



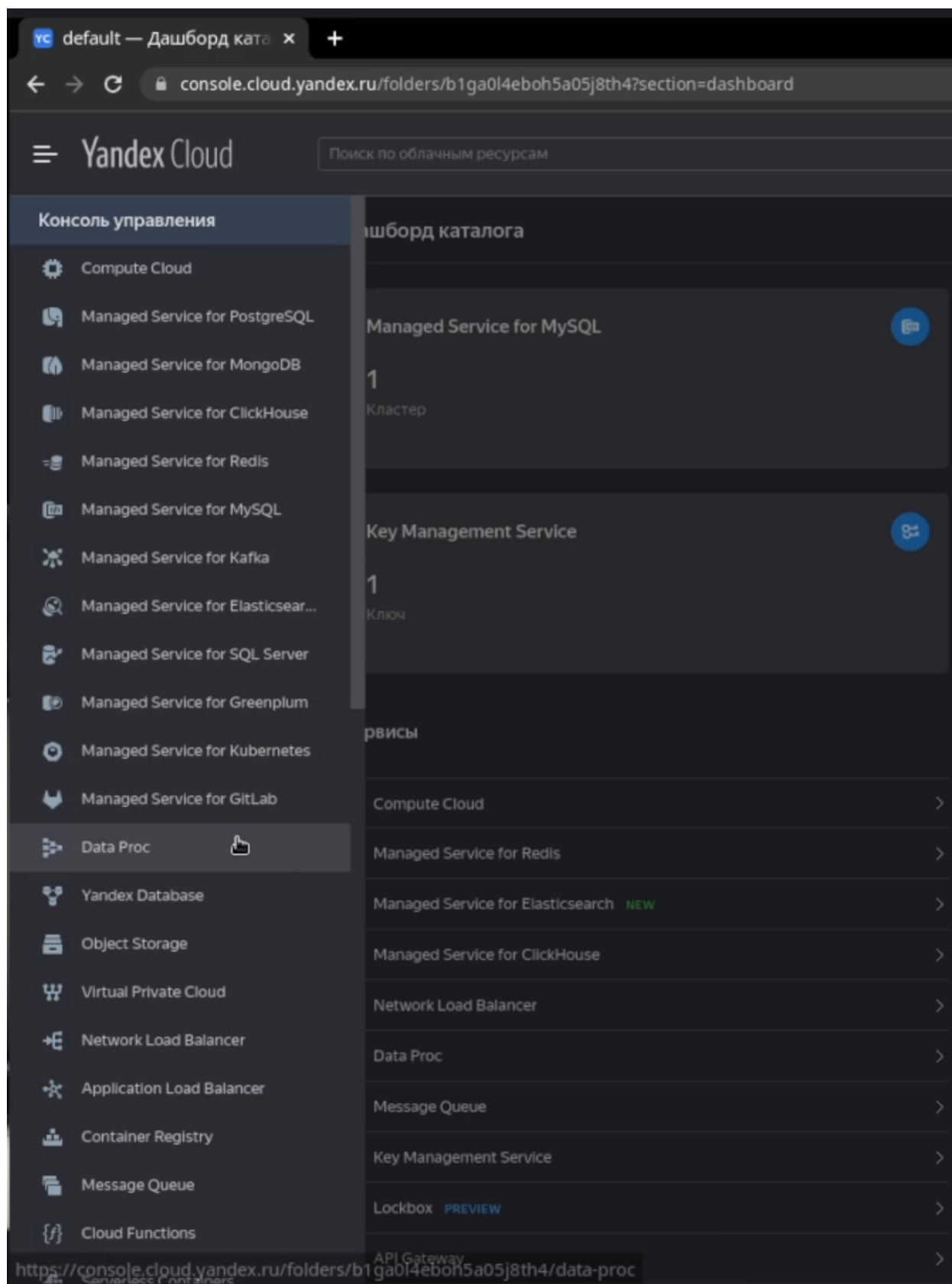
> Конспект > Практика 0 > Как создать кластер

- > [Создание кластера Hadoop в яндекс облаке](#)
- > [Изменение характеристик кластера](#)

