Arenadata™ Hadoop

Версия - v2.0.0

Инструкция по установке кластера Arenadata Hadoop

2019 Arenadata™

Оглавление

1 Инструкция по установке ADCM **3**

2 Загрузка бандла хостпровайдера ADCM **5**

3 Выгрузка бандла хостпровайдера в ADCM **6**

4 Добавление хостов в ADCM **8**

5 Загрузка бандла кластера **9**

6 Установка ADH в ADCM **12**

6.1 Выгрузка бандла ADH в ADCM . . . 12
6.2 Создание экземпляра кластера . . . 12
6.3 Добавление сервисов кластера . . . 13
6.4 Конфигурирование сервиса . . . 13
6.5 Добавление хостов . . . 17
6.6 Размещение компонентов сервисов на хостах . . . 17
6.7 Конфигурирование кластера . . . 20
6.8 Установка сервисов кластера . . . 20

Инструкция может быть полезна администраторам, программистам, разработчикам и сотрудникам подразделений информационных технологий, осуществляющих внедрение и эксплуатацию кластера.

Кластер Arenadata Hadoop (ADH) устанавливается при помощи Arenadata Cluster Manager (ADCM) – отдельного продукта Arenadata, предназначенного для простого, удобного и быстрого развёртывания и эксплуатации программного обеспечения.

С документацией ADCM возможно ознакомиться по следующей ссылке: https://docs.arenadata.io/adcm/user/index.html

Глава 1

Инструкция по установке ADCM

Для установки ADCM необходимо выделить отдельный хост. Перед установкой ADCM необходимо убедиться, что целевой для установки хост подходит под требования.

Также рекомендуется ознакомиться с базовыми понятиями, используемыми внутри ADCM: https://docs.arenadata.io/adcm/user/concept.html

Далее необходимо установить ADCM, руководствуясь следующей инструкцией: https://docs.arenadata.io/adcm/user/install.html.

После установки ADCM его веб-интерфейс станет доступен на порте 8000 развёрнутого контейнера. Дальнейшая установка кластера ADH будет происходить через этот веб-интерфейс.

При установке ADCM по умолчанию создается пользователь с правами администратора – необходимо использовать эту учётную запись для первого логина в ADCM:

- Имя пользователя: **admin**
- Пароль: **admin**

Important: Компоненты кластера имеют возможность отсылать статусную информацию о своем состоянии в ADCM. Для корректной работы этой функциональности могут потребоваться ручные корректировки автоматических настроек ADCM.

Контейнер ADCM может оказаться в сети с использованием NAT, что исключает возможность автоматического определения адреса контейнера, видимого со стороны узлов кластера ADH (их установка описана далее). Поэтому может потребоваться указание адреса вручную, для этого необходимо перейти во вкладку *"Settings"* web-интерфейса ADCM и проверить на корректность автоматически определившийся URL, при необходимости заменв его на исправный:

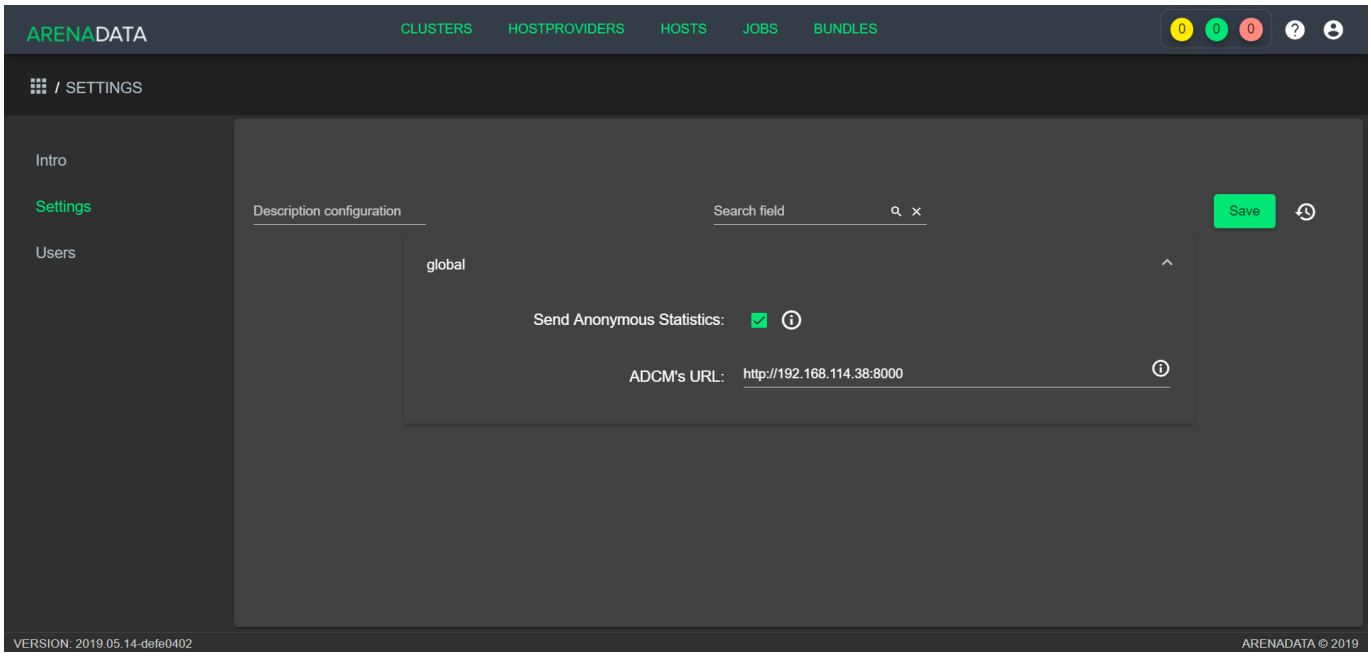


Рис.1.1.: Вкладка “Settings”

Глава 2

Загрузка бандла хостпровайдера ADCM

Для того, чтобы ADCM мог управлять кластером и его узлами, сначала необходимо подготовить хосты под кластер, а также связать эти хосты с ADCM. За эту функциональность отвечают хостпровайдеры ADCM – плагины ADCM, которые инкапсулируют сложный механизм взаимодействия ADCM и подконтрольных ему хостов.

Дистрибутивы хостпровайдеров для ADCM выполнены в виде так называемых бандлов. В случае простейшего хост провайдера ssh – это обычный архив, содержащий описание и программную логику подключения к хосту по протоколу ssh. Провайдер ssh будет использоваться дальше как пример в данной документации.

Для загрузки бандла хостпровайдера необходимо перейти на сайт https://arenadata.tech/products/adcm и выбрать вкладку *“Скачать”*. Далее отразиться краткое описание продукта, ссылки на компоненты дистрибутива и информация о релизе:

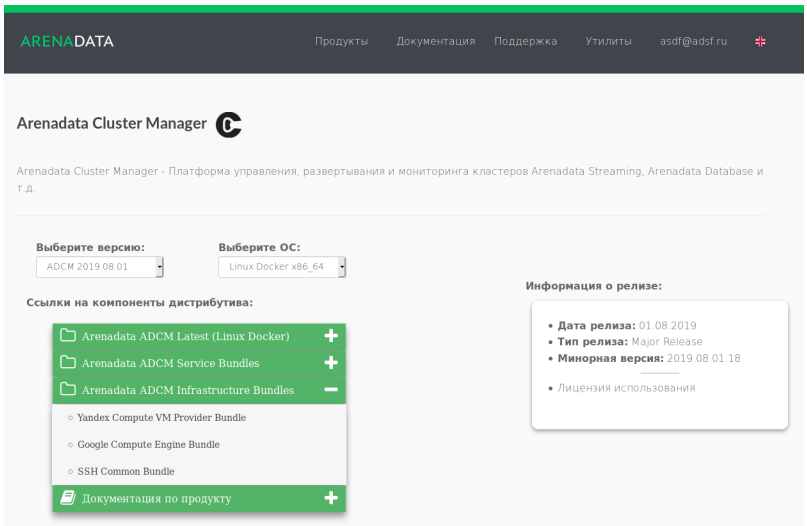


Рис.2.1.: Страница загрузки Arenadata Cluster Manager

После указания версии ADCM и операционной системы в раскрывающемся списке *“Arenadata ADCM Infrastructure Bundles”* возможен выбор вариантов загрузки бандлов хостпровайдеров – по нажатию на элемент списка бразузер начнёт загрузку. В блоке *“Информация о релизе”* отображается дата выхода, тип релиза выбранной версии и ссылки на лицензию.

