



Российская сервисная платформа виртуализации РУСТЭК

Работа с файловой системой OCFS2

Релиз 2.6

Оглавление

1	Увеличение размера	3
1.1	iSCSI	3
1.2	Fibre Channel	3
1.3	Расширение ФС	4
2	Подключение дополнительного блочного устройства	6
2.1	Использование нового бэкенда и существующего типа диска	6
2.2	Использование нового бэкенда с новым типом диска	7
2.2.1	Создание дополнительного типа диска	8
3	Настройка балансировки нагрузки между бэкендами	9
3.1	Распределение дисков	9
3.2	Стэкирование дисков	10
4	Настройка кэширования образов	11

1 Увеличение размера

Увеличение размера кластерной файловой системы (ФС) OCFS2 возможно на лету — без остановки кластера. Перед увеличением размера ФС нужно обновить размер блочного устройства в операционной системе.

В качестве примера будем увеличивать размер блочного устройства с WWID 36001405e18f5fe49009433495ea0e4e7, представленного дисками `sdd` и `sde`:

```
# lsblk
NAME                                MAJ:MIN RM  SIZE RO TYPE  MOUNTPOINT
...
sdb                                  8:16   0  100G  0 disk
└─36001405b6fdeda6d5dd4070a991c297e 253:0   0  100G  0 mpath /mnt/heartbeat
sdc                                  8:32   0  100G  0 disk
└─36001405b6fdeda6d5dd4070a991c297e 253:0   0  100G  0 mpath /mnt/heartbeat
sdd                                  8:48   0  300G  0 disk
└─36001405e18f5fe49009433495ea0e4e7 253:1   0  300G  0 mpath /mnt/ocfs2-36001405e18f5fe49009433495ea0e4e7
sde                                  8:64   0  300G  0 disk
└─36001405e18f5fe49009433495ea0e4e7 253:1   0  300G  0 mpath /mnt/ocfs2-36001405e18f5fe49009433495ea0e4e7
```

1.1 iSCSI

В случае подключения блочного устройства по протоколу iSCSI нужно сделать повторное сканирование iSCSI-таргетов.

В разделе **Дисковая подсистема** РУСТЭК.Конфигуратора переходим в следующие подразделы: **Настройка блочного хранилища** → **Настройки iSCSI-хранилища** и выбираем пункт **Подключить iSCSI-хранилище**.

После успешного завершения посмотрим вывод информации о блочных устройствах:

```
# lsblk
NAME                                MAJ:MIN RM  SIZE RO TYPE  MOUNTPOINT
...
sdb                                  8:16   0  100G  0 disk
└─36001405b6fdeda6d5dd4070a991c297e 253:0   0  100G  0 mpath /mnt/heartbeat
sdc                                  8:32   0  100G  0 disk
└─36001405b6fdeda6d5dd4070a991c297e 253:0   0  100G  0 mpath /mnt/heartbeat
sdd                                  8:48   0  400G  0 disk
└─36001405e18f5fe49009433495ea0e4e7 253:1   0  300G  0 mpath /mnt/ocfs2-36001405e18f5fe49009433495ea0e4e7
sde                                  8:64   0  400G  0 disk
└─36001405e18f5fe49009433495ea0e4e7 253:1   0  300G  0 mpath /mnt/ocfs2-36001405e18f5fe49009433495ea0e4e7
```

Размер блочного устройства с WWID 36001405e18f5fe49009433495ea0e4e7, представленного дисками `sdd` и `sde`, увеличился с 300Gb до 400Gb.

После этого необходимо увеличить размер ФС для отображения нового пространства, как написано в разделе [Расширение ФС](#).

1.2 Fibre Channel

В случае подключения блочного устройства по протоколу Fibre Channel нужно сделать повторное сканирование SCSI-шины.

В разделе **Дисковая подсистема** РУСТЭК.Конфигуратора переходим в подраздел **Настройка блочного хранилища** и выбираем пункт "Обновить информацию о блочных устройствах".

После успешного завершения посмотрим вывод информации о блочных устройствах:

```
# lsblk
NAME                                MAJ:MIN RM  SIZE RO TYPE  MOUNTPOINT
...
sdb                                  8:16   0  100G  0 disk
└─36001405b6fdeda6d5dd4070a991c297e 253:0   0  100G  0 mpath /mnt/heartbeat
sdc                                  8:32   0  100G  0 disk
└─36001405b6fdeda6d5dd4070a991c297e 253:0   0  100G  0 mpath /mnt/heartbeat
sdd                                  8:48   0  400G  0 disk
└─36001405e18f5fe49009433495ea0e4e7 253:1   0  300G  0 mpath /mnt/ocfs2-36001405e18f5fe49009433495ea0e4e7
sde                                  8:64   0  400G  0 disk
└─36001405e18f5fe49009433495ea0e4e7 253:1   0  300G  0 mpath /mnt/ocfs2-36001405e18f5fe49009433495ea0e4e7
```

Размер блочного устройства с WWID 36001405e18f5fe49009433495ea0e4e7, представленного дисками sdd и sde, увеличился с 300Gb до 400Gb.

