ООО «АИДАТех»
119021, Г.Москва, ул Льва Толстого, д. 2/22 стр. 6
ИНН 9704261020
КПП 770401001

**ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ «KAGECORE ML PLATFORM»**

# Руководство администратора

Листов 373

# Содержание

# Обозначения и сокращения

В настоящем документе применяют следующие сокращения и обозначения:

| | | |
|---|---|---|
| AD | - | Active Directory, служба каталога |
| API | - | Application Programming Interface, программный интерфейс взаимодействия |
| CA | - | Certificate Authority, доверенный орган, который выдает цифровые сертификаты |
| CLI | - | Command Line Interface, интерфейс командной строки |
| CN | - | Common Name, обычное название (атрибут с одним значением) |
| CNI | - | Container Networking Interface, сетевой интерфейс контейнера |
| CPU | - | Central Processing Unit, процессор |
| CRD | - | Custom Resource Definition, специальный ресурс, который позволяет создавать новые типы ресурсов для расширения функционала |
| CVE | - | Common Vulnerabilities and Exposures, база данных общеизвестных уязвимостей информационной безопасности |
| DB | - | Database, база данных |
| DC | - | Domain Controller, контроллер домена |
| DN | - | Distinguished Name, отличительное имя |
| DNS | - | Domain Name System, система доменных имен |
| DPI | - | Deep Packet Inspection, технология глубокой фильтрации сетевых пакетов, которая позволяет анализировать не только заголовки пакетов, но и их содержимое, то есть данные, которые передаются внутри |
| DWH | - | Data Warehouse, хранилище, предназначенное для сбора и аналитической обработки исторических данных организации |
| EOF | - | End of File, конец файла |
| FQDN | - | Fully Qualified Domain Name, полное доменное имя |
| GB | - | Gigabyte, гигабайт |
| GPU | - | Graphics Processing Unit, графический процессор |
| HA | - | High Availability, высокая доступность |
| HTTP | - | HyperText Transfer Protocol, протокол передачи гипертекста |

| | | |
|---|---|---|
| IAM | - | Identity and Access Management, управление идентификацией и контролем доступа |
| ID | - | Identifier, идентификатор |
| IOPS | - | Input/Output Operations Per Second, количество операций ввода-вывода в секунду |
| IP | - | Internet Protocol, межсетевой протокол |
| IPAM | - | IP Address Management, служба управления IP-адресами |
| IT | - | Information Technology, информационные технологии |
| JSON | - | JavaScript Object Notation, текстовый формат обмена данными, основанный на JavaScript |
| KSM | - | Kubernetes State Metrics |
| LDAP | - | Lightweight Directory Access Protocol, легковесный протокол доступа к каталогам |
| LDAPS | - | LDAP over SSL, легковесный протокол доступа к каталогам с поддержкой шифрования SSL/TLS |
| MAC-адрес | - | Media Access Control address, уникальный идентификатор сетевого оборудования, который назначается каждому устройству при его производстве |
| MB | - | Mbyte, мегабайт |
| ML | - | Machine Learning, машинное обучение |
| NAT | - | Network Address Translation, технология, которая позволяет нескольким устройствам в локальной сети использовать один публичный IP-адрес для доступа в интернет |
| NFS | - | Network File System, протокол сетевого доступа к файловым системам |
| NTP | - | Network Time Protocol, протокол сетевого времени |
| NUMA | - | Non-Uniform Memory Access, неоднородный доступ к памяти |
| NVD | - | National Vulnerability Database, Национальная база данных уязвимостей |
| OIDC | - | OpenID Connect, безопасный механизм, позволяющий приложению связаться со службой идентификации, чтобы получить необходимые данные о пользователе и вернуть их обратно в приложение, обеспечив полную защиту данных |

| | | |
|---|---|---|
| OS | - | Operating System, операционная система |
| OU | - | Organization Unit, организационная единица |
| PAAS | - | Platform as a Service, платформа как услуга, модель предоставления облачных вычислений, при которой потребитель получает доступ к использованию информационно-технологических платформ: операционных систем, систем управления базами данных, связующему программному обеспечению, средствам разработки и тестирования, размещённым у провайдера |
| PKI | - | Public Key Infrastructure, инфраструктура открытых ключей |
| QCOW | - | QEMU Copy On Write, формат используемый QEMU/KVM для хранения образов виртуальных машин |
| RADIUS | - | Remote Authentication Dial-In User Service, расширенный протокол удаленной аутентификации пользователей |
| RAM | - | Random Access Memory, оперативная память |
| RBAC | - | Role-Based Access Control, управление доступом на основе ролей |
| RDMA | - | Remote Direct Memory Access, удаленный прямой доступ к памяти |
| RHEL | - | Red Hat Enterprise Linux, коммерческий дистрибутив Linux, разработанный компанией Red Hat |
| RSA | - | Rivest–Shamir–Adleman, криптографический алгоритм с открытым ключом, используемый для безопасной передачи данных и цифровой подписи |
| SAML | - | Security Assertion Markup Language, язык разметки декларации безопасности |
| SDN | - | Software-Defined Networking, программно-определяемая сеть |
| SIEM | - | Security Information and Event Management, система управления информацией и событиями безопасности |
| SMTP | - | Simple Mail Transfer Protocol, открытый протокол обмена сообщениями на транспортном уровне |
| SNAT | - | Source Network Address Translation, преобразование сетевого адреса источника |
| SSH | - | Secure Shell, безопасная оболочка |
| SSL | - | Secure Sockets Layer, уровень защищенных сокетов |
| STP | - | Spanning Tree Protocol, протокол остовного дерева |

| | | |
|---|---|---|
| TCP | - | Transmission Control Protocol, протокол управления передачей |
| TLS | - | Transport Layer Security, протокол защиты транспортного уровня |
| TSC | - | Time Stamp Counter, счетчик метки времени |
| TTL | - | Time To Live, предельный период времени или число итераций или переходов, которые набор данных (пакет) может осуществить (прожить) до своего исчезновения |
| UDP | - | User Datagram Protocol, протокол пользовательских датаграмм |
| UI | - | User Interface, пользовательский интерфейс |
| UID | - | User Identifier, идентификатор пользователя |
| URL | - | Uniform Resource Locator, унифицированный указатель ресурса |
| VLAN | - | Virtual Local Area Network, виртуальная локальная сеть |
| VM | - | Virtual Machine, виртуальная машина |
| YAML | - | YAML Ain't Markup Language, специальный язык для структурированной записи информации, обладающий простым синтаксисом |
| БД | - | база данных |
| ВМ | - | виртуальная машина |
| ГБ | - | гигабайт |
| ИТ | - | информационные технологии |
| МБ | - | мегабайт |
| ОЗУ | - | оперативное запоминающее устройство |
| ОС | - | операционная система |
| ПО | - | программное обеспечение |
| ТП | - | тарифный план |
| ТПО | - | тарифный план организации |
| ФСТЭК | - | Федеральная служба по техническому и экспортному контролю |
| ЦОД | - | центр обработки данных |
| ЦП | - | центральный процессор |

# Термины и определения

В настоящем документе применяют следующие термины с соответствующими определениями:

| | | |
|---|---|---|
| ALD Pro | - | это набор сетевых служб сервера Astra Linux для создания службы каталога и организации централизованного управления ИТ-инфраструктурой. Продукт построен на хорошо известных компонентах с открытым исходным кодом, которые используют только открытые протоколы для обмена информацией |
| CI/CD | - | методология разработки программного обеспечения, направленная на автоматизацию процессов интеграции, тестирования и развертывания |
| REST API | - | архитектурный стиль, определяющий правила взаимодействия между клиентом и сервером, часто используемый в веб-разработке для обмена данными |
| кластер | - | это логическая группа хостов с общими доменами хранения и ЦП одного типа (Intel или AMD). Если модели ЦП хостов относятся к разным поколениям, то используются только те функции, которые присутствуют во всех моделях. Виртуальные машины динамически распределяются между хостами кластера и могут перемещаться между ними в соответствии с политиками, заданными в кластере, и настройками виртуальных машин. Кластер является самым высоким уровнем, на котором могут определяться политики электропитания и разделения нагрузки |
| контейнер | - | легковесные запускаемые образы, в состав которых входит некоторое ПО и его зависимости. Поскольку в контейнерах виртуализируется операционная система, вы можете запускать контейнеры одинаково в любом совместимом окружении |

# Введение

Платформа KageCore ML Platform, включая модули KageCore ML Platform. Модуль тарификации, KageCore ML Platform. Модуль витрины сервисов и KageCore ML Platform. Модуль пользовательского мониторинга (далее – ПО, Система), предназначена для предоставления высокопроизводительных вычислительных ресурсов и совокупности сервисов (IaaS/PaaS/SaaS) в интересах одной или нескольких организаций и проектов, позволяя эффективнее обучать и эксплуатировать модели искусственного интеллекта, оптимизировать производственные и научно-исследовательские процессы и тем самым способствовать ускоренному развитию технологий ИИ.

ПО предназначено для решения следующих задач:

– формирование единого пула аппаратных средств (серверы с CPU и GPU, высокоскоростные сети, системы хранения), доступного пользователям на основе квотирования, ролевой модели и механизма биллинга;

– автоматизация процесса заказа ресурсов (vCPU, RAM, GPU, объём дисков) через портал самообслуживания и программный интерфейс (API);

– использование готовых шаблонов (маркетплейса) с преднастроенными библиотеками и фреймворками (TensorFlow, PyTorch, Scikit-learn, Keras), а также средствами разработки JupyterLab, VSCode;

– использование механизмов контейнеризации и виртуализации для гибкой оркестрации и оперативного масштабирования ML-заданий;

– предоставление инфраструктурных сервисов (виртуальные машины) и платформенных (среды разработки, БД, аналитические инструменты) в унифицированном виде;

– разграничение прав и ресурсов на уровне отдельных организаций, проектов и групп, что позволяет параллельно вести несколько сценариев инференса и обучения;

– внедрение роли и квот (RBAC), позволяющих ограничивать доступ и лимитировать объём ресурсов (CPU, GPU, память, хранилище);

– учёт и детальный биллинг (включая CPU, GPU, хранение данных), обеспечивающие прозрачность и справедливое распределение затрат между участниками;

– поддержка экспериментов и трассировки (логирование метрик, параметров, артефактов) с целью воспроизводимости и контроля качества обучаемых моделей;

– предоставление пользователям возможности оперативного выбора и автоконфигурации необходимых сервисов из унифицированного каталога (marketplace), содержащего преднастроенные модули, которые охватывают базовые и

прикладные сервисы, а также автоматизированного развертывания в контейнерной или виртуальной среде без ручной настройки;

– предоставление администраторам и пользователям платформы наблюдаемость за всеми компонентами платформы, от процессора, графического ускорителя и памяти до приложения. Предоставление возможности сбора любых доступных метрик и их отображения, а также сбора анализа журналов приложений, операционных систем при необходимости.

# 1 Компоненты Системы

## 1.1 Основные компоненты Системы

ПО KageCore ML Platform состоит из следующих основных составных частей:

– система управления виртуализацией (далее – СУВ);

– система контейнеризации и оркестрации (далее – СКО);

– система управление вычислительными ресурсами и обеспечивающая масштабирование уровня IaaS/PaaS/SaaS (далее – СУВР);

– инфраструктурный мониторинг;

– система машинного обучения.

В состав ПО также входят самостоятельные модули:

1) KageCore ML Platform. Модуль витрины сервисов.
2) KageCore ML Platform. Модуль тарификации.
3) KageCore ML Platform. Модуль пользовательского мониторинга.

# 2 Система управления виртуализацией

## 2.1 Администрирование и обслуживание платформы виртуализации

Для поддержания работы системы управления виртуализацией требуется администратор. В задачи администратора входит следующее:

1) Управление физическими и виртуальными ресурсами: хостами и виртуальными машинами. Сюда входит:
   – обновление и добавление хостов;
   – импортирование доменов;
   – преобразование виртуальных машин, созданных на внешних гипервизорах;
   – управление пулами виртуальных машин.
2) Мониторинг общих ресурсов системы на предмет потенциальных проблем:
   – чрезмерная нагрузка на один из хостов;
   – недостаточный объем памяти или дискового пространства;
   – выполнение требуемых действий (например, перенос виртуальных машин на другие хосты для уменьшения нагрузки или высвобождение ресурсов путем выключения машин).
3) Обеспечение соответствия виртуальных машин новым требованиям (например, обновление операционной системы или выделение большего объема памяти).
4) Управление свойствами настраиваемых объектов с помощью тегов.
5) Управление настройкой пользователей и задание уровней разрешений.
6) Поиск и устранение неполадок у определенных пользователей или виртуальных машин для обеспечения общей работоспособности системы.
7) Формирование общих отчетов и отчетов по отдельным срезам.

### 2.1.1 Глобальная конфигурация

Чтобы получить доступ к глобальной конфигурации откройте Управление (Administration) > Настройка (Configure). В окне Настройка (Configure) можно настроить ряд глобальных ресурсов для системы управления виртуализацией платформы виртуализации, например, роли, системные разрешения, политики планирования, типы экземпляров и пулы MAC-адресов. В этом окне можно настроить способ взаимодействия пользователей с ресурсами в системе, а также окно служит центром для настройки параметров, применимых к нескольким кластерам.

2.1.1.1 Расширенные настройки отображения портала

Менеджер управления предоставляет возможность гибкой настройки интерфейса с учетом индивидуальных потребностей.

Настройка интерфейса осуществляется в окне Параметры аккаунта. Для доступа к окну на Портале администрирования нажмите в боковой панели Управление (Administration) > Параметры аккаунта (Account settings) или в верхней панели > Параметры аккаунта (Account settings).

2.1.1.2 Роли

Роли – это заранее заданные наборы прав, которые можно конфигурировать в Менеджере управления. Роли предоставляют разрешения на доступ к различным уровням ресурсов в центре данных и к конкретным физическим и виртуальным ресурсам и управление ими.

Благодаря многоуровневому администрированию любые разрешения, применимые к какому-либо объекту в контейнере, также применимы ко всем отдельным объектам в этом контейнере. Например, когда роль администратора хоста назначается пользователю на каком-либо конкретном хосте, пользователь получает разрешения на выполнение любых доступных операций на хосте, но только на назначенном хосте. Но если роль администратора хоста назначается пользователю в центре данных, то пользователь получает разрешения на выполнение операций с хостами на всех хостах в кластерах центра данных.

2.1.1.2.1 Создание новой роли

Если нужная роль отсутствует в списке ролей по умолчанию платформы виртуализации, то вы можете создать новую роль и настроить ее в соответствии с вашими целями.

Порядок действий:

1) Нажмите Управление (Administration) > Настройка (Configure). Откроется окно Настройка (Configure) . Вкладка Роли (Roles) выбрана по умолчанию и отображает список имеющихся по умолчанию ролей Пользователя и Администратора, а также любых дополнительно созданных ролей.

2) Нажмите Новая (New).

3) Введите имя и описание новой роли в поля Имя (Name) и Описание (Description).

4) Выберите Администратор (Admin) или Пользователь (User) в качестве Типа учетной записи (Account Type).

5) Нажимайте кнопки Развернуть все (Expand All) или Свернуть все (Collapse All), чтобы показать больше или меньше разрешений для объектов в списке Выберите

опции для разрешенных действий (Check Boxes to Allow Action). Вы также можете развернуть или свернуть список опций каждого объекта.

6) Для каждого объекта выберите или удалите действия, которые хотите разрешить или запретить для настраиваемой роли.

7) Нажмите OK, чтобы закрыть окно, сохранить изменения и новая роль появилась в списке ролей или нажмите Закрыть (Cancel), чтобы закрыть окно без создания новой роли.

2.1.1.2.2 Изменение или копирование роли

Вы можете изменять настройки созданной роли, но не сможете вносить изменения в роли, присутствующие по умолчанию. Для изменения ролей по умолчанию их необходимо клонировать и отредактировать в соответствии с потребностями.

Порядок действий:

1) Нажмите Управление (Administration) > Настройка (Configure). Откроется окно Настройка (Configure), в котором отображается список ролей Пользователя и Администратора по умолчанию, а также любых дополнительно созданных ролей.

2) Выберите роль, которую хотите изменить.

3) Нажмите Изменить (Edit) или Копировать (Copy). Откроется окно Изменить роль (Edit Role) или Копировать роль (Copy Role).

4) При необходимости измените имя и описание роли в полях Имя (Name) и Описание (Description).

5) Нажимайте кнопки Развернуть все (Expand All) или Свернуть все (Collapse All), чтобы показать больше или меньше разрешений для объектов в списке. Вы также можете развернуть или свернуть список опций каждого объекта.

6) Для каждого из объектов выберите или удалите действия, которые хотите разрешить или запретить для изменяемой роли.

7) Нажмите OK, чтобы закрыть окно и сохранить внесенные изменения или Закрыть (Close), чтобы закрыть окно без сохранения внесённых данных.

2.1.1.3 Системные разрешения

Разрешения позволяют пользователям осуществлять действия с объектами, где в качестве объектов могут выступать как отдельные объекты, так и их контейнеры. Любые разрешения в отношении контейнера также распространяются на все его содержимое.

2.1.1.3.1 Свойства пользователей

Роли и разрешения являются свойствами пользователя. С помощью ролей осуществляется многоуровневое администрирование, которое позволяет организовать дифференцированную иерархию разрешений.

Например, администратор центра данных может управлять всеми объектами назначенного центра данных, а администратор хоста — только конкретным назначенным хостом.

Один пользователь может иметь доступ к виртуальной машине, но не обладать правами на изменение её настроек. Другой пользователь может получить полные системные права для работы с виртуальной машиной.

2.1.1.3.2 Роли пользователей и администраторов

В платформе виртуализации предусмотрено множество заранее сконфигурированных ролей: от администратора с разрешениями, распространяющимися на всю систему, до конечного пользователя с доступом к одной единственной виртуальной машине. Нельзя изменять или удалять роли по умолчанию, но можно клонировать их и вносить нужные изменения или настраивать новые нужные вам роли. Есть два типа ролей:

– Роль администратора: позволяет получать доступ к Порталу администрирования (Administration Portal) для управления физическими и виртуальными ресурсами. Роль администратора предоставляет разрешения на выполнение действий на пользовательском портале, но никак не влияет на то, что пользователь может там увидеть.

– Роль пользователя: позволяет получать доступ к Пользовательскому порталу (VM Portal) для управления и обращения к виртуальным машинам и шаблонам. То, что пользователь может увидеть на пользовательском портале, зависит от его роли. Разрешения, предоставленные пользователю с ролью администратора, отображаются в действиях, доступных для этого пользователя в пользовательском портале.

2.1.1.3.3 Описание административных ролей

В следующей таблице описаны основные роли администратора, которым разрешен доступ к ресурсам и их конфигурирование на портале администрирования.

Таблица 4 –Основные административные роли платформы виртуализации

| **Роль** | **Права** | **Примечания** |
|---|---|---|

| | | |
|---|---|---|
| **SuperUser** | Системный администратор среды платформы виртуализации | Обладает всеми разрешениями для всех объектов и уровней, может управлять всеми объектами во всех центрах данных |
| **ClusterAdmin** | Администратор кластера | Обладает административными разрешениями в отношении всех объектов в конкретном кластере |
| **DataCenterAdmin** | Администратор центра данных | Обладает административными разрешениями в отношении всех объектов в конкретном центре данных, за исключением хранилища |

2.1.1.3.4 Управление ролями для хостов

Будучи Суперпользователем (SuperUser), системный администратор управляет всеми аспектами портала администрирования. Другим пользователям могут быть назначены ограниченные административные роли. Ограниченные роли нужны, чтобы предоставить пользователю административные права, которые действуют только в отношении определенного ресурса. Например, роль администратора центра данных (DataCenterAdmin) предусматривает права администратора только для назначенного центра данных, за исключением хранилища для этого центра данных, а администратору кластера (ClusterAdmin) доступны права администратора только в отношении назначенного кластера.

Администратор хоста – это роль системного администратора только для определенного хоста. Это полезно в кластерах с несколькими хостами, где для каждого хоста требуется системный администратор.

Роль администратора хоста разрешает выполнять следующие действия:

– Изменять конфигурацию хоста.
– Настраивать логические сети.
– Удалять хост.

Системного администратора хоста также можно сменить, удалив существующего системного администратора и добавив нового.

2.1.1.3.4.1 Описание административных ролей для хостов

В приведенной ниже таблице описаны роли и права администратора, применимые к администрированию хоста.

Таблица 5 – Административные роли для хостов

| Роль | Права | Примечания |
|---|---|---|
| **HostAdmin** | Администратор хоста | Может конфигурировать, управлять и удалять конкретный хост. Может также выполнять связанные с сетью операции на конкретном хосте |

2.1.1.3.5 Управление ролями для доменов хранения

Будучи Суперпользователем (SuperUser), системный администратор управляет всеми аспектами портала администрирования. Другим пользователям могут быть назначены ограниченные административные роли. Ограниченные роли нужны, чтобы предоставить пользователю административные права, которые действуют только в отношении определенного ресурса. Например, роль администратора центра данных (DataCenterAdmin) предусматривает права администратора только для назначенного центра данных, за исключением хранилища для этого центра данных, а администратору кластера (ClusterAdmin) доступны права администратора только в отношении назначенного кластера.

Администратор хранилища – это роль системного администратора только для определенного домена хранения. Это полезно в центрах данных с несколькими доменами хранения, где для каждого домена хранения требуется системный администратор.

Роль администратора домена хранения разрешает выполнять следующие действия:

- Изменять конфигурацию домена хранения.
- Переводить домен хранения в режим обслуживания.
- Удалять домен хранения.

**Важно!** Вы можете назначать роли и разрешения только существующим пользователям.

Системного администратора домена хранения также можно сменить, удалив существующего системного администратора и добавив нового.

### 2.1.2 Экран мониторинга (Dashboard)

Экран мониторинга служит для обзора состояния системы платформы виртуализации и отображает сводную информацию о наличии и использовании ресурсов платформе виртуализации. Эта сводка может предупредить вас о проблеме и позволяет проанализировать проблемную область.

По умолчанию информация на экране мониторинга обновляется каждые 15 минут из Хранилища данных (Data Warehouse, DWH), каждые 15 секунд с помощью API менеджера управления или при каждом обновлении экрана мониторинга. Экран мониторинга обновляется, когда пользователь возвращается с другой страницы или обновляет экран вручную. Экран мониторинга не обновляется автоматически. Информацию об инвентарной карточке учета предоставляет API менеджера управления, а информацию об использовании ресурсов предоставляет Хранилище (DWH). Экран мониторинга реализован как компонент плагина пользовательского интерфейса, который автоматически устанавливается и обновляется вместе с менеджером управления.6

2.1.2.1 Предварительные условия

Для использования экрана мониторинга требуется установить и настроить Хранилище (DWH).

2.1.2.2 Глобальный перечень ресурсов

В верхней части экрана мониторинга представлен глобальный перечень ресурсов (Global Inventory) платформы виртуализации: центры данных, кластеры, хосты, домены хранения, виртуальные машины и события. Значки показывают статус каждого ресурса, а числа - количество каждого ресурса с этим статусом.

В заголовке отображается количество ресурсов данного типа, а под заголовком - их статус. По щелчку на заголовке ресурса откроется соответствующая страница в менеджере управления. Для кластеров (Clusters) статус всегда отображается как N/A.

### 2.1.3 Поиск

2.1.3.1 Как устроен поиск в системе управления виртуализацией

На портале администрирования можно управлять тысячами ресурсов, например, виртуальными машинами, хостами, пользователями и т.д. Для поиска введите поисковый запрос (в произвольной форме или по правилам синтаксиса) в строку поиска, расположенную на главной странице каждого ресурса. Поисковые запросы можно сохранять в закладках на будущее, чтобы не вводить конкретный запрос каждый раз заново. Регистр текста не учитывается при поиске.

2.1.3.2 Синтаксис и примеры поиска

Синтаксис поисковых запросов в ресурсах платформы виртуализации выглядит следующим образом:

```
result type: {criteria} [sortby sort_spec]
```

где result type: - не редактируемая визуальная часть, зависит от того, где находится администратор. Для Ресурсы (Compute) > Виртуальные машины (Virtual Machines) будет равна Vms:, для Ресурсы (Compute) > Хосты (Hosts) - Hosts: и т.д. Служит напоминанием о типе возвращаемого результата, в примерах ниже его вводить не надо.