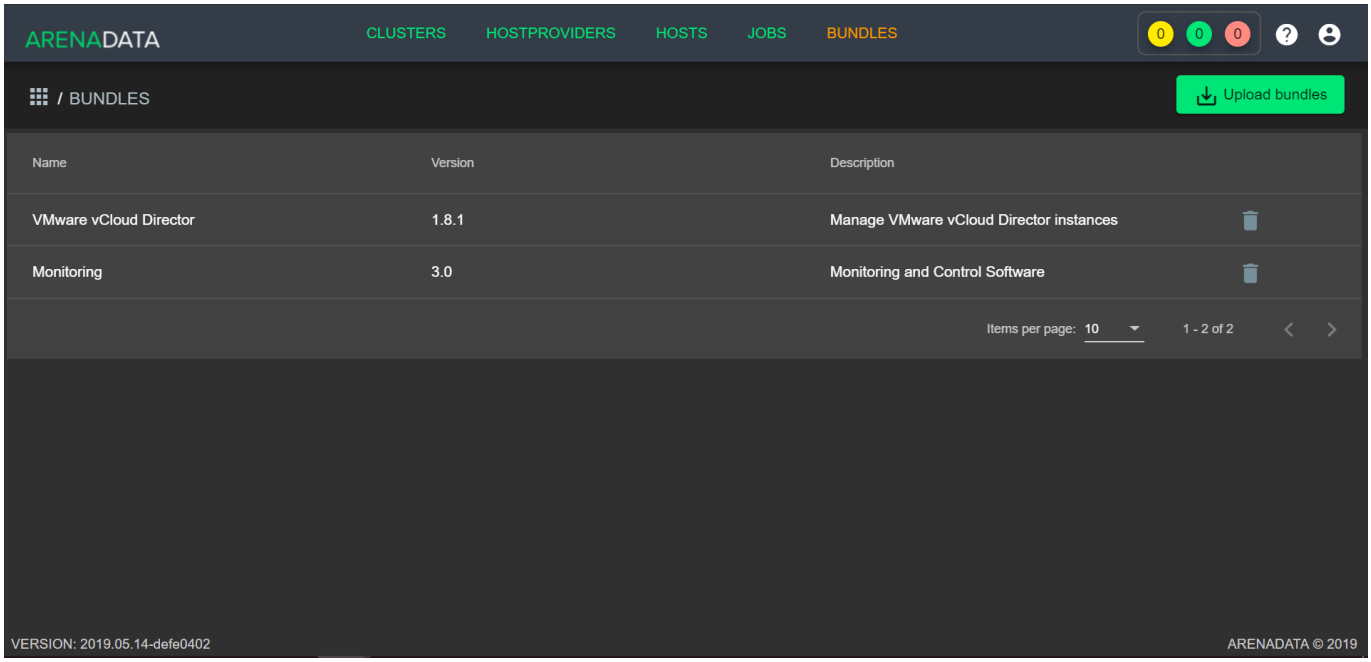
Бандл хостпровайдера ssh называется *“SSH Common Bundle”*.

Глава 3

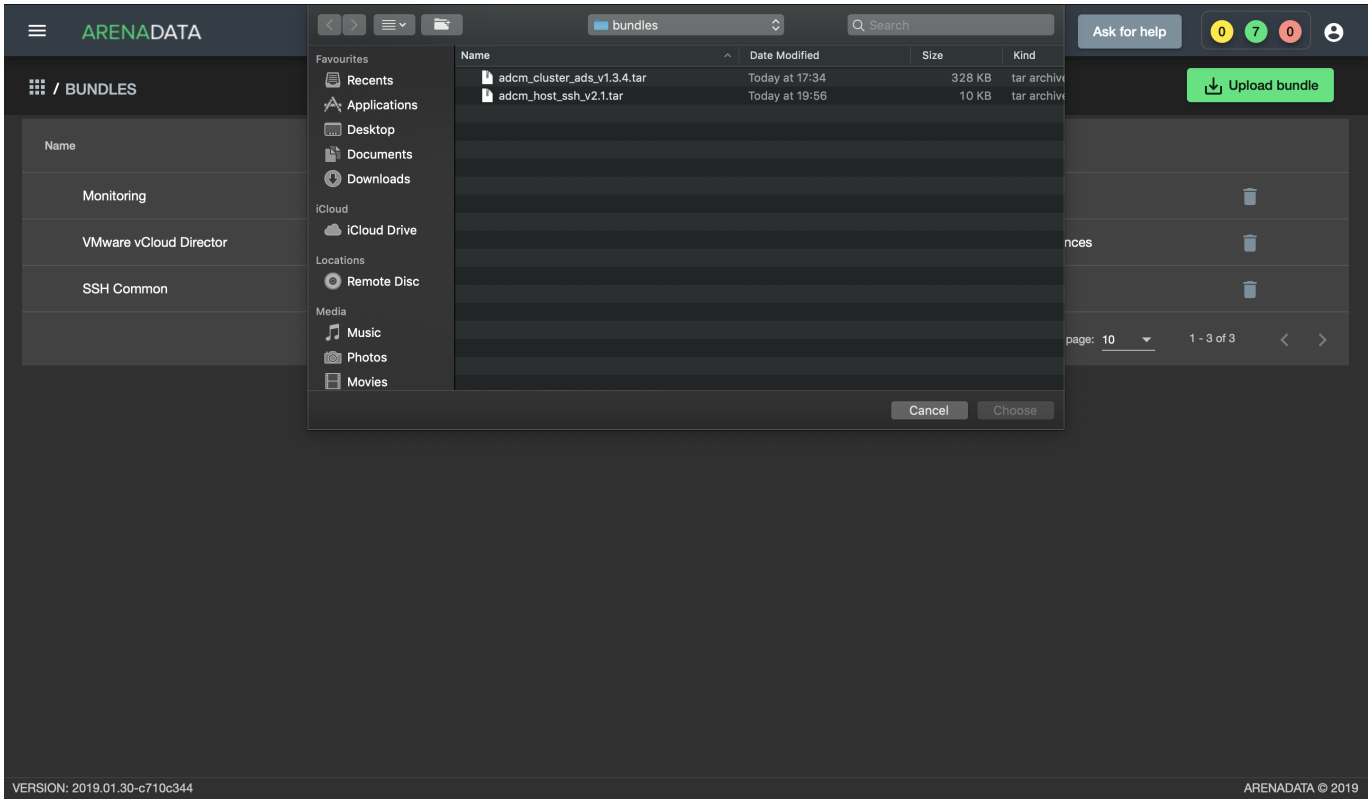
Выгрузка бандла хостпровайдера в ADCM

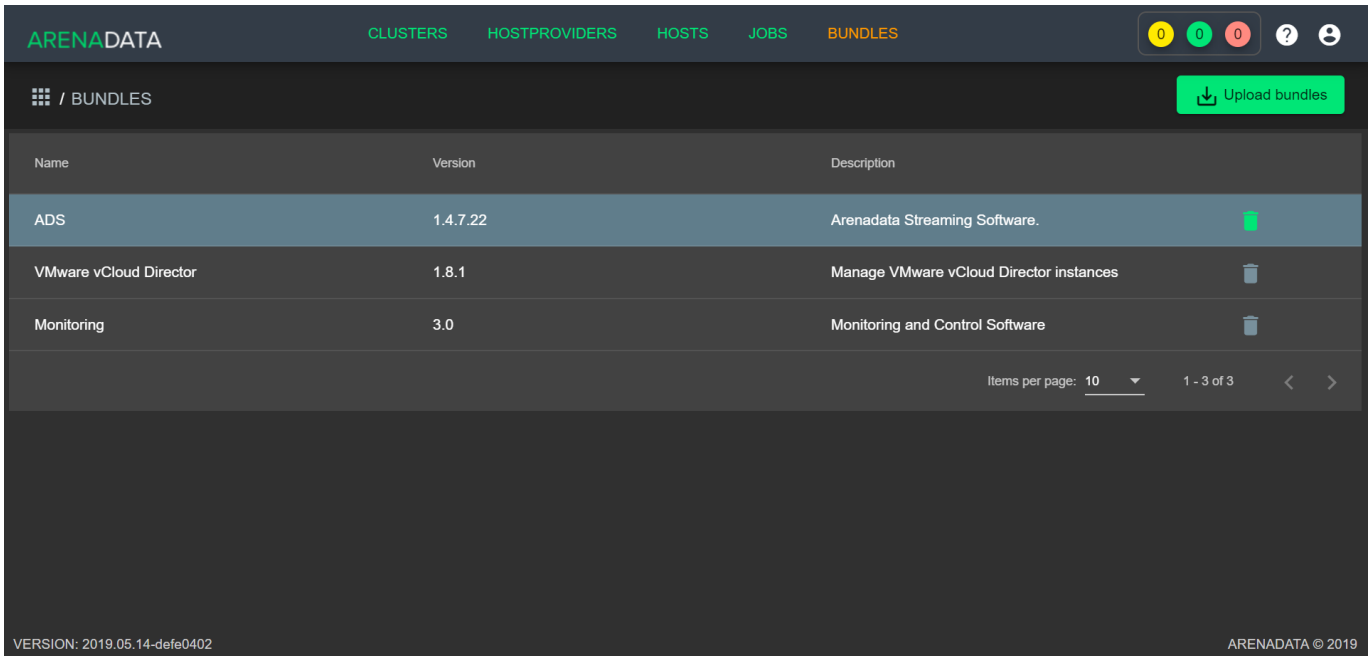
Для выгрузки бандла в ADCM следует выполнить следующие действия в веб-интерфейсе ADCM:

1. Открыть вкладку *“BUNDLES”*:



2. Нажать *“Upload bundle”* и в открывшейся форме выбрать файл бандла:
3. В результате выполненных действий факт успешной выгрузки отображается в общем списке бандлов на вкладке *“BUNDLES”*:


ARENADATA
bundles
Search
Ask for help
0 7 0
/ BUNDLES
Upload bundle
Favourites
Recents
Applications
Desktop
Documents
Downloads
iCloud
iCloud Drive
Locations
Remote Disc
Media
Music
Photos
Movies
Name
Date Modified
Size
Kind
adcm_cluster_ads_v1.3.4.tar
Today at 17:34
328 KB
adcm_host_ssh_v2.1.tar
Today at 19:56
10 KB
Cancel
Choose
Name
Monitoring
VMware vCloud Director
SSH Common
page: 10
1 - 3 of 3
VERSION: 2019.01.30-c710c344
ARENADATA © 2019


ARENADATA
CLUSTERS
HOSTPROVIDERS
HOSTS
JOBS
BUNDLES
0 0 0
/ BUNDLES
Upload bundles
Name
Version
Description
ADS
1.4.7.22
Arenadata Streaming Software.
VMware vCloud Director
1.8.1
Manage VMware vCloud Director instances
Monitoring
3.0
Monitoring and Control Software
Items per page: 10
1 - 3 of 3
VERSION: 2019.05.14-defe0402
ARENADATA © 2019

Глава 4

Добавление хостов в ADCM

В зависимости от выбранного хостпровайдера может потребоваться его изначальная настройка, а также настройка каждого добавленного хоста.

