

# ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ «HELIUS.КУБ»

ОПИСАНИЕ ПРОГРАММЫ

ЛИСТОВ 18

RU.ЕЦРТ.00057-01 13



## АННОТАЦИЯ

В настоящем документе приведено описание программного обеспечения «Helius.КУБ» RU.ЕЦРТ.00057-01 (далее – ПО).

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

Перечень терминов и сокращений приведён в конце документа.

## СОДЕРЖАНИЕ

1. ОБЩИЕ СВЕДЕНИЯ .....	4
1.1. Обозначение и наименование программы .....	4
1.2. Языки программирования.....	4
1.3. Используемые программные средства .....	4
2. ФУНКЦИОНАЛЬНОЕ НАЗНАЧЕНИЕ .....	5
2.1. Назначение программы.....	5
2.2. Функции программы .....	5
3. ОПИСАНИЕ ЛОГИЧЕСКОЙ СТРУКТУРЫ.....	9
3.1. Логическая структура ПО.....	9
3.2. Панель управления .....	11
4. ИСПОЛЬЗУЕМЫЕ ТЕХНИЧЕСКИЕ СРЕДСТВА .....	14
5. ЗАГРУЗКА И ЗАПУСК .....	15
ПЕРЕЧЕНЬ ТЕРМИНОВ.....	16
ПЕРЕЧЕНЬ СОКРАЩЕНИЙ .....	18

## 1. ОБЩИЕ СВЕДЕНИЯ

### 1.1. Обозначение и наименование программы

Обозначение программы: RU.ЕЦРТ.00057-01.

Наименование программы: Программное обеспечение «Helius.КУБ».

### 1.2. Языки программирования

Исходный язык программирования: Python3, DHTML, Javascript.

### 1.3. Используемые программные средства

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

## 2. ФУНКЦИОНАЛЬНОЕ НАЗНАЧЕНИЕ

### 2.1. Назначение программы

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.

### 2.2. 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 снапшотов 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 подключения и объединения внешних хранилищ данных по протоколам S3, AWS4, NFS, SMB/CIFS, iSCSI, FC.

### 3. ОПИСАНИЕ ЛОГИЧЕСКОЙ СТРУКТУРЫ

#### 3.1. Логическая структура ПО

Логическая структура ПО представлена на рисунке 1.

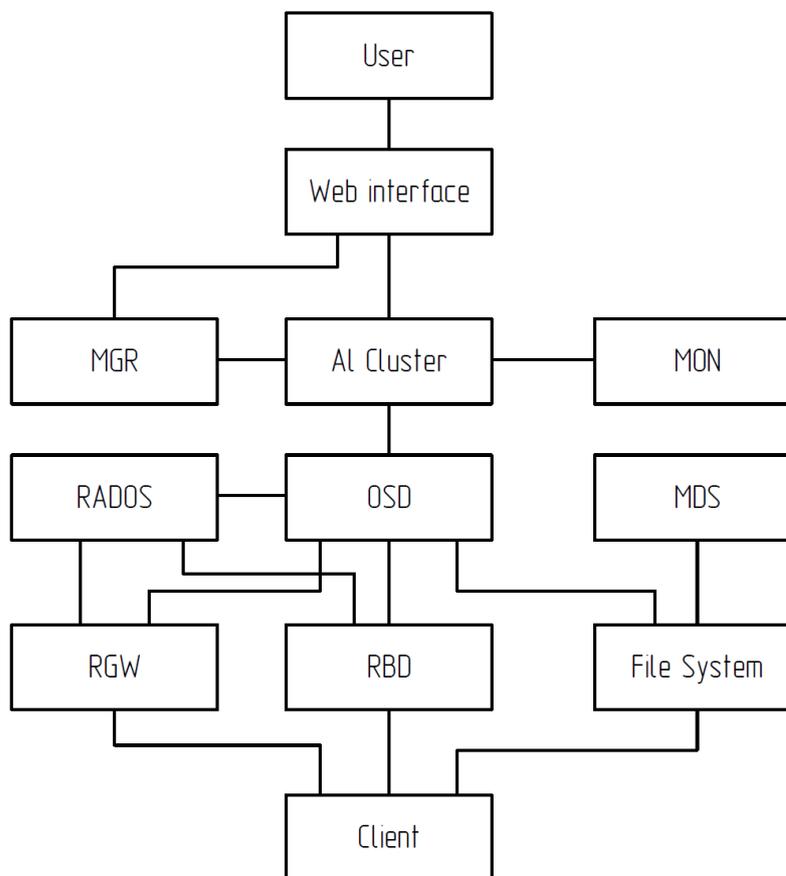


Рисунок 1

При установке на физические серверы ПО объединяет их в единый кластер, который можно легко масштабировать путём добавления дисков или хостов.

