



## **РУКОВОДСТВО АДМИНИСТРАТОРА**

**Программное средство интеллектуальной обработки информации и ведения  
баз знаний «РЕД Центр Знаний»**

Листов 17

Москва

2026

## СОДЕРЖАНИЕ

<b>1. ВВЕДЕНИЕ</b> .....	3
<b>2. СИСТЕМНЫЕ ТРЕБОВАНИЯ</b> .....	3
<b>3. УСТАНОВКА СИСТЕМЫ</b> .....	3
3.1 ПОЛУЧЕНИЕ ДИСТРИБУТИВА.....	3
3.2 РАСПАКОВКА ДИСТРИБУТИВА.....	4
3.3 ЗАГРУЗКА ОБРАЗОВ.....	4
<b>4. НАСТРОЙКА КОНФИГУРАЦИИ СЕРВИСОВ</b> .....	5
4.1 ОСНОВНЫЕ ПАРАМЕТРЫ .....	5
4.2 ПЕРЕМЕННЫЕ ОКРУЖЕНИЯ .....	7
<b>5. ЗАПУСК СИСТЕМЫ</b> .....	7
<b>6. ОБНОВЛЕНИЕ СИСТЕМЫ</b> .....	7
6.1 ПОДГОТОВКА ОБНОВЛЕНИЯ .....	7
6.2 ЗАГРУЗКА НОВЫХ ОБРАЗОВ .....	7
6.3 ОБНОВЛЕНИЕ КОНФИГУРАЦИИ.....	7
6.4 ЗАПУСК ОБНОВЛЕННОЙ СИСТЕМЫ.....	8
<b>7. МОНИТОРИНГ И ОБСЛУЖИВАНИЕ</b> .....	8
7.1 ПРОВЕРКА ЛОГОВ.....	8
7.2 РЕЗЕРВНОЕ КОПИРОВАНИЕ.....	8
7.3 ОЧИСТКА СТАРЫХ ОБРАЗОВ .....	8
<b>8. УСТРАНЕНИЕ НЕПОЛАДОК</b> .....	8
8.1 ПРОБЛЕМЫ С ЗАПУСКОМ КОНТЕЙНЕРОВ .....	9
8.2 ПРОБЛЕМЫ С ПОДКЛЮЧЕНИЕМ К LLM.....	9
8.3 ОШИБКИ ХРАНЕНИЯ ДАННЫХ .....	9
<b>9. РЕКОМЕНДАЦИИ ДЛЯ ПРОДАКШЕНА</b> .....	9
9.1 ИЗОЛЯЦИЯ СЕТЕЙ.....	9
9.2 ОГРАНИЧЕНИЕ ПОРТОВ .....	9
9.3 CI/CD ИНТЕГРАЦИЯ .....	9
9.4 РЕЗЕРВНОЕ КОПИРОВАНИЕ.....	9
<b>10. ОСТАНОВКА И УДАЛЕНИЕ СИСТЕМЫ</b> .....	10
10.1 ОСТАНОВКА СИСТЕМЫ .....	10
10.2 ПОЛНОЕ УДАЛЕНИЕ.....	10
10.3 ЧАСТИЧНОЕ УДАЛЕНИЕ.....	10
<b>11. АДМИНИСТРИРОВАНИЕ СИСТЕМЫ</b> .....	10
11.1 СОЗДАНИЕ ТЕХНИЧЕСКОГО ПОЛЬЗОВАТЕЛЯ СИСТЕМЫ .....	10
11.2 СОЗДАНИЕ ПОЛЬЗОВАТЕЛЯ СИСТЕМЫ .....	15
11.3 РОЛИ ПОЛЬЗОВАТЕЛЕЙ .....	16
<b>ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ</b> .....	17

## 1. ВВЕДЕНИЕ

Программное средство интеллектуальной обработки информации и ведения баз знаний «РЕД Центр Знаний» представляет собой комплекс микросервисов, работающих в Docker- или Kubernetes- окружении.

Администратор отвечает за управление инфраструктурой, настройку параметров, обновления, мониторинг работоспособности и администрирование пользователей системы.

## 2. СИСТЕМНЫЕ ТРЕБОВАНИЯ

Для развертывания системы требуется:

- **Docker Engine** версии 20.10 и выше;
- **Docker Compose** версии 2.18 и выше;
- **Доступ к LLM:** сервер с OpenAI-совместимым API (например, Qwen/Qwen3-32B-AWQ);
- **Оперативная память:** не менее 8 ГБ;
- **Видеоадаптер NVIDIA с оперативной памятью:** 8 ГБ и более.

## 3. УСТАНОВКА СИСТЕМЫ

### 3.1 ПОЛУЧЕНИЕ ДИСТРИБУТИВА

Дистрибутив программного средства интеллектуальной обработки информации и ведения баз знаний «РЕД Центр Знаний» предоставляется в виде файлов-архивов.

Ссылка на дистрибутив может быть предоставлена при личном обращении в техническую поддержку продукта.

