|  |
| --- |
|  |



Инструкция по установке

|  |
| --- |
|  |

|  |
| --- |
|  |

Москва

2016

NEOFLEX DATAGRAM

Компания «Неофлекс»

Адрес: Россия, г. Москва, 127015, ул. Вятская, д. 35, стр. 4, 1 подъезд, 2 этаж

Тел.: +7 (495) 984-25-13

# Содержание

[Содержание 2](#_Toc471734259)

[1 Требования 3](#_Toc471734260)

[1.1 Проверка наличия установленных программ 3](#_Toc471734261)

[2 Комплект поставки Neoflex Datagram 5](#_Toc471734262)

[3 Установка 6](#_Toc471734263)

[3.1 Подготовка к установке 7](#_Toc471734264)

[3.2 Загрузка компонентов программы 15](#_Toc471734265)

[3.3 Запуск компонентов программы 16](#_Toc471734266)

[3.4 Контроль запуска компонентов программы 17](#_Toc471734267)

[4. Приложение 17](#_Toc471734268)

[4.2 Приложение 1. Как узнать значение oozie\_api? 17](#_Toc471734269)

[4.3 Приложение 2. Как узнать значение hdfs\_defaultFS? 18](#_Toc471734270)

[4.4 Приложение 3. Проверка наличия корневого каталога HDFS 19](#_Toc471734271)

[4.5 Приложение 4. Как узнать значение hdfs\_default\_user? 20](#_Toc471734272)

[4.6 Приложение 5. Команды Docker 20](#_Toc471734273)

[4.7 Приложение 6. Пример создания базы данных и пользователя 21](#_Toc471734274)

# Требования

Предварительно в системе должны быть установлены и доступны программы:

1. **Hartoonworks Data Platform** (<http://hortonworks.com/products/data-center/hdp/>) с компонентами:

* **HDFS**;
* **Oozie**;
* **YARN**;
* **Hive**;
* **ZooKeeper.**

При установке Hartoonworks Data Platform необходимо сохранить (записать):

* Путь к каталогу, в котором находится конфигурация Hadoop;
* Путь к каталогу пользователя hdfs (на компьютере, а не в Hadoop).

1. **Docker** (<https://www.docker.com>) Для пользователя, под которым работает Docker, должны быть установите права на чтение служебных каталогов, входящих в комплект поставки (см. раздел «[Комплект поставки Neoflex Datagram](#_Комплект_поставки_Neoflex)»).
2. **PostgreSQL** (<https://www.postgresql.org>). В PostgreSQL должны быть созданы:

* База данных **teneo** (название базы может быть произвольным);
* Пользователь, от имени которого Neoflex Datagram будет работать с базой данных. Пользователю должны быть заданы соответствующие права (пример создания базы данных и пользователя представлен в «[Приложение 6](#_Приложение_6._Скрипты)»).

## Проверка наличия установленных программ

1. Проверьте, что в системе доступен сервер **Ambari**. Для этого в адресной строке браузера введите адрес:

http://<ambarihost>:8080/#/main/dashboard/metrics

На экране должна появиться форма регистрации пользователя.

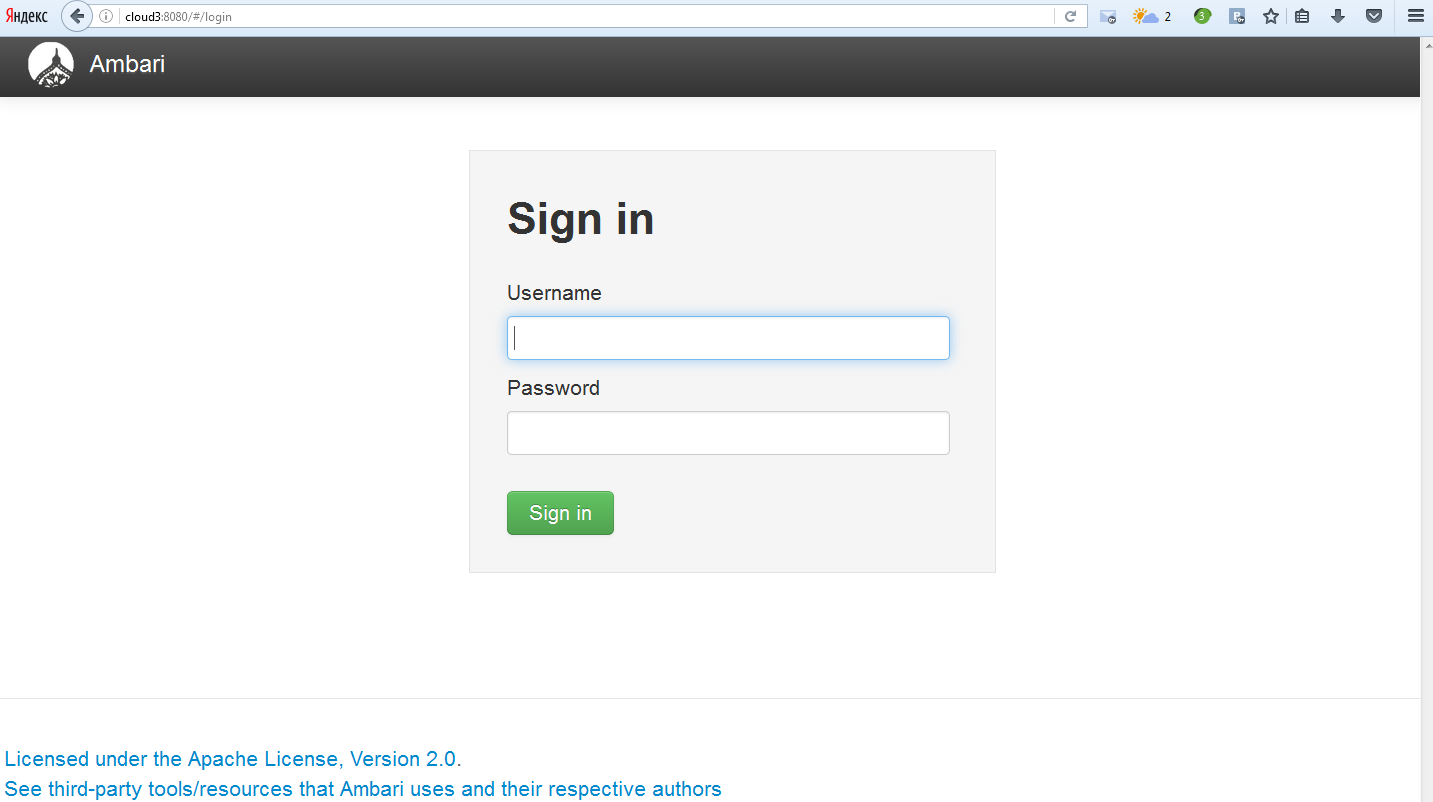


Рисунок 1. Форма регистрации пользователя Ambari

Выполните вход в систему. По умолчанию для входа используются значения:

* **Username – admin;**
* **Password – admin**.