Кластер администрируется через панель управления (веб-интерфейс) и через интерфейс командной строки.

Кластерная система хранения данных ПО состоит из следующих служб:

– мониторы (MON) – элемент архитектуры ПО, который обеспечивает адресацию данных внутри кластера и хранит информацию о топологии, состоянии и распределении данных внутри хранилища. Мониторы отслеживают состояние всего кластера путём хранения карты состояния кластера, которая включает в себя карту монитора, карту PG, карту менеджера, карту OSD, карту MDS и карту CRUSH. Все хосты кластера сообщают хостам монитора

информацию обо всех изменениях в своих состояниях. Монитор поддерживает отдельную карту информации для каждого компонента;

– OSD – элемент хранилища, который хранит данные, обрабатывает запросы клиентов и предоставляет информацию для мониторинга ПО. OSD отвечает за управление отдельной единицей хранения, которая обычно представляет собой отдельный диск. Кластер состоит из множества OSD. OSD используют CPU, RAM и сетевые возможности хостов для выполнения репликации данных, балансировки, восстановления, мониторинга и функций отчётности;

– MDS – серверы метаданных, которые хранят метаданные для файловой системы Helius.KUBFS. Серверы метаданных позволяют пользователям Helius.KUBFS запускать базовые команды (такие как *ls*, *find* и т.д.), не создавая нагрузки на кластер. Они управляют тем, как пользовательские файлы разделяются на фрагменты, и местом хранения этих фрагментов, а также обеспечивают наличие достаточного количества реплик для фрагментов. Чтобы обеспечить высокую доступность метаданных, службы метаданных должны выполняться как минимум на трех хостах кластера. При отказе одной службы остающиеся две продолжают контролировать кластер;

– менеджеры (MGR) – сервис мониторинга, который отвечает за отслеживание показателей времени выполнения и текущего состояния кластера, включая использование хранилища, текущие показатели производительности и загрузку ПО. Менеджеры также содержат модули для управления информацией о кластере и предоставления её доступа, включая панель управления ПО и REST API.

ПО предоставляет следующие типы хранения данных:

– объектное хранилище – данные хранятся в виде объектов (файлов) в корзинах (S3) или контейнерах (Swift). Каждый объект имеет метаданные и уникальный идентификатор. Интерфейсом объектного хранилища является объектный шлюз (RGW);

– блочное хранилище – предоставляет доступ к данным на уровне блоков. Оно реализовано через блочное устройство (RBD), которое управляет хранением блочных данных, разбивая их на фрагменты. RADOS хранит эти фрагменты в виде объектов. Блочное устройство управляет хранением этих объектов в кластере хранения;

– файловая система – используется для хранения файлов и предоставляет стандартный интерфейс файловой системы.

ПО поддерживает несколько вариантов использования (типов клиентов):

– блочное устройство (RBD) – позволяет создавать виртуальные блочные устройства, доступные для использования в виртуальных машинах, контейнерах или физических серверах;

– объектный шлюз (RGW) – служба, взаимодействующая с кластером хранения. Объектный шлюз имеет собственную систему управления пользователями, т.к. предоставляет интерфейсы, совместимые с Swift/S3. Объектный шлюз может хранить данные в том же кластере хранения,

который используется для хранения данных от клиентов блочных устройств, но для этого потребуются отдельные пулы и другая иерархия распределения. API-интерфейсы S3 и Swift используют общее пространство имён, поэтому можно записывать данные с помощью одного API и извлекать их с помощью другого;

- файловая система (Helius.KUBFS);
- встроенный API.

Независимо от типа клиента (блочного устройства, объектного шлюза, файловой системы, встроенного API или командной строки) ПО хранит данные в виде объектов RADOS в логических пулах хранения. Клиенты взаимодействуют напрямую со всеми OSD для чтения и записи с использованием распределённого алгоритма хранения CRUSH.

Кластер хранения обладает возможностями самовосстановления: при отказе хоста или диска кластер автоматически пытается восстановить утерянные данные. Благодаря последовательному обновлению без перерывов в работе данные остаются доступными даже во время обновления хостов. В случае обслуживания хоста или применения исправлений рабочая нагрузка переносится на другие доступные хосты.