Поставка ПО включает:

Набор образов docker (images.tar.bz2):

1. code.red-soft-center.ru/docker/red-database:3.0.20-ubi8 - образ Ред Базы Данных;
2. code.red-soft-center.ru/docker/minio-redos:ubi8 - образ s3 хранилища;
3. code.red-soft-center.ru/services/red-vector-db:latest - образ РЕД Вектор Данных;
4. code.red-soft-center.ru/ncore/red-archive:latest - образ Ред Платформы;
5. code.red-soft-center.ru/web/red-archive:latest - образ веб приложения РЕД Центр Знаний;
6. code.red-soft-center.ru/datasience/ml-ncore-adapter:latest - образ ML Адаптер;
7. code.red-soft-center.ru/datasience/rag-service:base-ubi8 - образ ML Chat.

Дополнительные файлы (files.tar):

8. docker-compose.yml - файл конфигурации сервисов для Docker Composes;
9. db/ncore-rack-dev.rdb - файл базы данных для Ред Базы Данных;
10. db/security3.fdb - база данных безопасности в Ред Базе Данных.

По умолчанию, дистрибутив состоит из двух файлов:

- 1.images.tar.bz2 - все образы docker;
- 2.files.tar - архив с базами данных и docker-compose.yml;
- 3.README.md - описание дистрибутива, а также инструкция по первоначальной настройке.

### 3.2 РАСПАКОВКА ДИСТРИБУТИВА

В начале установки нужно создать папку установки, в данном примере /srv/redarchive:

```
sudo mkdir -p /srv/redarchive
```

Распаковать файлы базы данных и настроек (files.tar) в данную папку:

```
$ sudo tar --same-owner -pxvf files.tar -C /srv/redarchive
```

```
db/
```

```
db/security3.fdb
```

```
db/ncore-rack-dev.rdb
```

```
docker-compose.yml
```

```
s3data/
```

После распаковки папка установки выглядит так (пример):

```
$ sudo tree -up /srv/redarchive
```

```
[drwxr-xr-x root  ] /srv/redarchive
```

```
|— [drwx----- root  ] db
```

```
| |— [-rw-rw-r-- root  ] ncore-rack-dev.rdb
```

```
| |— [-rw-r--r-- root  ] security3.fdb
```

```
|— [-rw-r--r-- root  ] docker-compose.yml
```

```
|— [drwxr-xr-x 998   ] s3data
```

Примечание: Обратите внимание на то что владельцем пустой папки s3data является пользователь с идентификатором 998. Это очень важно для работы s3 хранилища.

Включает базы данных (файлы fdb и rdb), а также файл конфигурации Docker Compose.

### 3.3 ЗАГРУЗКА ОБРАЗОВ

Загрузить образы из архива (images.tar.bz2) в локальный Docker-репозиторий:

```
sudo docker load -i images.tar.bz2
```

Выполнение команды может занимать значительное время. При успешной загрузке образов в локальный репозиторий docker сообщит соответствующими логами:

\$ sudo docker load -i images.tar.bz2
...
Loaded image: code.red-soft-center.ru/docker/minio-redos:ubi8
...
Loaded image: code.red-soft-center.ru/services/red-vector-db:latest
...
Loaded image: code.red-soft-center.ru/ncore/red-archive:latest
Loaded image: code.red-soft-center.ru/datasience/ml-ncore-adapter:latest
Loaded image: code.red-soft-center.ru/datasience/rag-service:base-ubi8
...
Loaded image: code.red-soft-center.ru/web/red-archive:latest
...
Loaded image: code.red-soft-center.ru/docker/red-database:3.0.20-ubi8

## 4. НАСТРОЙКА КОНФИГУРАЦИИ СЕРВИСОВ

### 4.1 ОСНОВНЫЕ ПАРАМЕТРЫ

Настройка конфигурации Docker Compose хранится в файле `docker-compose.yml`. Необходимо убедиться, что переменные окружения соответствуют локальной инфраструктуре.

Файл `docker-compose.yml` снабжен комментариями для удобства его настройки.

При этом критически важными параметрами являются настройки сервиса `ml-chat`, которые определяют доступ к LLM:

1. `OPENAI_BASEURL` - определяет доступ к OpenAI-совместимому API (LLM): по умолчанию, из переменной окружения Host сервера, `OPENAI_BASEURL`
2. `OPENAI_API_KEY` - определяет ключ доступа к OpenAI-совместимому API: по умолчанию, из переменной окружения Host сервера, `OPENAI_API_KEY`
3. `OPENAI_MODEL` - название модели: по умолчанию, из переменной окружения Host сервера, `OPENAI_MODEL` или `Qwen/Qwen3-32B-AWQ`

Дополнительно, могут потребоваться параметры, определяющие доступ к GPU адаптеру. Это блоки вида:

deploy:
resources:
reservations:

devices:
- driver: nvidia
count: -1
capabilities: [gpu]

И параметры с постфиксом `_DEVICE`, например:

- 'EMBEDDER_DEVICE=cuda'
- 'SPARSE_DEVICE=cuda'

В зависимости от конфигурации GPU адаптеров в системе, данные настройки могут потребовать изменения.