При этом на некоторых страницах ниже поля ввода фильтра могут присутствовать кнопки интерфейса пользователя для облегчения построения поискового запроса, которые дополняют его не редактируемую часть. На рисунке ниже приведён пример запроса, который выведет диски-образы с содержимым типа Данные, хранящиеся в доменах хранения центра данных Default:

2.1.3.3 Автозаполнение поиска

Функция автозаполнения на портале администрирования помогает формировать правильные и эффективные поисковые запросы. По мере ввода каждой части поискового запроса под строкой поиска открывается раскрывающийся список вариантов для следующей части поиска. Можно выбрать вариант из списка и затем продолжить набирать/выбирать следующую часть поискового запроса или набрать запрос вручную без подсказок.

2.1.3.4 Разные типы результатов поиска

Типы результатов делают возможным поиск любого из перечисленных ниже типов ресурсов:

- Vms для списка виртуальных машин.
- Host для списка хостов.
- Pools для списка пулов.
- Template для списка шаблонов.
- Events для списка событий.
- Users для списка пользователей и групп пользователей.
- Cluster для списка кластеров.
- DataCenter для списка центров данных.
- Storage для списка доменов хранения.
- Disk для списка дисков.
- Volumes для списка томов Gluster.

Поскольку у каждого типа ресурса есть уникальный набор свойств и набор других типов ресурсов, с которыми он ассоциирован, для каждого типа поиска предусмотрен набор допустимых синтаксических комбинаций. Кроме того, с помощью функции автозаполнения можно легко создавать корректные запросы.

2.1.3.5 Критерии поиска

Критерии поиска можно указать в запросе после двоеточия. Синтаксис {criteria} выглядит следующим образом:

```
<prop><operator><value>
```

или

```
<obj-type><prop><operator><value>
```

2.1.3.5.1 Поиск для виртуальных машин

В таблице ниже описаны все возможности поиска для виртуальных машин.

**Важно!** На текущий момент параметры поиска не поддерживают свойства Network Label (Метка сети), Custom Emulated Machine (Пользовательская эмулируемая машина) и Custom CPU Type (Пользовательский тип ЦП).

Таблица 25 – Поиск для виртуальных машин

| Свойство (ресурса или типа ресурса) | Тип | Описание (пример) |
|---|---|---|
| **Hosts.hosts-prop** | Зависит от типа свойства | Свойство хостов, ассоциированных с виртуальной машиной |
| **Templates.templates-prop** | Зависит от типа свойства | Свойство шаблонов, ассоциированных с виртуальной машиной |
| **Events.events-prop** | Зависит от типа свойства | Свойство событий, ассоциированных с виртуальной машиной |
| **Users.users-prop** | Зависит от типа свойства | Свойство пользователей, ассоциированных с виртуальной машиной |
| **Storage.storage-prop** | Зависит от типа свойства | Свойство устройств хранения, ассоциированных с виртуальной машиной |
| **Vnic.vnic-prop** | Зависит от типа свойства | Свойство vNIC, ассоциированных с виртуальной машиной |
| **name** | String (Строка) | Имя виртуальной машины |
| **comment** | String (Строка) | Комментарий к виртуальной машине |

| Свойство (ресурса или типа ресурса) | Тип | Описание (пример) |
|---|---|---|
| **status** | List (Список) | Доступность виртуальной машины |
| **ip** | Integer (Целое число) | IP-адрес виртуальной машины |
| **on_host** | | |
| **fqdn** | | |
| **uptime** | Integer (Целое число) | Сколько минут виртуальная машина уже работает |
| **domain** | String (Строка) | Домен (обычно домен Active Directory), в котором сгруппированы эти машины |
| **os** | String (Строка) | Операционная система, выбранная при создании виртуальной машины |
| **creationdate** | Date (Дата) | Дата создания виртуальной машины |
| **address** | String (Строка) | Уникальное имя, определяющее виртуальную машину в сети |
| **cpu_usage** | Integer (Целое число) | Процент использования вычислительной мощности CPU |
| **mem_usage** | Integer (Целое число) | Процент использования оперативной памяти |
| **network_usage** | Integer (Целое число) | Процент использования сети |
| **memory** | Integer (Целое число) | Максимальная заданная память |
| **guaranteed_memory** | | |
| **migration_progress_percent** | | |
| **apps** | String (Строка) | Приложения, установленные в настоящий момент на виртуальной машине |

| Свойство (ресурса или типа ресурса) | Тип | Описание (пример) |
|---|---|---|
| **cluster** | List (Список) | Кластер, которому относится виртуальная машина |
| **pool** | List (Список) | Пул виртуальных машин, к которому относится виртуальная машина |
| **loggedinuser** | String (Строка) | Имя пользователя, авторизованного в настоящий момент на виртуальной машине |
| **tag** | List (Список) | Теги, к которым относится виртуальная машина |
| **datacenter** | String (Строка) | Центр данных, к которому относится виртуальная машина |
| **type** | List (Список) | Тип виртуальной машины (сервер или рабочая станция) |
| **quota** | String (Строка) | Имя квоты, ассоциированной с виртуальной машиной |
| **id** | | |
| **description** | String (Строка) | Ключевые слова или текстовое описание виртуальной машины, которые дополнительно были использованы при создании виртуальной машины |
| **architecture** | | |
| **custom_emulated_machine** | | |
| **custom_cpu_type** | | |
| **compatibility_level** | | |
| **custom_compatibility_level** | | |
| **created_by_user_id** | | |

| Свойство (ресурса или типа ресурса) | Тип | Описание (пример) |
|---|---|---|
| **next_run_configuration_exists** | Bool (Логическая константа) | Есть ли изменения конфигурации, внесенные в виртуальную машину и ожидающие подтверждения |
| **has_illegal_images** | | |
| **bios_type** | | |
| **k8s_namespace** | | |
| **sortby** | List (Список) | Сортирует результаты поиска по одному из свойств ресурса |
| **page** | Integer (Целое число) | Отображаемый номер страницы результатов |

2.1.3.6 Метки

2.1.3.6.1 Использование меток для пользовательской настройки взаимодействий с платформы виртуализации

После того, как платформа виртуализации установлена и сконфигурирована согласно вашим требованиям, вы можете, используя метки, настроить способ работы с ней. Метки позволяют организовывать системные ресурсы в группы или категории. Это полезно, когда в среде виртуализации существует множество объектов и администратор хочет сосредоточиться на определенном их наборе.

В этом разделе описывается, как создавать и изменять метки, назначать их хостам или виртуальным машинам и выполнять поиск, используя их в качестве критериев.

В соответствии с потребностями организации метки можно организовать в иерархическую структуру.

Чтобы создать, изменить или удалить метки на Портале администрирования, нажмите значок Метки (Tags) (  ) в верхней панели.

2.1.3.6.2 Создание метки

Создавайте метки, чтобы с их помощью фильтровать результаты поиска.

Порядок действий:

1) Нажмите значок Метки (Tags) (  ) в верхней панели.

2) Нажмите Добавить (Add), чтобы создать новую метку, либо выберите метку и нажмите Новый (New), чтобы создать подчиненную метку.

3) Введите Имя (Name) и Описание (Description) новой метки.

4) Нажмите OK.

2.1.3.6.3 Изменение метки

Вы можете изменить имя и описание метки.

Порядок действий:

1) Нажмите значок Метки (Tags) (  ) в верхней панели.

2) Выберите метку, которую вы хотите изменить, и нажмите Изменить (Edit).

3) При необходимости измените поля Имя (Name) и Описание (Description).

4) Нажмите OK.

2.1.3.6.4 Удаление метки

Когда метка больше не нужна, удалите её.

Порядок действий:

1) Нажмите значок Метки (Tags) (  ) в верхней панели.

2) Выберите метку, которую хотите удалить, и нажмите Удалить (Remove). Появится предупреждение о том, что удаление метки также удалит все подчиненные метки.

3) Нажмите OK.

Вы удалили метку и все её подчиненные метки. Метки также удаляется со всех объектов, которым она была назначена.

2.1.3.6.5 Добавление меток к объектам и удаление меток с объектов

Вы можете назначать метки хостам, виртуальным машинам и пользователям, а также удалять метки с них.

Порядок действий:

1) Выберите объект(ы), на которых вы хотите назначить/удалить метки.

2) Нажмите Дополнительные действия (More Actions), затем – Назначить метки (Assign Tags).

3) Установите флажок, чтобы назначить метку объекту, либо снимите флажок, чтобы снять назначение метки с объекта.

4) Нажмите OK.

Указанная метка теперь добавляется или удаляется как пользовательское свойство выбранного объекта(-ов).

2.1.3.6.6 Поиск объектов с помощью меток

Введите поисковый запрос, используя метку (tag) как свойство, а также желаемое значение или набор значений как критерий для поиска.

Объекты с метками, соответствующими указанным критериям, перечислены в списке результатов.

Важно! Если вы ищете объекты, используя метку (tag) в качестве свойства и оператор неравенства (!=), например, Host: Vms.tag!=server1, то список результатов не будет включать в себя объекты без тегов.

2.1.3.6.7 Пользовательская настройка хостов с помощью меток

Вы можете использовать метки для хранения информации о ваших хостах. Затем вы можете искать хосты по меткам. Дополнительные сведения о поиске см. в разделе Поиск.

Порядок действий:

1) Нажмите Ресурсы (Compute) > Хосты (Hosts) и выберите хост.
2) Нажмите Дополнительные действия (More Actions) (), затем – Назначить метки(Assign Tags).
3) Установите флажки напротив применимых меток.
4) Нажмите OK.

Вы добавили дополнительную информацию о вашем хосте в виде меток, и по этой информации его можно будет искать.

# 3 Система контейнеризации и оркестрации

## 3.1 Управление узлами платформы

### 3.1.1 Общие сведения о процессах управления узлами

В процессе эксплуатации кластерной платформе контейнеризации вы можете столкнуться с различными задачами по управлению узлами кластера, например:

– Расширение количества узлов.
– Изменение объема выделенных ресурсов.
– Переустановка компонентов платформы на узле кластера.
– Перевод узла в режим обслуживания.

Текущая конфигурация узлов кластера определена в манифесте infrastructures.config.k8s-platform.io/cluster, ознакомиться с которым вы можете, выполнив команду:

```
kubectl describe infrastructures.config.k8s-platform.io cluster
```

Пример

```
$ kubectl describe infrastructures.config.k8s-platform.io cluster

Name:         cluster
Namespace:
Labels:       <none>
Annotations:  <none>
API Version:  config.k8s-platform.io/v1alpha2
Kind:         Infrastructure
Metadata:
  Creation Timestamp:  2023-09-05T15:14:06Z
  Generation:          1
  Resource Version:    3263
  UID:                 bfd04986-dca3-4361-b669-6a7e0da4e861
Spec:
  Cluster Configuration:
    Dns Base Domain: apps.cls01.k8s.internal
  Cluster Nodes:
    Infra:
      Network Spec:
        Hostname: infra.k8s.internal
        Ip:       172.31.101.24
      State:      present
    Master:
      Network Spec:
```

```
        Hostname: master.k8s.internal
        Ip:       172.31.101.25
      State:      present
    Worker:
      Network Spec:
        Hostname: worker.k8s.internal
        Ip:       172.31.101.26
      State:      present
  Customer ID:  5d4a2c84b6e22623
  Infrastructure Provider:
    None:
  License Key:  a098f0aefdc021b643c2eb76c1cad0a8
  Version:      v5.0.1
Status:
  API Server URL:  https://172.31.100.122:6443
  Platform:        none
Events:            <none>
```

**Важно!** Изменение манифеста infrastructures.config.k8s-platform.io/cluster вручную с помощью kubectl, curl и подобных утилит не поддерживается. Взаимодействие с объектом осуществляется только с помощью утилиты k8s-ctl.

3.1.1.1 Горизонтальное масштабирование кластера

Горизонтальное масштабирование заключается в добавлении новых узлов в кластер Kubernetes и выполняется с помощью утилиты k8s-ctl.

3.1.1.1.1 Масштабирование кластера, установленного универсальным методом (UPI)

Необходимые условия

– Вы подготовили виртуальный или физический узел для добавления в кластер согласно разделу Подготовка к установке платформы контейнеризации.

– На вашем локальном компьютере установлена утилита k8s-ctl.

– У вас есть доступ к Kubernetes API с привилегиями администратора кластера (cluster-admin).

– У вас есть закрытый ключ SSH на вашем локальном компьютере, который нужно предоставить утилите k8s-ctl.

– У вас есть токен доступа к хранилищу секретов Vault с привилегиями root.

**Важно!** При подготовке нового узла используйте такие же имя пользователя и открытую часть ключа SSH, как и на остальных узлах кластера, которые использовались на подготовительном этапе.

Процедура

1) Запустите процесс масштабирования узлов кластера с помощью команды:

```
k8s-ctl scale --ssh-user <имя_пользователя> --ssh-key <закрытый ключ SSH>
```

В качестве аргументов --ssh-key и --ssh-user укажите информацию, использованную на этапе конфигурации ключевой пары SSH.

2) Далее для временного редактирования будет открыт файл конфигурации кластера в текстовом редакторе vi.

3) Добавьте новый узел в блок конфигурации ClusterNodes и сохраните изменения.

Пример

```
k8s-ctl scale --ssh-user k8s-installer --ssh-key id_rsa.pem
    spec:
    ...
        clusterNodes:
        ...
        worker: # Роль узла в кластере Kubernetes.
        - hostGroup: "worker" # Существующий узел в кластере Kubernetes.
          networkSpec:
            hostname: "worker01.k8s.internal"
            ip: "172.31.101.26"
          state: "present"
        ...
        - hostGroup: "worker" # Добавляемый узел в кластер Kubernetes.
          networkSpec:
            hostname: "worker02.k8s.internal"
            ip: "172.31.101.27"
          state: "present"
```

4) На запрос Enter Vault root token: введите токен доступа к хранилищу секретов Vault с привилегиями root, после чего начнется процесс добавления узла в кластер.

5) Дождитесь сообщения об успешном выполнении операции.

6) Проверьте состояние узлов кластера Kubernetes после успешного масштабирования согласно руководству.

3.1.1.1.2 Масштабирование кластера, установленного в среде платформы виртуализации (IPI)

Необходимые условия

1) На вашем локальном компьютере установлена утилита k8s-ctl.

2) У вас есть доступ к Kubernetes API с привилегиями администратора кластера (cluster-admin).

3) У вас есть закрытый ключ SSH на вашем локальном компьютере, который нужно предоставить утилите k8s-ctl.

4) У вас есть токен доступа к хранилищу секретов Vault с привилегиями root.

**Важно!** Если вы решили использовать новый шаблон для виртуальной машины, то при его подготовке используйте такие же имя пользователя и открытую часть ключа SSH, как и на остальных узлах кластера, которые использовались на подготовительном этапе.

Процедура

1) Запустите процесс масштабирования узлов кластера с помощью команды:

```
k8s-ctl scale --ssh-user <имя_пользователя> --ssh-key <закрытый ключ SSH>
```

В качестве аргументов --ssh-key и --ssh-user укажите информацию, использованную на этапе конфигурации ключевой пары SSH.

2) Далее для временного редактирования будет открыт файл конфигурации кластера в текстовом редакторе vi.

3) Добавьте новый узел в блок конфигурации ClusterNodes и сохраните изменения.

4) При необходимости вы можете добавить в блок hostGroup описание новой группы узлов, в которой можно указать идентификатор нового шаблона виртуальной машины.

5) На запрос Enter Vault root token: введите токен доступа к хранилищу секретов Vault с привилегиями root, после чего начнется процесс добавления узла в кластер.

6) В процессе масштабирования будет запрошено подтверждение на создание виртуальных машин и сетевых интерфейсов в среде виртуализации платформы виртуализации.

7) Дождитесь сообщения об успешном выполнении операции.

8) Проверьте состояние узлов кластера Kubernetes после успешного масштабирования согласно руководству.

### 3.1.2 Вертикальное масштабирование узлов кластера

Вертикальное масштабирование заключается в добавлении физических или виртуальных ресурсов к текущим узлам кластера Kubernetes. В зависимости от метода установки платформы вертикальное масштабирование может выполняться как с помощью утилиты k8s-ctl, так и вручную.

3.1.2.1 Масштабирование узлов кластера, установленного универсальным методом (UPI)

Перед проведением работ по изменению ресурсов узла кластера Kubernetes необходимо вывести данный узел в режим обслуживания.

Необходимые условия

– На вашем локальном компьютере установлена утилита k8s-ctl.

– У вас есть доступ к Kubernetes API с привилегиями администратора кластера (cluster-admin).

– У вас есть закрытый ключ SSH на вашем локальном компьютере, который нужно предоставить утилите k8s-ctl.

Процедура

1) Выведите список всех узлов кластера и определите имя узла, которому необходимо изменить количество выделенных ресурсов:

```
kubectl get nodes
```

2) Переведите выбранный узел в режим обслуживания согласно процедуре, описанной в разделе Перевод узла в режим обслуживания.

3) После успешного перевода узла в режим обслуживания вы можете выключить узел для проведения работ по масштабированию ресурсов.

4) После завершения работ по масштабированию включите узел и дождитесь его загрузки.

5) Проверьте, что узел стал доступен для размещения рабочих нагрузок, следуя процедуре возврата узла в эксплуатацию.

3.1.2.2 Масштабирование кластера, установленного в среде платформы виртуализации (IPI)

Необходимые условия

– На вашем локальном компьютере установлена утилита k8s-ctl.

– У вас есть доступ к Kubernetes API с привилегиями администратора кластера (cluster-admin).

– У вас есть закрытый ключ SSH на вашем локальном компьютере, который нужно предоставить утилите k8s-ctl.

Процедура

1) Запустите процесс масштабирования узлов кластера с помощью команды:

```
k8s-ctl scale --ssh-user <имя_пользователя> --ssh-key <закрытый ключ SSH>
```

**Важно!** В качестве аргументов --ssh-key и --ssh-user укажите информацию, использованную на этапе конфигурации ключевой пары SSH.

2) Далее для временного редактирования будет открыт файл конфигурации кластера в текстовом редакторе vi.

3) Измените количество выделенных ресурсов для групп узлов определенных в блоке hostGroup и сохраните изменения.

Информация

В среде виртуализации платформы виртуализации ресурсы для виртуальных машин будут добавлены с помощью горячего подключения (hot plug) при следующих условиях:

– Увеличение объема ОЗУ (memory) на значение не превышающее установленный лимит (maximumMemory).

– Увеличение количества виртуальных процессоров (cpuSockets) без изменений их характеристик (cpuCores, cpuThreads).

Уменьшение количества ресурсов, изменение характеристик виртуальных процессоров (cpuCores, cpuThreads), а также изменение лимита ОЗУ (maximumMemory) требует последующего перезапуска виртуальных машин вручную. Для выполнения корректного перезапуска узлов кластера действуйте согласно процедуре вертикального масштабирования кластера, установленного универсальным методом (UPI).

4) В процессе масштабирования будет запрошено подтверждение на изменение виртуальных машин в среде виртуализации платформы виртуализации.

5) Дождитесь сообщения об успешном выполнении операции.

6) Перейдите в веб-интерфейс среды виртуализации платформы виртуализации и убедитесь, что ресурсы виртуальных машин были изменены.

7) Запланируйте и выполните процедуру перезапуска узла в среде платформы виртуализации. Перезапуск узла рекомендуется выполнять вместе с переводом узла в режим обслуживания.

### 3.1.3 Удаление узлов кластера

Удаление узлов из кластера может быть выполнено в автоматическом режиме с помощью утилиты k8s-ctl.

Необходимые условия

– На вашем локальном компьютере установлена утилита k8s-ctl.

– У вас есть доступ к Kubernetes API с привилегиями администратора кластера (cluster-admin).

– У вас есть закрытый ключ SSH на вашем локальном компьютере, который нужно предоставить утилите k8s-ctl.

– У вас есть токен доступа к хранилищу секретов Vault с привилегиями root.

Процедура

1) Запустите процесс удаления узлов кластера с помощью команды:

```
k8s-ctl scale --ssh-user <имя_пользователя> --ssh-key <закрытый ключ SSH>
```

В качестве аргументов --ssh-key и --ssh-user укажите информацию, использованную на этапе конфигурации ключевой пары SSH.

2) Далее для временного редактирования будет открыт файл конфигурации кластера в текстовом редакторе vi.

3) Отметьте узлы выбранные для удаления, изменив значение ключа state с present на absent и сохраните изменения.

4) В процессе будет запрошено подтверждение на удаление виртуальной машины.

Данный этап выполняется только для кластеров Kubernetes, развернутых в инфраструктуре, подготавливаемой установщиком (IPI). Виртуальная машина будет полностью удалена из платформы виртуализации.

5) Дождитесь сообщения об успешном выполнении операции.

6) Проверьте состояние узлов кластера Kubernetes после успешного масштабирования согласно руководству.

### 3.1.4 Повторное добавление узлов в кластер

В случае необходимости повторного добавления узла в кластер Kubernetes, например, в ситуациях, когда на узле требуется полная переустановка компонентов платформы, вам необходимо сначала удалить узел из кластера, следуя процедуре удаления узлов, а затем выполнить повторное добавление узла, используя процедуру горизонтального масштабирования кластера. На данном узле будет выполнен полный сброс и переустановка всех компонентоплатформе контейнеризации.

### 3.1.5 Перевод узла в режим обслуживания

В процессе эксплуатации платформы контейнеризации может возникнуть необходимость перевода узла в режим обслуживания. Находясь в режиме обслуживания, узел исключается из очереди Kubernetes Scheduler и более не может размещать какие-либо вновь запланированные рабочие нагрузки. Любые существующие рабочие нагрузки будут эвакуированы и запущенны на других подходящий узлах кластера Kubernetes. Исключение составляют нагрузки, запущенные контроллером DaemonSet, перезапуск которых на других узлах невозможен.

Перевод узла кластера Kubernetes в режим обслуживания может быть выполнен с помощью утилиты kubectl согласно процедуре, описанной далее.

**Важно!** Объекты Pod, которые не управляются контроллерами ReplicaSet, DaemonSet, StatefulSet и Job, не могут быть эвакуированы с узла и будут удалены при переводе узла в режим обслуживания. Убедитесь, что в кластере Kubernetes есть достаточное количество узлов, отвечающих тем же критериям, что и переводимый в режим обслуживания узел. В противном случае эвакуированные объекты Pod будут находиться в состоянии `Pending` до тех пор, пока узел не будет возвращен в работу.

3.1.5.1 Перевод узла кластера в режим обслуживания

Для перевода узла в режим эксплуатации следуйте процедуре ниже.

Необходимые условия

– На вашем локальном компьютере установлена утилита k8s-ctl.

– У вас есть доступ к Kubernetes API с привилегиями администратора кластера (cluster-admin).

– У вас есть закрытый ключ SSH на вашем локальном компьютере, который нужно предоставить утилите k8s-ctl.

Процедура

1) Получите список всех узлов кластера и определите имя узла, который необходимо вывести в режим обслуживания:

```
kubectl get nodes
```

2) Переведите выбранный узел в состояние SchedulingDisabled, выполнив команду:

```
kubectl cordon <имя узла>
```

3) Выполните эвакуацию всех рабочих нагрузок с выбранного узла, выполнив команду:

```
kubectl drain <имя узла> --ignore-daemonsets
```

Информация

Для удаления с узла объектов Pod, запущенных без использования контроллеров ReplicaSet, DaemonSet, StatefulSet и Job, используйте kubectl drain с опцией --force.

4) После успешного выполнения команды kubectl drain вы можете выключить узел для проведения запланированных работ.

3.1.5.2 Возврат узла кластера в режим эксплуатации

Для возврата узла в режим эксплуатации следуйте процедуре ниже.

Необходимые условия

– На вашем локальном компьютере установлена утилита k8s-ctl.

– У вас есть доступ к Kubernetes API с привилегиями администратора кластера (cluster-admin).

– У вас есть закрытый ключ SSH на вашем локальном компьютере, который нужно предоставить утилите k8s-ctl.

Процедура

1) После завершения работ включите ранее выключенный узел и дождитесь его загрузки.

2) Проверьте, что узел стал доступен для размещения рабочих нагрузок (статус узла Ready) выполнив команду:

```
kubectl get nodes
```

3) Разрешите запуск рабочих нагрузок на выбранном узле, выполнив команду:

```
kubectl uncordon <имя узла>
```

4) Узел готов к эксплуатации в кластере Kubernetes.

### 3.1.6 Выключение платформы контейнеризации

В данном разделе описана процедура корректного выключения платформы контейнеризации.

3.1.6.1 Подготовительные действия для выключения платформы контейнеризации

– Выполните полное резервное копирование мастер-узлов перед выключением кластера Kubernetes.

**Важно!** Убедитесь, что резервная копия надежна сохранена на внешнем хранилище. В случае каких-либо проблем с платформой контейнеризации данная резервная копия поможет вам восстановить кластер в прежнее состояние.

