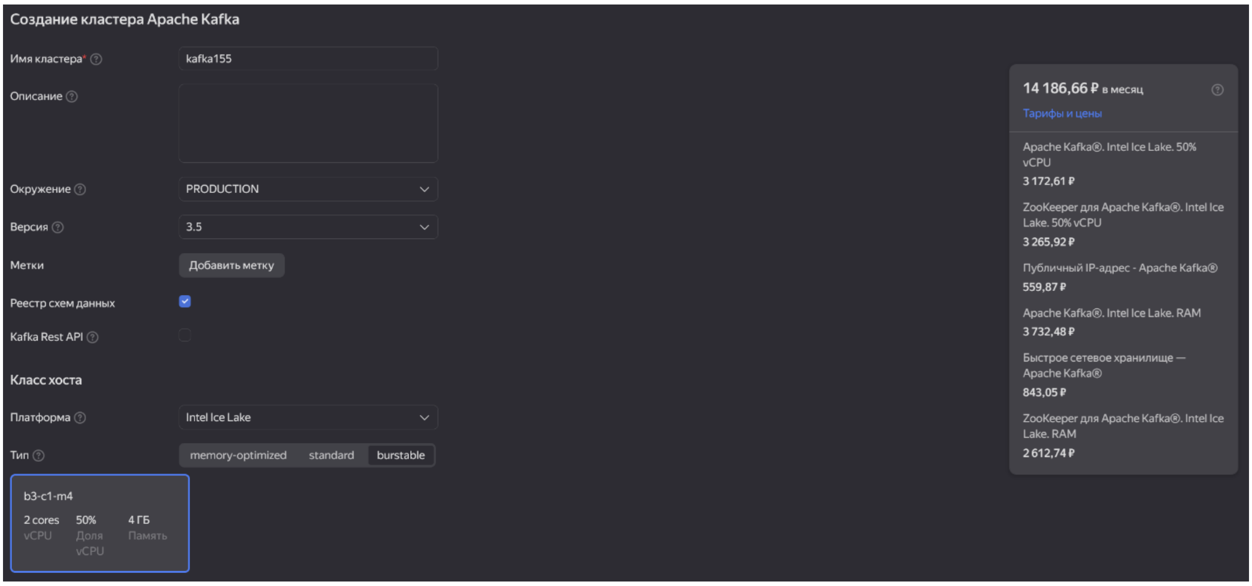
Задание 1. Развёртывание и настройка Kafka-кластера в Yandex Cloud

Создаем кластер Apache Kafka



Создание кластера Apache Kafka

Имя кластера*: kafka155

Описание:

Окружение: PRODUCTION

Версия: 3.5

Метки: Добавить метку

Реестр схем данных ☑

Kafka Rest API □

Класс хоста

Платформа: Intel Ice Lake

Тип: memory-optimized | standard | [✓ burstable]

b3-c1-m4

2 cores vCPU | 50% Доля vCPU | 4 ГБ Память

14 186,66 ₽ в месяц

Тарифы и цены

Apache Kafka®. Intel Ice Lake. 50% vCPU
3 172,61 ₽

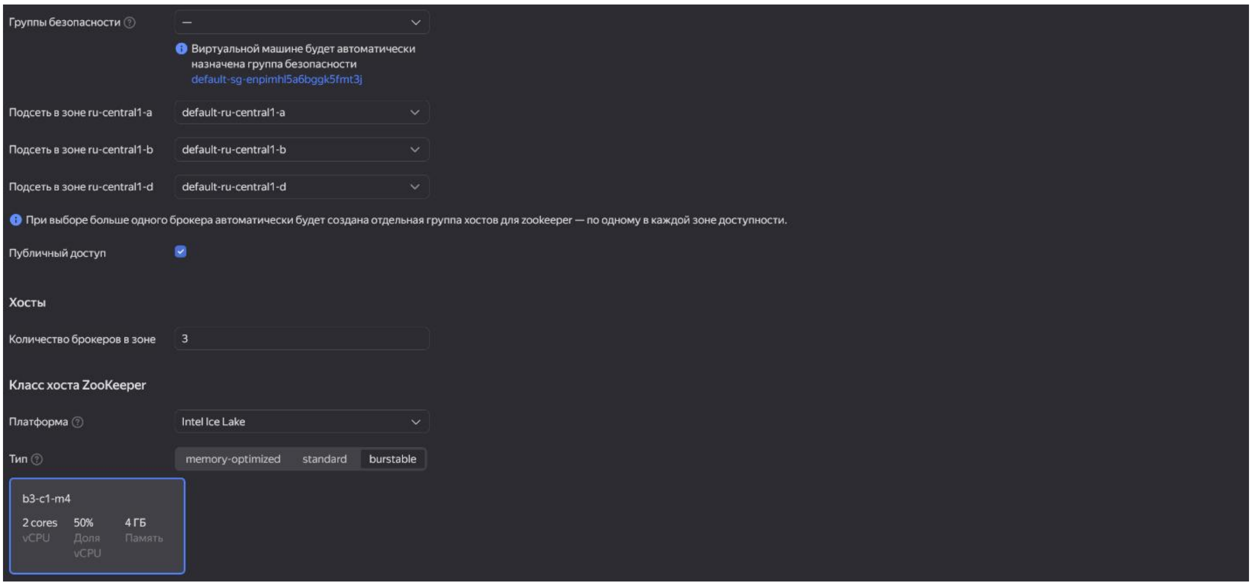
ZooKeeper для Apache Kafka®. Intel Ice Lake. 50% vCPU
3 265,92 ₽