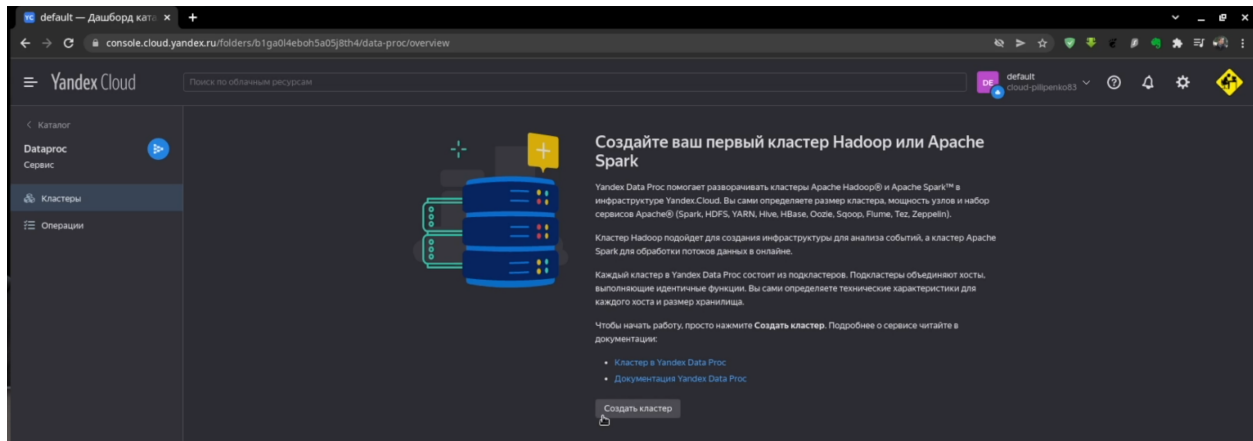
> Создание кластера Hadoop в яндекс облаке

Для создания кластера Hadoop выполним следующие шаги:

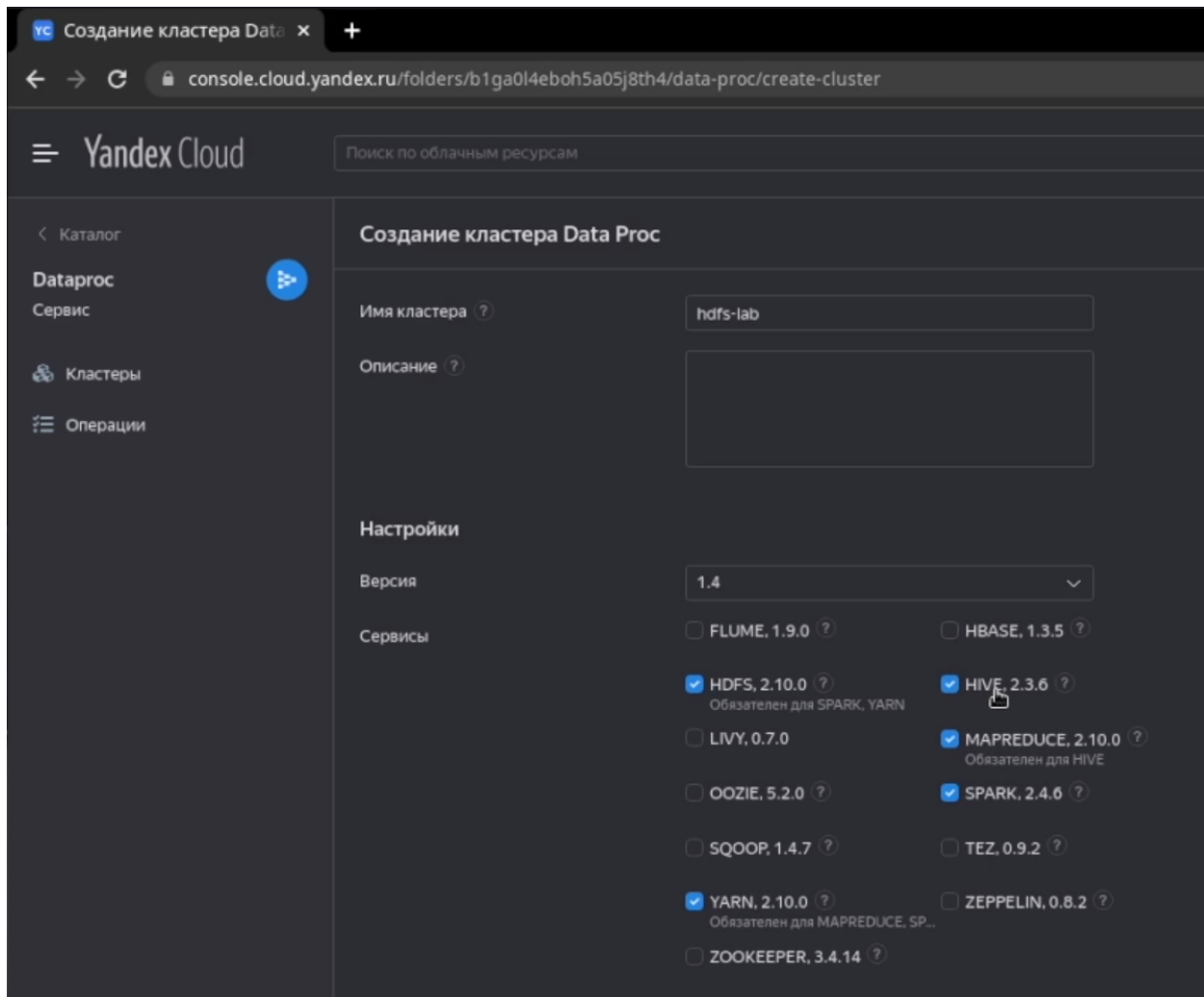
В консоле управления перейдем на **Data Proc.**