– Если вы планируете отключение кластера Kubernetes на довольно продолжительное время, то проверьте срок действия TLS-сертификатов.

  – Для проверки срока действия сертификатов используйте раздел документации Проверка срока действия сертификатов платформы.
  – При необходимости обновления сертификатов воспользуйтесь разделом документации Обновление сертификатов платформы.

– (Опционально) Выполните резервное копирование данных пользовательских приложений предусмотренным вами методом.

3.1.6.2 Выключение узлов кластера

**Предварительные условия:**

- Вы выполнили полное резервное копирование мастер-узлов

- У вас есть доступ к кластеру с учетной записью, имеющей роль `cluster-admin` в Kubernetes.
- Вы установили утилиту `kubectl` для работы с Kubernetes.

*Процедура*

1. Установите на каждом узле кластера Kubernetes запрет на запуск новых нагрузок:

```
for node in $(kubectl get nodes -o jsonpath='{.items[*].metadata.name}'); do echo ${node} ; kubectl cordon ${node} ; done
```

*Пример*

```
for node in $(kubectl get nodes -o jsonpath='{.items[*].metadata.name}'); do echo ${node} ; kubectl cordon ${node} ; done

k8s-infra-1.k8s.internal
node/k8s-infra-1.k8s.internal cordoned
k8s-master-1.k8s.internal
node/k8s-master-1.k8s.internal cordoned
k8s-worker-1.k8s.internal
node/k8s-worker-1.k8s.internal cordoned
```

2. Остановите существующие нагрузки и освободите от них рабочие узлы кластера Kubernetes:

```
for node in $(kubectl get nodes -l node-role.kubernetes.io/worker -o jsonpath='{.items[*].metadata.name}'); do echo ${node} ; kubectl drain ${node} --delete-emptydir-data --ignore-daemonsets=true --timeout=15s ; done
```

*Пример*

```
for node in $(kubectl get nodes -l node-role.kubernetes.io/worker -o jsonpath='{.items[*].metadata.name}'); do echo ${node} ; kubectl drain ${node} --delete-emptydir-data --ignore-daemonsets=true --timeout=15s ; done

k8s-worker-1.k8s.internal
node/k8s-worker-1.k8s.internal already cordoned
Warning: ignoring DaemonSet-managed Pods: kube-system/cilium-bhndt, kube-system/kube-proxy-4c5t6, k8s-csi-drivers/k8s-oauth-csi-provider-s4zk5, k8s-csi-drivers/k8s-secrets-store-csi-driver-t65px, k8s-ingress-public/k8s-ingress-public-controller-lzjhk, k8s-monitoring/k8s-cadvisor-s2pc4, k8s-monitoring/k8s-prometheus-node-exporter-nqhmq, k8s-secrets-webhook/k8s-oauth-secrets-webhook-dzvlk
node/k8s-worker-1.k8s.internal drained
```

3. Остановите существующие нагрузки и освободите от них инфраструктурные узлы кластера Kubernetes:

```
for node in $(kubectl get nodes -l node-role.kubernetes.io/infra -o jsonpath='{.items[*].metadata.name}'); do echo ${node} ; kubectl drain ${node} --delete-emptydir-data --ignore-daemonsets=true --timeout=15s ; done
```

*Пример*

```
for node in $(kubectl get nodes -l node-role.kubernetes.io/infra -o jsonpath='{.items[*].metadata.name}'); do echo ${node} ; kubectl drain ${node} --delete-emptydir-data --ignore-daemonsets=true --timeout=15s ; done

k8s-infra-1.k8s.internal
node/k8s-infra-1.k8s.internal already cordoned
Warning: ignoring DaemonSet-managed Pods: kube-system/cilium-vh7xq, kube-system/kube-proxy-qkdc8, k8s-csi-drivers/k8s-oauth-csi-provider-4rp6k, k8s-csi-drivers/k8s-secrets-store-csi-driver-2l7mj, k8s-ingress-internal/k8s-ingress-internal-controller-97mkb, k8s-monitoring/k8s-cadvisor-pdncn, k8s-monitoring/k8s-prometheus-node-exporter-4626w, k8s-secrets-webhook/k8s-oauth-secrets-webhook-sc8ms
evicting pod k8s-monitoring/prometheus-main-0
evicting pod kube-system/k8s-dns-6fb4489cd7-qn8mw
evicting pod k8s-gitops/image-automation-controller-5ffdb68d9d-t4flv
evicting pod kube-system/coredns-6ccbfbdc9d-6fk7l
...
```

4. Выключите узлы кластер Kubernetes:

```
for node in $(kubectl get nodes -o jsonpath='{.items[*].metadata.name}'); do kubectl debug node/${node} --image=hub.k8s-platform.io/registry/k8s/support-tools:v1.0.0 -n kube-system -- chroot /host shutdown -h 1; done
```

*Пример*

```
for node in $(kubectl get nodes -o jsonpath='{.items[*].metadata.name}'); do kubectl debug node/${node} --image=hub.k8s-platform.io/registry/k8s/support-tools:v1.0.0 -n kube-system -- chroot /host shutdown -h 1; done

Creating debugging pod node-debugger-k8s-infra-1.k8s.internal-n6vjk with container debugger on node k8s-infra-1.k8s.internal.
Creating debugging pod node-debugger-k8s-master-1.k8s.internal-78gfx with container debugger on node k8s-master-1.k8s.internal.
Creating debugging pod node-debugger-k8s-worker-1.k8s.internal-xd6wc with container debugger on node k8s-worker-1.k8s.internal.
```

5. Выключите при необходимости связанные с платформой контейнеризации внешние ресурсы (внешние хранилища, службы каталогов, хранилища образов и т.п.).

Для включения кластера платформы контейнеризации следуйте процедуре перезапуска узлов платформы.

### 3.1.7 Перезапуск платформы контейнеризации

В данном разделе описана процедура корректного перезапуска (включения) платформы контейнеризации.

3.1.7.1 Предварительные условия для перезапуска платформы контейнеризации

– Вы корректно выключили кластер платформы контейнеризации согласно процедуре выключения.

– У вас есть доступ к кластеру с учетной записью, имеющей роль cluster-admin платформе контейнеризации.

– Вы установили утилиту kubectl для работы с платформой контейнеризации.

Процедура

1) Включите связанные с платформой контейнеризации внешние ресурсы (внешние хранилища, службы каталогов, хранилища образов и т.п.).

2) Включите узлы кластера платформы контейнеризации и дождитесь загрузки ОС. Подождите не менее 5 минут перед как проверять статус мастер-узлов платформы.

3) Проверьте, узлы запущены и готовы к работе:

```
kubectl get nodes
```

4) Разрешите планирование (запуск) рабочих нагрузок на мастер-узлах:

```
kubectl uncordon -l node-role.kubernetes.io/control-plane
```

5) Дождитесь запуска рабочих нагрузок на мастер-узлах.

6) Разрешите запуск (планирование) рабочих нагрузок на инфраструктурных узлах:

```
kubectl uncordon -l node-role.kubernetes.io/infra
```

7) Дождитесь запуска рабочих нагрузок на инфраструктурных узлах.

8) Разрешите запуск рабочих нагрузок на рабочих узлах:

```
kubectl uncordon -l node-role.kubernetes.io/worker
```

9) Проверьте, что в кластере Kubernetes все необходимые сервисы запустились корректно:

```
kubectl get pods -A -o wide | grep -v "Run\|Comp"
```

Пустой вывод данной команды означает, что в кластере Kubernetes отсутствуют какие-либо сервисы, находящиеся в статусах, отличных от Running или Completed.

### 3.1.8 Распределение рабочей нагрузки в процессе работы

Базовая функциональность Kubernetes не предусматривает перераспределение рабочей нагрузки в процессе работы. Платформе контейнеризации для поддержки оптимального распределения и использования ресурсов задействуется компонент descheduler.

Descheduler на основе заданных критериев периодически оценивает нагрузку и вытесняет поды.

Descheduler запускается платформе контейнеризации с политикой, заданной в ConfigMap k8s-descheduler в namespace k8s-descheduler.

Descheduler каждые 15 минут вытесняет поды, которые удовлетворяют включенной в ConfigMap политике.

**Важно!** Время, после которого будут вытеснены поды завист от настройки minPodLifetimeSeconds.

3.1.8.1 Настройки политики Descheduler

3.1.8.1.1 DefaultEvictor

DefaultEvictor обеспечивает вытеснение подов таким образом, чтобы результирующий под поместился хотя бы на одном доступном узле, если указана опция nodeFit: true.

```
- name: "DefaultEvictor"
  args:
    nodeFit: true
```

3.1.8.1.2 RemoveDuplicates

Политика RemoveDuplicates следит за тем, чтобы на узле был запущенн только один под с одним контроллера. Если подов с контроллером два на одном узле, descheduler удаляет дублирующий под.

```
- name: ProfileName
  pluginConfig:
  - name: "RemoveDuplicates"
  plugins:
    balance:
      enabled:
        - "RemoveDuplicates"
```

3.1.8.1.3 RemoveFailedPods

Политика RemoveFailedPods вытесняет поды, которые завершились с ошибкой и в течении установленного времени minPodLifetimeSeconds не восстановились, за исключением тех, которые принадлежат типам, перечисленным в excludeOwnerKinds.

```
- name: "RemoveFailedPods"
  args:
    excludeOwnerKinds:
    - "Job"
    minPodLifetimeSeconds: 900
```

3.1.8.1.4 PodLifeTime

Политика PodLifeTime гарантирует, что поды в состояниях Pending и/или PodInitializing старше 30 минут будут вытеснены с узлов.

```
- name: "PodLifeTime"
  args:
    maxPodLifeTimeSeconds: 1800
    states:
    - "Pending"
    - "PodInitializing"
```

3.1.8.1.5 RemovePodsHavingTooManyRestarts

Политика RemovePodsHavingTooManyRestarts гарантирует, что поды, имеющие больше 50 перезапусков контейнеров (включая init-контейнеры), будут удалены с узлов.

```
- name: "RemovePodsHavingTooManyRestarts"
  args:
    podRestartThreshold: 50
    includingInitContainers: true
```

3.1.8.1.6 RemovePodsViolatingNodeTaints

Политика RemovePodsViolatingNodeTaints гарантирует, что поды, нарушающие определённые условия (taint) на узлах, будут удалены. Например, под имеющий toleration и запущенный на узле с соответствующим taint будет будет вытеснен с узла, если taint на узле будет изменен или удален.

Настроенные taint можно посмотреть в настройке политики.

```
- name: "RemovePodsViolatingNodeTaints"
  args:
    excludedTaints:
    - node.kubernetes.io/memory-pressure
    - node.kubernetes.io/disk-pressure
    - node.kubernetes.io/pid-pressure
    - node.kubernetes.io/unschedulable
    - node.kubernetes.io/network-unavailable
```

3.1.8.1.7 RemovePodsViolatingNodeAffinity

Политика RemovePodsViolatingNodeAffinity обеспечивает соблюдение подами affinity rules. Если поды не соответствуют affinity rules, то поды будут вытеснены с узла. Настроенные affinity rules:

– requiredDuringSchedulingIgnoredDuringExecution;
– preferredDuringSchedulingIgnoredDuringExecution.

```
- name: "RemovePodsViolatingNodeAffinity"
  args:
    nodeAffinityType:
    - "requiredDuringSchedulingIgnoredDuringExecution"
    - "preferredDuringSchedulingIgnoredDuringExecution"
```

3.1.8.1.8 RemovePodsViolatingInterPodAntiAffinity

Политика RemovePodsViolatingInterPodAntiAffinity следит за тем, чтобы все поды нарушившие правила anti-affinity были удалены.

```
- name: "RemovePodsViolatingInterPodAntiAffinity"
```

3.1.8.1.9 LowNodeUtilization

Политика LowNodeUtilization находит нагруженные узлы и перераспределяет их поды на ненагруженные узлы. Узлы в кластере отслеживаются по CPU/памяти/подам (в процентах). После превышения порога указанного в политике — перераспределяет поды.

**Важно!** Настройка учитывает не реальное потребление ресурсов на узле, а реквесты у подов.

```
- name: ProfileName
  pluginConfig:
  - name: "LowNodeUtilization"
    args:
      thresholds:
        "cpu" : 40
        "memory": 40
        "pods": 20
      targetThresholds:
        "cpu" : 80
        "memory": 80
        "pods": 60
  plugins:
    balance:
      enabled:
        - "LowNodeUtilization"
        - "RemoveDuplicates"
```

При обновлении платформы старая политика сохраняется в ConfigMap k8s-descheduler-deprecated.

## 3.2 Обновление платформы

Платформе контейнеризации обновление всех компонентов платформы выполняется с помощью одной команды k8s-ctl cluster update. Администраторы платформы могут получать статус обновления в терминале, а также подтверждать готовность к обновлению каждого узла кластера Kubernetes.

Утилита k8s-ctl содержит матрицу возможных обновлений платформы и выполняет необходимые проверки совместимости компонентов. Перед началом обновления k8s-ctl проверяет доступность необходимых публичных сервисов (репозиториев) и актуальность лицензионной информации.

Для начала обновления администратор платформы должен иметь следующую информацию:

– Иметь доступ в кластер Kubernetes с ролью cluster-admin.

– Иметь SSH-ключ для доступа к узлам кластера.

– Знать имя пользователя, под которым выполняется SSH-подключение к узлам кластера.

– Предоставить номер новой версии платформы, на которую необходимо выполнить обновление.

– Предоставить токен доступа к Vault.

– (Опционально) Указать, требуется ли в процессе обновления кластера также выполнить обновление (перевыпуск) всех TLS-сертификатов Control Plane в Kubernetes.

**Важно!** k8s-ctl требует наличие конфигурационного файла для доступа к кластеру Kubernetes. По умолчанию путь к конфигурационному файлу берется из переменной окружения KUBECONFIG. Если переменная отсутствует либо файл не найден, то используется локальный файл kubeadmin.conf.

### 3.2.1 Стадии и этапы обновления

Обновление платформы начинается с момента, когда администратор выполняет команду k8s-ctl cluster update и подтверждает запуск процесса обновления.

Процесс обновления глобально делится на две стадии, которые выполняются друг за другом:

– Стадия 1: Обновление Core-компонентов платформы, используя сценарии Configuration Manager.

– Стадия 2: Обновление платформенных сервисов (в том числе дополнительных модулей), используя службу непрерывного развертывания ПО

FluxCD.

На первой стадии выполняются следующие этапы:

1) Выполняются необходимые проверки, а именно:

– Доступность необходимых репозиториев.

– Актуальность лицензионной информации, если платформа установлена с использованием сети Интернет.

– Проверка подключения к кластеру Kubernetes.

– Доступность ресурсов CRD Infrastructure[config.k8s-platform.io] и ClusterVersion[config.k8s-platform.io].

– Проверка подключения к узлам кластера с помощью SSH.

– Проверка подключения к Vault.

2) Выполняются подготовительные действия, в том числе настройка Vault, и по очереди запускаются агенты Host Agent.

3) k8s-ctl по очереди для каждого узла запрашивает у администратора подтверждение начала обновления.

4) Агенты Host Agent передают серверу Configuration Manager обновленные факты о себе, в том числе желаемую версию платформы. Сервер Configuration Manager в свою очередь направляет агентам обновленные сценарии для настройки. Результат выполнения сценария агентами передается в Configuration Manager в виде отчета.

В среде Kubernetes на данном этапе выполняются следующие последовательные действия:

– Устанавливается запрет на запуск сервисов (Cordon) и происходит эвакуация всех сервисов с узла кластера Kubernetes (Drain).

– Выполняется обновление необходимых пакетов ОС, установка новых конфигураций, обновление компонентов Kubernetes.

– В случае успешного обновления снимается запрет на запуск сервисов, после чего узел кластера становится доступным для аллокации сервисов.

– k8s-ctl приостанавливает процесс обновления и запрашивает подтверждение администратора на обновление следующего узла кластера Kubernetes.

На второй стадии Source-контроллер службы FluxCD переключается на новую версию Git-репозитория платформенных сервисов, и запускается процесс обновления данных сервисов в Kubernetes:

1) k8s-ctl активирует ресурсы Kustomization.

2) Kustomize-контроллер службы FluxCD анализирует описание объектов в Git-репозитории и приводит в соответствие текущее состояние данных объектов в Kubernetes (процесс реконсиляции).

3) Реконсиляция некоторых ресурсов Kustomization выполняется последовательно друг за другом согласно установленным зависимостям. Завершением реконсиляции ресурса является успешная Healthcheck-проверка обновляемого платформенного сервиса.

4) k8s-ctl ожидает завершения реконсиляции всех ресурсов Kustomization.

5) По завершении процесса обновления k8s-ctl предоставляет администратору обновленные учетные данные (kubeadmin.conf), если выполнялся перевыпуск TLS-сертификатов.

**Важно!** В случае прерывания процесса обновления платформы на любом из этапов администратору достаточно повторно запустить команду k8s-ctl cluster update. При повторном запуске будут обновлены только оставшиеся узлы.

### 3.2.2 Продолжительность обновления

Продолжительность обновления платформы контейнеризации может меняться в зависимости от топологии кластера и специфики запущенных пользовательских сервисов. В данном разделе предоставлена информация, позволяющая понять и оценить факторы, влияющие на продолжительность обновления платформы в вашем окружении.

3.2.2.1 Влияющие факторы

Следующие факторы могут влиять на продолжительность обновления платформы:

Тип обновления платформы

Установка патч-релиза занимает меньшее количество времени, поскольку объем изменений и затрагиваемых компонентов, как правило, невелик. Минорные и мажорные релизы требуют больше времени в зависимости от количества обновляемых компонентов и сценариев обновления.

Количество узлов в кластере Kubernetes

Чем больше узлов в кластере Kubernetes, тем дольше осуществляется процесс обновления. Обновление узлов выполняется последовательно по одному узлу каждой роли в порядке:

– Мастер-узлы: Продолжительность обновления одного узла достаточно предсказуема и занимает около 5 минут.

– Инфраструктурные узлы: Продолжительность обновления зависит от количества установленных дополнительных модулей, перемещение которых на другие узлы занимает время. В среднем обновление одного инфраструктурного узла занимает

до 10 минут.

– Узлы балансировки нагрузки: Продолжительность обновления данных узлов минимальна и занимает около 5 минут на один узел.

– Рабочие узлы: Продолжительность обновления зависит от количества размещаемых на узлах пользовательских сервисов. Примерное время обновления узла без учета эвакуации сервисов - 5 минут.

**Важно!** Платформе контейнеризации процесс освобождения узла от каких-либо нагрузок имеет ограничение по времени - 5 минут. Если процессы Pod Eviction не завершаются за отведенный промежуток времени, то происходит принудительное освобождение узла от всех размещаемых на нем сервисов.

Количество установленных дополнительных модулей платформы

Каждый установленный дополнительный модуль содержит набор ресурсов Kustomization, реконсиляция которых с учетом всех зависимостей может занимать дополнительное время.

Обновление базового модуля платформы занимает до 10 минут, а каждого дополнительного модуля - около 5 минут.

Параметры пользовательских сервисов на рабочих узлах платформы

Если пользовательские сервисы имеют специфические либо некорректные настройки, влияющие на процессы их остановки, а также эвакуации с узла кластера (например, PreStop-хуки, Pod Disruption Budgets и т.п.), то такие сервисы могут блокировать операции освобождения узла от нагрузок. В данном случае, для каждого узла, где размещаются подобные сервисы, будет выделяться дополнительное ожидание в течение 5 минут до начала принудительного освобождения узла.

3.2.2.1.1 Расчет продолжительности обновления

Вы можете предварительно оценить время обновления вашего кластера платформы контейнеризации, если какие-либо исторические данные по обновлению похожих кластеров отсутствуют.

### 3.2.3 Обновление системы оркестрации и контейнеризации

Вы можете обновить кластер платформs контейнеризации с помощью утилиты k8s-ctl.

3.2.3.1 Каналы обновлений и релизы

В настоящий момент Платформы контейнеризации поддерживает 3 канала обновлений. Новый функционал выходит, в первую очередь, для последней мажорной версии. При наличии обратной совместимости может быть добавлена в предыдущую мажорную версию. Последняя поддерживаемая версия получает патчи, направленные на устранение ошибок и уязвимостей.

Релизы платформы бывают следующих видов:

- **Патч** (x.y.Z) - релиз, направленный на устранение мелких недочетов, ошибок конфигурации, обновление некоторых программных компонентов, в том числе патч-версии среды Kubernetes.
- **Минорный** (x.Y.z) - релиз, в котором добавляется новый функционал Платформы контейнеризации.
- **Мажорный** (X.y.z) - релиз, в котором повышается минорная версия Kubernetes.

Утилита k8s-ctl автоматически проверяет текущую версию платформы и возможность обновления на запрашиваемую версию. k8s-ctl имеет аналогичное версионирование релизов и **поддерживает три последних минорных релиза** Платформы контейнеризации.

**Пример**

У вас установлена Платформы контейнеризации версии v2.1.0. Вы можете использовать k8s-ctl версии v2.4.0 для обновления кластеров платформы до версии v2.4.0.

**Информация**

Минорный и мажорные релизы могут включать также изменения, направленные на устранение мелких недочетов, ошибок конфигурации, обновление некоторых программных компонентов.

В таблице указаны версии k8s-ctl, с помощью которых можно выполнить обновление платформы до версии, равной версии k8s-ctl.

### 3.2.4 Обновление ОС

Платформе контейнеризации управление пакетами ОС выполняется независимо от самой платформы. Вам не потребуется использовать новые образы ОС, однако необходимо будет иметь доступ к публичным репозиториям производителя ОС либо к корпоративным зеркалам для загрузки пакетов.

Перед установкой Платформы контейнеризации рекомендуется обновлять ОС до последней доступной и поддерживаемой версии.

| |
|---|
| Если необходимо добавить в ОС собственные пакеты, дополнительное ПО или выполнить обновление — исключите из списка следующие пакеты: |

- kubelet
- kubectl
- kubeadm
- cri-tools
- containerd.io
- k8s-storage-agent
- iptables*
- puppet*
- vault

## 3.3 Резервное копирование и восстановление

Резервное копирование и восстановление - это процессы создания резервной копии компонентов платформы и последующего восстановления данных компонентов из этой копии, если что-то пойдет не так.

Мы рекомендуем регулярно делать копии мастер-узлов платформы, с использованием внешних систем резервного копирования. В этом случае процедуры и политика резервного копирования определяется клиентом.

## 3.4 Системы хранения данных

### 3.4.1 Добавление oVirt CSI в платформе установленной методом UPI

В руководстве описан процесс добавления oVirt CSI в платформе, но только на Worker узле, который располагается в системе виртуализации платформы виртуализации.

3.4.1.1 Предварительные условия

Платформы контейнеризации установлена методом UPI в минимальной конфигурации с двумя Worker узлами. Один из узлов является виртуальной машиной в системе виртуализации платформы виртуализации, а второй - физическим сервером.

3.4.1.2 Описание процесса добавления oVirt CSI

1. Настройка платформы виртуализации.
2. Настройка Vault.
3. Настройка Платформы контейнеризации.
4. Проверка.

3.4.1.3 Настройка платформы виртуализации

Настройте пользователя и получите данные для настройки интеграции с платформы виртуализации согласно статье Интеграция с платформы виртуализации.

3.4.1.4 Настройка Vault

1. Зайдите в веб-консоль Vault с использованием root токена.
2. На вкладке **Secrets** перейдите в **k8s-secrets credentials** и создайте секрет с именем ovirt-csi.
3. Добавьте следующие ключи-значения в новый секрет:
   - ovirt_ca_bundle — цепочка корневых TLS-сертификатов для подключения к интерфейсу платформы виртуализации API.
4. Перейдите на вкладку **Policies** и создайте новую политику с именем k8s-system-ovirt-csi и следующей настройкой:

   path "k8s-secrets/data/credentials/ovirt-csi" { capabilities = ["read"] }
5. Перейдите на вкладку **Access** в раздел **Auth Methods k8s-kubernetes** и создайте новую роль со следующими параметрами:
   - **Name:** k8s-system-ovirt-csi
   - **Alias name source:** serviceaccount_uid
   - **Bound service account names:**
     - ovirt-csi-driver-controller-sa
     - ovirt-csi-driver-node-sa
   - **Bound service account namespaces:** k8s-csi-drivers
   - Раскройте блок Tokens
     - Включите опцию Generated Token’s Maximum TTL и поставьте значение в 1 час
     - Включите опцию Do Not Attach 'default' Policy To Generated Tokens
     - В поле **Generated Token’s Policies** добавьте имя политики созданной ранее
     - Включите опцию Generated Token’s Initial TTL и поставьте значение в 1 час
     - В поле Generated Token’s Type поставьте значение default

3.4.1.5 Настройка Платформы контейнеризации

1. В веб-консоли Платформы контейнеризации перейдите на вкладку **Узлы кластера Nodes**.
2. Повторите следующие действия для всех **Worker** узлов на платформы виртуализации:

a. Нажмите на выбранный узел
b. Перейдите на вкладку YAML и добавьте метку:
c.
```
labels:
    ovirt: 'true'
```

3. Создайте сервисные аккаунты и роли для них.
4. Установите контроллер и агенты.
5. Установите манифест для StorageClass

```
apiVersion: storage.k8s.io/v1
kind: StorageClass
metadata:
  name: ovirt-csi-sc
  namespace: k8s-csi-drivers
  annotations:
    storageclass.kubernetes.io/is-default-class: "true"
provisioner: csi.ovirt.org
reclaimPolicy: "Delete"
volumeBindingMode: WaitForFirstConsumer
allowVolumeExpansion: true
parameters:
  storageDomainName: "платформы виртуализацииCsiStorageDomainName"
  thinProvisioning: "true"
                fsType: "xfs"
```

a. Измените на имя вашего домена хранения в платформе виртуализации.

