

Arenadata™ Streaming

Версия - v1.6-RUS

Инструкция по установке

Оглавление

1	Загрузка дистрибутива платформы	3
2	Предварительные действия	6
3	Загрузка бандла ADS	8
4	Создание кластера ADS	11
4.1	Создание экземпляра кластера	12
4.2	Конфигурация кластера	12
4.3	Добавление хостов	14
4.4	Добавление сервисов	16
4.5	Размещение компонентов сервисов на хостах	19
4.6	Установка кластера	19

Кластер ADS устанавливается при помощи Arenadata Cluster Manager, описание которого приведено на сайте: [ADCM](#). При этом обеспечивается следующее:

1. Вся логика по установке и конфигурированию находится внутри бандла ADS:
 - Пользователю нет необходимости вручную вводить множество команд в консоли;
 - Сконфигурированы все важные настройки ОС и сервисов.
2. Единый интерфейс доступа и управления конфигурациями:
 - Возможность управления несколькими сервисами через единый интерфейс;
 - Возможность управления несколькими кластерами через единый интерфейс.
3. Возможность использования как облачной, так и физической инфраструктуры;
4. Все rpm-пакеты для ADS берутся из репозитория Arenadata:
 - Все rpm протестированы;
 - Используется фиксированный набор пакетов в помощь пользователям и для составления баг-репортов.

Important: Контактная информация службы поддержки – e-mail: info@arenadata.io

Глава 1

Загрузка дистрибутива платформы

Для загрузки дистрибутива **Arenadata Streaming** необходимо на сайте arenadata.io выбрать вкладку “Скачать”. При этом происходит переход на страницу *Магазин Программного Обеспечения Arenadata* (store.arenadata.io, Рис.1.3), и открывается экранная форма заявки для загрузки продукта (Рис.1.1.).

Заполните, пожалуйста, форму для продолжения загрузки

ФИО:

E-Mail:

Компания:

Цель использования:

Отправить

Рис.1.1.: Заявка для загрузки продукта

В экранной форме заявки следует заполнить следующие поля:

- *ФИО* – ФИО пользователя;
- *E-Mail* – адрес электронной почты пользователя;
- *Компания* – наименование компании пользователя;
- *Цель использования* – описание цели использования платформы.

Все поля являются обязательными для заполнения. После ввода данных необходимо нажать кнопку *Отправить*, в результате чего выдается соответствующее сообщение (Рис.1.2.).

При закрытии окна происходит переход на страницу *Магазин Приложений* с перечнем продуктов **Arenadata** (Рис.1.3.).

Спасибо! Вы можете теперь закрыть окно и продолжить.



Рис.1.2.: Заявка принята

ARENADATA Продукты Документация Поддержка Утилиты Пользователь

Магазин Приложений

В данном разделе вы можете загрузить программное обеспечение компании Arenadata, а так же получить электронную версию документации по продуктам.

Управление сервисами:

- Arenadata Cluster Manager**
Платформа управления, развертывания и мониторинга кластеров.
[Документация](#)
- Arenadata Ambari**
Сервис для автоматического развертывания стека Hadoop.
[Документация](#)

Хранение и обработка данных:

- Arenadata Hadoop**
Набор сервисов Hadoop стека.
[Документация](#)
- Arenadata DB**
MPP-система на базе Greenplum DB.
[Документация](#)
- Arenadata Streaming**
Платформа для обработки потоковых данных.
[Документация](#)
- Arenadata QuickMarts**
Кластерная колоночная система управления базами данных (СУБД).
[Документация](#)

Рис.1.3.: Магазин Приложений

Для загрузки дистрибутива платформы следует перейти на страницу продукта, кликнув на продукт **Arenadata Streaming**. При этом происходит переход на соответствующую страницу с кратким описанием продукта, ссылками на компоненты дистрибутива и информацией о релизе (Рис.1.4.).

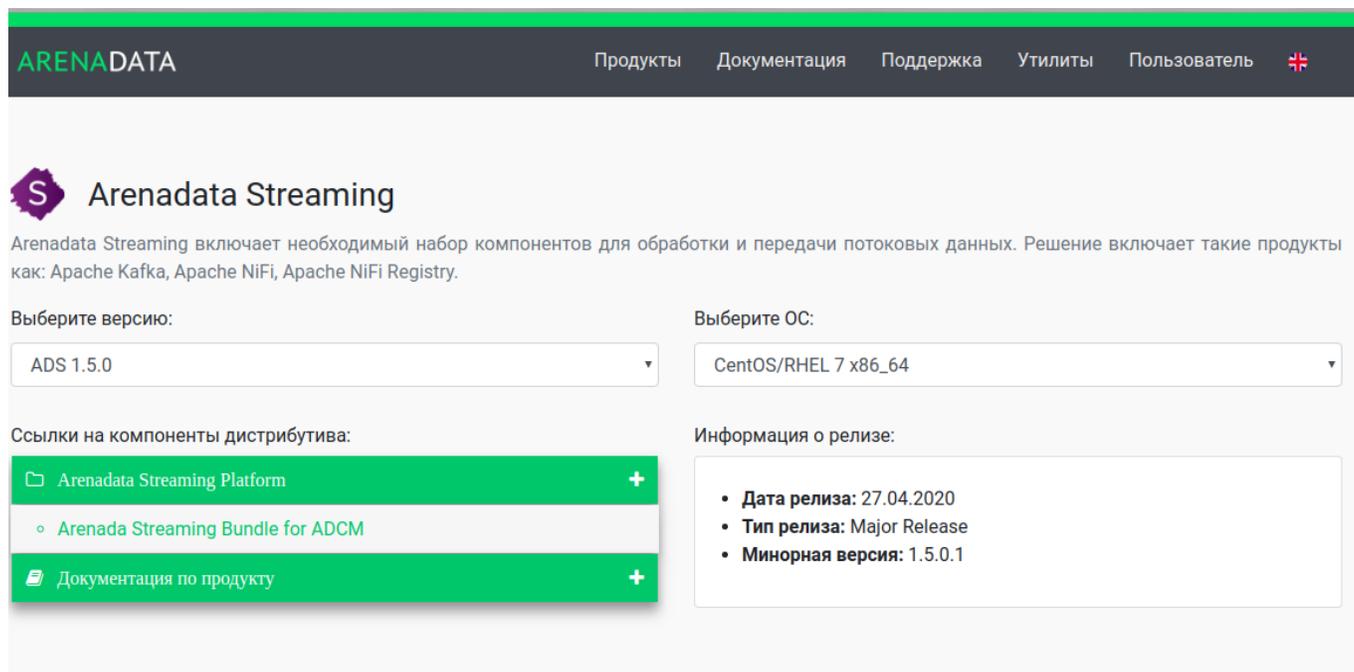


Рис.1.4.: Страница загрузки Arenadata Streaming

В разделе “*Выберите версию продукта*” следует указать интересующую версию **ADS**, а в поле “*Выберите ОС*” – используемую операционную систему. При этом в разделе “*Информация о релизе*” отображается дата выхода и тип релиза выбранной версии, ее номер, ссылки на лицензию использования и Release Notes.

В разделе *Ссылки на компоненты дистрибутива* находятся ссылки для загрузки компонентов выбранной версии платформы. После загрузки компонентов следует установить их, действуя в соответствии с последующими разделами настоящего документа.

Глава 2

Предварительные действия

Для установки ADS посредством ADCM необходимо:

1. Установить ADCM (документация продукта приведена по [ссылке](#)).
2. Проверить настройки ADCM:
 - Перейти во вкладку “Settings”;
 - Проверить на корректность определившийся автоматически URL и при необходимости заменить его на исправный (Рис.2.1).

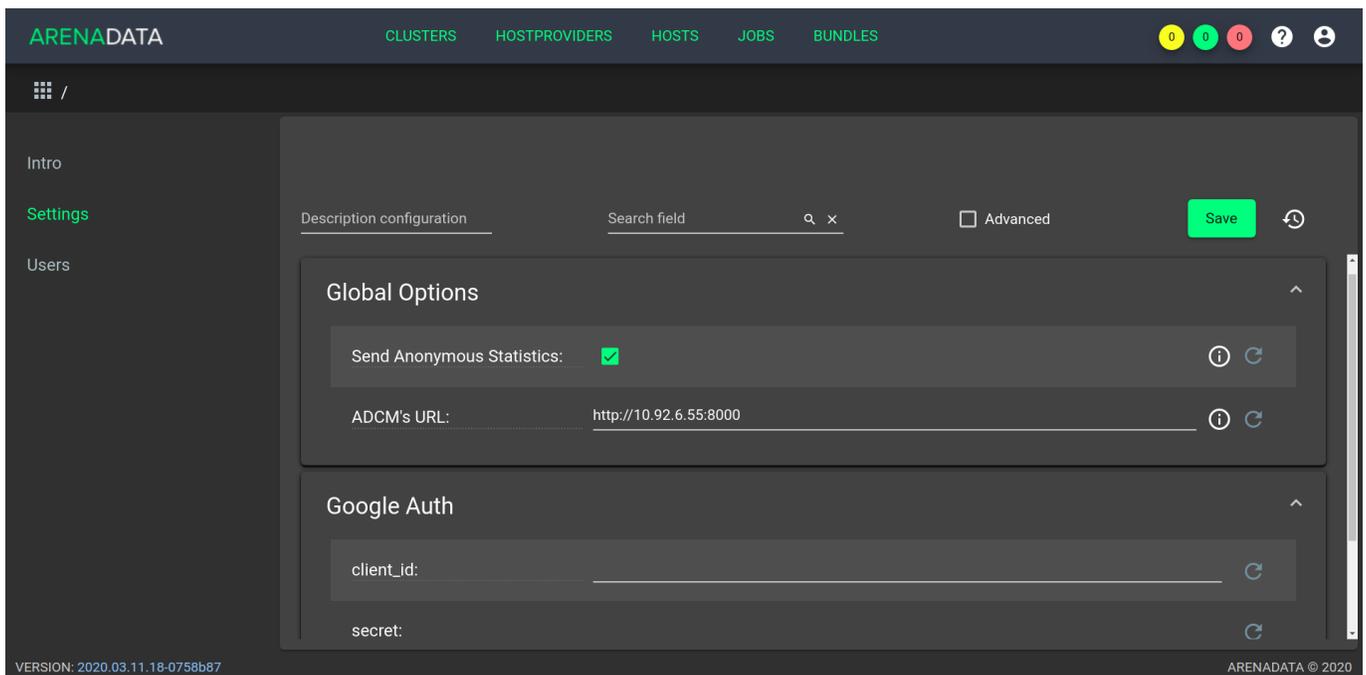


Рис.2.1.: Проверка настроек ADCM

Important: Каждый компонент сервиса кластера имеет возможность отсылать статусную информацию о своем состоянии (keep alive) процессу ADCM в докер-контейнере

В ряде случаев **ADCM** может оказаться за NAT, и тогда исключается очевидный вариант автоматического определения его адреса, видимого со стороны сервисного компонента на хосте кластера. Поэтому данный адрес указывается вручную. Во время установки **ADS** адрес **ADCM** используется для заполнения конфигурации компонентов, отвечающих за передачу статусной информации.