Примечание: рекомендуется завести переменные окружения, связанные с настройкой доступа к OpenAI API (`OPENAI_BASEURL`, `OPENAI_API_KEY`, `OPENAI_MODEL`) или подключить эти переменные окружения через **env** файл и параметр **env\_file** (пример):

\$ cat /srv/redarchive/docker-compose.yml
...
ml-chat:
image: code.red-soft-center.ru/datasience/rag-service:base-ubi8
volumes:
- ./chat/.cache:/root/.cache
environment:
- 'LOG_LEVEL=info'
- 'REDVECTOR_HOST=http://vector:6333'
- 'REDVECTOR_INDEX=red_archive'
- 'RAY_DISABLED=true'
- 'EMBEDDER_DEVICE=cuda'
- 'RERANKER_DEVICE=cuda'
- 'FILDERPROD_DEVICE=cuda'
env_file:
- ./env
...
\$ cat ./env

OPENAI_BASEURL=http://llm.red-soft-center.ru/v1
OPENAI_MODEL=Qwen/Qwen3-32B-AWQ
OPENAI_API_KEY=12345

Несмотря на то, что файл конфигурации построен таким образом, чтобы быть готовым к немедленному запуску, рекомендуется уточнить параметры:

- volumes – места хранения данных;
- ports – порты;
- environment – переменные окружения;
- переменные, отвечающие за аутентификацию.

#### 4.2 ПЕРЕМЕННЫЕ ОКРУЖЕНИЯ

Критически важные переменные приведены в Таблице 1:

**Таблица 1. Переменные окружения.**

Сервис	Переменная	Описание
adapter-load	EMBEDDER_DEVICE	Устройство для модели dense-эмбеддингов, например: `cuda`
adapter-load	SPARSE_DEVICE	Устройство для sparse-эмбеддингов, например: `cuda`
ml-chat	EMBEDDER_DEVICE	Устройство для модели dense-эмбеддингов, например: `cuda`
ml-chat	RERANKER_DEVICE	Устройство для модели реранкера, например: `cuda`
ml-chat	FILDERPROD_DEVICE	Устройство для модели filterprod, например: `cuda`
ml-chat	OPENAI_BASEURL	URL к LLM-серверу, например: http://LLM_HOST:PORT/v1
ml-chat	OPENAI_API_KEY	Ключ доступа к LLM
ml-chat	OPENAI_MODEL	Название модели (по умолчанию: Qwen/Qwen3-32B-AWQ)

Рекомендуется использовать **.env** файл для хранения секретов:

env_file:
- ./env

## 5. ЗАПУСК СИСТЕМЫ

Для первоначального запуска системы рекомендуется обратиться в документацию **Руководство по развертыванию тестового экземпляра**.

## 6. ОБНОВЛЕНИЕ СИСТЕМЫ

### 6.1 ПОДГОТОВКА ОБНОВЛЕНИЯ

Скачайте новые архивы образов.

### 6.2 ЗАГРУЗКА НОВЫХ ОБРАЗОВ

Повторите шаги из **разд. 3 Установка системы**.

### 6.3 ОБНОВЛЕНИЕ КОНФИГУРАЦИИ

Если в новом дистрибутиве изменился `docker-compose.yml`:

1. Сравните текущий и новый файлы.
2. Обновите переменные окружения и параметры под вашу инфраструктуру.

### 6.4 ЗАПУСК ОБНОВЛЕННОЙ СИСТЕМЫ

```
docker compose up -d
```

Текущий статус сервисов можно проверить командой:

```
docker compose ps #
```

## 7. МОНИТОРИНГ И ОБСЛУЖИВАНИЕ

### 7.1 ПРОВЕРКА ЛОГОВ

Отслеживайте логи для диагностики:

```
docker compose logs -f redvector
```

```
docker compose logs -f rag-service
```

### 7.2 РЕЗЕРВНОЕ КОПИРОВАНИЕ

- Ред База Данных: Используйте инструменты Ред Базы Данных.
- MinIO (S3): Настройте бэкап через CLI/API.
- РЕД Вектор Данных: Используйте инструменты РЕД Вектор Данных.

### 7.3 ОЧИСТКА СТАРЫХ ОБРАЗОВ

Удалите устаревшие образы:

```
docker image prune -a --force
```

## 8. УСТРАНЕНИЕ НЕПОЛАДОК

### 8.1 ПРОБЛЕМЫ С ЗАПУСКОМ КОНТЕЙНЕРОВ

- Проверьте логи:

```
docker compose logs -f <SERVICE_NAME>
```

- Убедитесь, что порты не заняты другими процессами.

### 8.2 ПРОБЛЕМЫ С ПОДКЛЮЧЕНИЕМ К LLM

- Проверьте доступность сервера LLM по указанному URL.
- Убедитесь, что OPENAI\_API\_KEY корректен.

### 8.3 ОШИБКИ ХРАНЕНИЯ ДАННЫХ

- Проверьте права доступа к директориям volumes.
- Убедитесь, что MinIO и база данных запущены.

## 9. РЕКОМЕНДАЦИИ ДЛЯ ПРОДАКШЕНА

### 9.1 ИЗОЛЯЦИЯ СЕТЕЙ

Используйте отдельные Docker-сети:

networks:
red-net:
driver: bridge
services:
redvector:
networks:
- red-net

### 9.2 ОГРАНИЧЕНИЕ ПОРТОВ

Открывайте только необходимые порты (например, 80 для UI, 443 для HTTPS).

