

Краткое руководство по установке ПО «Тезис»

Система автоматической расшифровки и протоколирования записей совещаний. Установка из поставки с готовыми образами контейнеров на сервер заказчика, без доступа в сеть Интернет.

Редакция документа: 31.08.2026

Применимо к поставке: каталог `insight-kit`

Кто выполняет: системный администратор заказчика

Время установки: 20–40 минут

Полная версия: «Руководство по установке ПО «Тезис»» (`docs/install-guide.md`)

1. О программе

«Тезис» принимает записи совещаний, автоматически преобразует их в текст, определяет говорящих и формирует структурированный протокол по заданному шаблону. Источники записей: серверы видео-конференц-связи Vinteo (опрос по расписанию), ручная загрузка через веб-интерфейс, внешние источники через коннектор записей.

ПО выполнено набором контейнеров под управлением Docker Compose. Интерфейс — веб, на русском языке, доступ по HTTPS. Все компоненты, включая модели искусственного интеллекта, входят в поставку и работают локально.

2. Требования к серверу

Параметр	Вариант А — локальный контур (рекомендуемый)	Вариант Б — внешние сервисы ИИ
Процессор	от 16 ядер x86_64	от 4 ядер x86_64
Оперативная память	от 32 ГБ	от 8 ГБ
Диск под систему и образы	от 150 ГБ	от 50 ГБ
Диск под медиафайлы	≈1 ГБ на час аудио с учётом срока хранения (RETENTION_DAYS , по умолчанию 30 дней)	
Графический ускоритель	для языковой модели (ориентир — от 24 ГБ видеопамати); для распознавания речи — желателен	не требуется
ОС и ПО	Linux x86_64 (проверено: Ubuntu 22.04, Debian 12/13), Docker 24.0+ и Docker Compose v2	

Сетевые порты: входящие **443/tcp** (веб-интерфейс и API) и **80/tcp** (перенаправление на 443). Порты 5432, 6379, 8080, 9000–9003 слушают только 127.0.0.1 и наружу не публикуются. Доступ в Интернет при установке не требуется.

3. Подготовка к установке

- Проверить Docker:** `docker --version` и `docker compose version` (нужен Docker 24.0+ и Compose v2). Если Docker отсутствует, на Ubuntu/Debian установщик поставит его сам из официального репозитория.
- Перенести каталог поставки insight-kit** на сервер — по сети или на съёмном носителе (носитель форматировать в exFAT или ext4: FAT32 не хранит файлы больше 4 ГБ).

```
rsync -a insight-kit/ user@server:~/insight-kit/
```

3. **Получить отдельным каналом** (не с того же носителя!) публичный ключ подписи поставки и доверенный сценарий запуска:

```
sudo install -m 0755 run-signed-kit.sh /usr/local/sbin/insight-kit-run
sudo install -m 0644 kit-signing-public.pem /etc/insight/kit-signing-public.pem
```

