33bc2cce--3372--42e2--bbfa--2709e1eas0b9 253:67 0 128M 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-05a2e833--e147--44bf--abc4--ac52bb3a04b8 253:76 0 2G 0 lvm
├─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-954270c2--9aa6--4547--947f--21be378564ab 253:97 0 10G 0 lvm
└─f596e06c--2eb2--4d3f--b5a4--248f9473c7ca-c7hccfa2--c498--4090--b811--b5aab83055df 253:100 0 11.5G 0 lvm
```

Рисунок 1. Пример:

1. Является необходимым **multipath** устройством,
`/dev/mapper/36d039ea0002b2950000011ae65580912`

2. **Volume Group**, которая создана на базе этого устройства, **f596e06c—2eb2—4d3f—b5a4—248f9473c7ca**

```
parted /dev/mapper/36d039ea0002b2950000011ae65580912 print
```

Вывод должен быть следующим:

```
Error: /dev/mapper/36d039ea0002b2950000011ae65580912: unrecognised disk
label
Model: Linux device-mapper (multipath) (dm)
Disk /dev/mapper/36d039ea0002b2950000011ae65580912: 10.7TB
Sector size (logical/physical): 512B/4096B
Partition Table: unknown
Disk Flags:
```

В случае, если вывод отличается:

- **присутствует метка, таблица разделов, разделы** - это и является причиной возникновения проблемы.

Ее возникновением могло послужить **прикрепление блочного ус-ва к другому узлу**, который выполнил разметку на нем.

2. Решение



Действия необходимо выполнять для каждого хоста, на котором наблюдаются проблемы с VG.



Во время выполнения процедуры необходимо организовать простой для VM на данном домене хранения.



Далее в примере будет все также использоваться блочное ус-во **/dev/mapper/36d039ea0002b2950000011ae65580912** - вам необходимо будет заменять на **свое**

1. Удостоверяемся, что в бэкапе **lvm** присутствуют все необходимые слои(**logical volumes**), которые используются в домене, т.к иначе метаданные по ним восстановлены не будут:
 - Файл бэкапа - **/etc/lvm/backup/STORAGE_ID**
 - Все слои, находящиеся в домене хранения, - **/rhev/data-center/HE_MNT_ДИРЕКТОРИЯ/STORAGE_ID(VOLUME GROUP)/DISK_ID**(Диски из вкладки **Хранилище > Диски**)/**СЛОИ_ДИСКА**(Именно они и содержат данные)
2. Выполняем удаление раздела, если он есть:

```
parted /dev/mapper/36d039ea0002b2950000011ae65580912 rm PARTITION_ID ①
```

① **PARTITION_ID** номер раздела.

3. Перезагружаем хост (**В случае, если на нем есть ВМ из других доменов хранения - предварительно мигрируем их**)

4. Очищаем метаданные блочного устройства:

```
wipefs --all --backup /dev/mapper/36d039ea0002b2950000011ae65580912
```

5. Выполняем восстановление метаданных **physical volume**, предварительно необходимо получить его ID из файла бэкапа.

а. Сверяем, что адрес найденного блочного ус-ва соответствует ус-ву, указанному в метаданных

```
939d6cac-cc99-4216-b1dd-34c2fc5052 {
  id = "r10t0M-Rawc-0Edy-nea0-Ybm0-ZxbA-pQrJMj"
  seqno = 21
  format = "lvm2"
  status = ["RESIZABLE", "READ", "WRITE"]
  flags = []
  tags = ["MDT LEASETIMESEC=60", "MDT ALIGNMENT=1048576", "MDT_SDUUID=939d6cac-cc99-4216-b1dd-34c2fc5052",
4b5400a9a9864fd844450d1:IFGAgi-6161-dwG0-cAXa-tiGx-Fi9Q-godqix440p6sstart:064485pecount:3896446mapoffset:0", "MDT_ID
ESCRPTION=iscsi", "MDT_BLOCK_SIZE=512", "MDT_TYPE=ISCSI", "MDT_VGUUID=r10t0M-Rawc-0Edy-nea0-Ybm0-ZxbA-pQrJMj", "MDT_
ERSION=1", "MDT_POOL_DESCRIPTION=Default", "MDT_POOL_DOMAINS=939d6cac-cc99-4216-b1dd-34c2fc5052:Active", "MDT_POOL
_A_CKSUM=2b1744d1c67a31b91767ef98b3c92542029", "MDT_POOL_UUID=27718008-11d4-11ef-beff-566f02a1039c"]
  extent_size = 262144 # 128 Megabytes
  max_lv = 0
  max_pv = 0
  metadata_copies = 0