Кликаем на кнопку **Создать кластер**.

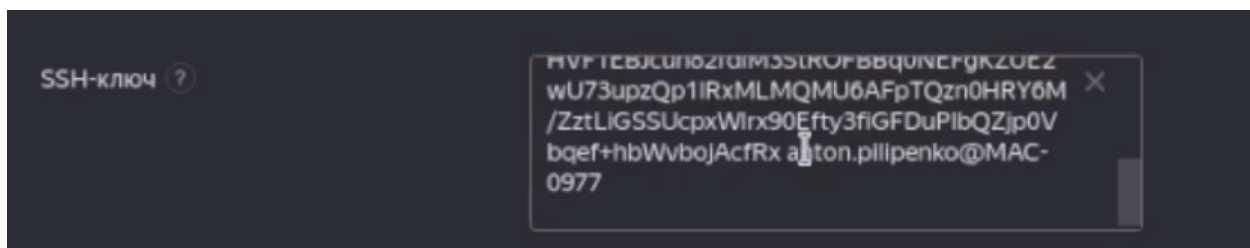


В окне создания кластера указываем его имя, описание (по желанию), версию (выберем 1.4, это соответствует HDFS 2 версии), выбираем нужные компоненты (HDFS, YARN, MAPREDUCE, SPARK, HIVE).



Далее в этом же окне нужно ввести **ssh-ключ**. Это некоторая последовательность байт, которая сохранена локально. Ключ состоит из 2 частей - приватной и публичной. Здесь требуется указать публичную (открытую) часть. Указание ключа требуется, чтобы в дальнейшем мы могли подключаться к нашему кластеру по протоколу. Скопировать публичную часть ключа в буфер обмена можно командой:

```
ssh. cat ~/.ssh/id_rsa.pub | pbcop
```



Выбираем значение сервисного аккаунта, либо создаем новый. Указываем зону доступности (в нашем случае выбор неважен). Выбираем сеть, если ее нет, то нужно создать новую. Нажимаем на галочку напротив **UI Proxy**, это нужно, чтобы у нас была веб консоль для HDFS и YARN.

The screenshot shows the configuration page for a new Databricks cluster. The interface is in Russian. On the left, there is a list of configuration items: 'Сервисный аккаунт' (Service account), 'Зона доступности' (Availability zone), 'Свойства' (Properties), 'Формат указания бакета' (Bucket format), 'Имя бакета' (Bucket name), 'Формат указания сети' (Network format), 'Сеть' (Network), 'Группы безопасности' (Security groups), 'UI Proxy', and 'Лог-группа' (Log group). On the right, the corresponding values are set: 'hadoop-service-a...' for the service account, 'ru-central1-b' for the availability zone, 'hdfs-cool' for the bucket name, 'hadoop-net' for the network, and 'лог-группа по умолчанию' (default log group) for the log group. The 'UI Proxy' checkbox is checked. A message indicates that only accounts with the 'mdb.dataproc.agent' role can be used. A 'Создать новую' (Create new) button is available for the service account. A 'Создать группу' (Create group) button is available for security groups, with a note that no groups are currently available.