## 3.5 Логирование

### 3.5.1 Обзор

На текущий момент платформе контейнеризации можно установить дополнительный модуль OpenSearch.

### 3.5.2 OpenSearch

OpenSearch — это легко масштабируемая система поисковых и аналитических инструментов с открытым исходным кодом. Она была ответвлена от кодовой базы Elasticsearch 7.10.2. Из OpenSearch убрали компоненты, распространяемые не под лицензией Apache 2.0. В систему включены: движок хранения и поиска, веб-интерфейс, среда визуализации данных OpenSearch Dashboards, а также дополнения, которые ранее поставлялись в продукте Open Distro for Elasticsearch.

### 3.5.3 Custom Resource Definitions

В API Kubernetes *resource* — это конечная точка, которая хранит коллекцию объектов API определенного типа. Например, встроенный ресурс *Pods* содержит коллекцию объектов *Pod*.

*Custom resource definition (CRD)* определяет новый уникальный тип объекта в кластере и позволяет серверу API Kubernetes управлять всем его жизненным циклом.

Объекты *custom resource (CR)* создаются из *CRDs*, добавленных в кластер администратором, что позволяет всем пользователям кластера добавлять новый тип ресурса в проекты.

#### 3.5.3.1 Custom Resource Definitions платформе контейнеризации

Платформе контейнеризации *CRDs* представлены в веб-консоли на вкладке **Administration CustomResourceDefinitions**.

На этой странице можно найти *Custom Resources* и созданные для него экземпляры. Например, в статье по установке Logging operator описывается добавление ресурса _Syslog-ng, в манифесте которого можно увидеть apiVersion и kind.

apiVersion: logging.banzaicloud.io/v1beta1

kind: Logging

В поиске на странице **CustomResourceDefinitions** найдите нужный ресурс, используя значение kind из манифеста.

Также убедитесь, что в колонке *Группа* значение соответствует значению apiVersion из манифеста. Перейдите в этот ресурс и откройте вкладку **Экземпляры**. На этой странице можно увидеть все экземпляры выбранного ресурса, а также создать новый экземпляр.

Нажмите на кнопку **Создать <тип ресурса>**. В открывшемся окне виден общий код для этого ресурса. Также справа находится поле *Схема*, которое содержит общую информацию о свойствах объекта, эти данные помогут вам составить собственный манифест

### 3.5.4 Opensearch

#### 3.5.4.1 Обзор

Платформе контейнеризации есть возможность установить сервис OpenSearch, который автоматически начинает собирать логи кластера.

#### 3.5.4.2 Архитектура

3.5.4.2.1 Компоненты модуля

Модуль OpenSearch в кластере Платформы контейнеризации включает компоненты, представленные в таблице ниже:

| Компонент | Категория | Количество |
|---|---|---|
| Opensearch | Кластерный компонент | 1 |
| Opensearch Dashboard | Кластерный компонент | 1 |
| Provisioner | Кластерный компонент | 1 |
| FluentbitAgent | Хостовой компонент (агент) | 1 на узел |
| HostTailer | Хостовой компонент (агент) | 1 на узел |
| Fluentd | Кластерный компонент | 1 |

Компоненты OpenSearch являются контейнерами, которые взаимодействуют как между собой, так и с кластером Kubernetes и его узлами. Все компоненты OpenSearch запускаются и функционируют в среде Kubernetes.

- **Opensearch**: основной компонент. Платформе контейнеризации контроллер разворачивается в виде сущности StatefulSet. Также контроллер предоставляет REST API для управления сущностями OpenSearch.
- **Opensearch Dashboard**: компонент предоставляет веб-интерфейс для управления сущностями Opensearch.
- **Provisioner**: компонент, запускаемый однократно в виде сущности Job для первоначальной настройки Opensearch.
- **FluentbitAgent**: хостовый компонент (агент), который разворачивается в виде сущности DaemonSet на каждом узле кластера Kubernetes. Предназначен для сбора логов с контейнеров.
- **HostTailer**: аналог FluentbitAgent, но настроенный на сбор логов с узлов кластера.
- **Fluentd**: компонент сервиса сборщика логов. Предназначен для сбора логов с Fluentd в соответствии с заданными правилами (в виде *Custom Resources* Flow и Output) и отправки их в Opensearch.

3.5.4.2.2 Предварительно настроенные параметры

При установке модуля OpenSearch происходит установка Fluentd, его агентов и правил для сбора логов кластера.

В таблице представлены настройки для сбора и отправки логов в OpenSearch с использованием Logging Operator.

| Имя ресурса | Тип ресурса | Название индекса в OpenSearch |
|---|---|---|
| k8s-node-audit-flow | Flow | k8s-node-audit-<текущая дата> |
| k8s-node-audit-output | Output | k8s-node-audit-<текущая дата> |
| k8s-k8s-audit-apiserver-flow | Flow | k8s-k8s-audit-apiserver-<текущая дата> |
| k8s-k8s-audit-apiserver-output | Output | k8s-k8s-audit-apiserver-<текущая дата> |

| Имя ресурса | Тип ресурса | Название индекса в OpenSearch |
|---|---|---|
| k8s-kubelet-flow | Flow | k8s-kubelet-<текущая дата> |
| k8s-kubelet-output | Output | k8s-kubelet-<текущая дата> |
| k8s-Vault-audit-logs-flow | Flow | k8s-Vault-audit-logs-<текущая дата> |
| k8s-Vault-audit-logs-output | Output | k8s-Vault-audit-logs-<текущая дата> |
| k8s-scheduler-flow | ClusterFlow | k8s-scheduler-<текущая дата> |
| k8s-scheduler-output | ClusterOutput | k8s-scheduler-<текущая дата> |

Также устанавливается дашборд в Grafana в виде *ConfigMap* k8s-grafana-dashboard-opensearch, который размещается в Namespace k8s-monitoring, и используется для визуализации и мониторинга основных метрик OpenSearch.

3.5.4.2.3 Предварительно настроенные параметры внутри OpenSearch

После применения Kustomization необходимо дождаться его успешного завершения. В результате установки в кластер OpenSearch будут импортрованы следующие сущности:

- Index pattern
  - k8s-k8s-audit-apiserver-* - Аудит лог Kubernetes. Собирается с каждого мастер-узла;
  - k8s-node-audit-* - Аудит лог системы. Собирается с каждого узла кластера;
  - k8s-kubelet-* - Аудит лог kubelet. Собирается с каждого узла кластера;
  - k8s-Vault-audit-logs* - Аудит лог Vault.
- Dashboards
  - [платформы контейнеризации] Kubernates API audit log
  - [платформы контейнеризации] Vault audit logs
- Policies
  - k8s_policy_1
- Templates
  - k8s-base-template

| |
|---|
| Для доступа ко всем вышеперечисленным ресурсам, необходимо сменить Tenant с Private на Global. |

3.5.4.3 Высокая доступность и непрерывность работы

Для обеспечения высокой доступности и непрерывной работы компонентов OpenSearch рекомендуется увеличить количество реплик OpenSearch и Opensearch Dashboard. Это необходимо для предотвращения прерывания работы сервисов в случае недоступности отдельных узлов или при перезапуске сервисов.

Далее рассмотрены особенности работы каждого компонента OpenSearch в режиме высокой доступности.

**Opensearch**

Для повышения доступности Opensearch рекомендуется устанавливать модуль в рекомендуемой конфигурации.

**Opensearch Dashboard**

Веб-интерфейс управления сущностями Opensearch взаимодействует с контроллером через REST API и не влияет на работу ключевых сервисов Opensearch. Для повышения доступности рекомендуется устанавливать модуль в рекомендуемой конфигурации. Поддерживается постоянство пользовательских сессий с помощью настройки балансировки в Ingress Controller (nginx.ingress.kubernetes.io/upstream-hash-by: "$binary_remote_addr").

**Provisioner**

Компонент Provisioner является сущностью типа *Job* в Kubernetes и запускается только один раз при первоначальной установке, поэтому для него высокая доступность и большое количество реплик не требуются.

**FluentbitAgent**

Разворачивается в виде сущности DaemonSet на каждом узле кластера Kubernetes, поэтому сценарий с увеличением реплик здесь не применим.

**HostTailer**

Разворачивается в виде сущности DaemonSet на каждом узле кластера Kubernetes, поэтому сценарий с увеличением реплик здесь не применим.

**Fluentd**

Компонент требует обязательного увеличения количества реплик, если вы обрабатываете большой объем логов.

3.5.4.4 Планирование установки и системные требования

3.5.4.4.1 Общие требования к установке

Для установки модуля Opensearch платформе контейнеризации должны быть выполнены следующие условия:

- Платформы контейнеризации должна иметь версию v2.0.0 и выше.
- У вас есть доступ к кластеру с учетной записью, имеющей роль cluster-admin в Kubernetes.
- На инфраструктурных узлах кластера достаточно ресурсов для запуска компонентов Opensearch, либо для размещения Opensearch подготовлены отдельные узлы кластера Kubernetes.

3.5.4.4.2 Системные требования

Общие требования по вычислительным ресурсам можно найти в статье.

3.5.4.4.3 Конфигурация

Платформе контейнеризации по умолчанию не установлены лимиты на потребление ресурсов (Resource Limits) компонентами Opensearch. В таблице ниже указаны ресурсы, которые требуются для запуска компонентов Opensearch.

| Компонент | CPU Requests | Memory Requests |
|---|---|---|
| Vault Agent | 500m | 512Mi |
| Opensearch Dashboard | 100m | 512Mi |
| Init Containers | 100m | 128Mi |
| Opensearch | 500m | 1024Mi |
| Provisioner | 250m | 256Mi |
| Fluentd | 1000m | 1024Mi |

Эта таблица содержит данные для установки в минимальной конфигурации. Для рекомендуемой конфигурации требуется установить по три экземпляра компонентов Opensearch и Opensearch Dashboard. Соответственно, необходимые ресурсы для запуска этих компонентов следует увеличить в три раза.

В таблице ниже представлены дополнительные компоненты Opensearch с установленными лимитами на потребление ресурсов.

| Компонент | CPU Requests | CPU Limits | Memory Requests | Memory Limits |
|---|---|---|---|---|
| FluentbitAgent | 100m | 200m | 50M | 100M |
| HostTailer | 100m | 200m | 50M | 100M |

3.5.4.5 Запрет на удаление индексов

Чтобы запретить ручное удаление индексов (*immutability*) в OpenSearch, выполните следующие шаги:

1. В веб-консоли Платформы контейнеризации перейдите на вкладку **Administration CustomResourceDefinitions**.
2. Найдите ресурс *Kustomization* и зайдите в него.
3. Перейдите на вкладку **Экземпляры**, найдите k8s-release-logs-main и зайдите в него.
4. Перейдите на вкладку **YAML** и добавьте patch, который будет содержать текущие настройки из *ConfigMap* k8s-logs-config в Namespace k8s-logs, а также ваши изменения.

3.5.4.6 Настройка уведомлений в Opensearch

В данной статье описывается процесс настройки уведомлений в OpenSearch, включая создание оповещений и проверку.

3.5.4.6.1 Предварительные условия

- Вы ознакомились с архитектурой и концепциями Opensearch платформе контейнеризации.
- У вас есть доступ к кластеру с учетной записью, имеющей роль cluster-admin в Kubernetes.
- Вы установили утилиту kubectl для работы с Kubernetes.
- Вы установили модуль Opensearch.

3.5.4.6.2 Настройка почтовых уведомлений

1. Перейдите на вкладку **Management Notifications Email senders** и нажмите на кнопку Create SMTP sender.
2. Укажите имя, адрес электронной почты, адрес SMTP-сервера, порт и метод шифрования для почтового ящика, с которого будут отправляться уведомления.
3. Перейдите на вкладку **Management Notifications Email recipient groups** и нажмите на кнопку Create recipient group.
4. Укажите имя группы и адреса электронной почты получателей уведомлений.
5. Перейдите на вкладку **Management Notifications Channels** и нажмите на кнопку **Create channel**.
6. Укажите имя канала уведомлений, тип канала Email, тип отправителя SMTP sender и выберите, от кого будут отправляться уведомления, а также кто будет их получать.
7. Если SMTP-сервер требует авторизации, откройте веб-консоль Платформы контейнеризации, зайдите в pod k8s-logs-0, перейдите на вкладку **Терминал** и выберите контейнер opensearch.

Также вы можете зайти в консоль контейнера через kubectl (если он настроен), используя следующую команду:

```
kubectl exec --stdin --tty k8s-logs-0 -c opensearch -n k8s-logs -- /bin/bash
```

8. Для авторизации почтового адреса отправителя выполните две команды, заменив **<имя созданного ранее Email sender>**, **почтовый адрес отправителя** и **пароль от почтового адреса отправителя**.
9. /usr/share/opensearch/bin/opensearch-keystore add opensearch.notifications.core.email.<имя созданного ранее Email sender>.username <<< 'почтовый адрес отправителя'

   /usr/share/opensearch/bin/opensearch-keystore add opensearch.notifications.core.email.<имя созданного ранее Email sender>.password <<< 'пароль от почтового адреса отправителя'
10. Для обновления учётных данных в Opensearch перейдите на вкладку **Management** > **Dev Tools** и выполните команду:

    POST _nodes/reload_secure_settings

```
{
  "secure_settings_password": ""
}
```

11. Откройте созданный канал и нажмите на кнопку **Actions**, а далее **Send test message** для отправки тестового письма.

3.5.4.6.3 Настройка оповещения

1. Создайте новый индекс. Для этого перейдите на вкладку **Management** > **Index Management** > **Indexes** и нажмите на кнопку **Create Index** и укажите имя индекса.
2. Откройте веб-консоль Платформы контейнеризации, зайдите в под k8s-logs-0, перейдите на вкладку **Терминал** и выберите контейнер opensearch.

Также вы можете зайти в консоль контейнера через kubectl (если он настроен), используя следующую команду:

```
kubectl exec --stdin --tty k8s-logs-0 -c opensearch -n k8s-logs -- /bin/bash
```

3. Добавьте лог в новый индекс. Выполните следующую команду, заменив **<имя индекса>**.
4. Перейдите на вкладку **OpenSearch Plugins** > **Alerting** > **Monitors** и нажмите на кнопку **Create monitor**.
5. Заполните следующие поля:
   - Monitor name - укажите имя.
   - Выберите Per document monitor.
   - В блоке Schedule выберите интервал: каждую минуту.
   - Выберите созданный индекс в поле Index.
   - Query name - укажите имя.
   - Field - выберите *loglevel is CRITICAL*.
   - Нажмите на кнопку **Add trigger**.
   - Trigger name - укажите имя.
   - Specify queries or tags - выберите Query name из предыдущего пункта.
   - Action name - укажите имя.
   - Channels - выберите канал, который был настроен ранее.
6. Выполните пункты 2 и 3, чтобы сгенерировать новое сообщение.
7. Проверьте почту получателя уведомлений. Обратите внимание, что оповещение проверяется каждую минуту, поэтому письмо может не прийти сразу.
8. Перейдите на вкладку **OpenSearch Plugins** > **Alerting** > **Alerts** и убедитесь, что появилось оповещение.

## 3.6 Мониторинг

### 3.6.1 Обзор

На текущий момент платформе контейнеризации используются сервисы:

- Prometheus
- Grafana
- Alert Manager

#### 3.6.1.1 Архитектура

##### 3.6.1.1.1 Компоненты

Компоненты мониторинга в кластере Платформы контейнеризации представлены в таблице ниже:

| Компонент | Категория | Количество |
|---|---|---|
| Prometheus | Кластерный компонент | 1 |
| Prometheus Operator | Кластерный компонент | 1 |
| Alert Manager | Кластерный компонент | 1 |
| Prometheus Adapter | Кластерный компонент | 1 |
| Prometheus Node Exporter | Хостовой компонент (агент) | 1 на узел |
| Cadvisor | Хостовой компонент (агент) | 1 на узел |
| Grafana | Кластерный компонент | 1 |
| Kubernetes State Metrics (KSM) | Кластерный компонент | 1 |
| Kubernetes Metrics Server | Кластерный компонент | 1 |

Компоненты мониторинга являются контейнерами, которые взаимодействуют как между собой, так и с кластером Kubernetes и его узлами. Все они запускаются и функционируют в среде Kubernetes.

- **Prometheus**: основной компонент системы мониторинга и сбора метрик с встроенной базой данных временных рядов. Платформе контейнеризации контроллер разворачивается в виде сущности StatefulSet.
- **Prometheus Operator**: компонент, который упрощает развертывание и управление экземплярами Prometheus в Kubernetes. Платформе контейнеризации контроллер разворачивается в виде сущности Deployment.
- **Alert Manager**: компонент, ответственный за обработку оповещений.
- **Prometheus Adapter**: компонент предоставляет возможность использовать метрики Prometheus в качестве источника для горизонтального масштабирования в Kubernetes.
- **Prometheus Node Exporter**: компонент, предназначенный для сбора метрик со всех узлов кластера.

- **Cadvisor**: компонент, предназначенный для сбора метрик со всех контейнеров.
- **Grafana**: компонент для визуализации метрик, который позволяет создавать настраиваемые панели с графиками на основе данных, собранных из Prometheus.
- **Kubernetes State Metrics (KSM)**: компонент, который экспонирует метрики о внутренних объектах Kubernetes, таких как pod'ы, контроллеры, узлы, через Prometheus. Он предоставляет расширенные данные о состоянии и функционировании кластеров Kubernetes.
- **Kubernetes Metrics Server**: компонент, который в режиме реального времени собирает основные ресурсные метрики Kubernetes, такие как использование процессора и памяти на уровне pod'ов и узлов.

### 3.6.2 Особенности работы Prometheus платформе контейнеризации

Платформе контейнеризации вы можете настроить Prometheus для сбора метрик с определённых pod'ов в Namespace. Для этого используется Prometheus Operator и ServiceMonitor.

ServiceMonitor — это специальный пользовательский ресурс (*Custom Resource* / *CR*) в Kubernetes, который используется в связке с Prometheus Operator для автоматизации настройки сбора метрик от сервисов, работающих в Kubernetes.

3.6.2.1 Настройка пространства имён

Чтобы Prometheus Operator начал работать с ServiceMonitor в Namespace, необходимо добавить метку при создании или затем в существующий Namespace:

```
 labels:
   k8s-platform.io/cluster-monitoring: 'true'
```

3.6.2.2 Создание ServiceMonitor

Далее необходимо содать ServiceMonitor, с которым Prometheus Operator начнет работу.

```
apiVersion: monitoring.coreos.com/v1
kind: ServiceMonitor
metadata:
  name: example
spec: {}
```

3.6.2.3 Настройка pod'а

Создайте pod, который будет содержать сервис и встроенный экспортер. Pod должен содержать метку, указанную в ServiceMonitor.

### 3.6.3 Prometheus Adapter

Prometheus Adapter — это инструмент, позволяющий использовать метрики из Prometheus для масштабирования приложений в среде Kubernetes на основе динамических метрик. Он выступает связующим звеном между сервером метрик Kubernetes (Metrics API) и Prometheus, преобразуя метрики, собранные Prometheus, в формат, понятный Kubernetes, что обеспечивает автоматическое горизонтальное масштабирование (HPA — Horizontal Pod Autoscaler).

#### 3.6.3.1 Настройка Prometheus Adapter

По умолчанию Prometheus Adapter настраивается только на метрики CPU и Memory. Для добавления своих метрик выполните следующие шаги:

1. В веб-консоли Платформы контейнеризации перейдите на вкладку **Administration CustomResourceDefinitions** и найдите ресурс *Kustomization*. Далее перейдите на вкладку **Экземпляры** и найдите k8s-release-prometheus-adapter-main.
2. На вкладке **YAML** добавьте патч, который будет содержать текущие настройки из *ConfigMap* k8s-prometheus-adapter в Namespace k8s-monitoring и ваши изменения.
3. Убедитесь, что pod k8s-prometheus-adapter успешно перезапустился.
4. Убедитесь, что следующая команда выводит значение метрики:

   kubectl get --raw "/apis/custom.metrics.k8s.io/v1beta1/namespaces/<имя пространства имён>/pods/*/<название метрики из конфига Prometheus Adapter>"

#### 3.6.3.2 Настройка Horizontal Pod Autoscalers

После того как необходимая метрика становится доступной в Prometheus Adapter, можно приступать к настройке Horizontal Pod Autoscalers (HPA) для выбранного ресурса.

Alertmanager — это компонент экосистемы Prometheus, который управляет срабатыванием и обработкой оповещений. Он играет ключевую роль в системе мониторинга на базе Prometheus, организуя оповещения о различных событиях и аномалиях в работе кластера.

##### 3.6.3.2.1 Особенности Alertmanager платформе контейнеризации

###### 3.6.3.2.1.1 Оповещения в веб-консоли Платформы контейнеризации

Платформе контейнеризации оповещения настроены на странице **Observer** > **Система оповещений**.

- В блоке **Оповещения** видны все текущие проблемы в кластере.
- В блоке **Управление оповещениями** видны все правила остановки оповещения. Также можно создать правило для заглушения оповещений.
- В блоке **Правила оповещения** видны все настроенные правила для оповещений, включая как предустановленные, так и созданные пользователем.

3.6.3.2.1.2 Оповещения в Alertmanager

Для Alertmanager платформе контейнеризации создан ресурс *Ingress* с именем k8s-alertmanager-main в Namespace k8s-monitoring. При использовании ссылки из этого ресурса вы попадёте в веб-консоль Alertmanager, где доступны текущие оповещения, настройки Alertmanager и другие функции.

## 3.6.4 Grafana

Grafana — это платформа с открытым исходным кодом для визуализации и анализа данных, которая позволяет создавать динамические и интерактивные дашборды. Она поддерживает интеграцию с различными источниками данных, такими как базы данных, системы мониторинга и облачные сервисы. Grafana позволяет пользователям строить графики, диаграммы и настраивать предупреждения в реальном времени, что делает её популярным инструментом для мониторинга и анализа системных метрик, бизнес-аналитики и других видов данных.

3.6.4.1 Основные сведения

Grafana устанавливается в виде ресурса *Deployment* в Namespace k8s-monitoring. Для доступа к Grafana из внешней сети установлен ресурс *Ingress*.

В Grafana предустановлен источник Prometheus и дашборды. Данные из этих дашбордов отображаются на странице **Observe Панели мониторинга** в веб-консоли Платформы контейнеризации. Данные с других дашбордов не будут выведены.

Рекомендуется добавлять новые дашборды и источники данных через ConfigMap. Grafana автоматически начнёт с ними работать.

3.6.4.2 Добавление нового плагина

В настоящее время установка плагинов на постоянной основе невозможна. После перезапуска pod’a настройки возвращаются к исходному состоянию. Данный функционал будет реализован в одной из будущих версий платформы.

3.6.4.3 Добавление нового источника данных (Datasource)

Для добавления нового источника данных создайте ресурс *ConfigMap* в Namespace k8s-monitoring с именем k8s-grafana-datasources-<имя_источника> и меткой k8s-platform.io/cluster-dashboard-datasource: 'true'.

3.6.4.4 Добавление нового дашборда

Для добавления нового дашборда создайте ресурс *ConfigMap* в Namespace k8s-monitoring с именем k8s-grafana-dashboard-<имя_источника> и меткой grafana_dashboard: '1'.

# 4 Система управления вычислительными ресурсами

## 4.1.1 Начало работы

### 4.1.1.1 Авторизация на портале Control-panel

Авторизация в интерфейсе control-panel по адресу https://product-manager.<домен>/ (где <домен> - домен, заданный на этапе установки продукта) доступна с учетной записью, созданной на этапе post-deploy. Для авторизации нужно ввести имя пользователя/e-mail и пароль и нажать на кнопку Вход.

