

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ «HELIUS.КУБ»

РУКОВОДСТВО АДМИНИСТРАТОРА
RU.ЕЦРТ.00057-01 36

ЛИСТОВ 87



АННОТАЦИЯ

Настоящее руководство предназначено для администратора программного обеспечения «Helius.КУБ» RU.ЕЦРТ.00057-01 (далее – ПО). В нем содержатся сведения, необходимые для настройки и эксплуатации ПО.

Администрирование ПО производится через веб-интерфейс с удалённого автоматизированного рабочего места, подключённого к ПО по сети, или через командную строку.

Перечень сокращений приведён в конце документа.

В связи с постоянной работой по совершенствованию ПО в настоящее руководство могут быть внесены изменения, связанные с улучшением ПО. Внесённые изменения будут опубликованы в новой редакции руководства и на сайте компании: URL: <https://гелиус.рф>

СОДЕРЖАНИЕ

| | |
|---|----|
| 1. ОБЩИЕ СВЕДЕНИЯ | 5 |
| 1.1. Обозначение и наименование программы | 5 |
| 1.2. Назначение и область применения | 5 |
| 1.3. Функции программы | 5 |
| 1.4. Сервисное обслуживание и техническая поддержка | 8 |
| 2. УСЛОВИЯ ВЫПОЛНЕНИЯ ПРОГРАММЫ | 9 |
| 2.1. Требования к аппаратному обеспечению..... | 9 |
| 2.2. Требования к программному обеспечению | 9 |
| 2.3. Требования к уровню подготовки администратора | 9 |
| 3. НАЧАЛО ИСПОЛЬЗОВАНИЯ ПРОГРАММЫ..... | 10 |
| 3.1. Вход в веб-интерфейс..... | 10 |
| 3.2. Обзор интерфейса и навигация | 10 |
| 3.3. Настройка представлений. Механизмы поиска и фильтрации | 14 |
| 3.4. Доступные операции на панели инструментов | 15 |
| 3.5. Выход из веб-интерфейса | 15 |
| 4. ПАНЕЛЬ УПРАВЛЕНИЯ..... | 16 |
| 5. НАСТРОЙКИ ПАНЕЛИ..... | 20 |
| 5.1. Пользователи..... | 20 |
| 5.2. Роли..... | 23 |
| 5.3. Конфигурация телеметрии..... | 26 |
| 6. КЛАСТЕР ХРАНЕНИЯ | 30 |
| 6.1. Хосты | 30 |
| 6.2. Физические диски..... | 36 |
| 6.3. Мониторы..... | 36 |
| 6.4. Сервисы | 38 |
| 6.5. OSD | 44 |
| 6.6. Редактор конфигурации..... | 59 |
| 6.7. CRUSH карта..... | 61 |
| 6.8. Модули менеджера..... | 62 |
| 6.9. Пользователи кластера хранения..... | 64 |
| 6.10. Логи..... | 68 |

| | |
|--------------------------------------|----|
| 6.11. Оповещения..... | 69 |
| 7. ПУЛЫ..... | 73 |
| 7.1. Создание пула..... | 75 |
| 7.2. Редактирование пула..... | 84 |
| 7.3. Просмотр информации о пуле..... | 85 |
| 7.4. Удаление пула..... | 86 |
| 7.5. Общая производительность..... | 86 |
| ПЕРЕЧЕНЬ СОКРАЩЕНИЙ | 87 |

1. ОБЩИЕ СВЕДЕНИЯ

1.1. Обозначение и наименование программы

Обозначение программы: RU.ЕЦРТ.00057-01.

Наименование программы: Программное обеспечение «Helius.КУБ».

1.2. Назначение и область применения

ПО предназначено для:

- создания масштабируемой системы хранения данных, которая позволяет легко увеличивать объем хранимой информации и добавлять ресурсы по мере необходимости;
- обеспечения единой точки мониторинга и управления;
- обеспечения отказоустойчивости, гарантируя целостность данных и их доступность даже при сбоях оборудования и/или ПО;
- поддержки универсальности, что позволяет работать с различными форматами и протоколами доступа к данным;
- автоматизации управления данными, что снижает затраты на администрирование и повышает производительность;
- обеспечения безопасности, защищая данные от несанкционированного доступа и потерь;
- визуализации и отслеживания утилизации системных ресурсов;
- своевременного реагирования (с помощью уведомлений) на внештатные ситуации, потенциальные проблемы и неисправности, чтобы минимизировать время простоя.

1.3. Функции программы

ПО обладает следующим функционалом:

- предоставление подробной информации о состоянии каждого хоста, включая статус (онлайн/офлайн), загрузку CPU, использование памяти, сетевую активность и состояние дисков;
- мониторинг изменений в конфигурации кластера в реальном времени;
- механизм уведомлений о критических событиях;
- отображение списка всех пулов, доступных в кластере;
- фильтрация и сортировка отображаемой информации по различным критериям;
- отображение статистики использования ресурсов кластера (диск, память, сеть);
- проверка и отображение учётных записей с подробной информацией;
- идентификация неактивных или устаревших учётных записей;
- проверка и отображение текущего состояния кворума кластера;

- механизм автоматического уведомления при потере кворума;
- отображение структуры дерева OSD кластера;
- возможность проверки состояния каждого OSD;
- предоставление статистики работы каждого OSD (загрузка, ошибки);
- механизм мониторинга производительности OSD;
- отображение статистики по всем PG, включая их состояние и распределение;
- механизм анализа производительности PG;
- предоставление полного списка PG с возможностью фильтрации отображаемой информации по статусу и другим параметрам;
- создание и настройка виртуального IP-адреса для доступа к кластеру;
- создание пула с репликацией данных;
- создание пула с механикой «erasure coding»;
- удаление пула с подтверждением операции;
- изменение имени пула;
- создание CRUSH-правил для оптимизации распределения данных на SSD-накопителях, на HDD-дисках и для гибридных пулов (включающих SSD накопители и HDD-диски);
- создание и настройка файловой системы в кластере;
- создание NFS-сервиса для доступа к файловой системе;
- проверка доступности NFS-сервиса и корректности настроек;
- настройка и запуск NFS-сервера;
- создание пула, обеспечивающего блочный доступ к данным;
- создание объектного шлюза для доступа к объектам;
- проверка доступности объекта через протокол S3;
- горизонтальное масштабирование, позволяющее наращивать вычислительные мощности и объём хранения данных без прерывания обслуживания;
- высокая доступность данных благодаря механизмам репликации и самовосстановления;
- автоматическое восстановление данных после выхода из строя отдельных хостов или компонентов;
- предоставление функционала объектного хранилища, соответствующего стандартам (Amazon S3 или OpenStack Swift) для удобной интеграции с существующими сервисами;
- возможность организации данных в виде файловой системы, которая поддерживает стандартные операции;
- защита данных от несанкционированного доступа с помощью шифрования на уровне диска и/или сетевого трафика;

- предоставление программных интерфейсов (API), позволяющих интегрировать ПО с другими системами и приложениями;
- добавление метаданных к объектам в файловом хранилище для упрощения поиска и категоризации данных;
- создание резервных копий данных и восстановление их в случае необходимости;
- возможность устранять дублирующие блоки данных для более эффективного использования пространства хранения (дедупликация данных);
- возможность поддерживать зеркалирование данных между несколькими кластерами для повышения отказоустойчивости;
- создание снимков данных для быстрого отката к предыдущей версии;
- поддержка комбинаций HDD-дисков и SSD-накопителей для оптимизации производительности и стоимости;
- возможность перемещения данных между уровнями хранения (например, быстрые SSD и медленные HDD) в зависимости от частоты их использования;
- выделение «горячих» (часто используемых) и «холодных» (редко используемых) данных для оптимизации хранения и затрат;
- возможность создания, редактирования и удаления OSD через веб-интерфейс;
- автоматическая балансировка данных между OSD для равномерного распределения нагрузки и использования ресурсов;
- предоставление информации о состоянии каждого OSD, включая загрузку CPU, использование памяти, занятость диска и сетевые показатели;
- поддержка работы с различными типами дисков, включая HDD, SSD;
- добавление новых хостов в существующий кластер без прерывания работы остальных хостов;
- удаление хостов из кластера с минимальным воздействием на остальные хосты и работоспособность кластера;
- автоматическое обнаружение и регистрацию новых хостов при их подключении к сети;
- возможность управления состоянием хостов (включение, выключение, перевод в режим обслуживания и т.д.) через веб-интерфейс;
- множественный доступ с индивидуальными правами для каждого пользователя;
- широкие возможности для настройки прав доступа, включая роли и разрешения для различных категорий пользователей;
- строгие меры аутентификации пользователей;

- генерация и отправка уведомлений администраторам ПО в случае обнаружения проблем или неисправностей;

- учёт всех логов, фиксирование каждой операции и событий, происходящих в кластере, для обеспечения полноты и достоверности данных;

- поддержка технологии SSD- и RAM-кэширование данных;

- возможность подключение и объединение внешних хранилищ данных по протоколам S3, AWS4, NFS, SMB/CIFS, iSCSI, FC.

1.4. Сервисное обслуживание и техническая поддержка

Если в процессе эксплуатации возникнут вопросы, обращайтесь в службу сервиса и технической поддержки предприятия-изготовителя.

Перед обращением в службу технической поддержки подготовьте следующую информацию:

- чёткое описание возникшей проблемы;

- состав используемых аппаратных и программных средств;

- номер используемой версии программного обеспечения «Helius.КУБ».

Техническая поддержка осуществляется по ссылке URL: <https://helius.graviton.ru/>, по электронной почте TechSupport@helius.graviton.ru и по телефону 8 (800) 551-75-57.

2. УСЛОВИЯ ВЫПОЛНЕНИЯ ПРОГРАММЫ

2.1. Требования к аппаратному обеспечению

ПО функционирует на сервере архитектуры x64 под управлением ОС семейства Linux со следующими минимальными техническими характеристиками:

- четырёх-ядерный процессор Intel или AMD с базовой частотой не менее 2,4 ГГц;
- оперативная память не менее 32 ГБ;
- объём свободного дискового пространства не менее 512 ГБ;
- доступ к серверу точного времени (NTP).

Для работы с веб-интерфейсом ПО рабочее место администратора должно быть оборудовано персональным компьютером (тонкий клиент, ноутбук) с сетевой платой со скоростью передачи данных не менее 100 Мбит/с, на котором должен быть установлен веб-браузер, а также должны быть выполнены настройки доступа к сети, позволяющие беспрепятственно выполнять работу в веб-интерфейсе ПО.

2.2. Требования к программному обеспечению

ПО функционирует под управлением ОС «РЕД ОС 8» (допускается использование другой ОС семейства Linux).

Веб-интерфейс ПО функционирует на следующих веб-браузерах:

- Яндекс.Браузер (версии 23.1.1 и выше);
- Google Chrome (версия 120 и выше);
- Microsoft Edge (версия 120 и выше).

2.3. Требования к уровню подготовки администратора

Для выполнения задач администратор ПО должен иметь опыт работы по администрированию:

- серверного оборудования;
- процессов резервного копирования и восстановления данных.

3. НАЧАЛО ИСПОЛЬЗОВАНИЯ ПРОГРАММЫ

3.1. Вход в веб-интерфейс

Для входа в веб-интерфейс ПО необходимо выполнить следующие действия:

– в адресной строке браузера ввести IP-адрес (указан в приложении к договору) ПО через http-доступ, в открывшейся вкладке появится форма входа (рис. 1);

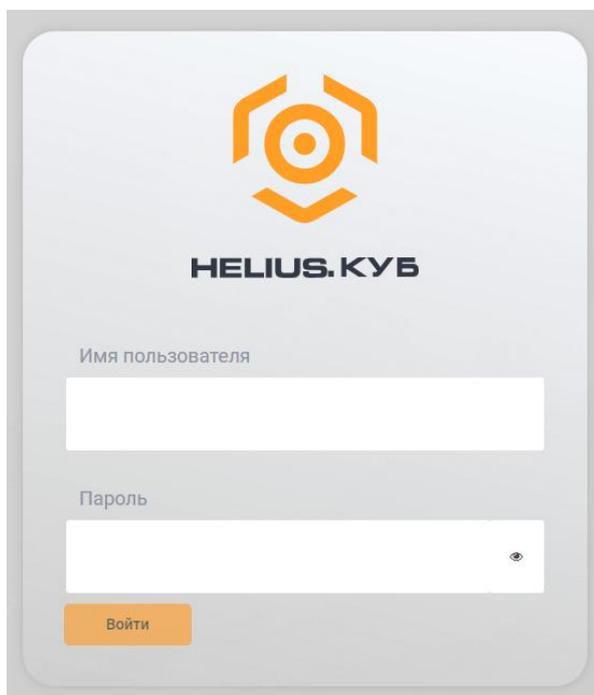


Рисунок 1

– в поле «Имя пользователя» ввести логин, а в поле «Пароль» ввести пароль для входа;
– нажать «Войти».

3.2. Обзор интерфейса и навигация

Интерфейс начальной страницы ПО состоит из нескольких условных частей (рис. 2).

Верхнее меню состоит из нескольких меню: «Зарегистрированный пользователь», «Настройки панели», «Помощь» и «Задачи и уведомления».

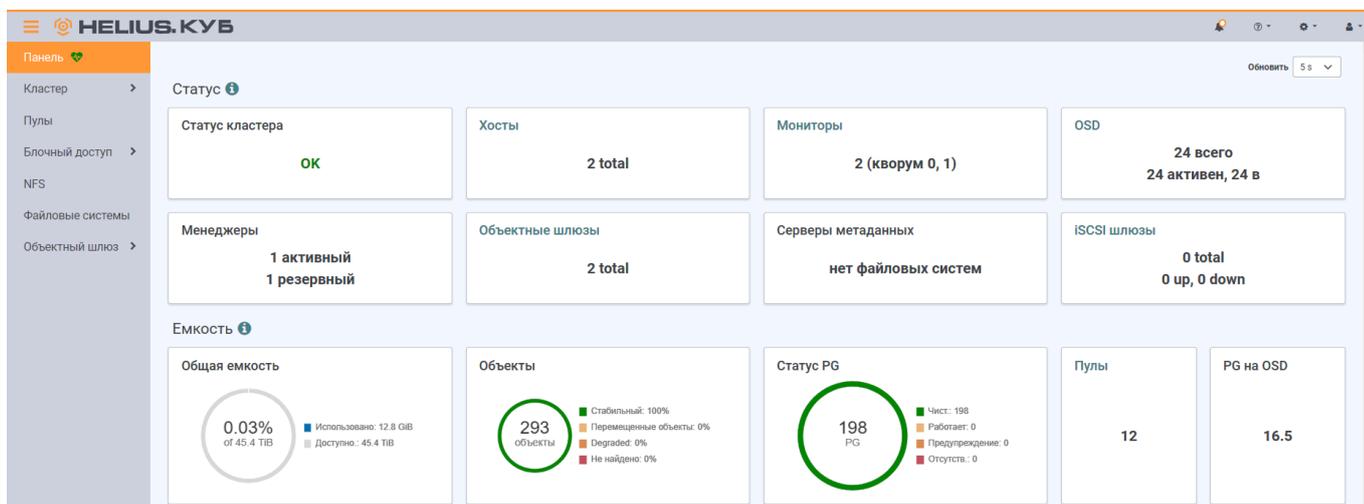


Рисунок 2

3.2.1. Меню «Зарегистрированный пользователь»

В правом верхнем углу отображается имя зарегистрированного пользователя, под которым был произведён вход в ПО. В меню действий отображается список возможных действий с зарегистрированным пользователем (рис. 3).

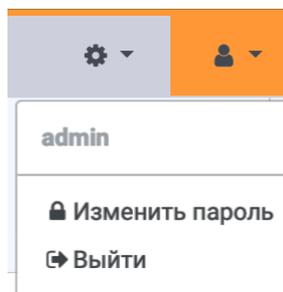


Рисунок 3

3.2.2. Меню «Настройки панели»

Меню «Настройки панели» предоставляет доступ к управлению пользователями (см. подробнее 5.1) и настройке телеметрии (см. подробнее 5.3) (рис. 4).

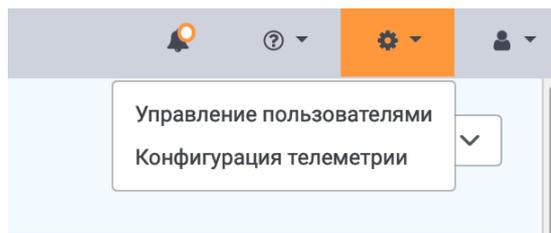


Рисунок 4

3.2.3. Меню «Помощь»

Меню «Помощь» содержит ссылки на документацию и REST API, а также сведения о ПО (рис. 5).

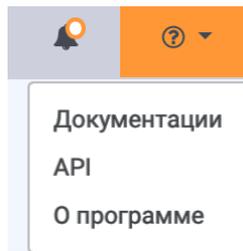


Рисунок 5

При выборе «Документации» будет инициировано скачивание настоящего руководства в формате PDF.

При выборе «API» будет инициировано скачивание файла в формате PDF, содержащего описание API ПО.

При выборе «О программе» появится окно с подробной информацией о ПО (рис. 6).

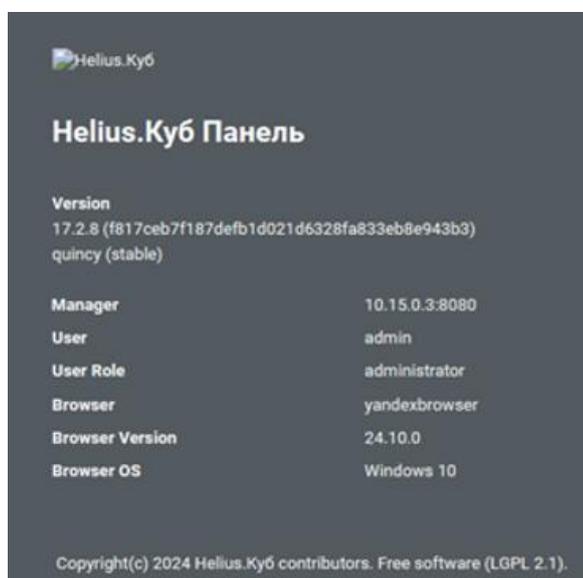


Рисунок 6

3.2.4. Меню «Задачи и уведомления»

В меню «Задачи и уведомления» хранятся и отображаются уведомления о последних задачах текущего пользователя. Уведомления отображаются только для задач, выполненных во время текущего пользовательского сеанса, и очищаются после выхода пользователя (рис. 7).

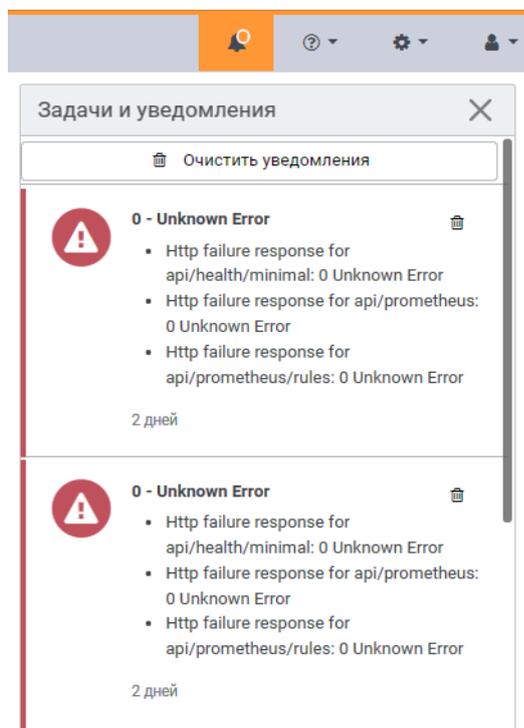


Рисунок 7

Пользователь информируется о каждой задаче с помощью всплывающего уведомления в правом верхнем углу. После закрытия всплывающего окна уведомление будет доступно в меню «Задачи и уведомления».

3.2.5. Главное меню

Навигация по разделам ПО осуществляется в главном меню (рис. 8). При выборе определённого раздела из выпадающего списка (при наличии) открывается форма для работы с данными этого раздела.

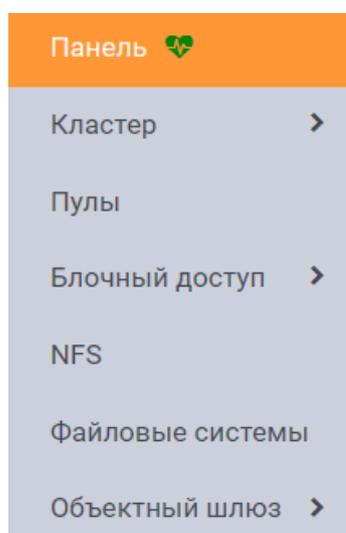


Рисунок 8

3.2.6. Область отображения информации

В области отображения информации выводится содержимое раздела (рис. 9).

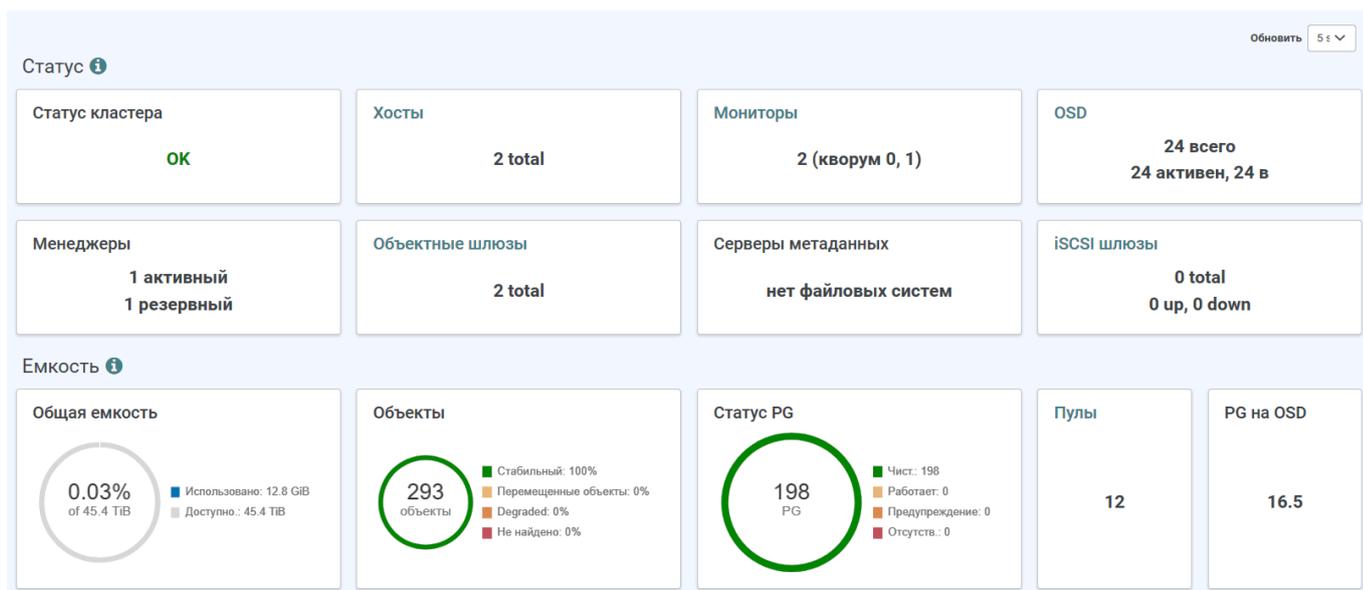


Рисунок 9

3.3. Настройка представлений. Механизмы поиска и фильтрации

3.3.1. Инструмент сортировки

Для всех отображающихся полей доступен инструмент сортировки. Поля сортируются по возрастанию и убыванию.

3.3.2. Поиск в таблице

Некоторые страницы с таблицами имеют дополнительное поле поиска, расположенное над таблицей (рис. 10). Для поиска необходимо ввести ключевое слово и нажать <Enter> на клавиатуре.



Рисунок 10

Результат поиска отобразится в таблице.

3.3.3. Групповые действия

Действие можно запустить в отношении группы предварительно выбранных элементов. Для этого необходимо отметить выбранные элементы (включить флаги, расположенные слева от имени элемента) и выбрать действие на панели инструментов.

3.3.4. Работа с выпадающими списками

При нажатии некоторых кнопок в интерфейсе открываются выпадающие списки параметров и опций, которые могут закрывать другие элементы интерфейса. Чтобы закрыть список после того, как был выбран параметр, нажмите кнопку ещё раз или в любое другое место на странице.

3.4. Доступные операции на панели инструментов

На панели инструментов доступны следующие операции:

- обновить;
- переключить столбы;
- поиск;
- фильтр;
- изменить количество строк для отображения;
- переключение на следующую страницу.

Примечания:

1. Для выбора элемент администрирования необходимо «кликнуть» по его имени.
2. Для просмотра подробной информации об элементе администрирования необходимо развернуть содержимое строки, нажав «>».
3. Все поля, отмеченные звёздочкой (*), являются обязательными для заполнения.

3.5. Выход из веб-интерфейса

Для выхода из веб-интерфейса необходимо перейти в меню «Зарегистрированный пользователь», а затем в меню действий выбрать «Выход», при этом текущий пользователь завершит сеанс работы в ПО.

4. ПАНЕЛЬ УПРАВЛЕНИЯ

Панель управления (панель) – инструмент управления и мониторинга ПО, который используется для проверки ресурсов в кластере и их администрирования.

Панель управления состоит из набора виджетов, которые отображают информацию об общем состоянии кластера, производительности и пропускной способности. Она предоставляет обновления (каждые 5 секунд) в режиме реального времени о любых изменениях в кластере и обеспечивает быстрый доступ к другим разделам панели управления (рис. 11).

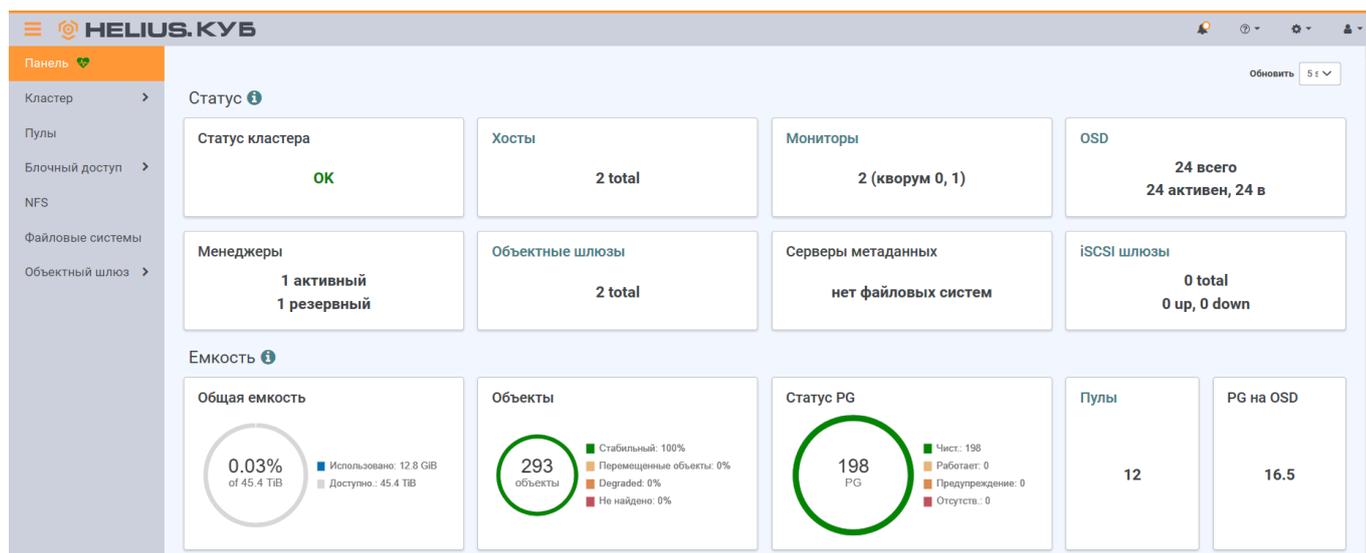


Рисунок 11

Группа виджетов «Статус» предоставляет краткую информацию о текущем состоянии кластера и его компонентов (рис. 12):

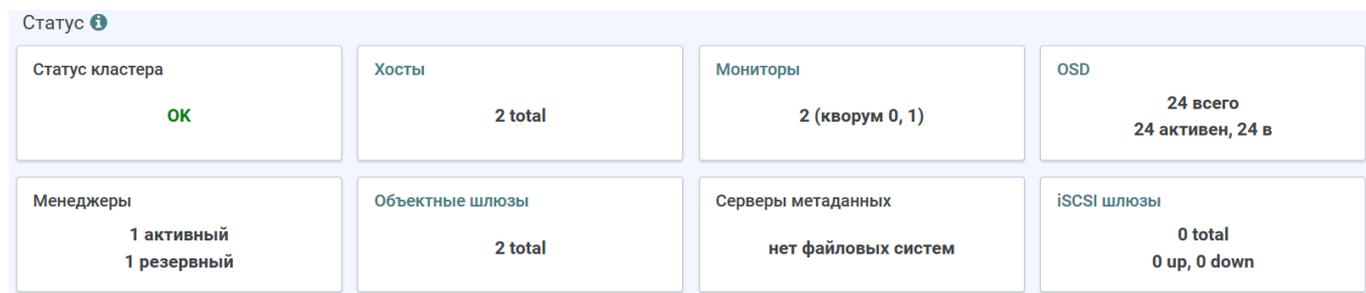


Рисунок 12

- статус кластера – основная информация о состоянии кластера;
- хосты – общее количество хостов кластера;
- мониторы – общее количество запущенных MON и их кворум;

– OSD – общее количество OSD, а также количество включённых («in») и выключенных («out») OSD;

– менеджеры – количество активных и резервных модулей менеджера;

– объектные шлюзы – количество запущенных объектных шлюзов;

– серверы метаданных – количество MDS;

– iSCSI шлюзы – количество настроенных шлюзов iSCSI.

Кластер находится в работоспособном состоянии, если его статус «ОК» и статусы всех PG «active+clean».

Статус кластера может быть «Health_warn» в ряде случаев, связанных с пиригом PG (в таких случаях кластер восстановится самостоятельно):

– пул только создан, и PG ещё не выполнили пириг;

– PG восстанавливаются;

– только что было добавлено OSD в кластер или удалено OSD из кластера;

– CRUSH карта была изменена и PG перемещаются;

– в разных репликах PG имеются противоречивые данные;

– копии PG удаляются;

– ПО не располагает достаточным объёмом хранилища для завершения операции резервного копирования.

Группа виджетов «Ёмкость» отображает соотношение использованных и доступных объёмов хранения кластера, количество объектов данных, хранящихся в кластере, диаграмму групп размещения в соответствии с их статусом, количество пулов в кластере и среднее количество PG на OSD (рис. 13).

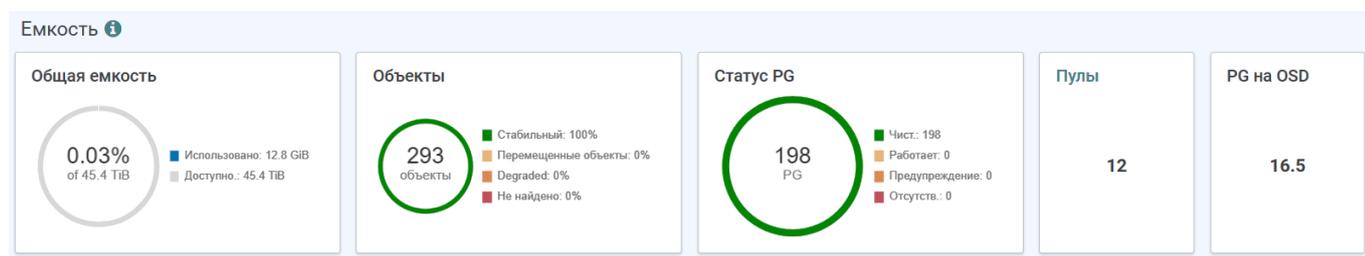


Рисунок 13

Виджет «Общая ёмкость» отображает следующую информацию:

– использовано – использованный объём из общего физического объёма;

– доступно – общий физический объём, предоставленный хостами хранения (OSD);

– предупреждение – пороговое значение (0,85 или 85% от объёма использования);

– опасность – полное пороговое значение (0,95 или 95% от объёма использования).

Когда объём использования кластера приближается к максимальному, в качестве меры безопасности для предотвращения потери данных запрещается запись или чтение в OSD. По умолчанию коэффициент полной загрузки составляет 0,95 или 95% от объёма использования.