Сервисный аккаунт	hadoop-service-a... или Создать новый
<i>Можно использовать только аккаунты со следующими ролями: mdb.dataproc.agent</i>	
Зона доступности	ru-central1-b
Свойства	Ключ : Значение Добавить свойство
Формат указания бакета	Список ID
Имя бакета	hdfs-cool
Формат указания сети	Список ID
Сеть	hadoop-net
Группы безопасности	Создать группу Нет доступных групп
UI Proxy	<input checked="" type="checkbox"/>
Лог-группа	лог-группа по умолчанию

В разделе подкластеры мы можем отредактировать параметры NameNode (здесь она называется мастером) и DataNode (здесь Data). Для редактирования нужно нажать на **...**, тогда откроются параметры, чьи значения можно редактировать. Изменим параметр Мастера, выберем машину **s2.small**.

Имя подкластера ?

Хосты

Класс хоста

Платформа

standard memory-optimized burstable

s2.micro 2 cores vCPU 8 ГБ Память	s2.small 4 cores vCPU 16 ГБ Память
s2.medium 8 cores vCPU 32 ГБ Память	s2.large 12 cores vCPU 48 ГБ Память
s2.xlarge 16 cores vCPU 64 ГБ Память	s2.2xlarge 24 cores vCPU 96 ГБ Память

В параметрах Data изменим количество хостов с 1 до 3 (т.е. у нас будет 3 датаноды).

Имя подкластера ?

Хосты 1 32

Класс хоста

Платформа

standard memory-optimized burstable

s2.micro 2 cores vCPU 8 ГБ Память	s2.small 4 cores vCPU 16 ГБ Память
s2.medium 8 cores vCPU 32 ГБ Память	s2.large 12 cores vCPU 48 ГБ Память

После выполнения всех пунктов нажимает на кнопку **Создать кластер**.