3. Создать хосты для кластера ADS:

- Загрузить выбранный бандл хоста. В текущем примере используется бандл *SSH*;
- Инициализировать необходимое количество хостов (в текущем примере данный шаг пропускается, так как используются готовые хосты, а не облачный провайдер);
- Добавить хосты в ADCM. В текущем примере используется 4 хоста: для Zookeeper (*zk*), для Kafka брокеров (*kafka1* и *kafka2*), для Nifi (*nifi*).

4. (Опционально) Создать кластер мониторинга:

- Загрузить бандл мониторинга;
- Создать экземпляр кластера мониторинга и установить его.

5. (Опционально) Для установки без доступа к сети Интернет (offline):

- Запросить дистрибутив [Arenadata Enterprise Tools](#);
- Развернуть Arenadata Enterprise Tools на выделенном хосте.

Глава 3

Загрузка бандла ADS

Загрузка бандла **ADS** необходима для создания в **ADCM** прототипа кластера, из которого в дальнейшем возможно создание его экземпляров.

Для загрузки бандла следует выполнить следующие действия:

1. Открыть в ADCM вкладку “BUNDLES” (Рис.3.1.).

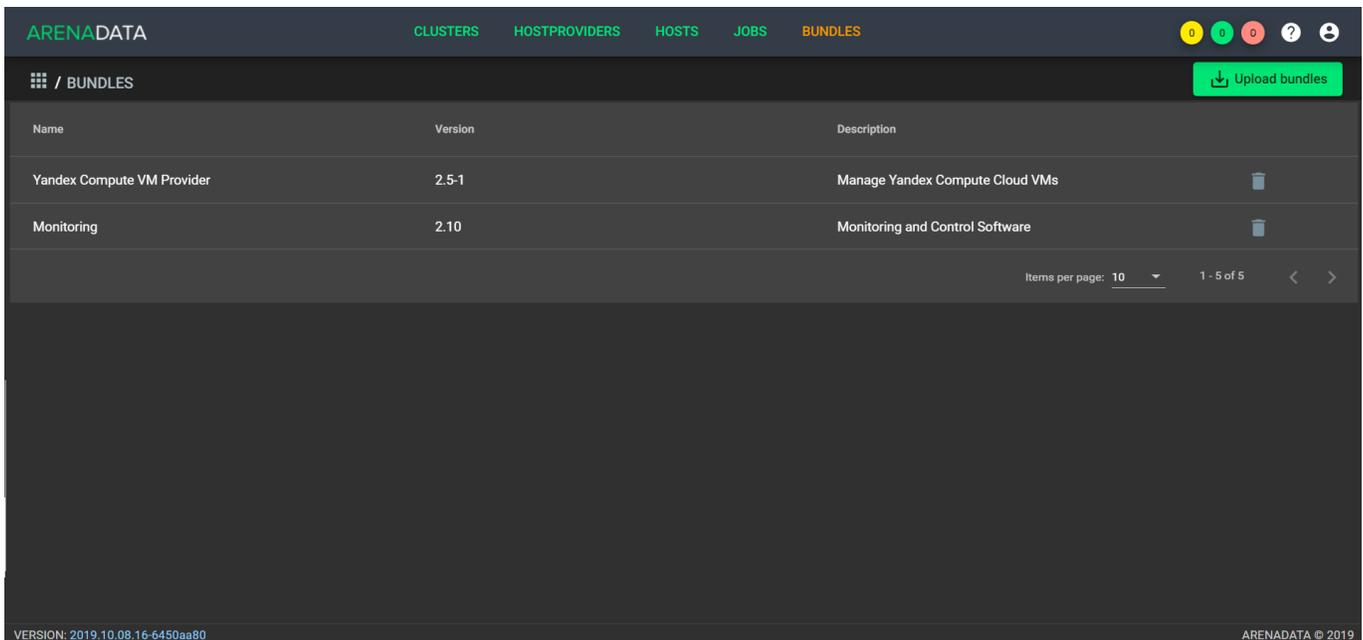


Рис.3.1.: Вкладка “BUNDLES”

2. Нажать “Upload bundle” и в открывшейся форме выбрать файл бандла ADS (Рис.3.2.).
3. После успешной загрузки бандла необходимо принять лицензионное соглашение (Рис.3.3.).
4. В результате выполненных действий факт успешной загрузки отображается в общем списке бандлов на вкладке “BUNDLES” (Рис.3.4.).