Группа виджетов «Производительность» отображает основные данные о производительности клиентов (рис. 14):

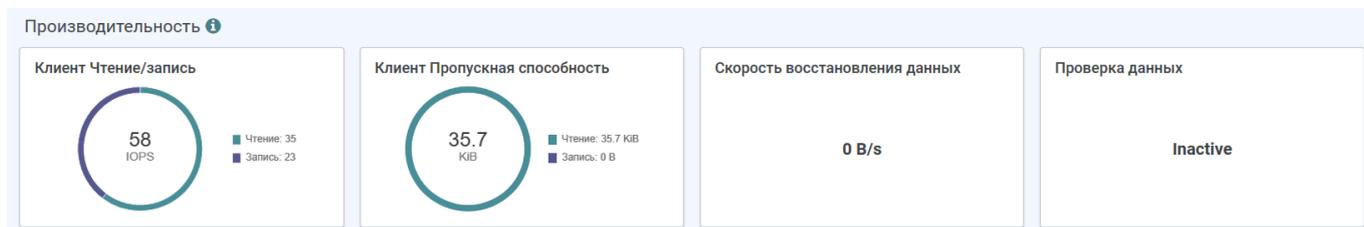


Рисунок 14

– чтение/запись клиента – количество операций чтения и записи (IOPS), выполняемых клиентами в секунду;

– пропускная способность клиента – объём данных, передаваемых клиентами ПО в байт в секунду;

– скорость восстановления данных – пропускная способность восстанавливаемых данных в секунду.

– проверка данных – статус «Inactive» означает, что PG неактивна и не может обрабатывать запросы (требуется диагностика и восстановление). В ПО статусы проверки данных относятся к процессу проверки и восстановления данных (см. подробнее 6.5.6).

В процессе проверки данных объекты и PG могут находиться в следующих состояниях:

- «active»+«clean» – PG активна и все данные согласованы;
- «active»+«scrubbing» – PG находится в процессе лёгкой проверки;
- «active»+«deep»+«scrubbing» – PG находится в процессе глубокой проверки;
- «active»+«degraded» – PG активна, но некоторые объекты недоступны или повреждены (требуется восстановление данных);
- «active»+«recovering» – PG активна и восстанавливает данные (восстановление данных после обнаружения ошибок);
- «active»+«recovery_wait» – PG ожидает начала восстановления данных (восстановление данных запланировано, но ещё не началось);
- «active»+«inconsistent» – PG активна, но обнаружены несоответствия в данных (требуется восстановление данных);
- «active»+«repair» – PG активна и находится в процессе восстановления данных (восстановление данных после обнаружения несоответствий);

– «inactive» – PG неактивна и не может обрабатывать запросы (требуется диагностика и восстановление).

5. НАСТРОЙКИ ПАНЕЛИ

Меню «Настройки панели» позволяет управлять пользователями и настраивать параметры телеметрии.

5.1. Пользователи

ПО поддерживает многопользовательское управление, когда несколько учётных записей пользователей с разными разрешениями (ролями).

Для управления пользователями необходимо в меню «Настройки панели» выбрать «Управление пользователями» (рис. 15).

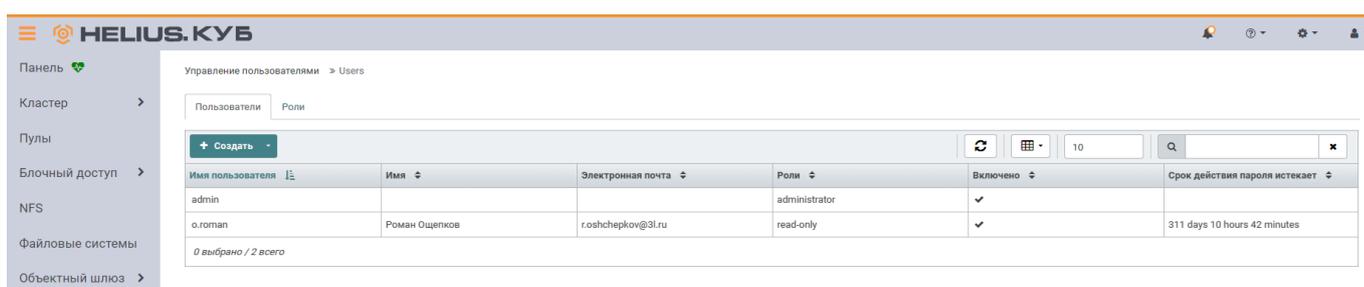


Рисунок 15

5.1.1. Создание пользователя

Для создания нового пользователя (новой учётной записи) необходимо на вкладке «Пользователи» на панели инструментов нажать «Создать». В открывшемся окне заполнить предлагаемую форму (рис. 16).

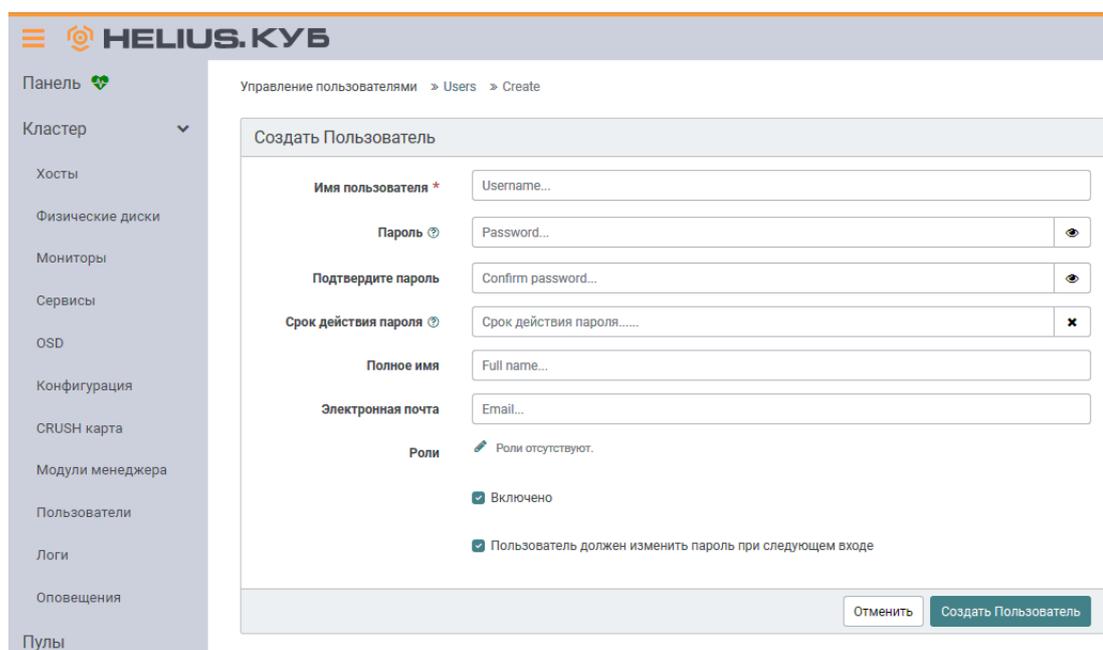


Рисунок 16

В поле «Имя пользователя» указать имя пользователя (пробелы в начале и в конце будут удалены).

В поле «Пароль» указать пароль (хранится в зашифрованном виде) для создаваемого пользователя.

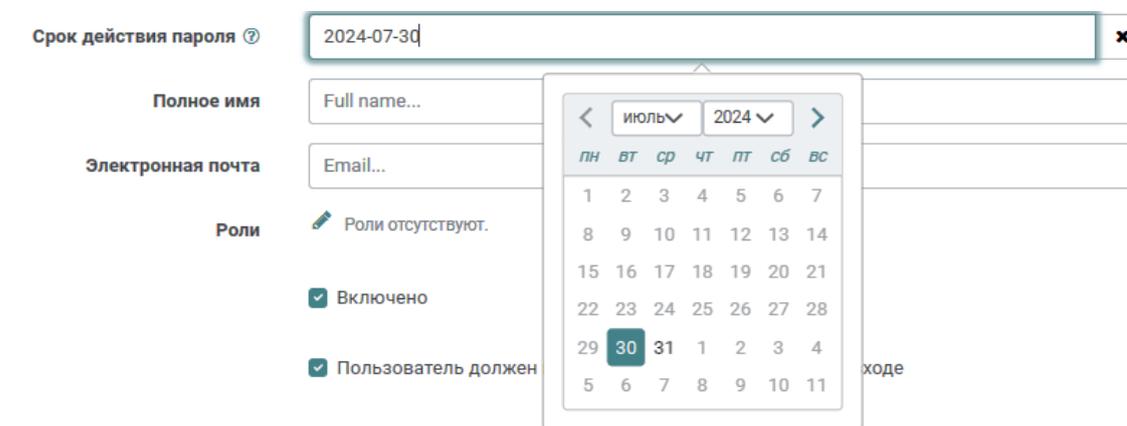
К паролям предъявляются следующие требования:

- должен содержать не менее 8 символов;
- не должен совпадать с предыдущим.

В поле «Подтвердите пароль» повторно ввести указанный выше пароль.

При создании пользователя можно настроить правила сложности паролей, требующие от пользователей менять свой пароль после первого входа в ПО или по истечении настраиваемого периода времени.

В поле «Срок действия пароля» установить срок действия пароля (по умолчанию – 0). При необходимости можно указать дату, при наступлении которой срок действие пароля истечёт, и пользователю будет предложено сменить пароль (рис. 17).



Срок действия пароля ⓘ 2024-07-30 ✕

Полное имя Full name...

Электронная почта Email...

Роли  Роли отсутствуют.

- Включено
- Пользователь должен

хоче

Рисунок 17

В поле «Полное имя» указать полное имя пользователя.

В поле «Электронная почта» указать адрес электронной почты пользователя.

Поле «Роли» позволяет назначить определённые роли пользователю. Пользователю можно назначить предопределённые системные роли или создать новую роль с необходимыми разрешениями для областей безопасности (см. подробнее 5.2.1).

Ниже приведён список системных ролей:

- administrator – предоставляет полные разрешения для всех областей безопасности (см. подробнее 5.2.1);
- block-manager – предоставляет полные разрешения для областей безопасности «rbd-image», «rbd-mirroring» и «iscsi»;
- fs-manager – предоставляет полные разрешения для области безопасности «fs»;

– cluster-manager – предоставляет полные разрешения для областей безопасности «hosts», «osd», «monitor», «manager» и «config-opt»;

– ganesha-manager – предоставляет полные разрешения для области безопасности «nfs-ganesha»;

– pool-manager – предоставляет полные разрешения для области безопасности «pool»;

– read-only – разрешает «чтение» для всех областей безопасности, кроме «dashboard-settings»;

– rgw-manager – предоставляет полные разрешения для области безопасности «rgw».

Опция «Включено» позволяет включить учётную запись пользователя.

Опция «Пользователь должен изменить пароль при следующем входе» обязывает пользователя сменить пароль после первого входа в ПО.

Нажать «Создать Пользователь».

5.1.2. Редактирование пользователя

Для редактирования пользователя необходимо выбрать пользователя и на панели инструментов нажать «Редактировать». В открывшемся окне обновить изменяемые свойства пользователя и нажать «Редактировать Пользователь».

5.1.3. Смена пароля пользователя

Для смены пароля пользователя необходимо в меню действий выбрать «Изменить пароль». В открывшемся одноименном окне заполнить предлагаемую форму (рис. 18).

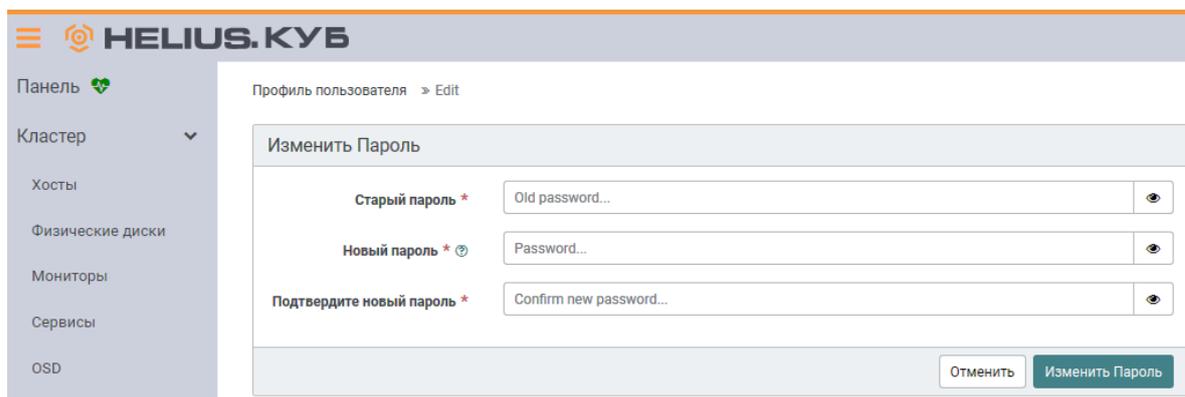


Рисунок 18

В поле «Старый пароль» ввести старый пароль пользователя.

В поле «Новый пароль» ввести новый пароль пользователя.

К паролям предъявляются следующие требования:

– должен содержать не менее 8 символов;

– не должен совпадать с предыдущим.

В поле «Подтвердите новый пароль» повторить введение нового пароля.

Нажать «Изменить Пароль».

5.1.4. Удаление пользователя

Для удаления пользователя необходимо выбрать пользователя и в меню действий выбрать «Удалить». Подтвердить удаление в появившемся окне.

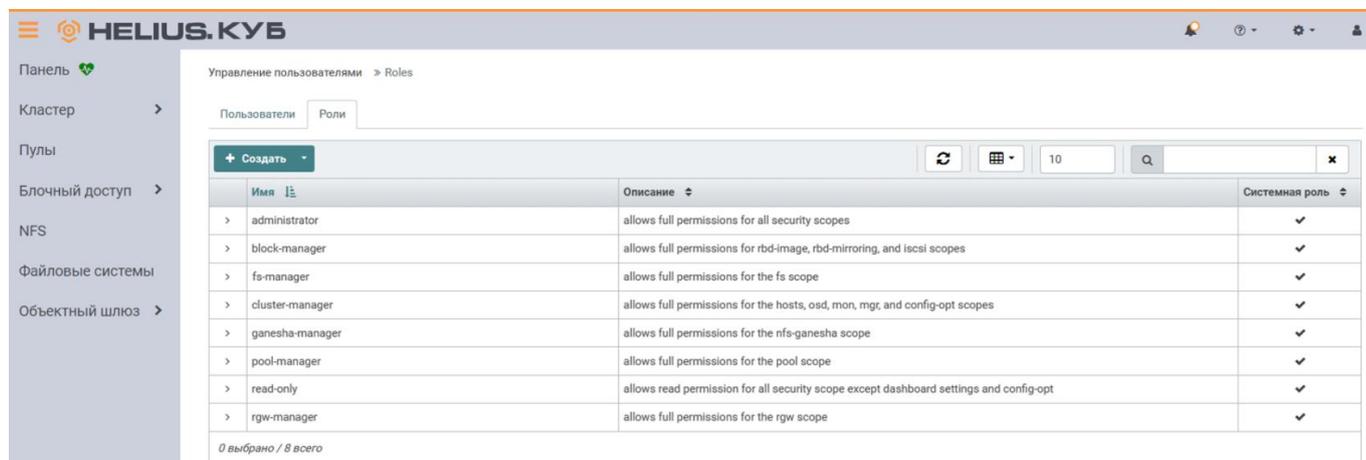
5.2. Роли

Учётные записи пользователей связаны с набором ролей, которые определяют, к каким функциям панели управления можно получить доступ.

Роль определяет набор сопоставлений между областью безопасности и набором разрешений (набор назначенных прав и привилегий пользователя для выполнения определённого набора операций).

Панель управления предоставляет набор уже predefined ролей, которые называются системными ролями (см. подробнее 0).

Для управления ролями необходимо перейти на вкладку «Роли» (рис. 19).

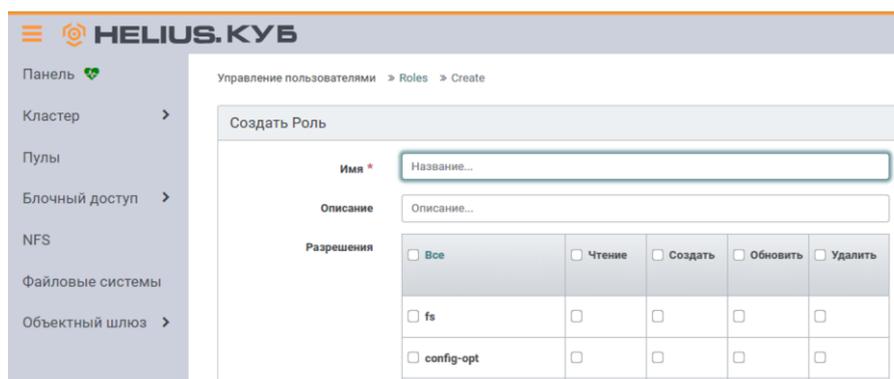


| Имя | Описание | Системная роль |
|-----------------|--|----------------|
| administrator | allows full permissions for all security scopes | ✓ |
| block-manager | allows full permissions for rbd-image, rbd-mirroring, and iscsi scopes | ✓ |
| fs-manager | allows full permissions for the fs scope | ✓ |
| cluster-manager | allows full permissions for the hosts, osd, mon, mgr, and config-opt scopes | ✓ |
| ganesha-manager | allows full permissions for the nfs-ganesha scope | ✓ |
| pool-manager | allows full permissions for the pool scope | ✓ |
| read-only | allows read permission for all security scope except dashboard settings and config-opt | ✓ |
| rgw-manager | allows full permissions for the rgw scope | ✓ |

Рисунок 19

5.2.1. Создание роли

Для создания новой роли необходимо нажать «Создать» на панели инструментов. В открывшемся окне заполнить предлагаемую форму (рис. 20).



| Имя * | Описание | Все | Чтение | Создать | Обновить | Удалить |
|-------------|-------------|-------------------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Название... | Описание... | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| fs | | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| config-opt | | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |

Рисунок 20

В поле «Имя» указать название роли.

В поле «Описание» указать описание создаваемой роли.

В поле «Разрешения» выбрать разрешения для функций панели управления. Существует четыре типа разрешений: чтение, создание, обновление и удаление.

Функции панели управления сгруппированы в области безопасности, которые предопределены и статичны. Доступны следующие области безопасности:

- fs – функции, связанные с управлением файловой системой Helius.КубFS;
- config-opt – функции, связанные с управлением параметрами конфигурации;
- dashboard-settings – позволяет изменять настройки панели управления;
- grafana – функции, связанные с управлением прокси-сервером Grafana;
- hosts – функции, связанные с управлением хостами;
- iscsi – функции, связанные с управлением iSCSI;
- log – функции, связанные с управлением журналами (логами);
- manager – функции, связанные с управлением модулями менеджера;
- monitor – функции, связанные с управлением MON;
- nfs-ganesha – функции, связанные с управлением NFS Ganesha;
- osd – функции, связанные с управлением OSD;
- pool – функции, связанные с управлением пулом;
- prometheus – функции, связанные с управлением оповещениями Prometheus;
- rbd-image – функции, связанные с управлением RBD-образами;
- rbd-mirroring – функции, связанные с управлением RBD-зеркалированием;
- rgw – функции, связанные с управлением объектным шлюзом;
- user – функции, связанные с управлением пользователями.

Активировав флаг «Все» (в поле «Разрешения»), можно активировать сразу все разрешения для всех областей безопасности.

Нажать «Создать Роль». После создания роли, её можно назначить определённому пользователю.

5.2.2. Просмотр информации о роли

Для просмотра информации о роли необходимо развернуть содержимое строки (рис. 21).

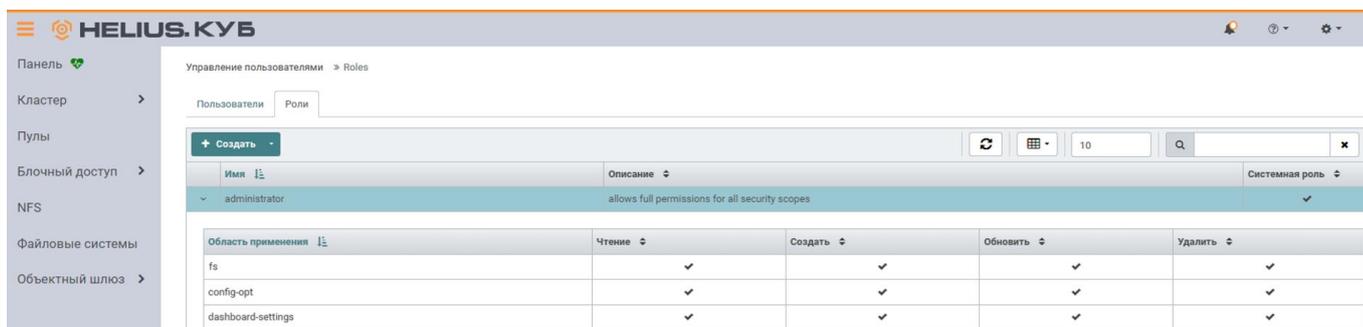


Рисунок 21

5.2.3. Клонирование роли

Клонирование роли позволяет назначить разрешения уже существующим системным ролям.

Для клонирования роли необходимо выполнить следующие действия:

- выбрать роль, которую необходимо клонировать;
- в меню действий выбрать «Клон.»;
- в открывшемся окне в поле «Новое имя» указать новое имя для клонированной роли и нажать «Клонировать роль» (рис. 22).

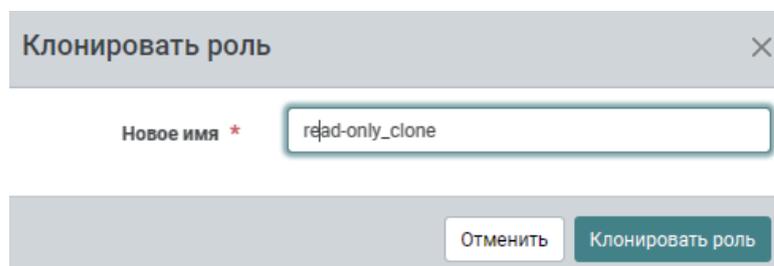


Рисунок 22

После клонирования роли можно настроить дополнительные разрешения в соответствии с требованиями.

5.2.4. Редактирование роли

Для редактирования роли необходимо выбрать роль и нажать «Редактировать» на панели инструментов. В открывшемся окне обновить изменяемые свойства роли и нажать «Редактировать Роль».

5.2.5. Удаление роли

Для удаления роли необходимо выбрать роль и в меню действий выбрать «Удалить». Подтвердить удаление в появившемся окне.

Примечание. Системные роли недоступны для удаления.

5.3. Конфигурация телеметрии

Для настройки конфигурации телеметрии необходимо в меню действий (меню «Настройки панели») выбрать «Конфигурация телеметрии» (рис. 23).

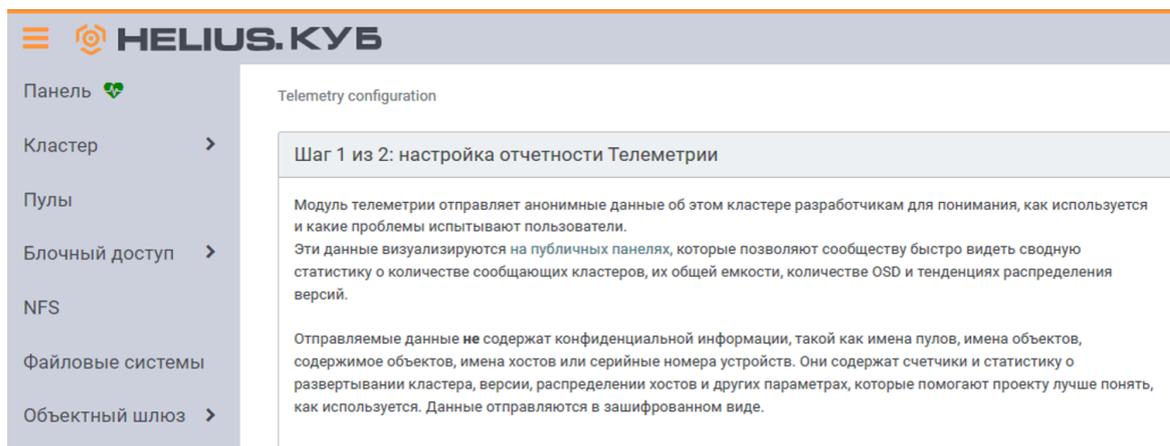


Рисунок 23

Настройка конфигурации телеметрии состоит из двух шагов.

Шаг 1 из 2: Настройка отчётности телеметрии:

Модуль телеметрии отправляет анонимные данные об этом кластере разработчикам для понимания, как используется кластер и какие проблемы испытывают пользователи.

Эти данные визуализируются на публичных панелях, которые позволяют видеть сводную статистику о количестве кластеров, их общей производительности, количестве OSD и тенденциях распределения версий.

Отправляемые данные не содержат конфиденциальной информации, такой как имена пулов, имена объектов, содержимое объектов, имена хостов или серийные номера устройств. Они содержат счётчики и статистику о развёртывании кластера, версии, распределении хостов и других параметров. Данные отправляются в зашифрованном виде.

Отчёт телеметрии разделён на несколько каналов, каждый из которых содержит различные типы информации, которые можно настроить (рис. 24).

Канал «Базовый» содержит базовую информацию о кластере:

- ёмкость кластера;
- количество мониторов, менеджеров, OSD, MDS, объектных шлюзов или других служб;
- используемая версия ПО;
- количество и типы пулов RADOS и файловых систем;
- названия параметров конфигурации, которые были изменены относительно их значений по умолчанию (но не значения этих параметров).

Каналы

Отчет телеметрии разделен на несколько "каналов", каждый из которых содержит различные типы информации, которые можно настроить ниже.

| | |
|-----------------|-------------------------------------|
| Базовый ? | <input checked="" type="checkbox"/> |
| Сбой ? | <input checked="" type="checkbox"/> |
| Устройство ? | <input checked="" type="checkbox"/> |
| Идентификация ? | <input type="checkbox"/> |
| Производ. ? | <input type="checkbox"/> |

Рисунок 24

Канал «Сбой» содержит информацию о сбоях служб:

- тип службы;
- версия службы;
- ОС (дистрибутив ОС, версия ядра);
- трассировка стека, указывающая место в коде, где произошёл сбой.

Канал «Устройство» содержит информацию о метриках устройства (например, анонимизированные метрики SMART).

Канал «Идентификация» содержит идентификационную информацию о кластере, предоставленную пользователем:

- описание кластера;
- контактный адрес электронной почты.

При выборе канала «Идентификация» на панели «Контактная информация» в полях «Контакт», «Описание» и «Организация» будет отображена контактная информация (рис. 25).

Контактная информация ?

| | |
|-------------|--|
| Контакт | <input type="text" value="Example User <user@example.com>"/> |
| Описание | <input type="text" value="Мой первый кластер"/> |
| Организация | <input type="text" value="Название организации"/> |

Рисунок 25

Примечание. Предоставление контактной информации необязательно и по умолчанию отключено.

Опция «Производ.» содержит различные метрики производительности кластера, которые можно использовать для следующих целей:

- отображение общей работоспособности кластера;

- определение шаблонов рабочей нагрузки;
- устранение неполадок, связанных с задержкой, регулированием, управлением памятью и другими подобными проблемами;
- контроль производительность кластера с помощью служб.

Также для заполнения доступны дополнительные настройки (рис. 26).

Дополнительные настройки



Рисунок 26

В поле «Интервал» указать интервал (в часах) составления и отправки отчёта. Модуль телеметрии по умолчанию составляет и отправляет новый отчёт каждые 24 часа.

В поле «Прокси» указать адрес прокси-сервера. Если кластер не может напрямую подключиться к настроенной конечной точке телеметрии, то можно настроить HTTP/HTTPS прокси-сервер. При необходимости можно включить имя пользователя и пароль.

Нажать «Далее».

Шаг 2 из 2: Предварительный просмотр отчёта телеметрии (рис. 27).

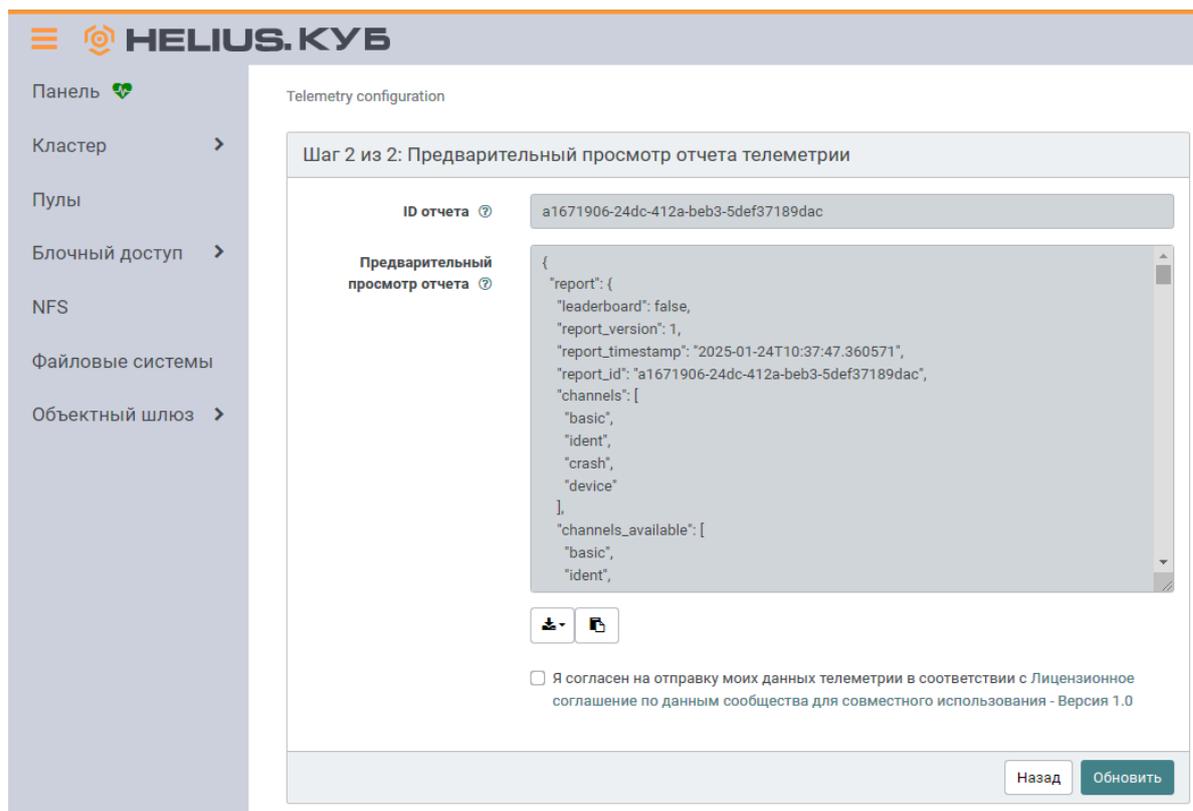


Рисунок 27

Поле «ID отчета» отображает случайный UUID для идентификации определённого кластера в течение нескольких отчетов телеметрии (поле заполнено по умолчанию).

Поле «Предварительный просмотр отчёта» отображает фактические данные телеметрии, которые будут отправлены (поле заполнено по умолчанию).

Полный отчёт, включая метрики из канала производительности, можно загрузить (выбрав «Скачать») или скопировать в буфер обмена. Файл отчёта скачается в формате JSON.

Перед отправкой данных телеметрии необходимо предварительно ознакомиться с полным отчётом. При необходимости подтвердить передачу данных телеметрии в соответствии с лицензионным соглашением. Нажать «Обновить».

Передаваемые данные не содержат конфиденциальных данных, таких как имена пулов, имена объектов, содержимое объектов, имена хостов или серийные номера устройств. Они содержат счётчики и статистику о том, как развернут кластер, версию ПО, распределение хостов и другие параметры.

Данные защищены и отправляются на сервер для обработки.