### 9.3 CI/CD ИНТЕГРАЦИЯ

Настройте автоматическое обновление образов, например, через CI/CD.

### 9.4 РЕЗЕРВНОЕ КОПИРОВАНИЕ

Настройте ежедневные бэкапы для данных, см. подразд. 7.3 Резервное копирование.

## 10. ОСТАНОВКА И УДАЛЕНИЕ СИСТЕМЫ

### 10.1 ОСТАНОВКА СИСТЕМЫ

```
docker compose down
```

Эта команда:

- останавливает контейнеры;
- удаляет их и сети;
- сохраняет тома (данные).

### 10.2 ПОЛНОЕ УДАЛЕНИЕ

```
docker compose down --volumes --rmi all
```

**Внимание!** Это удалит все данные и образы. Рекомендуется предварительно сделать бэкап.

### 10.3 ЧАСТИЧНОЕ УДАЛЕНИЕ

```
docker compose stop <SERVICE_NAME>
```

```
docker compose rm <SERVICE_NAME>
```

Данная команда позволяет остановить и удалить отдельный сервис комплекса.

**Внимание!** Любое удаление отдельного сервиса сделает систему не работоспособной! Данная команда может быть полезна для отдельных административных задач.

## 11. АДМИНИСТРИРОВАНИЕ СИСТЕМЫ

### 11.1 СОЗДАНИЕ ТЕХНИЧЕСКОГО ПОЛЬЗОВАТЕЛЯ СИСТЕМЫ

Для нужд сервиса ML Адаптер необходим доступ к системе через права технического пользователя.

В представленном в комплекте дистрибутива файле конфигураций Docker Compose `docker-compose.yml` предполагается использование пользователя (в переменных окружения `ADAPTER_NCORE_USER`, `ADAPTER_NCORE_PASSWORD`), по умолчанию:

1. Имя: `TECHNICAL_USER`
2. Пароль: `12345`

Этого пользователя не существует после установки, потому работа ML Адаптеров (сервисы настроены для работы через технического пользователя системы) после запуска системы происходит не корректно. Поэтому требуется вручную после запуска системы создать технического пользователя.

Для начала надо открыть страницу системы в браузере: `http://localhost:8080`

Далее войти под учетной записью системного администратора Ред Базы Данных SYSDBA

(Рисунок 1):