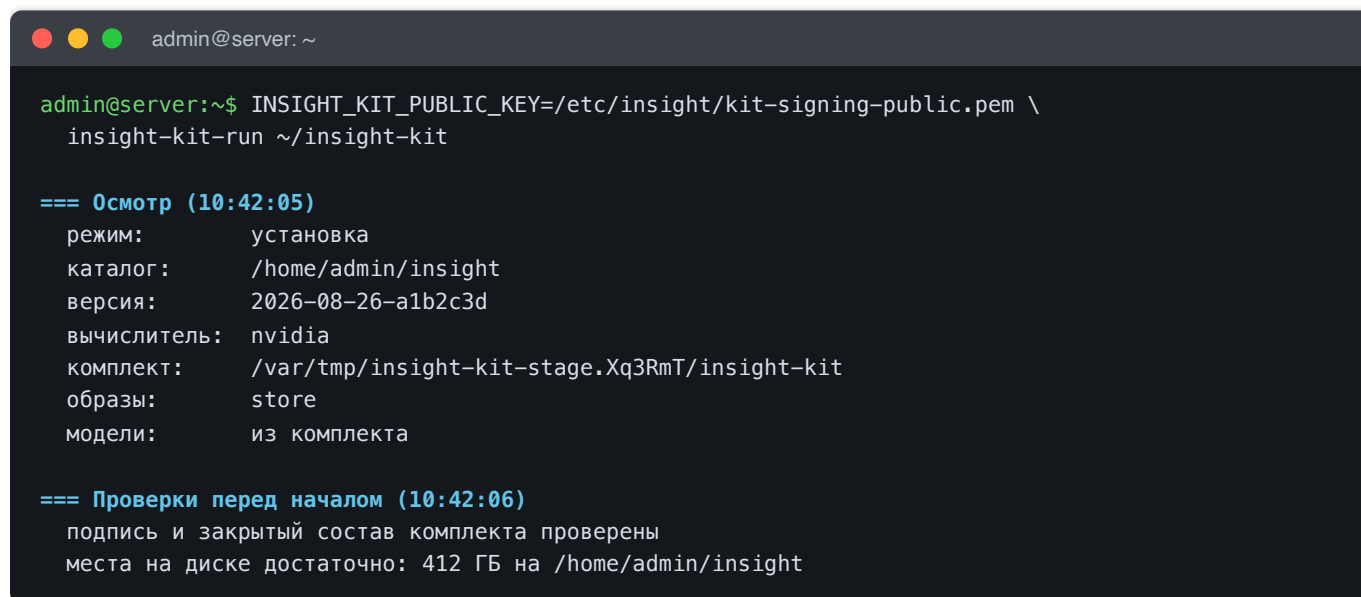
4. **При желании проверить подпись вручную** (установщик сделает это сам):

```
openssl dgst -sha256 -verify /etc/insight/kit-signing-public.pem \
  -signature ~/insight-kit/SHA256SUMS.sig ~/insight-kit/SHA256SUMS
(cd ~/insight-kit && sha256sum -c --strict SHA256SUMS)
```

Ожидается `Verified OK` и `OK` по каждому файлу.

4. Установка

Вся установка выполняется одной командой. Сценарий сам проверяет подпись и состав поставки, Docker, место и порты, задаёт несколько вопросов, раскладывает образы и модели, создаёт конфигурацию и сертификат, инициализирует базу, запускает службы и завершается приёмочной проверкой — прогоном записи через все стадии обработки.



```
admin@server: ~
admin@server:~$ INSIGHT_KIT_PUBLIC_KEY=/etc/insight/kit-signing-public.pem \
insight-kit-run ~/insight-kit

=== Осмотр (10:42:05)
режим:          установка
каталог:        /home/admin/insight
версия:         2026-08-26-a1b2c3d
вычислитель:    nvidia
комплект:       /var/tmp/insight-kit-stage.Xq3RmT/insight-kit
образы:         store
модели:         из комплекта

=== Проверки перед началом (10:42:06)
подпись и закрытый состав комплекта проверены
места на диске достаточно: 412 ГБ на /home/admin/insight
```

Рис. 1. Запуск установки: осмотр машины и проверки перед началом.

Проверить сервер, ничего не меняя, можно заранее — той же командой с ключом `--check`. Скрипт выполнит все проверки и завершится сообщением «Проверки пройдены. Ничего не менялось (`--check`)».

4.1. Вопросы установщика

Установщик задаёт два основных вопроса — какие службы распознавания поднять на этой машине и где взять языковую модель, — затем показывает план действий и ждёт подтверждения.

admin@server: ~

=== Настройка (10:42:11)

Какие службы поднять на этой машине?

- 1) всё локально: распознавание, детектор речи, разделение и опознание говорящих
- 2) только распознавание и детектор речи (разделение говорящих отключено)
- 3) ничего — все модели снаружи, здесь только приложение

Выбор [1]: 1

службы: asr,vad,diarizer,identity

токен Hugging Face не нужен: модели разделения говорящих в комплекте

Где взять языковую модель (нужна для протоколов и ассистента)?