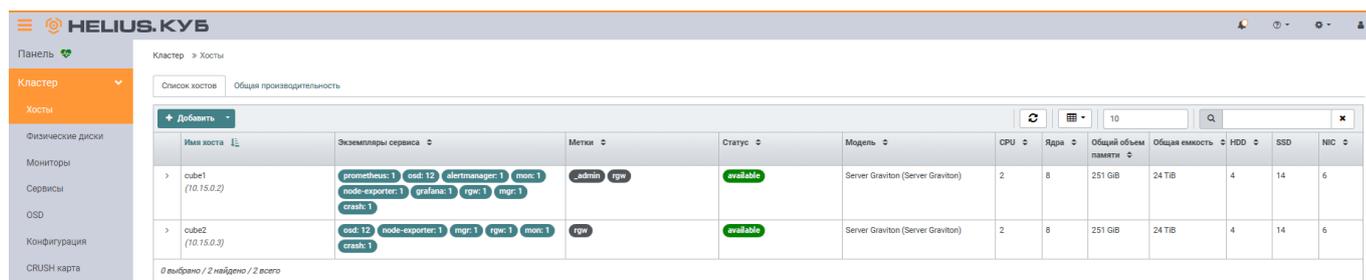
6. КЛАСТЕР ХРАНЕНИЯ

Кластерная система хранения данных вмещает большое количество хостов, которые взаимодействуют друг с другом для динамической репликации и перераспределения данных.

6.1. Хосты

Основной функцией хоста является предоставление физических ресурсов (дисковое пространство, RAM, CPU).

В разделе «Кластер» из выпадающего списка выбрать подраздел «Хосты» (рис. 28).



| Имя хоста | Экземпляры сервиса | Метки | Статус | Модель | CPU | Ядра | Общий объем памяти | Общая емкость | HDD | SSD | NIC |
|----------------------|---|-------------|-----------|-----------------------------------|-----|------|--------------------|---------------|-----|-----|-----|
| cube1 (10.15.0.2) | prometheus: 1, osd: 12, alertmanager: 1, mon: 1, node-exporter: 1, grafana: 1, rgw: 1, mgr: 1, crash: 1 | _admin, rgw | available | Server Graviton (Server Graviton) | 2 | 8 | 251 GIB | 24 TIB | 4 | 14 | 6 |
| cube2 (10.15.0.3) | osd: 12, node-exporter: 1, mgr: 1, rgw: 1, mon: 1, crash: 1 | rgw | available | Server Graviton (Server Graviton) | 2 | 8 | 251 GIB | 24 TIB | 4 | 14 | 6 |

Рисунок 28

6.1.1. Создание хоста

На вкладке «Список хостов» нажать «Добавить» на панели инструментов. В открывшемся окне заполнить предлагаемую форму (рис. 29).

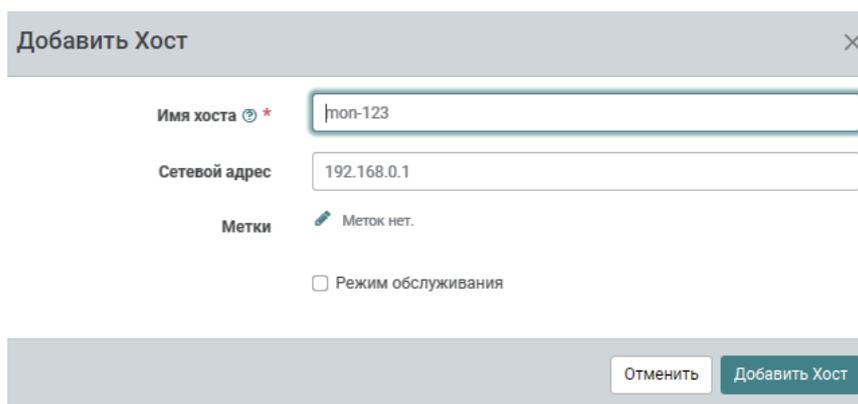


Рисунок 29

В поле «Имя хоста» указать имя хоста. Чтобы добавить несколько хостов одновременно, необходимо ввести:

- список имён хостов, разделённых запятыми;
- выражение диапазона;
- выражение диапазона с запятыми.

В поле «Сетевой адрес» указать сетевой адрес.

В поле «Метки» указать метки. Метки используются для указания размещения служб. Метки имеют свободную форму и не имеют определённого значения. Каждый хост может иметь несколько меток.

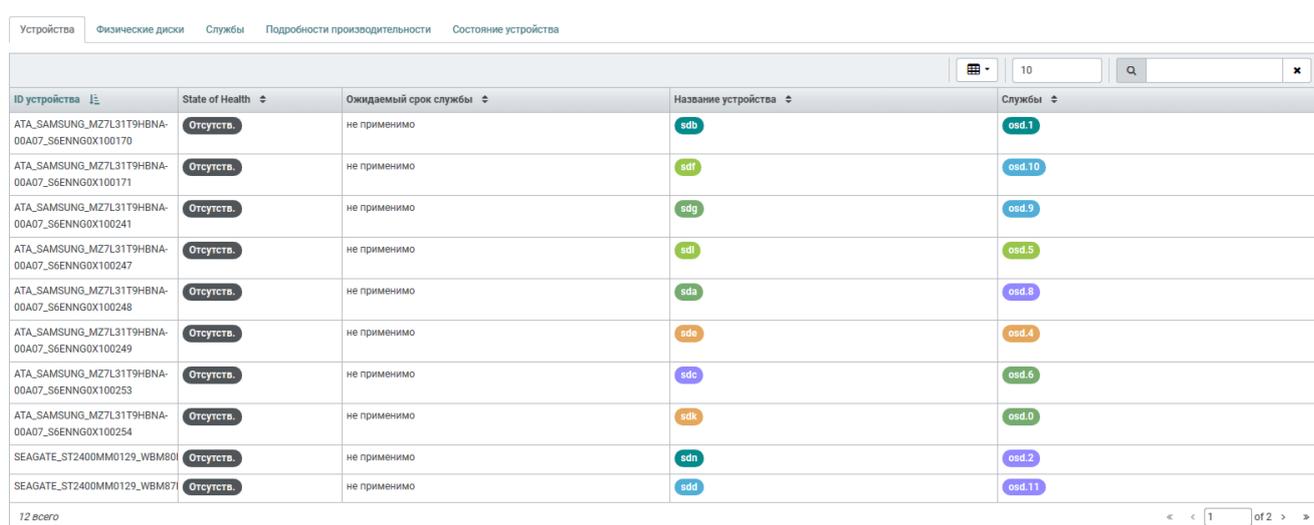
При необходимости выбрать опцию «Режим обслуживания» (см. подробнее 6.1.4).

Нажать «Добавить Хост», при этом созданный хост отобразится в общем списке.

6.1.2. Просмотр информации о хосте

Для просмотра информации о хосте необходимо развернуть содержимое строки. Информация о хосте представлена на вкладках «Устройство», «Физические диски», «Службы», «Подробности производительности» и «Состояние устройства».

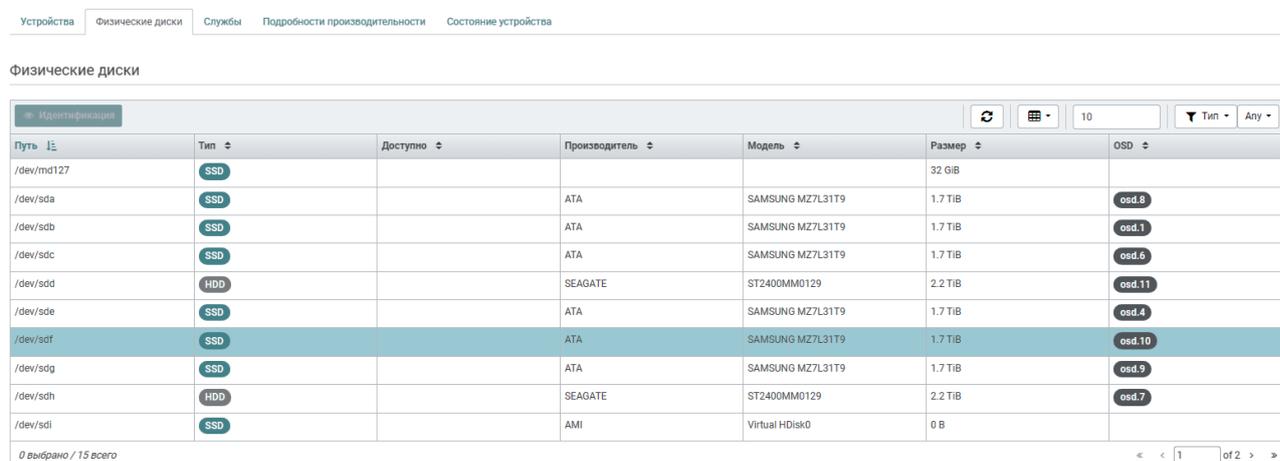
Вкладка «Устройство» отображает подключённые устройства и информацию о них (рис. 30).



| ID устройства | State of Health | Ожидаемый срок службы | Название устройства | Службы |
|--|-----------------|-----------------------|---------------------|--------|
| ATA_SAMSUNG_MZ7L31T9HBN-00A07_S6ENNGDX100170 | Отсутств. | не применимо | sdb | osd.1 |
| ATA_SAMSUNG_MZ7L31T9HBN-00A07_S6ENNGDX100171 | Отсутств. | не применимо | sdf | osd.10 |
| ATA_SAMSUNG_MZ7L31T9HBN-00A07_S6ENNGDX100241 | Отсутств. | не применимо | sdg | osd.9 |
| ATA_SAMSUNG_MZ7L31T9HBN-00A07_S6ENNGDX100247 | Отсутств. | не применимо | sdi | osd.5 |
| ATA_SAMSUNG_MZ7L31T9HBN-00A07_S6ENNGDX100248 | Отсутств. | не применимо | sda | osd.8 |
| ATA_SAMSUNG_MZ7L31T9HBN-00A07_S6ENNGDX100249 | Отсутств. | не применимо | sde | osd.4 |
| ATA_SAMSUNG_MZ7L31T9HBN-00A07_S6ENNGDX100253 | Отсутств. | не применимо | sdh | osd.6 |
| ATA_SAMSUNG_MZ7L31T9HBN-00A07_S6ENNGDX100254 | Отсутств. | не применимо | sdk | osd.0 |
| SEAGATE_ST2400MM0129_WBM801 | Отсутств. | не применимо | sdn | osd.2 |
| SEAGATE_ST2400MM0129_WBM871 | Отсутств. | не применимо | sdd | osd.11 |

Рисунок 30

Вкладка «Физические диски» отображает список подключённых физических дисков и подробную информацию о них (рис. 31).



| Путь | Тип | Доступно | Производитель | Модель | Размер | OSD |
|------------|-----|----------|---------------|------------------|---------|--------|
| /dev/md127 | SSD | | | | 32 GiB | |
| /dev/sda | SSD | | ATA | SAMSUNG MZ7L31T9 | 1.7 TiB | osd.8 |
| /dev/sdb | SSD | | ATA | SAMSUNG MZ7L31T9 | 1.7 TiB | osd.1 |
| /dev/sdc | SSD | | ATA | SAMSUNG MZ7L31T9 | 1.7 TiB | osd.6 |
| /dev/sdd | HDD | | SEAGATE | ST2400MM0129 | 2.2 TiB | osd.11 |
| /dev/sde | SSD | | ATA | SAMSUNG MZ7L31T9 | 1.7 TiB | osd.4 |
| /dev/sdf | SSD | | ATA | SAMSUNG MZ7L31T9 | 1.7 TiB | osd.10 |
| /dev/sdg | SSD | | ATA | SAMSUNG MZ7L31T9 | 1.7 TiB | osd.9 |
| /dev/sdh | HDD | | SEAGATE | ST2400MM0129 | 2.2 TiB | osd.7 |
| /dev/sdi | SSD | | AMI | Virtual Hdisk0 | 0 B | |

Рисунок 31

Идентификация устройства облегчает замену неисправных физических дисков.

Для идентификации устройства необходимо выбрать устройство из списка и на панели инструментов нажать «Идентификация». В открывшемся окне «Идентификация устройства» в поле «Пожалуйста, введите продолжительность мигания LED» из выпадающего списка выбрать продолжительность мигания LED и нажать «Выполнить» (рис. 32).

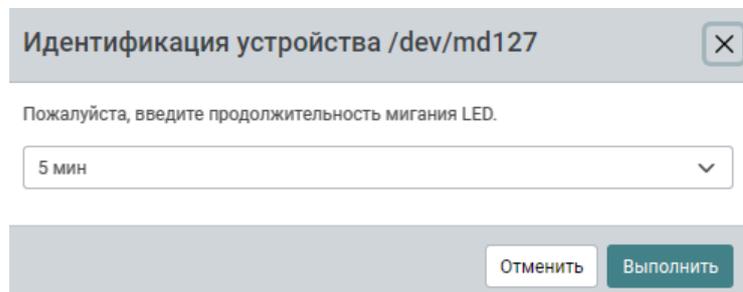
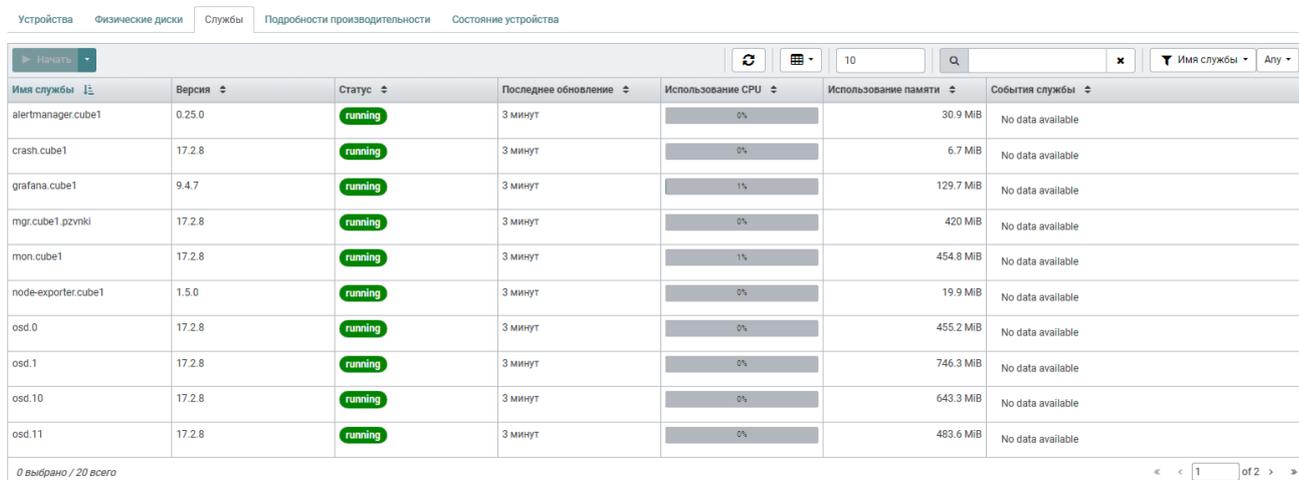


Рисунок 32

Вкладка «Службы» отображает службы, развёрнутые на хосте, и подробную информацию о них (рис. 33).



| Имя службы | Версия | Статус | Последнее обновление | Использование CPU | Использование памяти | События службы |
|---------------------|--------|---------|----------------------|-------------------|----------------------|-------------------|
| alertmanager.cube1 | 0.25.0 | running | 3 минут | 0% | 30.9 MIB | No data available |
| crash.cube1 | 17.2.8 | running | 3 минут | 0% | 6.7 MIB | No data available |
| grafana.cube1 | 9.4.7 | running | 3 минут | 1% | 129.7 MIB | No data available |
| mgm.cube1.pzvnki | 17.2.8 | running | 3 минут | 0% | 420 MIB | No data available |
| mon.cube1 | 17.2.8 | running | 3 минут | 1% | 454.8 MIB | No data available |
| node-exporter.cube1 | 1.5.0 | running | 3 минут | 0% | 19.9 MIB | No data available |
| osd.0 | 17.2.8 | running | 3 минут | 0% | 455.2 MIB | No data available |
| osd.1 | 17.2.8 | running | 3 минут | 0% | 746.3 MIB | No data available |
| osd.10 | 17.2.8 | running | 3 минут | 0% | 643.3 MIB | No data available |
| osd.11 | 17.2.8 | running | 3 минут | 0% | 483.6 MIB | No data available |

Рисунок 33

Для того чтобы запустить службу, необходимо выбрать службу и на панели инструментов нажать «Начать», при этом статус службы изменится на «running».

Для того чтобы остановить службу, необходимо выбрать службу и в меню действий выбрать «Остановить», при этом статус службы измениться на «stop».

Перезапуск службы может потребоваться для применения изменений конфигурации, устранения сбоев или выполнения технического обслуживания. Для того чтобы перезапустить службу, необходимо выбрать службу и в меню действий выбрать «Запустить заново», при этом статус службы измениться на «running».

Развёртывание службы заново – процесс повторного создания и настройки службы после её удаления, сбоя или изменения конфигурации. Для того чтобы повторно развернуть службу,

необходимо выбрать службу и в меню действий выбрать «Развернуть заново», при этом статус службы изменится на «running».

Вкладка «Подробности производительности» показывает подробную информацию о производительности выбранного хоста (рис. 34).

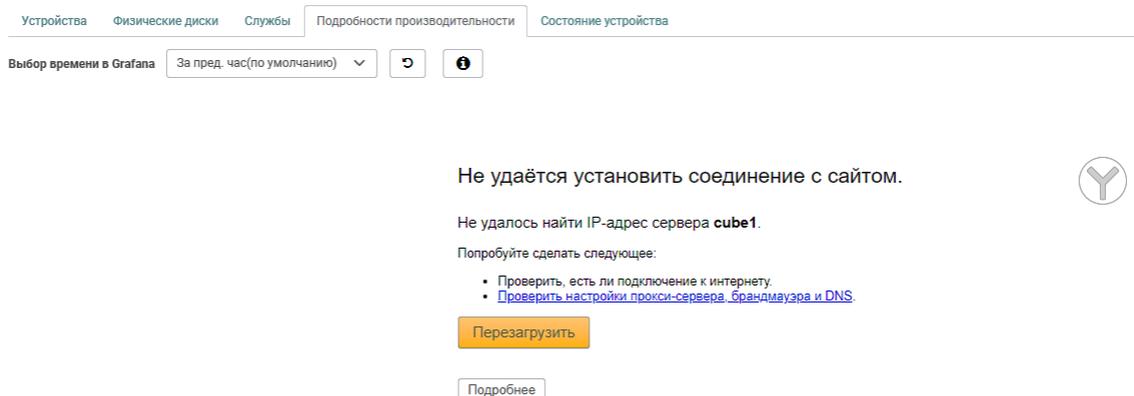


Рисунок 34

В поле «Выбор времени в Grafana» из выпадающего списка выбрать интервал времени отображения подробной информации о производительности выбранного хоста.

Кнопка «Сбросить настройки» позволяет сбросить настройки выбора времени в Grafana.

Кнопка «Показать скрытую информацию» позволяет просмотреть скрытую информацию.

Примечание. Если ниже не появилась встроенная панель Grafana, пожалуйста, перейдите по указанной в скрытой информации ссылке, чтобы проверить доступность Grafana и отсутствие проблем с сертификатами HTTPS. Возможно, вам потребуется перезагрузить эту страницу после принятия исключений сертификатов браузера.

Вкладка «Состояние устройства» отображает состояние подключённых устройств (отображённых на вкладке «Устройства») и состоит из вкладок «Информация об устройстве» (рис. 35) и «SMART» (рис. 36).

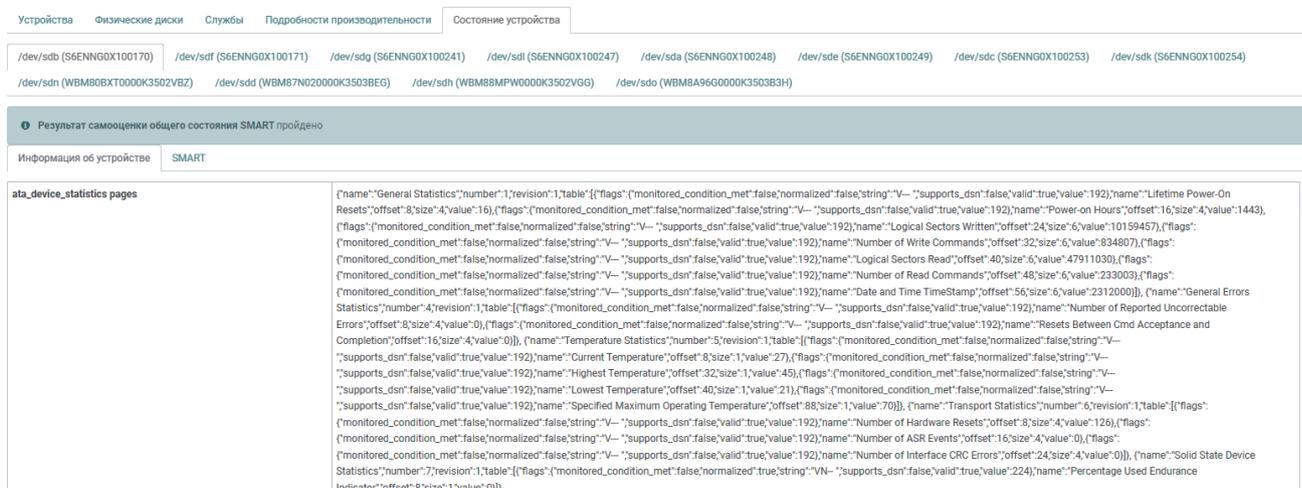


Рисунок 35

Устройства Физические диски Службы Подробности производительности Состояние устройства

/dev/sdb (S6ENNG0X100170) /dev/sdf (S6ENNG0X100171) /dev/sdg (S6ENNG0X100241) /dev/sdl (S6ENNG0X100247) /dev/sda (S6ENNG0X100248) /dev/sde (S6ENNG0X100249) /dev/sdc (S6ENNG0X100253) /dev/sdk (S6ENNG0X100254)
 /dev/sdn (WBM80BXT0000K3502VBZ) /dev/sdd (WBM87N02000K3503BEG) /dev/sdh (WBM88MPW0000K3502VGG) /dev/sdo (WBM8A96G0000K3503B3H)

● Результат самооценки общего состояния SMART пройден

Информация об устройстве SMART

| ID | Имя | Необработанные данные | Порог | Значение | При сбое | Худшее значение |
|-----|-------------------------|-----------------------|-------|----------|----------|-----------------|
| 5 | Reallocated_Sector_Ct | 0 | 10 | 100 | | 100 |
| 9 | Power_On_Hours | 1443 | 0 | 99 | | 99 |
| 12 | Power_Cycle_Count | 16 | 0 | 99 | | 99 |
| 177 | Wear_Leveling_Count | 1 | 5 | 99 | | 99 |
| 179 | Used_Rsvd_Blk_Cnt_Tot | 0 | 10 | 100 | | 100 |
| 180 | Unused_Rsvd_Blk_Cnt_Tot | 2638 | 10 | 100 | | 100 |
| 181 | Program_Fail_Cnt_Total | 0 | 10 | 100 | | 100 |
| 182 | Erase_Fail_Count_Total | 0 | 10 | 100 | | 100 |
| 183 | Runtime_Bad_Block | 0 | 10 | 100 | | 100 |
| 184 | End-to-End_Error | 0 | 97 | 100 | | 100 |

26 всего < 1 of 3 >

Рисунок 36

Для устройств с поддержкой SMART можно получать индивидуальное состояние работоспособности и данные SMART.

6.1.3. Редактирование хоста

Для редактирования хоста необходимо выбрать хост и на панели инструментов нажать «Редактировать». В открывшемся окне обновить изменяемые свойства хоста и нажать «Редактировать Хост».

6.1.4. Режим обслуживания хоста

Плановое обслуживание кластера подразумевает установку обновлений, замену отказавших дисков (или устройств), резервное копирование и восстановление данных. Также оно включает постепенное освобождение хостов из кластера без потери данных и их удаление из инфраструктуры. Для обслуживания хост необходимо перевести в специальный режим (режим обслуживания).

Режим обслуживания хоста – это состояние, при котором хост временно исключается из активной работы кластера для выполнения технического обслуживания, обновления или устранения проблем. В этом режиме данные перераспределяются на другие хосты, что позволяет безопасно выполнять работы на хосте без ущерба для доступности данных.

Для того чтобы перейти в режим обслуживания, необходимо выбрать хост и в меню действий выбрать «Перейти в режим обслуживания».

Примечания:

1. Если хост невозможно перевести в режим обслуживания, то выводится уведомление о причине сбоя с необходимыми действиями, которое необходимо предпринять перед отключением хоста.
2. Когда хост переходит в режим обслуживания, все службы, развёрнутые на этом хосте, останавливаются, а статус хоста меняется на «На обслуживании».

После перевода хоста в режим обслуживания необходимо отключить его и выполнить требуемое техническое обслуживание. Завершив работы, необходимо включить хост и вернуть его в эксплуатацию.

6.1.5. Освобождение хоста

Освобождение хоста – процесс временного исключения хоста из активной работы кластера для выполнения технического обслуживания, обновления или устранения проблем. После инициирования освобождения хоста кластер начнёт реплицировать фрагменты данных, которые хранились на освобождаемом хосте, а затем распределять их между другими хостами в кластере. В зависимости от объёма реплицируемых данных этот процесс может занять до нескольких часов. При необходимости можно освободить хост принудительно, то есть без репликации.

Примечание. Принудительное освобождение хостов может привести к потере данных.

Для того чтобы инициировать освобождение хоста, необходимо выбрать хост и в меню действий выбрать «Начать освобождение».

6.1.6. Остановка освобождение хоста

Остановка освобождения хоста означает прекращение процесса временного исключения хоста из активной работы кластера. Остановка освобождения хоста используется, если освобождение хоста было начато по ошибке или если работы на хосте завершены раньше запланированного времени. Для того чтобы остановить освобождение хоста, необходимо выбрать хост и в меню действий выбрать «Остановить освобождение».

6.1.7. Удаление хоста

Перед удалением хоста следует эвакуировать запущенные на нем службы на другой работоспособный хост. Если на предназначенном для удаления хосте, запущены OSD, то сначала необходимо удалить OSD с этого хоста и убедиться, что на этом хосте не осталось запущенных OSD. После удаления всех OSD на хосте удалить хост из карты CRUSH.

Для удаления хоста необходимо на вкладке «Список хостов» выбрать хост и в меню действий выбрать «Удалить». Подтвердить удаление в появившемся окне.

6.1.8. Общая производительность хоста

Для просмотра общей производительности хоста необходимо в подразделе «Хосты» (раздел «Кластер») перейти на вкладку «Общая производительность» (рис. 37).

В поле «Выбор времени в Grafana» из выпадающего списка выбрать интервал времени отображения подробной информации об общей производительности выбранного хоста.

Кнопка «Сбросить настройки» позволяет сбросить настройки выбора времени в Grafana.

Кнопка «Показать скрытую информацию» позволяет просмотреть скрытую информацию.

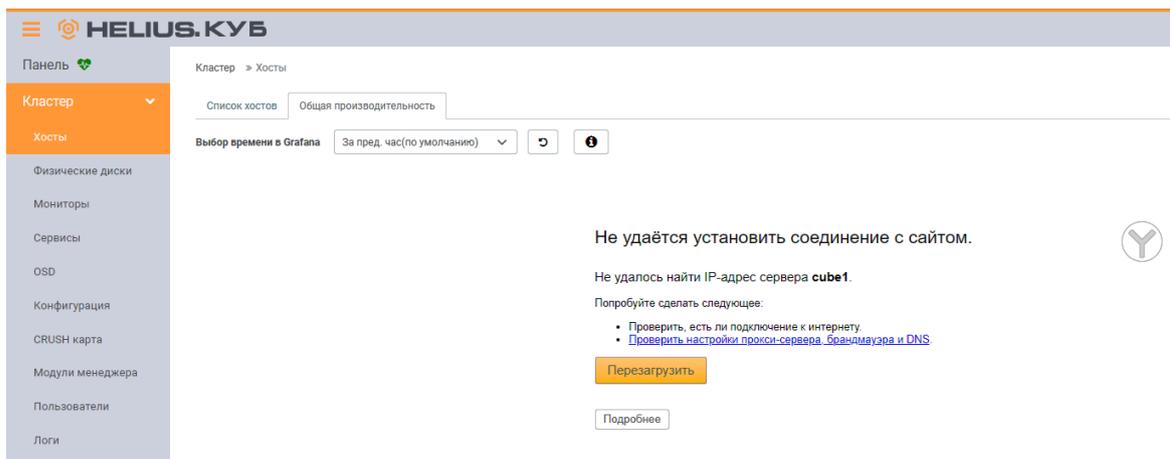
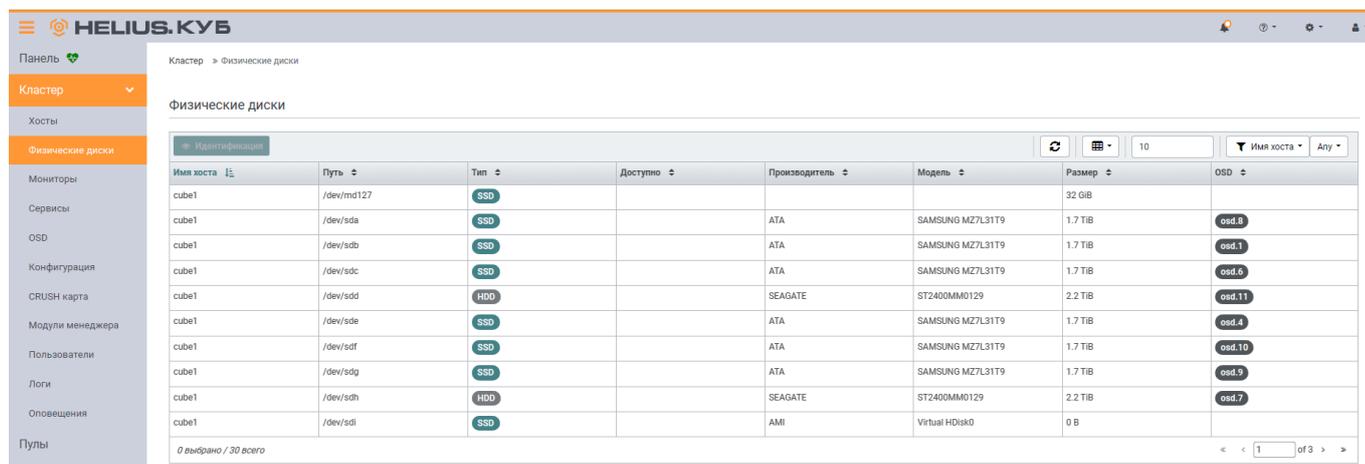


Рисунок 37

Примечание. Если ниже не появилась встроенная панель Grafana, пожалуйста, перейдите по указанной в скрытой информации ссылке, чтобы проверить доступность Grafana и отсутствие проблем с сертификатами HTTPS. Возможно, вам потребуется перезагрузить эту страницу после принятия исключений сертификатов браузера.

6.2. Физические диски

В разделе «Кластер» из выпадающего списка выбрать подраздел «Физические диски» (рис. 38).



| Имя хоста | Путь | Тип | Доступно | Производитель | Модель | Размер | OSD |
|-----------|------------|-----|----------|---------------|------------------|---------|--------|
| cube1 | /dev/md127 | SSD | | | | 32 GiB | |
| cube1 | /dev/sda | SSD | | ATA | SAMSUNG MZ7L31T9 | 1.7 TiB | osd.8 |
| cube1 | /dev/sdb | SSD | | ATA | SAMSUNG MZ7L31T9 | 1.7 TiB | osd.1 |
| cube1 | /dev/sdc | SSD | | ATA | SAMSUNG MZ7L31T9 | 1.7 TiB | osd.6 |
| cube1 | /dev/sdd | HDD | | SEAGATE | ST2400MM0129 | 2.2 TiB | osd.11 |
| cube1 | /dev/sde | SSD | | ATA | SAMSUNG MZ7L31T9 | 1.7 TiB | osd.4 |
| cube1 | /dev/sdf | SSD | | ATA | SAMSUNG MZ7L31T9 | 1.7 TiB | osd.10 |
| cube1 | /dev/sdg | SSD | | ATA | SAMSUNG MZ7L31T9 | 1.7 TiB | osd.9 |
| cube1 | /dev/sdh | HDD | | SEAGATE | ST2400MM0129 | 2.2 TiB | osd.7 |
| cube1 | /dev/sdi | SSD | | AMI | Virtual HDisk0 | 0 B | |

Рисунок 38

На данной вкладке отображена информация о подключённых физических дисках.

Для идентификации устройства необходимо выбрать устройство из списка и на панели инструментов нажать «Идентификация». В открывшемся окне «Идентификация устройства» в поле «Пожалуйста, введите продолжительность мигания LED» из выпадающего списка выбрать продолжительность мигания LED и нажать «Выполнить».