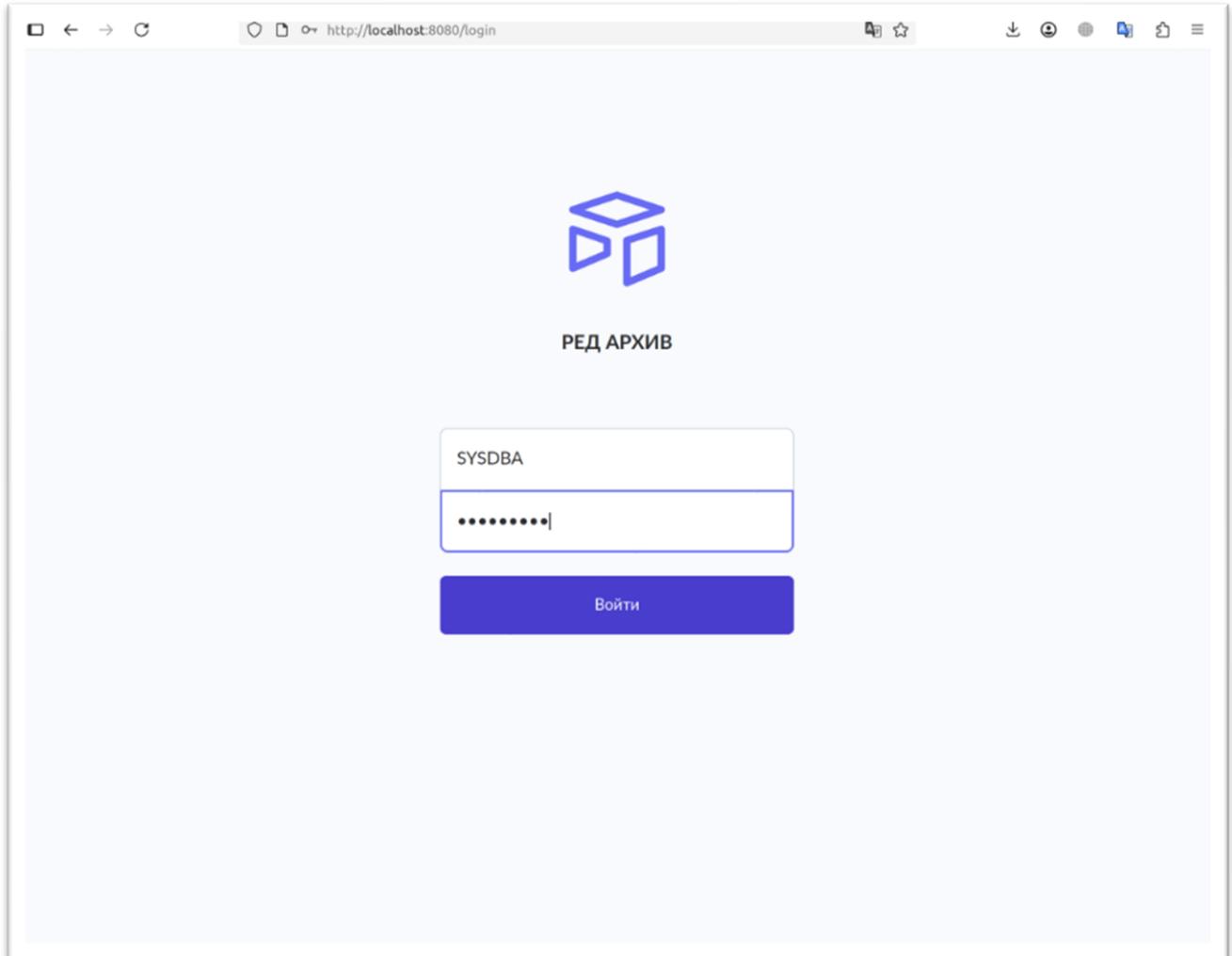


Рисунок 1 – Страница авторизации

После входа необходимо перейти в настройки системы (пункт меню Настройки) и далее пункт