Публичный IP-адрес - Apache Kafka®
559,87 ₽

Apache Kafka®. Intel Ice Lake. RAM
3 732,48 ₽

Быстрое сетевое хранилище — Apache Kafka®
843,05 ₽

ZooKeeper для Apache Kafka®. Intel Ice Lake. RAM
2 612,74 ₽



Группы безопасности: —

Виртуальной машине будет автоматически назначена группа безопасности default-sg-enpimhl5a6bggk5fmt3j

Подсеть в зоне ru-central1-a: default-ru-central1-a

Подсеть в зоне ru-central1-b: default-ru-central1-b

Подсеть в зоне ru-central1-d: default-ru-central1-d

При выборе больше одного брокера автоматически будет создана отдельная группа хостов для zookeeper — по одному в каждой зоне доступности.

Публичный доступ ☑

Хосты

Количество брокеров в зоне: 3

Класс хоста ZooKeeper

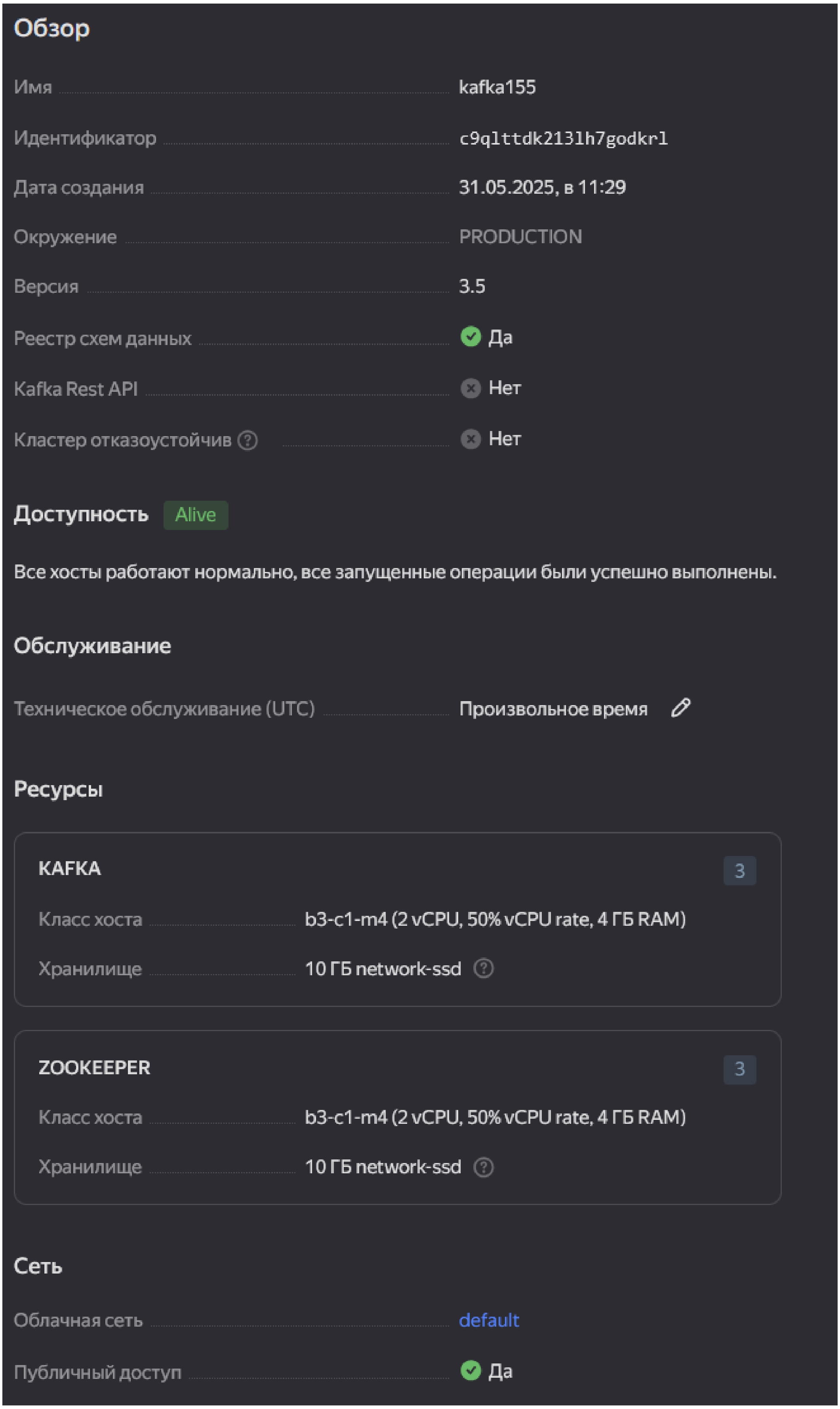
Платформа: Intel Ice Lake

Тип: memory-optimized | standard | [✓ burstable]

b3-c1-m4

2 cores vCPU | 50% Доля vCPU | 4 ГБ Память

Результат



Обзор

Имя	kafka155
Идентификатор	c9qlttdk2131h7godkrl
Дата создания	31.05.2025, в 11:29
Окружение	PRODUCTION
Версия	3.5
Реестр схем данных	✓ Да
Kafka Rest API	✕ Нет
Кластер отказоустойчив ?	✕ Нет

Доступность Alive

Все хосты работают нормально, все запущенные операции были успешно выполнены.