6.3. Мониторы

Монитор (MON) – демон, который обеспечивает адресацию данных внутри кластера и хранит информацию о топологии, состоянии и распределении данных.

MON отслеживают состояние всего кластера путём хранения карты состояния кластера, которая включает в себя карту MON, карту PG, карту менеджера, карту OSD, карту MDS и карту CRUSH. Все хосты кластера сообщают хостам MON информацию обо всех изменениях в своих состояниях. MON поддерживает отдельную карту информации для каждого компонента.

Карта MON содержит полную информацию о хосте MON, которая включает в себя ID кластера, имя хоста MON, а также IP-адрес MON с номером порта. Она обновляется в соответствии с текущим состоянием кластера и поддерживает историю предыдущих состояний кластера.

Карта OSD содержит информацию об ID кластера, периоде создания карты OSD и её последнем изменении, информацию об OSD (количество, состояние, вес, интервал последней очистки и информацию о хосте OSD), а также информацию, относящуюся к пулам (имена пулов, ID пулов, тип, уровень репликации и PG).

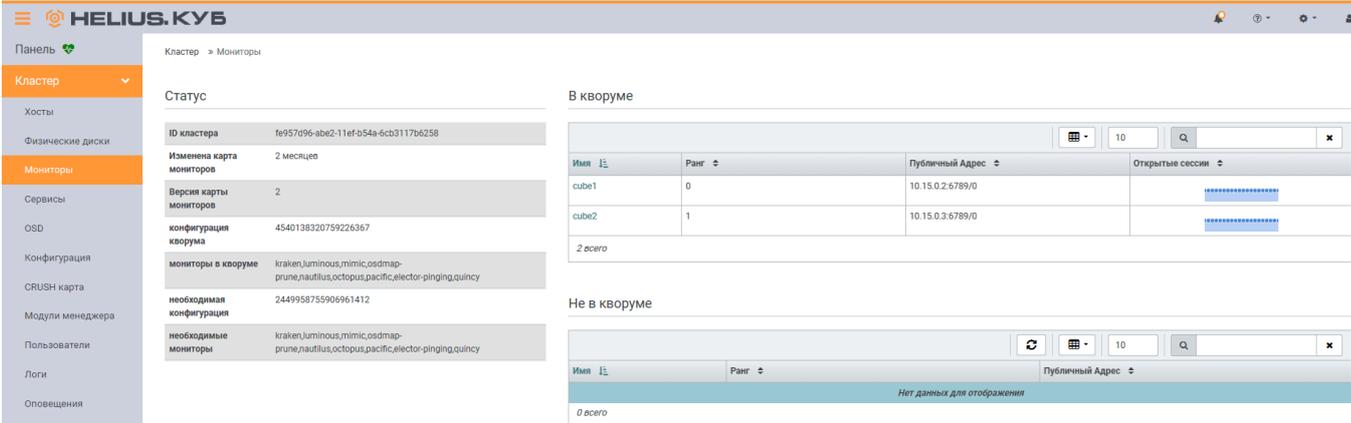
Карта PG содержит информацию о ID и состоянии каждой PG, количестве объектов, действующем наборе OSD.

Карта CRUSH содержит информацию об устройствах хранения кластера, иерархию домена отказов и определённых для домена отказов правил при хранении данных.

Карта MDS хранит информацию о текущем периоде карты MDS, времени создания и изменения карты, ID пула данных и метаданных, количестве серверов метаданных кластера и состоянии MDS.

Наличие нескольких MON в кластере обеспечивает доступность в случае сбоя одного из MON или его хоста.

В разделе «Кластер» из выпадающего списка выбрать подраздел «Мониторы» (рис. 39).



The screenshot shows the HELIUS.KYB web interface. On the left is a navigation menu with 'Мониторы' selected. The main content area is titled 'Кластер > Мониторы' and shows the following information:

- Статус:** ID кластера: fe957d96-abe2-11ef-054a-6cb3117b6258
- Изменена карта мониторов:** 2 месяцев
- Версия карты мониторов:** 2
- конфигурация кворума:** 4540138320759226367
- мониторы в кворуме:** kraken,luminous,mimic,osdmap-prune,nautilus,octopus,pacific,elector-pinging,quincy
- необходимая конфигурация:** 2449958755906961412
- необходимые мониторы:** kraken,luminous,mimic,osdmap-prune,nautilus,octopus,pacific,elector-pinging,quincy

Below this, there are two sections for monitors:

- В кворуме:** A table with columns 'Имя', 'Ранг', and 'Публичный Адрес'. It lists 'cube1' (rank 0) and 'cube2' (rank 1), both with address '10.15.0.2:6789/0'. There are '2 всего'.
- Не в кворуме:** A section showing 'Нет данных для отображения' and '0 всего'.

Рисунок 39

На данной вкладке отображена информация об общем состоянии MON и хостах MON, которые находятся в кворуме или вне его.

Кворум – минимальное количество активных MON, необходимое для поддержания работоспособности кластера. Кворум обеспечивает консенсус между MON.

Консенсус – механизм, который обеспечивает согласованность данных и принятие решений между MON кластера. Консенсус реализуется с помощью алгоритма «Paxos», который используется MON для поддержания согласованного состояния кластера, включая конфигурацию, состояние OSD и другие метаданные.

Если количество активных MON меньше кворума, то кластер переходит в состояние «не в кворуме» и становится недоступным для записи (данные доступны для чтения, но изменения невозможны). Кворум гарантирует, что кластер может продолжать работу даже при сбое некоторых MON. Кворум рассчитывается как $(n / 2) + 1$, где n – общее количество мониторов.

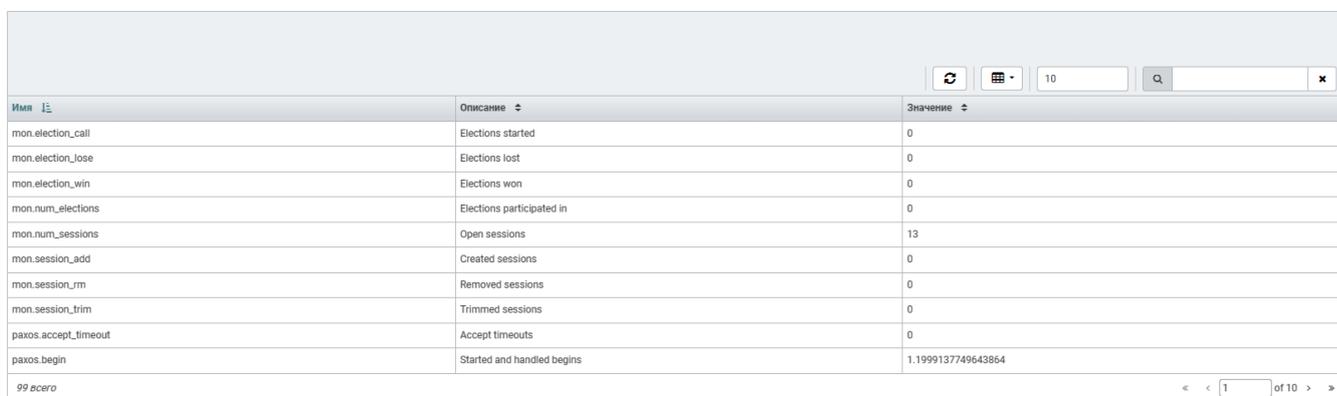
Примечание. Для обеспечения отказоустойчивости используйте нечётное количество мониторов (3, 5, 7).

MON используют ранги для выбора лидера (MON, который координирует процесс консенсуса) и принятия решений. Ранг MON – уникальный идентификатор, который присваивается каждому MON в кластере. Ранг MON назначается автоматически при добавлении MON в кластер и обычно соответствует порядку добавления MON в кластер (например, первый MON получает ранг 0, второй – ранг 1 и т.д.). При удалении MON его ранг освобождается и может быть повторно использован.

Для того чтобы увидеть количество открытых сессий необходимо навести курсор на поле «Открытые сессии».

При выборе MON будет отображена информация о его счётчиках производительности (рис. 40).

Cluster > Monitors > Performance Counters



| Имя | Описание | Значение |
|----------------------|----------------------------|--------------------|
| mon.election_call | Elections started | 0 |
| mon.election_lose | Elections lost | 0 |
| mon.election_win | Elections won | 0 |
| mon.num_elections | Elections participated in | 0 |
| mon.num_sessions | Open sessions | 13 |
| mon.session_add | Created sessions | 0 |
| mon.session_rm | Removed sessions | 0 |
| mon.session_trim | Trimmed sessions | 0 |
| paxos.accept_timeout | Accept timeouts | 0 |
| paxos.begin | Started and handled begins | 1.1999137749643864 |

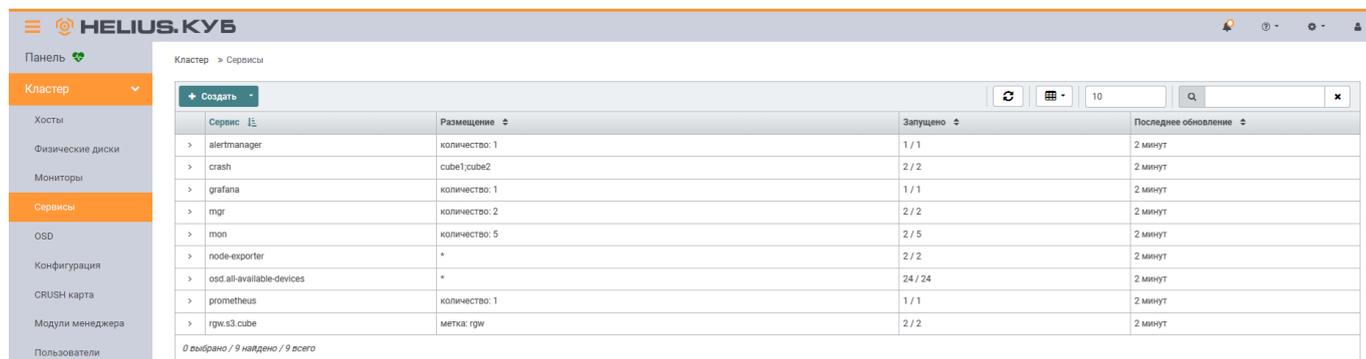
Рисунок 40

6.4. Сервисы

Сервис – группа демонов, настроенных совместно, которая выполняют различные функции в кластере. Демон – системный модуль, который запущен и является частью службы. Каждый

сервис отвечает за определённые задачи, такие как хранение данных, управление метаданными, мониторинг состояния кластера или предоставление доступа к данным через различные интерфейсы. Сервисы отвечают за работу кластера, а службы выполняют конкретные задачи внутри кластера.

В разделе «Кластер» из выпадающего списка выбрать подраздел «Сервисы» (рис. 41).



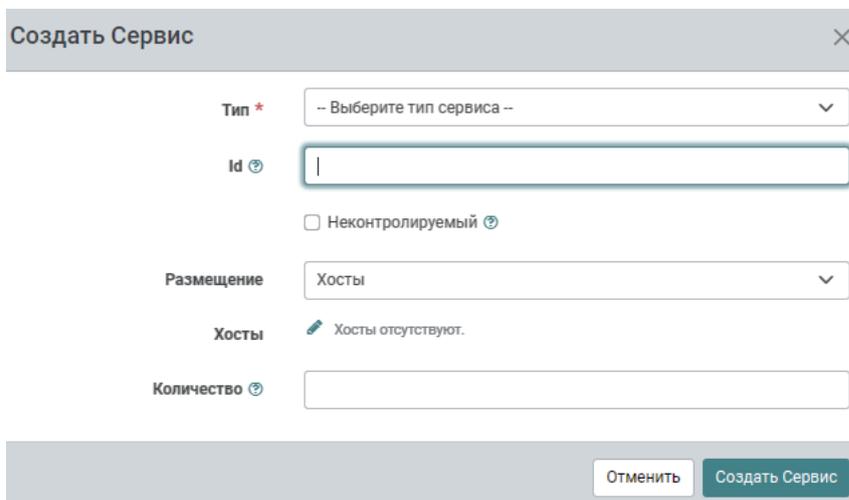
| Сервис | Размещение | Запущено | Последнее обновление |
|-----------------------------|---------------|----------|----------------------|
| > alertmanager | количество: 1 | 1 / 1 | 2 минут |
| > crash | cube1;cube2 | 2 / 2 | 2 минут |
| > grafana | количество: 1 | 1 / 1 | 2 минут |
| > mgr | количество: 2 | 2 / 2 | 2 минут |
| > mon | количество: 5 | 2 / 5 | 2 минут |
| > node-exporter | * | 2 / 2 | 2 минут |
| > osd.all-available-devices | * | 24 / 24 | 2 минут |
| > prometheus | количество: 1 | 1 / 1 | 2 минут |
| > rgw.s3.cube | метки: rgw | 2 / 2 | 2 минут |

0 выбрано / 9 найдено / 9 всего

Рисунок 41

6.4.1. Создание сервиса

В подразделе «Сервисы» нажать «Создать» на панели инструментов. В открывшемся окне заполнить предлагаемую форму (рис. 42).



Создать Сервис

Тип * -- Выберите тип сервиса --

Id

Неконтролируемый

Размещение: Хосты

Хосты: Хосты отсутствуют.

Количество

Отменить Создать Сервис

Рисунок 42

В поле «Тип» из выпадающего списка выбрать тип сервиса. Это может быть служба (mon, crash, mds, mgr, osd или rbd-mirror), шлюз (nfs или rgw), часть стека мониторинга (alertmanager, grafana, node-exporter или prometheus) или пользовательский контейнер (container).

При выборе типа сервиса «grafana» необходимо заполнить следующие дополнительные поля (рис. 43):

Grafana Порт 

Grafana пароль   

Рисунок 43

- в поле «Grafana Порт» указать стандартный порт, используемый Grafana;
- в поле «Grafana пароль» указать пароль администратора Grafana по умолчанию (устанавливается при первом запуске).

При выборе типа сервиса «ingress» необходимо заполнить следующие дополнительные поля (рис. 44):

Создать Сервис ×

Тип *

Сервис бэкенда *

Id  *

Неконтролируемый 

Размещение

Хосты  Хосты отсутствуют.

Количество 

Виртуал. IP  *

Порт внешнего интерфейса  *

Порт Монитора  *

Сети CIDR 

SSL

Рисунок 44

- в поле «Сервис бэкенда» из выпадающего списка выбрать сервис бэкенда, при этом поле «Id» заполнится по умолчанию;
- в поле «Виртуал. IP» указать виртуальный IP-адрес и подсеть (в нотации CIDR), где будет доступен сервис входа;
- в поле «Порт внешнего интерфейса» указать порт для доступа к сервису входа;
- в поле «Порт Монитора» указать порт, используемый haproxy для статуса балансировщика нагрузки;

– в поле «Сети CIDR» указать список сетей для определения, какой сетевой интерфейс использовать для виртуального IP-адреса;

– при выборе опции «SSL» в дополнительном поле «Сертификат» выбрать локальный файл для загрузки SSL-сертификата (в формате PEM), нажав «Выберите файл» (рис. 45).

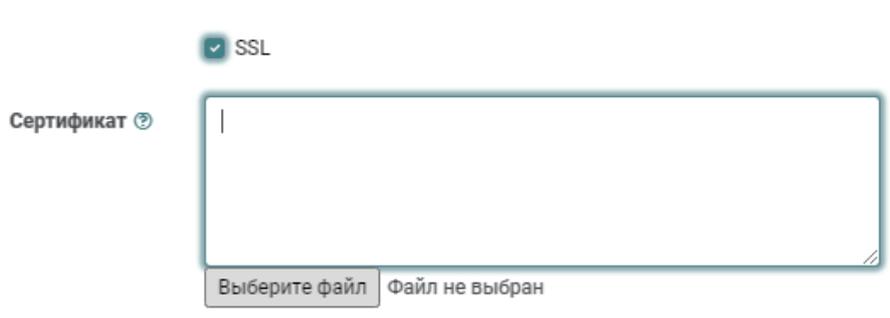


Рисунок 45

При выборе типа сервиса «iscsi» необходимо заполнить следующие дополнительные поля (рис. 46):

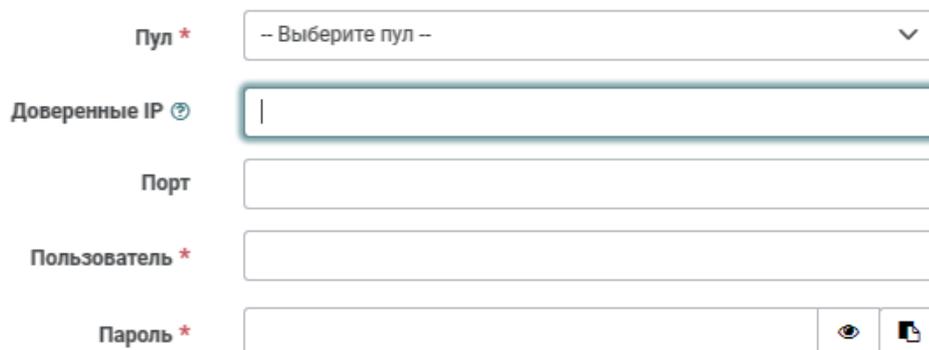


Рисунок 46

– в поле «Пул» из выпадающего списка выбрать пул;
 – в поле «Доверенные IP» указать список IP-адресов (разделённых запятыми) для доступа к iSCSI-шлюзу;

– в поле «Порт» указать используемый порт iSCSI-шлюза;

– в поле «Пользователь» указать пользователя iSCSI-шлюза;

– в поле «Пароль» указать пароль пользователя iSCSI-шлюза;

– при выборе опции «SSL» в дополнительных полях «Сертификат» и «Приватный ключ» выбрать локальные файлы для загрузки SSL-сертификата (в формате PEM) и приватного ключа SSL (в формате PEM) соответственно, нажав «Выберите файл» (рис. 47).

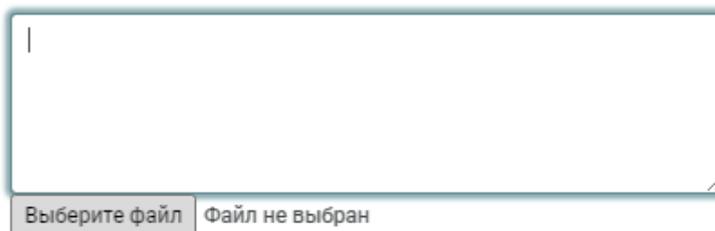
Приватный ключ 


Рисунок 47

При выборе типа сервиса «rgw» необходимо заполнить следующие дополнительные поля:

- в поле «Порт» указать используемый порт RGW;
- при выборе опции «SSL» в дополнительном поле «Сертификат» выбрать локальный файл для загрузки SSL-сертификата (в формате PEM), нажав «Выберите файл».

При выборе типа сервиса «snmp-gateway» необходимо заполнить следующие дополнительные поля (рис. 48):

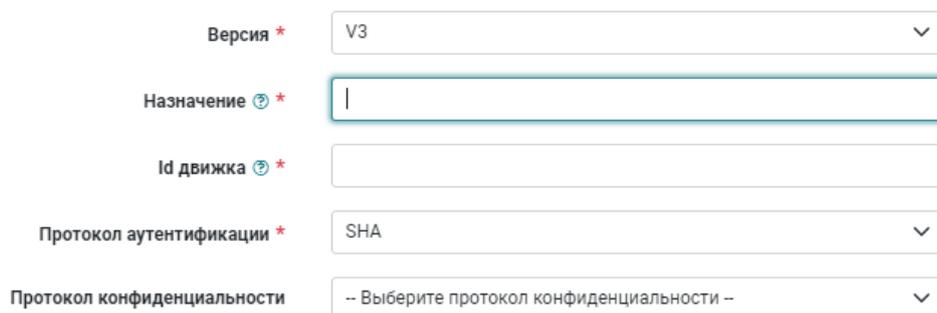


Рисунок 48

- в поле «Версия» из выпадающего списка выбрать версию SNMP («V2c» или «V3»);
- в поле «Назначение» указать назначение (в формате имя хоста:порт);
- в поле «Id движка» (поле доступно при выборе версии SNMP «V3») указать уникальный идентификатор устройства (в шестнадцатеричном формате);
- в поле «Протокол аутентификации» (поле доступно при выборе версии SNMP «V3») из выпадающего списка выбрать протокол аутентификации («SHA» или «MD5»);
- в поле «Протокол конфиденциальности» (поле доступно при выборе версии SNMP «V3») из выпадающего списка выбрать протокол конфиденциальности («DES» или «AES»);
- при выборе версии SNMP «V2c» необходимо на панели «Учётные данные» заполнить поле «SNMP сообщество» (рис. 49);

Учётные данные

SNMP сообщество *



Рисунок 49

– при выборе версии SNMP «V3» необходимо на панели «Учётные данные» заполнить дополнительные поля «Имя пользователя» и «Пароль»;

– при выборе версии SNMP «V3» и выборе протокола конфиденциальности (поле «Протокол конфиденциальности») необходимо на панели «Учётные данные» заполнить дополнительное поле «Шифрование» (рис. 50).

Учётные данные

Имя пользователя *

Пароль *

Шифрование *

Рисунок 50

В поле «Id» указать идентификатор сервиса (он используется в названии сервиса, которое имеет формат `service_type.service_id`).

При необходимости выбрать опцию «Неконтролируемый». При выборе данной опции оркестратор не будет запускать или останавливать службу, связанную с этим сервисом. Расположение и все другие свойства будут игнорироваться.

В поле «Размещение» из выпадающего списка выбрать тип размещения (хосты или метка):

– при выборе «Хосты» в дополнительном поле «Хосты» выбрать хост;

– при выборе «Метка» в дополнительном поле «Метка» указать метку.

В поле «Количество» указать количество служб, которые нужно развернуть.

Нажать «Создать Сервис».

6.4.2. Просмотр информации о сервисе

Для просмотра подробной информации необходимо развернуть содержимое строки. Информация будет представлена на вкладках «Службы» и «События сервиса».

Вкладка «Службы» отображает развёрнутые службы и информацию о них (рис. 51).

| Имя хоста | Имя службы | Версия | Статус | Последнее обновление | Использование CPU | Использование памяти | События службы |
|-----------|--------------------|--------|---------|----------------------|-------------------|----------------------|-------------------|
| cube1 | alertmanager.cube1 | 0.25.0 | running | 3 минут | 5% | 31.1 MiB | No data available |

0 выбрано / 1 всего

Рисунок 51

Для того чтобы запустить службу необходимо выбрать службу и на панели инструментов нажать «Начать», при этом статус службы изменится на «running».

Для того чтобы остановить службу необходимо выбрать службу и в меню действий выбрать «Остановить», при этом статус службы изменится на «stop».

Перезапуск службы может потребоваться для применения изменений конфигурации, устранения сбоев или выполнения технического обслуживания. Для того чтобы перезапустить службу, необходимо выбрать службу и в меню действий выбрать «Запустить заново», при этом статус службы изменится на «gunning».

Развёртывание службы заново – процесс повторного создания и настройки службы после её удаления, сбоя или изменения конфигурации. Для того чтобы повторно развернуть службу, необходимо выбрать службу и в меню действий выбрать «Развернуть заново», при этом статус службы изменится на «gunning».

Вкладка «События сервиса» отображает информацию о событиях сервиса (рис. 52).



| Имя сервиса | Тип сервиса | События сервиса |
|--------------|--------------|-------------------|
| alertmanager | alertmanager | No data available |
| Итого | | |

Рисунок 52

6.4.3. Редактирование сервиса

Для редактирования сервиса необходимо выбрать сервис и на панели инструментов нажать «Редактировать». В открывшемся окне обновить изменяемые свойства сервиса и нажать «Редактировать Сервис».

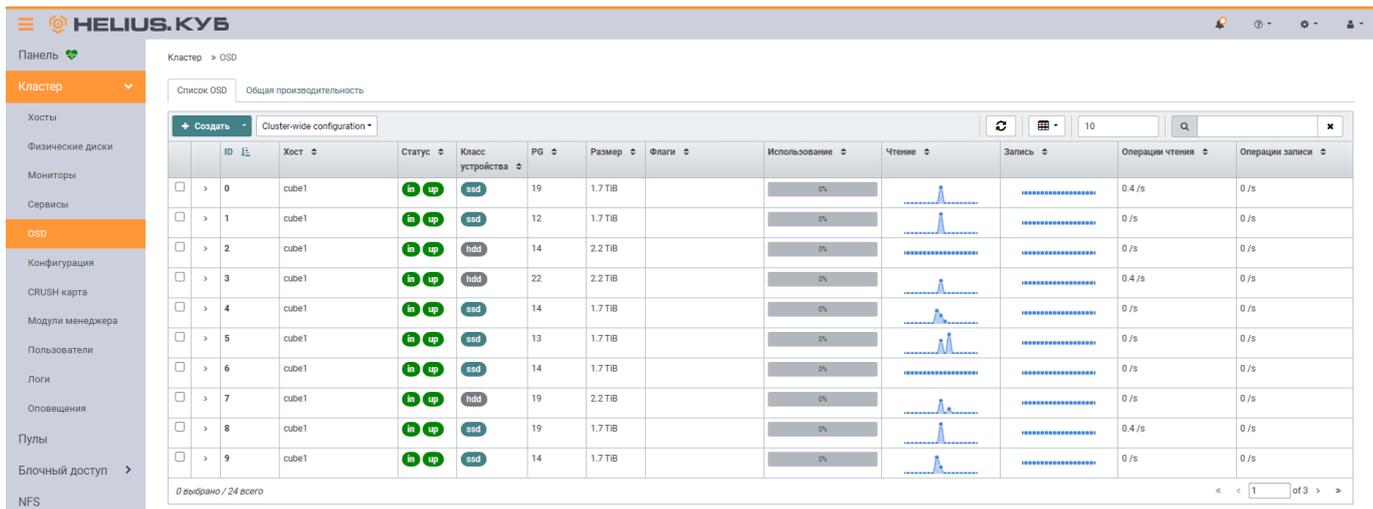
6.4.4. Удаление сервиса

Для удаления сервиса необходимо выбрать сервис и в меню действий выбрать «Удалить». Подтвердить удаление в появившемся окне.

6.5. OSD

OSD – устройство хранения объектов, которое хранит данные, обрабатывает запросы клиентов и предоставляет информацию для мониторинга ПО. OSD отвечает за управление отдельной единицей хранения, которая представляет собой отдельный физический или логический диск.

В разделе «Кластер» из выпадающего списка выбрать подраздел «OSD». Данный подраздел состоит из вкладок «Список OSD» и «Общая производительность» (рис. 53).



| ID | Хост | Статус | Класс устройства | PG | Размер | Использование | Чтение | Запись | Операции чтения | Операции записи |
|----|-------|--------|------------------|----|---------|---------------|--------|--------|-----------------|-----------------|
| 0 | cube1 | in up | ssd | 19 | 1.7 TiB | 0% | 0.4 /s | 0 /s | 0.4 /s | 0 /s |
| 1 | cube1 | in up | ssd | 12 | 1.7 TiB | 0% | 0 /s | 0 /s | 0 /s | 0 /s |
| 2 | cube1 | in up | hdd | 14 | 2.2 TiB | 0% | 0 /s | 0 /s | 0 /s | 0 /s |
| 3 | cube1 | in up | hdd | 22 | 2.2 TiB | 0% | 0.4 /s | 0 /s | 0.4 /s | 0 /s |
| 4 | cube1 | in up | ssd | 14 | 1.7 TiB | 0% | 0 /s | 0 /s | 0 /s | 0 /s |
| 5 | cube1 | in up | ssd | 13 | 1.7 TiB | 0% | 0 /s | 0 /s | 0 /s | 0 /s |
| 6 | cube1 | in up | ssd | 14 | 1.7 TiB | 0% | 0 /s | 0 /s | 0 /s | 0 /s |
| 7 | cube1 | in up | hdd | 19 | 2.2 TiB | 0% | 0 /s | 0 /s | 0 /s | 0 /s |
| 8 | cube1 | in up | ssd | 19 | 1.7 TiB | 0% | 0.4 /s | 0 /s | 0.4 /s | 0 /s |
| 9 | cube1 | in up | ssd | 14 | 1.7 TiB | 0% | 0 /s | 0 /s | 0 /s | 0 /s |

Рисунок 53

Статус OSD разделён на две части: в первой части отображено состояние процессов OSD («up» и «down»), а во второй – состояние данных на OSD («in» и «out»):

- OSD запускается и переходит в состояние «up», при выключении OSD переходит в состояние «down»;

- при включении OSD входит в состояние «in», а при выключении и долгом отсутствии – ВЫХОДИТ В «out».

Примечание. По умолчанию таймаут – 20 мин. В течении этого времени кластер считает, что OSD находится в состоянии «in». По истечении этого времени кластер автоматически переведёт данные OSD в состояние «out».

Важно понимать разницу между состояниями «up/down» и «in/out»:

- если OSD в состоянии «up» – это не означает, что кластер использует данные OSD;
- если OSD в состоянии «down» – кластер всё ещё может использовать данные OSD;
- если OSD в состоянии «in» – на OSD реплицируются данные, т.е. OSD включён в процесс обмена данными;

- если OSD в состоянии «out» – данные выводятся с OSD.

6.5.1. Создание OSD

Если на хосте кластера есть устройства хранения данных (HDD, SSD или NVME), статус которых «available», то можно создать OSD. На вкладке «Список OSD» нажать «Создать» на панели инструментов. В открывшемся окне заполнить предлагаемую форму (рис. 54).

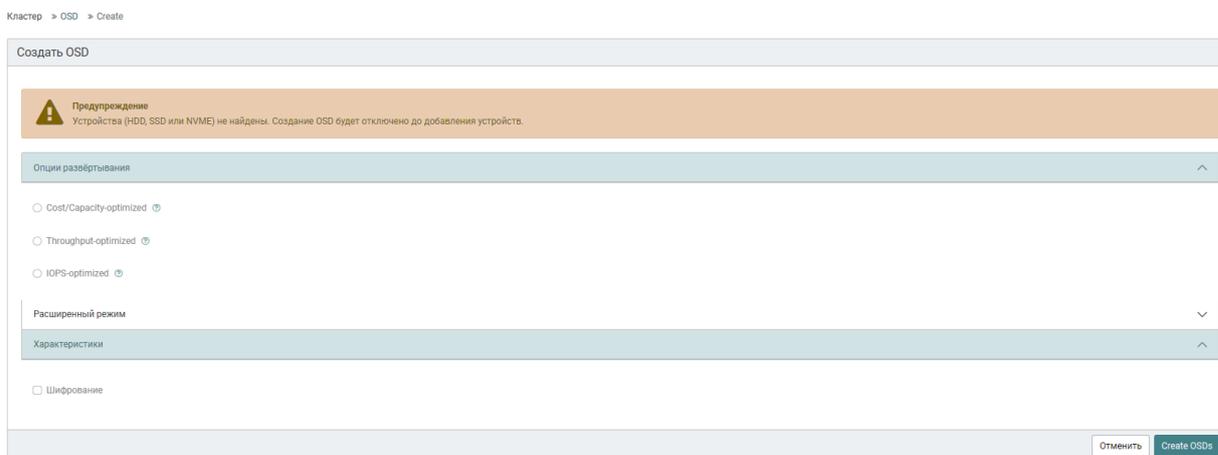


Рисунок 54

Примечание. Если устройства (HDD, SSD или NVME) не найдены, то создание OSD будет отключено до добавления устройств.

На панели «Опции развёртывания» выбрать необходимые опции развёртывания:

- Cost/Capacity-optimized – выбраны все доступные HDD-диски;
- Throughput-optimized – HDD-диски или SSD-накопители выбираются для устройств передачи данных, а SSD-накопители или NVMe-накопители – для устройств DB/WAL;
- IOPS-optimized – выбраны все доступные NVMe-накопители.

Для добавления устройств необходимо выбрать «Расширенный режим», а затем добавить необходимые устройства, нажав «Добавить» (рис. 55).

