

Подготовка к установке

Проверка установленных компонентов

Настройка Spark

Установка программы

Настройка Livy server

Подготовка к установке

Перед началом установки Neoflex Datagram установите следующее ПО:

1. Arenadata Hadoop 1.6.1 CentOS/RHEL 7 (рег. номер ПО: 4074, сайт разработчика: <https://arenadata.tech>) с компонентами:
 - o HDFS;
 - o Oozie;
 - o Livy;
 - o YARN;
 - o Hive;
 - o ZooKeeper.

Скачать дистрибутив Arenadata Hadoop можно по ссылке: <https://store.arenadata.io/#products/arenadata-hadoop>

Примечание.

При установке Arenadata Hadoop необходимо сохранить (записать) путь к каталогу, в котором находится конфигурация Hadoop и путь к каталогу пользователя hdfs (на компьютере, а не в Hadoop).

2. Postgres Pro (рег. номер ПО: 104, сайт разработчика: <https://postgrespro.ru>) или PostgreSQL (open source, сайт разработчика: <https://www.postgresql.org>). В базе данных должны быть созданы:
 - o База данных teneo (название базы может быть произвольным);
 - o Пользователь, от имени которого Neoflex Datagram будет работать с базой данных. Пользователю должны быть заданы максимальные права доступа.

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

```
postgres=# CREATE DATABASE teneo;  
CREATE DATABASE  
postgres=# CREATE USER datagram WITH PASSWORD 'pass';  
CREATE USER  
postgres=# GRANT ALL privileges ON DATABASE teneo TO  
datagram;  
GRANT  
postgres=# \q
```

3. Maven (open source, сайт разработчика <https://mvnrepository.com>).

Команда запуска установки для CentOS:

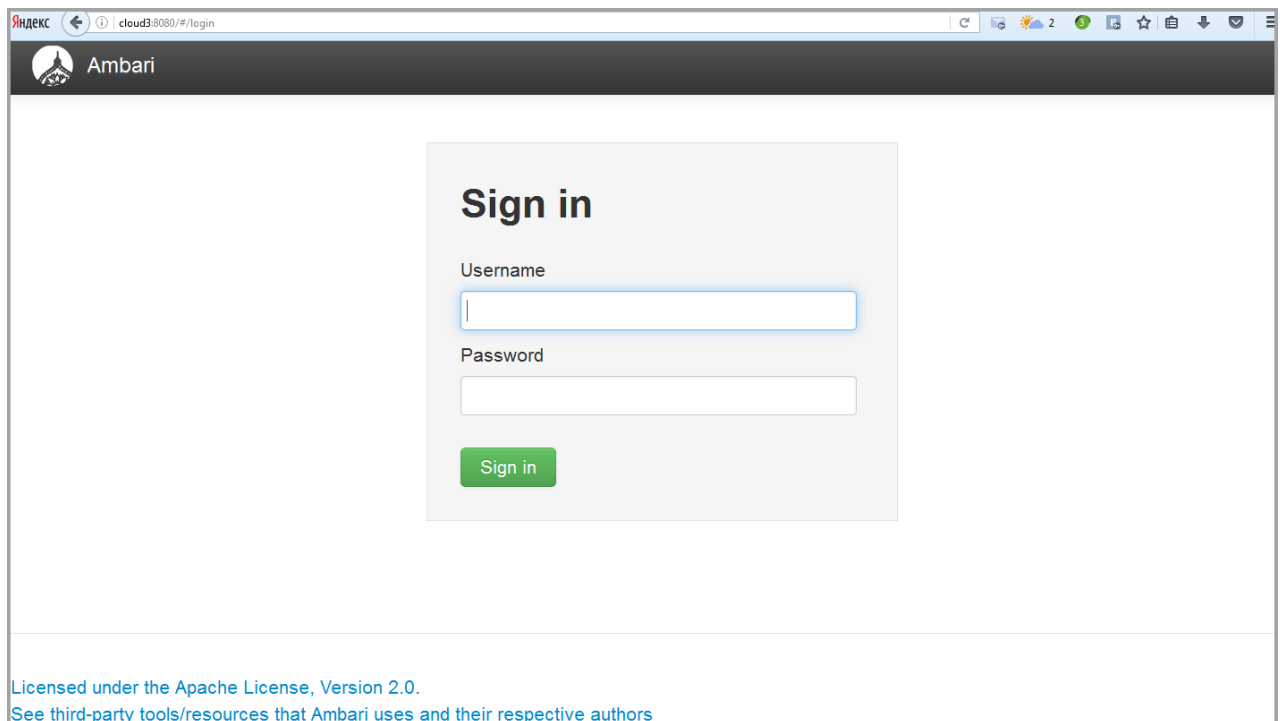
yum install maven.