Администратор (глобальная роль Суперадминистратор в control-panel) может:

- настраивать подключение к платформам виртуализации и облачным провайдерам в разделе Инфраструктура
- настраивать права доступа и ресурсные правила в пространстве IAM и управление
- проводить аудит и отладку заказов в разделе Утилиты;
- управлять тарифными классами и пополнять счета организаций в пространстве Биллинг;
- управлять доступными для заказа на портале Продуктами, их ограничения и действия, а также настраивать графические адаптеры в пространстве Инструменты.

### 4.1.1.2 Инфраструктура

### 4.1.1.3 Облака

Раздел Облака отображает подключенные к инсталляции платформы виртуализации, позволяет добавить новые и управлять ими.

#### 4.1.1.3.1 Добавление подключения

Для добавления нового подключения выполните следующие действия:

Порядок действий

1. Перейдите в Инфраструктура → Облака
2. Нажмите Добавить подключение и выберите подходящий вариант (в текущей версии поддерживается Платформы виртуализации).
3. В открывшемся окне введите параметры подключения:
    - Название — уникальное имя подключения
    - URL — адрес подключения к системе управления
    - Доступ: имя и пароль администратора системы, к которой осуществляется подключение
    - При необходимости можно настроить дополнительные параметры, активировав переключатель Расширенные настройки
4. Нажмите Добавить.

5. При успешном подключении новая платформа отобразится в списке на экране Инфраструктура → Облака.

#### 4.1.1.3.2 Редактирование подключения

На экране Облака можно изменить параметры существующих подключений. Порядок действий

1. Перейдите в Инфраструктура → Облака
2. На карточке нужного подключения нажмите … и выберите Редактировать подключение
3. В открывшемся окне измените необходимые параметры
4. Нажмите Сохранить

#### 4.1.1.3.3 Запуск разведки ВМ

Разведка ВМ - это получение информации о виртуальных машинах, созданных на платформе виртуализации напрямую, без использования портала KAGECORE ML PLATFORM.

Запуски разведки ВМ можно выполнить двумя способами:

Способ 1

Разведка ВМ включается либо при создании, либо путём редактирования подключения:

1. Перейдите в Инфраструктура → Облака
2. В окне настройки подключения (открывается либо при создании нового подключения, либо при запуске редактирования) активируйте опцию Включить разведку ВМ и выберите нужный Воркер
3. Нажмите Добавить (Сохранить, в случае редактирования)

Способ 2

Актуален только для существующих подключений.

1. Перейдите в Инфраструктура → Облака
2. На карточке нужного подключения нажмите … и выберите Запустить разведку ВМ.

#### 4.1.1.3.4 Запуск разведки айтемов

Разведка айтемов — это получение информации о существующих ресурсных пулах платформ виртуализации:

- Кластеры
- Сети
- Домены хранения
- Образы

- Прочие пулы.

Запуски разведки айтемов можно выполнить двумя способами:

Способ 1

Разведка ВМ включается либо при создании, либо путём редактирования подключения.

1. Перейдите в раздел Инфраструктура → Облака.
2. В окне настройки подключения (открывается либо при создании нового подключения, либо при запуске редактирования) активируйте опцию Включить разведку айтемов и выберите нужный воркер.
3. Нажмите Добавить (Сохранить, в случае редактирования).

Способ 2

Актуален только для существующих подключений.

1. Перейдите в Инфраструктура → Облака
2. На карточке подключения нажмите … и выберите Запустить разведку айтемов.

4.1.1.3.5 Получение подробной информации о подключении

На экране Облака можно получить подробную информацию об имеющемся подключении.

Порядок действий

1. Перейдите в Инфраструктура → Облака
2. Нажмите на нужное подключение

В подробном представлении отображается общая информация о подключении, а также параметры обнаруженных айтемов. Информация об айтемах представлена в виде таблиц распределённых по соответствующим вкладкам.

4.1.1.3.6 Удаление подключения

Если подключение больше не требуется, его можно удалить. Для этого:

1. Перейдите в раздел Инфраструктура → Облака.
2. На карточке нужного подключения нажмите … и выберите Удалить.

4.1.1.4 Дата-центры

Дата-центр (ЦОД) — сущность, позволяющая разделить вычислительные ресурсы клиента по территориально/административному признаку (например по принадлежности ресурса к физическому ЦОДу).

На экране Дата-центры отображается список доступных центров данных.

Поля в разделе Фильтры позволяют настроить фильтрацию вывода известных дата-центров.

Различные элементы управления в списке дата-центров позволяют создать, изменить и удалить дата-центр.

4.1.1.4.1 Создание дата-центра

Порядок действий

1. Перейдите в Инфраструктура — Дата-центры
2. Нажмите Создать дата-центр
3. В появившемся окне введите необходимые параметры:
    a. Код — уникальное буквенно-цифровое значение
    b. Обозначение латиницей
    c. Организация — организация, к которой будет прикреплён дата-центр
    d. Вес (Вес определяет приоритет дата-центра. Чем больше вес, тем выше приоритет)
    e. При необходимости активируйте опцию Доступен в форме заказа
    f. Сегменты сети
4. Нажмите Создать

В случае успешного создания дата-центр отобразится в списке.

4.1.1.4.2 Изменение дата-центра

Порядок действий

1. Перейдите в Инфраструктура → Дата-центры
2. Щелкните по ID необходимого дата-центра и нажмите Редактировать
3. В появившемся окне измените необходимые параметры
4. Нажмите Сохранить

4.1.1.4.3 Удаление дата-центра

Если дата-центр больше не нужен, его можно удалить.

Порядок действий

1. Перейдите в Инфраструктура → Дата-центры
2. Щелкните по ID необходимого дата-центра и нажмите Удалить

Дата-центр будет помечен на удаление в списке.

4.1.1.5 Платформы

Платформа является объектом организации и представляет собой абстракцию, объединяющую сегменты сети и центры данных.

На экране Платформы отображается список добавленных платформ, а также присутствуют элементы управления платформами.

4.1.1.5.1 Создание платформы

Перед созданием платформы необходимо создать запись в справочнике для описания платформы.

Порядок действий

1. Перейти в меню Инструменты — Справочники
2. Выбрать справочник — resource_pool
3. На вкладке страницы нажать Добавить страницу

в открывшемся окне указать следующие параметры:

1. имя — символьное имя кластера из системы виртуализации. Например, в платформе виртуализации для получения имени нужно перейти в вкладку Ресурсы → Кластеры → Выбрать из списка кластер → во вкладке Общее скопировать имя
2. степень важности — определяет приоритет использования для размещения ресурсов
3. page data — содержит описание платформы в формате json,со следующей структурой:
    - "resource_pool" — описывает пул ресурсов для создания платформы:
        - name — имя кластера хранения из системы виртуализации.
        - uuid — идентификатор кластера хранения.
        - domain — имя домена хранения, который будет доступен для заказа ресурсов в системе KAGECORE ML PLATFORM.
        - endpoint — FQDN — системы виртуализации.
    - platform — описывает тип платформы, обычно в зависимости от используемой системы виртуализации.
    - category — описывает категорию использования, например значение vm указывает, что платформа используется для размещения виртуальных машин.
4. Указать теги, которые используются как Код при дальнейшем создании платформы (при указании тегов, один должен быть уникальный для однозначной идентификации новой платформы)

4.1.1.5.2 Порядок действий по созданию платформы

1. Перейдите в Инфраструктура → Платформы
2. Нажмите Создать платформу
3. В появившемся окне введите необходимые параметры:
    a. Название — Имя платформы
    b. Код — уникальный тег указанный в описании платформы в справочнике
    c. При необходимости выберите Организацию и задайте Вес
4. Нажмите Создать

Новая платформа появится в списке.

1. Нажать на ⋮ → Редактировать
2. В окне Редактирование платформы указать Сегменты сети и Дата-центры
3. Для сохранения параметров нажмите Сохранить

4.1.1.5.3 Изменение платформы

Порядок действий

1. Перейдите в Инфраструктура → Платформы
2. Напротив нужной платформы в списке нажмите ⋮и выберите Редактировать
3. В появившемся окне измените необходимые параметры
4. Нажмите Сохранить

4.1.1.5.4 Удаление платформы

Если платформа больше не требуется, её можно удалить.

Порядок действий

1. Перейдите в Инфраструктура → Платформы
2. Напротив нужной платформы в списке нажмите ⋮и выберите Удалить

Платформа будет помечена в списке, как удаленная.

4.1.1.6 Домены

Домены являются объектом организации и представляет собой абстракцию, назначаемую сегментам сетей.

На экране Домены отображается список добавленных доменов, а также присутствуют элементы управления доменами.

4.1.1.6.1 Создание домена

Порядок действий

1. Перейдите в Инфраструктура → Домены
2. Нажмите Создать домен
3. В появившемся окне введите необходимые параметры:
    a. Название — Имя домена
    b. Код — уникальное буквенно-цифровое значение
    c. Описание – выбрать Организацию
    d. Задайте Вес
4. Нажмите Создать

Новый домен появится в списке.

4.1.1.6.2 Изменение домена

Порядок действий

1. Перейдите в Инфраструктура — Домены
2. Напротив нужного домена в списке нажмите ⋮ и выберите Редактировать
3. В появившемся окне измените необходимые параметры
4. Нажмите Сохранить

4.1.1.6.3 Удаление домена

Если домен больше не требуется, его можно удалить.

Порядок действий

1. Перейдите в Инфраструктура — Домены
2. Напротив нужного домена в списке нажмите ⋮ и выберите Удалить

Домен будет помечен на удаление в списке.

4.1.1.7 Сегменты сети

Сегмент сети — сущность, позволяющая выделить часть сетевых ресурсов клиента для разделения групп сервисов на базе систем виртуализации, реализуется за счет виртуальных локальных сетей (VLAN).

Сегменты сети являются объектом организации и представляют собой сетевую абстракцию, которая может быть назначена дата-центрам и платформам.

На экране Сегменты сети отображается список добавленных сегментов, а также присутствуют элементы управления сегментами.

4.1.1.7.1 Добавление сегмента сети

Перед добавлением сегмента сети необходимо выполнить его настройку в сервисе NetBox, доступном по адресу https://netbox.<domain>/ с учётными правами администратора.

Порядок действий

Перейти в боковом меню слева во вкладку IPAM → VRFS → нажать на Add

Заполнить обязательные параметры:

1.
   a. Name — символьное имя объекта VRF, содержит только символы: "A-Z","a-z", "0-9", "_", "-", ":", "."
   b. RD — уникальный различитель маршрутов, содержит только символы: "A-Z","a-z", "0-9", "_", "-", ":", "." (RD важно запомнить, так как используется для связывания объектов NetBox и Сервиса заказов(order-service))
   c. Tenant — арендатор представляет собой группу ресурсов, используемую в административных целях.
   d. Нажмите Create

Перейти в боковом меню слева во вкладку IPAM → VLANs → нажать на Add

Заполнить обязательные параметры:

1.
    a. ID — цифровой идентификатор VLAN в диапазоне от 1 до 4094.
    b. Name — символьное имя VLAN, содержит только символы: "A-Z","a-z", "0-9", "_", "-", ":", "."
    c. Status — рабочее состояние VLAN (имеет значения: active-активно, reserved — зарезервировано, deprecated-устарело)
    d. Tenant — арендатор представляет собой группу ресурсов, используемую в административных целях.
    e. Assignment — описывает группу VLAN или сайт, к которому привязана VLAN.
    f. Custom virt subnet uuid — настраиваемый идентификатор виртуальной подсети

*Custom virt subnet uuid указывается из системы виртуализации, подключенной в качестве платформы к KAGECORE ML PLATFORM. Например, в платформе виртуализации в качестве данного идентификатора используется vnicprofile виртуальной сети (VLAN).*

*В результате, например, будет получен следующий вывод:*

*<vnic_profile href="/ovirt-engine/api/vnicprofiles/4bc56c78-784c-43da-b823-d897e134074a" id="4bc56c78-784c-43da-b823-d897e134074a">*

*<name>VM_VLAN2302</name>*

*<link href="/ovirt-engine/api/vnicprofiles/4bc56c78-784c-43da-b823-d897e134074a/permissions" rel="permissions"/>*

    g. Нажмите Create

Перейти в боковом меню слева во вкладку IPAM → PREXIES → нажать на Add

Заполнить обязательные параметры:

1. Prefix — IP-адрес сети и маска.
2. Status — рабочее состояние (имеет значения: active-активно, reserved — зарезервировано,deprecated-устарело)
3. VRF — имя VRF
4. Tags — набор меток-идентификаторов для объектов, в качестве которых могут выступать названия платформ, кластеров, дата-центров и т.д.
5. VLAN — имя VLAN
6. Tenant — арендатор представляет собой группу ресурсов, используемую в административных целях.
7. Defaul gw — IP-адрес шлюза по умолчанию.
8. Nameservers — IP-адрес DNS-сервера в JSON-формате.

Нажмите Create

1. Рекомендуется зарезервировать IP-адрес DNS-сервера, для этого необходимо перейти IPAM → IP Addresses → нажать на Add и заполнить следующие параметры:
    a. Address — IP-адрес DNS-сервера
    b. Status — рабочее состояние (имеет значения: active-активно, reserved — зарезервировано,deprecated-устарело)
    c. VRF — имя VRF
    d. Tenant — арендатор представляет собой группу ресурсов, используемую в административных целях.

Создать новый сетевой сегмент в сервисе заказов.

1. Перейти в Инфраструктура → Сегменты сети, нажать Создать сегмент сети и заполнить следующие параметры:
    a. Название
    b. Код — RD для VRF (ранее созданной в NetBox)
    c. Вес — числовое значение от 0 до 100, определяет приоритет использования сетевого сегмента. Чем больше вес, тем выше приоритет.
    d. Организация — имя организации, для которой будет использоваться данный сетевой сегмент.
2. Задать параметры для связи с центром-данных и средой, для этого перейдите в сегмент сети и нажмите ⋮и выберите Редактировать сегмент сети. Далее в окне Редактирование сегмента сети укажите значения для полей Дата-центры и Среды и нажмите Сохранить.
3. Прикрепить созданный сегмент сети к домену, для этого перейти во вкладку Домены → нажмите ⋮и выберите Редактировать напротив имени необходимого домена → в поле Сегменты сети добавить из списка, ранее созданный сегмент сети → нажмите Сохранить.
4. Прикрепить созданный сегмент сети к центру данных, для этого перейти во вкладку Дата-центры → нажмите ⋮и выберите Редактировать напротив имени необходимого дата-центра → в поле Сегменты сети добавить из списка, ранее созданный сегмент сети → нажмите Сохранить.
5. Прикрепить созданный сегмент сети к платформе, для этого перейти во вкладку Платформы → нажмите ⋮и выберите Редактировать напротив имени необходимой платформы→ в поле Сегменты сети добавить из списка, ранее созданный сегмент сети → нажмите Сохранить.

Новый сегмент появится в списке на административном портале, для просмотра перейти в Инфраструктура → Сегменты сети.

4.1.1.7.2 Изменение сегмента сети

Порядок действий

1. Перейдите в Инфраструктура → Сегменты сети
2. Напротив нужного сегмента в списке нажмите ⋮и выберите Редактировать
3. В появившемся окне измените необходимые параметры
4. Нажмите Сохранить

#### 4.1.1.7.3 Удаление сегмента сети

Если сегмент больше не требуется, его можно удалить.

Порядок действий

1. Перейдите в Инфраструктура — Сегменты сети
2. Напротив нужного сегмента в списке нажмите ⋮и выберите Удалить

Сегмент будет помечен на удаление в списке.

### 4.1.1.8 Виртуальные машины

Портал позволяет проводить обнаружение ранее созданных виртуальных машин и добавлять их на Портал.

Добавление ВМ в проект

Порядок действий

1. Перейдите в Инфраструктура — Виртуальные машины
2. Напротив нужной ВМ поставьте галочку и выберите Перенести
3. В открывшейся панели Перевод в проект выберите нужную Организацию и Проект
4. Нажмите Применить

В фоновом режиме происходит создание заказа. Добавленные ВМ будут отображены на пользовательском портале в меню «Базовые вычисления».

### 4.1.1.9 Менеджер ресурсов

На экране Менеджер ресурсов отображается информация об обнаруженных ресурсах.

При переходе на экран Менеджер ресурсов список может оказаться пустым (несмотря на наличие подключений).

Для отображения ресурсов обязательно необходимо выбрать тип ресурсов, которые нужно отобразить. Для этого кликните в поле ввода для фильтрации, выберите нужный тип и нажмите Найти.

Для более точного отображения список ресурсов можно отфильтровать. Для этого в поле ввода для фильтрации нажмите Добавить параметр, в выбранном параметре нажмите на + выберите нужное значение из предоставленных и нажмите Найти

Значения показателей различных ресурсов также можно увидеть в графическом представлении.

Для этого в списке ресурсов нажмите на название нужного ресурса.

#### 4.1.1.10 Атрибуты

Атрибуты предназначены для описания сущностей и их свойств, используемых на Портале.

Экран Атрибуты предназначен для отображения и управления системными и пользовательскими атрибутами.

Отображением атрибутов можно управлять с помощью переключателя Системные атрибуты:

1. В состоянии Выключено отображаются только пользовательские атрибуты
2. В состоянии Включено отображаются пользовательские и системные атрибуты

#### 4.1.1.11 Создание пользовательского атрибута

Создание новых атрибутов выполняется на экране Атрибуты

Порядок действий

1. Перейдите в Инфраструктура → Атрибуты
2. Нажмите Добавить атрибут
3. В появившемся окне введите необходимые параметры:
    a. Идентификатор — уникальное имя атрибута
    b. Описание — дополнительная информация об атрибуте
    c. Тип — тип значения, которое может принимать атрибут. Доступны следующие типы:
        i. String — строковое значение
        ii. Number — числовое значение
        iii. Boolean — булево значение
    d. Значение — позволяет задать предопределённый набор значений. Если ничего не задано, атрибут может принимать любое значение, соответствующее установленному типу
    e. Множественный выбор значений - активация этой опции позволит ввести несколько значений этого атрибута.
4. Нажмите Добавить

##### 4.1.1.11.1 Редактирование пользовательского атрибута

Порядок действий

1. Перейдите в Инфраструктура → Атрибуты.
2. Напротив нужного атрибута нажмите … и выберите Редактировать
3. В открывшемся окне измените необходимые параметры
4. Нажмите Сохранить

4.1.1.11.2 Копирование пользовательского атрибута

Можно создать новый атрибут на основе имеющегося.

Порядок действий

1. Перейдите в Инфраструктура → Атрибуты
2. Напротив нужного атрибута нажмите … и выберите Дублировать
3. В открывшемся окне измените необходимые параметры
4. Нажмите Добавить

4.1.1.11.3 Удаление пользовательского атрибута

Если атрибут больше не нужен его можно удалить.

Порядок действий

1. Перейдите в Инфраструктура - Справочники - Атрибуты
2. Напротив нужного атрибута нажмите … и выберите Удалить

4.1.1.12 Управление доступом

4.1.1.12.1 Сервисы

Экран Сервисы содержит список всех доступных на Портале сервисов. Сервис создается по определенным правилам для ввода в эксплуатацию.

Сервис регистрирует свои ресурсы в IAM для того, чтобы можно было разграничивать права доступа.

Каждый сервис имеет:

1. Ресурсные типы — объекты, предоставляемые и управляемые через API
2. Ресурсные действия — предоставляемые операции, которые могут совершать объекты
3. Ресурсы — экземпляры объектов сервиса
4. Ресурсные правила — специальные правила, которые указывают для какого API-запроса какое действие совершается

4.1.1.12.1.1 Создание сервиса

Порядок действий

1. Перейдите в Управление доступом → Сервисы
2. Нажмите Создать сервис
3. В появившемся окне введите необходимые параметры:
    a. Кодовое название
    b. Название
    c. Описание
4. Нажмите Создать

#### 4.1.1.12.1.2 Изменение сервиса

Порядок действий

1. Перейдите в Управление доступом → Сервисы
2. Нажмите значок редактирования в строке нужного сервиса
3. В открывшемся окне измените необходимые параметры
4. Нажмите Сохранить

#### 4.1.1.12.1.3 Удаление сервиса

Порядок действий

1. Перейдите в Управление доступом → Сервисы
2. Нажмите значок Корзина
3. Подтвердите удаление

### 4.1.1.12.2 Учетные записи

Экран Учетные записи содержит список всех созданных пользователей.

Для упрощения поиска нужных пользователей можно использовать фильтрацию.

В качестве единой точки аутентификации и авторизации используется Keycloak.

#### 4.1.1.12.2.1 Создание учетной записи

Порядок действий

1. Подключиться к сервису Keycloak по адресу https://auth.<domain>/auth/ с учётными данными администратора
2. Выбрать пространство (Realm)
3. Перейти в вкладку Users → нажать Add user
4. Указать следующие параметры для нового пользователя:
    a. Username
    b. Email
    c. First name
    d. Last name

Для сохранения параметров нажмите Create.

1. При успешном создании пользователя будут отображены его свойства
2. Перейдите во вкладку Credentials → нажмите Set Password и задайте пароль (отключите опцию Temporary если не требуется смена пароля при следующем входе)

После завершения процедуры создания, пользователь отобразиться в списке пользователей в Control-Panel. Далее пользователя можно будет отнести к организации и назначить ему роли. Подробнее рассмотрено в разделах «Назначение роли пользователю» и «Добавление пользователя к организации».

#### 4.1.1.12.2.2 Совместимость с LDAP

Для использования внешнего источника учетных записей, необходимо подключить в Keycloak новый LDAP провайдер (в качестве LDAP провайдера может использоваться ALD Pro из состава KAGECORE ML PLATFORM).

Порядок действий

1. Развернуть Active directory и создать пользователя, который будет иметь роль администратора AD. В данном случае имя пользователя — KAGECORE ML PLATFORM (все пользователи обязательно должны иметь почту)
2. Подключиться к сервису Keycloak
3. Выбрать пространство (Portal)
4. Перейти во вкладкуUser Federation → нажать Add new provider → LDAP
5. В открывшемся окне задать следующие параметры:
    - Console display name — отображаемое имя провайдера при ссылке в консоли администратора.
    - Vendor — поставщик (провайдер) LDAP, в данном случае выбираем Active Directory.
    - Connection URL — URL-адрес подключения к вашему серверу LDAP, указывается в формате ldap://IP-адрес сервера.252:389.
    - Use Truststore SPI — указывает, будет ли LDAP-соединение использовать Truststore SPI с truststore, настроенным в standalone.xml/domain.sml. 'Always' означает, что оно всегда будет его использовать. 'Никогда' означает, что оно не будет его использовать. 'Only for ldaps' означает, что он будет использовать его, если URL вашего соединения использует ldaps.
    - Connection timeout — таймаут соединения LDAP в миллисекундах.
    - Bind type — тип метода аутентификации, используемого во время операции связывания LDAP. Он используется в большинстве запросов, отправляемых на LDAP-сервер. В настоящее время доступны только механизмы 'none' (анонимная аутентификация LDAP) или 'simple' (аутентификация по привязке учетных данных + привязка пароля).
    - Bind DN — DN администратора LDAP, который будет использоваться Keycloak для доступа к серверу LDAP.
    - Bind credentials — пароль администратора LDAP.

После основных параметров задайте параметры поиска и обновления по LDAP.

- Edit mode — задает параметры редактирования.
  Значение READ_ONLY означает что, LDAP-хранилище только для чтения. Значение WRITABLE означает, что данные будут синхронизироваться с LDAP по требованию. Значение UNSYNCED означает, что пользовательские данные будут импортированы, но не будут синхронизированы с LDAP.
- Users DN — указывается полный DN дерева LDAP, в котором находятся ваши пользователи. Этот DN является родительским для пользователей LDAP. Это может быть, например, 'ou=users,dc=example,dc=com', предполагая, что ваш пользователь будет иметь DN типа 'uid='john',ou=users,dc=example,dc=com.
- Username LDAP attribute — имя атрибута LDAP, который сопоставляется с именем пользователя Keycloak. Для многих поставщиков LDAP-серверов это может быть 'uid'. Для Active directory это может быть 'sAMAccountName' или 'cn'. Атрибут должен быть заполнен для всех записей пользователей LDAP, которые вы хотите импортировать из LDAP в Keycloak.
- RDN LDAP attribute — имя LDAP-атрибута, который используется в качестве RDN (верхнего атрибута) типичного DN пользователя. Обычно это то же самое, что и LDAP-атрибут Username, однако это не обязательно.
- UUID LDAP attribute — Имя атрибута LDAP, который используется в качестве уникального идентификатора объекта (UUID) для объектов в LDAP. Для Active directory должно быть задано objectGUID.
- User object classes — все значения атрибута LDAP objectClass для пользователей в LDAP, разделенные запятыми. Например: 'inetOrgPerson, organizationalPerson'. Вновь созданные пользователи Keycloak будут записаны в LDAP со всеми этими объектными классами, а существующие записи пользователей LDAP будут найдены только в том случае, если они содержат все эти объектные классы.