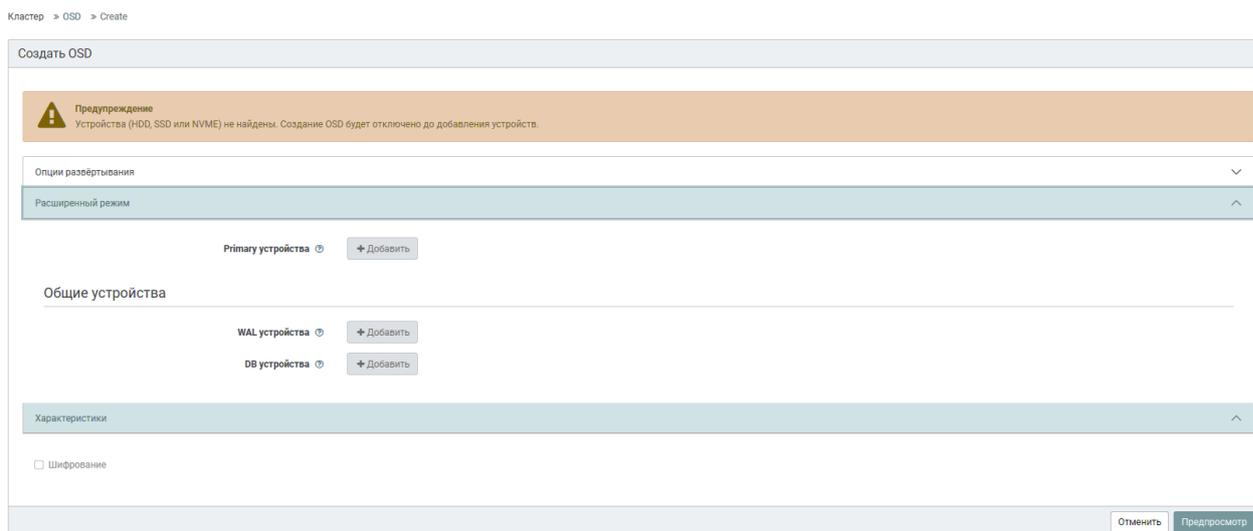


Рисунок 55

«Primary устройства» (основные устройства хранения) – устройства (обычно HDD, SSD или NVMe), которые используются для хранения основных данных в кластере. Эти устройства являются основным местом хранения объектов и их реплик, в отличие от DB- и WAL-устройств, которые используются для ускорения операций чтения и записи.

«WAL устройства» – высокопроизводительные устройства (обычно SSD или NVMe), которые используются для хранения журналов операций записи (WAL). WAL-устройства помогают ускорить операции записи и повысить производительность кластера, особенно в пулах кодирования с восстановлением данных, где данные разделяются на фрагменты и добавляются контрольные суммы.

«DB устройства» – высокопроизводительные устройства (обычно SSD или NVMe), которые используются для ускорения операций чтения и записи в пулах кодирования с восстановлением данных. DB-устройства хранят метаданные и журналы (WAL), что позволяет значительно повысить производительность пулов и эффективность использования ресурсов.

Нажать «Предпросмотр», чтобы предварительно посмотреть спецификации групп дисков для ранее добавленных устройств. Нажать «Создать», при этом новые устройства будут добавлены в список OSD.

Примечание. Для создания OSD требуется некоторое время, т.к. OSD появятся в списке после их развёртывания. Статус развёртывания можно посмотреть в логах кластера (см. подробнее 6.10.1).

При необходимости на панели «Характеристики» выбрать опцию «Шифрование» (выбор опции доступен при выходе из расширенного режима (выбор «Опции развёртывания»)). Данная опция позволяет включить шифрование данных, хранящихся на устройствах.

Примечание. Включённое шифрование снижает производительность.

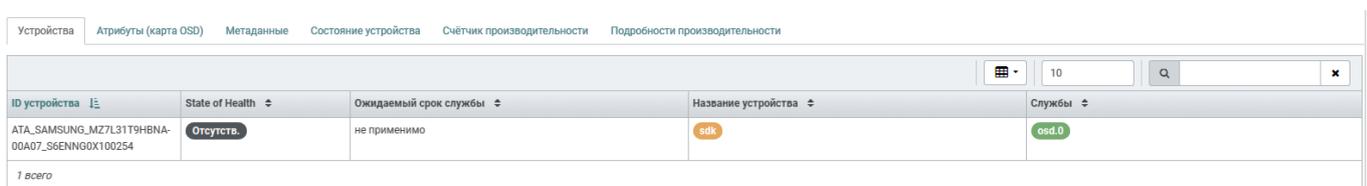
Нажать «Создать OSD».

При создании OSD автоматически включается в CRUSH-иерархию (см. подробнее 6.7)

6.5.2. Просмотр информации об OSD

Для просмотра подробной информации необходимо развернуть содержимое строки. Информация будет представлена на вкладках «Устройства», «Атрибуты (карта OSD)», «Метаданные», «Состояние устройства», «Счётчик производительности» и «Подробности производительности».

Вкладка «Устройства» отображает подключённые устройства и информацию о них (рис. 56).



| ID устройства | State of Health | Ожидаемый срок службы | Название устройства | Службы |
|--|-----------------|-----------------------|---------------------|--------|
| ATA_SAMSUNG_MZ7L31T9HBN-00A07_S6ENNG0X100254 | Оругув. | не применимо | sdk | osd.0 |

Рисунок 56

Вкладка «Атрибуты (карта OSD)» отображает адрес кластера, сведения о тактовом сигнале, состоянии OSD и другие атрибуты (рис. 57).

| Устройство | Атрибуты (карта OSD) | Метаданные | Состояние устройства | Счётчик производительности | Подробности производительности |
|----------------------|---------------------------|------------|----------------------|----------------------------|--------------------------------|
| cluster_addr | 10.15.0.2:6875/2861980667 | | | | |
| down_at | 559 | | | | |
| heartbeat_back_addr | 10.15.0.2:6881/2861980667 | | | | |
| heartbeat_front_addr | 10.15.0.2:6878/2861980667 | | | | |
| id | 0 | | | | |
| in | 1 | | | | |
| last_clean_begin | 399 | | | | |
| last_clean_end | 558 | | | | |
| lost_at | 0 | | | | |
| osd | 0 | | | | |

Рисунок 57

Вкладка «Метаданные» отображает метаданные OSD (рис. 58).

| Устройство | Атрибуты (карта OSD) | Метаданные | Состояние устройства | Счётчик производительности | Подробности производительности |
|--------------------------------|---|------------|----------------------|----------------------------|--------------------------------|
| arch | x86_64 | | | | |
| back_addr | [v2:10.15.0.2:6874/2861980667,v1:10.15.0.2:6875/2861980667] | | | | |
| back_iface | | | | | |
| bluefs | 1 | | | | |
| bluefs_dedicated_db | 0 | | | | |
| bluefs_dedicated_wal | 0 | | | | |
| bluefs_single_shared_device | 1 | | | | |
| bluestore_allocation_from_file | 1 | | | | |
| bluestore_bdev_access_mode | blk | | | | |
| bluestore_bdev_block_size | 4096 | | | | |

Рисунок 58

Вкладка «Состояние устройства» отображает состояние подключённых устройств (устройства отображены на вкладке «Устройства») и состоит из вкладок «Информация об устройстве» (рис. 59) и «SMART» (рис. 60).

Устройства Атрибуты (карта OSD) Метаданные **Состояние устройства** Счётчик производительности Подробности производительности

/dev/sdk (S6ENNG0X100254)

🔍 Результат самооценки общего состояния SMART пройдено

Информация об устройстве SMART

```
ata_device_statistics pages
{"name":"General Statistics","number":1,"revision":1,"table":{"flags":{"monitored_condition_met":false,"normalized":false,"string":"V--";"supports_dsn":false,"valid":true,"value":192},"name":"Lifetime Power-On Resets","offset":8,"size":4,"value":17},"flags":{"monitored_condition_met":false,"normalized":false,"string":"V--";"supports_dsn":false,"valid":true,"value":192},"name":"Power-on Hours","offset":16,"size":4,"value":1445},"flags":{"monitored_condition_met":false,"normalized":false,"string":"V--";"supports_dsn":false,"valid":true,"value":192},"name":"Logical Sectors Written","offset":24,"size":6,"value":5180673},"flags":{"monitored_condition_met":false,"normalized":false,"string":"V--";"supports_dsn":false,"valid":true,"value":192},"name":"Number of Write Commands","offset":32,"size":6,"value":288108},"flags":{"monitored_condition_met":false,"normalized":false,"string":"V--";"supports_dsn":false,"valid":true,"value":192},"name":"Logical Sectors Read","offset":40,"size":6,"value":47602616},"flags":{"monitored_condition_met":false,"normalized":false,"string":"V--";"supports_dsn":false,"valid":true,"value":192},"name":"Number of Read Commands","offset":48,"size":6,"value":232536},"flags":{"monitored_condition_met":false,"normalized":false,"string":"V--";"supports_dsn":false,"valid":true,"value":192},"name":"Date and Time TimeStamp","offset":56,"size":6,"value":292000}}}, {"name":"General Errors Statistics","number":4,"revision":1,"table":{"flags":{"monitored_condition_met":false,"normalized":false,"string":"V--";"supports_dsn":false,"valid":true,"value":192},"name":"Number of Reported Uncorrectable Errors","offset":8,"size":4,"value":0},"flags":{"monitored_condition_met":false,"normalized":false,"string":"V--";"supports_dsn":false,"valid":true,"value":192},"name":"Resets Between Cmd Acceptance and Completion","offset":16,"size":4,"value":0}}}, {"name":"Temperature Statistics","number":5,"revision":1,"table":{"flags":
```

Рисунок 59

Устройства Атрибуты (карта OSD) Метаданные **Состояние устройства** Счётчик производительности Подробности производительности

/dev/sdk (S6ENNG0X100254)

Результат самооценки общего состояния SMART пройдено

Информация об устройстве SMART

| ID | Имя | Необработанные данные | Порог | Значение | При сбое | Худшее значение |
|-----|-------------------------|-----------------------|-------|----------|----------|-----------------|
| 5 | Reallocated_Sector_Ct | 0 | 10 | 100 | | 100 |
| 9 | Power_On_Hours | 1445 | 0 | 99 | | 99 |
| 12 | Power_Cycle_Count | 17 | 0 | 99 | | 99 |
| 177 | Wear_Leveling_Count | 1 | 5 | 99 | | 99 |
| 179 | Used_Rsvd_Blk_Cnt_Tot | 0 | 10 | 100 | | 100 |
| 180 | Unused_Rsvd_Blk_Cnt_Tot | 2617 | 10 | 100 | | 100 |
| 181 | Program_Fail_Cnt_Total | 0 | 10 | 100 | | 100 |
| 182 | Erase_Fail_Count_Total | 0 | 10 | 100 | | 100 |
| 183 | Runtime_Bad_Block | 0 | 10 | 100 | | 100 |
| 184 | End-to-End_Error | 0 | 97 | 100 | | 100 |

26 всего

Рисунок 60

Вкладка «Счётчик производительности» отображает имя, описание и значение счётчиков производительности (рис. 61).

Устройства Атрибуты (карта OSD) Метаданные Состояние устройства **Счётчик производительности** Подробности производительности

| Имя | Описание | Значение |
|---------------------------------|---|----------|
| bluefs.alloc_slow_fallback | Amount of allocations that required fallback to slow/shared device | 0 |
| bluefs.alloc_slow_size_fallback | Amount of allocations that required fallback to shared device's regular unit size | 0 |
| bluefs.alloc_unit_db | Allocation unit size (in bytes) for standalone DB device | 0 |
| bluefs.alloc_unit_main | Allocation unit size (in bytes) for primary/shared device | 0 |
| bluefs.alloc_unit_wal | Allocation unit size (in bytes) for standalone WAL device | 0 |
| bluefs.bytes_written_slow | Bytes written to WAL/SSTs at slow device | 0 |
| bluefs.bytes_written_sst | Bytes written to SSTs | 0 |
| bluefs.bytes_written_wal | Bytes written to WAL | 0 |
| bluefs.compact_lat | Average bluefs log compaction latency | 0 |
| bluefs.compact_lock_lat | Average lock duration while compacting bluefs log | 0 |

183 всего

Рисунок 61

Вкладка «Подробности производительности» показывает подробную информацию о производительности выбранного OSD (рис. 62).

В поле «Выбор времени в Grafana» из выпадающего списка выбрать интервал времени отображения подробной информации о производительности OSD.

Кнопка «Сбросить настройки» позволяет сбросить настройки выбора времени в Grafana.

Кнопка «Показать скрытую информацию» позволяет просмотреть скрытую информацию.

[Устройства](#)
[Атрибуты \(карта OSD\)](#)
[Метаданные](#)
[Состояние устройства](#)
[Счётчик производительности](#)
[Подробности производительности](#)

Выбор времени в Grafana:

Не удаётся установить соединение с сайтом.



Не удалось найти IP-адрес сервера **cube1**.

Попробуйте сделать следующее:

- Проверить, есть ли подключение к интернету.
- [Проверить настройки прокси-сервера, брандмауэра и DNS.](#)

Рисунок 62

Примечание. Если ниже не появилась встроенная панель Grafana, пожалуйста, перейдите по указанной в скрытой информации ссылке, чтобы проверить доступность Grafana и отсутствие проблем с сертификатами HTTPS. Возможно, вам потребуется перезагрузить эту страницу после принятия исключений сертификатов браузера.

6.5.3. Редактирование OSD

Для редактирования OSD необходимо выбрать OSD и на панели инструментов нажать «Редактировать». В открывшемся окне в поле «Класс устройств» указать новый класс устройства и нажать «Редактировать OSD» (рис. 63).

Редактировать OSD: 0
✕

Класс устройства *

Рисунок 63

6.5.4. Установка индивидуальных флагов OSD

Для установки индивидуальных флагов OSD необходимо выбрать OSD и в меню действий выбрать «Флаги». В открывшемся окне «Индивидуальные флаги OSD» выбрать индивидуальные флаги и нажать «Обновить» (рис. 64).

Индивидуальные флаги OSD
✕

Запрет на запуск
Запуск OSD не разрешён

Запрет на выключение
Отчёты о сбоях OSD игнорируются, мониторы не помечают OSD как выключенные

Запрет на включение
Ранее исключённые OSD не будут включены обратно при запуске

Запрет на исключение
OSD не будут автоматически исключены после настроенного интервала

Восстановить предыдущий выбор
Отменить
Обновить

Рисунок 64

Кнопка «Восстановить предыдущий выбор» позволяет восстановить предыдущий выбор индивидуальных флагов.

Ниже приведено описание некоторых индивидуальных флагов:

- запрет на запуск – запуск OSD не разрешён;
- запрет на выключение – отчёты о сбоях OSD игнорируются, мониторы не помечают OSD как выключенные;
- запрет на включение – ранее исключённые OSD не будут включены обратно при запуске;
- запрет на исключение – OSD не будут автоматически исключены после настроенного интервала.

Примечание. Флаги «Запрет на запуск», «Запрет на исключение» и «Запрет на выключение» являются временными – как только флаги будут сброшены, действие, которое они блокировали, произойдёт. Флаг «Запрет на включение» не позволяет помечать OSD как «in» при загрузке, поэтому любые службы, запущенные во время установки флага, останутся в том же состоянии.

6.5.5. Замена неисправных OSD

При выходе OSD из строя его можно заменить и повторно использовать тот же идентификатор OSD, чтобы избежать необходимости перенастройки карты CRUSH.

Замена OSD идентична удалению OSD, за исключением того, что OSD не удаляется из иерархии CRUSH навсегда, а вместо этого получает статус «destroyed». Этот статус используется для определения идентификаторов OSD, которые могут быть повторно использованы при следующем развёртывании OSD.

Ниже приведены возможные причины замены OSD:

- OSD вышло из строя или скоро выйдет из строя (согласно информации SMART) и больше не может использоваться для безопасного хранения данных;

- необходимо обновить OSD, например, увеличить его размер;
- необходимо изменить расположение OSD на устройстве.

Неисправные OSD можно определить следующими способами:

- группа виджетов «Статус» на панели управления, отображающая статус кластера «HEALTH_WARN» и статус OSD;
- статус OSD на вкладке «Список OSD» (подраздел «OSD» раздела «Кластер»);
- всплывающие уведомления;
- индикация (обычно «красным») светодиодов на физическом диске.

Примечание. Цвет светодиода состояния диска может отличаться (зависит от конкретной модели диска и его производителя).

Для замены неисправных OSD следует выполнить следующие действия:

- на вкладке «Список OSD» (подраздел «OSD» раздела «Кластер») выбрать неисправный OSD со статусом «out»+«down»;
- в меню действий выбрать «Флаги»;
- в открывшемся окне «Индивидуальные флаги OSD» выбрать флаг «Запрет на включение» и нажать «Обновить»;
- затем в меню действий выбрать «Удалить». В появившемся окне выбрать «Сохранить ID OSD для замены», а затем подтвердить удаление, нажав «Да, я уверен»;
- дождаться смены статуса OSD на «out»+«destroyed»;
- при необходимости установить флаг «Запрет на включение» для всего кластера (подробнее см. 6.5.15.);
- если OSD вышло из строя из-за сбоя физического диска, то необходимо заменить физический диск. Если физический диск поддерживает «горячую» замену, то заменить неисправный диск на новый. Если физический диск не поддерживает «горячую» замену, а хост содержит несколько OSD, то необходимо выключить весь хост и заменить физический диск;
- создать новое OSD (см. подробнее 6.5.1), затем включить его (установить статус «in») (см. подробнее 6.5.9);
- в меню действий выбрать «Флаги» и в открывшемся окне снять выбор опций «Запрет на включение» для OSD;
- если ранее был установлен флаг «Запрет на включение» для всего кластера, то снять выбор данного флага (подробнее см. 6.5.15);
- убедиться, что OSD создано и ID OSD сохранен.

Примечания:

1. В случае замены OSD после сбоя рекомендуется выполнить глубокую чистку (проверку) PG.

2. В случае, если общее устройство для DB/WAL выйдет из строя, потребуется выполнить процедуру замены для всех OSD, которые используют это устройство.

6.5.6. Проверка и восстановление данных (чистка OSD)

Процесс проверки и восстановления данных (за счёт чистки PG) выполняется для обеспечения целостности и согласованности данных в кластере. Данный процесс помогает обнаружить и исправить ошибки в данных, такие как повреждённые объекты или несоответствия между репликами.

Для каждой PG создаётся каталог всех объектов и сравнивается каждый основной объект и его реплики (копии), чтобы убедиться, что ни один объект не пропал и реплика соответствует основному объекту.

Процесс проверки данных включает:

- лёгкую проверку метаданных и контрольных сумм;
- глубокую проверку, которая сравнивает данные и контрольные суммы всех реплик.

6.5.6.1. Лёгкая проверка

Лёгкая проверка (ежедневная) проверяет размер и атрибуты объекта, чтобы убедиться, что PG хранят одни и те же данные объектов.

Для лёгкой проверки необходимо выбрать OSD и в меню действий выбрать «Чистка». В открывшемся окне подтвердить чистку OSD, нажав «Обновить».

6.5.6.2. Глубокая проверка

Глубокая проверка (еженедельная) считывает данные и использует контрольные суммы для обеспечения целостности данных. Глубокая проверка проверяет содержимое объекта и его реплик, чтобы убедиться, что фактическое содержимое совпадает.

Примечание . Во время данной процедуры происходит повышенная нагрузка на кластер.

Для глубокой проверки необходимо выбрать OSD и в меню действий выбрать «Глубокая чистка». В открывшемся окне подтвердить глубокую чистку OSD, нажав «Обновить».

6.5.7. Изменение веса OSD

Вес – параметр, который определяет, какая доля данных должна храниться на конкретном OSD. Чем выше вес, тем больше данных будет храниться на OSD. Вес OSD влияет на распределение данных в кластере и используется алгоритмом CRUSH для определения, где именно будут храниться объекты и их реплики.

Вес OSD обычно находится в диапазоне от 0 до 1 (рассчитывается автоматически), но может быть больше для OSD с большим объёмом дискового пространства.

Для того чтобы вручную изменить вес OSD, необходимо выбрать OSD и в меню действий выбрать «Перевзвесить». В открывшемся окне в поле «Weight» указать новый вес OSD и нажать «Перевзвесить» (рис. 65).



Рисунок 65

6.5.8. Исключение OSD из кластера

Иногда необходимо временно исключить OSD из кластера (при сбое или техническом обслуживании), чтобы OSD не участвовало в распределении данных. Для временного исключения OSD из кластера необходимо в меню действий выбрать «Отмечено как исключено». При этом статус OSD изменить на «out».

6.5.9. Включение OSD

Иногда необходимо включить OSD, который был возвращён в кластер после временного исключения (например, после технического обслуживания). Включение OSD означает, что OSD снова участвует в распределении данных и обработке запросов. Для включения OSD необходимо в меню действий выбрать «Отмечено как включено». При этом статус OSD изменить на «in».

6.5.10. Неработоспособный OSD

Если OSD недоступен или не отвечает на запросы (сбой устройства, проблемы с сетью или ошибки в конфигурации), то его следует пометить OSD как неработоспособный. Такой OSD не может обрабатывать запросы и не участвует в распределении данных. Для того чтобы отметить OSD как неработоспособный, необходимо выбрать OSD в меню действий выбрать «Отмечено как нерабочий». При этом статус OSD изменить на «down».

6.5.11. Недоступный OSD

Если OSD в кластере недоступен в течении длительного времени и не отвечает на запросы, то его следует пометить как «потерянный». Данные, которые хранились на этом OSD, должны быть восстановлены на других OSD. Для того чтобы отметить OSD как «потерянный», необходимо в меню действий выбрать «Отмечено как потеряно». При этом статус OSD изменить на «down».

6.5.12. Очистка OSD

Для того чтобы очистить OSD, необходимо выбрать OSD и в меню действий выбрать «Очистить».

6.5.13. Уничтожение OSD

Уничтожение OSD – процесс полного удаления OSD из кластера, включая его конфигурацию и данные. Уничтожение OSD используется в случае, когда OSD вышел из строя или требуется заменить устройство.

Прежде чем уничтожить OSD, его необходимо остановить. Для уничтожения OSD необходимо выбрать OSD и в меню действий выбрать «Уничтожить».

При этом статус OSD изменить на «destroyed». После этого необходимо удалить OSD из CRUSH карты, чтобы оно больше не участвовало в распределении данных.

6.5.14. Удаление OSD

Под удалением OSD подразумевается вывод OSD из эксплуатации и удаление из кластера.

Прежде чем удалять OSD из кластера, убедитесь, что в кластере достаточно свободного места на диске, чтобы удалить OSD. Удаление OSD приводит к ребалансировке всего кластера.

Для удаления OSD необходимо выбрать OSD, не содержащий PG, и в меню действий выбрать «Удалить». Подтвердить удаление в появившемся окне.

Можно остановить удаление только тех OSD, которые поставлены в очередь на удаление. При этом произойдёт сброс начального состояния OSD и удаление OSD из очереди на удаление. Если OSD находится в процессе удаления, то остановить процесс невозможно.

6.5.15. Флаги OSD для всего кластера

Для того чтобы установить флаги OSD для всего кластера, необходимо выбрать «Cluster-wide configuration», затем в меню действий выбрать «Флаги» (рис. 66).

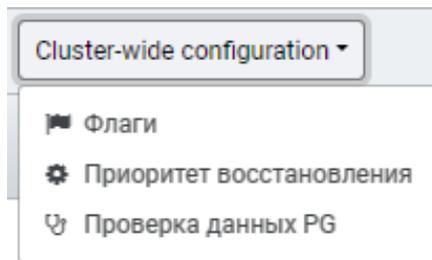


Рисунок 66

В открывшемся окне выбрать необходимые флаги (рис. 67).

Ниже приведено описание флагов:

- запрет на включение – ранее исключённые OSD не будут включены обратно при запуске;
- запрет на исключение – OSD не будут автоматически исключены после настроенного интервала;
- запрет на запуск – запуск OSD не разрешён;
- запрет на выключение – отчёты о сбоях OSD игнорируются, MON не помечают OSD как выключенные («down»);

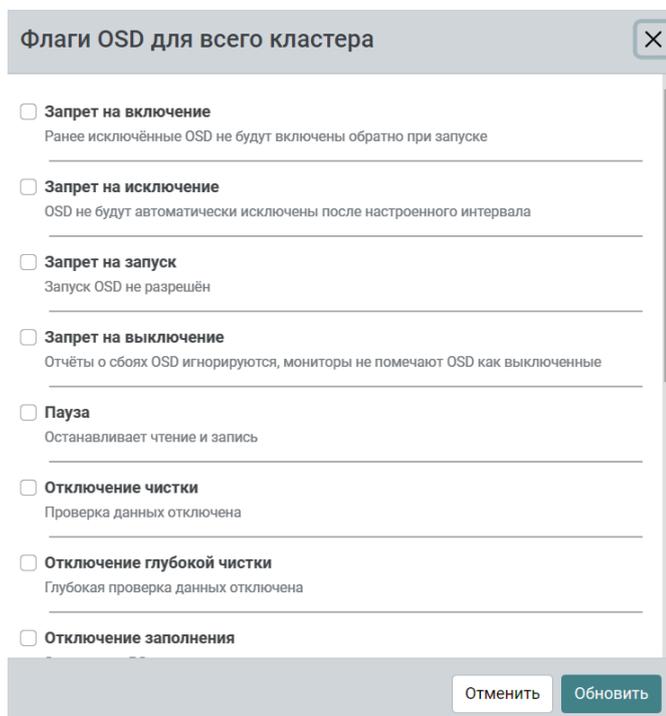


Рисунок 67

- пауза – останавливает чтение и запись;
- отключение чистки – проверка данных отключена;
- отключение глубокой чистки – глубокая проверка данных отключена;
- отключение заполнения – заполнение PG приостановлено;
- отключение перебалансировки – OSD не будет проводить восстановление данных, если только PG также не находится в неисправном состоянии;
- отключение восстановления – восстановление PG приостановлено;
- побитовая сортировка – использовать побитовую сортировку;
- очищенные директории снапшотов – у OSD конверт. наборы снапшотов;
- удаления при восстановлении – удаления выполняются во время восстановления, а не во время проверки;
- жёсткий лимит лога PG – устанавливает жёсткий лимит на длину лога PG.

Нажать «Обновить».

6.5.16. Приоритет восстановления OSD

Для того чтобы установить приоритет восстановления OSD, необходимо выбрать «Cluster-wide configuration», затем в меню действий выбрать «Приоритет восстановления». В открывшемся окне заполнить предлагаемую форму (рис. 68).

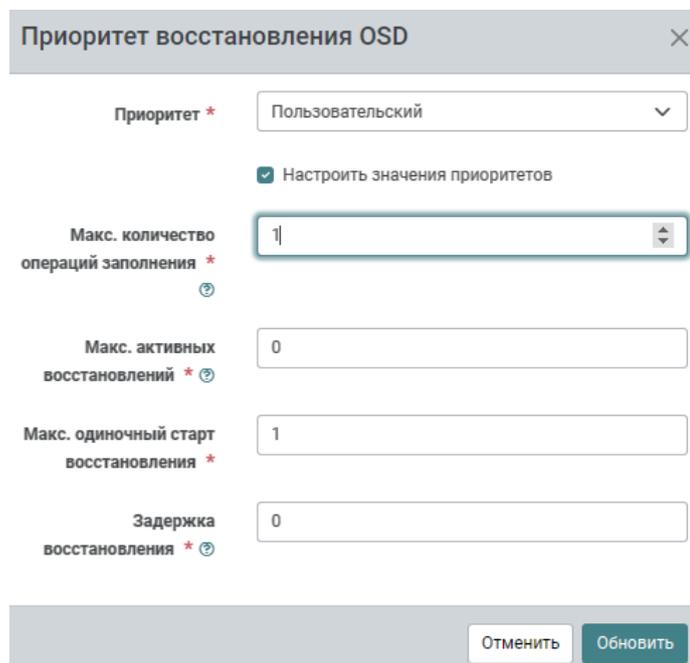


Рисунок 68

В поле «Приоритет» из выпадающего списка выбрать приоритет («Низ.», «По умолчанию», «Выс.» или «Пользовательский»).

При необходимости выбрать опцию «Настроить значения приоритетов».

В поле «Макс. количество операций заполнения» указать максимальное количество одновременных локальных и удалённых повторных заполнений или восстановлений на OSD.

В поле «Макс. активных восстановлений» указать количество одновременных активных операций восстановления на OSD (если значения `overrides_ssd` и `_hdd` не равны 0).

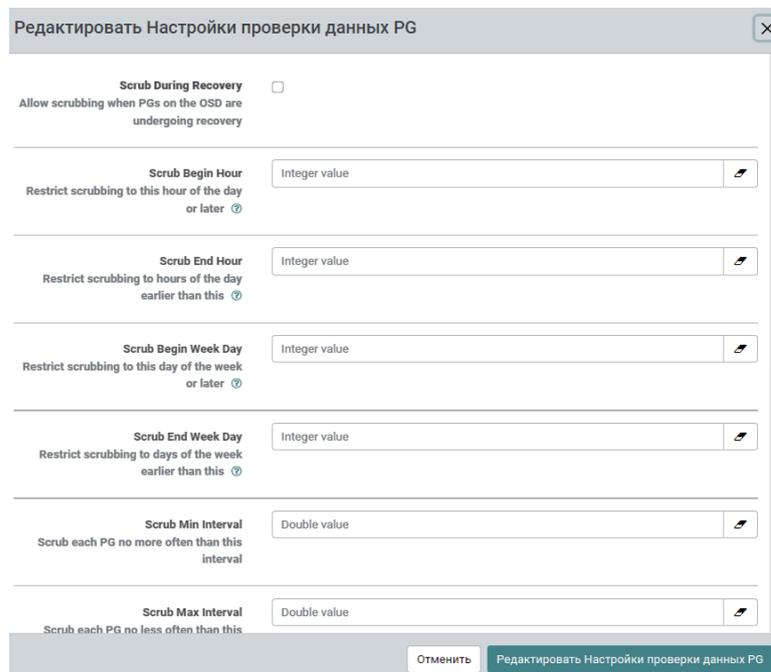
В поле «Макс. одиночный старт восстановления» указать максимальный одиночный старт восстановления.

В поле «Задержка восстановления» указать время (в секундах), необходимое для перехода в спящий режим перед следующей операцией восстановления или обратного заполнения.

Нажать «Обновить».

6.5.17. Проверка данных PG

Для проверки данных PG необходимо выбрать «Cluster-wide configuration», затем в меню действий выбрать «Проверка данных PG». В открывшемся окне заполнить предлагаемую форму (рис. 69).



Редактировать Настройки проверки данных PG

Scrub During Recovery
Allow scrubbing when PGs on the OSD are undergoing recovery

Scrub Begin Hour Integer value
Restrict scrubbing to this hour of the day or later ⓘ

Scrub End Hour Integer value
Restrict scrubbing to hours of the day earlier than this ⓘ

Scrub Begin Week Day Integer value
Restrict scrubbing to this day of the week or later ⓘ

Scrub End Week Day Integer value
Restrict scrubbing to days of the week earlier than this ⓘ

Scrub Min Interval Double value
Scrub each PG no more often than this interval

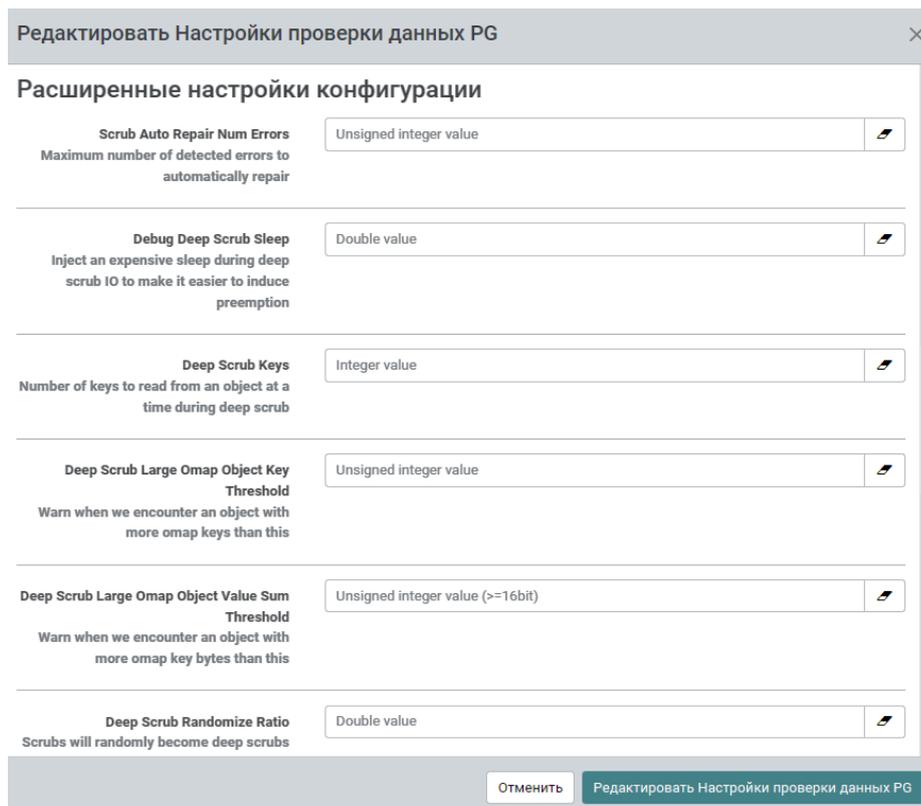
Scrub Max Interval Double value
Scrub each PG no less often than this

Отменить Редактировать Настройки проверки данных PG

Рисунок 69

Примечание. Проверка данных важна для поддержания целостности данных, но она снижает производительность.