Проверка установленных компонентов

1. Проверьте, что в системе доступен сервер **Ambari**, входящий в комплект поставки Arenadata Hadoop. Для этого в адресной строке браузера введите:

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

На экране должна появиться форма регистрации пользователя.

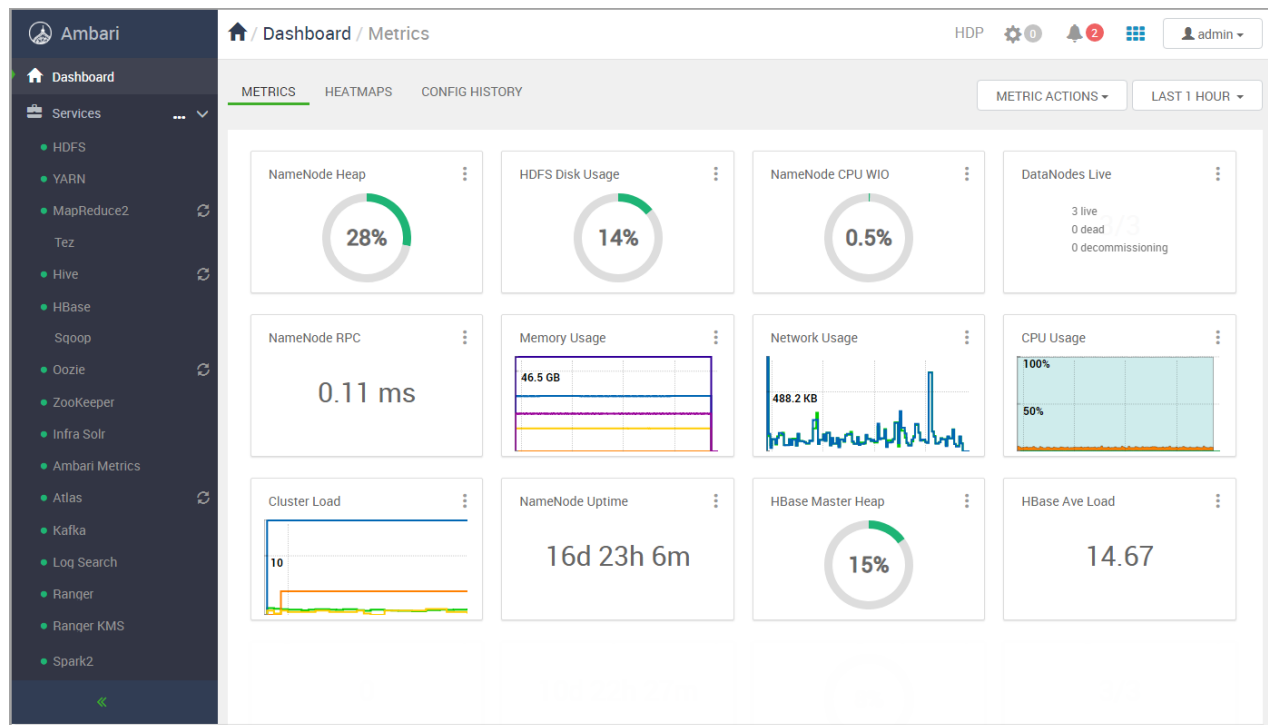



Выполните вход в систему. По умолчанию для входа используются значения:

Username – admin;

Password – admin.

В окне браузера отобразится главная страница сервера Ambari.