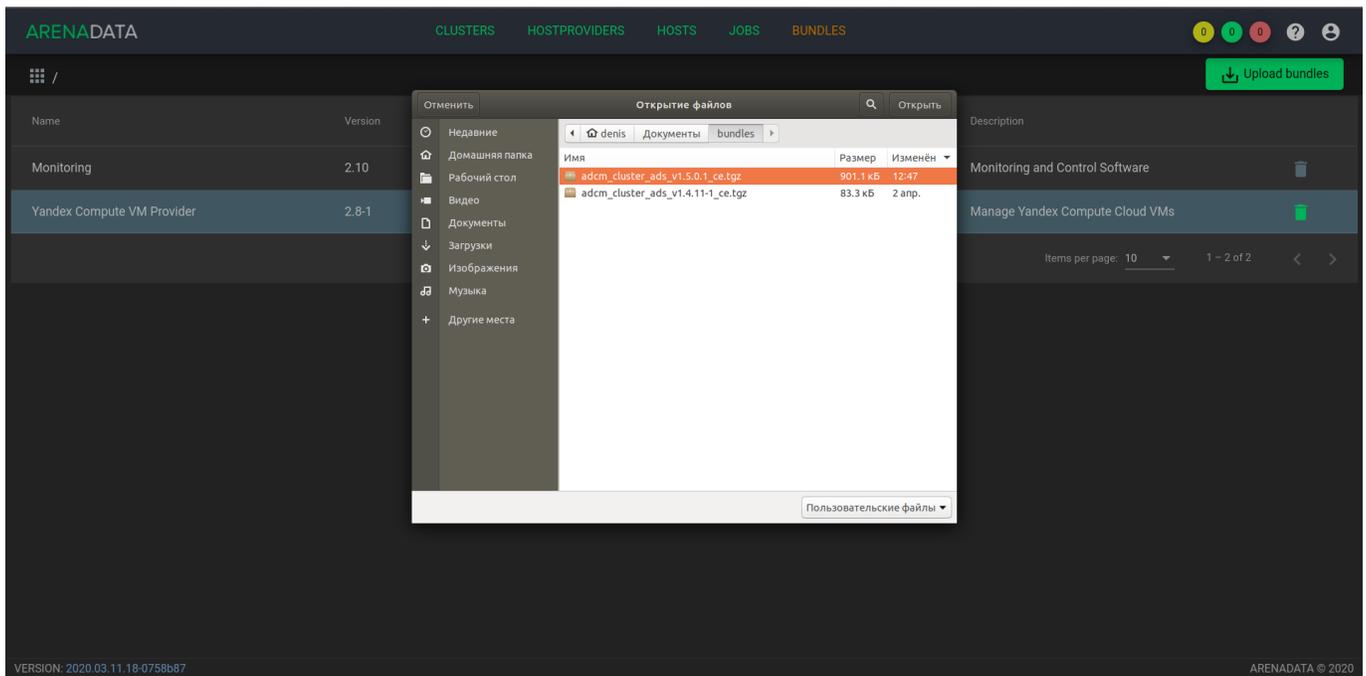


Рис.3.2.: Выбор бандла

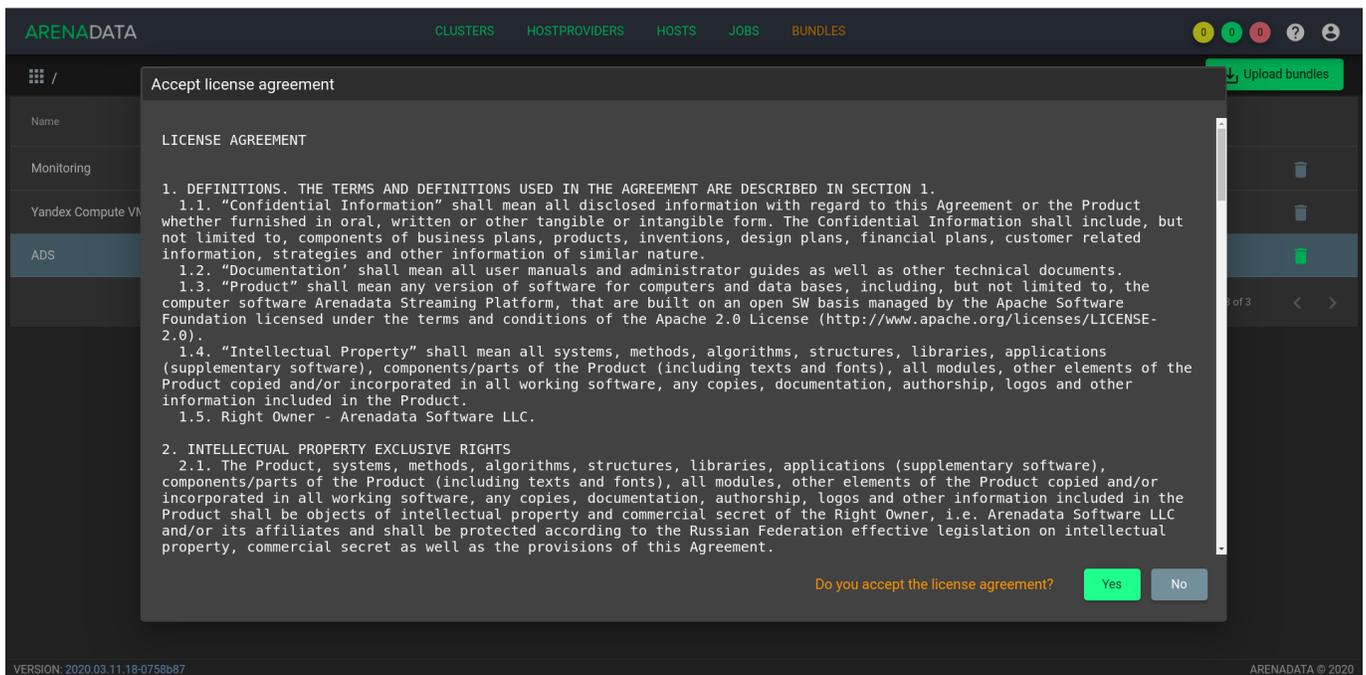


Рис.3.3.: Лицензионное соглашение

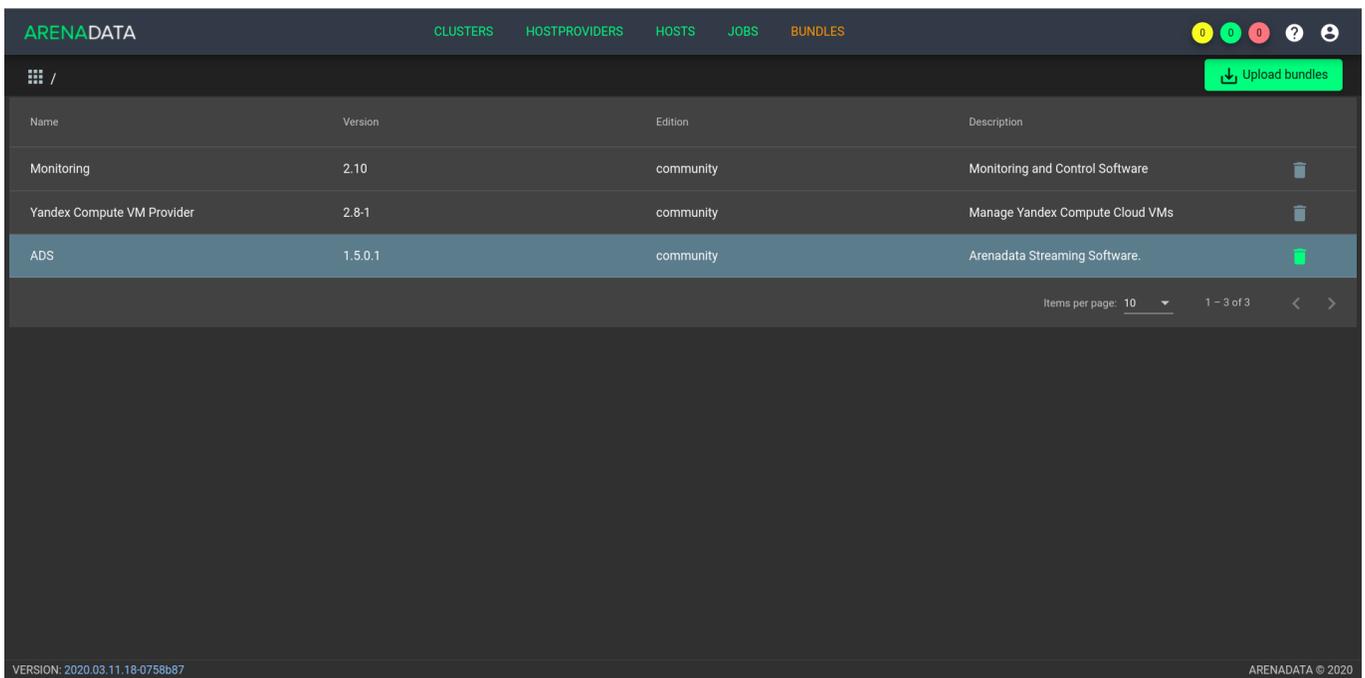


Рис.3.4.: Результат успешной загрузки бандла

Глава 4

Создание кластера ADS

После выполнения *предварительных действий* и *загрузки бандла* в кластер-менеджере ADCM содержится следующий список объектов (Рис.4.1.):

- Прототип кластера мониторинга и созданный на его основе экземпляр (программа мониторинга развернута на хосте, ADCM содержит записи о его результатах и настройках);
- Прототип кластера *ADS* для создания экземпляров;
- Предварительно подготовленные хосты, в базе данных ADCM хранятся записи о них и их учетных данных – ssh-ключами или паролями.

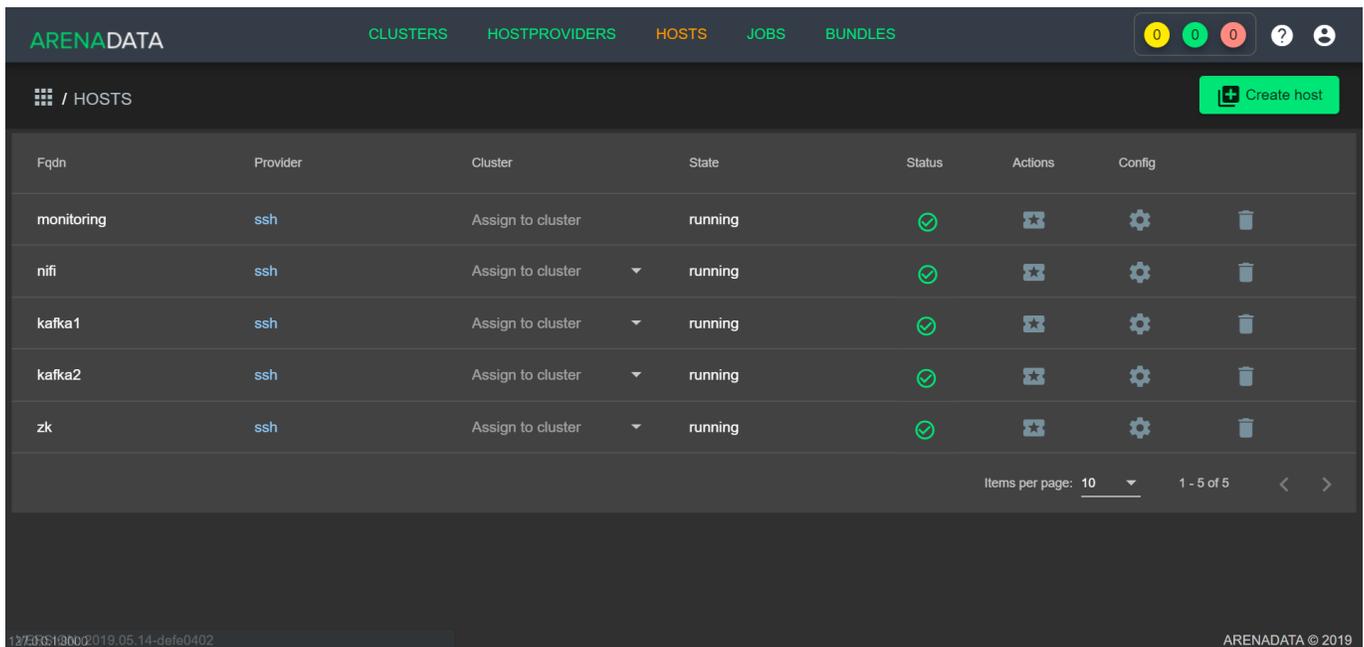


Рис.4.1.: Список объектов в ADCM

Объектам доступен следующий функционал:

- *Создание экземпляра кластера;*
- *Конфигурация кластера;*
- *Добавление хостов;*

- *Добавление сервисов;*
- *Размещение компонентов сервисов на хостах;*
- *Установка кластера.*

4.1 Создание экземпляра кластера

При создании кластера в веб-интерфейсе **ADCM** генерируется новый экземпляр кластера *ADS*, что означает только добавление данных о нем в базу данных **ADCM** – на этом этапе не производится установка *ADS* на хосты.

1. Открыть в **ADCM** вкладку “**CLUSTERS**” (Рис.4.2.).

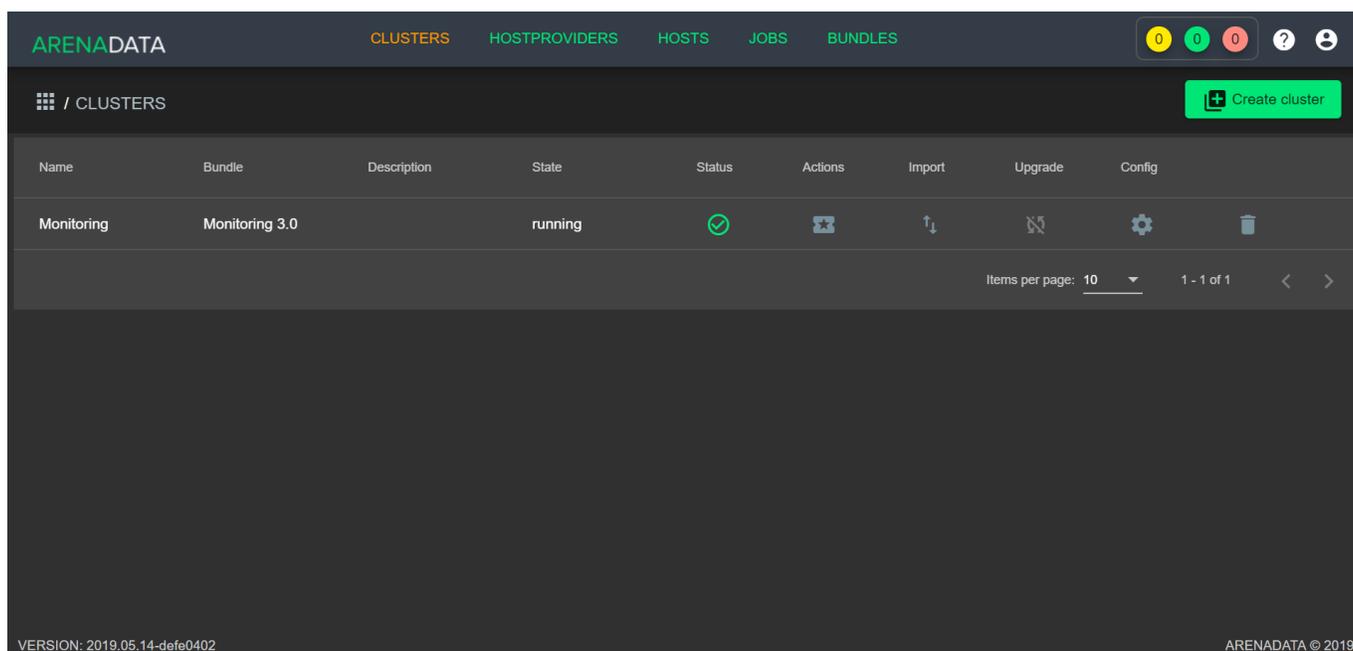


Рис.4.2.: Вкладка “**CLUSTERS**”

2. Нажать “**Create cluster**” и в открывшейся форме создать экземпляр кластера из прототипа *ADS*, полученного из бандла (Рис.4.3.).
3. В результате выполненных действий факт создания экземпляра кластера отображается в базе данных **ADCM** на вкладке “**CLUSTERS**” (Рис.4.4.).