В окне браузера отобразится главная страница сервера **Ambari**.

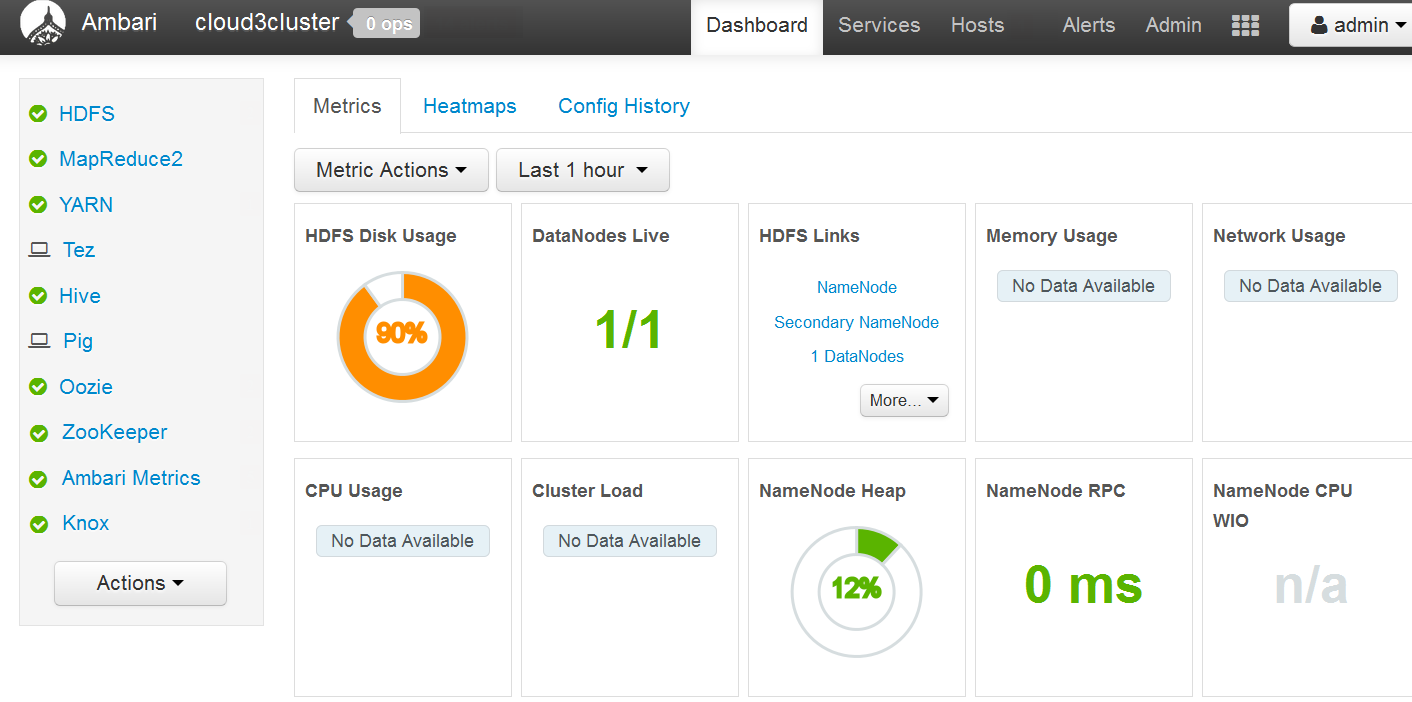


Рисунок 2. Главная страница Ambari

1. Убедитесь, что установлены и запущены компоненты: **HDFS, YARN, Hive, Oozie, ZooKeeper** – названия данных компонентов отображаются на панели, расположенной в левой части экрана, и рядом с названием компонента установлена пиктограмма  (см. рис 2).
2. Проверьте, что на компьютере, где будет проводиться инсталляция Neoflex Datagram, доступен **Docker** и установлена **PostgreSQL**. Для этого в командной строке консоли выполните команду: ***docker ps***



Рисунок 3. Результат выполнения команды docker ps, если Docker доступен

# Комплект поставки Neoflex Datagram

Представителями компании Neoflex передается заказчику каталог поставки, в котором находятся:

1. Набор **docker image** **файлов**, которые будут загружаться в Docker:
   * **mserver.tar**;
   * **dbmonitor.tar**;
   * **jobServer.tar**.
2. Файл настройки запуска контейнеров: **install\_env**.
3. Файл настройки подключения к корпоративному серверу каталогов (LDAP): **ldap.properties**;
4. Каталоги, в которых расположены скрипты запуска каждого из контейнеров **run.sh**.
5. Служебные каталоги:
   * **MavenRepository**;
   * **OozieShareLib**;
   * **SparkJarLib**;
   * **MSpaceDeployments** –входит в поставку, если заказчику поставляются индивидуальные настройки системы (см. раздел «Подсистема переноса метаданных» Руководства пользователя).