1. Для проверки подключения провайдера перейдите во вкладку Users → в строке поиска введите символ * и будут отображены как локальные пользователи, так и из каталога Users в Active Directory.
2. Для назначения пользователю роли перейдите в административную панель во вкладку Управление доступом → Глобальные роли. Щелкните по необходимой роли и во вкладке Участники нажмите Добавить участника и выберите пользователя.
3. Для проверки входа пользователя перезайдите в административную панель с учетными данными пользователя. В качестве учетных данных используйте параметр cn и пароль указанный в Active Directory.

#### 4.1.1.12.3 Сервисные аккаунты

Сервисный аккаунт — аккаунт, от имени которого можно управлять ресурсами в KAGECORE ML PLATFORM.

Экран Сервисные аккаунты содержит список всех созданных Сервисных аккаунтов.

Для упрощения поиска нужных сервисных аккаунтов, можно использовать фильтрацию.

4.1.1.12.4 Роли

Пользователи портала получают роль в рамках контекста — на уровне организации, папки или проекта. При получении роли на проект, действия, доступные данной роли, пользователь может выполнять только в этом проекте. Если часть действий, доступных роли, можно совершать только на уровне организации — пользователь не сможет их совершить. Если роль получена на папку или организацию, то доступные действия пользователь может совершать в выбранной папке, а также во всех дочерних папках и проектах вниз по дереву. Причем не только в тех, которые существовали при назначении прав пользователю, но и во всех вновь созданных.

Экран Роли содержит список всех известных ролей.

Роли могут быть следующих типов:

- Базовые
- Пользовательские
- Сервисные
- Глобальные

Для удобства поиска на экране доступны поля для фильтрации вывода.

Для обновления соответствующего типа ролей используйте кнопки Обновить сервисные роли и Обновить базовые роли

4.1.1.12.4.1 Базовые роли

К базовым ролям относятся:

4.1.1.12.4.1.1 Администратор

Роль дает все разрешения для управления порталом внутри организации:

1. Заказ и управление услугами Портала
2. Управление доступом пользователей к услугами Портала
3. Просмотр, создание и удаление папок и проектов внутри организации
4. Просмотр списка пользователей организации и их ролей, назначение ролей пользователям
5. Приглашение новых пользователей в организацию
6. Просмотр истории переводов средств на счет организации
7. Аудит действий пользователей на портале, подраздел Аудит раздела Аналитика
8. Просмотр подраздела Расходы раздела Аналитика
9. Просмотр существующих ролей в организации
10. Создание кастомных ролей для своей организации
11. Управление сервисными аккаунтами и ключами сервисных аккаунтов
12. Управление SSH-ключами для доступа к виртуальным машинам

4.1.1.12.4.1.2 Редактор

Роль дает разрешения для управления порталом за исключением управления доступом к порталу:

1. Заказ и управление сервисом Compute (в интерфейсе пользователя – базовые вычисления)
2. Заказ и управление услугами Портала
3. Просмотр, создание и удаление папок и проектов внутри организации
4. Просмотр списка пользователей организации и их ролей
5. Просмотр истории переводов средств на счет организации
6. Аудит действий пользователей на портале, подраздел Аудит раздела Аналитика
7. Просмотр подраздела Расходы раздела Аналитика
8. Просмотр существующих ролей в организации
9. Создание кастомных ролей для своей организации
10. Управление сервисными аккаунтами и ключами сервисных аккаунтов
11. Управление SSH-ключами для доступа к виртуальным машинам

4.1.1.12.4.1.3 Наблюдатель

Роль дает разрешения на просмотр функционала портала:

1. Просмотр заказанных виртуальных машин
2. Просмотр созданных заказов услуг
3. Просмотр папок и проектов внутри организации
4. Просмотр списка пользователей организации и их ролей
5. Просмотр истории переводов средств на счет организации
6. Аудит действий пользователей на портале, подраздел Аудит раздела Аналитика
7. Просмотр подраздела Расходы раздела Аналитика
8. Просмотр существующих ролей в организации
9. Просмотр сервисных аккаунтов

4.1.1.12.4.2 Сервисные роли

К сервисным ролям относятся:

4.1.1.12.4.2.1 Администратор организации

Роль дает разрешения для управления доступом к проектам, папкам и организации:

1. Просмотр списка пользователей организации и их ролей, назначение ролей пользователям
2. Просмотр тарифного плана организации

4.1.1.12.4.2.2 Администратор IAM проекта

Роль дает разрешения для управления доступом к проекту, на который она назначена (либо ко всем проектам, находящимся внутри папки/организации, если роль назначена на уровне папки/ организации):

1. Просмотр раздела орг.структура
2. Назначение ролей существующим пользователям организации на проект

4.1.1.12.4.2.3 Администратор сервисных аккаунтов

Роль дает разрешения для управления сервисными аккаунтами:

1. Создание, редактирование и удаление сервисных аккаунтов в проектах
2. Управление ключами сервисных аккаунтов

4.1.1.12.4.2.4 Наблюдатель сервисных аккаунтов

Роль дает разрешения для просмотра сервисных аккаунтов:

1. Просмотр сервисных аккаунтов (доступны на уровне проекта)
2. Просмотр ключей сервисных аккаунтов

4.1.1.12.4.2.5 Администратор ролей

Роль дает разрешения для управления доступом к проектам, папкам и организации:

1. Создание, редактирование и удаление кастомных ролей в организации

4.1.1.12.4.2.6 Наблюдатель аудита

Роль дает разрешения на просмотр действий пользователей на портале и просмотр подраздела Аудит и раздела Аналитика.

4.1.1.12.4.3 Глобальные роли

К глобальным ролям относятся:

4.1.1.12.4.3.1 Суперадминистратор

Роль дает доступ к следующим функциям:

1. Управление всеми организациями облака
2. Просмотр и управление продуктовым каталогом

3. Просмотр и управление биллинговыми аккаунтами всех организаций облака, перевод средств.
4. Аудит действий участников команды облака
5. Просмотр данных и управление сервисом новостей и рассылок
6. Назначение глобальных ролей участникам команды сопровождения
7. Просмотр списка ролей и глобальных ролей любого пользователя облака
8. Просмотр и управление тарифными планами и тарифными классами всех организаций облака
9. Выгрузка счетов всех организаций облака

#### 4.1.1.12.4.3.2 Системный Суперадминистратор

Роль даёт полный доступ к управлению сервисными операциями

#### 4.1.1.12.4.3.3 Супер Редактор

Роль дает возможность управления всеми действиями в облаке, за исключением назначения глобальных ролей:

1. Управление всеми организациями облака
2. Просмотр и управление продуктовым каталогом
3. Просмотр и управление биллинговыми аккаунтами всех организаций облака, перевод средств.
4. Аудит действий участников команды облака
5. Просмотр данных и управление сервисом новостей и рассылок
6. Назначение глобальных ролей участникам команды сопровождения
7. Просмотр списка ролей и глобальных ролей любого пользователя облака
8. Просмотр и управление тарифными планами и тарифными классами всех организаций облака
9. Выгрузка счетов всех организаций облака

#### 4.1.1.12.4.3.4 Presale Manager, менеджер по настройке организаций

Роль дает доступ к следующим функциям:

1. Просмотр, создание и управление организациями, папками и проектами
2. Просмотр списка пользователей организации и их ролей, назначение ролей пользователям
3. Приглашение новых пользователей в организацию
4. Просмотр и управление тарифными планами и тарифными классами всех организаций облака

#### 4.1.1.12.4.3.5 Администратор Продуктового каталога

Роль дает доступ к следующим функциям:

1. Просмотр и управление Продуктовым конструктором в Control-Panel
2. Просмотр логов Продуктового конструктора
3. Управление тарифными классами
4. Просмотр базового тарифного плана и тарифных планов организаций

4.1.1.12.4.3.6 Наблюдатель продуктового каталога

Роль дает доступ к следующим функциям:

1. Просмотр всех разделов Продуктового конструктора в Control-Panel
2. Просмотр логов Продуктового конструктора

4.1.1.12.4.3.7 Администратор сервиса справочников

Роль дает доступ к управлению сервисом справочников.

4.1.1.12.4.3.8 Администратор тарифных классов

Роль дает доступ к следующим функциям:

1. Управление тарифными классами
2. Просмотр базового тарифного плана и тарифных планов организаций

4.1.1.12.4.3.9 Администратор тарифных планов

Пользователь с данной ролью имеет доступ к просмотру и управлению базовым тарифным планом и тарифными планами организаций.

4.1.1.12.4.3.10 Администратор тарифов

Роль дает доступ к следующим функциям:
Управление тарифными классами

1. Просмотр и управление базовым тарифным планом и тарифными планами организаций

4.1.1.12.4.3.11 Администратор управления доступом к организации

Роль дает доступ к следующим функциям:

1. Просмотр, создание и управление организациями
2. Просмотр списка пользователей организации и их ролей, назначение ролей пользователям

4.1.1.12.4.3.12 Администратор управления доступом к папке

Роль дает доступ к следующим функциям:

1. Просмотр, создание и управление папок в организациях
2. Просмотр списка пользователей на уровне папки и их ролей, назначение ролей пользователям

4.1.1.12.4.3.13 Супер наблюдатель

Роль дает доступ к просмотру всех ресурсов.

4.1.1.12.4.3.14 Наблюдатель сервиса состояний (State Service)

Роль дает доступ на просмотр данных сервиса состояний

4.1.1.12.4.4 Назначение роли пользователю

Порядок действий

1. Перейти во вкладку Управление доступом → Глобальные роли
2. Выбрать из списка необходимую роль → перейти во вкладку Участники
3. Нажать на Добавить участников → выбрать участников из списка пользователей.
4. Нажать Сохранить

4.1.1.12.5 Организации

Организация — объект, который описывает самый верхний уровень иерархии организационной структуры Портала и ассоциируется с реальной организацией или инфраструктурой.

На экране Организации отображается список организаций, а также доступны инструменты управления ими.

Для удобства поиска на экране доступны поля для фильтрации вывода.

4.1.1.12.5.1 Создание организации

Пользователь с правами администратора может создавать отдельные организации.

Порядок действий

1. Перейдите в Управление доступом → Организации
2. Нажмите Создать организацию
3. В появившемся окне введите необходимые параметры:
    - Название

- Email владельца
- При необходимости указать вручную ID организации.
- Описание
- DNS серверы
- Дополнительно внести параметры для ресурсных квот, которые доступны после нажатия на Добавить ресурсную квоту.

4. Нажмите Создать

Созданная организация появится в списке.

#### 4.1.1.12.5.2 Изменение организации

Порядок действий

1. Перейдите в Управление доступом → Организации
2. Нажмите в строке с нужной организацией нажмите ⋮и выберите Редактировать
3. В появившемся окне измените необходимые параметры.
4. Нажмите Сохранить.

#### 4.1.1.12.5.3 Удаление организации

Если организация больше не требуется, её можно удалить.

Порядок действий

1. Перейдите в Управление доступом → Организации.
2. Нажмите в строке с нужной организацией нажмите ⋮и выберите Удалить
3. Подтвердите удаление, нажав кнопку Удалить

#### 4.1.1.12.5.4 Добавление пользователя к организации

Порядок действий

1. Пройти авторизацию на Портале KAGECORE ML PLATFORM
2. Активируйте необходимый контекст соответствующей организации
3. Перейдите на экран Управление доступом - Учетные записи
4. Нажмите +
5. В появившемся окне:
    - Выберите одного или нескольких пользователей в поле Пользователь. Для поиска пользователя начните набирать его имя
    - В поле Роли из выпадающего меню выберите роли, которые хотите назначить пользователю для этой организации/папки/проекта

4.1.1.12.6 Управление ресурсными квотами

Создание ресурсных квот доступно при создании организации и на странице уже существующей организации на портале Control-panel. Добавление ресурсных квот для конкретного проекта осуществляется через пользовательский портал KAGECORE ML PLATFORM.

4.1.1.12.6.1 Создание ресурсных квот организации

Порядок действий

- Перейдите в Управление доступом → Организации.
- Во время создания организации нажмите на Добавить ресурсную квоту.

или

- Выберите из списка уже созданную организацию и в окне управления, перейдите на вкладку Ресурсные квоты и нажмите на Добавить ресурсную квоту.
- В окне добавления ресурсной квоты:
  - Выберите в параметре Платформа тип платформы.
  - Выберите из списка Подключение.

| | |
|---|---|
| | Ресурсные квоты применяются отдельно к каждому подключению. |

  - Укажите значения параметров CPU, RAM, Диски.

| | |
|---|---|
| | Необходимо указать хотя бы один из параметров:<br>§ CPU - число, указывающее количество CPU, доступное для заказов на организацию.<br>§ RAM - число, указывающие количество памяти в ГБ, доступное для заказов на организацию.<br>§ Диски - число, ограничивающее размер всех дисков для заказов на организацию. |

- Нажмите Добавить.

Созданная квота будет отображена в свойствах организации на вкладке Ресурсные квоты

#### 4.1.1.12.6.2 Редактирование ресурсных квот

Порядок действий

- Перейдите в Управление доступом → Организации.
- Выберите из списка организацию и перейдите в ее свойства.
- Перейдите на вкладку Ресурсные квоты.
- В строке с необходимой квотой нажмите ⋮ и выберите Редактировать.
- Измените необходимые параметры и нажмите Сохранить.

В случае изменения параметров квот, нужно учитывать, что квоту можно уменьшить только до значений используемых ресурсов. Например, если в квоте было указано значение 20 CPU и были заказаны ресурсы с такими параметрами, то данную квоту уменьшить нельзя.

#### 4.1.1.12.6.3 Удаление ресурсных квот

Порядок действий

- Перейдите в Управление доступом → Организации.
- Выберите из списка организацию и перейдите в ее свойства.
- Перейдите на вкладку Ресурсные квоты.
- В строке с необходимой квотой нажмите ⋮ и выберите Удалить.
- Подтвердите удаление.

При удалении ресурсной квоты для организации будут удалены все ресурсные квоты для папок и проектов внутри организации в данном подключении.

### 4.1.1.12.7 Настройки IAM

Identity and Access Management — сервис предоставления прав доступа к Порталу различным пользователям. В этом меню можно управлять настройками IAM. Для этого нужно внести необходимые изменения и нажать Сохранить.

### 4.1.1.12.8 Настройки Unit Manager

В этом меню можно управлять настройками ресурсного менеджера. Для этого нужно внести необходимые изменения и нажать Сохранить.

## 4.1.1.13 Утилиты

### 4.1.1.13.1 Аудит

Утилита Аудит предназначена для фильтрации/поиска операций.

Для упрощения поиска нужных операций доступны фильтры.

В графе Отображать можно выбрать информацию для отображения.

Для дополнительной фильтрации можно использовать Дополнительные фильтры, позволяющие уточнить запрос.

4.1.1.13.2 Отладка заказов

Сервис уникален для каждой организации. Инструкции по эксплуатации сервиса будут доставлены вместе с дистрибутивом.

4.1.1.13.3 Сервис состояний

Утилита Сервис состояний предназначена для просмотра состояний в рамках Events/Items/Actions.

Данные отображаются в формате JSON.

Для получения необходимых сведений:

1. Выберите тип объекта:
   - Events - события
   - Items — айтемы (ВМ, диски и т.п.)
   - Actions — действия
2. Для уточнения запроса используйте фильтры
3. Для поиска нужных значений или ключей используйте поле поиска

4.1.1.14 Инструменты

4.1.1.14.1 Справочники

Для получения описание объектов, таких как список ресурсов, список образов можно воспользоваться справочником. Справочник содержит группы объектов и описание атрибутов объектов, которые можно использовать в графах и в системе.

Для просмотра перейдите Инструменты → Справочники.

Откройте необходимый справочник и выберите необходимую страницу для просмотра.

Пример создания новой страницы в справочнике:

1. Перейти в меню Инструменты - Справочники
2. Выбрать необходимый справочник, например Resource_pool.
3. На вкладке страницы нажать Добавить страницу. В данном случае новая станица будет описывать параметры для добавления новой платформы в систему.

в открывшемся окне указать следующие параметры:

1. имя - символьное имя кластера из системы виртуализации.
2. степень важности - определяет приоритет использования для размещения ресурсов
3. page data - содержит описание платформы в формате json,со следующей структурой:
   - "resource_pool" - описывает пул ресурсов для создания платформы:
     - name - имя кластера хранения из системы виртуализации.
     - uuid - идентификатор кластера хранения.

    - domain - имя домена хранения, который будет доступен для заказа ресурсов в системе KAGECORE ML PLATFORM.
    - endpoint - FQDN - системы виртуализации.
  - platform - описывает тип платформы, обычно в зависимости от используемой системы виртуализации.
  - category - описывает категероию использования, например значение vm - указывает, что платформа используется для размещения виртуальных машин.
4. Указать теги, которые используются как Код при дальнейшем создании платформы (при указании тегов, один должен быть уникальный для однозначной идентификации новой платформы)

## 4.2 Система машинного обучения

Среды разработки представлены в виде готовых образов контейнеров, размещенных в среде контейнеризации.

### 4.2.1 Выполнять на ВМ mlflow-db (БД для хранения метаданных экспериментов)

1. Установка и настройка СУБД PostgreSQL (БД хранится на отдельном диске /dev/sdb1. В проде возможна другая конфигурация)

```
sudo apt install postgresql-11
      sudo pg_ctlcluster 11 main stop

      sudo pg_dropcluster 11 main

      sudo mount /etc/sdb1 /var/lib/postgresql

      sudo pg_createcluster 11 main

      sudo echo "/dev/sdb1     /var/lib/postgresql     ext4     defaults     0
0" >> /etc/fstab
```

2. Включить в конф. Файле Postgres /etc/postgresql/11/main/postgresql.conf поддержку huge_pages и увеличить размер shared_buffers до 25% от RAM на ВМ:

```
shared_buffers = 1024MB                 # min 128kB
                                              # (change requires restart)
      huge_pages = try                        # on, off, or try
```

3. Задать в ОС размер huge pages в 75% от RAM на ВМ:

```
sudo echo 'vm.nr_hugepages = 1500' >> /etc/sysctl.d/30-postgresql.conf
      sudo sysctl -p --system
```

4. Запустить postgresql

```
sudo pg_ctlcluster 11 main start
```

5. Создать БД для хранения метаданных и пользователей:

```
su - postgres
      psql -c "alter user postgres with password '!QAZxsw2'"
      createuser mlflowdbuser
      createdb mlflow-db -O mlflowdbuser
      psql -c "alter user mlflowdbuser with password '!QAZxsw2'"
```

### 4.2.2 Выполнять на ВМ mlflow

1. Установка conda

```
mkdir -p /opt/miniconda3
      wget     https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-
x86_64.sh -O /opt/miniconda3/miniconda.sh
      bash /opt/miniconda3/miniconda.sh -b -u -p /opt/miniconda3
      rm ~/opt/miniconda3/miniconda.sh
```

2. Установка MLflow

```
source /opt/miniconda3/bin/activate
      conda install python=3.10

      pip3 install mlflow[extras]==2.18.0 psycopg2-binary boto3 pip-system-
certs mlflow_oidc_auth
```

3. Настройка MLflow. В конфигурационный файл /etc/default/mlflow-ts записываем

```
cat > /etc/default/mlflow-ts << EOF
      MLFLOW_S3_ENDPOINT_URL="http://minio-api.tech:9000"
      MLFLOW_S3_IGNORE_TLS=true
      AWS_ACCESS_KEY_ID=NU3nH2kp8hWlDN6rppZv
      AWS_SECRET_ACCESS_KEY=19uwsdnCaWQiFAMkrGMgXv7gNuE1dOIpvVv1rucH
      MLFLOW_AUTH_CONFIG_PATH=/etc/mlflow/auth_conf.ini

      OIDC_DISCOVERY_URL=https://auth.k2int.tech/auth/realms/Portal/.well-
known/openid-configuration
      OIDC_CLIENT_ID='mlflow'
```

```
      OIDC_CLIENT_SECRET='ax9xWyg0HURhNDCMD0oXbZ4AWIt9tiu7'
      OIDC_PROVIDER_DISPLAY_NAME="Login with Keycloak"
      OIDC_REDIRECT_URI=http://mlflow.tech:5001/callback
      OIDC_SCOPE="openid profile email groups"
      OIDC_GROUP_NAME="ml_development_group,ml_production_group"
      OIDC_ADMIN_GROUP_NAME="mlflow-admins"
      OIDC_USERS_DB_URI="postgresql://mlflowdbuser:12345678@mlflow-
db.tech:5432/mlflow-users"
      DEFAULT_MLFLOW_PERMISSION="NO_PERMISSIONS"

      SSL_CERT_DIR=/etc/ssl/certs

      EOF
```

4. OIDC_GROUP_NAME - список групп из AD, пользователям которых доступен вход в MLflow в кач-ве пользователя
   OIDC_ADMIN_GROUP_NAME - список групп из AD, пользователям которых доступен вход в MLflow в кач-ве администратора
   DEFAULT_MLFLOW_PERMISSION - режим доступа по умолчанию NO_PERMISSIONS (пользователи не имеют доступа к рез-ам экспериментов других пользователей). Доступные варианты READ, EDIT, MANAGE

5. Конфигурационный файл с данными для подключения к БД для авторизации:

```
cat > /etc/mlflow/auth_conf.ini <<EOF
      [mlflow]
      # данные для подключения к БД mlflow-users
      database_uri       =       postgresql://mlflowdbuser:!QAZxsw2@mlflow-
db.tech:5432/mlflow-users
      EOF
```

6. Файл сервиса запуска MLflow (s3://**mlflow-bucket** - имя бакета в S3 для MLflow, параметр **-w** - значение vCPU***3** на ВМке с MLflow):

```
cat > /etc/systemd/system/mlflow-ts.service << EOF
      [Unit]
      StartLimitBurst=5
      StartLimitIntervalSec=33

      [Service]
      User=root
      EnvironmentFile=-/etc/default/mlflow-ts
      Environment="PATH=/opt/miniconda3/bin"
      WorkingDirectory=/opt
      Restart=always
      RestartSec=5
```

```
      ExecStart=/opt/miniconda3/bin/mlflow server       --backend-store-uri
postgresql://mlflowdbuser:!QAZxsw2@mlflow-db.tech:5432/mlflow-db --
artifacts-destination  s3://mlflow-bucket -w  8  --gunicorn-opts "--timeout
600" -h 0.0.0.0 -p 5001 --app-name oidc-auth
      KillMode=mixed

      [Install]
      WantedBy=multi-user.target
      EOF
```

7. Установить сертификат

```
cp _.k2int.tech.* /usr/local/share/ca-certificates/
      update-ca-certificates
```

8. Настроить garbage collector в cron

```
10    *    *    *    * /opt/miniconda3/bin/mlflow gc    --backend-store-uri
postgresql://mlflowdbuser:12345678@mlflow-db.tech:5432/mlflow-db --
artifacts-destination s3://mlflow-bucket
```

9. Запуск MLflow

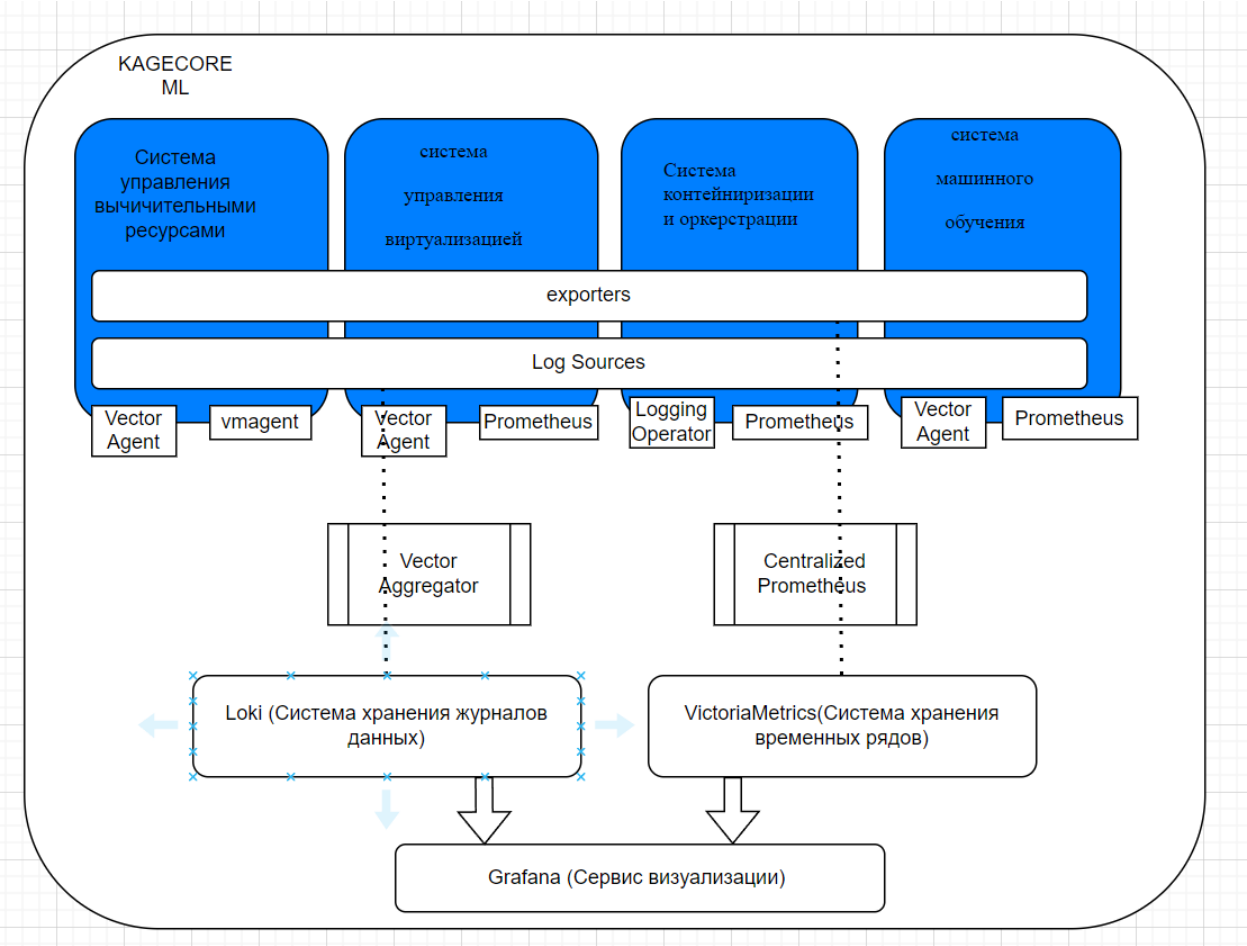
```
sudo systemctl daemon-reload

      sudo systemctl enable -now mlflow-ts
```

10. Доступ к MLFlow http://mlflow.tech:5001/.