4.2 Конфигурация кластера

Для перехода к настройкам экземпляра кластера *ADS* необходимо нажать кнопку с пиктограммой шестеренки в соответствующей строке вкладки “**CLUSTERS**” (Рис.4.4.) и перейти в раздел меню “**Configuration**”. При этом открывается окно конфигурации выбранного экземпляра (Рис.4.5.).

В блоке *RedHat repos* указываются требуемые для установки *ADS* пакеты из различных yum-репозиторий. В блоке *AltLinux repos* указываются требуемые для установки *ADS* пакеты из различных apt-репозиторий. При этом в каждом из параметров можно изменить заданный по умолчанию url:

- *ADS;*
- *monitoring;*

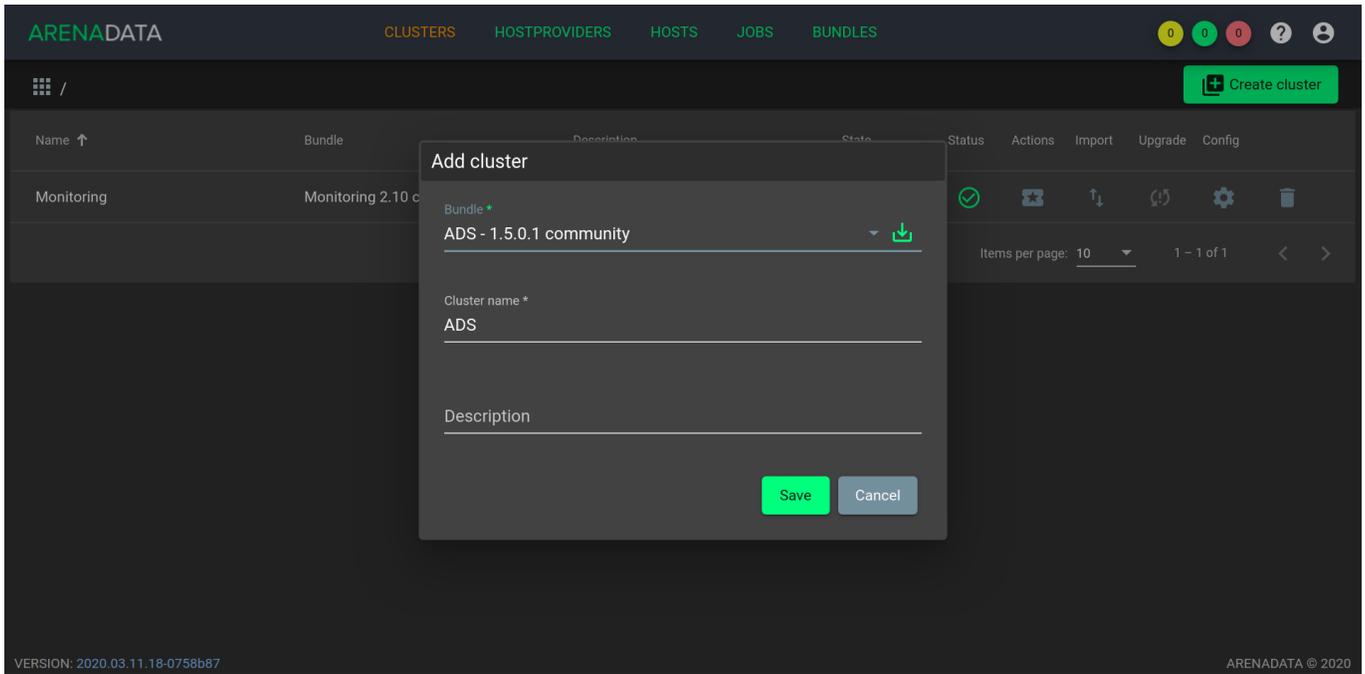


Рис.4.3.: Создание экземпляра кластера

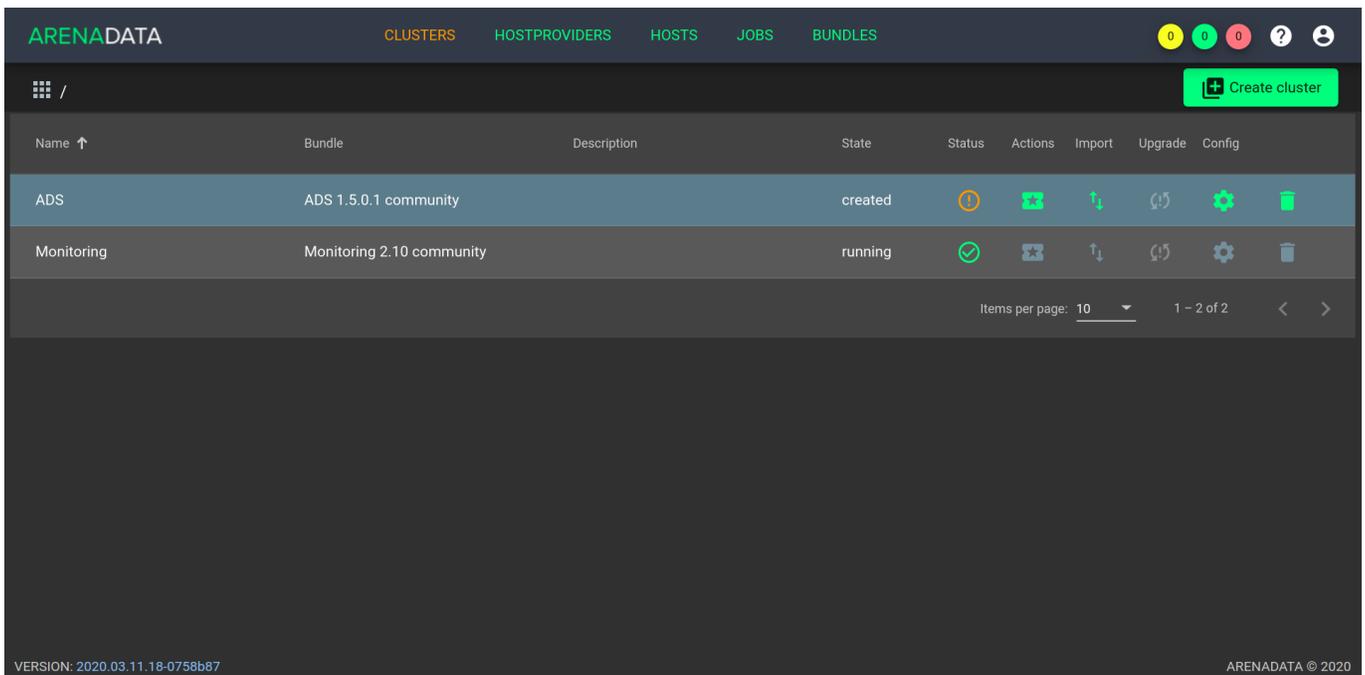


Рис.4.4.: Результат успешного создания экземпляра кластера

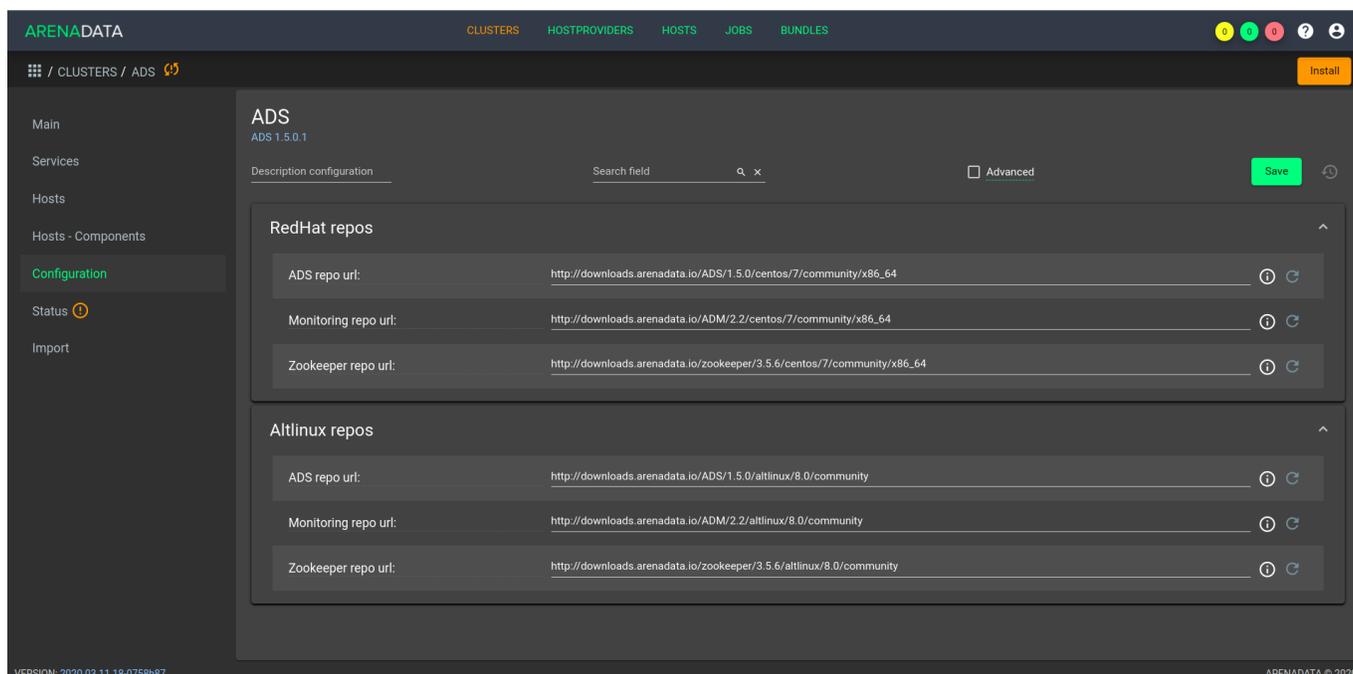


Рис.4.5.: Окно конфигурации кластера

- *Zookeeper*;
- *Ranger plugins*. Только для *Enterprise* версии.

Important: При установке в окружении без доступа к сети Интернет (offline) необходимо предварительно развернуть кластер **Arenadata Enterprise Tools** и импортировать настройки HTTP Mirror

В таком случае настройки offline-репозитория применяются автоматически. Для этого на странице кластера *ADS* необходимо перейти на вкладку *Import* и выбрать сервис http кластера Enterprise Tools (Рис.4.6.). Альтернативно, в эти поля можно указать url локальных репозитория с необходимыми пакетами.