В простейшем случае при использовании хостпровайдера ssh (который является не более чем коннектором к уже созданным хостам) необходимо сначала создать хосты с возможностью подключения по протоколу ssh – это операция производится вручную и данным хостпровайдером не поддерживается.

Далее необходимо создать эти хосты внутри ADCM:

- Загрузить выбранный бандл хостпровайдера как это описано в документации ADCM: https://docs.arenadata.io/adcm/user/quick.html;
- На вкладке *“HOSTPROVIDERS”* добавить экземляр загруженного бандла, воспользовавшись кнопкой *“Create provider”*;
- На вкладке *“HOSTS”* добавить хосты, воспользовавшись кнопкой *“Create host”*.

Настройки для каждого добавленного в ADCM хоста возможно изменять. Для этого на вкладке *“HOSTS”* необходимо нажать на иконку шестерёнки в строке необходимого хоста. Так, в частности, при использовании хостпровайдера ssh в настройках хоста возможно указать приватный ssh ключ в строке ansible_ssh_private_key_file:

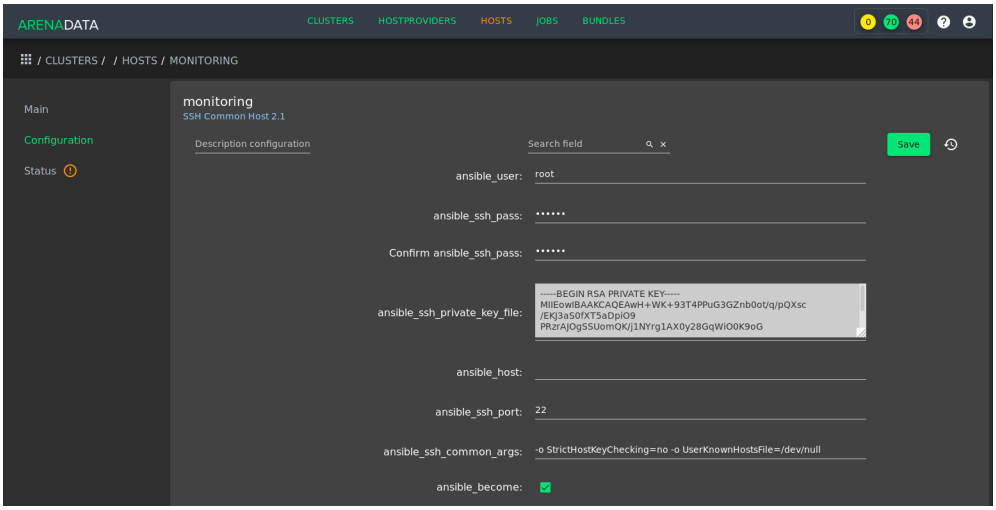


Рис.4.1.: Страница настройки для хоста в ADCM на примере хостпровайдера ssh

Глава 5

Загрузка бандла кластера

Дистрибутивы кластеров для ADCM выполнены в виде бандлов. В случае кластера ADH – это обычный архив, содержащий описание и программную логику развёртывания и эксплуатации кластера.

Для загрузки бандла ADH необходимо перейти на сайт https://arenadata.io и выбрать вкладку *"Скачать"*. Произойдёт переход на страницу Магазин Программного Обеспечения Arenadata и откроется форма заявки для загрузки продукта:

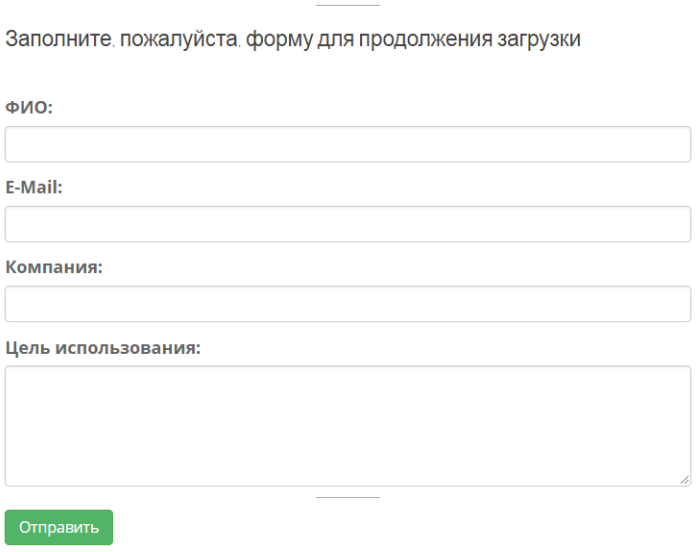


Рис.5.1.: Заявка для загрузки продукта

Все поля являются обязательными для заполнения. После отправки данных формы происходит переход на страницу *"Магазин Приложений"* с перечнем продуктов Arenadata:

При наведении курсора на продукт ARENADATA HADOOP откроется список ссылок:

- *Скачать* – указывает на страницу загрузки дистрибутива ADH;
- *Документация* – указывает на страницу онлайн документации ADH.

По прохождению по адресу ссылки *"Скачать"* отразиться краткое описание продукта, ссылки на компоненты дистрибутива и информация о релизе:

Послу указания версии ADH и операционной системы в раскрывающемся списке *"Ссылки на компоненты дистрибутива"* возможен выбор вариантов загрузки дистрибутива – по нажатию на элемент списка бразузер