# 5 Инфраструктурный мониторинг

## 5.1 Базовая схема взаимодействия компонентов мониторинга KageCore ML Platform




## 5.2 Добавление или изменение новых графиков в сервисе Grafana

Grafana позволяет визуализировать метрики, которые собираются Prometheus с помощью экспортеров, таких как GPU Exporter (например, NVIDIA DCGM Exporter или nvidia_gpu_exporter). Дашборды в Grafana состоят из панелей (panels), каждая из которых отображает определённые метрики в виде графиков, числовых индикаторов, таблиц и других визуальных элементов.

Ниже описан пример по созданию и настройке дашборда и панели на основе метрик GPU, собранных через Prometheus с помощью GPU Exporter.

## 5.3 Шаги создания дашборда и панели в Grafana

### 5.3.1 Добавление источника данных VictoriaMetrics (или другой TSDB)

1. Зайдите в веб-интерфейс Grafana (https://grafana.domain/).
2. В меню слева выберите Configuration (Настройки) → Data Sources (Источники данных) → Add data source (Добавить источник данных).
3. Выберите Prometheus.
4. В поле URL укажите адрес Prometheus, где доступны метрики GPU Exporter, например:
   http://<victoriametrics>:8428
5. Нажмите Save & Test для проверки подключения.

### 5.3.2 Создание нового дашборда

1. В меню слева выберите Create (Создать) → Dashboard (Дашборд).
2. Для нового дашборда нажмите Add new panel (Добавить новую панель).

### 5.3.3 Настройка панели

1. В панели редактирования перейдите на вкладку Query (Запрос).
2. В поле выбора источника данных выберите Prometheus.
3. Напишите PromQL-запрос для получения GPU метрик. Примеры метрик из nvidia_gpu_exporter:

   - Использование GPU в процентах:
     avg(nvidia_gpu_utilization_gpu) by (instance, gpu)
   - Температура GPU:
     avg(nvidia_gpu_temperature) by (instance, gpu)
   - Использование памяти GPU:
     avg(nvidia_gpu_memory_used_bytes) by (instance, gpu)

### 5.3.4 Выбор типа панели и визуализации

На правой панели выберите тип визуализации:

- Time series — график временных рядов.
- Stat — числовой индикатор текущих значений.
- Gauge — аналоговый индикатор, показывающий текущий уровень (полезно для процента загрузки).

- Bar gauge — столбчатый индикатор для сравнения значений.
- Table — табличный вид для отображения детальной информации.

Настройте отображение: легенды, цвета, пороги оповещений и др.

### 5.3.5 Определение переменных для дашборда (необязательно)

Для более удобной работы можно добавить переменные, чтобы переключать мониторинг между GPU, экземплярами, подами и т.п.

Например, переменная с именем instance с запросом:

label_values(nvidia_gpu_utilization_gpu, instance)

Это позволит в панели выбирать нужный сервер или ноду.

### 5.3.6 Сохранение дашборда и панели

1. После настройки панели нажмите Apply (Применить).
2. На странице дашборда вверху укажите название дашборда, например "GPU Monitoring".
3. Нажмите Save dashboard (Сохранить дашборд).

### 5.3.7 Пример базового PromQL-запроса в панели Grafana для NVIDIA GPU

Данный запрос выводит среднее использование GPU по каждому экземпляру и GPU:

avg(nvidia_gpu_utilization_gpu) by (instance, gpu)

### 5.3.8 Рекомендации

Вы можете импортировать готовые дашборды по ID, примеры импортируемых json вы можете найти в руководстве по установке.

## 5.4 Добавление сбора логов из файла журнала приложения в конфигурацию Vector

1. Для конфигурирования Vector с источником файловых логов (source file) откройте файл конфигурации vector:
   # /etc/vector/vector.toml

2. #Добавьте новый source:
   [sources.file_logs]
   type = "file"
   include = ["/var/log/myservice/*.log"] # Путь к логам, укажите актуальный
   ignore_older_secs = 86400 # Игнорировать файлы старше 1 дня
   #Добавьте трансформ для добавления необходимых меток (индексирование в Loki)
   [transforms.add_labels]
   type = "remap" # Используем remap transform для обработки событий
   inputs = ["file_logs"]
   source = '''
   .level = if .message =~ /(?i)error/ {
     "error"
   } else if .message =~ /(?i)warn(ing)?/ {
     "warning"
   } else if .message =~ /(?i)info/ {
     "info"
   } else {
     "unknown"
   }

   # Указываем статический лейбл сервиса, можно заменить значением или параметром
   .service = "myservice"

   # Формируем labels
   .labels = {
    service: .service,
    level: .level,
   }

   # По желанию можно выделить основное сообщение, если надо
   .message = .message
   # Добавьте новый трансформ в существующий Sync или сконфигурируйте новый
   [sinks.loki]
   type = "loki"

```
inputs = ["add_labels"]
endpoint = "http://localhost:3100/loki/api/v1/push"
encoding.codec = "json"
labels.service = "{{ service }}"
labels.level = "{{ level }}"
healthcheck.enabled = true
```

3. Сохраните изменения и протестируйте конфигурацию на корректность:
   vector test --config vector.toml
4. Если тест пройден, перезапустите сервис vector:
   Systemctl restart vector.service

Пояснения:

- sources.file_logs — читает логи из файлов по указанному пути. Можно настраивать include для точного указания местоположения логов.
- transforms.add_labels — с помощью языка remap (VRL) анализирует поле .message и пытается определить уровень лога (error, warning, info), добавляет статический лейбл сервиса .service.
- Лейблы формируются в поле .labels, чтобы Loki правильно индексировал логи.
- sinks.loki — отправляет данные в Loki по HTTP API в JSON формате, используя добавленные в трансформации лейблы.

## 5.5 Изменение сроков хранения метрик в tsdb VictoriaMetrics

Ограничения времени хранения метрик в VictoriaMetrics используется параметр командной строки -retentionPeriod. Этот параметр задаёт период хранения данных на диске и указывается в формате длительности, например, в днях, неделях или месяцах.

## 5.6 Основные детали по настройке времени хранения метрик

Флаг для установки времени хранения:
-retentionPeriod=<duration>
Форматы длительности, поддерживаемые VictoriaMetrics:
s (секунды), m (минуты), h (часы), d (дни), w (недели), y (годы)
Например:

- -retentionPeriod=30d — хранить данные 30 дней.
- -retentionPeriod=12w — хранить 12 недель.

- -retentionPeriod=1y — хранить 1 год.

Минимальное значение — 1 день (1d), максимальное — до сотен лет (например, 100y), хотя фактически для очень долгого хранения могут использоваться другие архитектурные подходы.

Все метрики, данные по которым старше указанного периода, автоматически удаляются системой (на уровне vmstorage) во время выполнения.

1. Чтобы применить ограничение, нужно указать этот параметр при старте компонента VictoriaMetrics, отредактировав system unit:
   lib/systemd/system/victoria-metrics.service
2. Измените значение -retentionPeriod=<duration>
   Сохраните изменение и выполните команду:
   systemctl daemon-reload
   затем
   systemctl restart Victoria-metrics.service
3. Проверьте состояние сервиса после изменений:
   journalctl -xe ( нет ошибок связанных с Victoria-metrics)
   systemctl status Victoria-metrics.service
4. Состояние должно быть Active: active (running).

### 5.6.1 Изменение сроков хранения метрик в хранилище журналов данных Loki

5.6.1.1 Настройка периода хранения

Период хранения настраивается в разделе конфигурации limits_config.

Существует два способа задания политик хранения:

1. retention_period, которая применяется глобально ко всем потокам логов.
2. retention_stream, которая применяется только к потокам логов, соответствующим селектору.
   Примечание
   Минимальный период хранения составляет 24 часа.

В данном примере настраивается глобальное хранение, которое применяется ко всем тенантам (если не переопределено настройкой для каждого тенанта):

```
yaml
...
limits_config:
  retention_period: 744h
  retention_stream:
  - selector: '{namespace="dev"}'
    priority: 1
```

```
    period: 24h
  per_tenant_override_config: /etc/overrides.yaml
...
```

Примечание

Вы можете использовать только сопоставители меток в поле selector определения retention_stream. Произвольные выражения LogQL не поддерживаются.

Хранение для каждого тенанта можно определить путем настройки динамических переопределений. Например:

yaml

```
overrides:
  "29":
    retention_period: 168h
    retention_stream:
    - selector: '{namespace="prod"}'
      priority: 2
      period: 336h
    - selector: '{container="loki"}'
      priority: 1
      period: 72h
  "30":
    retention_stream:
    - selector: '{container="nginx", level="debug"}'
      priority: 1
      period: 24h
```

Период хранения для данного потока определяется на основе первого совпадения в этом списке:

Если несколько селекторов retention_stream для каждого тенанта соответствуют потоку, выбирается период хранения с наибольшим приоритетом.

Если несколько глобальных селекторов retention_stream соответствуют потоку, выбирается период хранения с наибольшим приоритетом. Это значение не учитывается, если установлено retention_stream для каждого тенанта.

Если указан retention_period для каждого тенанта, он будет применен.

Глобальный retention_period будет применен, если ни одно из вышеперечисленных условий не выполнено.

Если глобальный retention_period не указан, используется значение по умолчанию — 744 часа (30 дней).

Примечание

Чем выше значение приоритета, тем выше приоритет.

Сопоставление потоков использует тот же синтаксис, что и сопоставление меток Prometheus:

=: Выбирает метки, которые точно равны предоставленной строке.

!=: Выбирает метки, которые не равны предоставленной строке.

=~: Выбирает метки, которые соответствуют регулярному выражению, предоставленному строкой.

!~: Выбирает метки, которые не соответствуют регулярному выражению, предоставленному строкой.

Примеры конфигураций, указанные выше, приведут к следующим периодам хранения:

1. Для тенанта 29:

   Потоки, имеющие метку namespace prod, будут иметь период хранения 336 часов (2 недели), даже если метка container имеет значение loki, поскольку приоритет правила prod выше.

   Потоки, имеющие метку container loki, но не находящиеся в пространстве имен prod, будут иметь период хранения 72 часа.

   Для остальных потоков этого тенанта применяется значение retention_period, указанное для каждого тенанта, равное 168 часам.

2. Для тенанта 30:

   Потоки, имеющие метки nginx и level debug, будут иметь период хранения 24 часа.

   Для остальных потоков этого тенанта применяется глобальный период хранения 744 часа, так как никакое переопределение не указано.

3. Все тенанты, кроме 29 и 30:

   Потоки, имеющие метку namespace dev, будут иметь период хранения 24 часа.

   Потоки, не имеющие метки namespace dev, будут иметь период хранения 744 часа.

Пример изменения Глобального периода хранения:

1. Откройте на редактирование файл конфигурации Loki /etc/loki/config.yml.
2. Измените значение retention_period в секции limits_config:

   limits_config:

   retention_period: <часы>h
3. Сохраните изменения и перезапустите сервис Loki:

   systemctl restart loki.service

4. Проверьте, что Loki корректно стартовал:
   journalctl -xe (нет ошибок связанных с loki)
   systemctl status loki.service
5. Состояние должно быть Active: active (running)

## 5.7 Добавление Scrape Target в Prometheus

1. Для добавления новой цели в Prometheus необходимо отредактировать или создать новый конфигурационный файл в директории /etc/prometheus/scrape_configs/
2. Добавление нового Job
   В секцию scrape_configs добавьте новый job. Пример для сбора метрик с Node Exporter:

```
node_exporter.yml
 scrape_configs:
  - job_name: 'node'       # Уникальное имя job
    scrape_interval: 15s           # Опционально: интервал сбора
    static_configs:
      - targets: ['localhost:9100'] # Адреса целей
        labels:                    # Дополнительные лейблы
          env: 'production'
          role: 'database_server'
```

Параметры:

job_name: Уникальный идентификатор группы целей

scrape_interval: Переопределение глобального интервала сбора (по умолчанию 15с)

static_configs: Для статических IP/DNS-адресов

targets: Список эндпоинтов в формате host:port

labels: Произвольные метки для группировки/фильтрации

3. Пример:

```
scrape_configs:
  - job_name: node
    static_configs:
      - labels:
          job: node
          instance: mon-visual
        targets:
          - 172.1.1.100:9100
```

4. Проверка конфигурации:
5. Выполните проверку перед перезагрузкой:
   promtool check config /etc/prometheus/prometheus.yml
6. Перезагрузка Prometheus
   Systemd:
   sudo systemctl reload Prometheus.service
7. Проверка работы
   Перейдите в веб-интерфейс: http://<prometheus-ip>:9090/targets
8. Убедитесь, что новый target в состоянии UP.

## 5.8 Добавление Alert в Prometheus Alertmanager для отправки уведомлений

### 5.8.1 Создание правила алертинга в Prometheus

1. Создайте файл с правилами.
   Создайте новый YAML-файл с правилами (например, /etc/prometheus/rules/node_alerts.yml):

```
yaml
groups:
- name: node-alerts
  rules:
  - alert: HighMemoryUsage
    expr: (node_memory_MemTotal_bytes - node_memory_MemAvailable_bytes) / node_memory_MemTotal_bytes * 100 > 90
    for: 5m
    labels:
      severity: critical
      team: infra
    annotations:
      summary: "High memory usage on {{ $labels.instance }}"
      description: "Memory usage is {{ $value | humanize }}% for 5 minutes. Instance: {{ $labels.instance }}"
      dashboard: "https://grafana.example.com/d/abcd1234"
```

2. Параметры правила
   - alert: Уникальное имя алерта
   - expr: PromQL-выражение для срабатывания
   - for: Длительность условия перед срабатыванием
   - labels: Произвольные метки для маршрутизации
   - annotations: Детали для уведомлений (поддерживают шаблонизацию)

3. Подключите файл правил в prometheus.yml
   yaml

```
rule_files:
  - "rules/*.yml"  # Путь к вашим правилам

alerting:
  alertmanagers:
  - static_configs:
      - targets: ['alertmanager:9093']  # Адрес Alertmanager
```

### 5.8.2 Настройка Alertmanager

1. Конфигурация Alertmanager (alertmanager.yml)/
   yaml

```
route:
  group_by: [alertname, cluster]  # Группировка уведомлений
  group_wait: 30s
  group_interval: 5m
  repeat_interval: 4h
  receiver: 'slack-infra-team'

  routes:  # Дополнительные маршруты
  - match:
      severity: critical
    receiver: 'pagerduty'

receivers:
- name: 'slack-infra-team'
  slack_configs:
  - api_url: "https://hooks.slack.com/services/XXXX/YYYY/ZZZZ"
    channel: "#alerts"
    text: "{{ range .Alerts }}<!channel> {{ .Annotations.summary }}\n{{ .Annotation
s.description }}\n{{ end }}"

- name: 'pagerduty'
  pagerduty_configs:
  - routing_key: "your-pagerduty-key"
```

2. Проверка конфигурации

```
amtool check-config alertmanager.yml
```

### 5.8.3 Запуск/перезагрузка сервисов

Для Prometheus:

```
# Проверка конфига
```

```
promtool check rules /etc/prometheus/rules/node_alerts.yml
promtool check config /etc/prometheus/prometheus.yml

# Перезагрузка
sudo systemctl reload Prometheus.service
```

Для Alertmanager:

```
sudo systemctl reload alertmanag
```

# 6 KageCore ML Platform. Модуль витрины сервисов

## 6.1 Процесс создания ссылочного продукта

### 6.1.1 Переход в панель администратора

1. Откройте админ-панель управления продуктами по адресу: https://product-manager.<base-domain>/kit/products.
2. Нажмите кнопку **"Создать".**

### 6.1.2 Заполнение основной информации

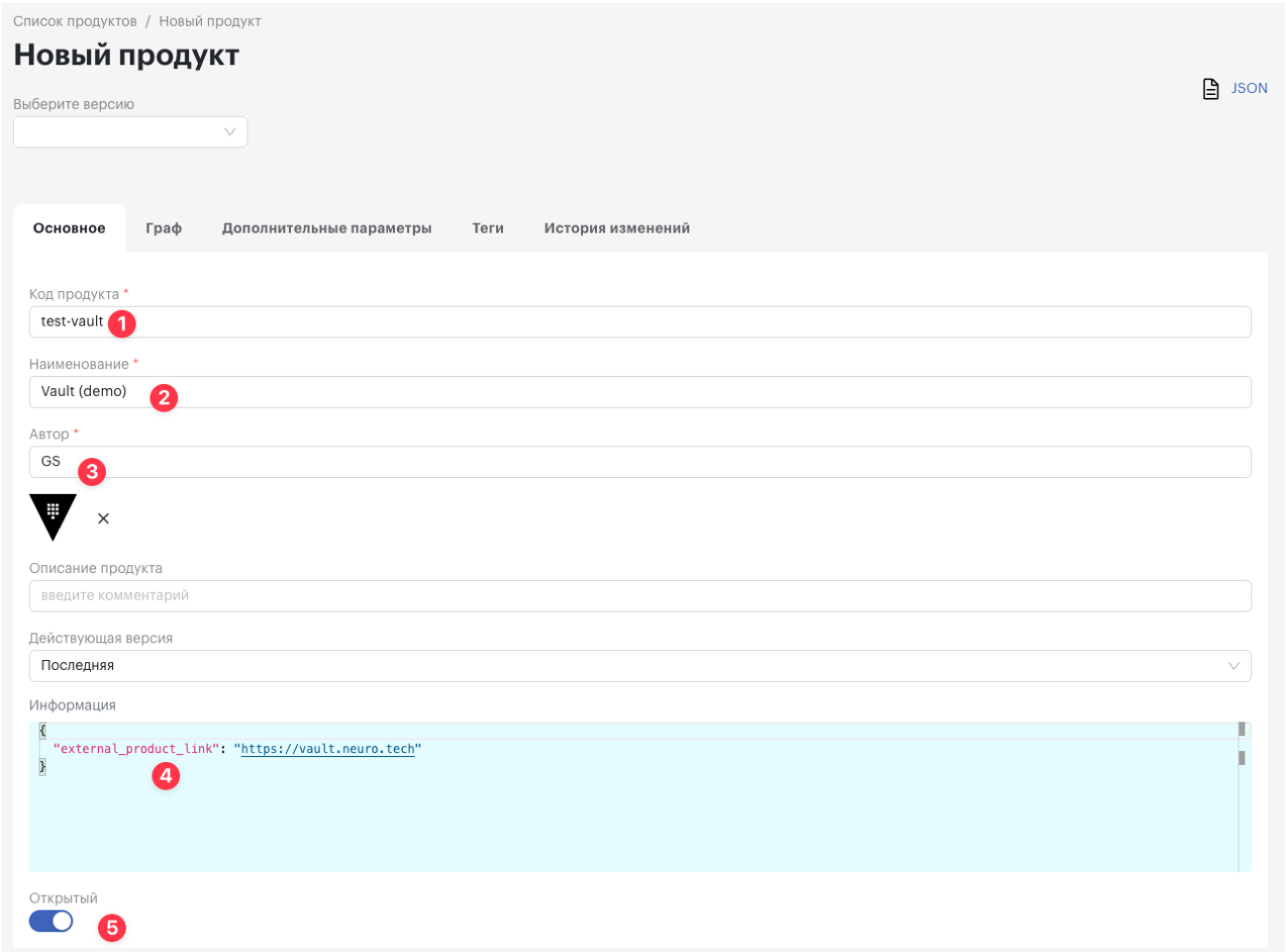
Заполните следующие обязательные поля:

- **Код продукта** - уникальный идентификатор продукта в системе
- **Наименование** - отображаемое название продукта на плитке
- **Автор** - информация о создателе продукта
- **Логотип** (опционально) - изображение, которое будет отображаться на плитке

**Важно:** В поле **"Информация"** необходимо добавить JSON-конфигурацию со ссылкой на внешнюю систему:

```
{
  "external_product_link": "https://vault.tech"
}
```

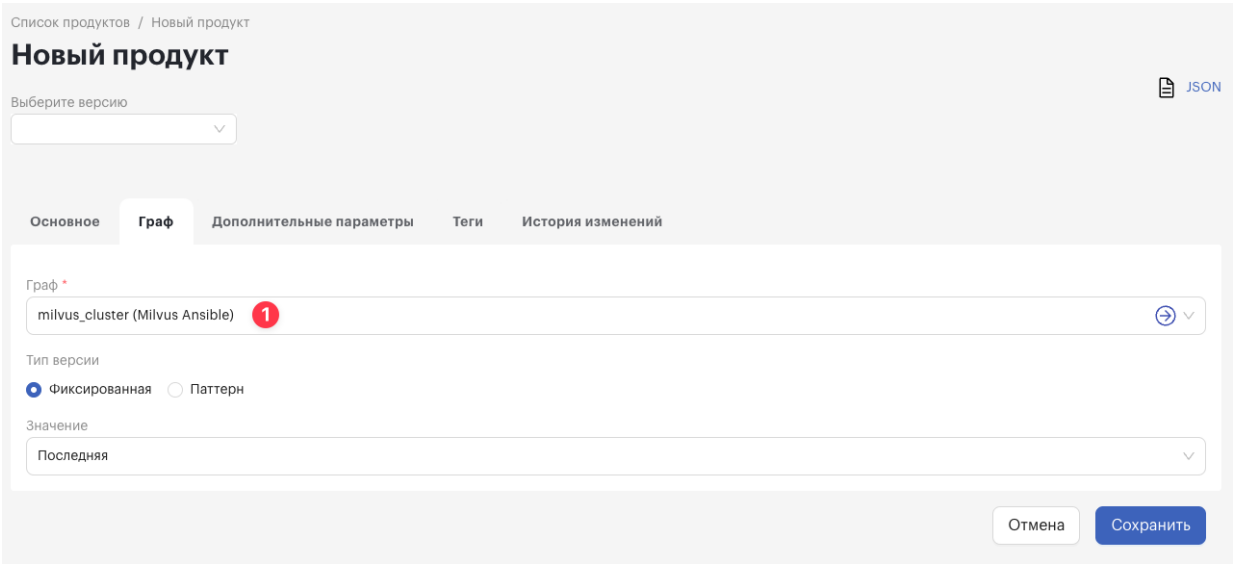
Включить переключатель "**Открытый**" для отображения в списке продуктов.