Important: Arenadata Enterprise Tools предоставляет локальную копию только репозитория Arenadata Streaming и Arenadata Monitoring. Доступ к репозиториям CentOS должен быть организован отдельно

4.3 Добавление хостов

Для добавления хостов в кластер *ADS* необходимо:

1. В меню кластера *ADS* открыть вкладку “Hosts” (Рис.4.7.).
2. Нажать “Add hosts” и в открывшейся форме выбрать необходимые хосты (Рис.4.8.).

Important: Необходимо использовать иерархические (FQDN) имена хостов. В противном случае возможны ошибки при установке/эксплуатации кластера.

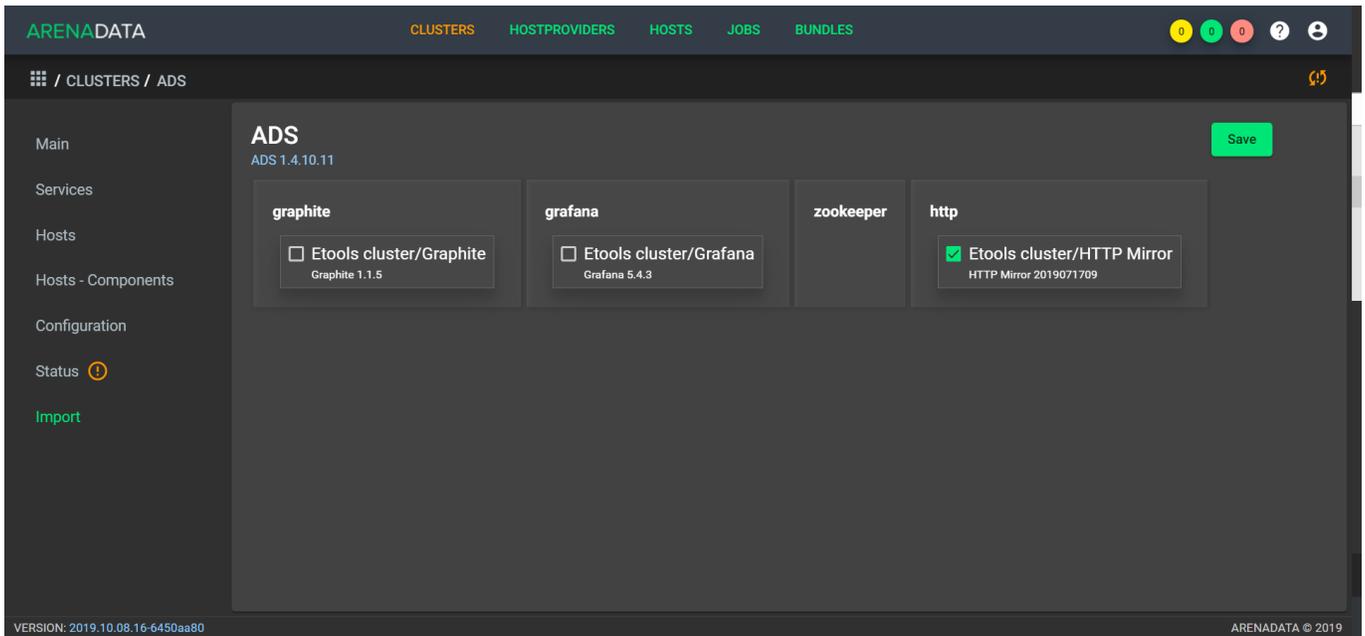


Рис.4.6.: Импорт конфигурации сервиса http

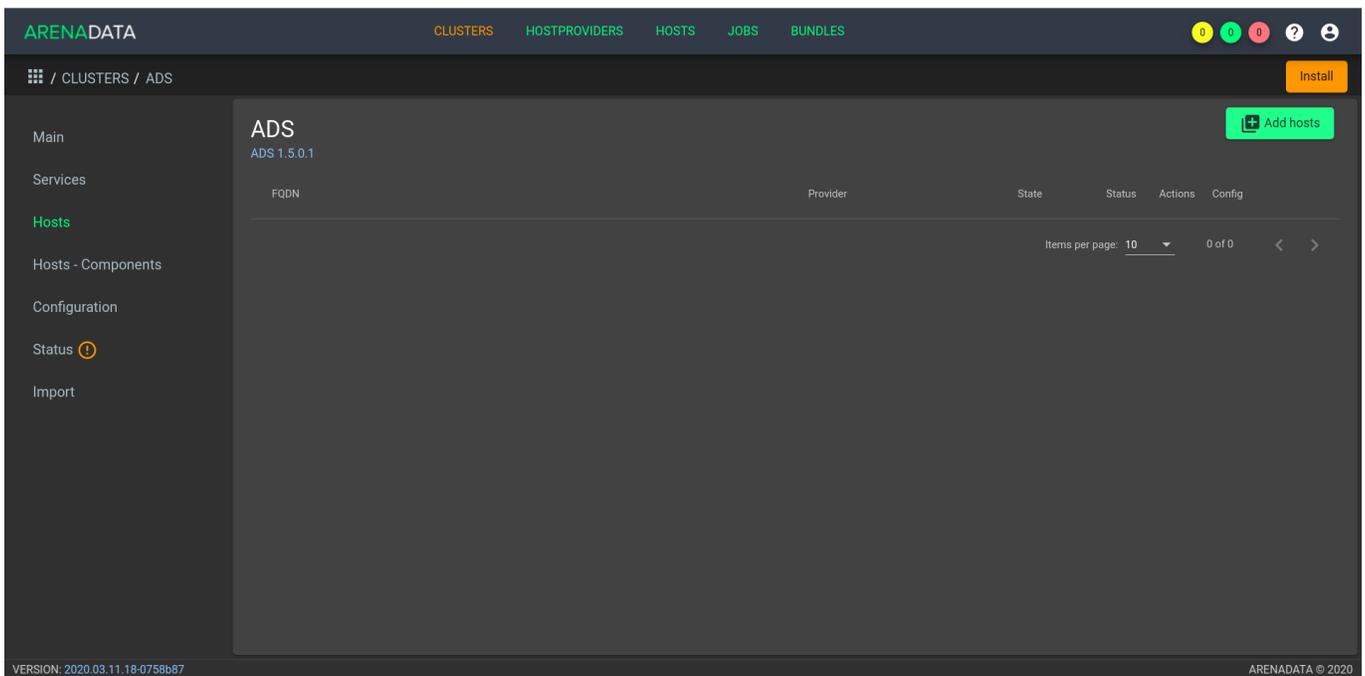


Рис.4.7.: Вкладка “Hosts” кластера ADS

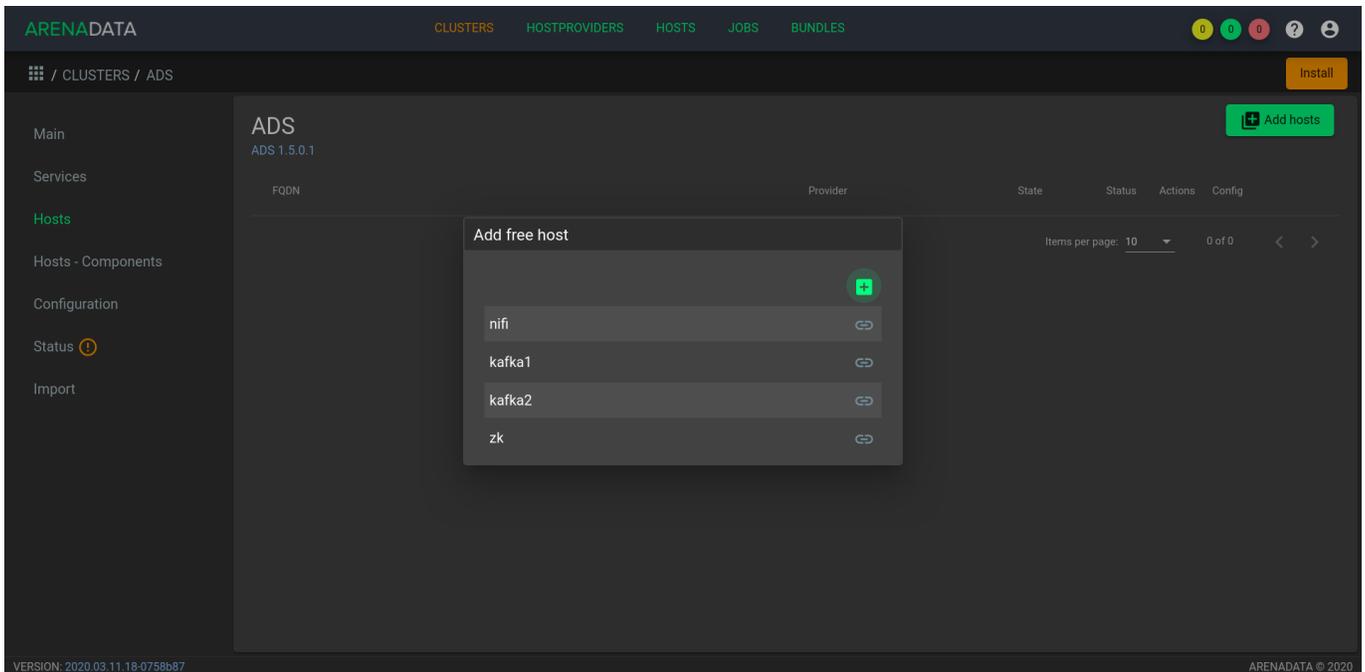


Рис.4.8.: Выбор хостов

3. В результате выполненных действий факт добавления хостов отображается в кластере *ADS* в списке вкладки “Hosts” (Рис.4.9).