- 1) здесь, в контейнере рядом с приложением — нужен файл .gguf
- 2) уже запущена на этой машине, вне docker (llama.cpp, vLLM, LM Studio)
- 3) на другой машине — укажу адрес
- 4) пока никак — расшифровки будут, протоколов и ассистента не будет

Выбор [1]: 3

Адрес модели [http://10.0.0.20:12345/v1]: http://10.0.0.20:12345/v1

Имя модели (как она отвечает в /v1/models): qwen3

Сейчас будет сделано:

- 1) подготовка каталога /home/admin/insight
- 2) получение образов (store)
- 3) подготовка моделей
- 4) создание базы и запуск
- 5) приёмочная проверка

Не выключайте сервер до сообщения «ГОТОВО».

Продолжить — введите ДА для продолжения: ДА

Рис. 2. Вопросы установщика и подтверждение плана действий.

4.2. Ход установки

```
admin@server: ~  
  
=== Конфигурация (10:43:02)  
  .env создан из .env.example  
  cookie сессии помечены Secure  
  
=== Образы (store) (10:43:04)  
  подгружаю точный служебный образ registry:2 из комплекта...  
  прочитан insight-api:2026-08-26-a1b2c3d  
  прочитан insight-worker:2026-08-26-a1b2c3d  
  прочитан insight-beat:2026-08-26-a1b2c3d  
  прочитан insight-asr:2026-08-26-a1b2c3d  
  прочитан insight-vad:2026-08-26-a1b2c3d  
  прочитан insight-diarizer:2026-08-26-a1b2c3d  
  прочитан insight-identity:2026-08-26-a1b2c3d  
  
=== Модели (10:47:31)  
  каталог: /home/admin/insight/models  
  копирую asr...  
  копирую vad...  
  копирую diarizer...  
  копирую identity...  
  
=== База данных (10:52:18)  
  структура базы приведена к версии 2026-08-26-a1b2c3d  
  
=== Запуск (10:52:40)
```

Рис. 3. Раскладка образов и моделей, инициализация базы, запуск служб.

4.3. Проверка и приёмка

Установщик сам убеждается, что все контейнеры поднялись, приложение здорово, обработчики очередей на связи, службы распознавания отвечают по тем адресам, по которым к ним обращается приложение, а языковая модель отзывается под именем из конфигурации. Затем он прогоняет приёмочную запись через все стадии и печатает время каждой. Если что-то не сошлось — установка считается невыполненной.

```
admin@server: ~  
  
=== Проверка (10:53:12)  
  контейнеры на месте: 16  
  /health: healthy  
  {"status":"healthy","version":"0.1.0","build":"2026-08-26-a1b2c3d",...}  
  обработчиков очередей: 9  
  очередь assistant-control: обслуживается  
  asr: отвечает (http://asr:8000)  
  vad: отвечает (http://vad:8000)  
  diarizer: отвечает (http://diarizer:8000)  
  identity: отвечает (http://identity:8000)  
  языковая модель: отвечает, имя совпадает
```

Рис. 4. Проверка служб: контейнеры, здоровье приложения, обработчики очередей, службы распознавания и языковая модель.

```
admin@server: ~  
  
=== Приёмочная проверка (10:54:01)  
    прогоняю запись через все стадии — это занимает несколько минут  
Приёмочная проверка установки  
  
— Служба  
  [  ок  ] приложение отвечает  
  [  ок  ] база и брокер здоровы  
           версия 0.1.0, сборка 2026-08-26-a1b2c3d, схема f4527964b45c  
  
— Запись насквозь (smoke-sample.mp3, 412 КБ)  
  [  ок  ] запись принята  
           pending:      2.9 с  
           converting:   4.1 с  
           transcribing: 41.7 с  
           summarizing: 28.3 с  
           итого 78.6 с  
  [  ок  ] обработка завершилась  
  
— Ассистент  
  [  ок  ] ассистент ответил  
=====
```