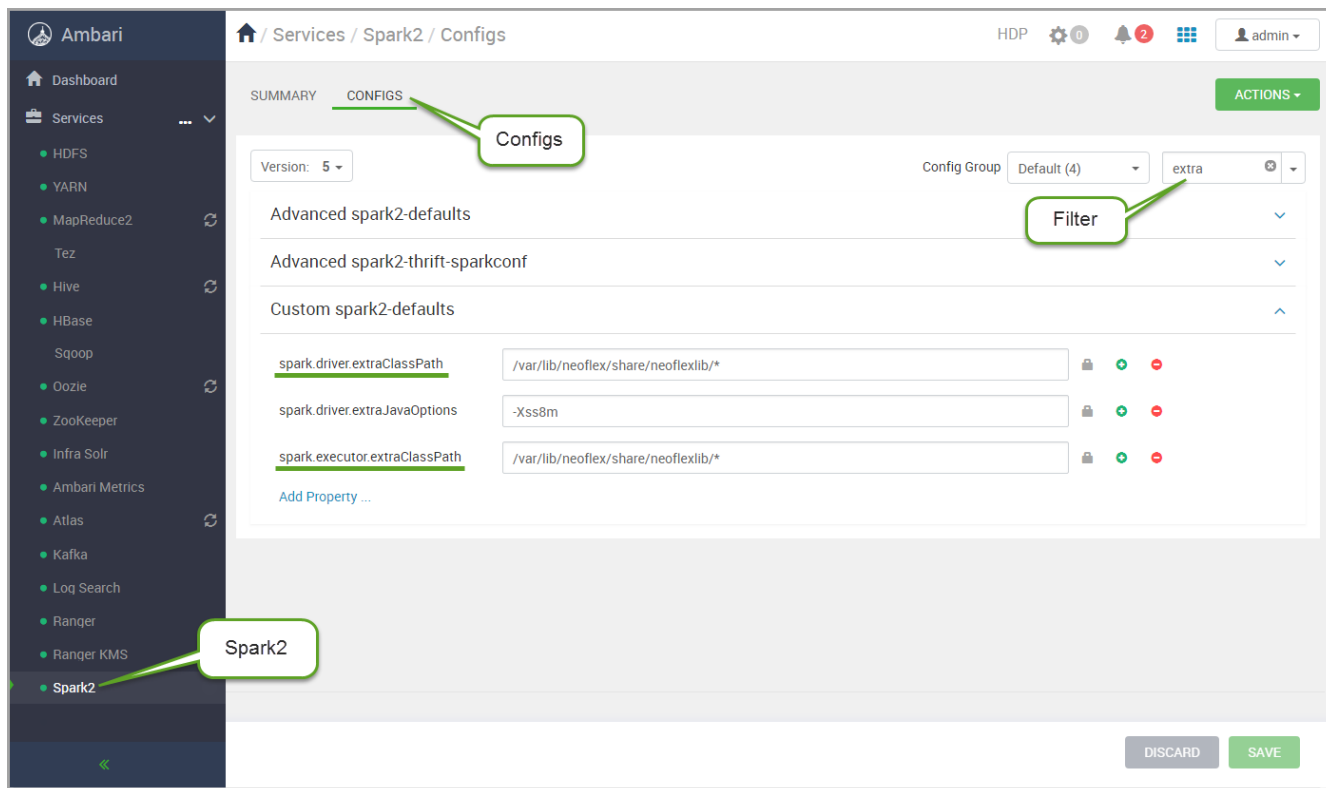


Убедитесь, что на панели, расположенной в левой части экрана, рядом с названиями компонентов Hadoop установлены значки .

2. Проверьте, что установлен Maven, для этого введите команду `mvn -version` в командной строке операционной системы.

Настройка Spark

Откройте главную страницу Ambari и на панели со списком компонентов выберите пункт Spark. На появившейся странице выберите CONFIGS и в фильтре введите значение extra.



1. В поле **spark.driver.extraClassPath** установите значение: /var/lib/neoflex/share/neoflexlib/*
2. В поле **spark.executor.extraClassPath** установите значение: /var/lib/neoflex/share/neoflexlib/*
3. Сохраните изменения кнопкой "SAVE".

Установка программы

На компьютере, где будет производиться установка, должен быть доступ к сети Интернет.

Представителями компании Neoflex поставляется каталог datagram с комплектом каталогов и файлов:

- mserver-version number-SNAPSHOT.jar;
- ldap.properties;
- neoflexlib.dir.