4.4 Добавление сервисов

Important: На текущий момент невозможно удалить из кластера уже добавленный сервис

Кластер **ADS** содержит следующие сервисы:

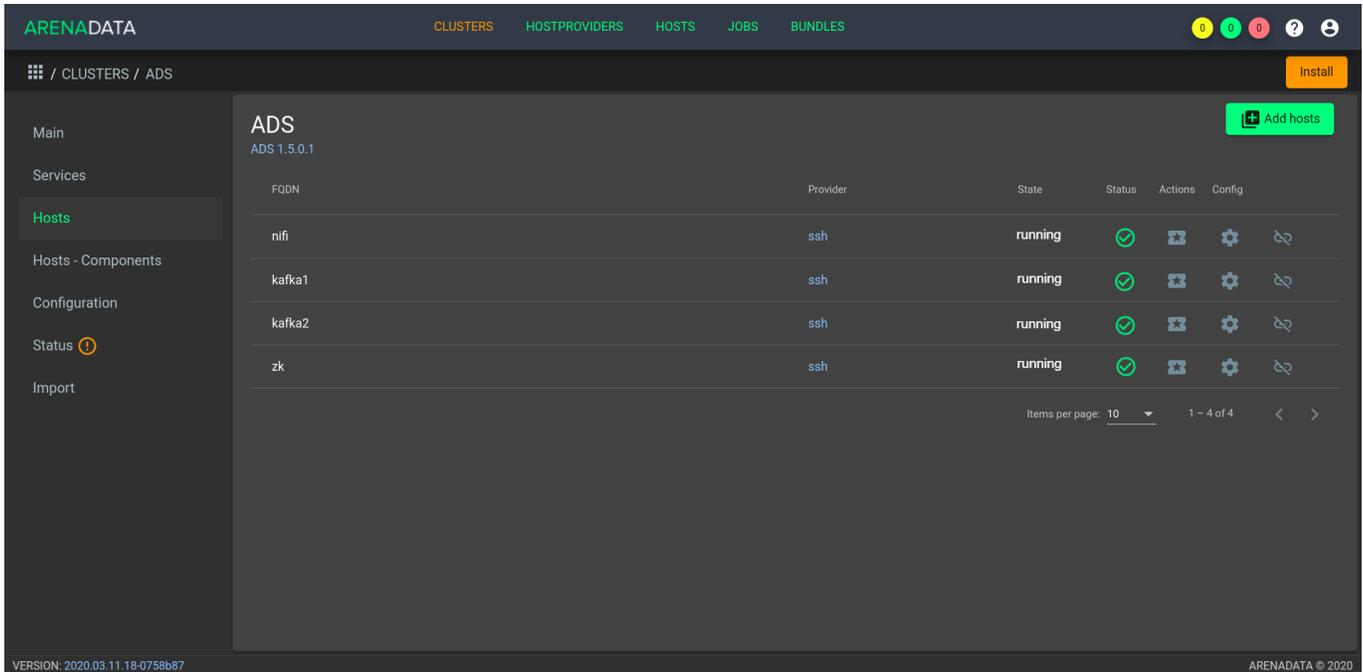


Рис.4.9.: Результат успешного добавления хостов

Таблица4.1.: Сервисы ADS

Сервис	Компоненты	Описание
Zookeeper	Zookeeper Server	Сервис, предназначенный для хранения конфигураций, выполнения распределенной синхронизации процессов
Kafka	Kafka Broker	Распределенная платформа для потоковых операций и данных
Nifi	Nifi Server, Nifi Registry	Распределенная платформа, предназначенная для построения и автоматизации потоков данных между различными системами
Schema-registry	Schema-registry	Сервис предоставляет обслуживающий слой для метаданных. Предоставляет интерфейс RESTful для хранения и извлечения схем Avro. Хранит версию историю всех схем, предоставляет несколько параметров совместимости и позволяет изменять схемы в соответствии с настроенными параметрами совместимости и расширенной поддержкой Avro. Предоставляет сериализаторы, подключаемые к клиентам Kafka, которые обрабатывают хранение и извлечение схемы для сообщений Kafka, отправляемых в формате Avro
Kafka REST Proxy	Kafka REST Proxy	Предоставляет RESTful-интерфейс для кластера Kafka, что упрощает создание и потребление сообщений, просмотр состояния кластера и выполнение административных действий без использования собственного протокола Kafka или клиентов
KSQL	KSQL Server, KSQL Client	Сервис представляет собой потоковый SQL-движок с открытым исходным кодом, который обеспечивает обработку данных в реальном времени с использованием Apache Kafka. Состоит из компонентов: Server; Client
Kafka-Manager	Kafka-Manager	Инструмент для управления Apache Kafka
MiNifi	MiNiFi C2 Server, MiNiFi Agent	Агент для сбора и отсылки данных, который поополняет основные принципы NiFi в управлении потоками данных, уделяя особое внимание сбору данных у их источника
Monitoring Clients	System metrics Agent, JMX metrics Collector	Агенты, отсылающие информацию о хосте и сервисах в мониторинг

Не все сервисы являются обязательными для установки. Например, если не планируется использование **Nifi**, то нет необходимости добавлять сервис. Или в случае, когда применяется сервис мониторинга (не на базе **Graphite**), незачем ставить агенты из *Monitoring Clients*. Однако, если планируется использование **Kafka**, одноименный сервис и *Zookeeper* обязательны, то же самое можно сказать и про сервис *Nifi*. При этом сервис может состоять из обязательных и необязательных компонентов. Например, сервис *KSQL* состоит из обязательного компонента *Server* и необязательного – *Client*.

Для добавления сервисов в кластер *ADS* необходимо:

1. В меню кластера *ADS* открыть вкладку “Services”.
2. Нажать “Add service” и в открывшейся форме выбрать необходимые сервисы (Рис.4.10.).

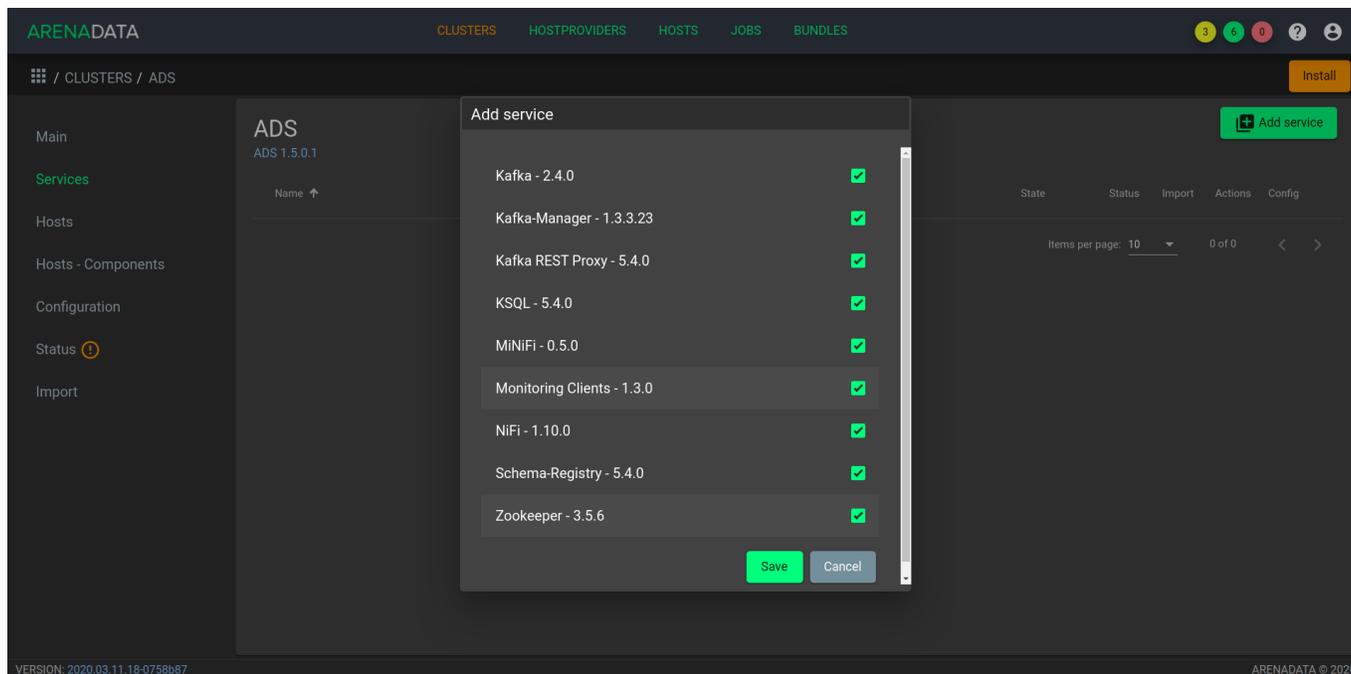


Рис.4.10.: Выбор сервисов для кластера ADS

В настоящем примере в кластер добавлены все сервисы, подробное описание конфигурации которых приведено по следующим ссылкам:

- [Настройка сервиса Zookeeper;](#)
- [Настройка сервиса Kafka;](#)
- [Настройка сервиса Nifi;](#)
- [Настройка сервиса Schema-registry.](#)
- [Настройка сервиса Kafka REST Proxy;](#)
- [Настройка сервиса KSQL;](#)
- [Настройка сервиса Kafka-Manager;](#)
- [Настройка сервиса MiNifi;](#)
- [Настройка сервиса Monitoring Clients.](#)

4.4.1 Особенности сервиса Monitoring Clients

Сервис *Monitoring Clients* требует обязательного импорта конфигурационных параметров кластера мониторинга (адрес, логин/пароль) при добавлении в кластер *ADS*.

Для импорта конфигурации мониторинга необходимо в кластере *ADS* выбрать сервис *Monitoring Clients*, затем перейти во вкладку *Import* и отметить импортируемые настройки сервисов (Рис.4.11.).