После этого необходимо увеличить размер ФС для отображения нового пространства, как написано в разделе [Расширение ФС](#).

1.3 Расширение ФС

Команды из этого раздела нужно выполнять на любом доступном узле с ролью "Управление дисками" или "Вычислительный узел".

Проверим размер ФС до расширения, указав в качестве имени ФС путь к блочному устройству с WWID 36001405e18f5fe49009433495ea0e4e7:

```
# df -h /dev/mapper/36001405e18f5fe49009433495ea0e4e7
Filesystem                                Size  Used Avail Use% Mounted on
/dev/mapper/36001405e18f5fe49009433495ea0e4e7 300G  2.4G  298G   1% /mnt/ocfs2-36001405e18f5fe49009433495ea0e4e7
```

Изменим размер multipath-устройства на всех доступных узлах с установленными ролями "Управление дисками" или "Вычислительный узел", запустив следующую команду на любом из узлов платформы, подставив в команду **multipathd resize map** нужный WWID:

```
# ansible -i /var/lib/rustack-ansible/inventory.yml '!unreachable,cinder,compute' -m shell -a 'multipathd resize map 36001405e18f5fe49009433495ea0e4e7'
```

Вывод информации о блочных устройствах после увеличения размера multipath-устройства:

```
# lsblk
...
sdb                8:16  0 100G  0 disk
└─36001405b6fdeda6d5dd4070a991c297e 253:0  0 100G  0 mpath /mnt/heartbeat
sdc                8:32  0 100G  0 disk
└─36001405b6fdeda6d5dd4070a991c297e 253:0  0 100G  0 mpath /mnt/heartbeat
sdd                8:48  0 400G  0 disk
└─36001405e18f5fe49009433495ea0e4e7 253:1  0 400G  0 mpath /mnt/ocfs2-36001405e18f5fe49009433495ea0e4e7
sde                8:64  0 400G  0 disk
└─36001405e18f5fe49009433495ea0e4e7 253:1  0 400G  0 mpath /mnt/ocfs2-36001405e18f5fe49009433495ea0e4e7
```

Размер multipath-устройства с WWID 36001405e18f5fe49009433495ea0e4e7 увеличился с 300Gb до 400Gb.

Расширим ФС, выполнив команду с указанием пути к блочному устройству:

```
# tuneefs.ocfs2 -vS /dev/mapper/36001405e18f5fe49009433495ea0e4e7
```

Проверим размер ФС после расширения, указав в качестве имени ФС путь к блочному устройству с WWID 36001405e18f5fe49009433495ea0e4e7:

```
# df -h /dev/mapper/36001405e18f5fe49009433495ea0e4e7
Filesystem                                Size  Used Avail Use% Mounted on
/dev/mapper/36001405e18f5fe49009433495ea0e4e7 400G  2.4G  398G   1% /mnt/ocfs2-36001405e18f5fe49009433495ea0e4e7
```

Размер ФС блочного устройства с WWID 36001405e18f5fe49009433495ea0e4e7 увеличился с 300Gb до 400Gb.

2 Подключение дополнительного блочного устройства

Перед настройкой бэкенда должно быть выполнено обнаружение блочного устройства в операционной системе на всех доступных узлах с установленными ролями "Управление дисками" или "Вычислительный узел".

Если блочное устройство презентуется по одному пути, то в операционной системе оно не обнаружится как multipath-устройство:

```
# lsblk
NAME MAJ:MIN RM SIZE RO TYPE MOUNTPOINT
...
sdb      8:16  0 100G  0 disk
```

Для присвоения такому устройству WWID в разделе **Дисковая подсистема** РУСТЭК.Конфигуратора переходим в следующие подразделы: **Настройка блочного хранилища** → **Настройки Multipath** и в пункте "Отображать только multipath-устройства" выбираем **Нет** и нажимаем **Применить**.

После этого нужно выполнить повторное обнаружение блочного устройства:

- Если блочное устройство подключено по протоколу iSCSI, то в разделе **Дисковая подсистема** РУСТЭК.Конфигуратора переходим в следующие подразделы: **Настройка блочного хранилища** → **Настройки iSCSI-хранилища** и выбираем пункт "Подключить iSCSI-хранилище".
- Если блочное устройство подключено по протоколу Fibre Channel, то в разделе **Дисковая подсистема** РУСТЭК.Конфигуратора переходим в подраздел **Настройка блочного хранилища** и выбираем пункт "Обновить информацию о блочных устройствах".