Для установки программы выполните действия:

1. Скопируйте каталог datagram на компьютер, где будет развернута Neoflex Datagram.
2. Отредактируйте файл ldap.properties:

Параметр	Обязательно заполнять	Описание
ldap.domain	Да	Доменное имя Ldap сервера. <i>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</i> <i>Пример: ldap.domain=ldapServer.ru</i>
ldap.host	Да	Имя хоста Ldap сервера <i>Пример: ldap.host=msk-ldserv1.company.ru</i>
ldap.port	Да	Порт Ldap сервера <i>Пример: ldap.port=789</i>
ldap.base	Да	Путь к каталогу для поиска пользователей <i>Пример: ldap.base=CN=Users,DC=company,DC=ru</i>
ldap.admin	Да	Имя группы пользователей, которым будут предоставлены права администратора
ldap.operator	Да	Имя группы пользователей, которым будут предоставлены права оператора
ldap.user	Да	Имя группы пользователей, которым будут предоставлены права пользователя

- Создайте каталог /var/lib/neoflex/share/neoflexlib/ и скопируйте в него файлы из каталога neoflexlib.dir, входящего в комплект поставки.
- Запустите файл mserver-version number-SNAPSHOT.jar при помощи стандартной команды запуска JAR-файлов Linux: **java -Dparameter=value ... -jar \${JAR_NAME}**, с указанием параметров. Каталог, указываемый в параметре -Dmspace.dir (см. таблицу "Параметры запуска") должен быть создан заранее.

Пример:

```
java -Xms2g -Xmx6g -Dfile.encoding=UTF-8 -Dmaven.home=/usr/share/maven -
Dmspace.dir=/opt/datagram/mspace -Dteneo.url=jdbc:postgresql://cloud:1111/teneodev -
Dteneo.user=postgres -Dteneo.password=pass -Dcust.code=dev.cloud -Dserver.port=8080 -jar
/root/Setup/mserver-2.0-SNAPSHOT.jar
```

Параметры запуска

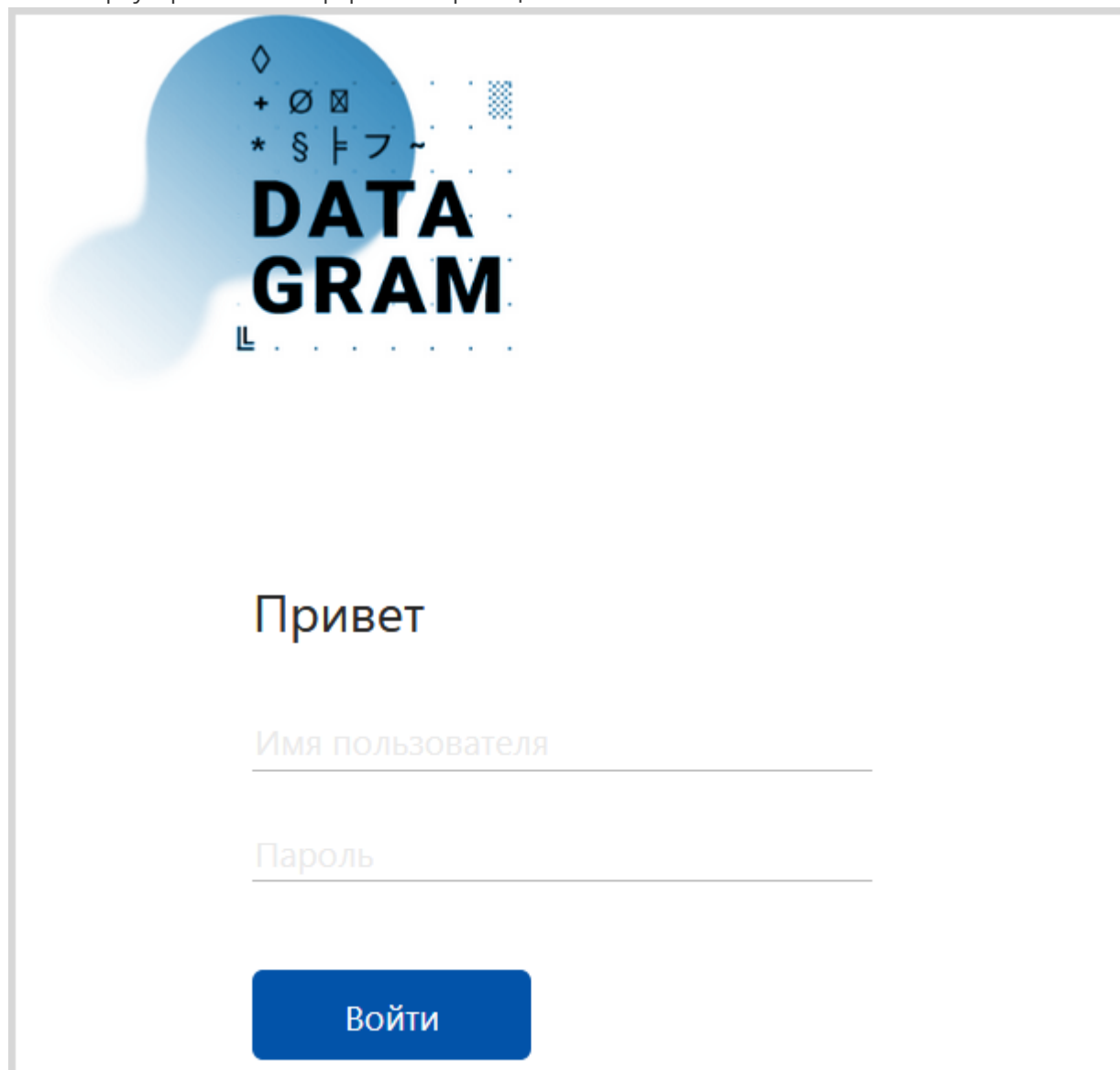
Параметр	Обязательно заполнять	Описание
-Xms2g	Да	Минимальный объем ОЗУ
-Xmx6g	Да	Максимальный объем ОЗУ
-Dfile.encoding	Да	Всегда используется кодировка UTF-8
-Dmaven.home	Да	Путь к инсталляции Maven (https://maven.apache.org/)
-Dmspace.dir	Да	Путь к каталогу программы <i>Примечание. Каталог должен быть создан до запуска файла .jar</i>
-Ddeploy.dir	Нет	Путь к каталогу, в котором хранятся ресурсы слоя сопровождения. Если параметр не задан, то каталог формируется по умолчанию: \${mspace.dir}/deployments/{cust.code}
-Dteneo.url	Да	Url-адрес для подключения к БД репозитория метаданных
-Dteneo.user	Да	Имя пользователя для подключения к БД репозитория метаданных
-Dteneo.password	Да	Пароль для подключения к БД репозитория метаданных
-Dcust.code	Нет	Код инсталляции (код клиента). По умолчанию: default
-Dserver.port	Нет	HTTP порт сервера метаданных
-Dldap.config	Нет	Путь к файлу конфигурации ldap
-Dpasswords	Нет	Путь к файлу хранения паролей. По умолчанию: \${user.dir}/passwords.properties
-Dencrypt.passwords	Нет	Опция шифрования паролей. Может принимать два значения: «false» (установлено по умолчанию) и true