Чтобы отредактировать расширенные настройки конфигурации необходимо нажать «Дополнительно...» (рис. 70).



Редактировать Настройки проверки данных PG

Расширенные настройки конфигурации

Scrub Auto Repair Num Errors Unsigned integer value
Maximum number of detected errors to automatically repair

Debug Deep Scrub Sleep Double value
Inject an expensive sleep during deep scrub IO to make it easier to induce preemption

Deep Scrub Keys Integer value
Number of keys to read from an object at a time during deep scrub

Deep Scrub Large Omap Object Key Threshold Unsigned integer value
Warn when we encounter an object with more omap keys than this

Deep Scrub Large Omap Object Value Sum Threshold Unsigned integer value (>=16bit)
Warn when we encounter an object with more omap key bytes than this

Deep Scrub Randomize Ratio Double value
Scrubs will randomly become deep scrubs

Отменить Редактировать Настройки проверки данных PG

Рисунок 70

Нажать «Редактировать Настройки проверки данных PG».

6.5.18. Общая производительность

Вкладка «Общая производительность» (раздел «Кластер» подраздел «OSD») показывает подробную информацию о производительности OSD (рис. 71).

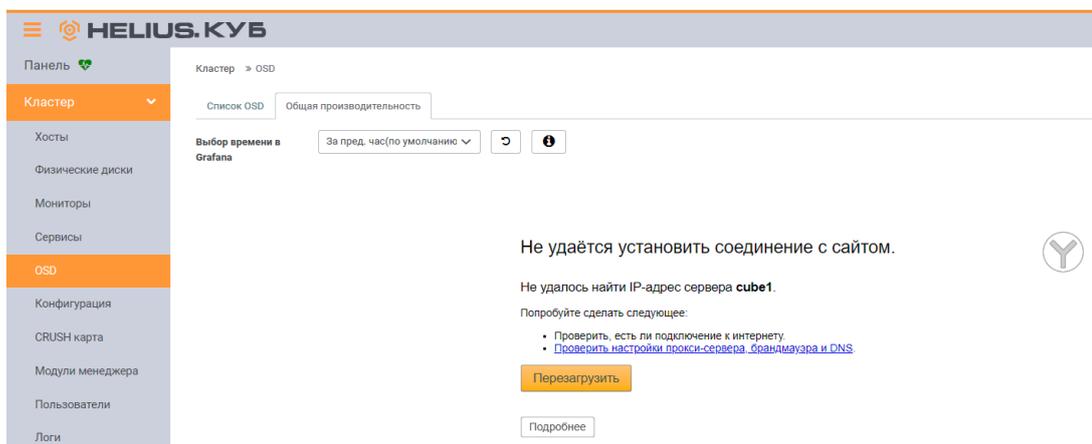


Рисунок 71

В поле «Выбор времени в Grafana» из выпадающего списка выбрать интервал времени отображения подробной информации о производительности OSD.

Кнопка «Сбросить настройки» позволяет сбросить настройки выбора времени в Grafana.

Кнопка «Показать скрытую информацию» позволяет просмотреть скрытую информацию.

Примечание. Если ниже не появилась встроенная панель Grafana, пожалуйста, перейдите по указанной в скрытой информации ссылке, чтобы проверить доступность Grafana и отсутствие проблем с сертификатами HTTPS. Возможно, вам потребуется перезагрузить эту страницу после принятия исключений сертификатов браузера.

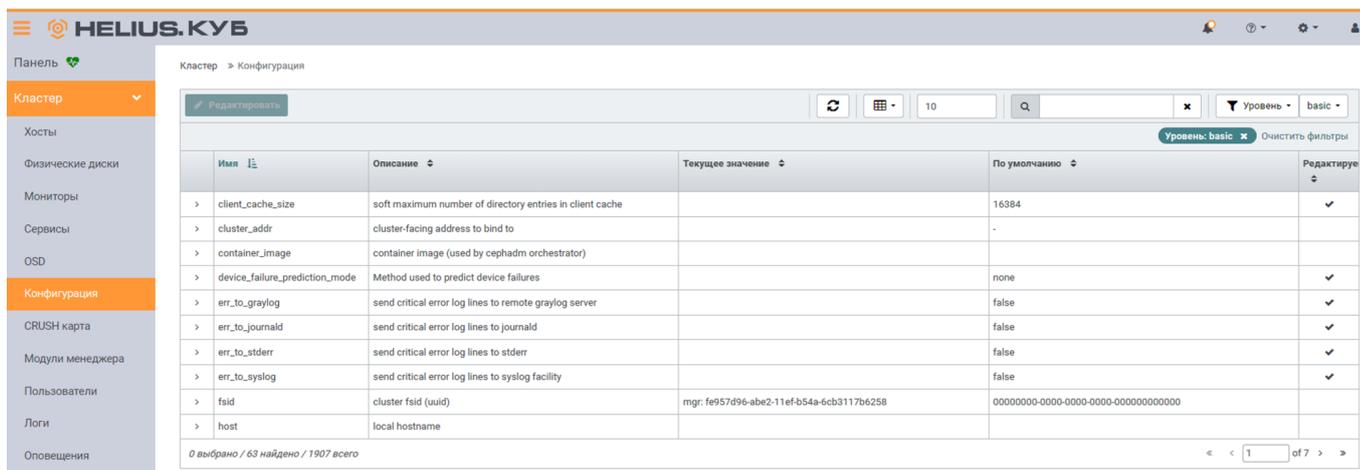
6.6. Редактор конфигурации

ПО позволяет вносить изменения в конфигурацию во время выполнения. Редактор конфигурации отображает полный список параметров конфигурации кластера, который содержит название параметра, его краткое описание, текущие значения и значения по умолчанию, а также информацию о том, можно ли изменить параметр.

В разделе «Кластер» из выпадающего списка выбрать подраздел «Конфигурация» (рис. 72).

Для параметров конфигурации доступны следующие фильтры:

- уровень: базовый (basic), продвинутый (advanced) и для разработчиков (dev);
- сервис: любой, mon, mgr, osd, mds, common, mds_client, rgw и аналогичные фильтры;
- источник: любой, mon и аналогичные фильтры;
- изменено: да или нет.

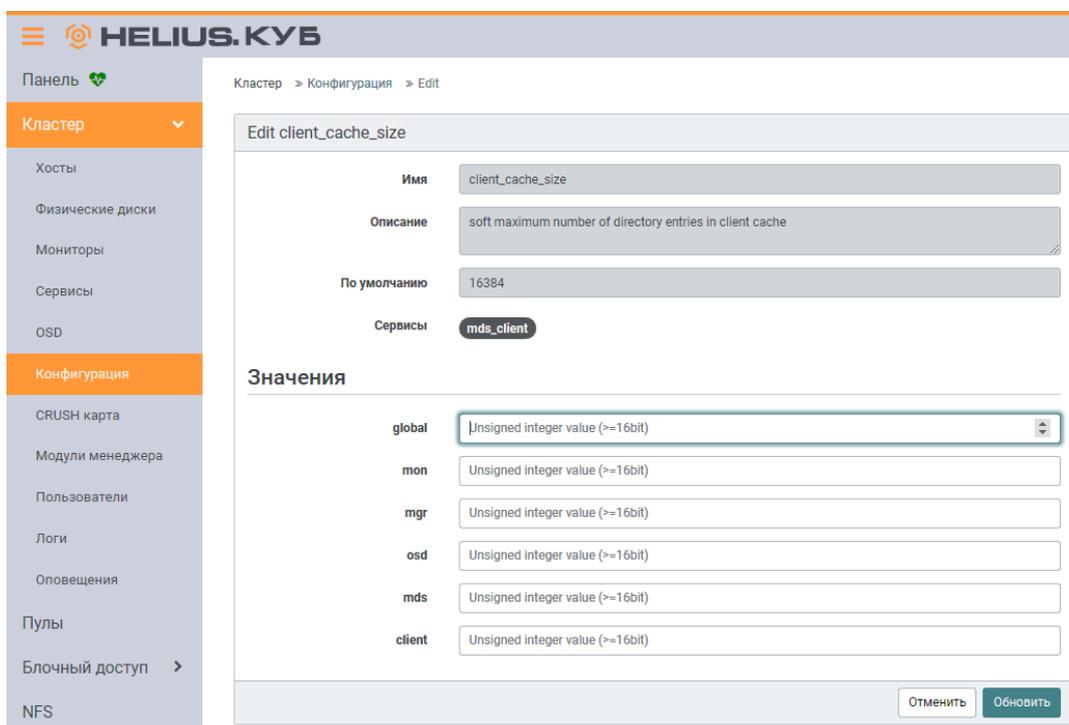


| Имя | Описание | Текущее значение | По умолчанию | Редактируе |
|--------------------------------|--|---|--------------------------------------|------------|
| client_cache_size | soft maximum number of directory entries in client cache | | 16384 | ✓ |
| cluster_addr | cluster-facing address to bind to | | - | |
| container_image | container image (used by cephadm orchestrator) | | | |
| device_failure_prediction_mode | Method used to predict device failures | | none | ✓ |
| err_to_graylog | send critical error log lines to remote graylog server | | false | ✓ |
| err_to_journald | send critical error log lines to journald | | false | ✓ |
| err_to_stderr | send critical error log lines to stderr | | false | ✓ |
| err_to_syslog | send critical error log lines to syslog facility | | false | ✓ |
| fsid | cluster fsid (uuid) | mgr: fe957d96-abe2-11ef-b54a-6cb3117b6258 | 00000000-0000-0000-0000-000000000000 | |
| host | local hostname | | | |

Рисунок 72

6.6.1. Редактирование конфигурации

Для редактирования конфигурации необходимо выбрать параметр конфигурации и на панели инструментов нажать «Редактировать». В открывшемся окне обновить изменяемые свойства и нажать «Обновить» (рис. 73).



Кластер > Конфигурация > Edit

Edit client_cache_size

Имя: client_cache_size

Описание: soft maximum number of directory entries in client cache

По умолчанию: 16384

Сервисы: mds_client

Значения

global: Unsigned integer value (>=16bit)

mon: Unsigned integer value (>=16bit)

mgr: Unsigned integer value (>=16bit)

osd: Unsigned integer value (>=16bit)

mds: Unsigned integer value (>=16bit)

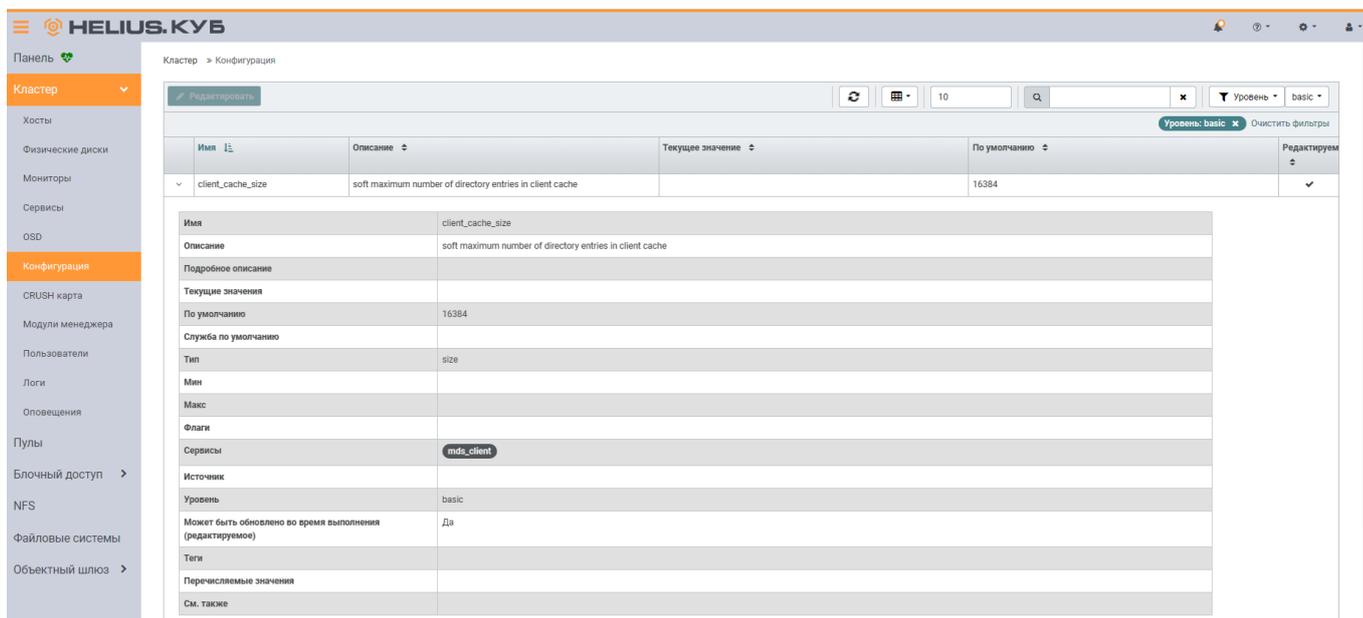
client: Unsigned integer value (>=16bit)

Отменить Обновить

Рисунок 73

6.6.2. Просмотр информации о конфигурации

Для просмотра подробной информации о параметре конфигурации необходимо развернуть содержимое строки (рис. 74).



Клиентская кеш-память (client_cache_size)

| Имя | Описание | Текущее значение | По умолчанию | Редактируем |
|-------------------|--|------------------|--------------|-------------|
| client_cache_size | soft maximum number of directory entries in client cache | | 16384 | ✓ |

| Имя | client_cache_size |
|--|--|
| Описание | soft maximum number of directory entries in client cache |
| Подробное описание | |
| Текущие значения | |
| По умолчанию | 16384 |
| Служба по умолчанию | |
| Тип | size |
| Мин | |
| Макс | |
| Флаги | |
| Сервисы | mds_client |
| Источник | |
| Уровень | basic |
| Может быть обновлено во время выполнения (редактируемое) | Да |
| Теги | |
| Перечисляемые значения | |
| См. также | |

Рисунок 74

6.7. CRUSH карта

CRUSH – алгоритм, который может вычислить физическое расположение данных на основе имени объекта, карты кластера и правил CRUSH. CRUSH требует наличия карты кластера и использует карту CRUSH для псевдослучайного хранения и извлечения данных в OSD с равномерным распределением данных по кластеру.

Карты CRUSH содержат информацию об устройствах хранения кластера, иерархию домена отказов и список правил CRUSH. Используя алгоритм CRUSH, ПО вычисляет, какая PG должна содержать объект, и в каком OSD должна храниться PG.

Карты CRUSH используют правила CRUSH (см. подробнее 7.1.2), которые представляют собой правила, определяющие размещение данных для пула. Когда CRUSH назначает PG OSD, он проверяет количество реплик для пула и назначает PG OSD таким образом, что каждая реплика PG назначается другому OSD. Набор OSD, которые должны содержать реплики конкретной PG, называется действующим набором. ПО обрабатывает запрос клиента с помощью активного набора, который представляет собой набор OSD, фактически обрабатывающих запросы. В большинстве случаев активный набор и действующий набор практически идентичны. В противном случае это указывает на то, что ПО переносит данные, OSD восстанавливается или возникла проблема (в таких случаях кластер имеет статус «HEALTH_WARN»).

Прежде чем записать данные в PG, она должна находиться в состоянии «active» и должна быть в состоянии «clean». Чтобы определить текущее состояние PG, первичный OSD (первый OSD в действующем наборе) согласовывает текущее состояние PG со вторичным и третичным OSD (при условии, что в пуле есть три реплики PG).

Примечание. Устройства хранения данных добавляются на карту CRUSH при добавлении OSD в кластер.

В разделе «Кластер» из выпадающего списка выбрать подраздел «CRUSH карта» (рис. 75).

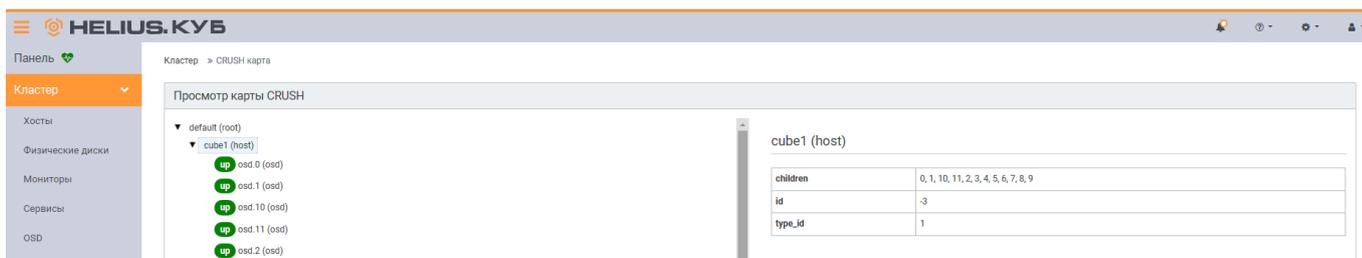


Рисунок 75

При выборе OSD отобразится информация об этом OSD (рис. 76).

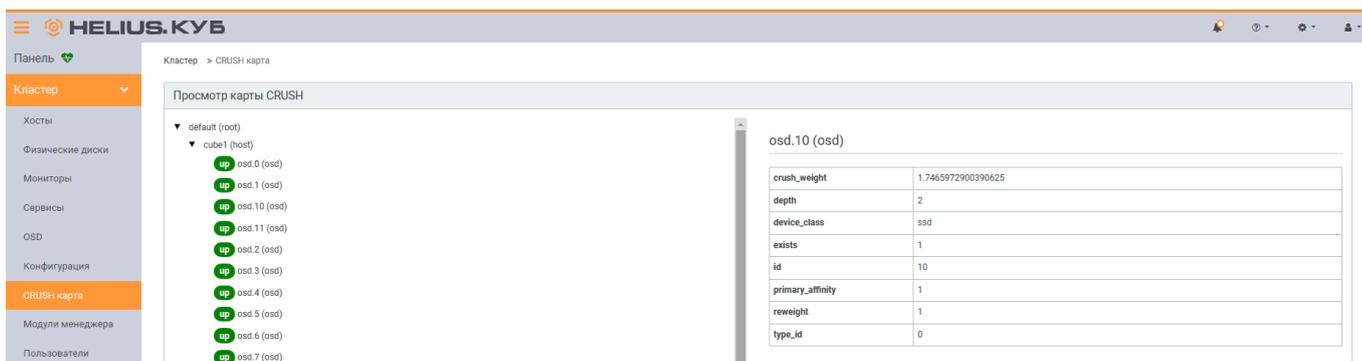


Рисунок 76

При добавлении OSD в кластер или удалении из кластера, алгоритм CRUSH перебалансирует кластер, перемещая PG в OSD или из них для восстановления баланса.

6.8. Модули менеджера

Менеджеры – сервис мониторинга, который отвечает за отслеживание показателей времени выполнения и текущего состояния кластера, включая использование хранилища, текущие показатели производительности и загрузку системы.

Модули менеджера позволяют управлять элементами администрирования на панели управления, включая или отключая функции по запросу. По умолчанию все функции включены. При отключении функции элементы администрирования панели управления (веб-интерфейса) становятся скрытыми, а связанные с ними конечные точки REST API отклоняют любые дальнейшие запросы к этой функции.

В разделе «Кластер» из выпадающего списка выбрать подраздел «Модули менеджера» (рис. 77).

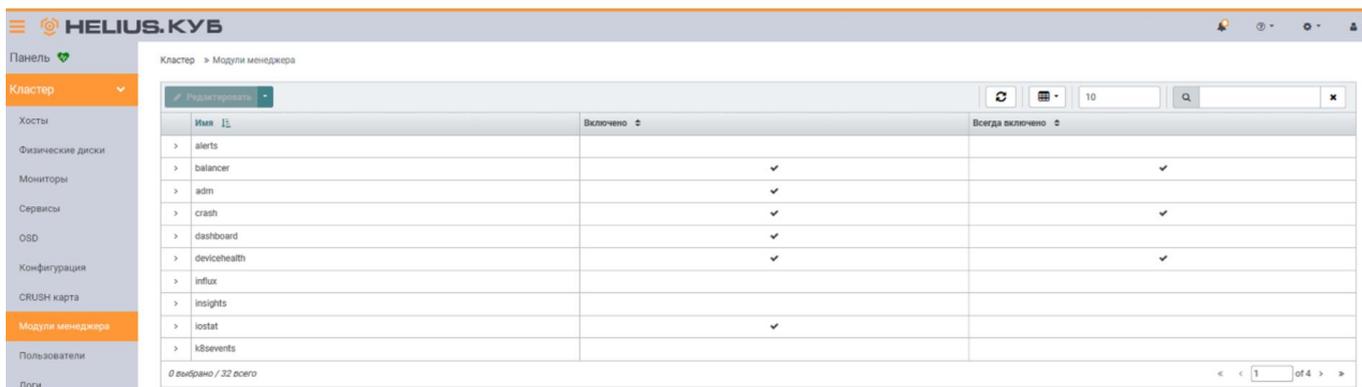


Рисунок 77

6.8.1. Просмотр информации о модулях менеджера

Для просмотра подробной информации о модуле менеджера необходимо развернуть содержимое строки (рис. 78).

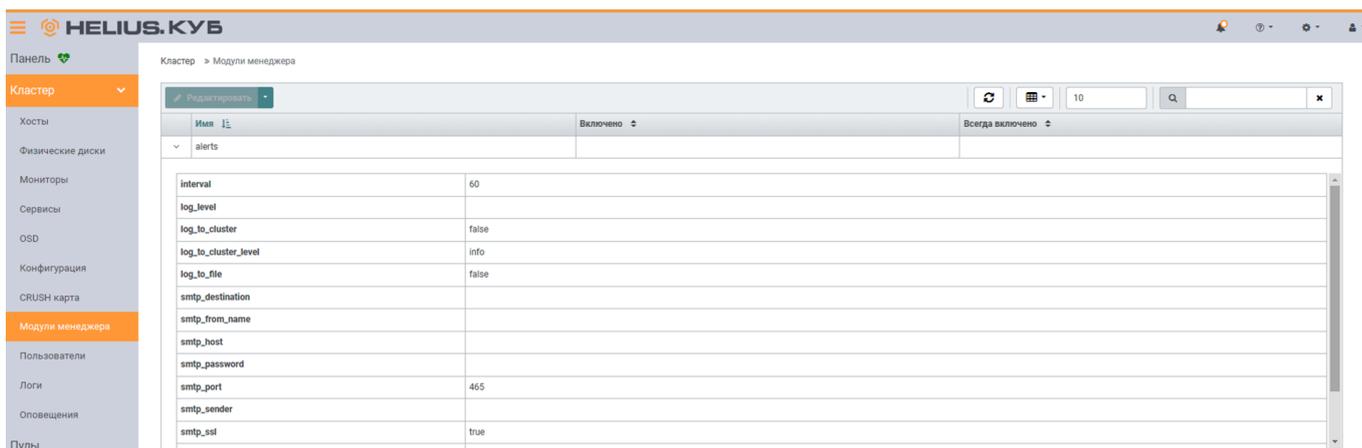


Рисунок 78

6.8.2. Редактирование модулей менеджера

Для редактирования необходимо выбрать модуль менеджера и на панели инструментов нажать «Редактировать». В открывшемся окне обновить изменяемые свойства и нажать «Обновить».

Выбрав опцию (сняв выбор опции), можно включить (отключить) функции отображения элементов администрирования на панели управления.

Примечания:

1. После отключения функции может потребоваться до 20 с чтобы отобразить изменения в веб-интерфейсе.
2. Не все модули менеджера имеют настраиваемые параметры. Если модуль менеджера не настраиваемый, то кнопка «Редактировать» неактивна.

6.8.3. Включение модулей менеджера

Для включения модуля менеджера необходимо выбрать модуль менеджера и в меню действий выбрать «Включить».

6.8.4. Отключение модулей менеджера

Для отключения модуля менеджера необходимо выбрать модуль менеджера и в меню действий выбрать «Отключить».

6.9. Пользователи кластера хранения

Администратор ПО может управлять базой пользователей кластера хранения, обеспечивая аутентификацию и контроль доступа к объектам в кластере. Пользователем кластера хранения могут быть как отдельные лица, так и приложения, которые взаимодействуют с службами кластера.

Пользователь кластера хранения – это не то же самое, что пользователь объектного хранилища или пользователь файловой системы. Объектный шлюз использует пользователя кластера хранения для взаимодействия между службами шлюза и кластером хранения, но у объектного шлюза есть собственная функция управления пользователями для конечных пользователей. Файловая система использует семантику POSIX, и пользовательское пространство, связанное с файловой системой, которое отличается от пользовательского пространства, связанного с пользователем кластера хранения.

В разделе «Кластер» из выпадающего списка выбрать подраздел «Пользователи» (рис. 79).

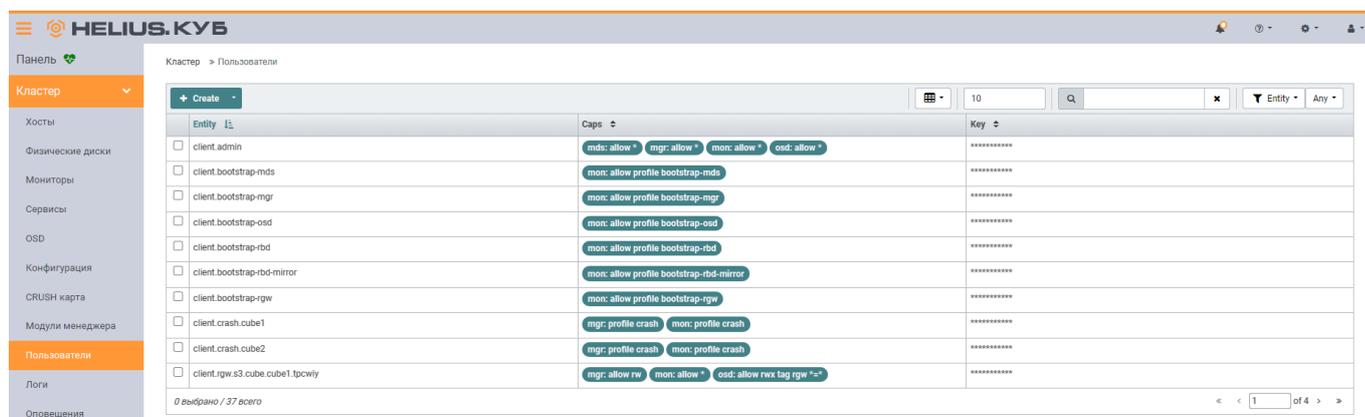


Рисунок 79

6.9.1. Создание пользователя кластера хранения

Для создания пользователя кластера хранения необходимо нажать «Создать» на панели инструментов. В открывшемся окне заполнить предлагаемую форму (рис. 80).

В поле «User entity» указать имя пользователя кластера хранения (тип и идентификатор).

Примечание. Обозначение пользователя кластера хранения состоит из двух частей: тип пользователя и его идентификатор. Например, client.admin – это пользователь типа client с идентификатором admin.

Чтобы предоставить пользователю кластера хранения возможность выполнять административные функции в кластере хранения, ему необходимо предоставить соответствующие административные права. Административные права обычно включают разрешения на управление кластером, пулами, OSD, MON и другими компонентами.

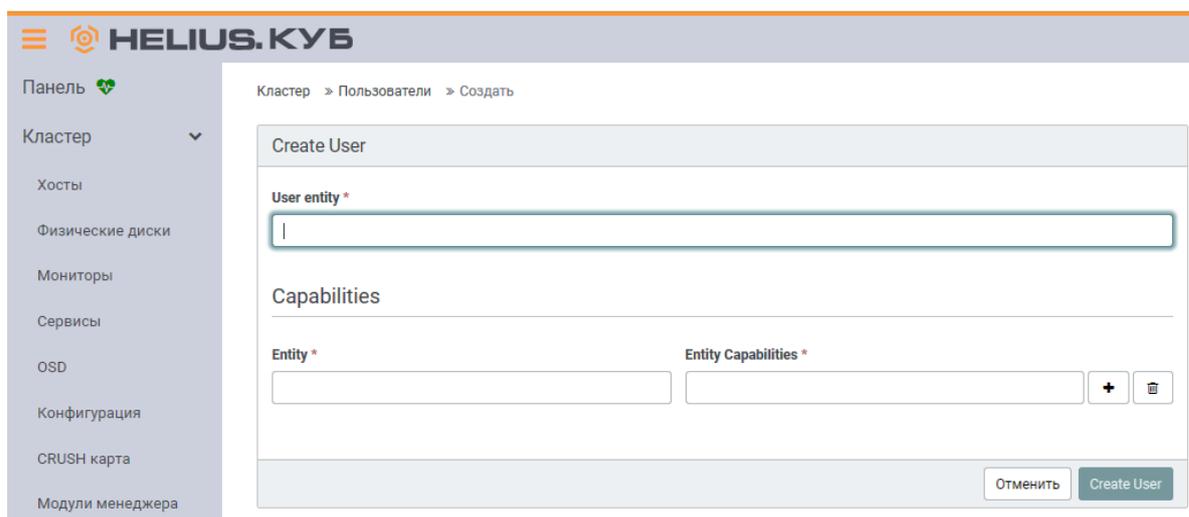


Рисунок 80

Для описания разрешений, предоставляемых аутентифицированному пользователю для использования функций MON, OSD и MDS использует термин «возможности». Также возможности могут ограничивать доступ к данным в пуле, пространстве имён в пуле. Администратор ПО определяет возможности пользователя кластера хранения при создании или обновлении этого пользователя.

Основные административные права (возможности) включают:

- mon – права для операций с MON («r», «w», «x» и настройки доступа);
- osd – права для операций с OSD («r», «w», «x», «class-read» и «class-write» и настройки доступа). Кроме того, возможности OSD позволяют настраивать пулы и пространства имён;
- mds – права для операций с MDS (если используется Helius.КубФС);
- mgr – права для операций с модулями менеджера («r», «w», «x» и настройки доступа);
- admin – позволяет доступ к административному API и другим командам высокого уровня.

В ПО используются следующие возможности доступа, которые могут быть предоставлены пользователю кластера хранения или клиенту ПО (блочное устройство, объектный шлюз, файловая система и собственный API):

- «allow» – предшествует настройкам доступа для служб. Подразумевает «rw» только для MDS;
- «r» – предоставляет пользователю доступ для чтения;
- «w» – предоставляет пользователю доступ на запись к объектам;
- «x» – предоставляет пользователю возможность вызывать методы класса (выполнять операции чтения и записи), а также выполнять «auth» операции на MON;
- «class-read» – предоставляет пользователю возможность вызывать методы чтения класса (подмножество «x»);

– «class-write» – предоставляет пользователю возможность вызывать методы записи класса (подмножество «x»);

– «*», «all» – предоставляет пользователю разрешения на чтение, запись и выполнение для конкретного демона или пула, а также возможность выполнять команды администратора.

Примечания:

1. Значение «allow *» предоставляет полный доступ к указанному компоненту.

2. Если предоставить пользователю кластера хранения возможности взаимодействия с OSD, но не ограничить его определённым пулом, то пользователь кластера хранения будет иметь доступ ко всем пулам в кластере хранения.

На панели «Capabilities» заполнить возможности, предоставляемые пользователю кластера хранения.

Для добавления Entity необходимо нажать «+», для удаления нажать на значок «Корзина».

В поле «Entity» указать службу, с которой будет взаимодействовать пользователь кластера хранения.

В поле «Entity Capabilities» указать возможности, которые назначены пользователю кластера хранения.

После предоставления административных прав пользователь кластера хранения может использовать REST API для выполнения административных задач.

Нажать «Создать Пользователя».

При необходимости можно отозвать административные права у пользователя кластера хранения, обновив его права и удалив разрешения.

Пользователю кластера хранения необходимы ключи для взаимодействия со службами ПО. Ключи позволяет пользователю кластера хранения проходить аутентификацию в кластере.