# Установка

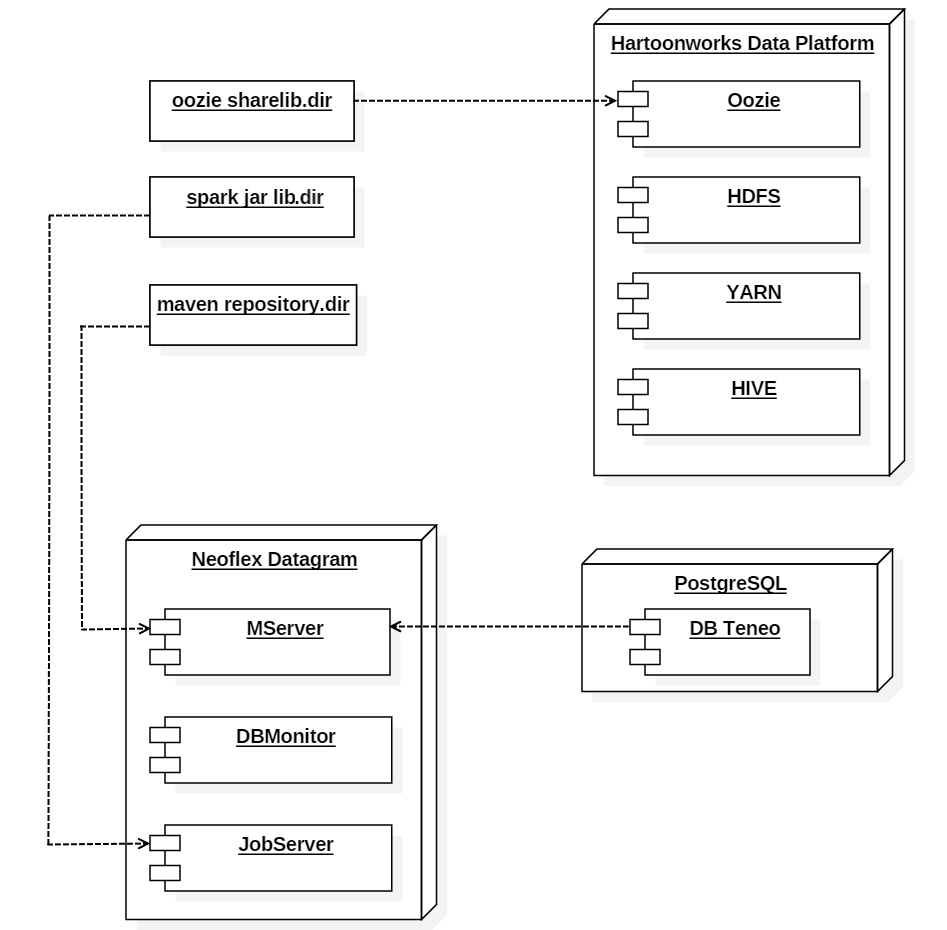


Рисунок 4. Диаграмма развертывания Neoflex Datagram

## Подготовка к установке

1. На компьютер, где находится **Docker**, скопируйте каталог поставки, предоставленный представителями компании Neoflex. При копировании должна быть сохранена структура.
2. Из каталога поставки переместите служебные каталоги в любое место на целевом компьютере.

Например:

/etc/DataGram/Maven repository

/etc/DataGram/OozieShareLib

/etc/DataGram/SparkJarLib

/etc/DataGram/MSpaceDeployments (может отсутствовать в поставке)

Если каталог **MSpaceDeployments** отсутствует в комплекте поставки, то на целевом компьютере его необходимо создать вручную.

1. На целевом компьютере создайте каталог для хранения конфигурационных файлов программы, например:

/etc/DataGram/conf

1. Шаг выполняется, если требуется настройка подключения к корпоративному серверу каталогов (LDAP). Из каталога поставки скопируйте файл **ldap.properties** в каталог хранения файлов конфигурации (см. шаг 3). Отредактируйте содержимое файла.

**Таблица 1.** Описание файла ldap.properties

|  |  |  |
| --- | --- | --- |
| Параметр | Обязательный | Описание |
| ldap.domain | Да | Доменное имя Ldap сервера  *На заметку.*  *Для авторизации по LDAP используется userPrincipalName вида:* ***username@domainname.com*.** *Если поле не заполнено, то при авторизации необходимо указывать userPrincipalName полностью. Если domain указан, то допускается ввод только userName* |
| ldap.host | Да | Имя хоста Ldap сервера |
| ldap.port | Да | Порт Ldap сервера |
| ldap.base | Да | Путь к каталогу для поиска пользователей |
| ldap.admin\_group | Да | Название группы пользователей, которым будут предоставлены права администратора |
| ldap.operator\_group | Да | Название группы пользователей, которым будут предоставлены права оператора |
| ldap.user\_group | Да | Название группы пользователей, которым будут предоставлены права пользователя |

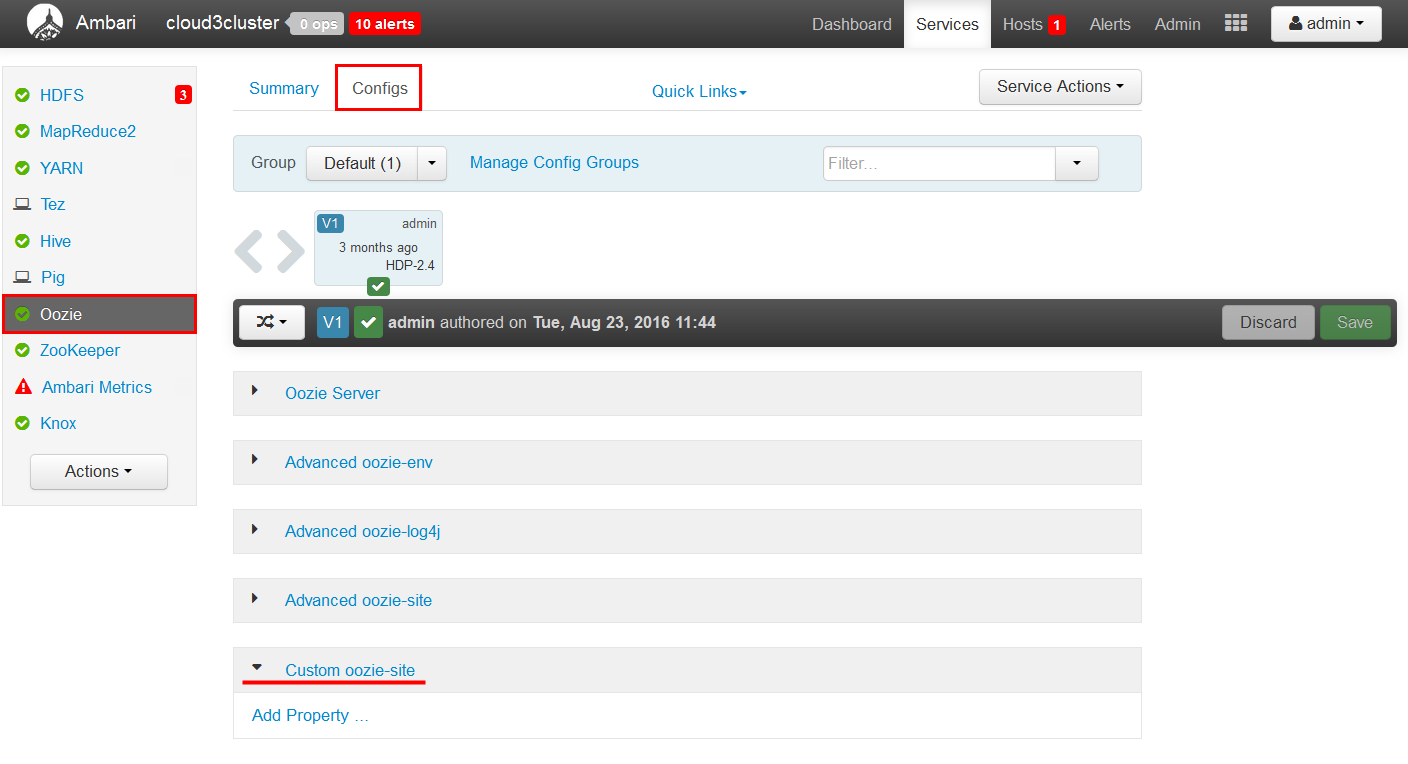
* Администратор – пользователи имеют полный доступ к функционалу программы;
* Оператор – пользователи могут просматривать логи программы, работать с консолями Oozie, Livy, HDFS, разрабатывать и запускать на исполнение объекты Workflow и Transformation;
* Пользователь – доступны консоли Oozie, Livy, HDFS, просмотр логов программы, запуск объектов Workflow и Transformation на исполнение.