После успешного завершения на всех доступных узлах с установленными ролями "Управление дисками" или "Вычислительный узел" блочное устройство, презентованное по одному пути, будет иметь WWID:

```
# lsblk
NAME                                MAJ:MIN RM  SIZE RO TYPE  MOUNTPOINT
...
sdb                                8:16  0 100G  0 disk
└─36001405b6fdeda6d5dd4070a991c297e 253:0  0 100G  0 mpath
```

2.1 Использование нового бэкенда и существующего типа диска

Для автоматического подключения дополнительного блочного устройства необходимо запустить РУСТЭК.Конфигуратор, перейти в раздел **Дисковая подсистема**, далее в подраздел **Настройки OCFS2** и в поле "Список WWID для OCFS2" выбрать WWID нового блочного устройства, а после этого нажать **Применить**.

После этого выбрать пункт "Применить конфигурацию РУСТЭК" в главном меню РУСТЭК.Конфигуратора.

Во время применения конфигурации для нового бэкенда в директории `/etc/openstack/cinder_backends/` будет создано два новых файла с именем, совпадающим с WWID блочного устройства, и расширениями `conf` и `share`:

```
# ls -l /etc/openstack/cinder_backends/
...
-rw-r----- 1 cinder cinder 458 Jul 10 21:42 3600140532361644ad15467195d3302b9.conf
-rw-r----- 1 cinder cinder 45 Jul 7 20:38 3600140532361644ad15467195d3302b9.share
```

После успешного применения конфигурации при использовании типа диска `ocfs2` при создании дисков, выбор бэкенда будет осуществляться драйвером на основании используемой стратегии (по умолчанию `CapacityWeigher`).

2.2 Использование нового бэкенда с новым типом диска

В качестве примера используем блочное устройство с WWID `3600140532361644ad15467195d3302b9`.

Для того чтобы настроить новый бэкенд на использование с новым типом диска, сначала нужно добавить бэкенд, как написано в разделе [Использование нового бэкенда и существующего типа диска](#).

После успешного добавления нового бэкенда нужно изменить значение параметра `volume_backend_name` в файле:

`/etc/openstack/cinder_backends/3600140532361644ad15467195d3302b9.conf` (где имя файла — это WWID блочного устройства):

```
[3600140532361644ad15467195d3302b9]
backend_host = node01
volume_backend_name = ocfs2
nas_secure_file_permissions = true
nas_secure_file_operations = true
volume_clear = zero
volume_clear_ionice = -c3
max_over_subscription_ratio = 1.0
volume_driver = rustack-volume-drivers.sharedfs.SharedFS
sharedfs_shares_config =
/etc/openstack/cinder_backends/3600140532361644ad15467195d3302b9.share
sharedfs_sparsed_volumes = true
sharedfs_reflink_cmd = cp --reflink
```

По умолчанию параметр `volume_backend_name` имеет значение `ocfs2`. Изменим его на произвольное значение, отличающееся от используемого в других бэкендах. Используемое значение всех бэкендов можно посмотреть командой:

```
# grep -r volume_backend_name /etc/openstack/cinder_backends/
/etc/openstack/cinder_backends/3600140532361644ad15467195d3302b9.conf:volume_backend_name = ocfs2
/etc/openstack/cinder_backends/360014056cfd72f5e176480199ad81276.conf:volume_backend_name = ocfs2
/etc/openstack/cinder_backends/nfs.conf:volume_backend_name = nfs
/etc/openstack/cinder_backends/36001405d8196806433a4f1781e75c189.conf:volume_backend_name = ocfs2
```

Изменим значение нового бэкенда на **new_ocfs2**. Для этого можно использовать один раз команду на любом из доступных узлов платформы:

```
# sed -i 's/volume_backend_name = ocfs2/volume_backend_name = new_ocfs2/g'
/etc/openstack/cinder_backends/3600140532361644ad15467195d3302b9.conf
```

После этого перезапустим сервис **cinder-volume**:

```
# ssh $(sc -p | grep "Сервис cinder-volume" | awk '{ print $4 }') rc-config restart
cinder-volume
Restarting init script
* Stopping cinder-volume ... [ ok ]
* starting cgroups cleanup ... [ ok ]
* Starting cinder-volume ... [ ok ]
```

Или вместо этого действия воспользуемся пунктом "Применить конфигурацию РУСТЭК" в главном меню РУСТЭК.Конфигуратора.

После перезагрузки сервиса или успешного применения конфигурации необходимо связать новый бэкенд **new_ocfs2** с новым типом диска, как написано в разделе Создание дополнительного типа диска.

2.2.1 Создание дополнительного типа диска

Создаём дополнительный тип диска с произвольным именем:

```
# openstack volume type create new_type_ocfs2
```

Далее новый тип диска необходимо связать с новым бэкендом, используя параметр **volume_backend_name** из конфигурационного файла бэкенда:

```
# openstack volume type set new_type_ocfs2 --property volume_backend_name=new_ocfs2
```

Теперь при создании дисков с указанием типа диска **new_type_ocfs2** они будут создаваться на бэкенде **new_ocfs2**.