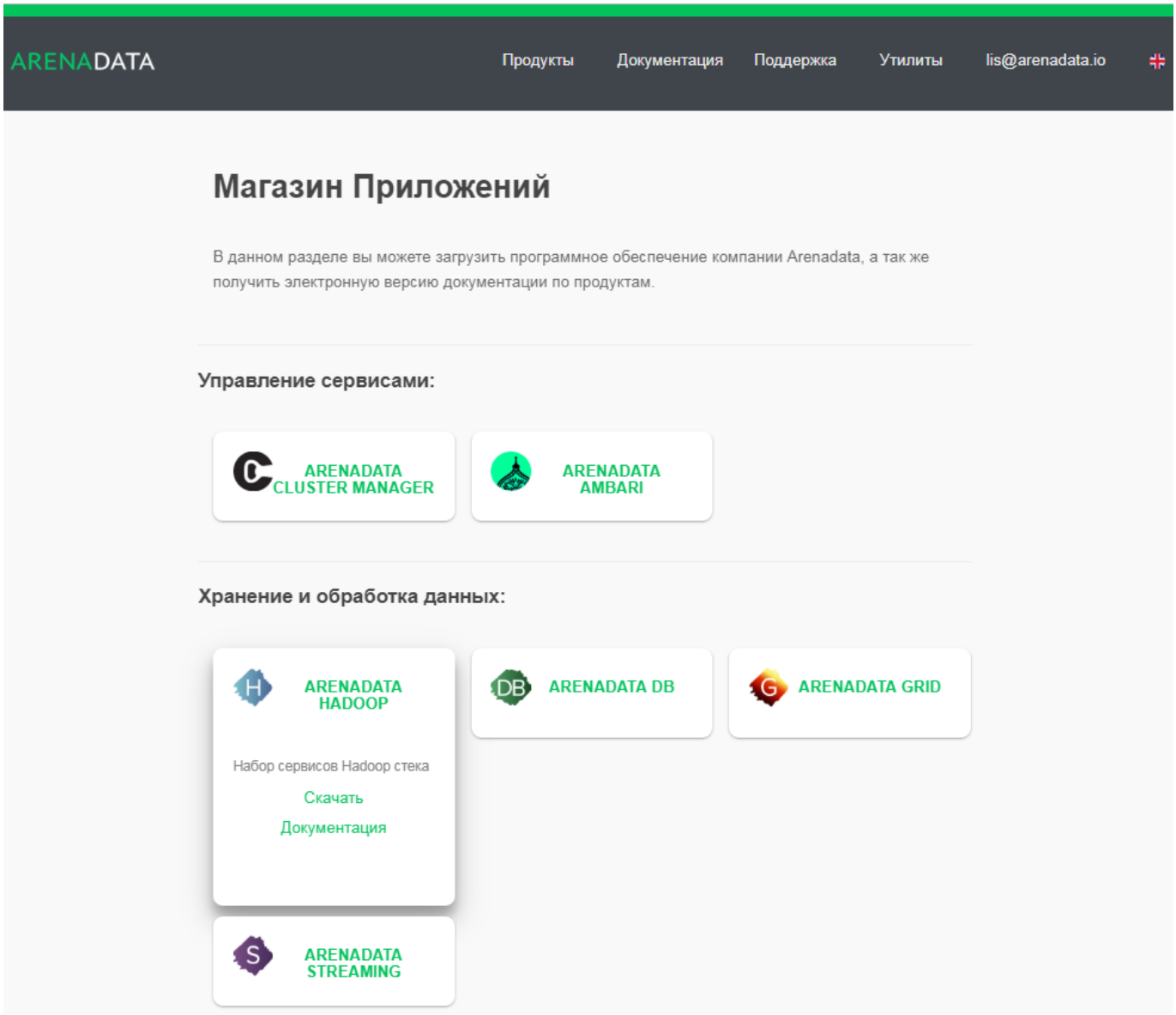


Рис.5.2.: Магазин Приложений

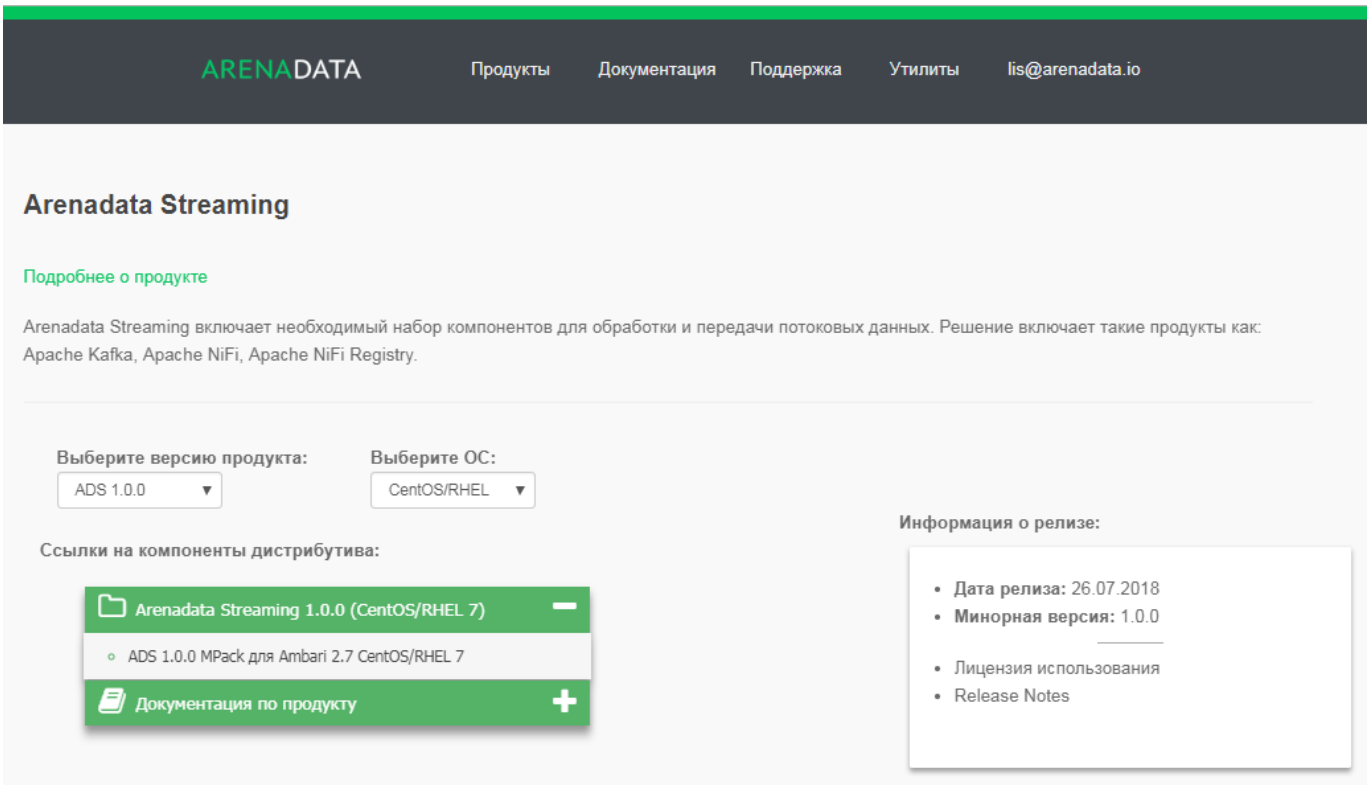


Рис.5.3.: Страница загрузки Arenadata Hadoop

начнёт загрузку. В блоке *“Информация о релизе”* отображается дата выхода, тип релиза выбранной версии, ссылки на лицензию и Release Notes.

Глава 6

Установка ADH в ADCM

6.1 Выгрузка бандла ADH в ADCM

Выгрузка бандла ADH необходима для создания в ADCM прототипа кластера, из которого в дальнейшем возможно создание его экземпляров.

Для выгрузки бандла следует выполнить действия аналогичные *выгрузке бандла хостпровайдера*.

6.2 Создание экземпляра кластера

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

1. Открыть в ADCM вкладку *"CLUSTERS"*:

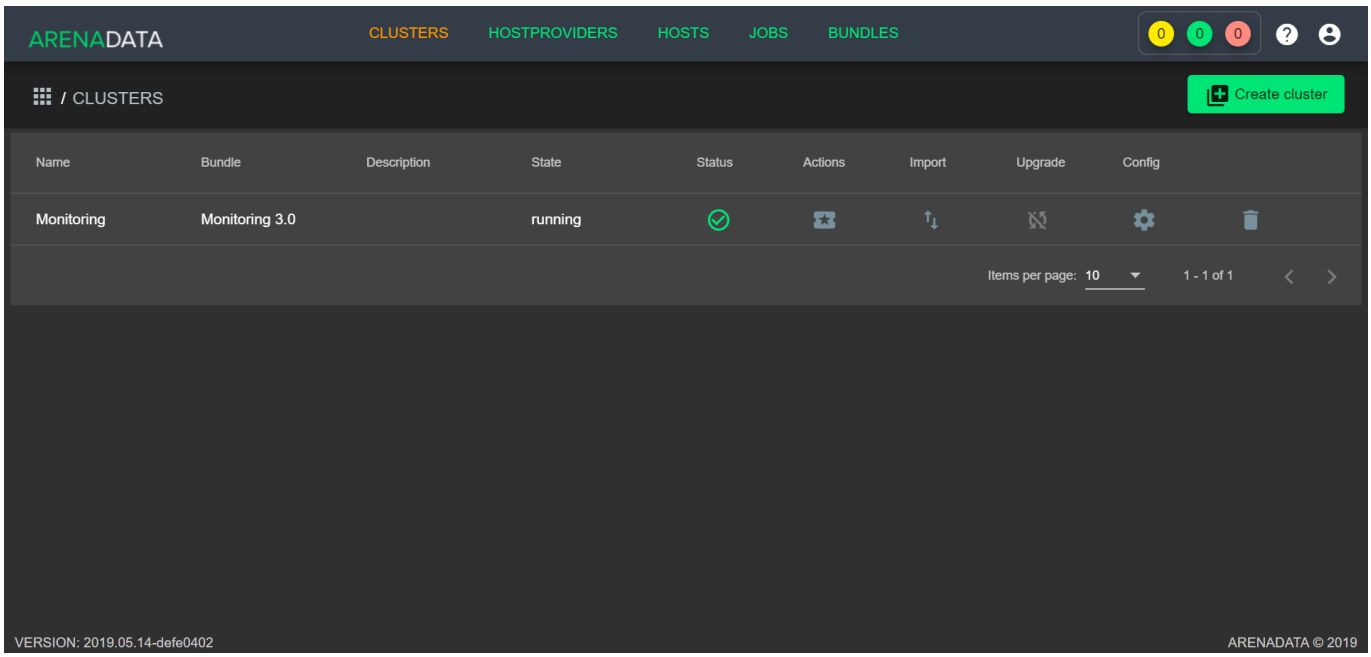


Рис.6.1.: Вкладка *"CLUSTERS"*

2. Нажать *“Add cluster”* и в открывшейся форме создать экземпляр кластера из прототипа ADH:

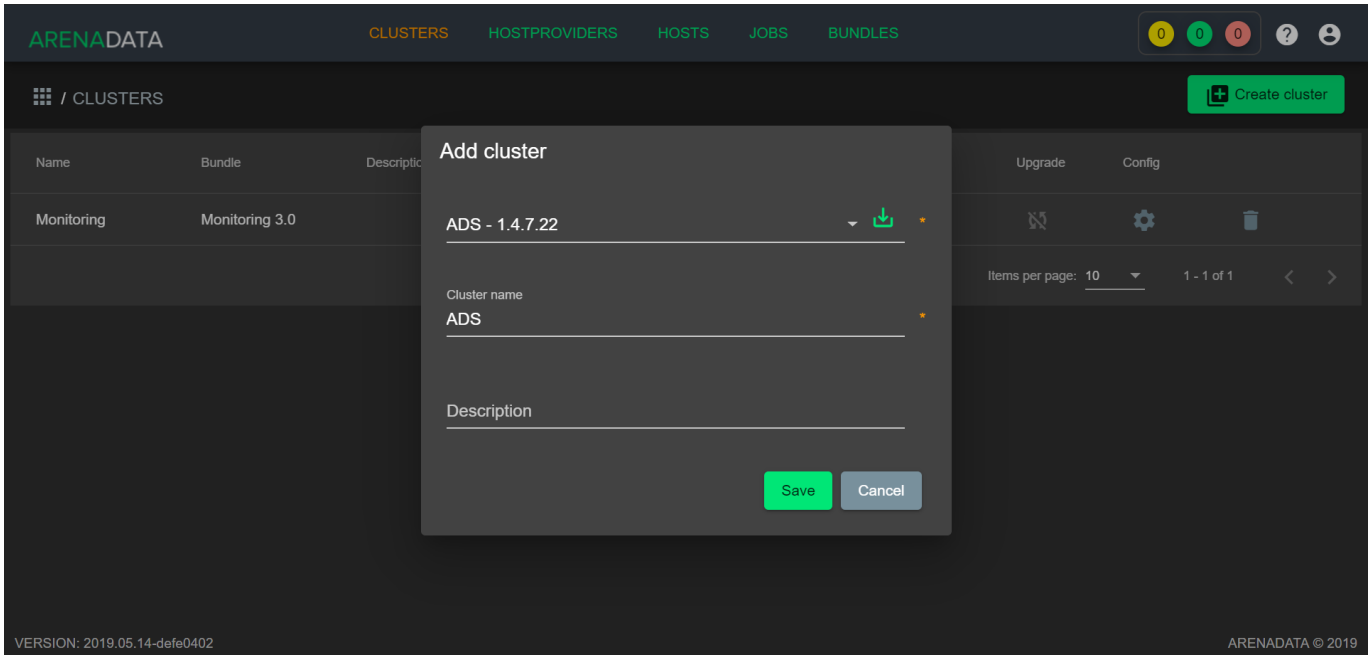


Рис.6.2.: Создание экземпляра кластера

3. В результате экземпляр кластера отразится в списке на вкладке *“CLUSTERS”*:

6.3 Добавление сервисов кластера

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

Для добавления сервисов в кластер ADH необходимо:

1. в меню кластера ADH открыть вкладку *“Services”*;
2. нажать *“Add service”*;

3.в открывшейся форме выбрать необходимые сервисы;

4. нажать *“Save”*.