Приёмка пройдена за 96 с.
приёмка пройдена

```
=== Открытие внешнего входа (10:55:40)  
    внешний вход открыт на 0.0.0.0 только после успешной приёмки  
  
=== ГОТОВО (10:55:44)  
  
Версия:      2026-08-26-a1b2c3d  
Веб-интерфейс: https://10.0.0.15/  
               (сертификат самоподписанный — браузер предупредит один раз)  
Первый вход:  admin / admin — пароль сменить сразу  
Каталог:     /home/admin/insight  
Модели:      /home/admin/insight/models  
  
Состояние:   cd /home/admin/insight && docker compose ps  
Журналы:     cd /home/admin/insight && docker compose logs -f api
```

Рис. 5. Приёмочный прогон записи и итоговое сообщение «ГОТОВО». Адреса, версии и времена стадий на вашем сервере будут свои.

Действие	Ориентировочное время
Установка из поставки с готовыми образами	20–30 минут
Первый старт служб распознавания	3–5 минут (модели грузятся в память)
Обновление	около 10 минут

5. Проверка работоспособности

Всё перечисленное установщик уже выполнил сам. Команды нужны, чтобы повторить проверку позже — после правки конфигурации или перезагрузки сервера:

```
cd ~/insight && sudo docker compose ps      # все контейнеры Up / Up (healthy)
curl -k https://localhost/health            # {"status":"healthy", ...}
python3 ~/insight-kit/smoke.py --dir ~/insight --audio ~/образец.mp3
```

Ответ `/health` сообщает и версию сборки (`build`), и версию структуры базы (`schema_revision`) — сверить два сервера можно с любой машины, не заходя на них. Значение `status: degraded` означает, что приложение отвечает, но база данных или брокер нездоровы.

6. Первый вход

Откройте в браузере `https://<IP-адрес сервера>/`. Самоподписанный сертификат вызовет предупреждение браузера — это ожидаемо; для постоянной эксплуатации замените его на сертификат удостоверяющего центра заказчика (§8).

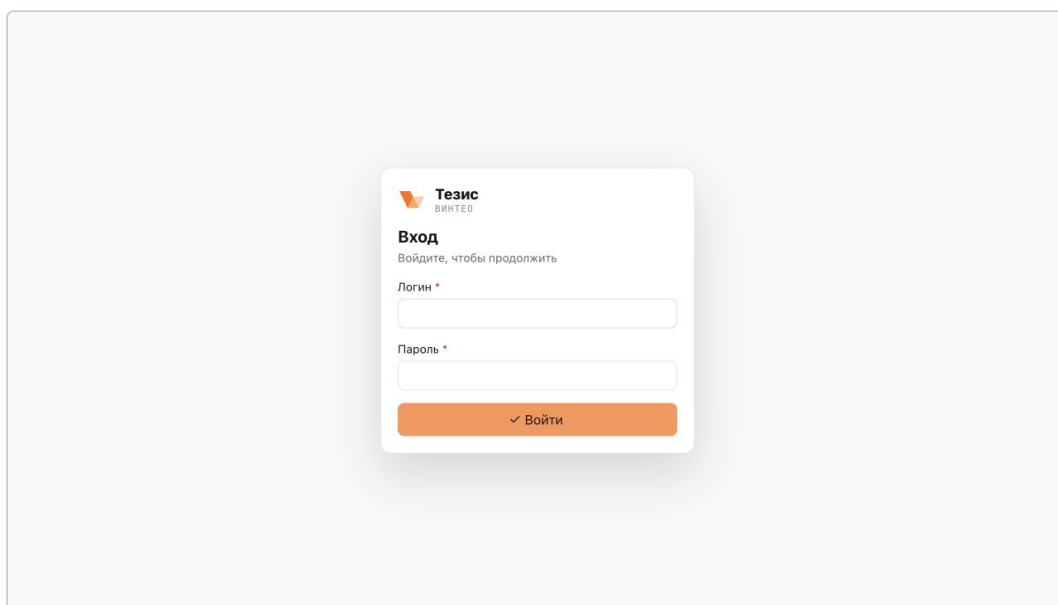


Рис. 6. Страница входа. Учётные данные при первом входе — **admin / admin**.