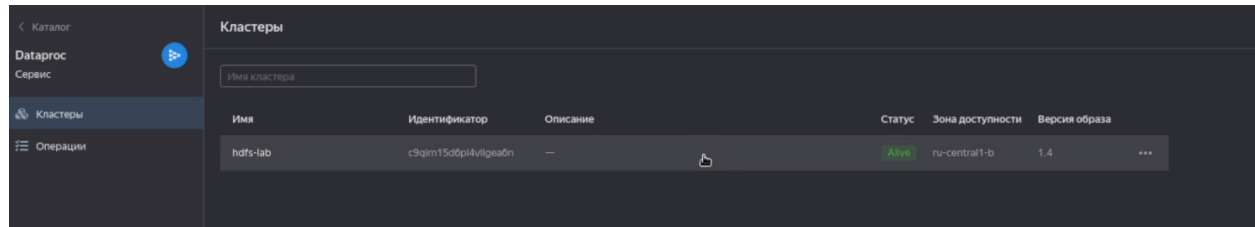
Дополнительные настройки

Защита от удаления ☐

Создать кластер

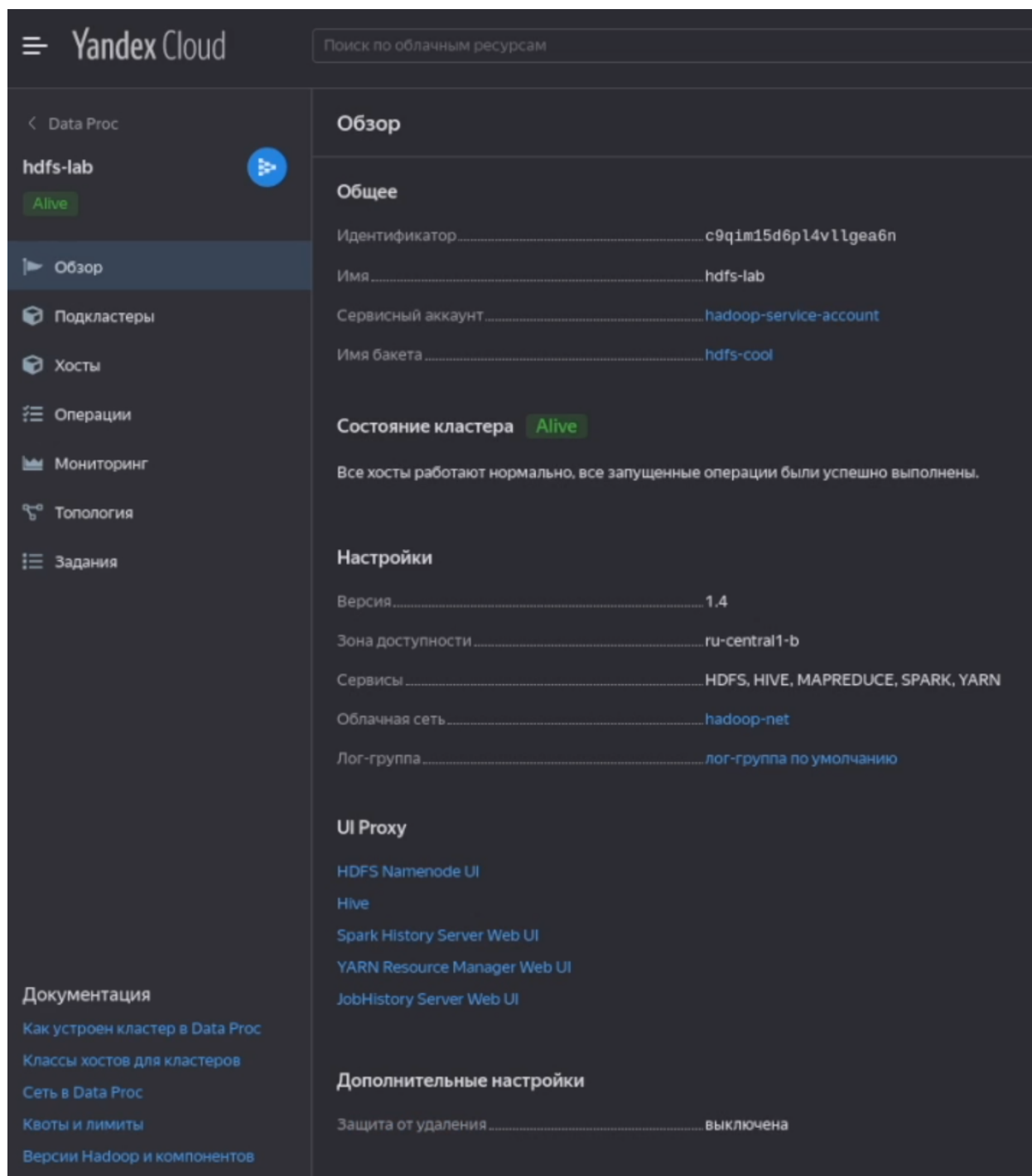
> Изменение характеристик кластера

На вкладке с кластерами мы видим наш созданный кластер в статусе **Alive**.

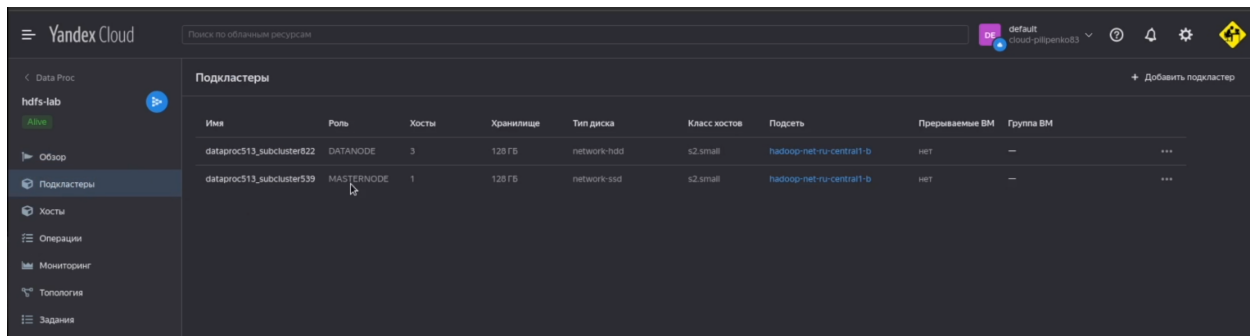


Имя	Идентификатор	Описание	Статус	Зона доступности	Версия образа
hdfs-lab	c9qm15d6p4vl9eabn	—	Alive	ru-central1-b	1.4