Обслуживание

Техническое обслуживание (UTC)	Произвольное время

Ресурсы

KAFKA 3

Класс хоста	b3-c1-m4 (2 vCPU, 50% vCPU rate, 4 ГБ RAM)
Хранилище	10 ГБ network-ssd ?

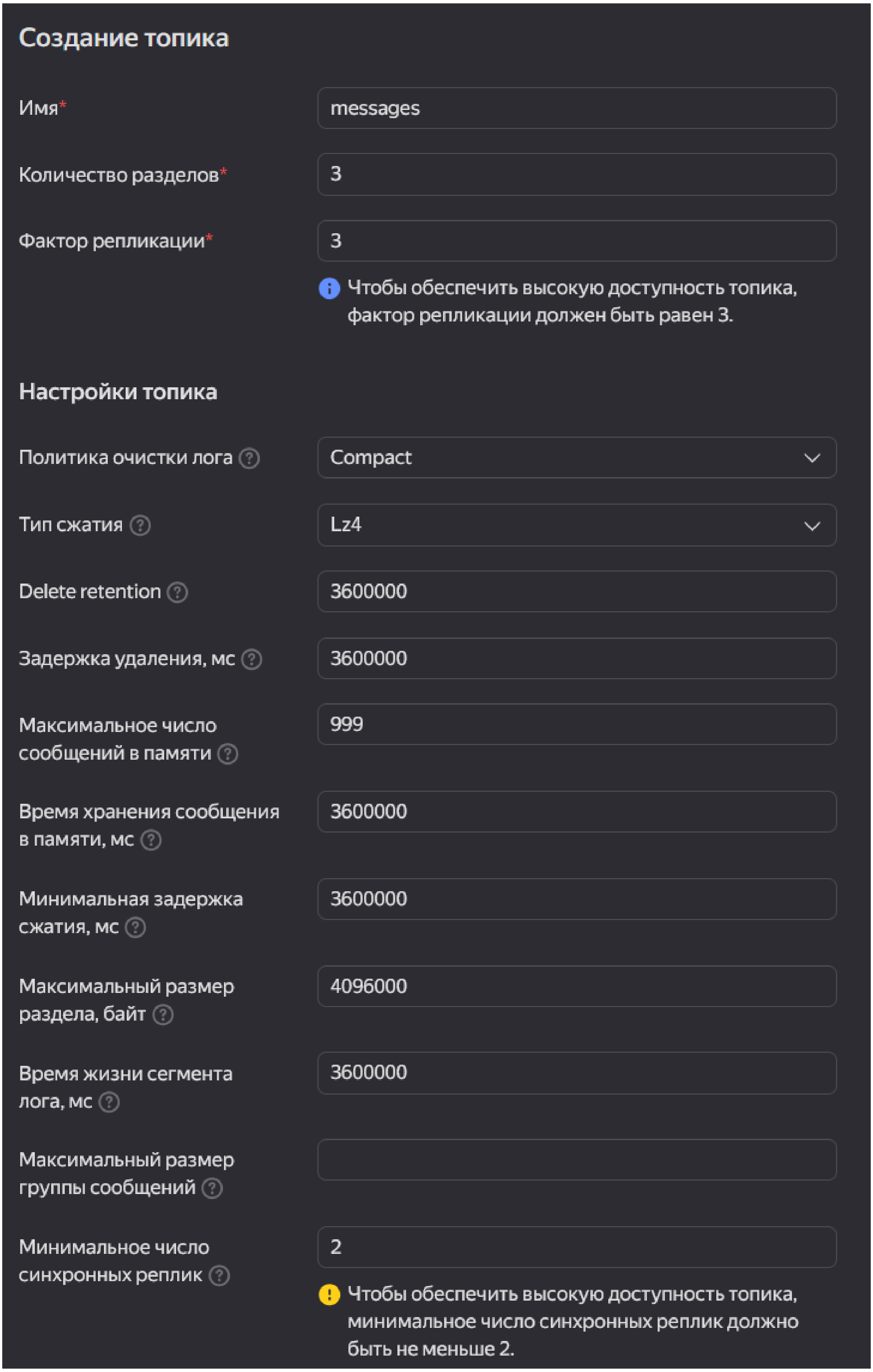
ZOOKEEPER 3

Класс хоста	b3-c1-m4 (2 vCPU, 50% vCPU rate, 4 ГБ RAM)
Хранилище	10 ГБ network-ssd ?