Пароль учётной записи admin необходимо сменить сразу после первого входа (меню «Пользователи» → профиль).

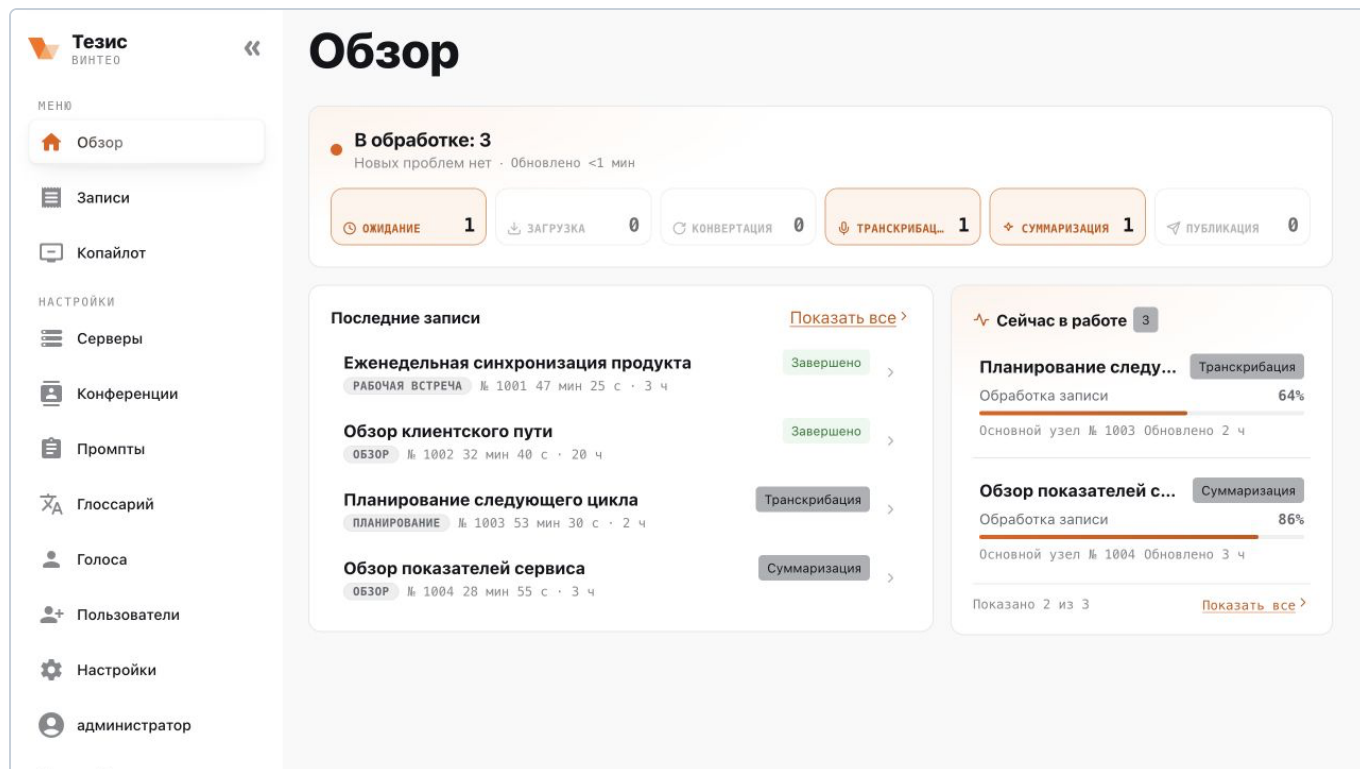


Рис. 7. Главная страница после входа: состояние обработки и последние записи.