Возможность добавления нового сервиса в уже работающий кластер не отличается от установки сервиса с нуля.

6.4 Конфигурирование сервиса

Для перехода к настройкам сервиса кластера необходимо нажать кнопку с пиктограммой шестеренки в соответствующей строке вкладки пункта меню *“Services”*.

По завершении конфигурирования сервиса нажмите *“Save”*.

В случае, если сервису требуется изначальная настройка, то в поле *“Actions”* соответствующего сервиса вместо иконки будет отражён оранжевый восклицательный знак:

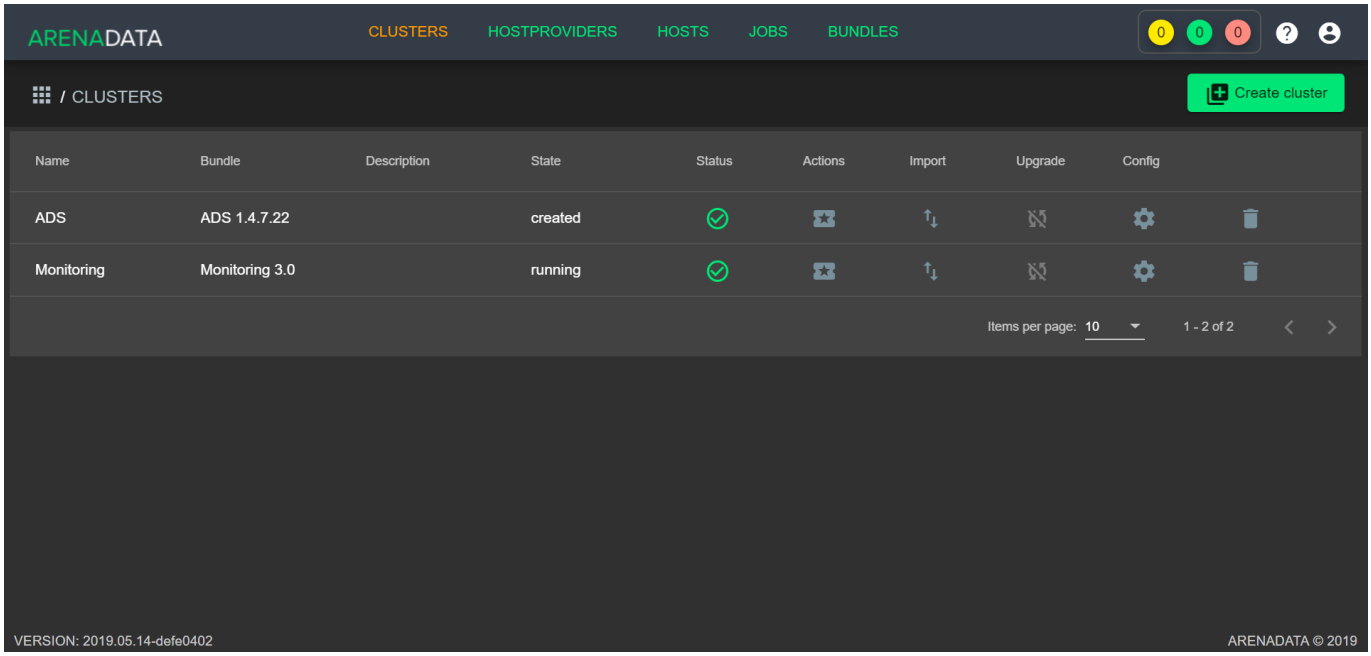


Рис.6.3.: Результат успешного создания экземпляра кластера

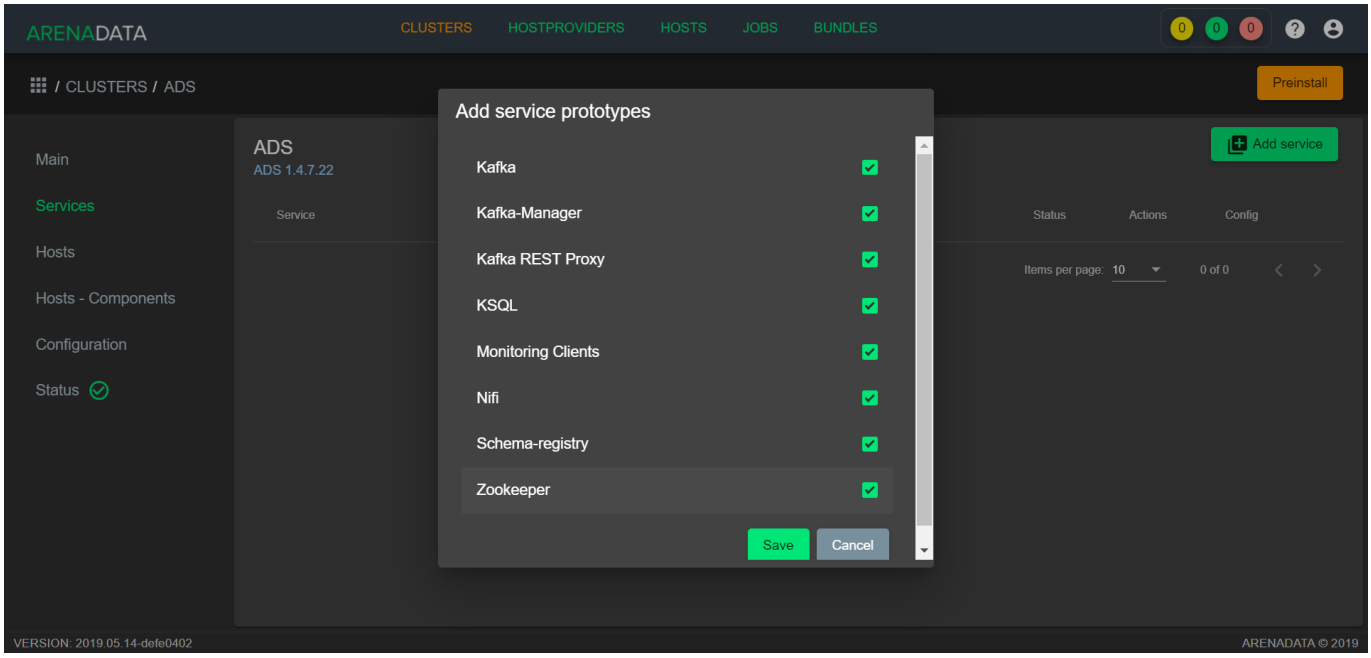


Рис.6.4.: Выбор сервисов

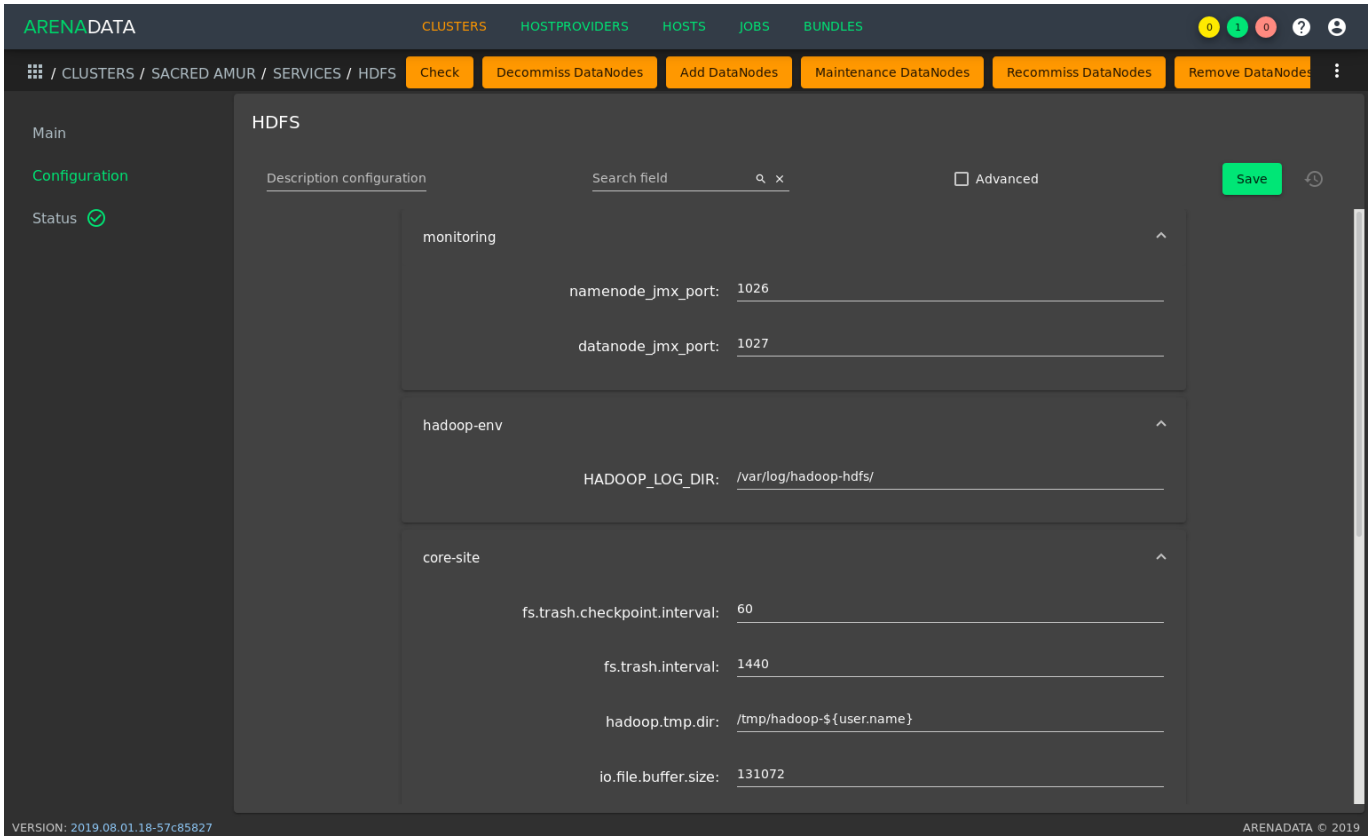


Рис.6.5.: Настройка сервиса

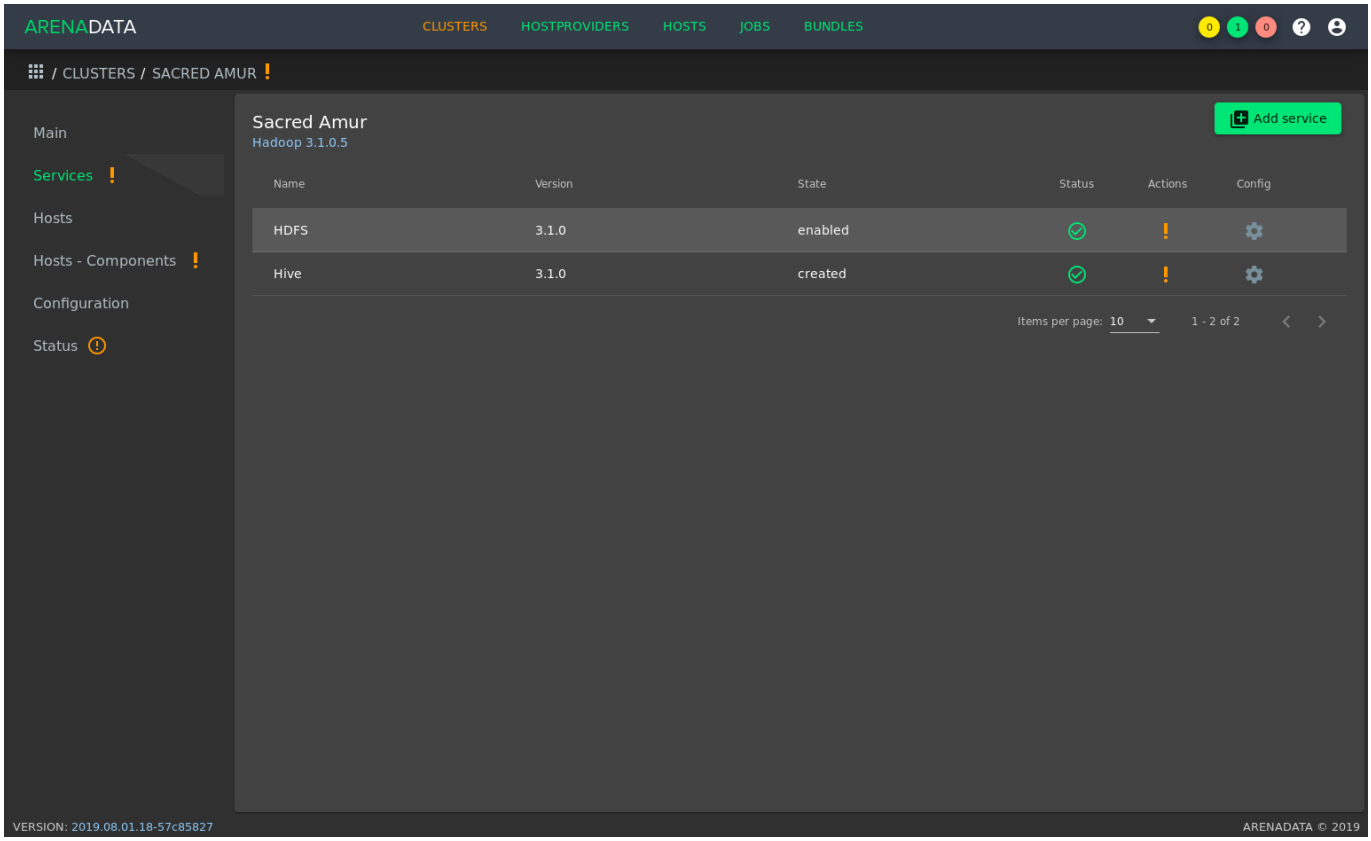


Рис.6.6.: Сервисы, требующие настройки