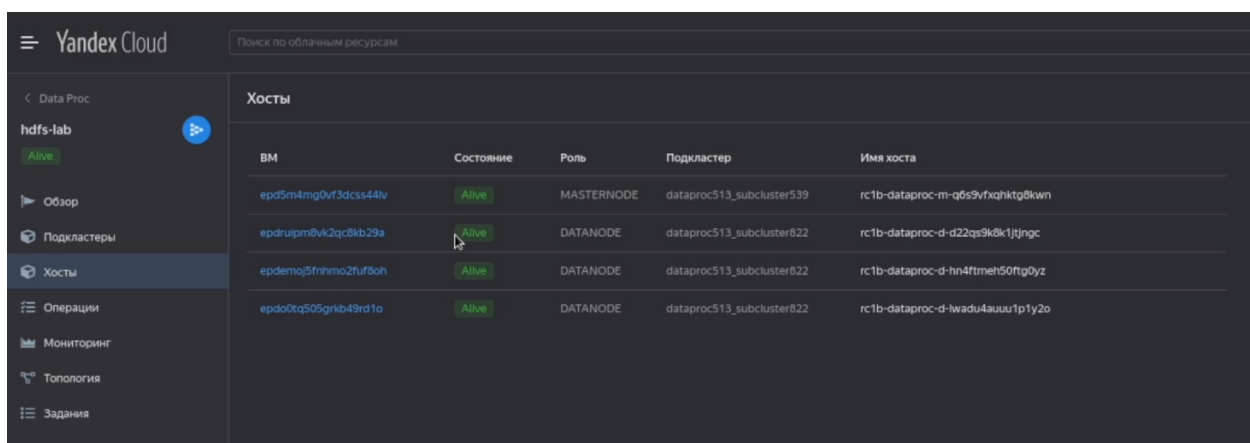
Нажмем на ..., чтобы увидеть его характеристики.



На вкладке **подкластеры** можно посмотреть и изменить характеристики машин выделенных в один покластер.



На вкладке **ХОСТЫ** видим список имеющихся хостов.



Чтобы мы смогли подключаться к хостам Hadoop кластера, добавим публичные IP-адреса для хостов. Для этого нужно выбрать некий хост, в блоке сеть нажимаем на кнопку **Добавить публичный IP-адрес**.