1. Отредактируйте файл **Install\_env**, расположенный в корне каталога поставки.

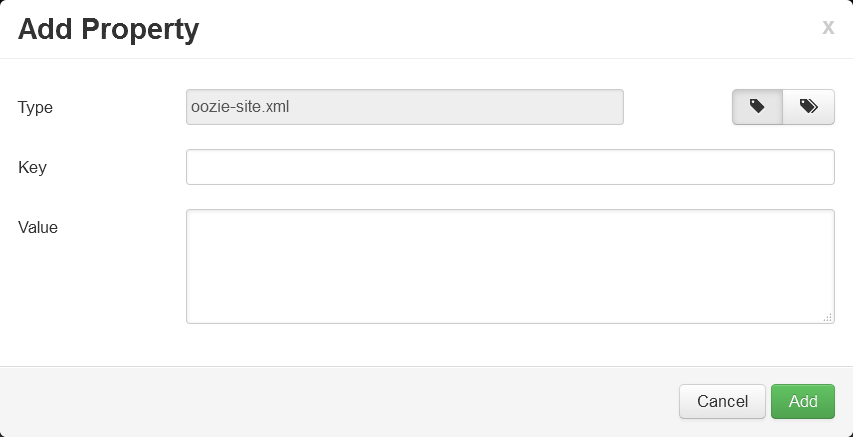
**Таблица** **2.** Описание параметров запуска в файле Install\_env

|  |  |
| --- | --- |
| Параметр | Описание |
| **Названия контейнеров**  В секции задаются имена, которые будут присваиваться контейнерам при запуске.  Указываемые имена контейнеров не должны совпадать между собой | |
| postgree\_name | Название контейнера postgree  Пример: postgree2 |
| dbmonitor\_name | Название контейнера dbmonitor  Пример: dbmonitor |
| mserver\_name | Название контейнера mserver  Пример: mserver |
| jobserver\_name | Название контейнера jobserver  Пример: jobserver |
| **external port on host**  В секции задаются номера портов целевого компьютера, на которых будут размещаться контейнеры.  Для контейнеров могут быть установлены любые свободные порты, но их номера не должны совпадать между собой | |
| dbmonitor\_port | Порт контейнера dbmonitor  Пример: 8092 |
| jobserver\_port | Порт контейнера jobserver  Пример: 8090 |
| mserver\_port | Порт контейнера mserver  Пример: 8084 |
| postgree\_port | Порт контейнера postgree  Пример: 5435 |
| **spark shared lib jars folder**  В секции задаются каталоги, в которых хранятся jar файлы | |
| sparklib | Каталог, в котором хранятся jar файлы для jobServer (см. шаг 2 данного пункта инструкции)  Пример: /etc/spark jar lib |
| maven\_rep | Каталог, где хранится репозиторий maven (см. шаг 2 данного пункта инструкции)  Пример: /etc/maven repository |
| hadoop\_conf\_dir | Каталог, в котором находится конфигурация Hadoop. Данный каталог создается в момент установки Hartoonworks Data Platform.  Расположение каталога должно быть известно системному администратору  Пример: /usr/hdp/2.4.2.0-258/hadoop/conf |
| postgree\_data | Каталог, где контейнер postgree будет хранить данные.  Значение не указывается, если PostgreSQL был установлен заранее. Расположение каталога должно быть известно системному администратору  Пример: /home/pdata2 |
| js\_userdir | Каталог пользователя hdfs (на компьютере, а не в Hadoop). Данный каталог создается в момент установки Hartoonworks Data Platform.  Расположение каталога должно быть известно системному администратору  Пример: /home/hdfs |
| customer\_code | Код заказчика. Значение должно быть установлено поставщиком и не требует редактирования  Пример: BPN |
| mspace\_dir | Каталог, где mserver будет хранить созданные артефакты (см. шаг 2 данного пункта инструкции)  Пример: /etc/mspace deployments |
| **teneo**  В секции задаются параметры подключения к базе данных | |
| host\_url | Хост, где размещается БД teneo (может указывать внутрь Docker контейнера Postgres, в таком случае – это хост, где находится Docker).  Значение должно быть известно системному администратору  Пример: cloud2.neoflex.ru |
| teneouser | Логин для подключения к БД. Значение должно быть известно системному администратору  Пример: postgres (значение по умолчанию) |
| teneopwd | Пароль для подключения к БД. Значение должно быть известно системному администратору  Пример: postgres (значение по умолчанию) |
| teneoname | Имя базы данных teneo  Пример: teneo |
| **dbmonitor**  В секции задаются параметры dbmonitor | |
| oozie\_api | API Oozie.  Значение можно узнать при помощи Oozie Web Console (см. «[Приложение 1](#_Приложение_1._Как)»)  Пример: http://cloud3:11000/oozie/v2 |
| hdfs\_defaultFS | Параметры доступа к hdfs.  Значение можно узнать в Ambari (см. «[Приложение 2](#_Приложение_2._Как)»)  Пример: hdfs://192.168.3.90:8020 |
| hdfs\_baseJobPath | Корневой каталог hdfs.  По умолчанию: /user. Ее наличие можно проверить при помощи Ambari (см «[Приложение 3](#_Приложение_3._Проверка)»)  Пример: /user |
| hdfs\_default\_user | Имя пользователя hdfs.  Значение можно узнать в Ambari (см. «[Приложение 4](#_Приложение_4._Как)»)  Пример: hdfs |
| mserver\_api | URL mserver.  Значение складывается из адреса хоста Docker и порта, на котором будет хостится mserver  Пример: http://cloud:8084 |
| dbmonitor\_link | URL адрес BDMonitor, используемый для перехода из интерфейса MetaServer  Пример: http://cloud3:8090/cim/dbmonitor/pim/ui/psm/angular/WEB/index.html#/dbmonitor |
| **version**  В секции задаются версии устанавливаемых дистрибутивов | |
| mserver\_version | Версия дистрибутива mserver.  Значение устанавливается поставщиком и не требует редактирования  Пример: 0.1-SNAPSHOT |
| dbmonitor\_version | Версия дистрибутива dbmonitor.  Значение устанавливается поставщиком и не требует редактирования  Пример: 0.1-SNAPSHOT |
| **security data**  В секции задаются дополнительные параметры запуска | |
| config\_dir | Путь к каталогу, в котором хранятся файлы конфигурации  Пример: /etc/DataGram/conf |
| jvm\_args | Дополнительный параметр запуска. Может принимать значения:   * **-Dldap.enabled=false** – не использовать подключение к корпоративному серверу каталогов (LDAP); * **-Dldap.config=/root/conf/ldap.properties** – использовать параметры подключения к корпоративному серверу каталогов (LDAP). Значение должно быть именно таким, независимо от того, где находится файл ldap.properties |

1. Выполните настройку автоматической очистки списков отработанных заданий Oozie. Для этого, откройте главную страницу Ambari (см. рис 2). На панели, расположенной в левой части экрана, выберите пункт **Oozie**. На появившейся странице перейдите на вкладку «**Configs**», далее разверните список «**Custom oozie-site**».