Сеть

Облачная сеть	default
Публичный доступ	✓ Да

Создаем топик



Создание топика

Имя*	messages
Количество разделов*	3
Фактор репликации*	3

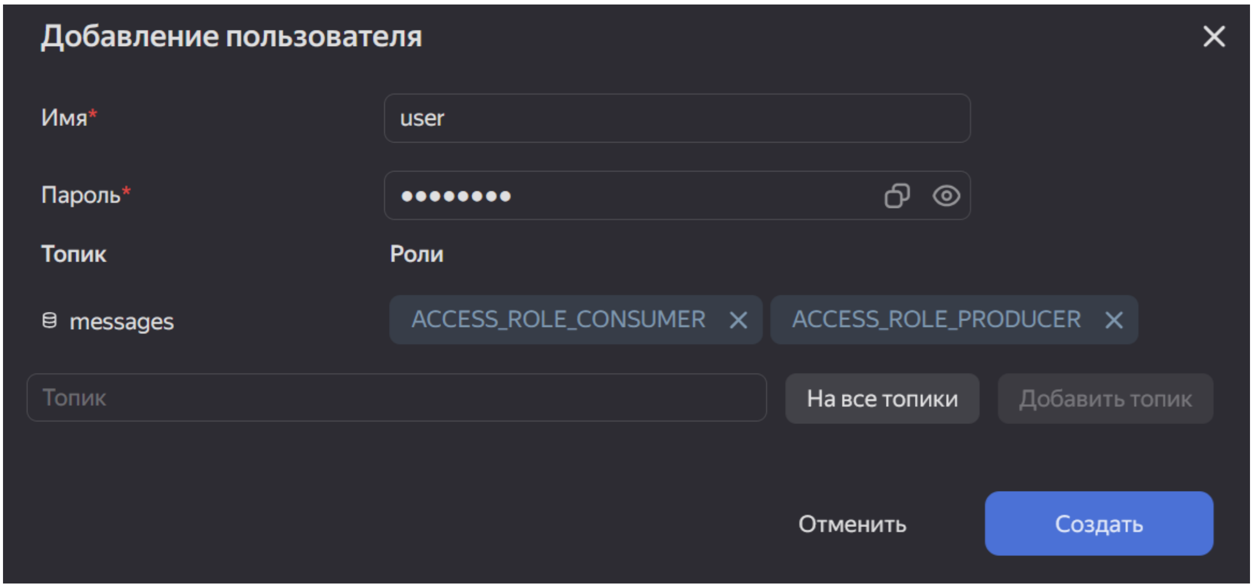
ⓘ Чтобы обеспечить высокую доступность топика, фактор репликации должен быть равен 3.

Настройки топика

Политика очистки лога	Compact
Тип сжатия	Lz4
Delete retention	3600000
Задержка удаления, мс	3600000
Максимальное число сообщений в памяти	999
Время хранения сообщения в памяти, мс	3600000
Минимальная задержка сжатия, мс	3600000
Максимальный размер раздела, байт	4096000
Время жизни сегмента лога, мс	3600000
Максимальный размер группы сообщений	
Минимальное число синхронных реплик	2

! Чтобы обеспечить высокую доступность топика, минимальное число синхронных реплик должно быть не меньше 2.