меню Пользователи, в котором нажать кнопку *Создать пользователя* (Рисунок 2):

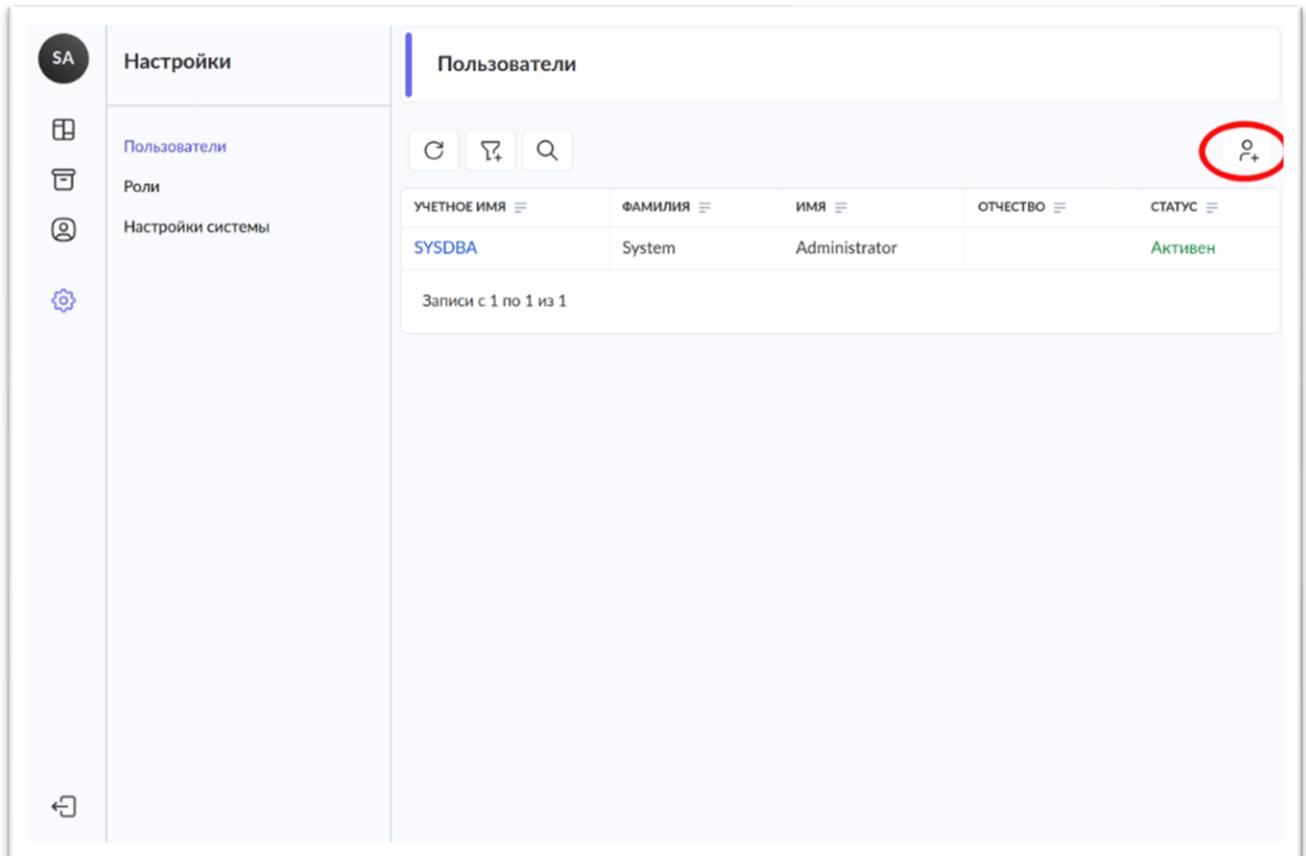
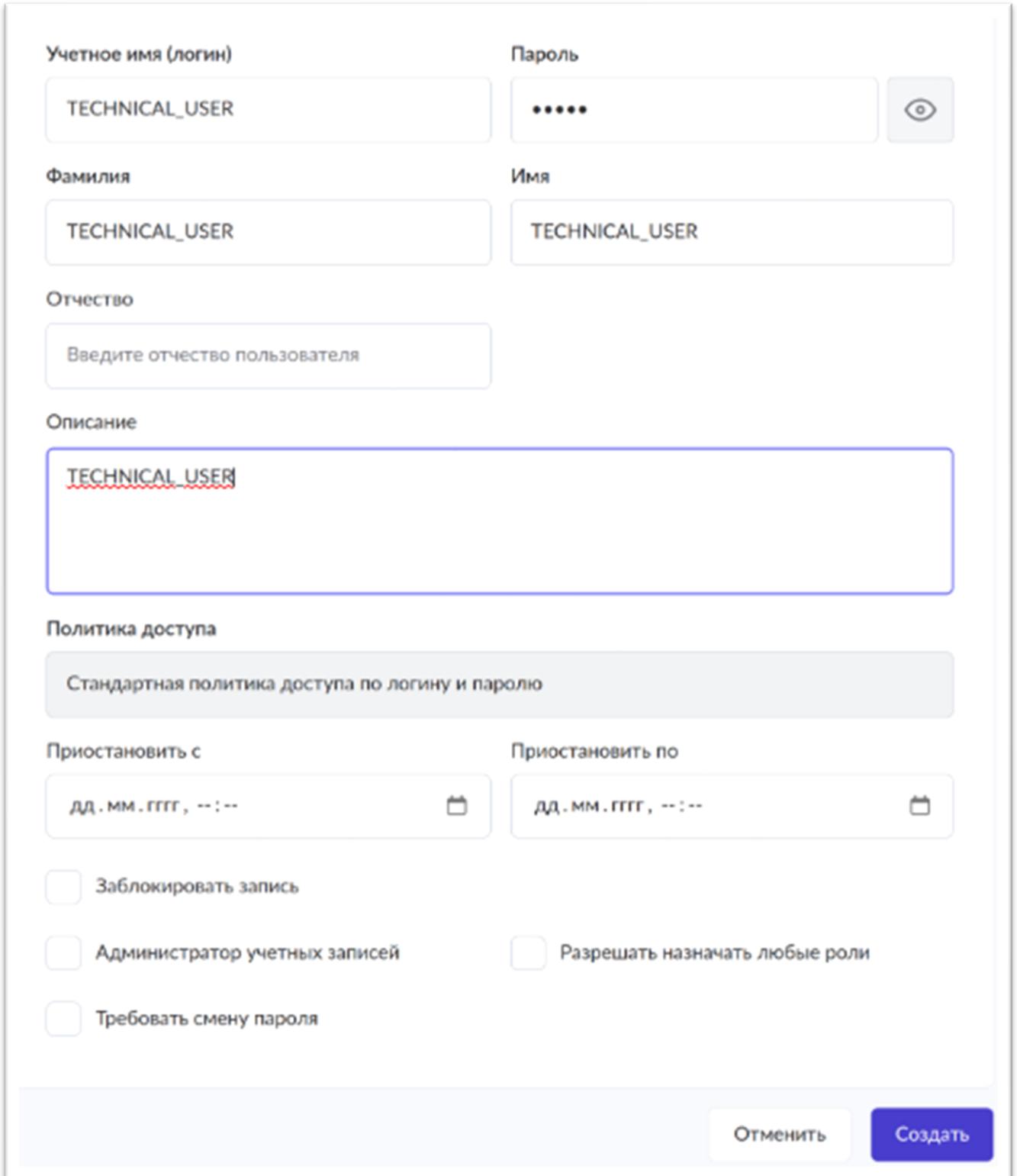