Рисунок 5. Конфигурация Oozie

Кликните по ссылке **«Add Property»**, на экране появится форма создания свойства.

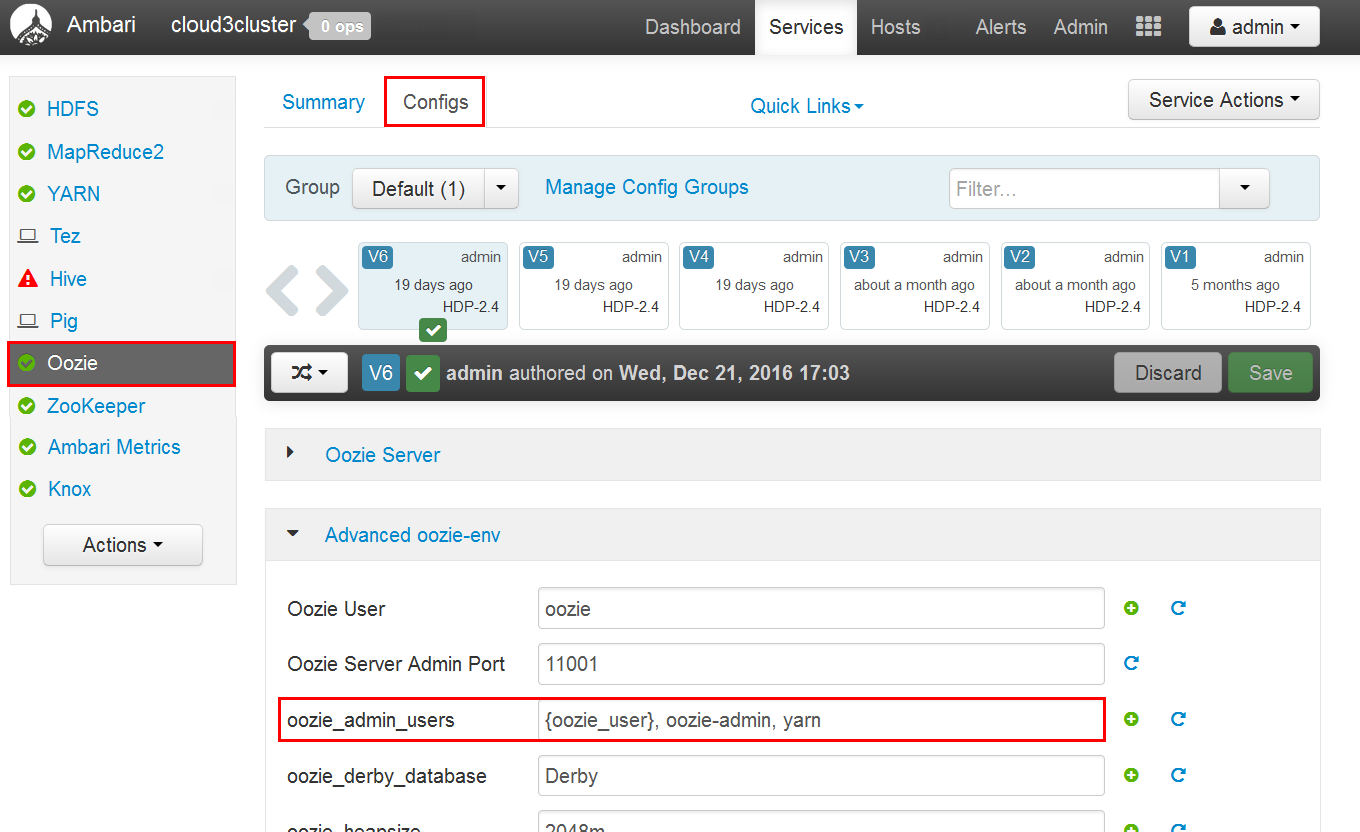
Рисунок 6. Форма создания свойства

Создайте свойства в соответствии со значениями, представленными в таблице ниже.

**Таблица 3.** Описание свойств Oozie для удаления списков отработанных заданий.

|  |  |  |
| --- | --- | --- |
| Key | Value | Описание |
| oozie.service.PurgeService.older.than | 14 | Длительность хранения отработанных workflow jobs. Значение задается в днях |
| oozie.service.PurgeService.coord.older.than | 30 | Длительность хранения отработанных coordinator jobs. Значение задается в днях |
| oozie.service.PurgeService.bundle.older.than | 30 | Длительность хранения отработанных bundle jobs. Значение задается в днях |
| oozie.service.PurgeService.purge.limit | 100 | Ограничение количества удаляемых заданий за одну итерацию очистки |
| oozie.service.PurgeService.purge.interval | 3600 | Интервал времени (в секундах) между запусками процедуры очистки |
| oozie.service.PurgeService.purge.old.coord.action | true или  falce | Если установлено значение true, то будет выполняться удаление устаревших workflow, даже если координатор, создавший их, все ещё работает.  Если установлено значение falce (по умолчанию), то такие workflow удаляться не будут. |

1. Для корректной работы программы в список администраторов **Oozie** необходимо добавить системное имя пользователя **YARN**. Для этого, откройте главную страницу Ambari (см. рис 2). На панели, расположенной в левой части экрана, выберите пункт **Oozie**. На появившейся странице перейдите на вкладку «**Configs**», далее разверните список «**Advanced oozie-env**».

Рисунок 7. Список администраторов Oozie

В поле «**oozie\_admin\_users»** к уже имеющимся там пользователям через запятую добавьте имя пользователя YARN в системе (на рисунке 7 имя пользователя YARN: «yarn»).

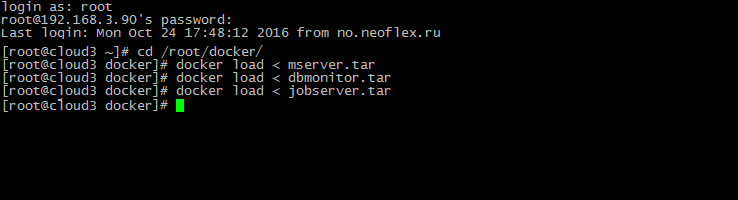
## Загрузка компонентов программы

Компоненты программы **Mserver**, **DBMonitor** и **JobServer** могут быть загружены одним из двух способов, в зависимости от наличия подключения к интернету на целевом компьютере.