3 Настройка балансировки нагрузки между бэкендами

Для балансировки нагрузки между бэкендами используются различные стратегии и параметры драйвера. По умолчанию используется стратегия **CapacityWeigher**, которая взвешивает бэкенды по их виртуальной или фактической свободной ёмкости. В случае тонких дисков бэкенды взвешиваются по их виртуальной свободной ёмкости, рассчитанной как общая ёмкость, умноженная на коэффициент оверкоммита за вычетом выделенной ёмкости. В случае толстых дисков бэкенды взвешиваются по их фактической свободной ёмкости с учетом зарезервированного пространства.

В качестве примера настройки используем стратегию **VolumeNumberWeigher**. У неё есть два режима: распределение и стэкирование. В первом режиме диски распределяются по бэкендам равномерно, а во втором режиме наоборот — диски создаются сначала на одном бэкенде, а после его заполнения на втором и так далее.

3.1 Распределение дисков

Сначала переопределим стратегию балансировки в конфигурационном файле любого из бэкендов в директории `/etc/openstack/cinder_backends/`, добавив секцию **[DEFAULT]**:

Стратегия будет применена для всех бэкендов.

```
[DEFAULT]
scheduler_default_weighers = VolumeNumberWeigher
```

После этого перезапустим сервис **cinder-volume**:

```
# ssh $(sc -p | grep "Сервис cinder-volume" | awk '{ print $4 }') rc-config restart
cinder-volume
Restarting init script
* Stopping cinder-volume ... [ ok ]
* starting cgroups cleanup ... [ ok ]
* Starting cinder-volume ... [ ok ]
```

А сервис **cinder-scheduler** перезапустим на всех узлах платформы с ролью "Управление дисками", запустив следующую команду на любом из узлов платформы:

```
# ansible -i /var/lib/rustack-ansible/inventory.yml '!unreachable,cinder' -m shell -a
'rc-config restart cinder-scheduler'
```

Или вместо этих действий воспользуемся пунктом "Применить конфигурацию РУСТЭК" в главном меню РУСТЭК.Конфигуратора.

После перезагрузки сервисов или успешного применения конфигурации диски по бэкендам будут распределяться равномерно.

3.2 Стэкирование дисков

Сначала переопределим стратегию балансировки в конфигурационном файле любого из бэкендов в директории `/etc/openstack/cinder_backends/`, добавив секцию **[DEFAULT]**:

Стратегия будет применена для всех бэкендов.

```
[DEFAULT]
scheduler_default_weighers = VolumeNumberWeigher
volume_number_multiplier = 1
```

После этого перезапустим сервис **cinder-volume**:

```
# ssh $(sc -p | grep "Сервис cinder-volume" | awk '{ print $4 }') rc-config restart
cinder-volume
Restarting init script
* Stopping cinder-volume ... [ ok ]
* starting cgroups cleanup ... [ ok ]
* Starting cinder-volume ... [ ok ]
```

А сервис **cinder-scheduler** перезапустим на всех узлах платформы с ролью "Управление дисками", запустив следующую команду на любом из узлов платформы:

```
# ansible -i /var/lib/rustack-ansible/inventory.yml '!unreachable,cinder' -m shell -a
'rc-config restart cinder-scheduler'
```

Или вместо этих действий воспользуемся пунктом "Применить конфигурацию РУСТЭК" в главном меню РУСТЭК.Конфигуратора.

После перезагрузки сервисов или успешного применения конфигурации диски сначала будут создаваться на одном бэкенде, а после его заполнения на втором и так далее.

4 Настройка кэширования образов

В конфигурационном файле бэкенда в основной секции (в примере она называется [3600140532361644ad15467195d3302b9]) активируем кэширование, задав параметр `image_volume_cache_enabled` значением `True`. При необходимости указываем дополнительные параметры для настройки кэширования:

```
[3600140532361644ad15467195d3302b9]
...
image_volume_cache_enabled = True
image_volume_cache_max_size_gb = 200
image_volume_cache_max_count = 50
```

После этого перезапустим сервис `cinder-volume`:

```
# ssh $(sc -p | grep "Сервис cinder-volume" | awk '{ print $4 }') rc-config restart
cinder-volume
Restarting init script
* Stopping cinder-volume ... [ ok ]
* starting cgroups cleanup ... [ ok ]
* Starting cinder-volume ... [ ok ]
```

Или вместо этого действия воспользуемся пунктом "Применить конфигурацию РУСТЭК" в главном меню РУСТЭК.Конфигуратора.

После перезагрузки сервиса или успешного применения конфигурации будет включено кэширование образов.