5. Запустите браузер и в адресной строке введите:

<http://host:port/cim/ddesigner/build/index.html?>

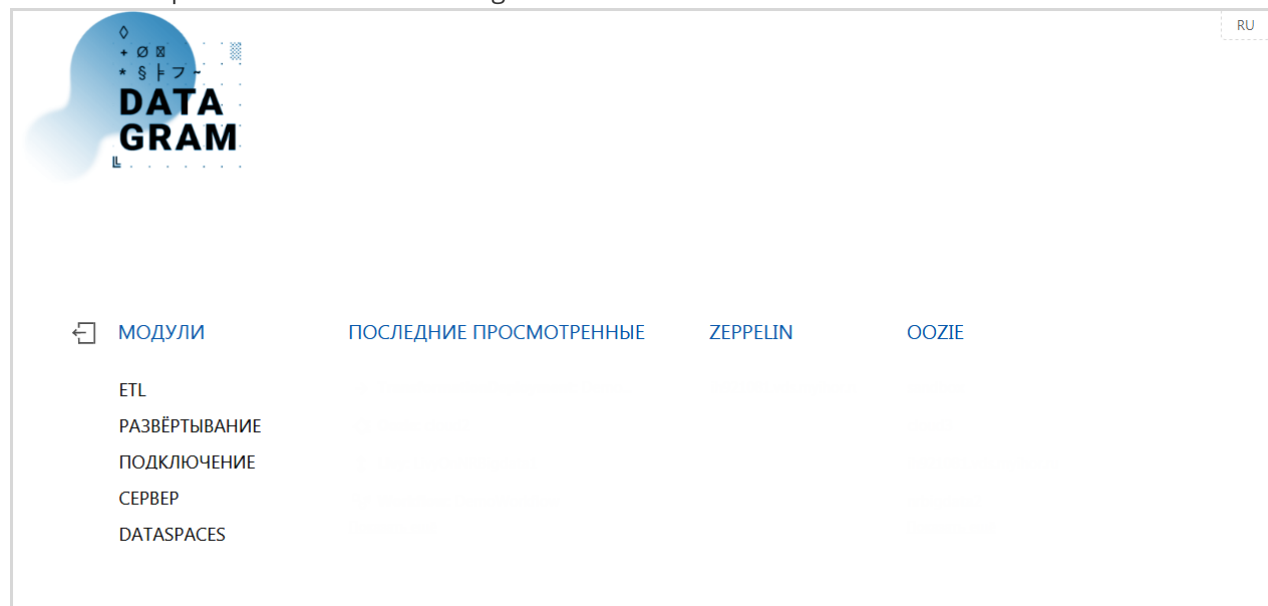
, где **host** - хост сервера, на котором установлена программа, **port** - номер порта сервера.