### 3.2. Панель управления

Панель управления – 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 (рис. 2). Она предоставляет обновления (каждые 5 секунд) в режиме реального времени о любых изменениях в кластере и обеспечивает быстрый доступ к другим разделам панели управления.

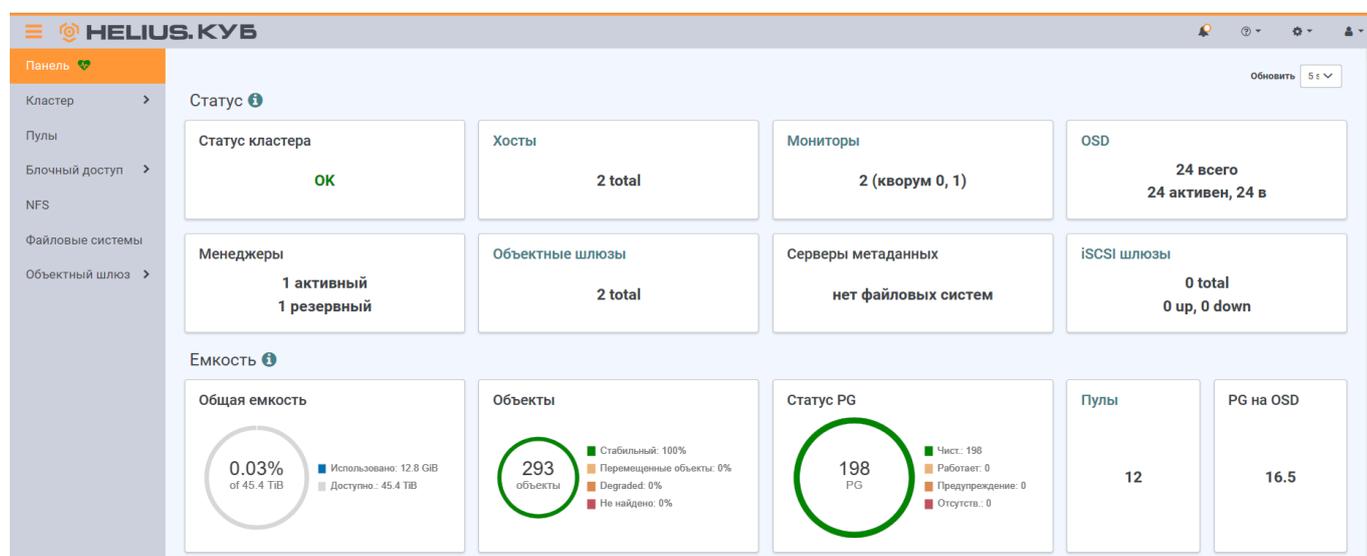


Рисунок 2

Панель управления ПО предлагает следующие возможности мониторинга и управления:

- общее состояние кластера – отображает показатели производительности и объёма использования, а также состояния кластера;
- встроенные панели мониторинга «Grafana» – отображают информацию и показатели производительности, собираемые модулем «Prometheus»;
- журналы (логи) кластера – отображают последние обновления файлов журналов (логов) событий кластера и аудита. Записи журнала (логи) можно фильтровать по приоритету, дате или ключевому слову;
- хосты – отображает список хостов кластера, а также подробную информацию о подключённых устройствах и запущенных службах. Доступно управление хостами;
- счётчики производительности – отображают подробную статистику по каждой запущенной службе;
- мониторы – отображает список мониторов, статус их кворума и открытые сессии;
- оповещения – отображает активные оповещения, позволяет настраивать и запускать оповещения, а также отключать уведомления о срабатывании оповещений;
- конфигурация – отображение (и редактирование при необходимости) всех доступных параметров конфигурации, их описаний, типов, значений по умолчанию и текущих настроек;
- пулы – отображает список пулов, информацию о них и их производительность. Доступно управление пулами;
- OSD – отображает список OSD, их состояние, статистику использования, гистограммы использования для операций чтения/записи, а также подробную информацию о подключённых устройствах (прогнозы работоспособности, карта OSD, метаданные, состояние устройства, счётчики и подробности производительности). Доступно управление OSD;
- физические диски – отображает список физических дисков, подключённых к хосту, и информацию о них. Доступна идентификация устройств (мигание LED);
- iSCSI – отображает RBD-образы и их характеристики производительности (операции чтения/записи, трафик), а также состояние iSCSI-шлюза и информацию об активных инициаторах;
- RBD-образы – отображает список RBD-образов и их свойства (размер, объекты, и т.д.), а также показатели общей производительности. Доступно управление RBD-образами, пространством имён и корзиной;
- зеркалирование RBD – отображает список служб и их состояние. Доступно управление зеркалированием RBD на удалённый сервер;