Yandex Cloud

Поиск по облачным ресурсам

Compute Cloud

rc1b-dataproc-m-q6s9vfxqhktg8kwn

Виртуальная машина

Обзор

Диски

Операции

Мониторинг

Серийная консоль

Последовательный порт

Документация

Создать виртуальную машину

Запустить сайт на LAMP или LEMP

Группы виртуальных машин

Доступные платформы

Гарантированная доля vCPU

Обзор

Идентификатор.....epd5m4mg0vf3dcss44lv

Статус.....Running

Имя.....rc1b-dataproc-m-q6s9vfxqhktg8kwn

Дата создания.....15 декабря 2021, в 13:36

Внутренний FQDN.....rc1b-dataproc-m-q6s9vfxqhktg8kwn.mdb.yandexcloud.net

Зона доступности.....ru-central1-b

Сервисный аккаунт.....hadoop-service-account

Ресурсы

Платформа.....Intel Cascade Lake

Гарантированная доля vCPU.....100%

vCPU.....4

RAM.....16 ГБ

Объём дискового пространства.....128 ГБ

Сеть

Сетевой интерфейс

Внутренний IPv4.....10.129.0.34

Подсеть.....hadoop-net-ru-central1-b

Настройки DNS для внутренних адресов

Зона FQDN TTL

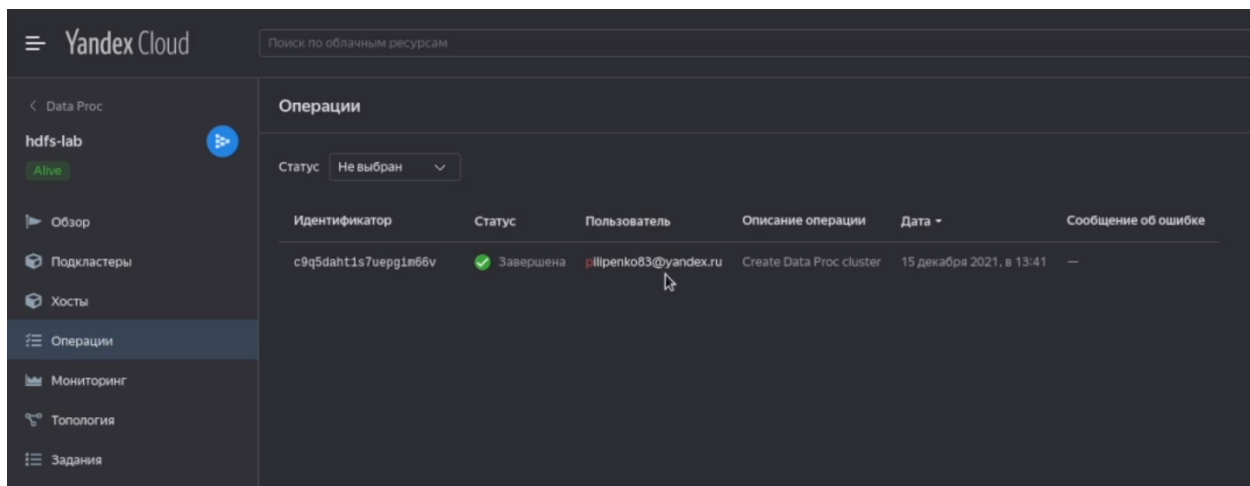
Нет данных

Изменить сетевой интерфейс

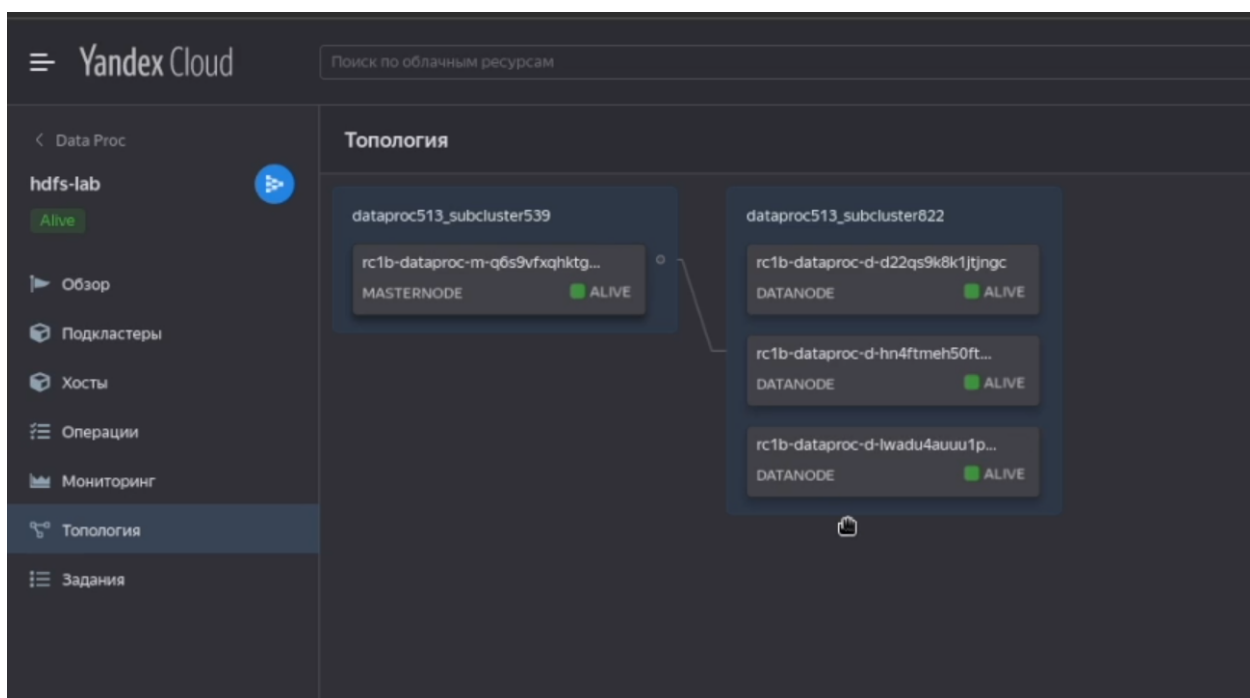
Добавить публичный IP-адрес

Дополнительно

Ставим галочку для параметра **Защита от DDoS-атак** и нажимаем на **добавить**.