Это означает, что необходимо открыть страницу конфигурирования сервиса и заполнить поля, выделенные оранжевым:

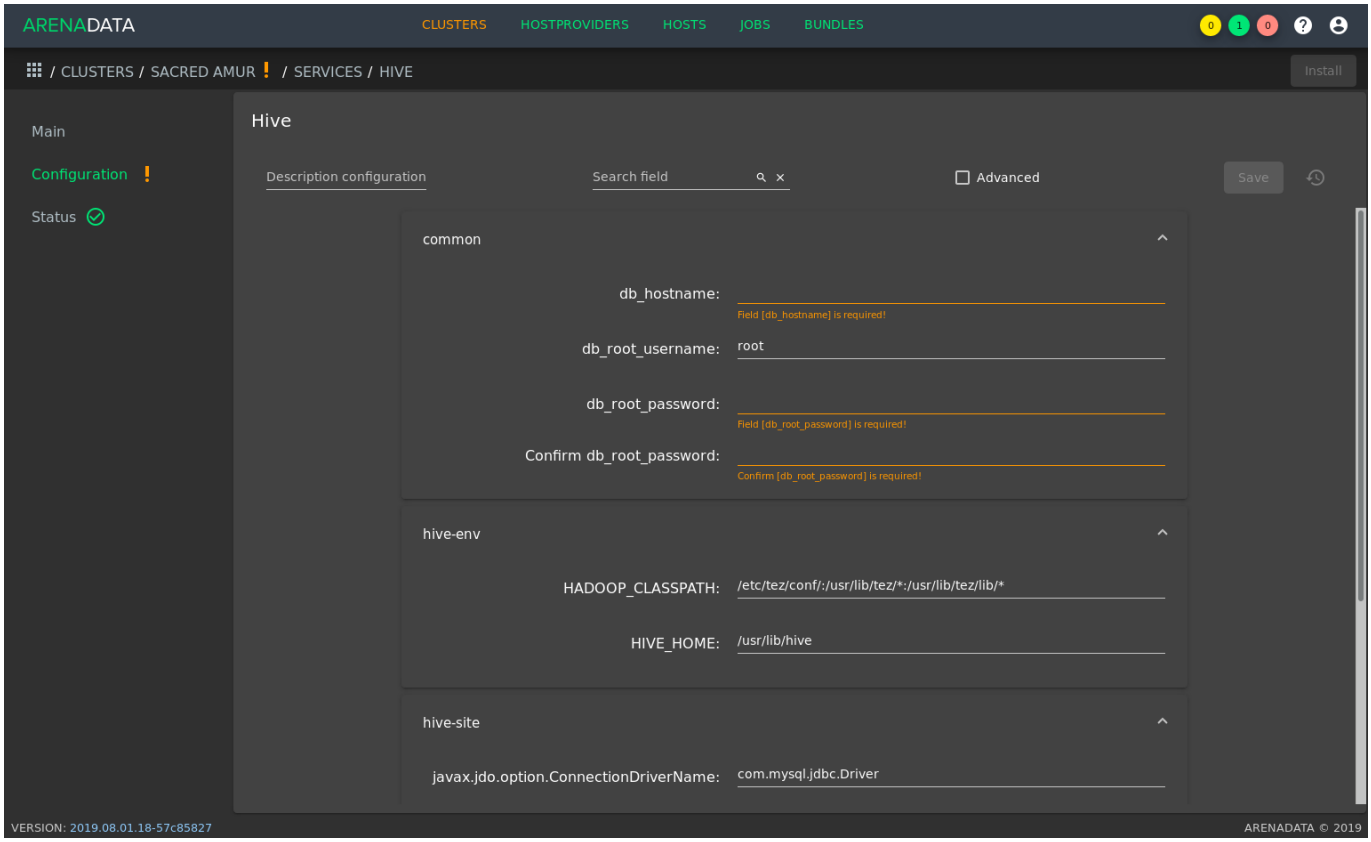


Рис.6.7.: Настройка сервиса, требующего настройки

6.5 Добавление хостов

Для добавления хостов в кластер ADH необходимо:

1. В меню кластера ADH открыть вкладку *“Hosts”*:
2. Нажать *“Add hosts”* и в открывшейся форме выбрать необходимые хосты:
3. В результате добавленные хосты отразятся в списке вкладки *“Hosts”*:

6.6 Размещение компонентов сервисов на хостах

Изначально ни на одном из хостов нет компонентов – распределение компонентов происходит вручную.

Для размещения компонентов необходимо перейти на вкладку *“Hosts - Components”*:

И распределить компоненты одним из двух способов:

1. выбрать компонент в колонке “Components” и определить для него хост в колонке “Hosts”;
2. выбрать хост в колонке “Hosts” и определить для него компонент в колонке “Components”.