Добавляем пользователя



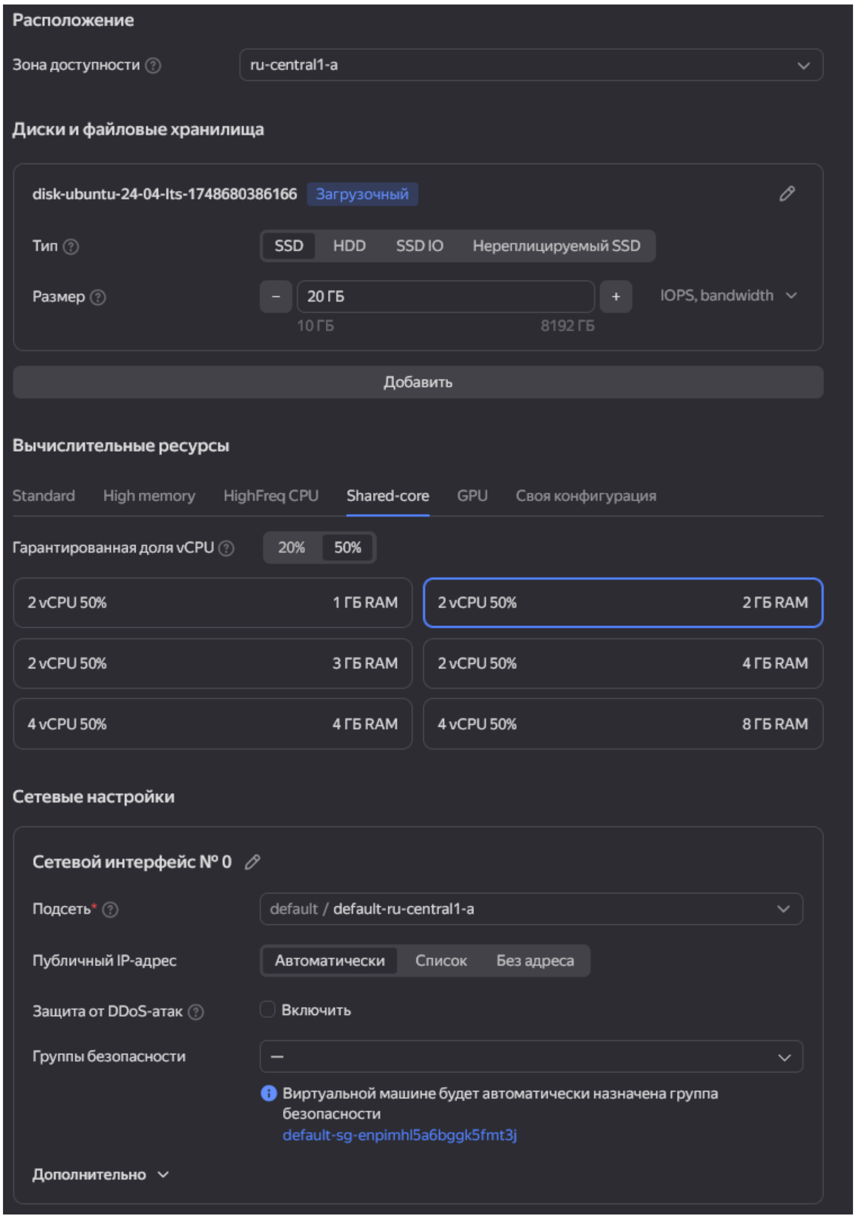
Добавление пользователя

Имя* user

Пароль* ●●●●●●●●

Топик	Роли
messages	ACCESS_ROLE_CONSUMER, ACCESS_ROLE_PRODUCER

Создаем виртуальную машину



Расположение

Зона доступности: ru-central1-a

Диски и файловые хранилища

disk-ubuntu-24-04-lts-1748680386166 Загрузочный

Тип: SSD | HDD | SSD IO | Нереплицируемый SSD

Размер: 20 ГБ (10 ГБ – 8192 ГБ) IOPS, bandwidth

Вычислительные ресурсы

Standard | High memory | HighFreq CPU | Shared-core | GPU | Своя конфигурация

Гарантированная доля vCPU: 20% | 50%

2 vCPU 50% 1 ГБ RAM	2 vCPU 50% 2 ГБ RAM
2 vCPU 50% 3 ГБ RAM	2 vCPU 50% 4 ГБ RAM
4 vCPU 50% 4 ГБ RAM	4 vCPU 50% 8 ГБ RAM

Сетевые настройки

Сетевой интерфейс № 0

Подсеть*: default / default-ru-central1-a

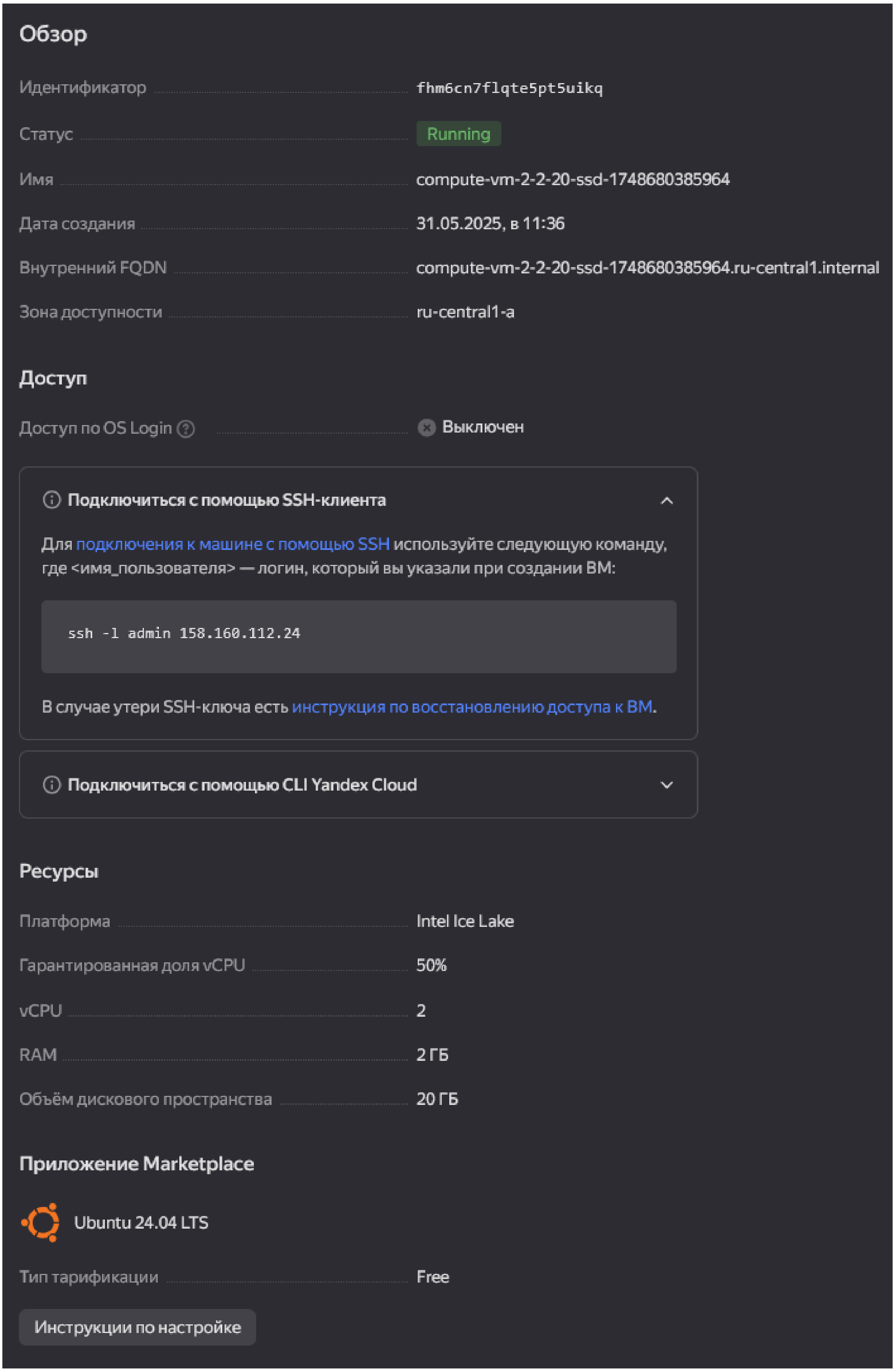
Публичный IP-адрес: Автоматически | Список | Без адреса