**1-ый способ – на целевом компьютере (где установлен Docker) отсутствует подключение к интернету.**

В этом случае загружаются уже собранные docker images. Для загрузки перейдите в каталог, где хранятся docker images, и выполните команды:

1. ***docker load < mserver.tar*** – загрузка Docker контейнера mserver;
2. ***docker load < dbmonitor.tar***– загрузка Docker контейнера dbmonitor;
3. ***docker load < jobserver.tar*** – загрузка Docker контейнера jobserver.

Рисунок 8. Результат выполнения команд загрузки контейнеров

**2-ой способ – целевой компьютер (где установлен Docker) имеет подключение к интернету.**

Если на целевом компьютере имеется подключение к интернету, то сборку контейнеров можно выполнить командой ***./build.sh****.*

*Внимание!*

*Команда* ***./build.sh*** *должна выполняться из соответствующих каталогов:*

*/docker/dbmonitor;*

*/docker/job-server;*

*/docker/mserver.*

Если в каталогах отсутствуют файлы **build.sh**, тогда выполните действия, описанные в следующем пункте инструкции («[Запуск компонентов программы](#_Запуск_компонентов_программы)»).

## Запуск компонентов программы

1. Для скриптов **run.sh**, расположенных в каталогах:

/docker/dbmonitor;

/docker/job-server;

/docker/mserver.

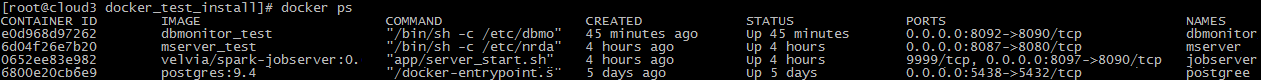
установите разрешение на запуск, выполнив команду ***chmod +x run.sh.***

1. Выполните запуск всех скриптов **run.sh** при помощи команды ***./run.sh***

Каждый скрипт запускает соответствующий контейнер с настройками, записанными в файле **Install\_env**.

## Контроль запуска компонентов программы

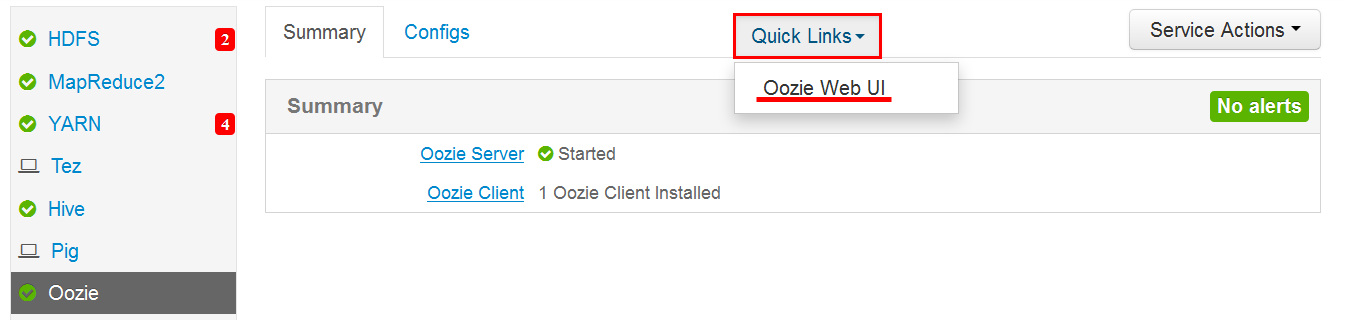
Выполните команду ***docker ps****,* которая выведет на экран таблицу со списком запущенных компонентов программы с указанием статусов. Убедитесь, что в столбце «**NAMES**» списка присутствуют компоненты: **postgee**, **mserver**, **dbmonitor**, **jobserver**. В столбце «**STATUS**» всех компонентов должен быть установлен статус: **Up**.

Рисунок 9. Результат выполнения команд загрузки контейнеров

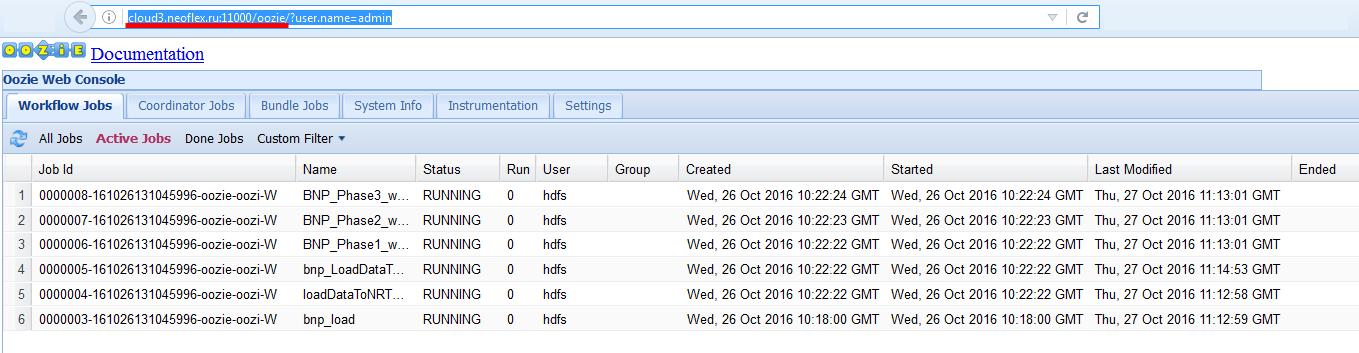
# Приложение

## Приложение 1. Как узнать значение oozie\_api?

Значение oozie\_api можно получить через интерфейс Oozie Web Console. Для перехода в Oozie Web Console зайдите на главную страницу Ambari (см. рис 2). На панели, расположенной в левой части главной страницы, кликните по строке **Oozie**.

Рисунок 10. Переход в интерфейс Oozie Web Console

На открывшейся странице кликните по кнопке «**Quick Links**» и в появившемся списке выберите «**Oozie Web UI**». Откроется новая вкладка с интерфейсом Oozie Web Console.