В окне браузера появится форма авторизации пользователя.



The login form for Neoflex Datagram features a blue abstract logo with the text "DATA GRAM" and a small "L" icon. Below the logo, the word "Привет" (Hello) is displayed. There are two input fields: "Имя пользователя" (Username) and "Пароль" (Password). A blue button labeled "Войти" (Login) is positioned at the bottom.


Для входа в программу укажите имя пользователя, пароль и нажмите кнопку **«Вход»**. На экране появится стартовое окно Neoflex Datagram.

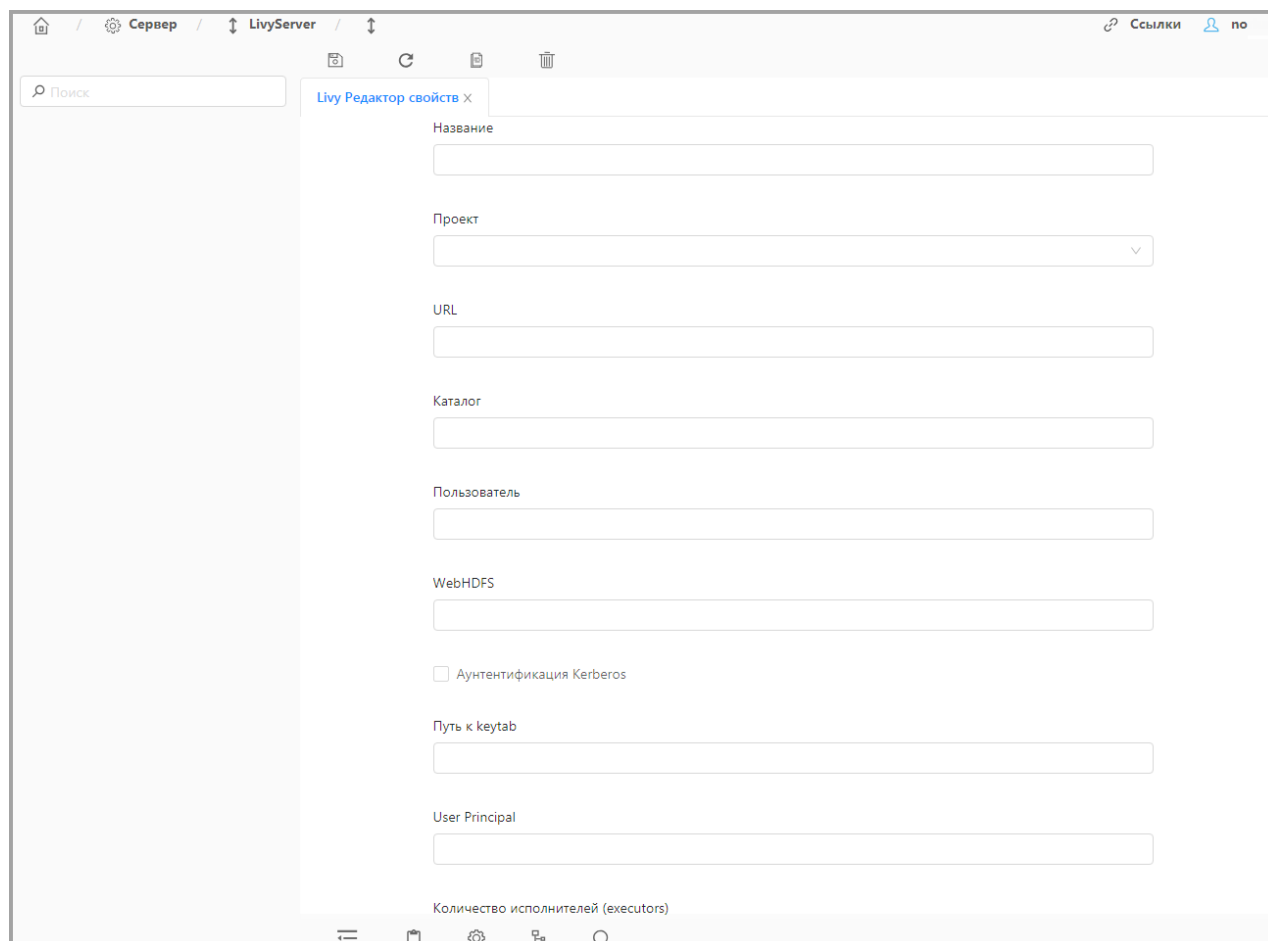


The dashboard interface includes the "DATA GRAM" logo in the top left and a language selector "RU" in the top right. A sidebar on the left contains a menu with the following items: "МОДУЛИ" (Modules), "РАЗВЁРТЫВАНИЕ" (Deployment), "ПОДКЛЮЧЕНИЕ" (Connection), "СЕРВЕР" (Server), and "DATASPACE". The main content area is titled "ПОСЛЕДНИЕ ПРОСМОТРЕННЫЕ" (Last Viewed) and displays a table with columns for "ZEPPELIN" and "OOZIE".