7. Первичная настройка

1. **Смена пароля** `admin` — обязательный первый шаг.
2. **Серверы ВКС** — добавьте адреса и ключи доступа серверов Vinteo, с которых забираются записи. После добавления опрос идёт автоматически по расписанию.
3. **Пользователи и права** — учётные записи сотрудников, рабочие пространства и области видимости записей.
4. **Шаблоны протоколов** — при необходимости скорректируйте тексты запросов к языковой модели под типы совещаний.
5. **Голосовые профили** — при включённой идентификации говорящих.

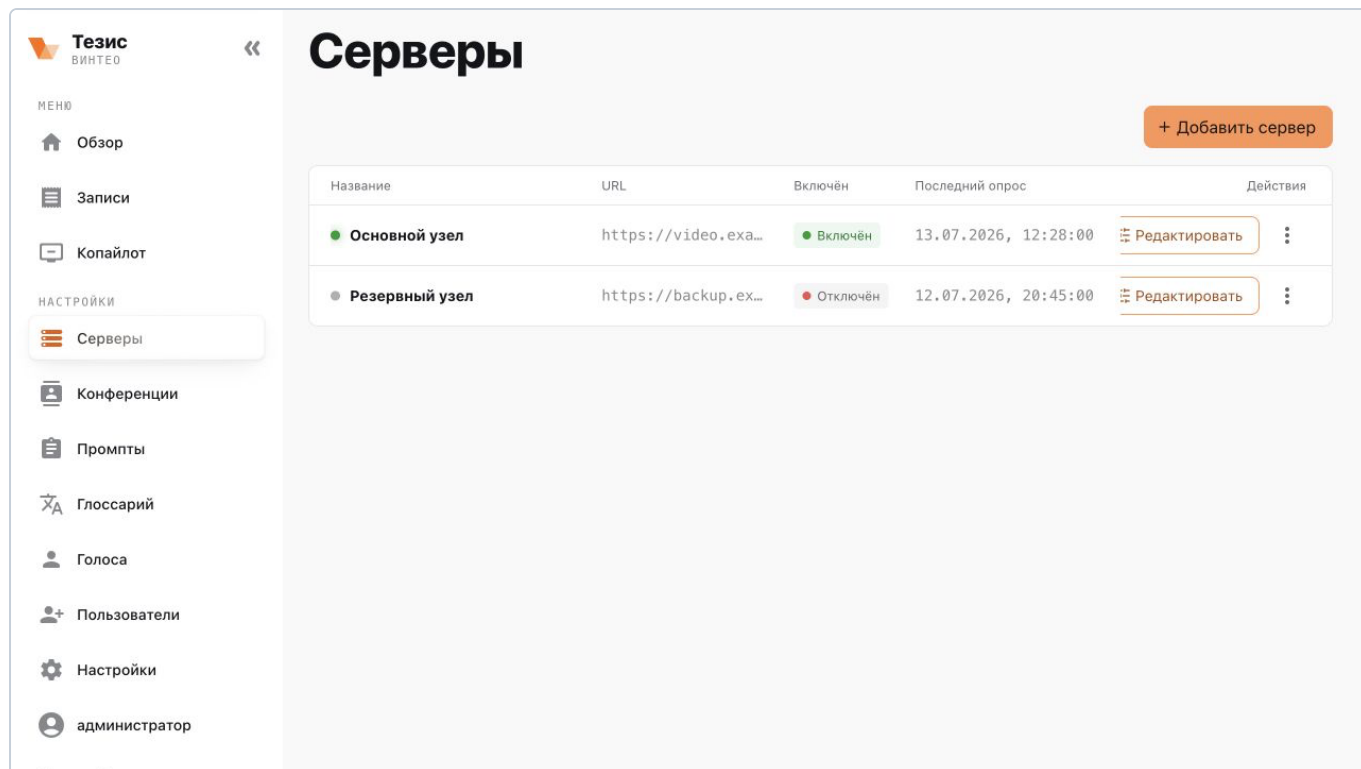


Рис. 8. Раздел «Серверы»: подключённые серверы Vinteo и их состояние.