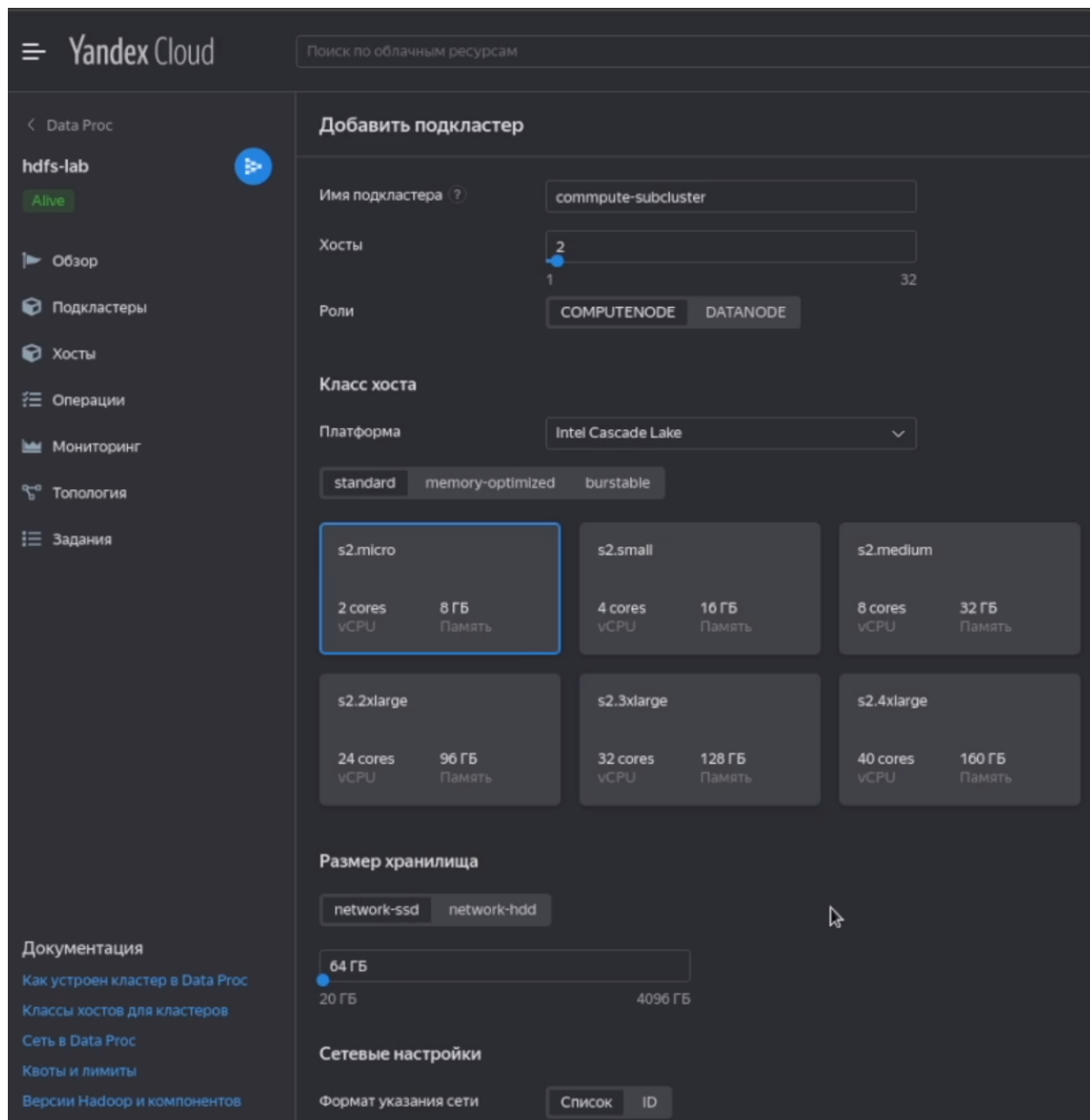


Перейдем в раздел **Топология**. В нем мы можем увидеть как наши машины связаны между собой.



Попробуем изменить наш кластер, а именно добавим к нему новый подкластер. Для этого перейдем на вкладку **Подкластеры**. Нажимаем на кнопку **Добавить подкластер**.

В окне добавления подкластера, указываем его имя, количество хостов, роли.



Нажимаем на кнопку **Добавить подкластер**. После этого наш кластер переходит в статус **Updating**. Статус снова изменится на **Alive**, когда все изменения будут успешно применены.

В разделе с операциями видим, что появилась новая запись:

Yandex Cloud

Поиск по облачным ресурсам

< Data Proc

hdfs-lab

Updating

Обзор

Подкластеры

Хосты

Операции

Мониторинг

Топология

Задания

Операции

Статус: Не выбран

Идентификатор	Статус	Пользователь	Описание операции	Дата	Сообщение об ошибке
c9qcv1281070cqn68t72	Не завершена	ipenko83@yandex.ru	Create Data Proc subcluster	15 декабря 2021, в 13:53	—
c9q5daht1s7uepg1w66v	Завершена	ipenko83@yandex.ru	Create Data Proc cluster	15 декабря 2021, в 13:41	—