– файловые системы – отображают список активных клиентов файловой системы и связанных пулов, включая статистику использования. Доступно управление файловыми системами, их квотами и снапшотами;

– объектный шлюз – отображает список активных объектных шлюзов и их счётчики производительности. Доступно управление пользователями объектного шлюза и просмотр сведений о них (например, квоты), а также управление сегментами пользователей и просмотр сведений о них (например, цели размещения, владельцы, квоты, управление версиями, многофакторной аутентификацией);

– NFS – обеспечивает управление экспортом NFS файловых систем Helius.КубFS и сегментов RGW S3 с помощью NFS Ganesha;

– модули менеджера – включение и выключение модулей менеджера, управление параметрами конфигурации для конкретного модуля.

ПО использует систему мониторинга «Prometheus» для отслеживания производительности и доступности кластера хранилища, хостов инфраструктуры и развёрнутых сервисов.

## 4. ИСПОЛЬЗУЕМЫЕ ТЕХНИЧЕСКИЕ СРЕДСТВА

ПО функционирует на сервере архитектуры x64 под управлением ОС «РЕД ОС 8» (допускается использование другой ОС семейства Linux) со следующими минимальными техническими характеристиками:

- четырёхъядерный 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.

## 5. ЗАГРУЗКА И ЗАПУСК

Загрузка и запуск ПО осуществляется способом, деталиные сведения о котором изложены в руководстве администратора RU.ЕЦРТ.00057-01 36.

## ПЕРЕЧЕНЬ ТЕРМИНОВ

Зеркалирование RBD – процесс асинхронной репликации образов блочных устройств между двумя или более кластерами.

Карта состояния кластера – набор карт, состоящий из карты монитора, карты PG, карты MDS, карты OSD и карты CRUSH, которые вместе сообщают о состоянии кластера.

Клиент – клиентское приложение, которое взаимодействует с RADOS-хранилищем для выполнения операций чтения и записи данных.

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

Мониторы (MON) – служба, которая обеспечивает адресацию данных внутри кластера и хранит информацию о топологии, состоянии и распределении данных.

Пространство имён – некоторое множество, под которым подразумевается модель, абстрактное хранилище или окружение, созданное для логической группировки уникальных идентификаторов (то есть имён).

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

Сервер метаданных (MDS) – сервер метаданных, который хранит метаданные для файловой системы.

Хост – физический сервер, на котором установлено ПО.

Экспорт NFS (Network File System) – способ предоставления доступа к файловой системе по сети.

API (Application Programming Interface) – программный интерфейс приложения.

CRUSH – механизм распределения объектов на OSD.

Grafana – платформа с открытым исходным кодом для визуализации, мониторинга и анализа данных.

iSCSI (Internet Small Computer System Interface) – протокол на базе TCP/IP для взаимодействия и управления системами хранения данных, серверами и клиентами.

NFS-Ganesha – это сервер сетевой файловой системы (NFS) с открытым исходным кодом, который позволяет экспортировать файловые системы с использованием протокола NFS.

OSD (Object Storage Daemon) – устройство хранения объектов, которое хранит данные, обрабатывает запросы клиентов и предоставляет информацию для мониторинга ПО.

PG (Placement Group) – группы размещения.

Prometheus – набор инструментов с открытым исходным кодом, представляющий собой полноценную систему мониторинга оповещения.

RADOS (Reliable Autonomic Distributed Object Store) – распределённое объектное хранилище.

RBD (RADOS Block Device) – программное обеспечение с открытым исходным кодом для хранения данных на основе блочного устройства в распределённых системах хранения.

RBD-образы – виртуальные диски, которые хранятся в кластере и могут быть подключены к клиентам через протокол iSCSI или напрямую через библиотеку librbd.

RGW (RADOS Gateway) – вспомогательная служба, исполняющая роль прокси для поддерживаемых API объектных хранилищ.

## ПЕРЕЧЕНЬ СОКРАЩЕНИЙ

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