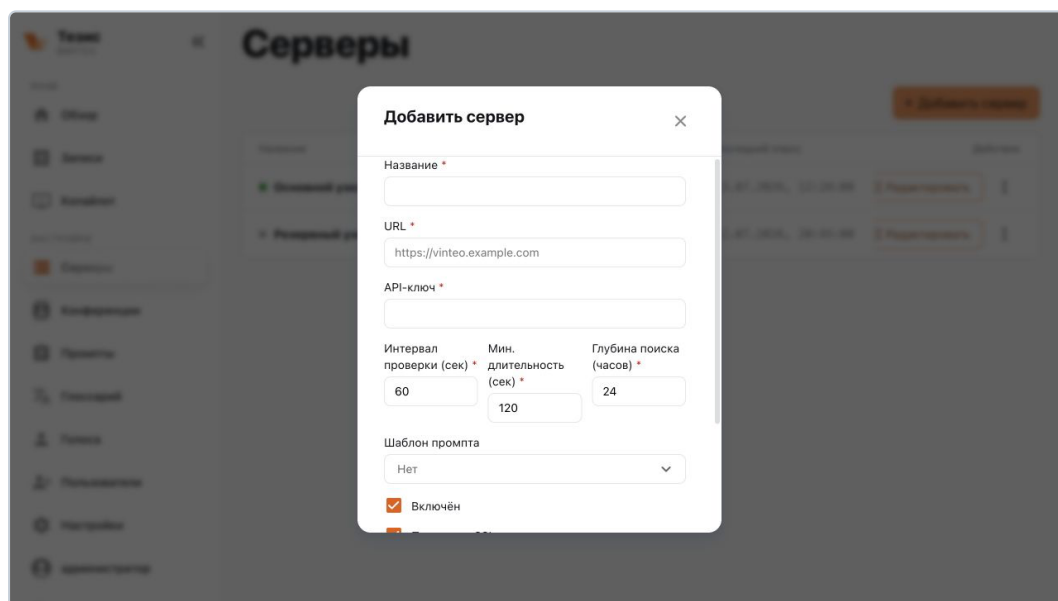


Рис. 9. Добавление сервера Vinteo: адрес, ключ API и период опроса.