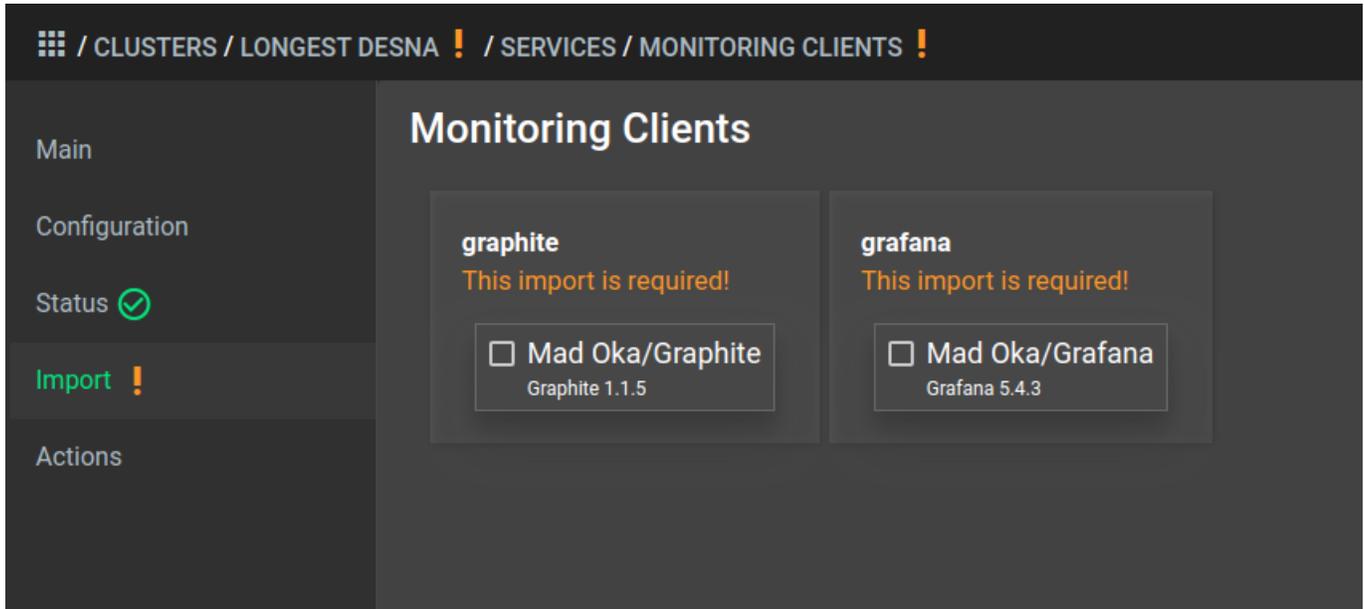


Рис.4.11.: Импорт конфигурации мониторинга

4.5 Размещение компонентов сервисов на хостах

Каждый сервис состоит из обязательных компонентов, которые должны быть размещены на хостах кластера, и необязательных, которые могут быть не размещены. Для размещения компонентов необходимо на вкладке “Hosts - Components” выбрать компонент в колонке “Components” и определить для него необходимый хост в колонке “Hosts” (Рис.4.12.).

Поскольку все сервисы добавлены в кластер *ADS*, но еще не размещены на хостах, то изначально ни на одном из хостов нет компонентов.

В качестве примера на Рис.4.13. показаны компоненты сервиса *Nifi*, размещенные на хосте *nifi*:

- *Nifi Server* – необходимо добавить на один или более хостов (возможна кластерная конфигурация);
- *Nifi Registry* – необходимо добавить ровно на один хост.

С версии *ADS 1.6.0.0* есть поддержка Service Dependency Tree. Данный функционал не позволяет пользователю забыть о зависимых сервисах. Так, для работы сервиса **Schema-Registry** необходим обязательный сервис **Kafka** (Рис.4.14.).

4.6 Установка кластера

Important: В случае если добавлен сервис *Monitoring Clients*, необходимо импортировать настройки из кластера *Monitoring* (см. *Особенности сервиса Monitoring Clients*)

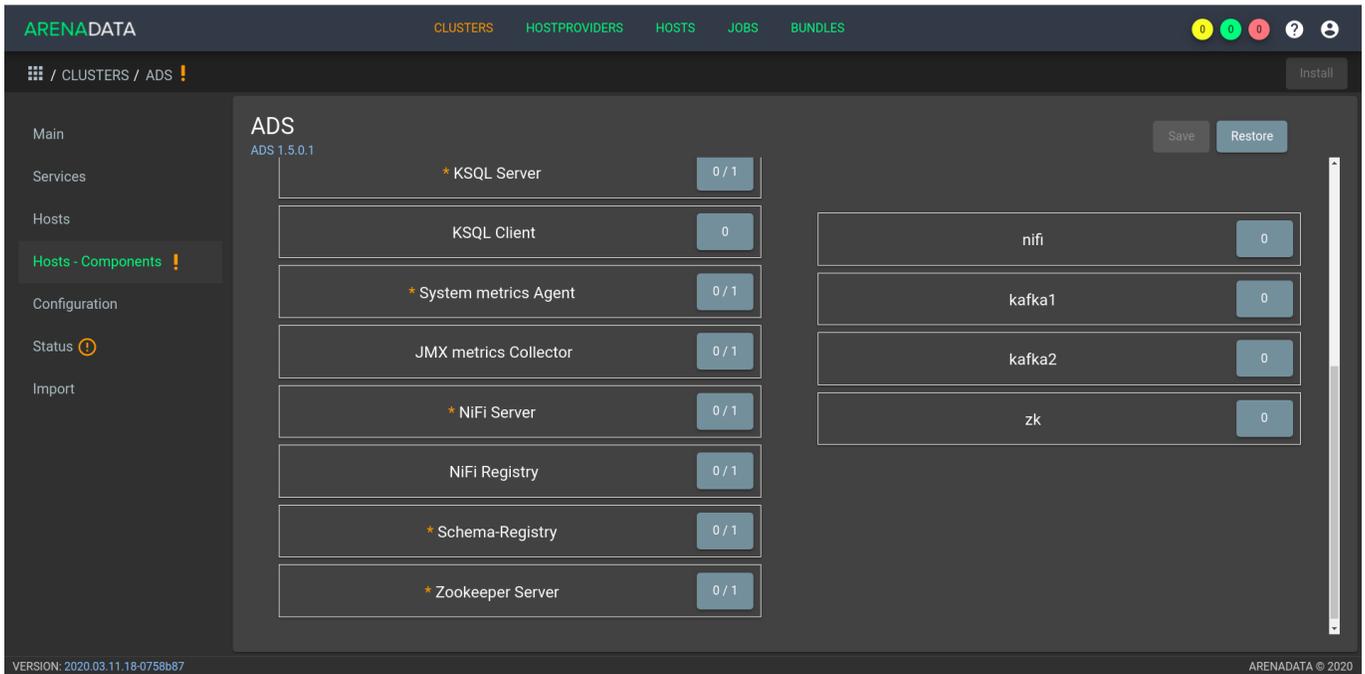
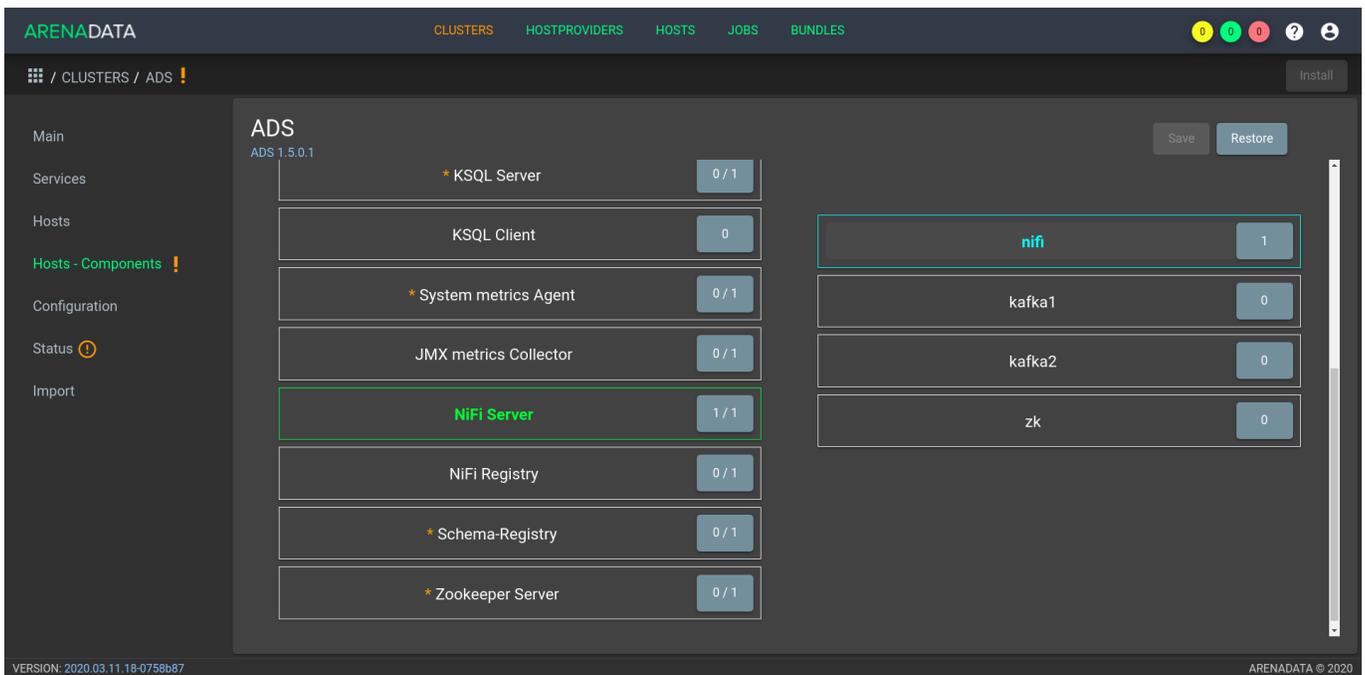


Рис.4.12.: Размещение компонентов сервисов на хостах

Рис.4.13.: Компоненты сервиса *Nifi*

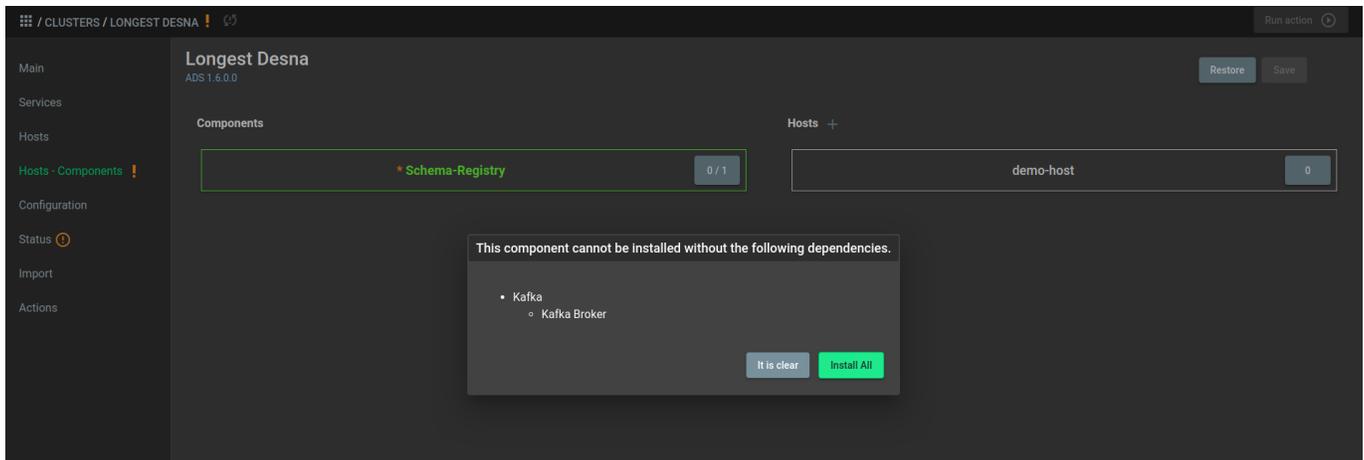


Рис.4.14.: Управление плагином

На данном этапе все хосты подготовлены и можно приступать к установке сервисов кластера:

- *Установка всех сервисов кластера;*
- *Установка сервиса в протестированный кластер.*

4.6.1 Установка всех сервисов кластера

Для установки всех добавленных сервисов в кластере *ADS* необходимо выбрать соответствующий кластер в **ADCM** и выполнить действие *Install* (Рис.4.15.).