Защита от DDoS-атак: □ Включить

Группы безопасности: —

Виртуальной машине будет автоматически назначена группа безопасности default-sg-enpimhl5a6bggk5fmt3j

Результат



Обзор

Идентификатор	fhm6cn7flqte5pt5uikq
Статус	Running
Имя	compute-vm-2-2-20-ssd-1748680385964
Дата создания	31.05.2025, в 11:36
Внутренний FQDN	compute-vm-2-2-20-ssd-1748680385964.ru-central1.internal
Зона доступности	ru-central1-a

Доступ

Доступ по OS Login ?	☒ Выключен

ⓘ Подключиться с помощью SSH-клиента

Для подключения к машине с помощью SSH используйте следующую команду, где <имя_пользователя> — логин, который вы указали при создании ВМ:

```
ssh -l admin 158.160.112.24
```

В случае утери SSH-ключа есть инструкция по восстановлению доступа к ВМ.

ⓘ Подключиться с помощью CLI Yandex Cloud

Ресурсы

Платформа	Intel Ice Lake
Гарантированная доля vCPU	50%
vCPU	2
RAM	2 ГБ
Объём дискового пространства	20 ГБ

Приложение Marketplace

Ubuntu 24.04 LTS

Тип тарификации	Free

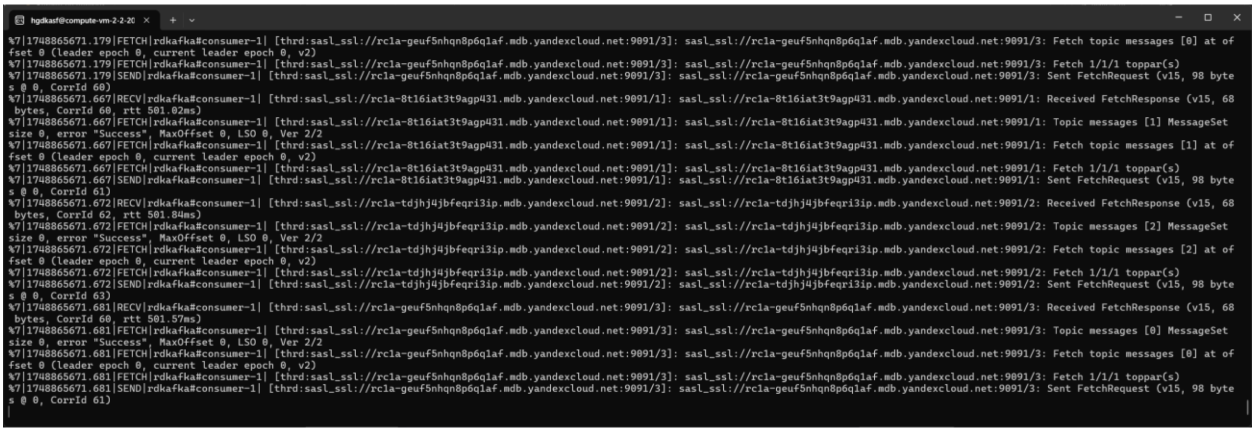
Код consumer.py

```
from confluent_kafka import Consumer

def error_callback(err):
    print('Something went wrong: {}'.format(err))

params = {
    'bootstrap.servers': 'rc1a-8t16iat3t9agp431.mdb.yandexcloud.net,rc1a-
geuf5nhqn8p6q1af.mdb.yandexcloud.net,rc1a-tdjhj4jbfeqri3ip.mdb.yandexcloud.net:9091',
    'security.protocol': 'SASL_SSL',
    'ssl.ca.location': '/usr/local/share/ca-
certificates/Yandex/YandexInternalRootCA.crt',
    'sasl.mechanism': 'SCRAM-SHA-512',
    'sasl.username': 'user',
    'sasl.password': '12345678',
    'group.id': 'test-consumer1',
    'auto.offset.reset': 'earliest',
    'enable.auto.commit': False,
    'error_cb': error_callback,
    'debug': 'all',
}

c = Consumer(params)
c.subscribe(['messages'])

while True:
    msg = c.poll(timeout=3.0)
    if msg:
        val = msg.value().decode()
        print(val)
```

Код producer.py

```
from confluent_kafka import Producer

def error_callback(err):
    print('Something went wrong: {}'.format(err))

params = {
    'bootstrap.servers': 'rc1a-8t16iat3t9agp431.mdb.yandexcloud.net,rc1a-
geuf5nhqn8p6q1af.mdb.yandexcloud.net,rc1a-tdjhj4jbfeqri3ip.mdb.yandexcloud.net:9091',
    'security.protocol': 'SASL_SSL',
    'ssl.ca.location': '/usr/local/share/ca-
certificates/Yandex/YandexInternalRootCA.crt',
    'sasl.mechanism': 'SCRAM-SHA-512',
    'sasl.username': 'user',
    'sasl.password': '12345678',
    'error_cb': error_callback,
}

p = Producer(params)
p.produce('messages', 'some payload1')
p.flush(10)
```