### 6.1.3 Настройка графа

1. Перейдите на вкладку **"Граф"**.
2. Выберите граф любого существующего продукта.

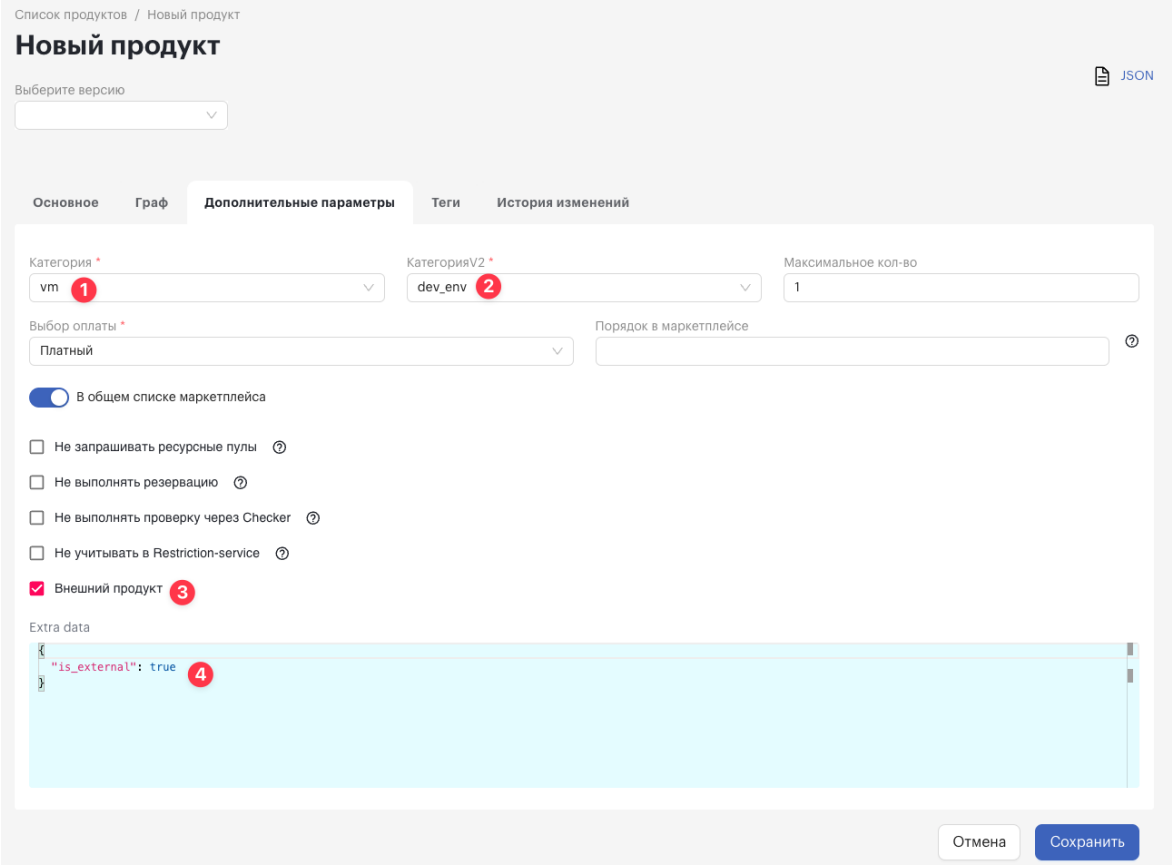
**Примечание:** Выбор графа является технической необходимостью для корректной регистрации продукта в системе. Сам граф не используется для ссылочных продуктов.




### 6.1.4 Дополнительные параметры

1. Перейдите на вкладку **"Дополнительные параметры"**
2. В поле **"Категория"** выберите **"vm"**
3. В поле **"КатегорияV2"** выберите **"dev_env"**
4. Установите галочку рядом с полем **"Внешний продукт"**
5. В поле **"Extra data"** добавьте следующий JSON:

```
{
        "is_external": true
      }
```




### 6.1.5 Сохранение и проверка

1. Сохраните созданный продукт
2. После успешного создания на главной странице портала отобразится новая плитка с созданным продуктом
3. При нажатии на плитку пользователь будет перенаправлен на указанный в конфигурации URL

### 6.1.6 Подключение Active Directory через Keycloak и синхронизация пользователей

ОС Windows 10 не позволяет создавать Active Directory. Если у вас установлена ОС Windows 10, то рекомендуется использовать Windows Server на виртуальной машине для создания Active Directory.

6.1.6.1 Создание нового realm

При первой регистрации в Keycloak в нем существует только master-область, поэтому нужно создать новую область для дальнейших действий. Для этого:

1. В верхнем левом углу нажмите Add realm.

2. Введите произвольное название, например, "active-directory" и нажмите Create.
3. Слева в панели выберите Realm Settings и заполните параметры нового realm:
    a. На вкладке General введите:
        - Display name — Active Directory.
    b. На вкладке Themes введите:
        - Login theme — KAGECORE ML PLATFORM.
        - Email theme — KAGECORE ML PLATFORM.
    c. На вкладке Localization введите:
        - Internationalization — Enabled
        - Supported locales — русский.
        - Default locale — русский.
    d. Нажмите Save.
    e. Скопируйте ссылку на OpenID Endpoint Configuration, она пригодится для дальнейшей настройки.
    Пример: https://auth.example.com/auth/realms/active-directory/.well-known/openid-configuration
    f. В панели слева выберите User federation и нажмите Add LDAP Provider.
    g. Укажите следующие параметры для федерации пользователей:
        - Console display name — active-directory
        - Vendor — Active Directory
        - Connection URL — ldap://<active-directory-ip-or-dns>
        - Bind type — simple
        - Bind DN — CN=dadmin,CN=Users,DC=test,DC=ad
        - Bind credentials — <password>
        - Edit mode — READ_ONLY
        - Users DN — CN=Users,DC=test,DC=ad
        - Trust email — On
    h. Завершите настройку, нажав Save.
    i. После ввода параметров проверьте подключение с помощью Test connection иTest authentication.
    j. По завершении настройки нажмите Sync all users и убедитесь в отсутствии ошибок.
    k. Переключитесь на realm "Portal" и добавьте новый Identity provider с параметрами:
        - Type — Keycloak OpenID Connect
        - Alias — active-directory
        - Display name — Active Directory
        - Discovery endpoint — введите сохранённую ранее ссылку наOpenID Endpoint Configuration
        - Client ID — active-directory
        - Client Secret — введите сохранённый ранееClient secret.
    l. Сохраните изменения.
    m. Повторно откройте настройки подключения и установите:
        - Trust Email — On
        - Sync mode — Force
    n. Нажмите Save.

6.1.6.2 Проверка синхронизации Keycloak с Active Directory

1. В интерфейсе Keycloak в левой панели управления нажмите Clients. Нажмите на URL-адрес консоли: В Keycloak Account Manager в правом верхнем углу нажмите Вход.
2. В появившемся окне введите учетные данные любого пользователя из Active Directory.
3. После успешной аутентификации вы попадете в аккаунт созданного пользователя и увидите его профиль.

### 6.1.7 Взаимодействие с пользовательским интерфейсом

Пользователи с правами администраторов по умолчанию имеют полные права на все действия в пользовательском интерфейсе. По необходимости, полномочия администраторов можно скорректировать.

Ниже приведены базовые возможности администратора портала в пользовательском интерфейсе.

Описанные далее процедуры выполняются на портале KAGECORE ML PLATFORM

6.1.7.1 Настройка учетной записи

Портал позволяет предоставлять доступ другим пользователям к папкам и проектам.

6.1.7.1.1 Добавление пользователя к организации/папке/проекту

Во вкладке Управление доступом можно добавить пользователя к нужному контексту.

Порядок действий

1. Активируйте необходимый контекст соответствующей организации
2. Перейдите на экран Управление доступом → Учетные записи
3. Нажмите +
4. В появившемся окне:
    - Выберите одного или нескольких пользователей в поле Пользователь. Для поиска пользователя начните набирать его имя
    - В поле Роли из выпадающего меню выберите роли, которые хотите назначить пользователю для этой организации/папки/проекта
5. Нажмите Добавить

Роли в каталоге создаются пользователями с правами администратора согласно ролевой модели организации.

При успешном предоставлении прав доступа к папке/проекту на странице с орг. структурой отобразится соответствующее уведомление.

Пользователи портала получают роль в рамках контекста - на уровне организации, папки или проекта.

При получении роли на проект, действия, доступные данной роли, пользователь может выполнять только в этом проекте. Если часть действий, доступных роли, можно совершать только на уровне организации - пользователь не сможет их совершить. Если роль получена на папку или организацию, то доступные действия пользователь может совершать в выбранной папке, а также во всех дочерних папках и проектах вниз по дереву. Причем не только в тех, которые существовали на момент создания роли, но и во всех вновь созданных.

6.1.7.1.2 Редактирование прав пользователя

При необходимости права пользователя в контексте можно изменить.

Порядок действий

1. Активируйте необходимый контекст соответствующей организации
2. Перейдите на экран Управление доступом → Учетные записи
3. Напротив нужного пользователя в списке нажмите ⋮и выберите Редактировать
4. В появившемся окне добавьте и/или удалите необходимые роли
5. Нажмите Применить

6.1.7.1.3 Отзыв прав пользователя

1. Порядок действий
2. Активируйте необходимый контекст соответствующей организации
3. Перейдите на экран Управление доступом → Учетные записи
4. Напротив нужного пользователя в списке нажмите ⋮и выберите Отозвать права

После выполнения операции пользователь будет уделён из списка пользователей организационной единицы.

6.1.7.2 Настройка организационной структуры

6.1.7.2.1 Организационная структура

В Организационной структуре пользователь может создавать папки, вложенные папки и проекты.

Доступ пользователей к проектам можно разграничивать на уровне папок.

Уместно создавать папки разного уровня вложенности в соответствии с активностями организационных подразделений или групп пользователей из разных подразделений, работающих в рамках одного проекта.

Проект – это информационная система с выбранной для нее средой. Непосредственно заказ продуктов доступен только на уровне Организационной структуры «Проект».

При формировании Организационной структуры необходимо учитывать следующие ограничения:

- уровней вложенных папок - не более 8
- папок на одном уровне - не более 50
- проектов на одном уровне - не более 15

Создание папок в Организационной структуре необходимо для разграничения прав доступа пользователей к проектам.

Выбрать Организацию, папку или проект можно следующими способами:

Способ 1 (выбор папки/проекта любой организации)

1. Находясь на любом экране, перейдите в контекст и выберите организацию.
2. В иерархическом списке найдите нужную папку/проект.
3. Кликните левой кнопкой мыши по нужной папке/проекту
4. При этом происходит переход на экран Организационная структура выбранной организации. Текущей станет папка/проект, выбранные в пункте 3.

Способ 2 (выбор папки/проекта текущей организации)

1. Находясь на любом экране, перейдите в контекст и выберите организацию.
2. На экране Управление доступом - Организация в иерархическом списке найдите нужную папку/проект.
3. Напротив необходимого объекта нажмите иконку ⋮и выберите команду Выбрать контекст.

Выбранный объект отобразится в верху экрана.

6.1.7.2.2 Настройка организационной структуры

Создание и настройка организационной структуры доступны только пользователю с соответствующими правами.

6.1.7.2.2.1 Создание папки/проекта

Управление папками и проектами осуществляется с экрана Управление орг. стурктурой или через вкладку меню Управление доступом → Организация.

Порядок действий

1. Находясь на любом экране, перейдите в контекст и выберите организацию.
2. Нажмите кнопку Управление орг. структурой
3. Нажмите на иконку ⋮напротив объекта, в котором хотите создать папку/проект.

4. Нажмите Создать папку или Создать проект
5. В появившемся окне введите необходимые данные
    a. Для новой папки:
        - Введите Название
        - При необходимости отключите опцию Создать собственный счет
        - Нажмите Сохранить и продолжить
        - Выберите одну или более Информационные системы из списка
        - Нажмите Создать
    b. Для нового проекта:
        - Введите Наименование
        - Из списка выберите нужную Информационную систему
        - Из списка выберите нужную Среду
        - Нажмите Создать

Созданная папка/проект отобразится в Организационной структуре

#### 6.1.7.2.2.2 Редактирование папки/проекта

При необходимости пользователь может внести изменения в наименование папки или проекта.

Порядок действий

1. Перейдите на экран Управление орг. структурой соответствующей организации.
2. Нажмите на иконку ⋮напротив необходимой папки/проекта.
3. В появившемся меню выберите пункт Редактировать.
4. В открывшейся форме внесите изменения в поле Наименование и примените изменение.

#### 6.1.7.2.2.3 Удаление папки/проекта

Пользовать с соответствующими правами может удалить папку или проект.

Удалять можно только те папки, которые не содержат дочерних элементов (папок/проектов).

Порядок действий

1. Перейдите на экран Перейдите на экран Управление орг. структурой соответствующей организации.
2. Нажмите на иконку ⋮напротив необходимой папки/проекта.
3. Скопируйте идентификатор объекта из уведомления
4. В появившемся меню выберите пункт Удалить.
5. Скопируйте идентификатор объекта из уведомления.
6. Подтвердите операцию удаления.

# 7 KageCore ML Platform. Модуль тарификации

## 7.1 Пополнение счета

Для пополнения счета конкретной организации:

1. Перейдите в раздел **Биллинг → Пополнение счёта**.
2. Выберите организацию, счёт которой необходимо пополнить.
3. Выберите из списка **Счёт отправитель** и **Счёт получатель**.
4. Укажите сумму перевода.
5. Опционально введите описание перевода.
6. Нажмите **Подтвердить**.

В меню также доступен просмотр журнала переводов.

## 7.2 Тарифные классы

Тарифный класс описывает стоимость ресурса, доступного для предоставления определенному продукту при заказе. Например, это может быть стоимость оперативной памяти, которую нужно предоставить при заказе виртуальной машины.

Несколько тарифных классов логически объединяются в тарифные планы.

## 7.3 Создание нового тарифного класса

Тарифный класс можно создать двумя способами: с помощью копирования тарифного класса или создания нового класса в интерфейсе.

| Если тарифный класс уже используются, то его нельзя редактировать. |
|---|

Для создания нового тарифного класса:

1. Перейдите в **Биллинг → Тарифные классы →** нажмите **+**.
2. Заполните поля в окне **Создание тарифного класса**:

| Поле | Описание | Значения и примеры |
|---|---|---|
| **Кодовое имя** | Уникальное название класса тарифа. Может содержать только символы: "a-z", "0-9", "_", "-", ":", "." | Например, "tariff_CPU" |
| **Описание** | Символьное описание тарифного класса, содержать символы: "A-Z","a- | Например, "тариф для расчета CPU" |

| Поле | Описание | Значения и примеры |
|---|---|---|
| | z","А-Я", "а-я", "0-9", "_", "-", ":", ".","," "." | |
| **Единица измерения** | Единица, за которую рассчитывается стоимость по тарифу. | - шт<br><br>- ГБ<br><br>- количество запросов put,post<br><br>- количество запросов get,head |
| **Единица измерения времени** | Время, за которое рассчитывается стоимость тарифа: **per_minute** — в минуту, **package** — за количество запросов, если в поле **Единица измерения** выбрано значение "количество запросов". | - per_minute<br><br>- package |
| **Тип элемента** | Тип элемента в items, к которому применяется тариф. **vm** — виртуальная машина, **app** — кластер. | - vm<br><br>- app |
| **Расчетный объект** | Расчетным объектом является тот объект в items, к которому применится тариф. **current** — текущий объект, **child** — дочерний объект от текущего объекта. Также может быть выбран родительский объект. | - current<br><br>- child |
| **Статусы элемента** | Статусы расчетного объекта, к выбранным статусам будет применяться тариф. | - On<br><br>- Off |

| Поле | Описание | Значения и примеры |
|---|---|---|
| | | - Reboot<br><br>- Deleted<br><br>- Problem |
| **Путь до множителя** | Путь в системе, где задан определенный множитель. Каталоги, в которых расположен множитель вводятся через запятую. Название множителя и его расположение в структуре item можно посмотреть в интерфейсе **control-panel**, разделе **Конструктор → Справочники → item_params**. | Например, data,config,flavor,cpu — для расчета тарифа по CPU, или data,config,boot_disk,size — для расчета тарифа по жесткому диску, или data,config,flavor,memory — для расчета тарифа по оперативной памяти и т.д. |
| **Добавочный множитель** | Добавочный множитель позволяет добавить к основной конфигурации тарифа дополнительное значение, на которое будет умножаться цена, полученная в результате расчета всех значений тарифа. | Например, "1" |
| **Скидка** | Опция активируется, если нужно добавить скидку в тариф. Скидки объединяются по тэгам, которые можно посмотреть и задать новые в интерфейсе **control-panel**, разделе **Конструктор** | Например, "ram_vdc_pay_as_you_go" |

| Поле | Описание | Значения и примеры |
|---|---|---|
| | **→ Справочники → item_params → tags** | |
| **Дополнительные параметры** | Поле позволяет добавить в тариф расширенные параметры, например, расчет стоимости в зависимости от выбранного провайдера.<br><br>Просмотреть структуру item и названия параметров можно в интерфейсе **Control-panel**, разделе **Конструктор → Справочники → item_params → item_structure**. | Пример кода:<br><br>`{ "provider": "redvirt", "provider_path": [ "data", "provider" ] }` |
| **Код** | Укажите тело тело функции в формате JavaScript для расчета множителя. Код нужно заключить в блок скобок {}, где:<br><br>- item — текущий элемент.<br><br>- items — массив всех элементов.<br><br>Код обязательно должен возвращать значение либо выражение с помощью return. | Пример: `{ var sum = item.data.config.extra_disks.reduce((accumulator, object) ⇒ { return accumulator + object.size; }, 0); return sum; }` |

Путь до множителя и дополнительные параметры являются системными и при создании собственных тарифных классов, необходимо их сравнивать с аналогичными в созданных по умолчанию тарифными классами.

3. Нажмите кнопку **Добавить**.

### 7.3.1 Импорт тарифных классов

В разделе **Тарифные планы организаций** можно импортировать файл с тарифными классами. Для этого:

1. Перейдите в **Биллинг → Тарифные планы организаций**.
2. Нажмите на название ТПО. Вы перейдете к странице с подробной информацией о тарифном плане организации.
3. Нажмите **Импорт** и загрузите файл с массивом данных о ТК в формате JSON. Эти данные обычно выгружаются из БД, в них можно указать стоимость своих продуктов или изменить имеющуюся. После этого данные можно загрузить в ТК.

И при загрузке файла создается задача на импорт. Состояние этой задачи можно просмотреть в разделе **Задачи на импорт**.

## 7.4 Тарифные планы организаций

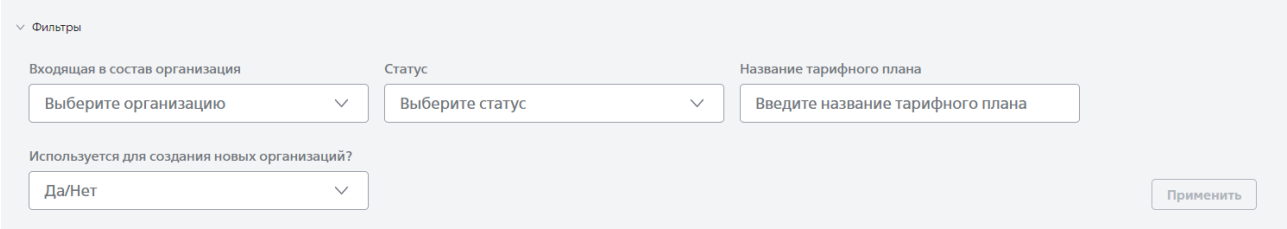
**Тарифный план** (далее - ТП) — это описание стоимости единицы товара для конкретной организации, папки или проекта.

**Тарифный план организации** (далее – **ТПО**) — это верхнеуровневый тарифный план для целой организации или нескольких организаций.

На экране **Тарифные планы организаций** отображается список существующих тарифных планов.

Для удобства поиска по тарифным планам организаций доступны фильтры, которые можно раскрыть над списком тарифных планов. Фильтры включают в себя поиск по:

- входящей в состав организации.
- статусу — "Активный", "Черновик", "Планируемый", "Архивный".
- названию тарифного плана
- использованию для создания новых организаций (тарифный план может быть использован при создании новых организаций).




### 7.4.1 Создание тарифного плана

Для одной организации должен существовать как минимум один активный тарифный план.

Для создания тарифного плана:

1. Перейдите в **Биллинг → Тарифные планы организаций**.
2. Нажмите **+**.
3. Введите название для нового тарифного плана.
4. Выберите из списка организацию.
5. Если тарифный план нужно использовать по умолчанию, то активируйте опцию **Использовать как дефолтный**.
6. Нажмите **Создать**.
7. Нажмите → для перехода в параметры тарифного плана.
8. Нажмите **+ Добавить** для добавления тарифных классов.
9. Вернитесь к списку тарифных планов → нажмите ⋮ напротив нужного тарифного плана и выберите **Активировать**.

### 7.4.2 Запланированная активация тарифного плана

Тарифный план организации (ТПО) можно не активировать сразу после создания, а отложить активацию до подходящего момента.

Для этого существует функция — **Запланировать**. С помощью этой функции можно отложить активацию нового тарифного плана. Также перевести существующие заказы на новые тарифы в конкретном тарифном плане организации, если активирована функция **Применить к текущим заказам**.

Для запланированной активации тарифного плана:

1. Перейдите в **Биллинг → Тарифные планы организаций**.
2. Напротив нужного ТПО нажмите ⋮ и выберите **Запланировать**.
3. Укажите:
   - **Дата** — выберите из календаря дату активации.
   - **Время** — задайте время активации. Оно должно быть минимум на 20 минут больше текущего.
   - **Применить к текущим заказам** (опционально) — активируйте функцию, если необходимо.
4. Нажмите **Запланировать**.

Если функция “Применить к текущим заказам” неактивна, то существующие заказы будут рассчитываться по старым тарифам из прошлого тарифного плана, а новые заказы — по новым тарифам. Таким образом в одной организации может быть несколько тарифных планов.

### 7.4.3 Просмотр информации о тарифном плане

При нажатии на строку ТП в списке, открывается подробное представление.

В подробном представлении отображается информация о сроке действия и статусе ТП, услугах и тарифных планах, связанных с выбранным ТП.

### 7.4.4 Копирование тарифного плана

Если необходимо создать новый тарифный план с параметрами существующего, можно использовать процедуру копирования.

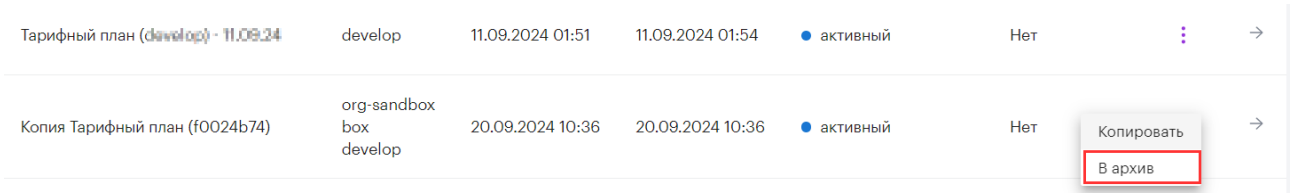
Для копирования тарифного плана:

1. Перейдите в **Биллинг → Тарифные планы организаций**.
2. Нажмите ⋮ напротив нужного ТПО и выберите **Копировать**.
3. Введите название для нового ТПО.
4. Нажмите **Создать**.

### 7.4.5 Удаление тарифного плана

Для удаления тарифного плана:

1. Перейдите в **Биллинг → Тарифные планы организаций**.
2. Нажмите ⋮ напротив нужного ТП и выберите **В архив**.




Статус тарифного плана изменится на "Архивный".

3. Нажмите ⋮ напротив нужного ТП и выберите **Удалить**.

Внутри тарифного плана организации также есть кнопки **В архив**, **Удалить**, **Активировать**, **Запланировать**, **В черновик**. Их наличие зависит от статуса тарифного плана. Для перехода к подробной информации о тарифном плане организации нажмите на ее название.

# 8 KageCore ML Platform. Модуль пользовательского мониторинга

Мониторинг стек – это система мониторинга и алертинга, построенная на базе современных open-source решений. Система включает сбор логов, метрик и алертинг, обеспечивая полноценный мониторинг стек для ИТ-инфраструктуры.

Пользовательский мониторинг настраивается в зависимости от требований конечного потребителя и является кастомизируемым ресурсом.