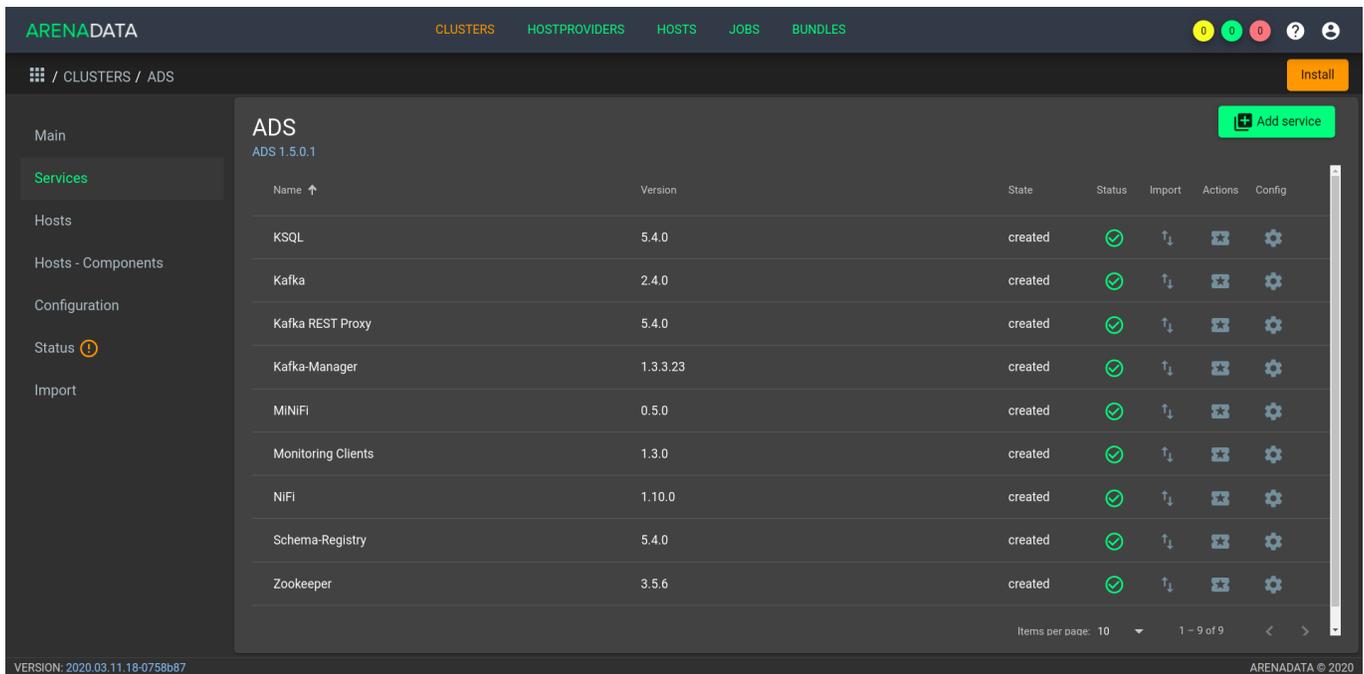


Рис.4.15.: Установка всех сервисов кластера

В появившемся диалоговом окне предоставляется выбор опций (Рис.4.16.):

- *Disable SELinux before cluster installation* – отключение SELinux на добавляемых хостах. Для того, чтобы данная настройка применилась, после завершения операции *Install* необходимо перезагрузить хосты вручную;
- *Disable Firewalld before cluster installation* – выключение firewalld на хостах;
- *Install OpenJDK before cluster installation* – установка пакета *java-1.8.0-openjdk* на хостах;
- *Set vm.swappiness to 0 for all hosts* – отключение *swapping* на хостах;
- *Append hosts into /etc/hosts file before cluster installation* – запись добавляемых нод в */etc/hosts* на хостах кластера. Данную опцию рекомендуется отключить, если настроен DNS.

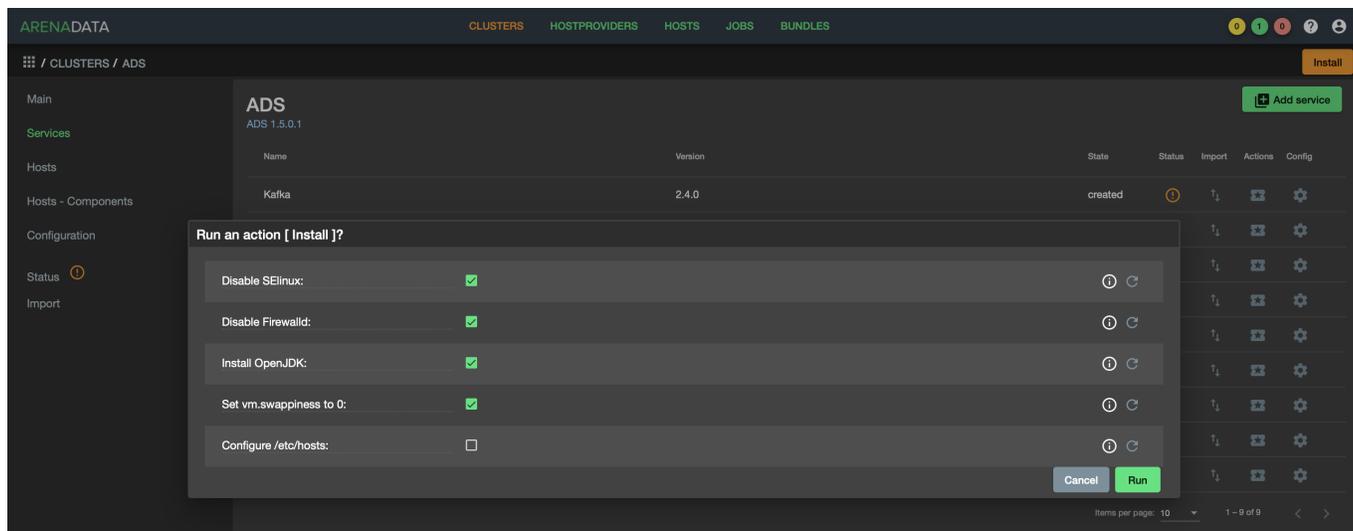


Рис.4.16.: Доступные при кластерной установке настройки

По результатам установки все добавленные сервисы меняют состояние с *created* на *installed* – установлен (Рис.4.17.).

Для запуска кластера необходимо нажать кнопку *Start*.

4.6.2 Установка сервиса в проинсталлированный кластер

В ADCM предусмотрена возможность добавления нового сервиса в уже работающий кластер. Для этого сервис необходимо добавить и произвести его установку. Например, для установки сервиса *Kafka* в проинсталлированный кластер необходимо:

- В меню кластера *ADS* открыть вкладку “Services”, нажать “Add service” и в открывшейся форме выбрать сервис *Kafka*;
- В строке сервиса *Kafka* в поле “Actions” нажать на пиктограмму и выбрать действие *Install*;
- По результатам инсталляции сервис *Kafka* меняет состояние с *created* – создан, на *installed* – установлен.

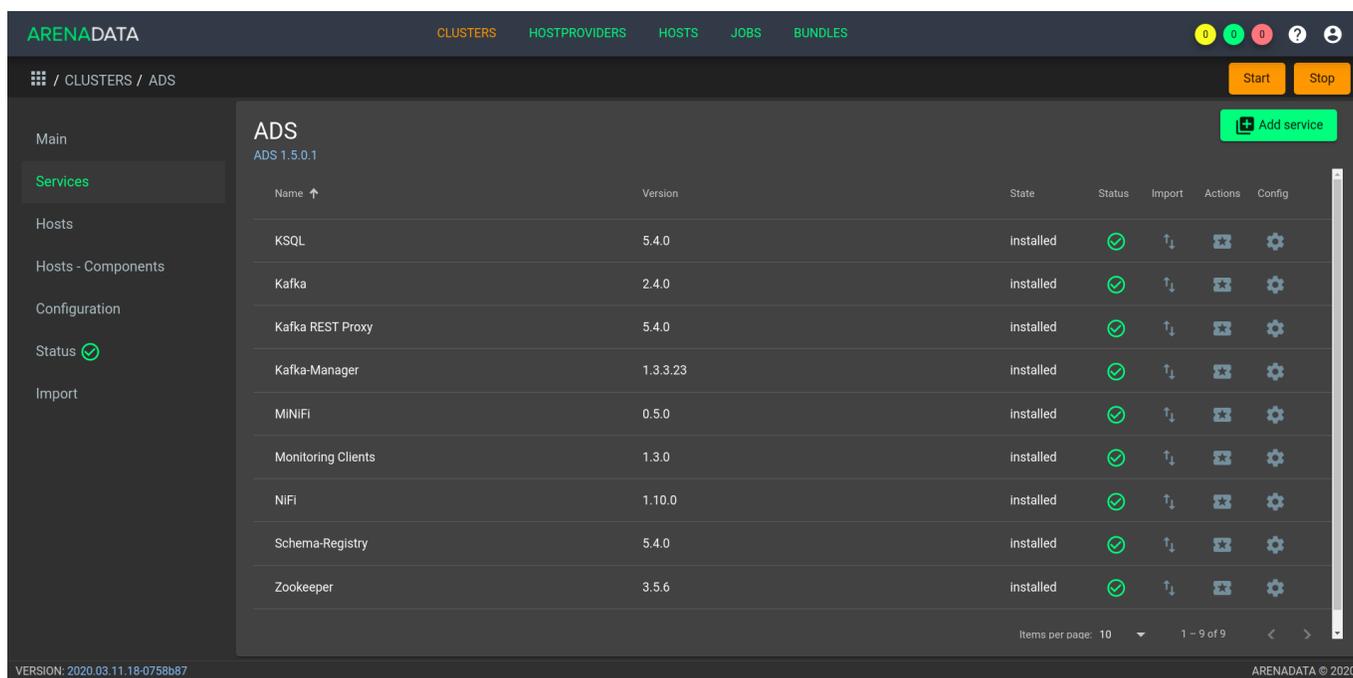


Рис.4.17.: Состояние сервисов кластера