Important: В сервисе могут быть обязательные и необязательные компоненты. Если компонент обязательный,

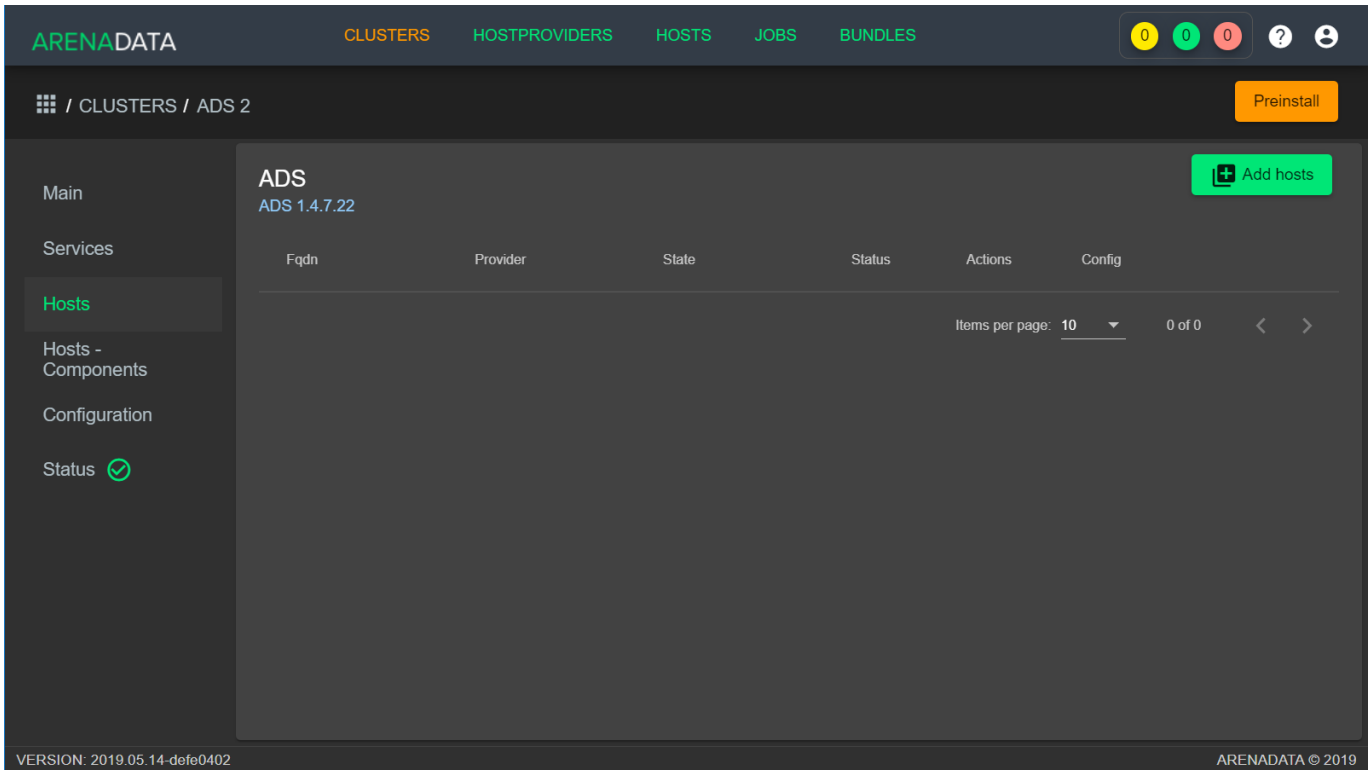


Рис.6.8.: Вкладка *"Hosts"* кластера ADH

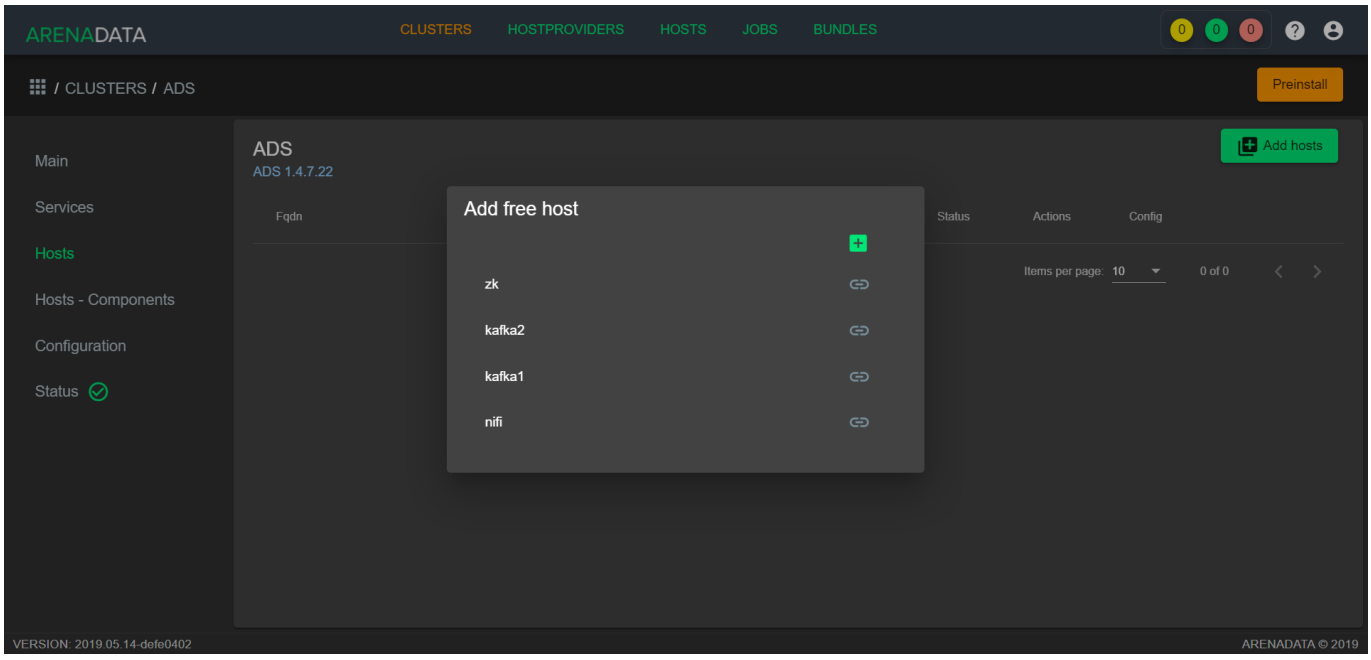


Рис.6.9.: Выбор хостов

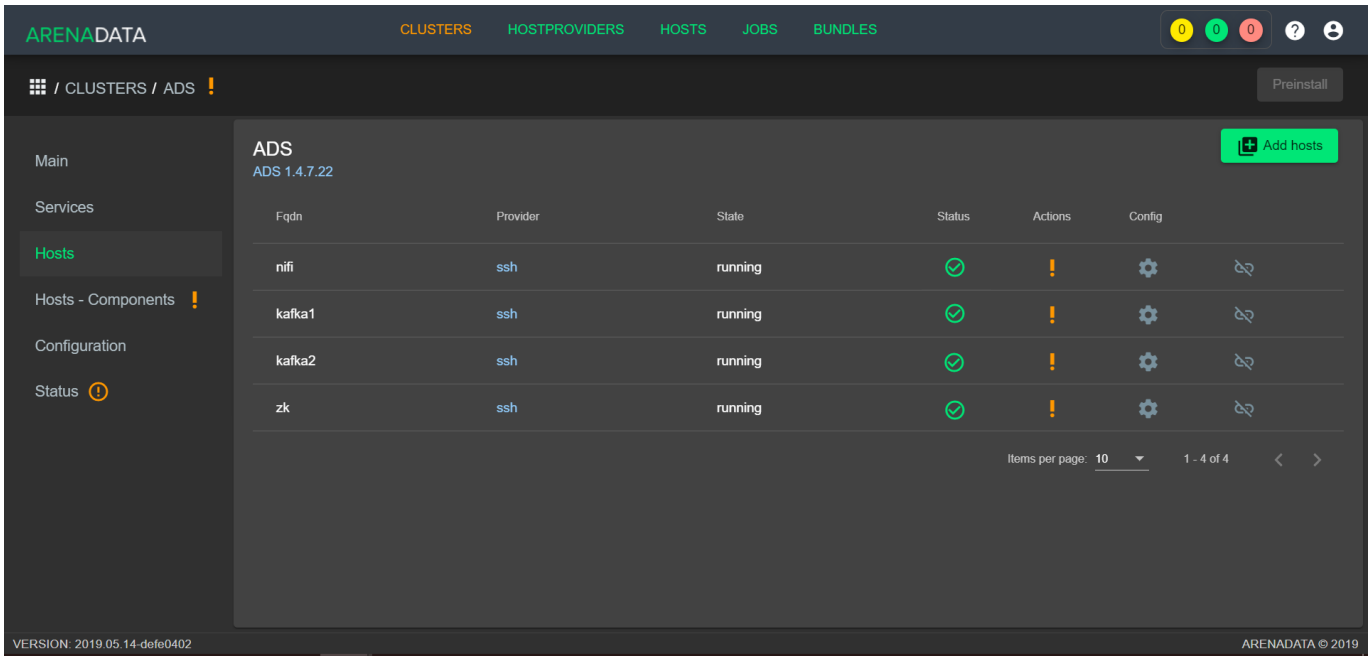


Рис.6.10.: Результат успешного добавления хостов

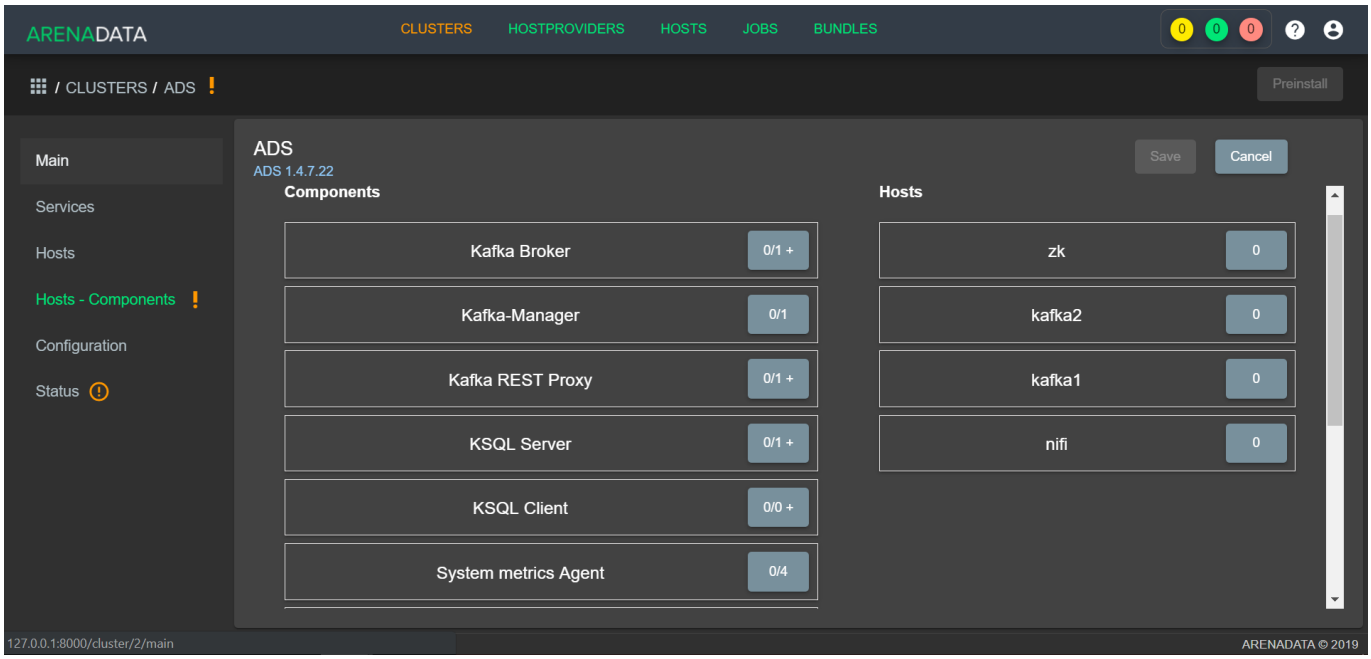


Рис.6.11.: Размещение компонентов сервисов на хостах

то в его количественном счётчике справа от названия присутствует символ ‘/’. Без назначения хоста обязательному компоненту карту размещения сервисов нельзя сохранить.

По завершении распределения хостов нажмите *“Save”*.

6.7 Конфигурирование кластера

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

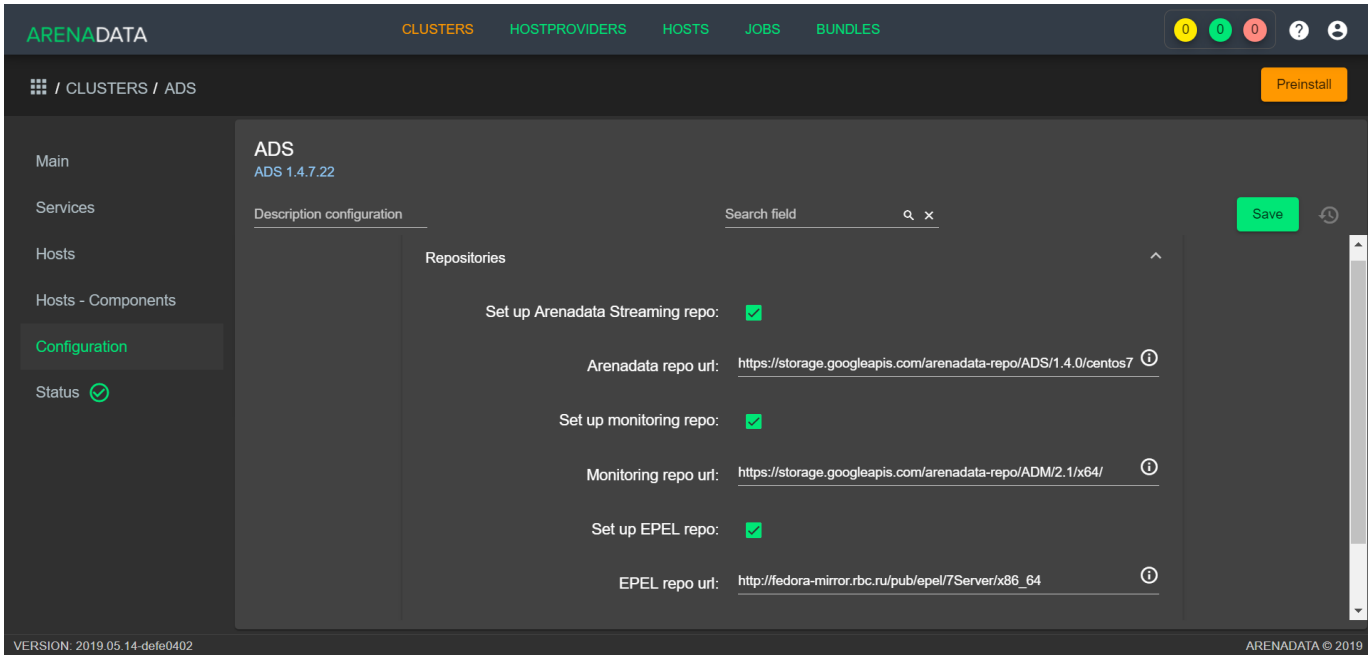


Рис.6.12.: Окно конфигурирования кластера

По завершении конфигурирования кластера нажмите *“Save”*.

6.8 Установка сервисов кластера

Important: Порядок установки сервисов и зависимости между ними на данный момент не ограничивается со стороны ADCM.

Устанавливать сервисы необходимо в следующем порядке:

1. Zookeeper
2. HDFS
3. YARN
4. HBase
5. Hive

6. Spark