Рисунок 11. Oozie Web Console

В адресной строке браузера будет указано значение,

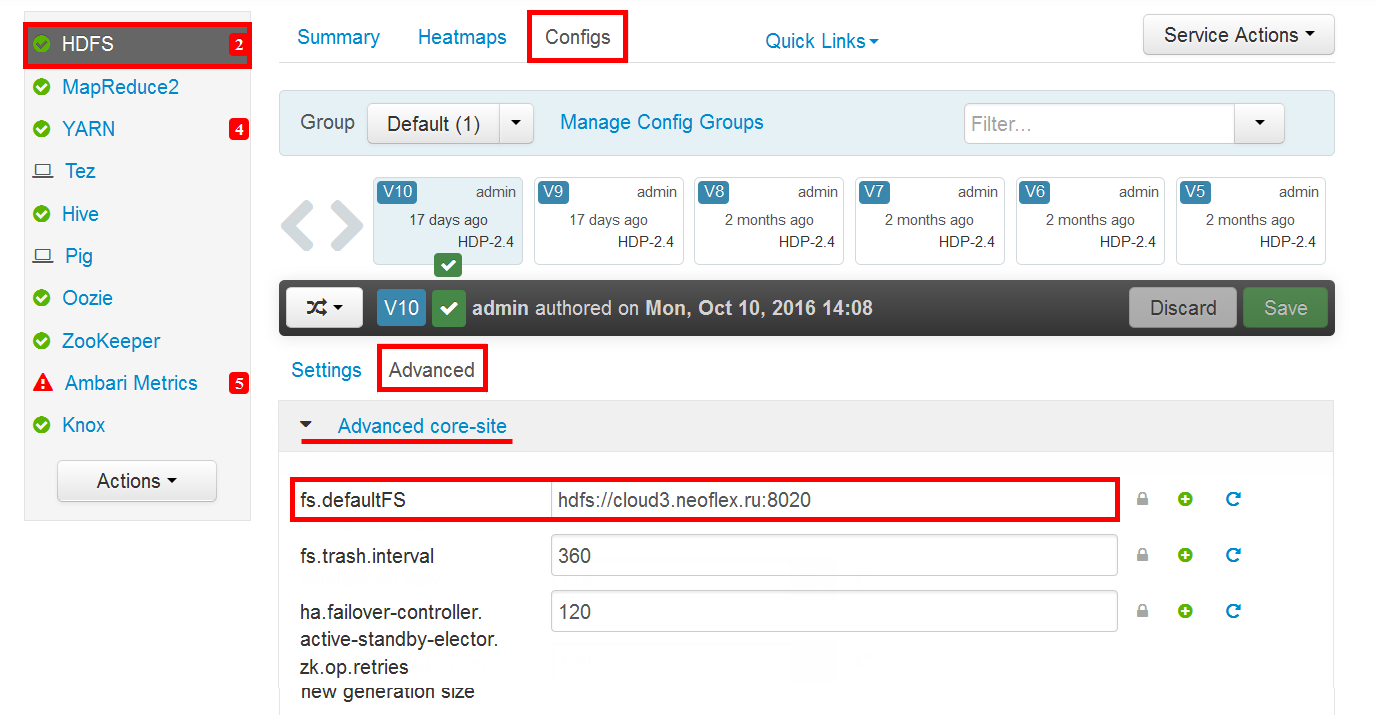
например: http://cloud3.neoflex.ru:11000/oozie/?user.name=admin

Чтобы получить значение oozie\_api, замените все, что идет после http://cloud3.neoflex.ru:11000/oozie/, на значение **v2.**

Таким образом, значение oozie\_api будет: **http://cloud3.neoflex.ru:11000/oozie/v2**

## Приложение 2. Как узнать значение hdfs\_defaultFS?

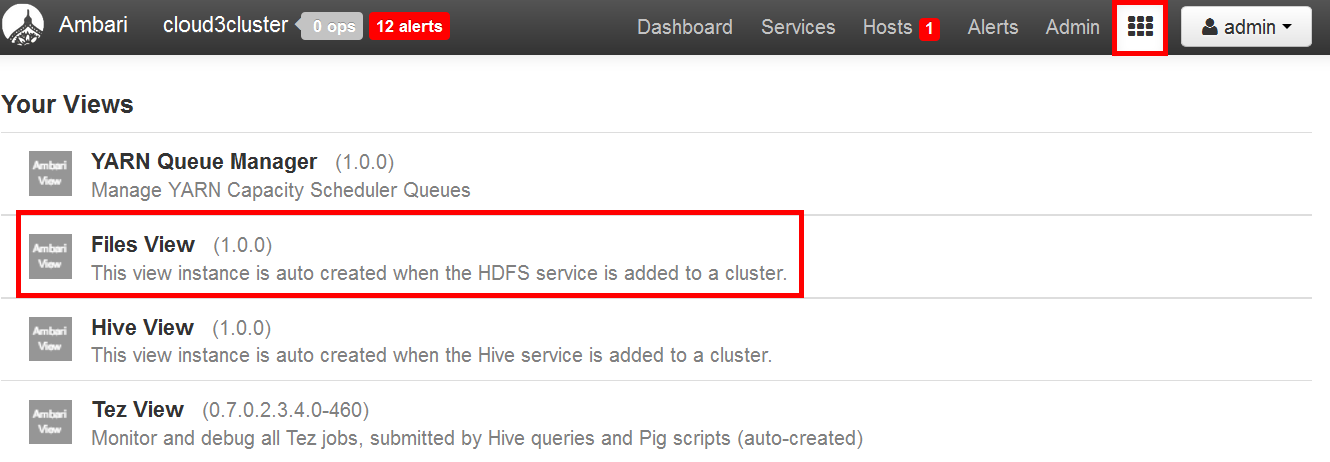
Зайдите на главную страницу Ambari (см. рис 2). На панели, расположенной в левой части экрана, выберите пункт **HDFS**. На появившейся странице перейдите на вкладку «**Configs**», далее выберите вкладку «**Advanced**» (расположена ниже). Промотайте страницу вниз и найдите список «**Advanced core-site**».

Рисунок 12. Значение hdfs.defaultFS

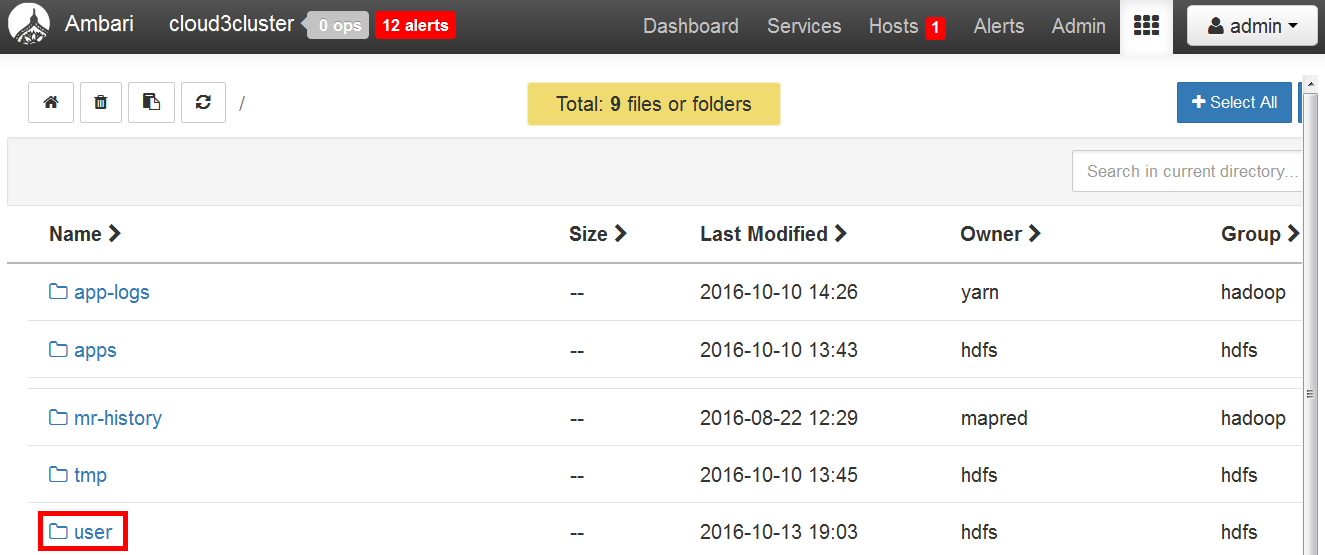
В поле «**fs.defaultFS**» будет указано искомое значение.