Результат



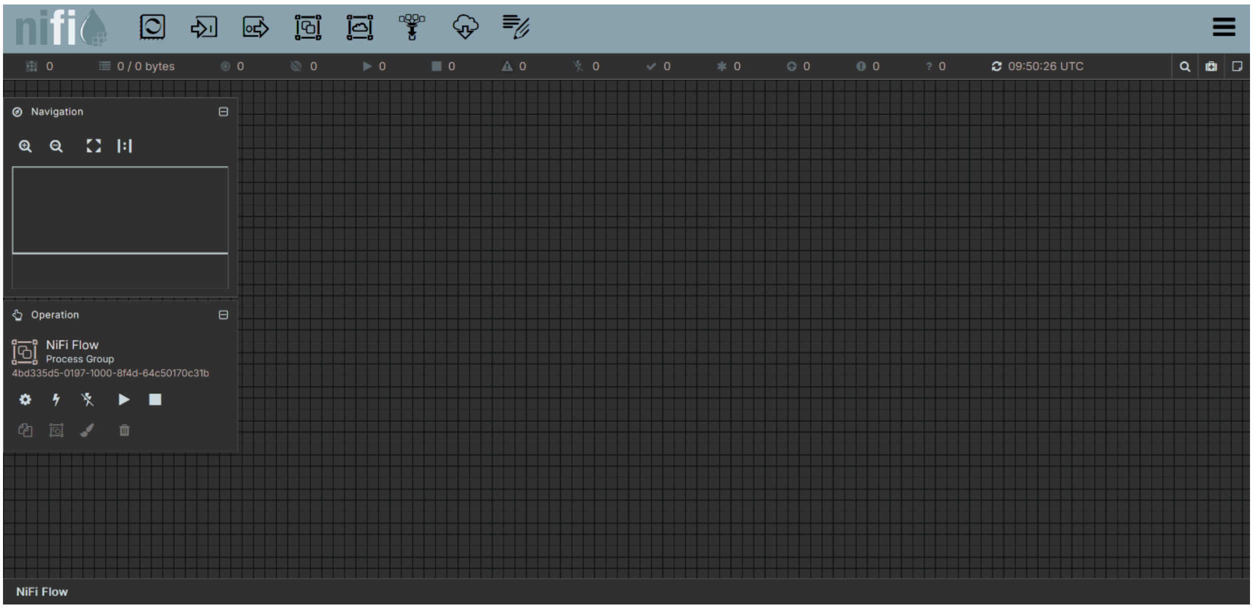
```
%7|1748865671.179|FETCH|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3]: sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3: Fetch topic messages [0] at of
fset 0 (leader epoch 0, current leader epoch 0, v2)
%7|1748865671.179|FETCH|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3]: sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3: Fetch 1/1/1 toppar(s)
%7|1748865671.179|SEND|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3]: sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3: Sent FetchRequest (v15, 98 byte
s @ 0, CorrId 60)
%7|1748865671.667|RECV|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-8t16iat3t9agp431.mdb.yandexcloud.net:9091/1]: sasl_ssl://rc1a-8t16iat3t9agp431.mdb.yandexcloud.net:9091/1: Received FetchResponse (v15, 68
 bytes, CorrId 60, rtt 501.02ms)
%7|1748865671.667|FETCH|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-8t16iat3t9agp431.mdb.yandexcloud.net:9091/1]: sasl_ssl://rc1a-8t16iat3t9agp431.mdb.yandexcloud.net:9091/1: Topic messages [1] MessageSet
size 0, error "Success", MaxOffset 0, LSO 0, Ver 2/2
%7|1748865671.667|FETCH|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-8t16iat3t9agp431.mdb.yandexcloud.net:9091/1]: sasl_ssl://rc1a-8t16iat3t9agp431.mdb.yandexcloud.net:9091/1: Fetch topic messages [1] at of
fset 0 (leader epoch 0, current leader epoch 0, v2)
%7|1748865671.667|FETCH|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-8t16iat3t9agp431.mdb.yandexcloud.net:9091/1]: sasl_ssl://rc1a-8t16iat3t9agp431.mdb.yandexcloud.net:9091/1: Fetch 1/1/1 toppar(s)
%7|1748865671.667|SEND|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-8t16iat3t9agp431.mdb.yandexcloud.net:9091/1]: sasl_ssl://rc1a-8t16iat3t9agp431.mdb.yandexcloud.net:9091/1: Sent FetchRequest (v15, 98 byte
s @ 0, CorrId 61)
%7|1748865671.672|RECV|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-tdjhj4jbfeqri3ip.mdb.yandexcloud.net:9091/2]: sasl_ssl://rc1a-tdjhj4jbfeqri3ip.mdb.yandexcloud.net:9091/2: Received FetchResponse (v15, 68
 bytes, CorrId 62, rtt 501.84ms)
%7|1748865671.672|FETCH|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-tdjhj4jbfeqri3ip.mdb.yandexcloud.net:9091/2]: sasl_ssl://rc1a-tdjhj4jbfeqri3ip.mdb.yandexcloud.net:9091/2: Topic messages [2] MessageSet
size 0, error "Success", MaxOffset 0, LSO 0, Ver 2/2
%7|1748865671.672|FETCH|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-tdjhj4jbfeqri3ip.mdb.yandexcloud.net:9091/2]: sasl_ssl://rc1a-tdjhj4jbfeqri3ip.mdb.yandexcloud.net:9091/2: Fetch topic messages [2] at of
fset 0 (leader epoch 0, current leader epoch 0, v2)
%7|1748865671.672|FETCH|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-tdjhj4jbfeqri3ip.mdb.yandexcloud.net:9091/2]: sasl_ssl://rc1a-tdjhj4jbfeqri3ip.mdb.yandexcloud.net:9091/2: Fetch 1/1/1 toppar(s)
%7|1748865671.672|SEND|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-tdjhj4jbfeqri3ip.mdb.yandexcloud.net:9091/2]: sasl_ssl://rc1a-tdjhj4jbfeqri3ip.mdb.yandexcloud.net:9091/2: Sent FetchRequest (v15, 98 byte
s @ 0, CorrId 63)
%7|1748865671.681|RECV|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3]: sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3: Received FetchResponse (v15, 68
 bytes, CorrId 60, rtt 501.57ms)
%7|1748865671.681|FETCH|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3]: sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3: Topic messages [0] MessageSet
size 0, error "Success", MaxOffset 0, LSO 0, Ver 2/2
%7|1748865671.681|FETCH|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3]: sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3: Fetch topic messages [0] at of
fset 0 (leader epoch 0, current leader epoch 0, v2)
%7|1748865671.681|FETCH|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3]: sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3: Fetch 1/1/1 toppar(s)
%7|1748865671.681|SEND|rdkafka#consumer-1| [thrd:sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3]: sasl_ssl://rc1a-geuf5nhqn8p6q1af.mdb.yandexcloud.net:9091/3: Sent FetchRequest (v15, 98 byte
s @ 0, CorrId 61)
```