7. Monitoring

Для установки добавленного сервиса необходимо в строке нужного сервиса в поле *“Actions”* нажать на пиктограмму и выбрать действие *Install*.

По результатам инсталляции сервис изменяет состояние (поле *“State”*) с *created* на *enabled*.

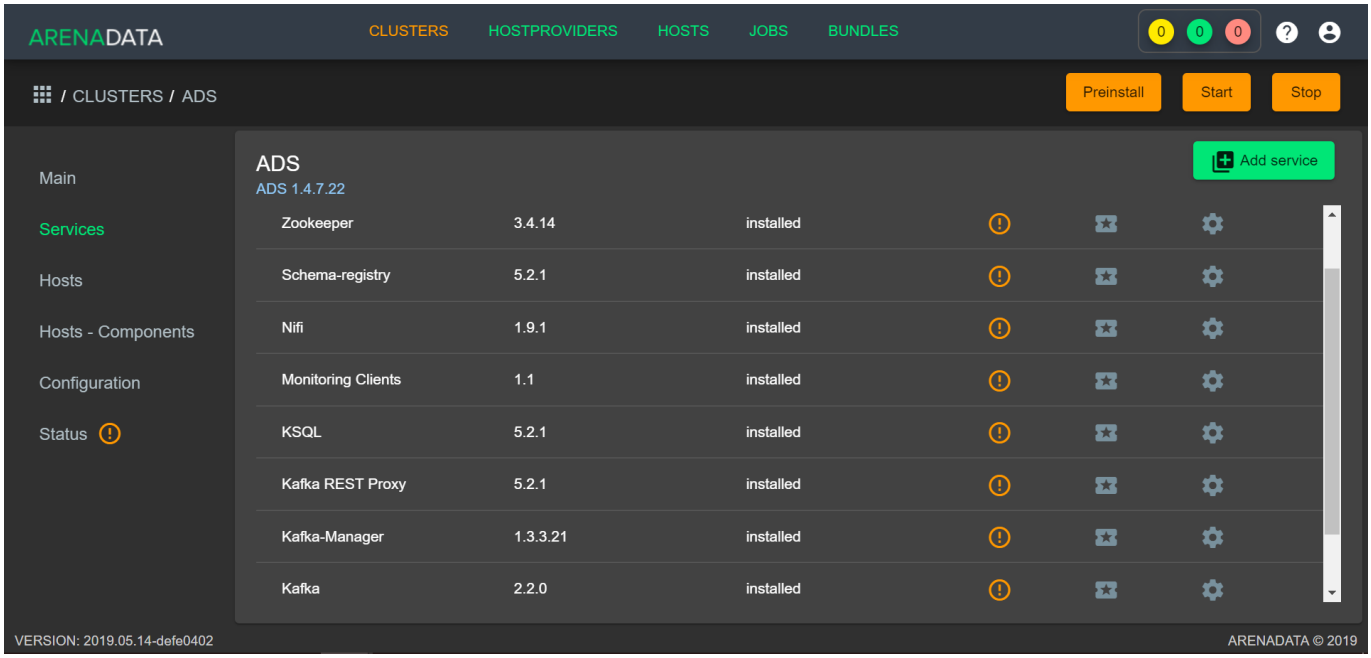


Рис.6.13.: Состояние сервисов кластера

Сервис запускается автоматически.

6.8.1 Особенности установки сервисов

Hive

Компонент Metastore сервиса Hive требует наличие БД MySQL/MariaDB и доступов к ней.

Сервис MySQL/MariaDB в бандл ADH не входит, его необходимо устанавливать и настраивать другими средствами.

Important: Помимо настройки самого сервиса MySQL/MariaDB для использования его сервисом Hive необходимо также настроить соответствующие доступы (GRANT).

Так, самый простой случай выдачи доступов выглядит следующим образом:

```
MariaDB [(none)]> GRANT ALL ON *.* TO 'root'@'<metastore_fqdn>' identified by '<password>' WITH GRANT␣
↪OPTION;
```

Где <metastore_fqdn> и <password> необходимо заменить реальными данными конкретного сервера MySQL/MariaDB.

Указать сервису Hive адрес и учётные данные для подключения к БД возможно в *настройках сервиса*.