## Приложение 3. Проверка наличия корневого каталога HDFS

Для того чтобы проверить наличие корневого каталога HDFS, зайдите на главную страницу Ambari (см. рис 2), нажмите кнопку , расположенную в правом верхнем углу страницы. На странице кликните по пункту «**Files View**».

Рисунок 13. Переход на страницу просмотра файлов

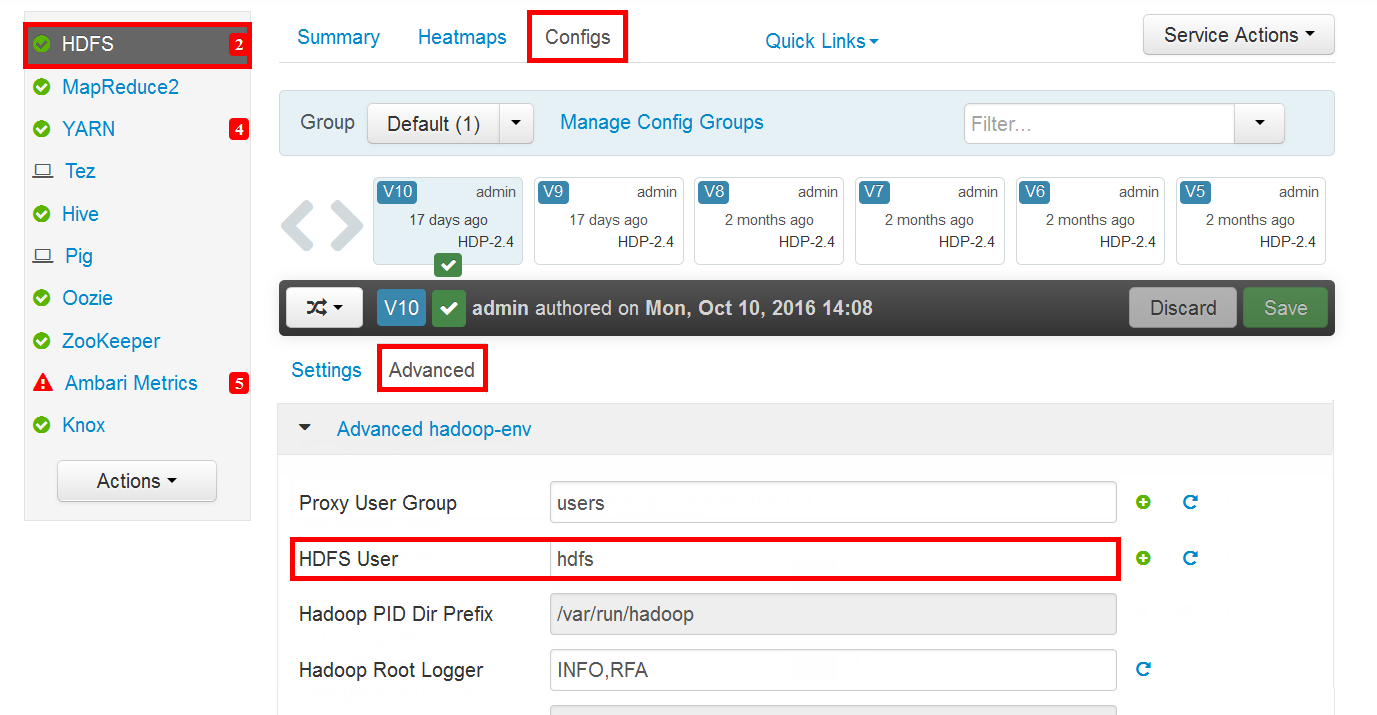
На экране появится страница просмотра файлов.

Рисунок 14. Страница просмотра файлов

В списке должен присутствовать каталог «**User**».

## Приложение 4. Как узнать значение hdfs\_default\_user?

Чтобы узнать значение hdfs\_default\_user, откройте главную страницу Ambari (см. рис 2). На панели, расположенной в левой части экрана, выберите пункт **HDFS**. На появившейся странице перейдите на вкладку «**Configs**», далее выберите вкладку «**Advanced**» (расположена ниже). Промотайте страницу вниз и найдите список «**Advanced hadoop-env**».

Рисунок 15. Значение hdfs.defaultFS

В поле «**HDFS User**» будет указано искомое значение.

## Приложение 5. Команды Docker

* **docker load < filename.tar** – загрузка контейнера.
* Если загрузка контейнера уже была выполнена, то для удаления контейнера необходимо выполнить команды:

**docker stop <id>** - остановка работы контейнера;

**docker rm <id>** - удаление контейнера.

* Данные о состоянии контейнеров можно получить, выполнив одну из команд:

**docker ps** – выводит список стартовавших контейнеров;

**docker ps -a** – выводит список всех контейнеров.

* **docker logs -f <id>** - просмотр логов внутри контейнера.
* **docker exec -it <id> /bin/bash** - вход в терминал контейнера.
* **doker start <id>** - запуск работы контейнера.

## Приложение 6. Пример создания базы данных и пользователя

***CREATE DATABASE [db name]*** – создание базы данных;

***CREATE USER [user name] WITH PASSWORD '[password]'*** – создание пользователя;

***GRANT ALL privileges ON DATABASE [db name] TO [user name]*** – назначение максимальных прав пользователю для работы с базой данных;

Для выхода используйте команду:**\q**.

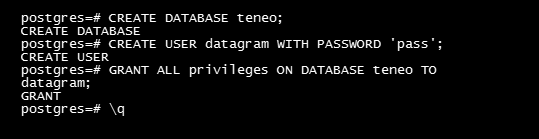


Рисунок 16. Пример успешного выполнения запросов PostgreSQL