	ZEPPELIN	OOZIE
ETL		
РАЗВЁРТЫВАНИЕ		
ПОДКЛЮЧЕНИЕ		
СЕРВЕР		
DATASPACE		

Настройка Livy server

1. Перейдите в раздел "Сервер/Livy" и по кнопке  откройте форму создания сервера Livy.



Настройка Livy server

Название

Проект

URL

Каталог

Пользователь

WebHDFS

☐ Аутентификация Kerberos

Путь к keytab

User Principal

Количество исполнителей (executors)

2. Заполните поля:

Название - укажите название создаваемого объекта Livy Server (например: NewLivy). Названия объектов в программе должно удовлетворять правилам формирования идентификаторов в языке Java.

URL - Url-адрес Livy Server AP (пример: <http://cloud.company.ru:8989>).


Каталог - каталог, используемый для развертывания "Transformation" (пример: /user).

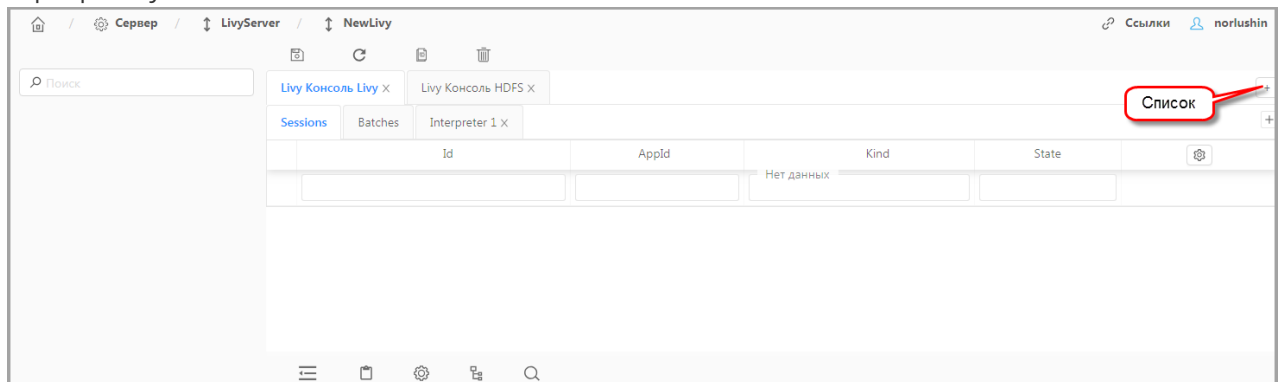
Пользователь - пользователь HDFS, от имени которого разворачиваются "Transformation" (пример: hdfs).


WebHDFS - Url-адрес HDFS API (пример: <http://cloud3.company.ru:50070/webhdfs/v1>).