6.9.2. Редактирование пользователя кластера хранения

Для редактирования пользователя кластера хранения необходимо выбрать пользователя и на панели инструментов нажать «Редактировать». В открывшемся окне обновить изменяемые свойства и нажать «Обновить Пользователя».

6.9.3. Экспорт данных пользователя кластера хранения

Данные пользователей кластера хранения или клиентов можно экспортировать из панели управления на локальный хост.

Для экспорта данных пользователя кластера хранения необходимо выбрать пользователя и в меню действий выбрать «Экспорт». В открывшемся окне нажать «Скопировать в буфер обмена», а затем нажать «Закрыть» (рис. 81).



Рисунок 81

6.9.4. Импорт пользователя кластера хранения

Данные пользователей кластера хранения или клиентов можно импортировать с локального хоста на панель управления.

Для импорта пользователя кластера хранения необходимо выбрать пользователя и в меню действий выбрать «Импорт». В открывшемся окне выбрать локальный файл для загрузки, нажав «Выберите файл», а затем нажать «Импорт Пользователя» (рис. 82).

Кластер > Пользователи > Импорт

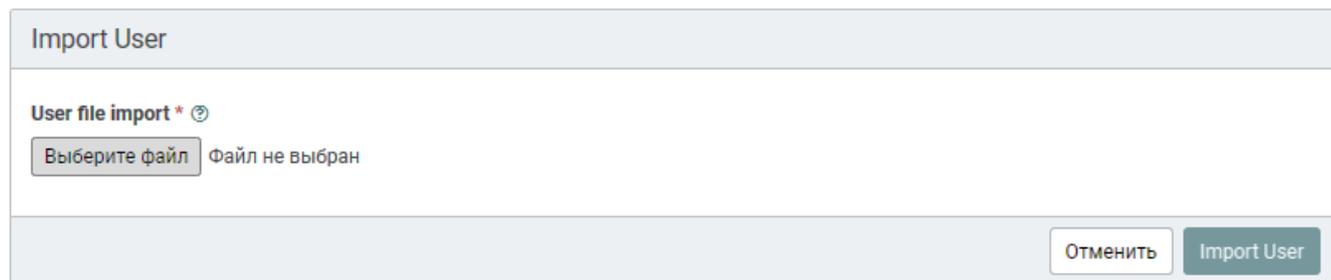


Рисунок 82

6.9.5. Удаление пользователя кластера хранения

Для удаления пользователя кластера хранения необходимо выбрать пользователя и в меню действий выбрать «Удалить». Подтвердить удаление в появившемся окне.

6.10. Логи

При возникновении проблемы в ПО можно отправить отчёт о проблеме. В отчёт необходимо включить все журналы (логи), необходимые для диагностики и устранения проблемы, которые затем отправить в службу технической поддержки.

В разделе «Кластер» из выпадающего списка выбрать подраздел «Логи». Данный подраздел состоит из вкладок «Логи кластера», «Логи аудита» и «Логи Служб».

На панели инструментов журналы (логи) можно отфильтровать по приоритету, ключевому слову, дате или временному диапазону, также их можно скачать (в формате «JSON» и «Text») или скопировать в буфер обмена.

6.10.1. Логи кластера

Вкладка «Логи кластера» показывает системные сообщения (логи), связанные с жизненным циклом кластера (рис. 83).

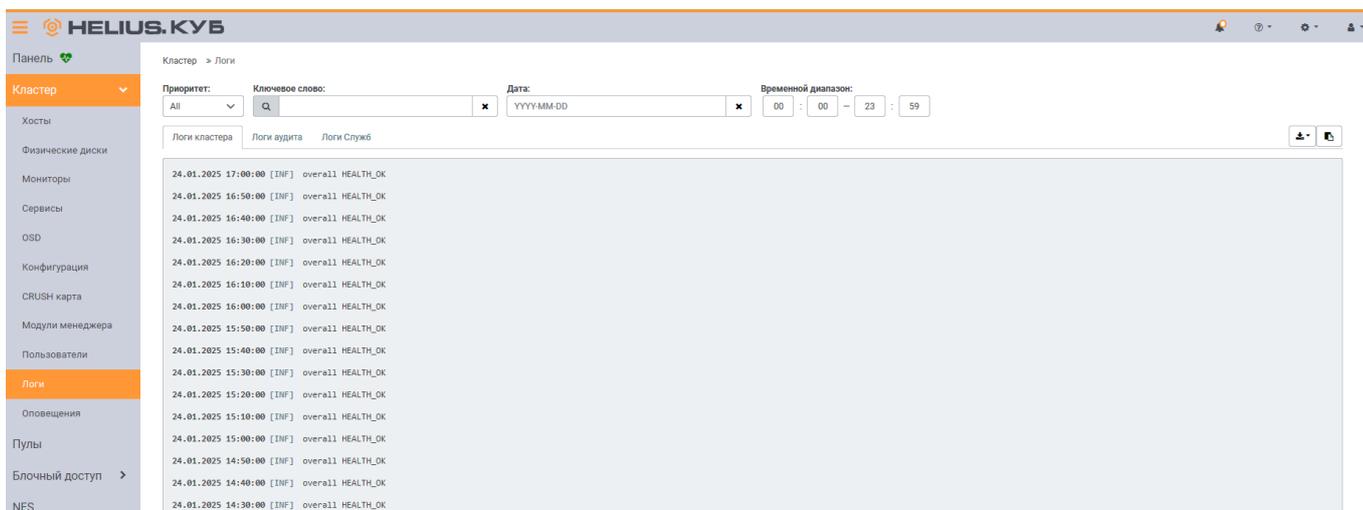


Рисунок 83

6.10.2. Логи аудита

Вкладка «Логи аудита» показывает все операции управления, выполненные пользователями, и события их активности (рис. 84).

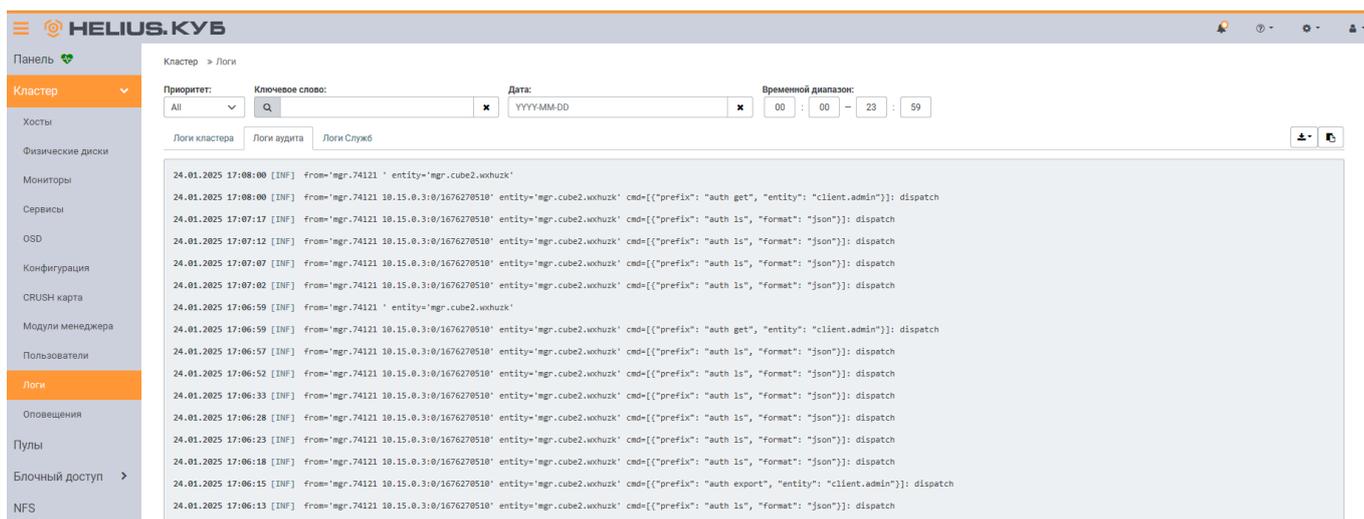


Рисунок 84

6.10.3. Логи Служб

Вкладка «Логи Служб» показывает системные сообщения (логи) служб, связанные с жизненным циклом служб (рис. 85).

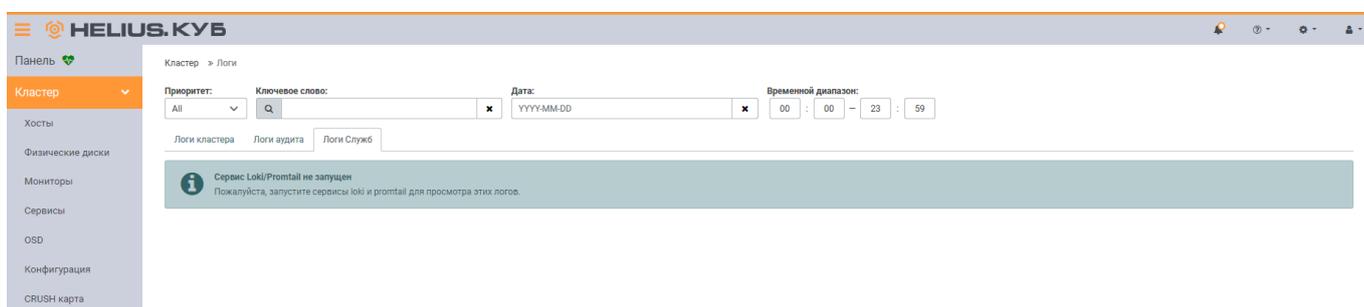


Рисунок 85

6.11. Оповещения

ПО использует систему мониторинга «Prometheus» для отслеживания производительности и доступности кластера хранения, хостов инфраструктуры и развёрнутых сервисов. Кроме того, система формирует оповещения, которые можно настроить для отправки в виде уведомлений.

Оповещение создаётся и регистрируется каждый раз, когда выполняется одно из следующих условий или происходит соответствующее событие:

- возникла критическая проблема с кластером, его компонентами (CS, MDS), дисками, хостами или сервисами;
- кластеру требуются настройка или дополнительные ресурсы для построения или восстановления работоспособности;
- сети требуется настройка, или в ней происходят неполадки, которые могут повлиять на производительность;
- в кластере скоро закончится или закончилось доступное пространство.

Оповещения можно сортировать по названию, заданию, уровню важности, состоянию и времени начала.

Оповещения можно проигнорировать (удалить из списка оповещений) или отложить на несколько часов. Отложенные оповещения повторно появляются в списке через некоторое время.

В разделе «Кластер» из выпадающего списка выбрать подраздел «Оповещения».

6.11.1. Активные оповещения

На данной вкладке отображены активные оповещения и информация о них (рис. 86).

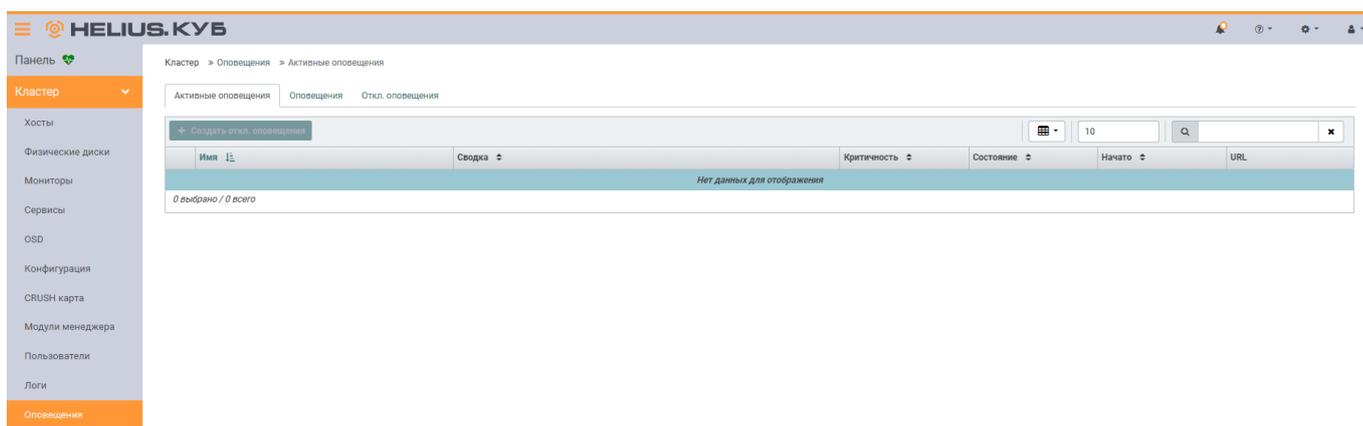


Рисунок 86

Для создания откл. оповещения необходимо на панели инструментов нажать «Создать откл. оповещения».

6.11.2. Оповещения

На данной вкладке отображены все оповещения (рис. 87).

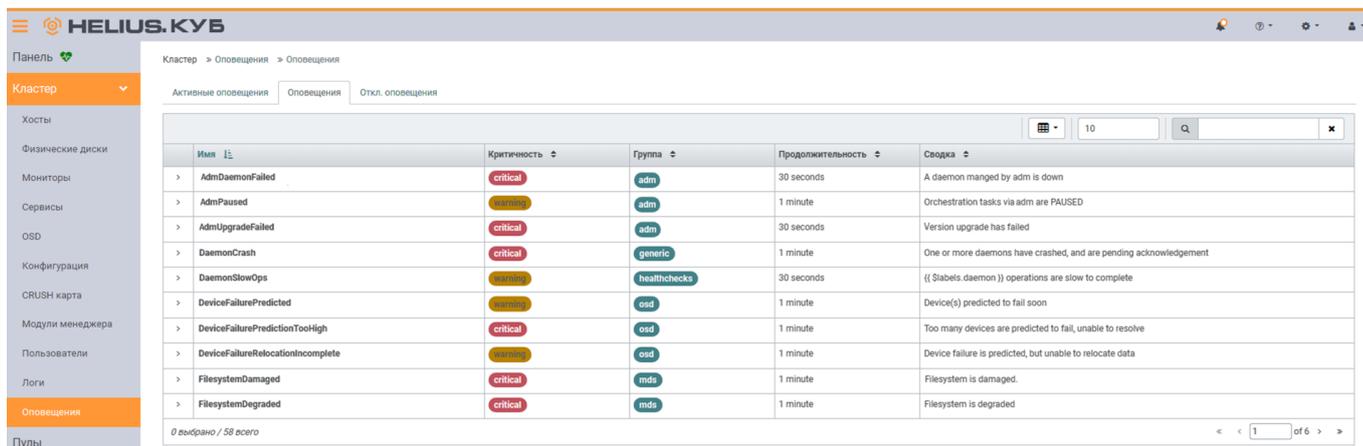


Рисунок 87

Для просмотра оповещения необходимо развернуть содержимое строки (рис. 88).

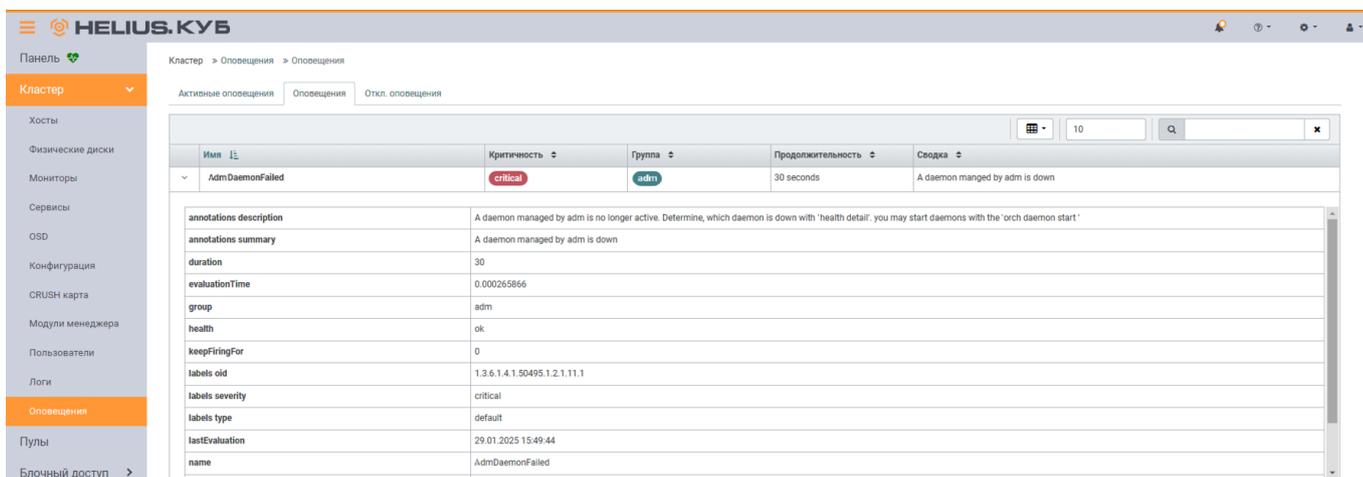


Рисунок 88

6.11.3. Откл. оповещения

Вкладка «Откл. оповещения» отображает отключённые оповещения (рис. 89).

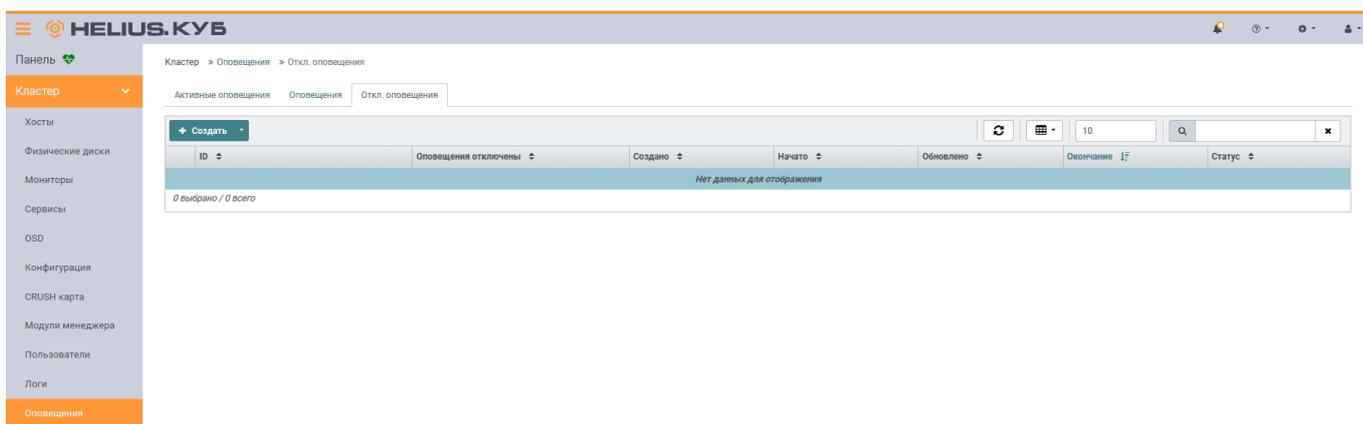


Рисунок 89

6.11.3.1. Создание откл. оповещения

Для создания откл. оповещения необходимо на вкладке «Откл. оповещения» нажать «Создать» на панели инструментов. В открывшемся окне заполнить предлагаемую форму (рис. 90).

В поле «Автор» указать автора откл. оповещения.

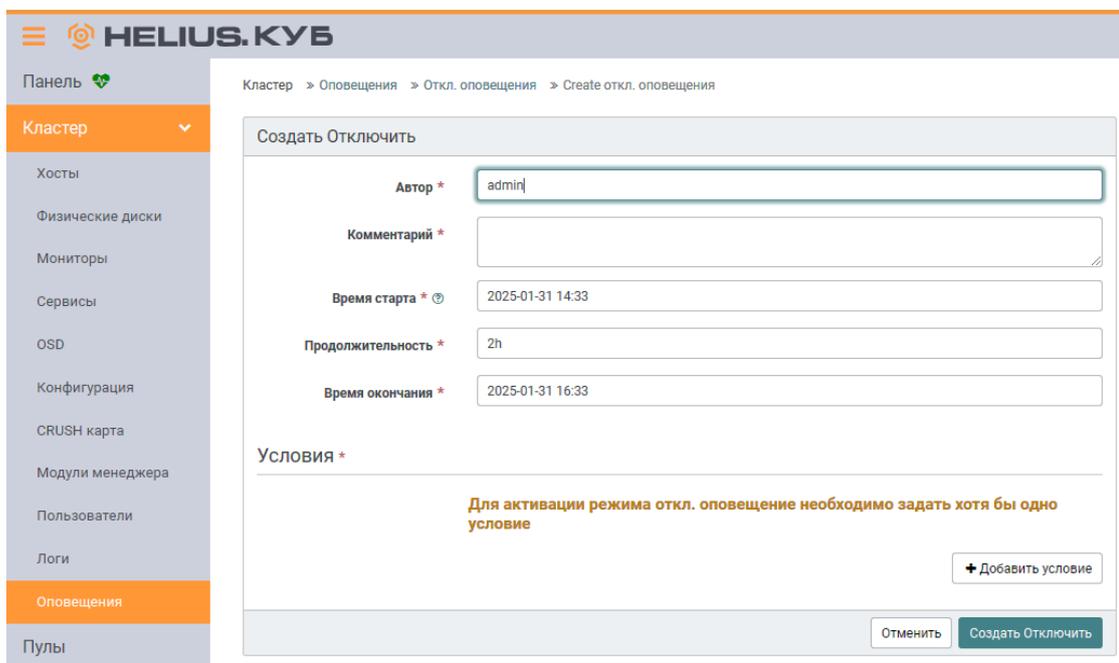
В поле «Комментарий» указать комментарий, используемый для описания откл. оповещения.

В поле «Время старта» указать дату и время старта откл. оповещения (если указанное время начала уже прошло, то будет использовано время создания записи).

В поле «Продолжительность» указать продолжительность откл. оповещения.

В поле «Время окончания» указать дату и время окончания откл. оповещения.

Для активации режима откл. оповещение необходимо задать хотя бы одно условие.



Панель  HELIUS.KUB

Кластер > Оповещения > Откл. оповещения > Create откл. оповещения

Создать Отключить

Автор * admin

Комментарий *

Время старта * 2025-01-31 14:33

Продолжительность * 2h

Время окончания * 2025-01-31 16:33

Условия *

Для активации режима откл. оповещение необходимо задать хотя бы одно условие

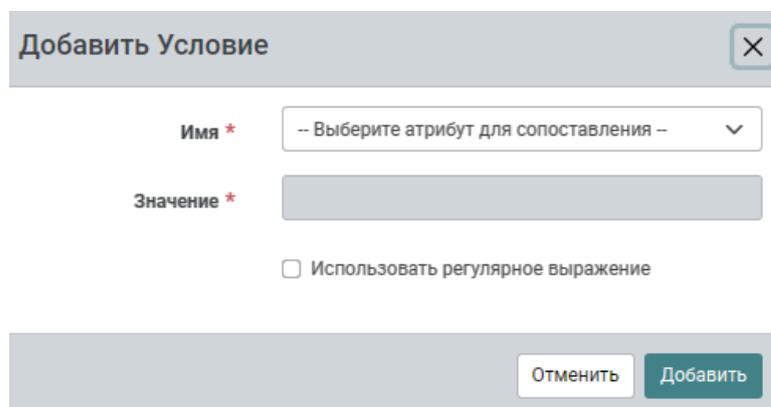
+ Добавить условие

Отменить Создать Отключить

Рисунок 90

Нажать «Добавить условие».

В открывшемся окне заполнить предлагаемую форму (рис. 91).



Добавить Условие

Имя * -- Выберите атрибут для сопоставления --

Значение *

Использовать регулярное выражение

Отменить Добавить

Рисунок 91

В поле «Имя» из выпадающего списка выбрать атрибут для сопоставления.

В поле «Значение» указать (или выбрать из предложенных) значение для выбранного атрибута.

При необходимости выбрать опцию «Использовать регулярное выражение».

Нажать «Добавить».

Нажать «Создать Отключить».

6.11.3.2. Создать заново откл. оповещения

6.11.3.3. Редактировать откл. оповещения

6.11.3.4. Вытеснить откл. оповещения

7. ПУЛЫ

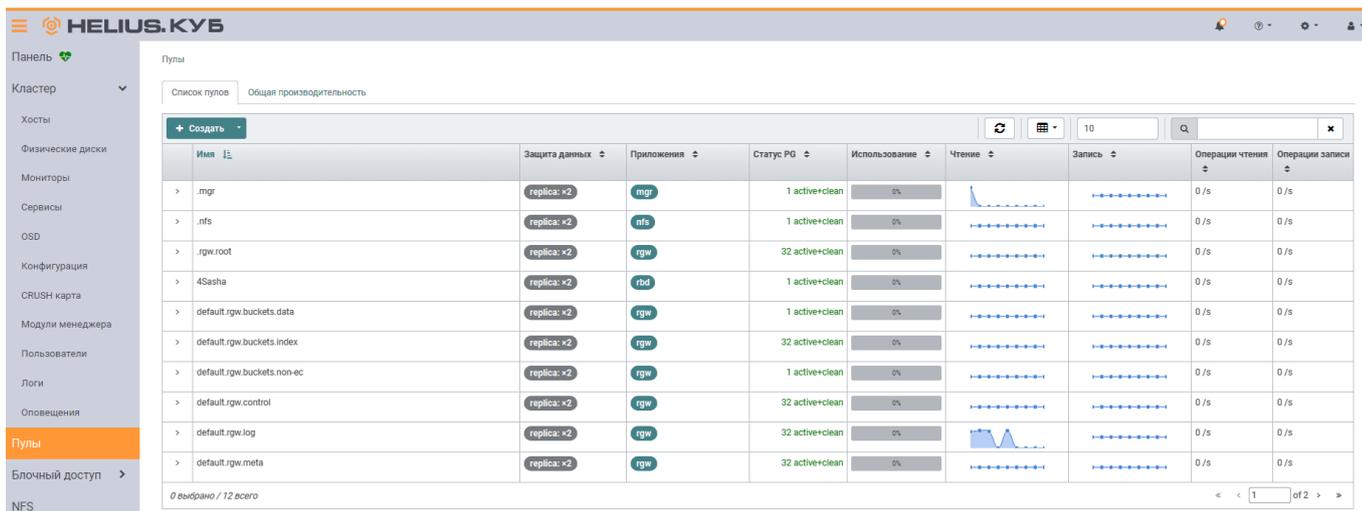
Пул – логическая группа для хранения объектов данных, которая предоставляет информацию об отказоустойчивости, группах размещения, правилах CRUSH и квотах.

Пул разделён на группы размещения (PG). При размещении данных в кластере объекты сопоставляются с группами размещения, а группы размещения – с OSD. Алгоритм CRUSH вычисляет группу размещения для хранения объекта и вычисляет действующий набор OSD для группы размещения. CRUSH псевдослучайным образом помещает каждый объект в группу размещения. Затем CRUSH сохраняет каждую группу размещения в наборе OSD. Администраторы ПО устанавливают количество групп размещения при создании или изменении пула.

Набор правил CRUSH определяет домены отказов и домены производительности для пула, а также может идентифицировать OSD по типу носителя информации. CRUSH позволяет OSD хранить копии объектов в доменах отказов. Если большая часть кластера выходит из строя, кластер все равно может работать в деградированном состоянии, пока не восстановится.

Администраторы ПО задают набор правил CRUSH при создании пула (см. подробнее 7.1.2).

В разделе «Пулы» перейти на вкладку «Список пулов» (рис. 92)



| Имя | Защита данных | Приложения | Статус PG | Использование | Чтение | Запись | Операции чтения | Операции записи |
|------------------------------|---------------|------------|-----------------|---------------|--------|--------|-----------------|-----------------|
| > .mgr | replica: *2 | mgr | 1 active+clean | 0% | | | 0 /s | 0 /s |
| > .nfs | replica: *2 | nfs | 1 active+clean | 0% | | | 0 /s | 0 /s |
| > .rgw.root | replica: *2 | rgw | 32 active+clean | 0% | | | 0 /s | 0 /s |
| > 4Sasha | replica: *2 | rbd | 1 active+clean | 0% | | | 0 /s | 0 /s |
| > default.rgw.buckets.data | replica: *2 | rgw | 1 active+clean | 0% | | | 0 /s | 0 /s |
| > default.rgw.buckets.index | replica: *2 | rgw | 32 active+clean | 0% | | | 0 /s | 0 /s |
| > default.rgw.buckets.non-ec | replica: *2 | rgw | 1 active+clean | 0% | | | 0 /s | 0 /s |
| > default.rgw.control | replica: *2 | rgw | 32 active+clean | 0% | | | 0 /s | 0 /s |
| > default.rgw.log | replica: *2 | rgw | 32 active+clean | 0% | | | 0 /s | 0 /s |
| > default.rgw.meta | replica: *2 | rgw | 32 active+clean | 0% | | | 0 /s | 0 /s |

Рисунок 92

Состояния групп размещения можно подразделить на три категории:

– критичные (Critical) – указывают на существенные разрушения в кластере, должны быть предприняты немедленные действия для разрешения выявленных проблем;

– предостерегающие (Warning) – указывают на проблемы в кластере, которые не приведут к полной его остановке;

– информационные (Info) – передают информацию относительно текущего состояния групп размещения.

Группы размещения имеют одно или несколько состояний (статус PG):

- «activating» – PG подключён, но ещё не активен;
- «active» – запросы к PG обрабатываются;
- «backfill_toofull» – ожидается операция обратного заполнения, поскольку OSD перезаполнено;
- «backfill_unfound» – заполнение остановлено из-за ненайденных объектов;
- «backfill_wait» – PG ждёт своей очереди, чтобы начать заполнение;
- «backfilling» – особый случай восстановления, при котором все содержимое PG сканируется и синхронизируется вместо того, чтобы делать вывод о том, какое содержимое из журналов последних операций необходимо синхронизировать;
- «clean» – все объекты в PG реплицируются правильное количество раз;
- «creating» – PG все ещё создаётся;
- «deerp» – данные PG проверяются по сохранённым контрольным суммам. В сочетании с «scrubbing» является глубокой очисткой;
- «degraded» – некоторые объекты в PG ещё не реплицированы достаточное количество раз;
- «down» – реплика с необходимыми данными отключена, поэтому PG находится в автономном режиме (отключена);
- «failed_repair» – попытка ремонта PG не удалась, требуется ручное вмешательство;
- «forced_backfill» – PG был отмечен как требующий наивысшего приоритета восстановления;
- «forced_recovery» – PG был отмечен как имеющий наивысший приоритет заполнения;
- «incomplete» – у PG не хватает информации о записях в журнале (сбой пиринга) или отсутствуют «здоровые» копии. При возникновении данного состояния необходимо запустить неисправные OSD, которые могут содержать необходимую информацию;
- «inconsistent» – реплики PG не согласованы (например, объекты имеют неправильный размер, объекты отсутствуют в одной реплике после завершения восстановления);
- «peering» – PG проходит процесс пирингового взаимодействия;
- «peered» – PG установил пиринг, но не может обслуживать клиентский ввод/вывод из-за отсутствия достаточного количества копий для достижения настроенного параметра пула;
- «recovering» – выполняется миграция или синхронизация объектов и их реплик;
- «recovery_toofull» – ожидается операция восстановления, поскольку OSD перезаполнен;
- «recovery_unfound» – восстановление остановлено из-за ненайденных объектов;

- «recovery_wait» – PG ожидает резервирования локального/удалённого восстановления;
- «remapped» – PG временно сопоставлен с другим набором OSD, отличным от указанного

CRUSH;

- «repair» – PG проверяется и исправляются любые обнаруженные несоответствия (если это возможно);
- «replay» – PG ожидает, пока клиенты воспроизведут операции после сбоя OSD;
- «snaptrim» – PG обрезает снапшоты;
- «snaptrim_error» – ошибка, при которой прекратилась обрезка снапшотов;
- «snaptrim_wait» – PG поставлен в очередь на обрезку снапшотов;
- «scrubbing» – проверяются метаданные PG на предмет несоответствий;
- «splitting» – PG разбивается на несколько PG;
- «stale» – PG находится в неизвестном состоянии – мониторы не получали обновлений с момента изменения отображения PG;
- «undersized» – PG содержит меньшее количество копий, чем настроенный уровень репликации пула;
- «unknown» – информация о состоянии PG от OSD ещё не поступала.

ПО будет отображать статус «creating» при создании одной или нескольких групп размещения. После завершения пиринга состояние группы размещения должно быть «active+clean», что означает, что клиент ПО может начать запись в группу размещения.