  physical_volumes {
    pv0 {
      id = "iFGAgi-6161-dwG0-cAXa-tiGx-Fi9Q-godqix"
      device = "/dev/mapper/36001445eaa9f8c60400a69864fd" # Hint only

      status = ["ALLOCATABLE"]
      flags = []
    }
  }
}
```

b. Предварительно запускаем в тестовом режиме (Метаданные не будут изменены).
Удостоверяемся, что операция прошла успешно

```
pvccreate --restore-file /etc/lvm/backup/939d6cac-cc99-4216-b1dd-d34ec2fc5052 --uuid PVID /dev/mapper/36001405eaafdfb604b5400a9a69864fd --config 'devices{filter=["a|.*/"]} -ff --pvmetadatatcopies 2 --test
```

с. Восстанавливаем метаданные PV

```
pvccreate --restore-file /etc/lvm/backup/939d6cac-cc99-4216-b1dd-  
d34ec2fc5052 --uuid PVID /dev/mapper/36001405eaa9dfb604b5400a9a69864fd -  
-config 'devices{filter=["a|.*|"]}' -ff --pvmetadatacopies 2
```

6. Восстанавливаем метаданные **Volume Group**



Возможно, чтобы восстановленный PV был найден, понадобится выполнить следующую команду:

```
pvscan --config 'devices{filter=["a|.*)"']}
```

а. Запускаем в тестовом режиме.

```
vgcfgrestore --file /etc/lvm/backup/939d6cac-cc99-4216-b1dd-d34ec2fc5052
939d6cac-cc99-4216-b1dd-d34ec2fc5052 --config 'devices{filter=
["a|.*)" --test
```

b. В случае успеха - запускаем восстановление:

```
vgcfgrestore --file /etc/lvm/backup/939d6cac-cc99-4216-b1dd-d34ec2fc5052
939d6cac-cc99-4216-b1dd-d34ec2fc5052 --config 'devices{filter=
["a|.*)"
```

c. Пересканируем **volume groups**:

```
vgscan --config 'devices{filter=["a|.*)"
```

7. Если все операции были выполнены успешно, проблемный домен хранения должен перейти в статус **UP** и все операции с ним должны проходить корректно.



Если какой-либо из слоев будет не найден, во вкладке События могут быть сообщения следующего характера

```
VDSM command SetVolumeDescriptionVDS failed: Volume does not exist:
('326303b8-9005-4632-8ab9-61a738b239ec',)
```

Отключение multipath для работы дополнительных дисков в локальной системе

Данная инструкция предназначена для отключения **multipath** при работе с локальными дисками, подключенными к хосту. При работе с iSCSI добавлять в black list устройства, обнаруженные iSCSI не нужно.

При подключении второго и последующих дисков система может применять к ним **multipath**, что не позволит монтировать их.

Для определения, применяется ли **multipath** к вашему устройству, выполните команду:

```
lsblk
```

BASH | 

Пример вывода команды:

```
NAME                                MAJ:MIN RM  SIZE RO TYPE  MOUNTPOINT
sda                                8:0    0 447,1G  0 disk  /storage
sdb                                8:16    0 111,8G  0 disk
├sdb1                             8:17    0  200M  0 part  /boot/efi
├sdb2                             8:18    0    1G  0 part  /boot
└sdb3                             8:19    0 110,6G  0 part
  ├zn-root                         253:0    0  75,7G  0 lvm    /
  ├zn-swap                         253:1    0   7,9G  0 lvm    [SWAP]
  ├zn-home                         253:7    0    1G  0 lvm    /home
  ├zn-tmp                         253:8    0    1G  0 lvm    /tmp
  ├zn-var                         253:9    0   15G  0 lvm    /var
  ├zn-var_log                     253:10   0    8G  0 lvm    /var/log
  └zn-var_log_audit               253:11   0    2G  0 lvm    /var/log/audit
nvme0n1                           259:0    0 465,8G  0 disk
└eui.0025385c01906511            253:2    0 465,8G  0 mpath
  ├eui.0025385c01906511p1        253:3    0  300M  0 part
  ├eui.0025385c01906511p2        253:4    0  100M  0 part
  ├eui.0025385c01906511p3        253:5    0  128M  0 part
  └eui.0025385c01906511p4        253:6    0 465,3G  0 part
```

BASH | 

Как видно из примера, **multipath** применяется к устройству **nvme0n1** (NVMeSSD).

Или выполните команду:

```
multipath -l
```

BASH | 

Пример вывода команды:

```
TS120GSSD220S_020231EFE17005140597 dm-3 ATA      ,TS120GSSD220S
size=112G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='service-time 0' prio=0 status=active
  `-- 1:0:0:0 sdb8:16  activeundef running
eui.0025385c01906511 dm-2 NVME,Samsung SSD 970 EVO Plus 500GB
size=466G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='service-time 0' prio=0 status=active
  `-- 0:0:1:0 nvme0n1 259:0 active undef running
```

BASH |

Как видно из примера, **multipath** применяется к устройствам **TS120GSSD220S** и **Samsung SSD 970 EVOPlus 500GB**.

Для отключения **multipath**:

1. Создайте новый конфигурационный файл в каталоге **/etc/multipath/conf.d**.



Для того, чтобы контролировать порядок выполнения, создавайте файлы с именем, которые начинаются с цифр, что позволит выполнить его последним по порядку. Например, **/etc/multipath/conf.d/90-fcdevice.conf**.

2. Добавьте секцию **blacklist** в созданный файл конфигурации со строкой вида:

```
devnode "^sd[a-z]"
```

BASH |

Например:

```
blacklist {
    devnode "^sd[a-z]"
}
```

BASH |

Данный параметр позволит не применять **multipath** для устройств с именами, заканчивающимися на **sda**, **sdb** и до **sdz**.

3. Выполните перезагрузку службы **multipath**:

```
systemctl reload multipathd.service
```

BASH |

4. Используя команду **lsblk** или **multipath -l** убедитесь, что **multipath** не применяется к вашему устройству.

Рекомендации по настройке multipath для систем хранения данных TATLIN.FLEX

1. Пояснение

При работе с системой хранения данных TATLIN.FLEX рекомендуется использовать настройки **multipath** (многопутевого ввода-вывода) от производителя.



Если система кластерная, то настройка выполняется на каждом узле.

2. Настройка multipath



Не вносите изменения в файл **/etc/multipath.conf** — это может привести к потере настроек при обновлении или перезапуске службы **VDSM**.

Чтобы настроить **multipath**:

1. Создайте новый конфигурационный файл **/etc/multipath/conf.d/flex-multipath.conf** со следующим содержанием:

```
defaults {
    fast_io_fail_tmo          5
    features                  "0"
    no_path_retry             10
    path_checker              tur
    polling_interval          5
    prio                      alua
}

devices {
    device {
        #For initiators with scsi_dh_alua
        #hardware_handler      "1 alua"
        detect_checker         no
        detect_prio            no
        failback                immediate
        no_path_retry          12
        path_grouping_policy    "group_by_prio"
        path_selector           "round-robin 0"
        path_checker            "tur"
        prio                    "alua"
    }
}
```



```
        product      ".*"
        rr_min_io     100
        rr_weight     "uniform"
        vendor        "YADRO"
    }
}
```



Не перезапускайте (`restart`) службу **multipathd** — это приводит к появлению ошибок в журналах службы **VDSM**.

2. Для применения настроек выполните перезагрузку службы **multipathd**:

```
systemctl reload multipathd
```