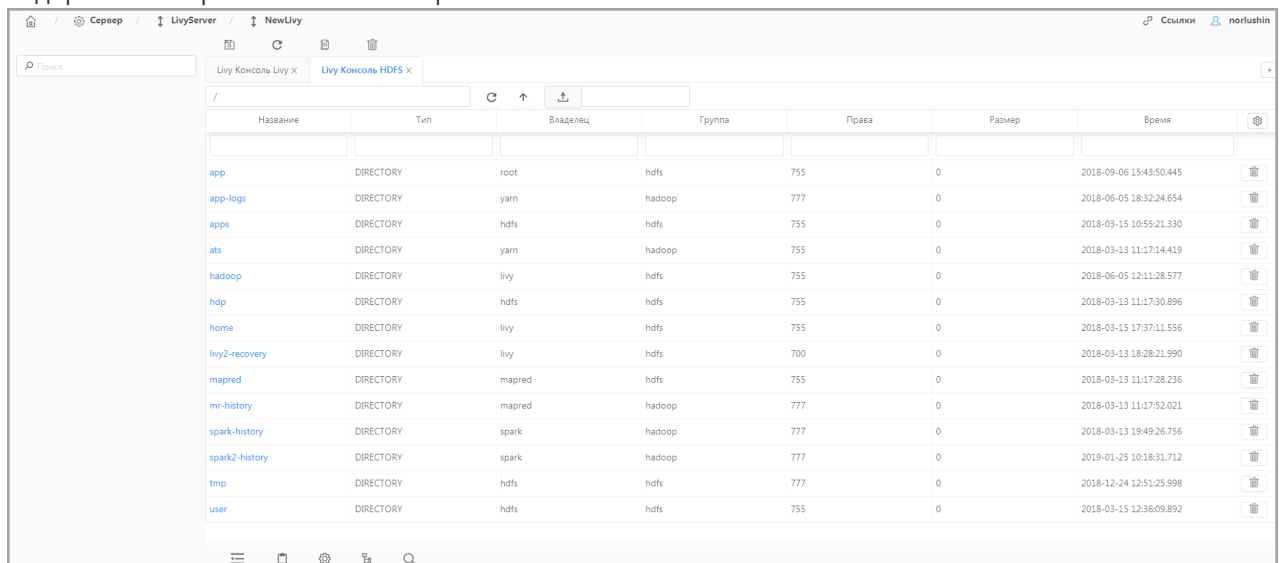
Количество исполнителей (executors) - количество ядер, задействованных для реализации исполняющего процесса Spark (пример: 1).














Использовать по умолчанию - включите чекбокс. Для остальных настроек оставьте значения по умолчанию.

3. Сохраните настройки кнопкой . После сохранения настроек на экране отобразится консоль сервера Livy.



4. Убедитесь, что установлено соединение с HDFS. Для этого откройте вкладку "Livy консоль HDFS" по кнопке  (см. рисунок выше). Если соединение установлено, то на вкладке отобразится содержимое корневого каталога файловой системы.



Название	Тип	Владелец	Группа	Права	Размер	Время	
app	DIRECTORY	root	hdfs	755	0	2018-09-06 15:43:50.445	
app-logs	DIRECTORY	yarn	hadoop	777	0	2018-06-05 18:32:24.654	
apps	DIRECTORY	hdfs	hdfs	755	0	2018-03-15 10:55:21.330	
ats	DIRECTORY	yarn	hadoop	755	0	2018-03-13 11:17:14.419	
hadoop	DIRECTORY	livy	hdfs	755	0	2018-06-05 12:11:28.577	
hdp	DIRECTORY	hdfs	hdfs	755	0	2018-03-13 11:17:30.896	
home	DIRECTORY	livy	hdfs	755	0	2018-03-15 17:37:11.556	
livy2-recovery	DIRECTORY	livy	hdfs	700	0	2018-03-13 18:28:21.990	
mapred	DIRECTORY	mapred	hdfs	755	0	2018-03-13 11:17:28.236	
mr-history	DIRECTORY	mapred	hadoop	777	0	2018-03-13 11:17:52.021	
spark-history	DIRECTORY	spark	hadoop	777	0	2018-03-13 19:49:26.756	
spark2-history	DIRECTORY	spark	hadoop	777	0	2019-01-25 10:18:31.712	
tmp	DIRECTORY	hdfs	hdfs	777	0	2018-12-24 12:51:25.998	
user	DIRECTORY	hdfs	hdfs	755	0	2018-03-15 12:36:09.892	