Задание 2. Интеграция Kafka с внешними системами (Apache NiFi / Hadoop)

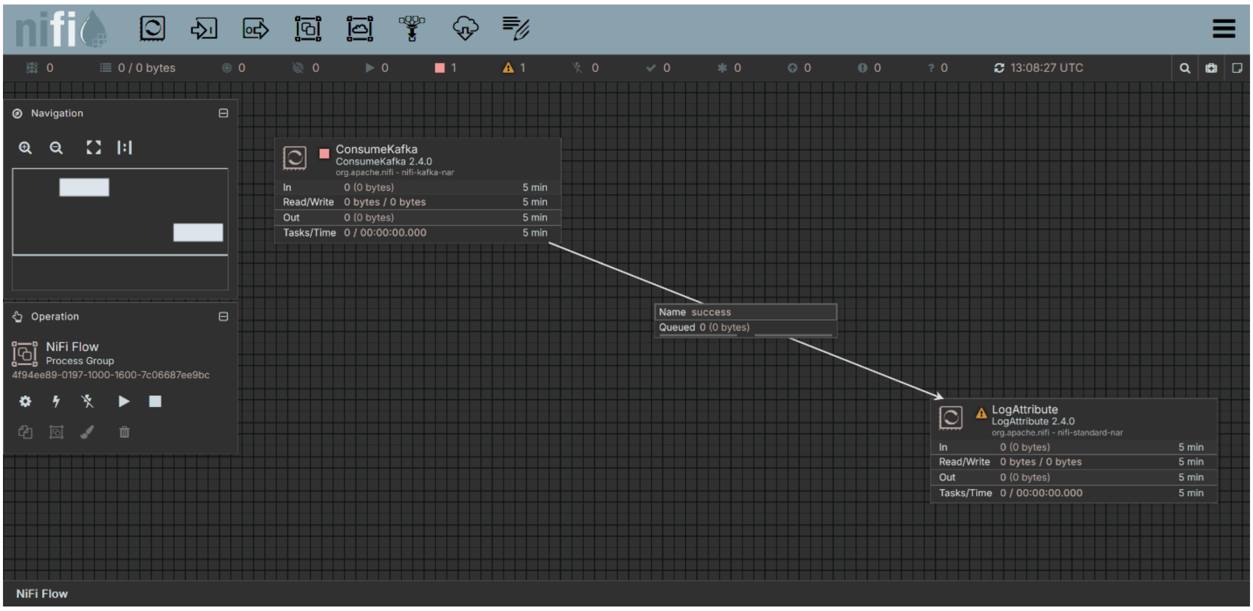
Установка Nifi

```
sudo apt update
sudo apt upgrade
sudo apt install unzip openjdk-21-jdk
wget https://dlcdn.apache.org/nifi/2.4.0/nifi-2.4.0-bin.zip
unzip nifi-2.4.0-bin.zip
rm nifi-2.4.0-bin.zip
mv nifi-2.4.0/ nifi/

sudo micro /etc/bash.bashrc
source /etc/bash.bashrc
echo $JAVA_HOME
cd nifi/bin/
micro ../conf/nifi.properties
bash nifi.sh start
tail ../logs/nifi-app.log

nifi.properties:

nifi.web.http.host=0.0.0.0
nifi.web.http.port=8484
nifi.web.https.host=
nifi.web.https.port=
nifi.security.keystore=
nifi.security.keystoreType=
nifi.security.keystorePasswd=
nifi.security.keyPasswd=
nifi.security.truststore=
nifi.security.truststoreType=
nifi.security.truststorePasswd=
nifi.security.user.login.identity.provider=
nifi.security.allow.anonymous.authentication=true
```

Результат



Создаем связи



ConsumeKafka
ConsumeKafka 2.4.0
org.apache.nifi - nifi-kafka-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min

Name success
Queued 0 (0 bytes)

LogAttribute
LogAttribute 2.4.0
org.apache.nifi - nifi-standard-nar

In	0 (0 bytes)	5 min
Read/Write	0 bytes / 0 bytes	5 min
Out	0 (0 bytes)	5 min
Tasks/Time	0 / 00:00:00.000	5 min