Рисунок 2 – Меню Настройки

Далее заполнить форму добавления нового пользователя (Рисунок 3):



Учетное имя (логин)

TECHNICAL\_USER

Пароль

.....

Фамилия

TECHNICAL\_USER

Имя

TECHNICAL\_USER

Отчество

Введите отчество пользователя

Описание

TECHNICAL\_USER

Политика доступа

Стандартная политика доступа по логину и паролю

Приостановить с

дд.мм.гггг, --:--

Приостановить по

дд.мм.гггг, --:--

Заблокировать запись

Администратор учетных записей

Разрешать назначать любые роли

Требовать смену пароля

Отменить

Создать

Рисунок 3 – Заполнение формы нового пользователя

После этого выбрать нового пользователя в списке пользователей (Рисунок 4):

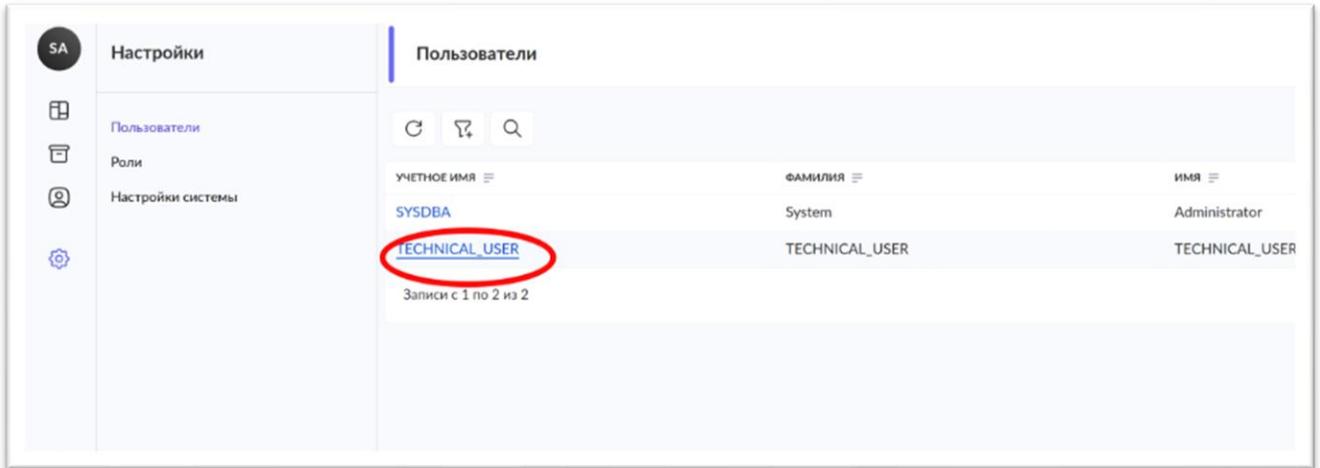


Рисунок 4 – Выбор карточки пользователя

Перейти на Форму пользователя в закладку *Назначить роли*, на которой добавить текущего пользователя во все роли нажатием кнопки  для данного пользователя (Рисунок 5):

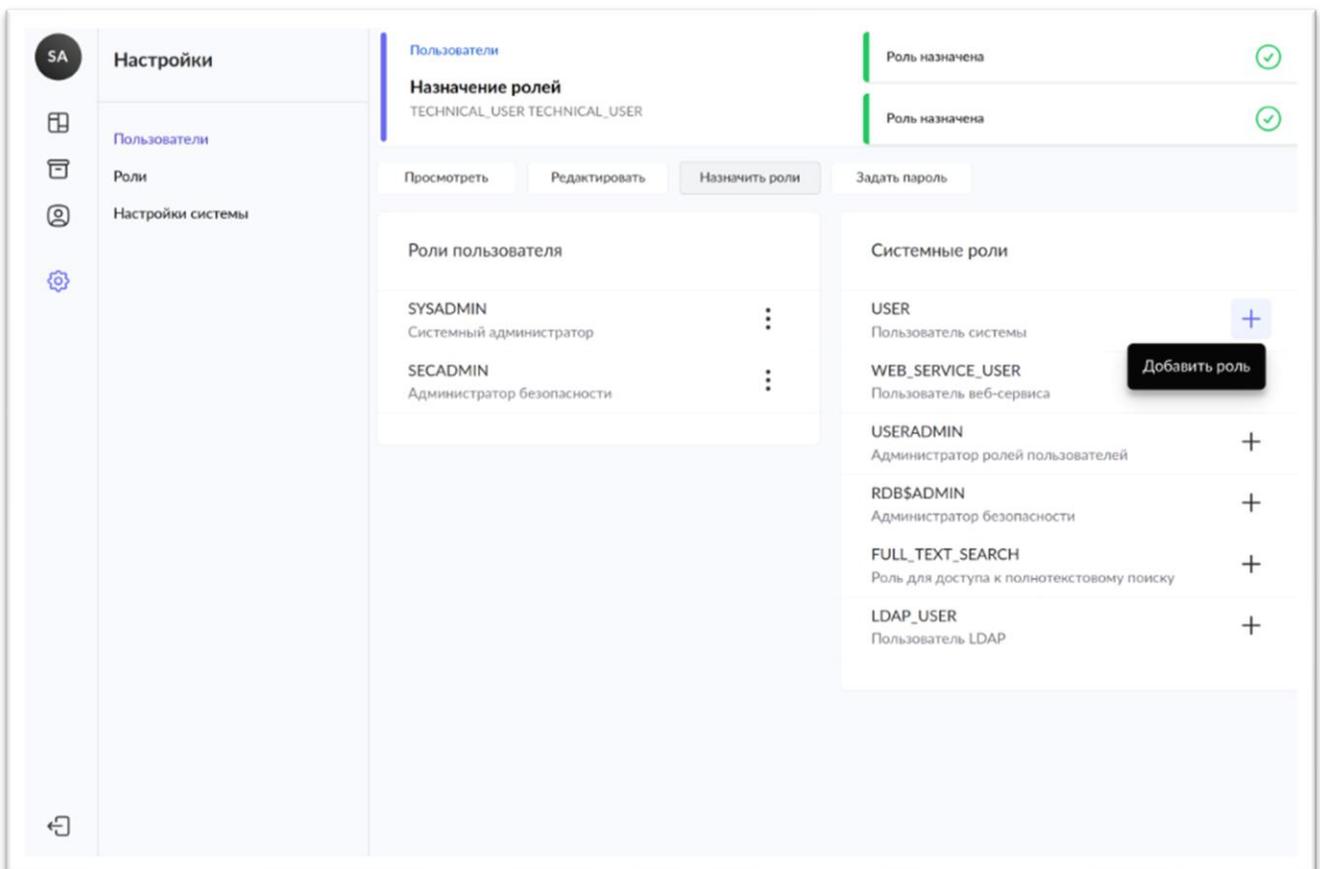


Рисунок 5 – Добавление роли пользователю

Таким образом у технического пользователя должны появиться все роли (Рисунок 6):