Статус «active+clean» является оптимальным состоянием для группы размещения.

7.1. Создание пула

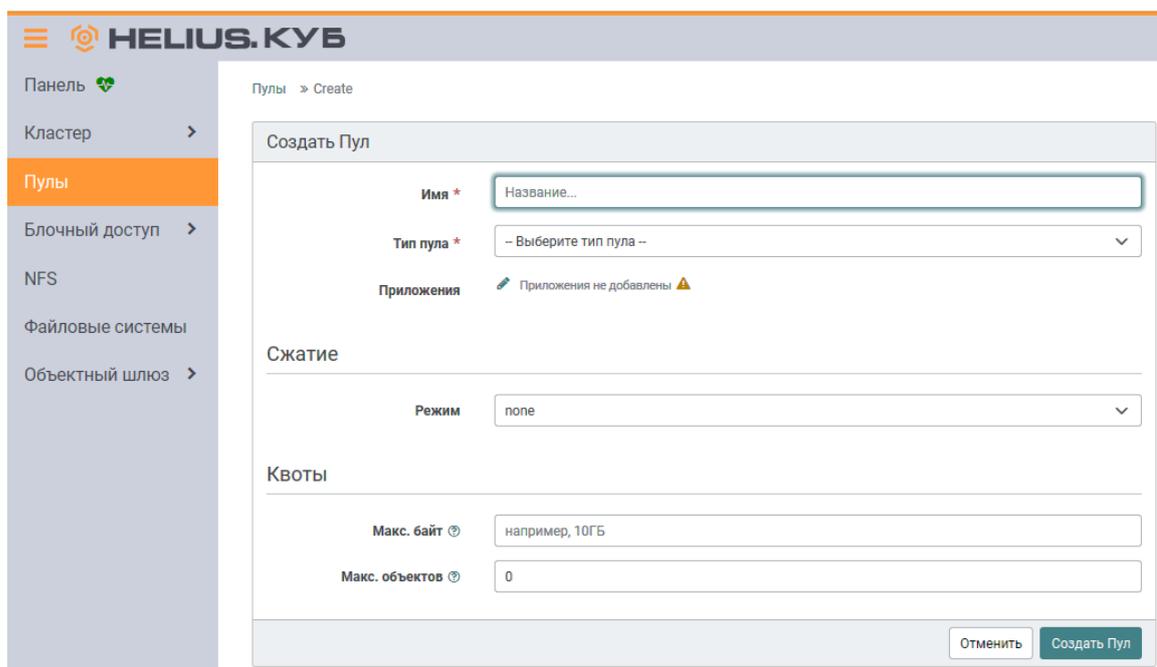
Для создания пула необходимо нажать «Создать» на панели инструментов. В открывшемся окне заполнить предлагаемую форму (рис. 93).

В поле «Имя» указать название пула.

В поле «Тип пула» из выпадающего списка выбрать тип пула («erasure» или «replicated»). Тип пула определяет метод хранения данных (размер частей данных и количество их копий).

В ПО используется два метода хранения данных – репликация (replicated) и кодирование с восстановлением данных (erasure).

При использовании метода репликации (тип пула «replicated») создаются копии входящих данных – реплики, которые хранятся на разных хостах кластера.



Панель 

Кластер >

Пулы

Блочный доступ >

NFS

Файловые системы

Объектный шлюз >

Пулы > Create

Создать Пул

Имя *

Тип пула *

Приложения  Приложения не добавлены 

Сжатие

Режим

Квоты

Макс. байт 

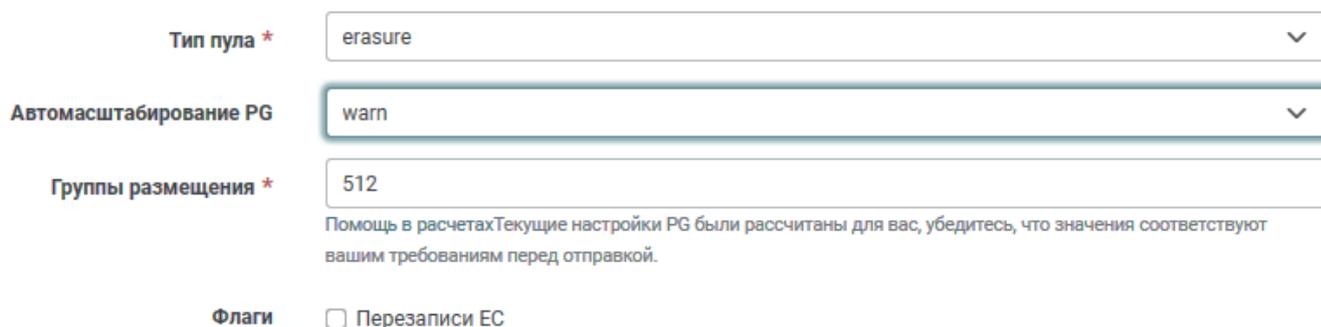
Макс. объектов 

Рисунок 93

При использовании метода кодирования с восстановлением данных (тип пула «erasure») входящие данные разбиваются на k частей одинакового размера (фрагменты данных). Дополнительно создаётся m частей того же размера для восстановления данных (фрагменты кодирования). Все части распределяются среди $k+m$ хостов кластера – одна часть на хост кластера. Кластер будет поддерживать работоспособность и сохранность данных, пока количество отказавших хостов меньше или равно m .

Примечание. При выборе типа пула появятся дополнительные поля для заполнения.

В поле «Автомасштабирование PG» из выпадающего списка выбрать режим автомасштабирования PG (рис. 94):



Тип пула *

Автомасштабирование PG

Группы размещения *

Помощь в расчетах Текущие настройки PG были рассчитаны для вас, убедитесь, что значения соответствуют вашим требованиям перед отправкой.

Флаги Перезаписи ЕС

Рисунок 94

- off – отключает автоматическое масштабирование;
- on – включает автоматическую корректировку количества групп размещения (PG);

– warn – выдаёт оповещение о состоянии «HEALTH_WARN», когда необходимо скорректировать количество групп размещения (PG).

В поле «Группы размещения» указать количество групп размещения (PG) (поле доступно при выборе «off» или «warn» в поле «Автомасштабирование PG»). При выборе «Помощь в расчётах» калькулятор PG рассчитает требуемое количество групп размещения.

Примечание. Количество PG влияет на производительность. Если в пуле слишком мало PG, то ПО будет перемещать большой процент данных одновременно, а нагрузка на сеть отрицательно скажется на производительности кластера. Если в пуле слишком много PG, то ПО будет использовать слишком много ресурсов (CPU и RAM) при перемещении небольшого процента данных, что также отрицательно скажется на производительности кластера.

По умолчанию пулы типа «erasure» работают только с объектным шлюзом ПО, который выполняет полную запись и добавление объектов. Выбор опции «Перезаписи ЕС» (опция доступна при выборе типа пула «erasure») позволяет блочным устройствам и файловым системам ПО хранить свои данные в пуле типа «erasure», при этом метаданные рекомендуется хранить в пуле типа «replicated».

Примечание. Создание пулов типа «erasure» с выбором опции «Перезаписи ЕС» доступно только в пуле, использующем OSD BlueStore.

При выборе типа пула «replicated» в поле «Размер репликации» указать размер репликации (по умолчанию – 3, минимальное значение – 2) (рис. 95).

Примечание. Размер репликации включает и сам объект. Если вы установите значение 2, то будет только одна копия ваших данных. Если вы потеряете один экземпляр объекта, вам нужно будет убедиться, что другая копия не была повреждена. Если в пуле установлена одна репликация, это означает, что в пуле есть ровно один экземпляр объекта данных. В случае сбоя OSD вы потеряете данные. Возможное применение пула с одной репликой — хранение временных данных в течение короткого времени.

| | |
|------------------------|--|
| Тип пула * | <input type="text" value="replicated"/> |
| Автомасштабирование PG | <input type="text" value="warn"/> |
| Группы размещения * | <input type="text" value="1024"/> <small>Помощь в расчетах Текущие настройки PG были рассчитаны для вас, убедитесь, что значения соответствуют вашим требованиям перед отправкой.</small> |
| Размер репликации * | <input type="text" value="3"/> |

Рисунок 95

В поле «Приложения» добавить приложения.

Далее необходимо заполнить поля на панели «CRUSH»:

– при создании пула типа «erasure» необходимо использовать профиль кодирования с восстановлением данных (рис. 96);

CRUSH

Профиль кодирования с восстановлением данных: default

Набор правил CRUSH: Новый набор правил CRUSH будет создан неявно.

Рисунок 96

– при создании пула типа «replicated» необходимо использовать набор правил CRUSH (рис. 97).

CRUSH

Набор правил CRUSH: - Выберите правило CRUSH -

Рисунок 97

В поле «Профиль кодирования с восстановлением данных» из выпадающего списка выбрать профиль кодирования (поле доступно при выборе типа пула «erasure»).

Для просмотра информации о доступных профилях кодирования с восстановлением данных (профиль EC) необходимо нажать «?». Информация будет представлена на вкладках «Профиль» и «Используется пулами» (рис. 98).

CRUSH

| | |
|-----------|--------------|
| k | 2 |
| m | 2 |
| plugin | jerasure |
| technique | reed_sol_van |

Рисунок 98

ПО создаёт профиль EC по умолчанию при инициализации кластера и обеспечивает тот же уровень избыточности, что и две копии в пуле с репликацией. Этот профиль по умолчанию определяет $k=2$ и $m=2$, что означает, что ПО распределяет данные объекта по четырём OSD ($k+m=4$), и ПО может потерять одно из этих OSD без потери данных. Такой профиль EC требует минимального объёма развёртывания в 4 хоста (рекомендуется 5 хостов) и может справиться с временной потерей 1 хоста OSD.

При необходимости можно создать новый профиль ЕС (см. подробнее 7.1.1) или удалить неиспользуемый.

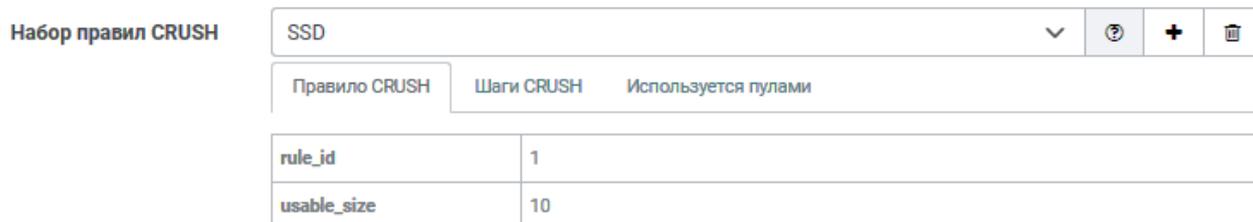
Примечание. Важно правильно выбрать профиль ЕС, так как его нельзя изменить после создания пула. Чтобы изменить профиль ЕС, необходимо создать новый пул с другим профилем ЕС и перенести объекты из старого пула в новый.

При выборе типа пула «erasure» поле «Набор правил CRUSH» будет недоступно для заполнения, т.к. набор правил CRUSH будет создан неявно.

При выборе типа пула «replicated» в поле «Набор правил CRUSH» из выпадающего списка выбрать стратегии размещения и репликации или политики распределения, позволяющие указать, как алгоритм CRUSH размещает реплики данных.

Для просмотра информации о выбранном правиле CRUSH необходимо нажать «?». Информация будет представлена на вкладках «Правило CRUSH», «Шаги CRUSH» и «Используется пулами» (рис. 99).

CRUSH



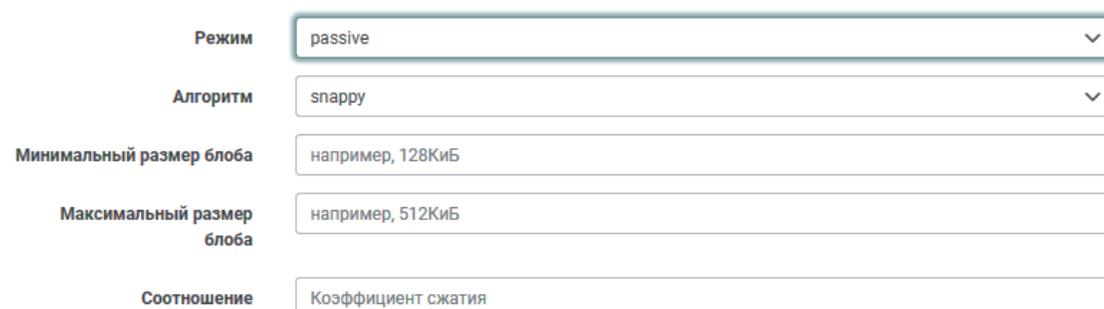
| Набор правил CRUSH | |
|--|----|
| SSD | |
| <div style="display: flex; justify-content: space-between; align-items: center;"> Правило CRUSH Шаги CRUSH Используется пулами </div> | |
| rule_id | 1 |
| usable_size | 10 |

Рисунок 99

При необходимости можно создать новое правило CRUSH (см. подробнее 7.1.2) или удалить неиспользуемое.

Далее необходимо заполнить поля на панели «Сжатие» (рис. 100).

Сжатие



| | |
|---------------------------|--------------------|
| Режим | passive |
| Алгоритм | snappy |
| Минимальный размер блока | например, 128КиБ |
| Максимальный размер блока | например, 512КиБ |
| Соотношение | Коэффициент сжатия |

Рисунок 100

В поле «Режим» из выпадающего списка выбрать политику встроенного алгоритма сжатия («none», «passive», «aggressive» или «force»).

В поле «Алгоритм» из выпадающего списка выбрать встроенный алгоритм сжатия («lz4», «snappy», «zlib» или «zstd»).

В поле «Минимальный размер BLOB» указать минимальный размер BLOB (фрагменты меньше указанного размера не будут подвергнуты сжатию).

В поле «Максимальный размер BLOB» указать максимальный размер BLOB (фрагменты, превышающие указанный размер, будут разбиты на более мелкие двоичные объекты перед сжатием данных).

В поле «Соотношение» указать коэффициент сжатия (отношение размера фрагмента данных после сжатия к исходному размеру).

Далее необходимо заполнить поля на панели «Квоты». При установке квот для пула можно ограничить максимальное количество объектов или максимальное количество байт, хранящихся в пуле (рис. 101).

Квоты



Рисунок 101

В поле «Макс. байт» указать максимальное количество байт, хранящихся в пуле. Чтобы отключить данную квоту необходимо оставить поле пустым или указать значение «0».

В поле «Макс. объектов» указать максимальное количество объектов, хранящихся в пуле. Чтобы отключить данную квоту необходимо оставить поле пустым или указать значение «0».

Нажать «Создать Пул», при этом созданный пул отобразится в общем списке.

7.1.1. Создание профиля кодирования с восстановлением данных

Создание профиля кодирования с восстановлением данных (профиль EC) доступно при создании пула типа «erasure» (см. подробнее 7.1.).

На панели «CRUSH» нажать «+». В открывшемся окне заполнить предлагаемую форму (рис. 102).

В поле «Имя» указать название профиля EC.

В поле «Плагин» из выпадающего списка выбрать плагин.

Плагин «jerasure» является наиболее универсальным и гибким и используется по умолчанию.

Создать Профиль ЕС
✕

| | |
|----------------------------|--|
| Имя | <input type="text" value="Name..."/> |
| Плагин * | <input type="text" value="jerasure"/> |
| Фрагменты данных (k) * | <input type="text" value="4"/> ✓ |
| Кодирующие фрагменты (m) * | <input type="text" value="2"/> ✓ |
| Домен отказа Crush | <input type="text" value="osd (24)"/> |
| Механизм | <input type="text" value="reed_sol_van"/> |
| Размер пакета | <input type="text" value="2048"/> |
| Корневой элемент Crush | <input type="text" value="default"/> |
| Класс устройства Crush | <input type="text" value="Позволить решить"/> <small>Доступные OSD: 24</small> |
| Директория | <input type="text" value="/usr/lib64/"/> |

Отменить
Создать Профиль ЕС

Рисунок 102

В поле «Фрагменты данных (k)» указать необходимое количество фрагментов данных. Каждый объект разделён на части (фрагменты данных), каждая из которых хранится на отдельном OSD.

В поле «Кодирующие фрагменты (m)» указать необходимое количество кодирующих фрагментов. Необходимо рассчитать кодирующие фрагменты (chunks) для каждого объекта и хранить их на разных OSD. Количество кодирующих фрагментов соответствует количеству OSD, которые могут выйти из строя без потери данных.

В поле «Домен отказа Crush» из выпадающего списка выбрать домен отказа CRUSH (хост или OSD). В качестве домена отказа рекомендуется выбрать OSD.

Примечание. Убедитесь, что ни один из двух блоков не находится в корзине с одной и той же зоной отказа. Например, если зоной отказа является хост, ни один из двух блоков не будет храниться на одном и том же хосте.

В поле «Механизм» из выпадающего списка выбрать технику, которая будет выбрана в рамках указанного плагина. Более гибкой техникой является reed_sol_van, т.к. достаточно установить k и m. Техника cauchy_good может быть более быстрой, но нужно тщательно выбирать размер пакета. Все техники reed_sol_r6_op, liberation, blaum_roth, liber8tion являются эквивалентами RAID6, так как они могут быть настроены только с m=2.

В поле «Размер пакета» указать размер пакета (в байтах). Кодирование будет проводиться пакетами определённого размера в байтах.

В поле «Корневой элемент Crush» из выпадающего списка выбрать имя корзины CRUSH, используемой на первом этапе правила CRUSH. Например, шаг CRUSH «take default».

В поле «Класс устройства Crush» из выпадающего списка выбрать класс устройства. Ограничьте размещение устройствами определённого класса (например, SSD или HDD), используя имена классов устройств в карте CRUSH. При выборе «Позволить решить» CRUSH использует все устройства независимо от класса.

В поле «Директория» установить имя каталога, из которого загружается плагин кодирования с восстановления данных.

Плагин «lrc» создаёт дополнительный локальный фрагмент чётности (1) или локальные группы, что обеспечивает восстановление с использованием меньшего количества не вышедших из строя OSD, при этом требуется больше ресурсов необработанного хранилища. При выборе плагина «lrc» необходимо заполнить дополнительные поля (рис. 103):

| | | |
|----------------------|---------------------------------------|---|
| Локальность (l) * ? | <input type="text" value="3"/> | ✓ |
| | Локальные группы: 2 | |
| Домен отказа Crush ? | <input type="text" value="osd (24)"/> | ▼ |
| Локальность Crush ? | <input type="text" value="Нет"/> | ▼ |

Рисунок 103

– в поле «Локальность (l)» указать размер локальной группы. Группируйте кодирующие фрагменты и фрагменты данных в группы определённого размера. Каждую локальную группу можно восстановить без чтения фрагментов из другой группы. Например, при $k=4$ и $m=2$, когда $l=3$, создаются две локальные группы, состоящие из трёх OSD. При потере одного OSD его можно восстановить с помощью трёх OSD (одной локальной группы) вместо шести;

– в поле «Локальность Crush» из выпадающего списка выбрать тип корзины CRUSH, в которой будет храниться каждая локальная группа (используется для создания шага правила CRUSH). Например, если выбран тип корзины «host», то локальные группы будут размещены на разных хостах. Если тип корзины не выбран, то группировка не производится.

При выборе плагина «shes» необходимо заполнить дополнительное поле «Оценщик надёжности (c)» (рис. 104):

Оценщик надёжности
(с) * ?

Рисунок 104

В поле «Оценщик надёжности (с)» указать количество фрагментов чётности, каждый из которых включает каждый фрагмент данных в своём диапазоне вычислений. Значение с используется как показатель надёжности, который приблизительно определяет количество OSD, которые могут быть отключены без потери данных. Например, если $s=2$, то 2 OSD могут выйти из строя без потери данных.

Плагин «slau» предназначен для обеспечения значительной экономии пропускной способности сети и ввода/вывода с диска при обслуживании (или ремонте), вышедшего из строя OSD. При выборе плагина «slau» необходимо заполнить дополнительные поля (рис. 105):

| | |
|--------------------------------------|--|
| Вспомогательные фрагменты (d) * ? | <input type="text" value="5"/> |
| | <small>D автоматически обновляется при изменениях k и m.</small> |
| Домен отказа Crush ? | <input type="text" value="osd (24)"/> |
| Скалярные mds ? | <input type="text" value="jerasure"/> |
| Механизм ? | <input type="text" value="reed_sol_van"/> |

Рисунок 105

– в поле «Вспомогательные фрагменты (d)» указать необходимое количество OSD, запрашиваемых для отправки данных при восстановлении одного фрагмента. Значение d должно быть выбрано таким образом, чтобы выполнялось условие $k+1 \leq d \leq k+m-1$. Чем больше значение d, тем больше экономия. Допускается устанавливать значение d вручную или использовать стандартный расчёт плагина, который максимизирует значение d;

– в поле «Скалярные mds» из выпадающего списка выбрать плагин (jerasure, isa, shes). «scalar_mds» используется в качестве стандартного блока в многоуровневой системе;

– в поле «Механизм» из выпадающего списка выбрать технику, которая будет выбрана в рамках плагина «scalar_mds». Поддерживаемые техники включают «reed_sol_van», «reed_sol_rb_or», «cauchy_orig», «cauchy_good», «liber8tion» для плагина jerasure (поле «Скалярные mds»), «reed_sol_van», «cauchy» для плагина isa (поле «Скалярные mds») и «single», «multiple» для плагина shes (поле «Скалярные mds»).

Нажать «Создать Профиль ЕС».

7.1.2. Создание нового правила CRUSH

Правила CRUSH и количество групп размещения являются основными факторами, определяющими, как будут размещаться данные.

Когда данные сохраняются в пуле, правила CRUSH, сопоставленные с пулом, позволяют алгоритму CRUSH идентифицировать правило для размещения каждого объекта и его реплик или фрагментов в кластере.

Создание нового правила CRUSH доступно при создании пула типа «replicated» (см. подробнее 7.1).

На панели «CRUSH» нажать «+». В открывшемся окне заполнить предлагаемую форму (рис. 106).

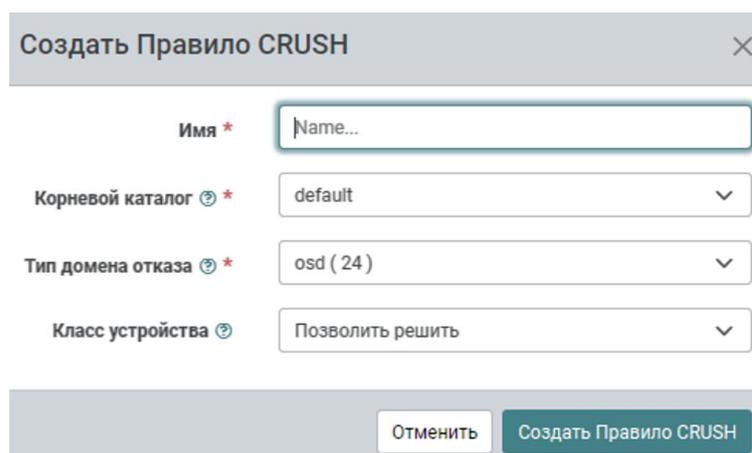


Рисунок 106

В поле «Имя» указать название правила.

В поле «Корневой каталог» из выпадающего списка выбрать имя хоста, на котором должны размещаться данные.

В поле «Тип домена отказа» из выпадающего списка выбрать тип узлов CRUSH (хост или OSD), между которыми следует распределить реплики.

В поле «Класс устройства» из выпадающего списка выбрать класс устройств (HDD или SDD), на котором должны размещаться данные. При выборе «Позволить решить» CRUSH использует все устройства независимо от класса.

Нажать «Создать Правило CRUSH».

Примечание. Администратор ПО не может изменить набор правил CRUSH после его создания.

7.2. Редактирование пула

Для редактирования пула необходимо выбрать пул и на панели инструментов нажать «Редактировать». В открывшемся окне обновить изменяемые свойства и нажать «Редактировать Пул».

Примечание. При переименовании пула необходимо обновить возможности (ограничения) аутентифицированного пользователя.

7.3. Просмотр информации о пуле

Для просмотра подробной информации о пуле необходимо развернуть содержимое строки. Подробная информация будет представлена на вкладках «Детали», «Подобности производительности» и «Конфигурация» (рис. 107).

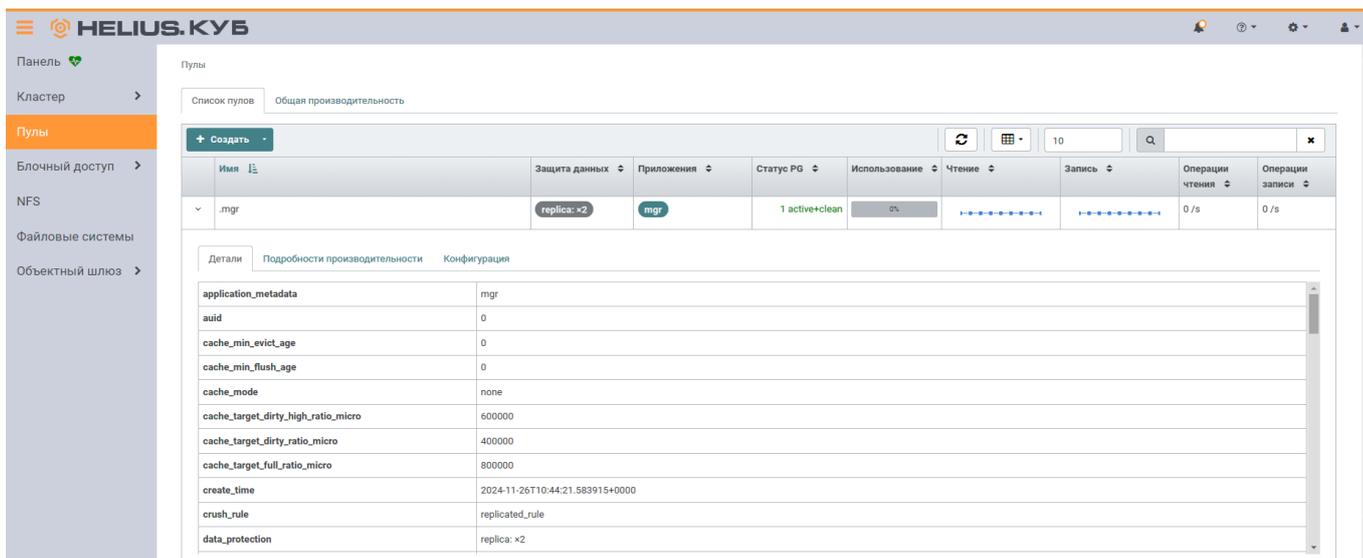


Рисунок 107

На вкладке «Детали» отображены метаданные пула.

На вкладке «Подобности производительности» отображена информация о производительности пула (рис. 108).

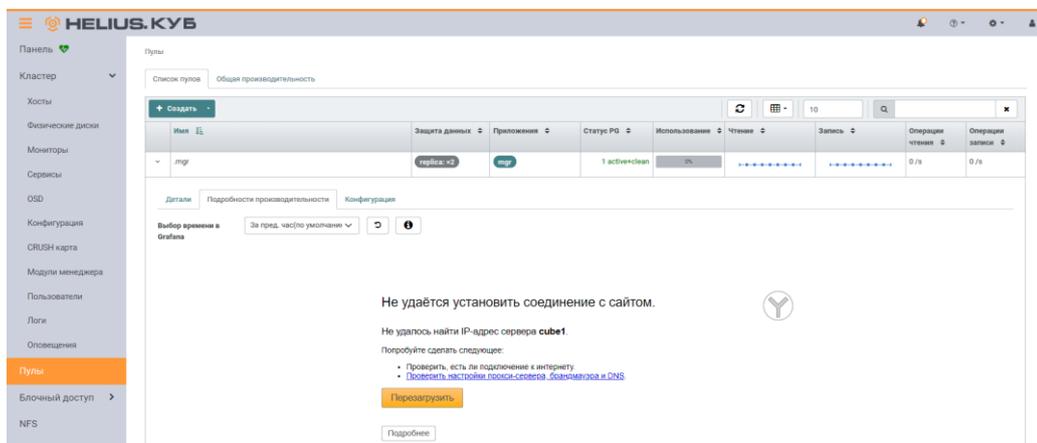


Рисунок 108

На вкладке «Конфигурация» отображена информация о конфигурации пула (рис. 109).

Детали Подробности производительности **Конфигурация**

| Имя | Описание | Ключ | Источник | Значение |
|---------------------------|--|-------------------------|------------|----------|
| Пиковый лимит BPS | Желаемый пиковый лимит байт ввода-вывода. | rbd_qos_bps_burst | Глобальное | 0 B/s |
| Лимит BPS | Желаемый лимит байт в секунду для операций ввода-вывода. | rbd_qos_bps_limit | Глобальное | 0 B/s |
| Пиковый лимит IOPS | Желаемый пиковый лимит операций ввода-вывода. | rbd_qos_iops_burst | Глобальное | 0 IOPS |
| Лимит IOPS | Желаемый лимит операций ввода-вывода в секунду. | rbd_qos_iops_limit | Глобальное | 0 IOPS |
| Пиковый лимит чтения BPS | Желаемый пиковый лимит чтения байт. | rbd_qos_read_bps_burst | Глобальное | 0 B/s |
| Лимит чтения BPS | Желаемый лимит чтения байт в секунду. | rbd_qos_read_bps_limit | Глобальное | 0 B/s |
| Пиковый лимит чтения IOPS | Желаемый пиковый лимит операций чтения. | rbd_qos_read_iops_burst | Глобальное | 0 IOPS |
| Лимит чтения IOPS | Желаемый лимит операций чтения в секунду. | rbd_qos_read_iops_limit | Глобальное | 0 IOPS |
| Пиковый лимит записи BPS | Желаемый пиковый лимит записи байт. | rbd_qos_write_bps_burst | Глобальное | 0 B/s |
| Лимит записи BPS | Желаемый лимит записи байт в секунду. | rbd_qos_write_bps_limit | Глобальное | 0 B/s |

12 всего

Рисунок 109

7.4. Удаление пула

Для удаления пула необходимо выбрать пул и в меню действий выбрать «Удалить». Подтвердить удаление в появившемся окне.

Примечание. При удалении пула рекомендуется также удалить правило CRUSH, созданное для данного пула, а также пользователей, которые его используют.

7.5. Общая производительность

Для просмотра общей производительности пула необходимо в разделе «Пулы» перейти на вкладку «Общая производительность» (рис. 110)

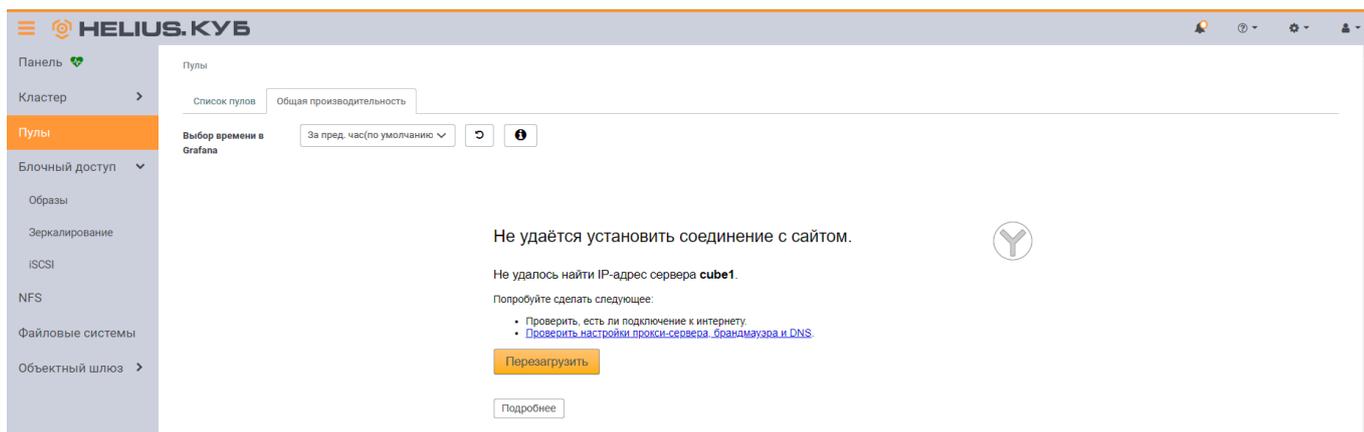


Рисунок 110

ПЕРЕЧЕНЬ СОКРАЩЕНИЙ

- ОС – операционная система
- ПО – программное обеспечение «Helius.КУБ»