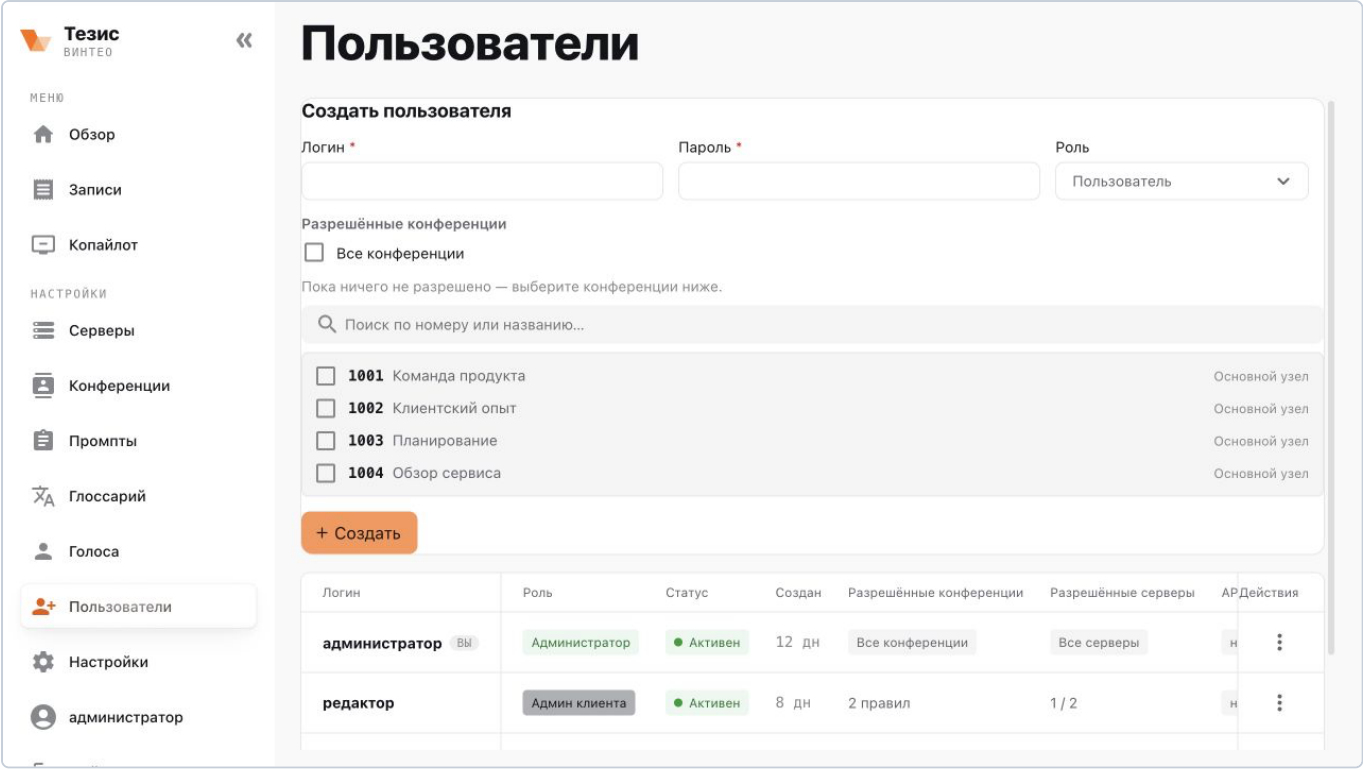


Рис. 10. Раздел «Пользователи»: учётные записи и права доступа.

8. Эксплуатация

Действие	Команда (из каталога <code>~/insight</code>)
Запуск	<code>sudo docker compose up -d</code>
Остановка (данные сохраняются)	<code>sudo docker compose down</code>
Состояние	<code>sudo docker compose ps</code>
Журнал API	<code>sudo docker compose logs -f api</code>
Журнал обработки	<code>sudo docker compose logs -f worker-transcription worker-summarization</code>
Применение изменений <code>.env</code>	<code>sudo docker compose up -d --force-recreate</code>

Команда `docker compose down -v` удаляет тома вместе со всеми данными: записями, расшифровками, протоколами и голосовыми профилями. В штатной эксплуатации она не применяется.

Резервное копирование

```
sudo docker exec insight-postgres pg_dump -U insight insight \
| gzip > insight-$(date +%F).sql.gz
```

Замена сертификата

```
sudo cp <сертификат>.crt traefik/certs/server.crt
sudo cp <ключ>.key      traefik/certs/server.key
sudo chmod 600 traefik/certs/server.key
sudo docker compose restart traefik
```

Обновление

Обновление выполняется тем же скриптом и той же командой, что и установка, — из каталога новой поставки. Скрипт сам определит, что перед ним рабочая установка: снимет точку отката и резервную копию базы, получит новые образы без остановки службы, выполнит обновление с коротким простоем (~5 минут), проверит новую версию за закрытым входом и откроет внешний доступ только после успешной приёмки. При любой ошибке до этого момента — автоматический возврат к прежней версии.

```
INSIGHT_KIT_PUBLIC_KEY=/etc/insight/kit-signing-public.pem \
insight-kit-run <путь-к-носителю>/insight-kit
```

Автозапуск после перезагрузки сервера обеспечивают включённая в systemd служба Docker и политика `restart: unless-stopped` у всех контейнеров; если сервис останавливали вручную (`docker compose down`), после перезагрузки нужен явный `docker compose up -d` . Подробности: полное руководство по установке, руководство пользователя и справочник по конфигурации — в комплекте документации поставки.