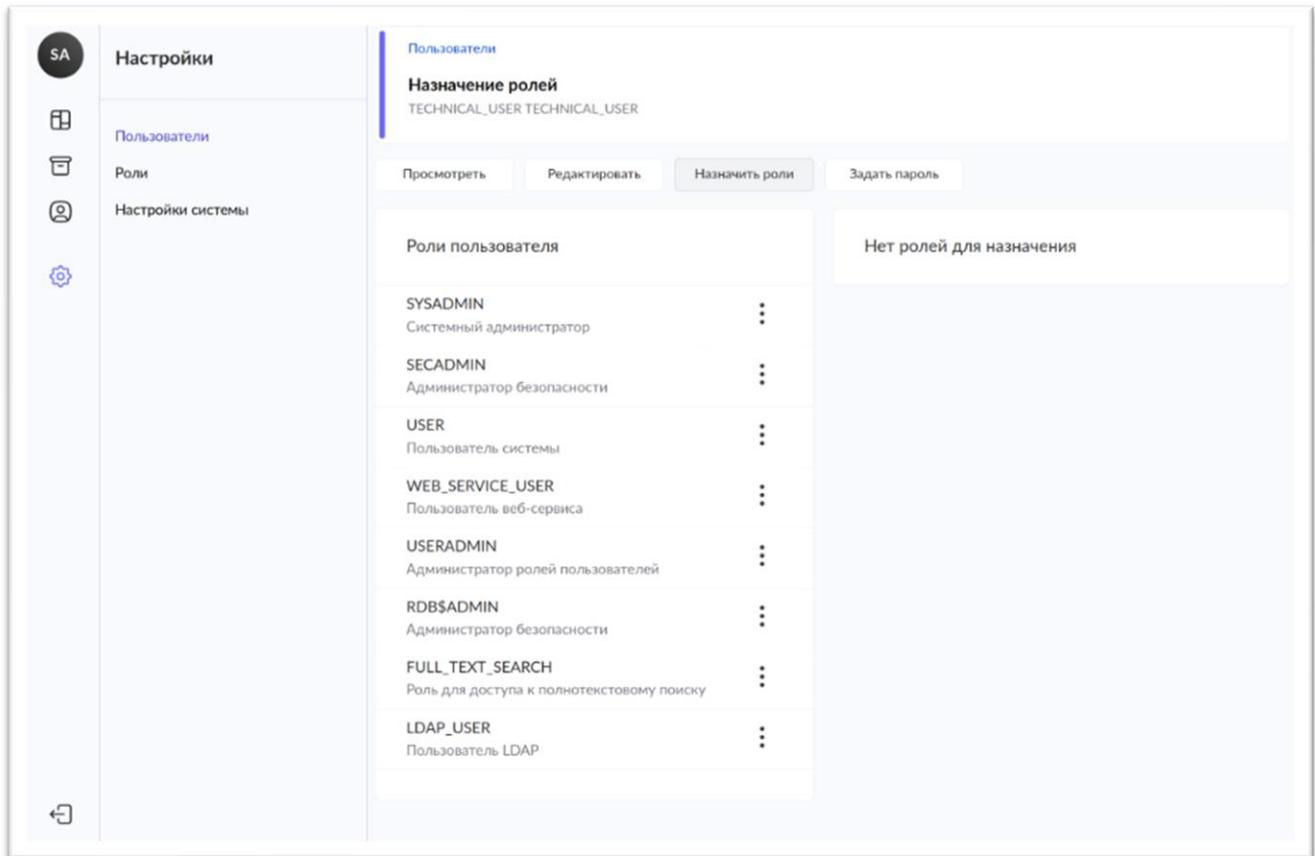


Рисунок 6 – Список ролей пользователя

Технический пользователь теперь может использоваться для работы с системой.

## 11.2 СОЗДАНИЕ ПОЛЬЗОВАТЕЛЯ СИСТЕМЫ

Процедура создания нового пользователя не отличается от создания технического пользователя.

Подробнее в подразд. 11.1 Создание технического пользователя системы.

Кратко:

1. Зайдите в веб-интерфейс по адресу <http://localhost:8080> под пользователем SYSDBA.

2. Перейдите в раздел Настройки → Пользователи → Создать пользователя.

3. Заполните форму:

- Имя пользователя;
- Пароль пользователя;
- Остальные поля.

4. Назначьте пользователю необходимые роли через кнопку . Описание ролей см. подразд. 11.3 Роли пользователей.

### 11.3 РОЛИ ПОЛЬЗОВАТЕЛЕЙ

Описание ролей пользователей представлено ниже в Таблице 2:

**Таблица 2. Описание ролей пользователей.**

<b>Роль</b>	<b>Описание</b>
SYSADMIN	Системный администратор
SECADMIN	Администратор безопасности
USER	Пользователь системы
WEB_SERVICE_USER	Пользователь веб-сервиса
USERADMIN	Администратор ролей пользователей
RDB\$ADMIN	Администратор безопасности БД
FULL_TEXT_SEARCH	Роль для доступа к полнотекстовому поиску
LDAP_USER	Пользователь LDAP

