Cognitum AI Platform

0.16.0 → **0.16.1** для администратора · 13 октября 2026

Новый сервис и одна кнопка проверки

Инструменты «Офис» и «Сайты» выполняет `nocode-tools` — из образа nocode-harness, через NATS. Долгие вызовы MCP становятся задачами, а «Самопроверка» за две минуты показывает, что сломалось и кого это затронуло.



nocode-tools
Office, DuckDB, Chromium

Самопроверка
30+ проверок, отчёт JSON

Задачи MCP
Ожидание в фоне, TTL

Права
mcp.manage, storages.manage

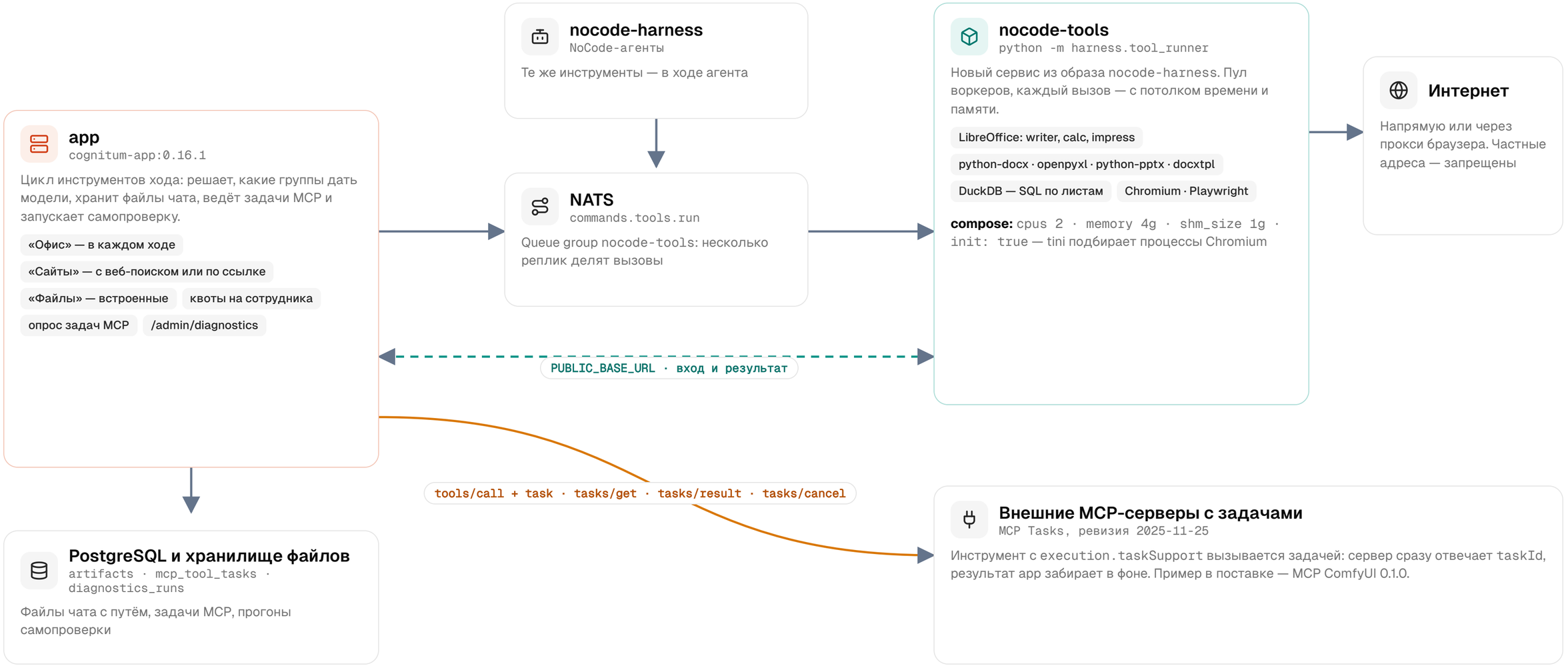
2 CPU · 4 ГБ
потолок `nocode-tools` под нагрузкой

30+
проверок в «Самопроверке»

3
лёгкие миграции, простой — около минуты

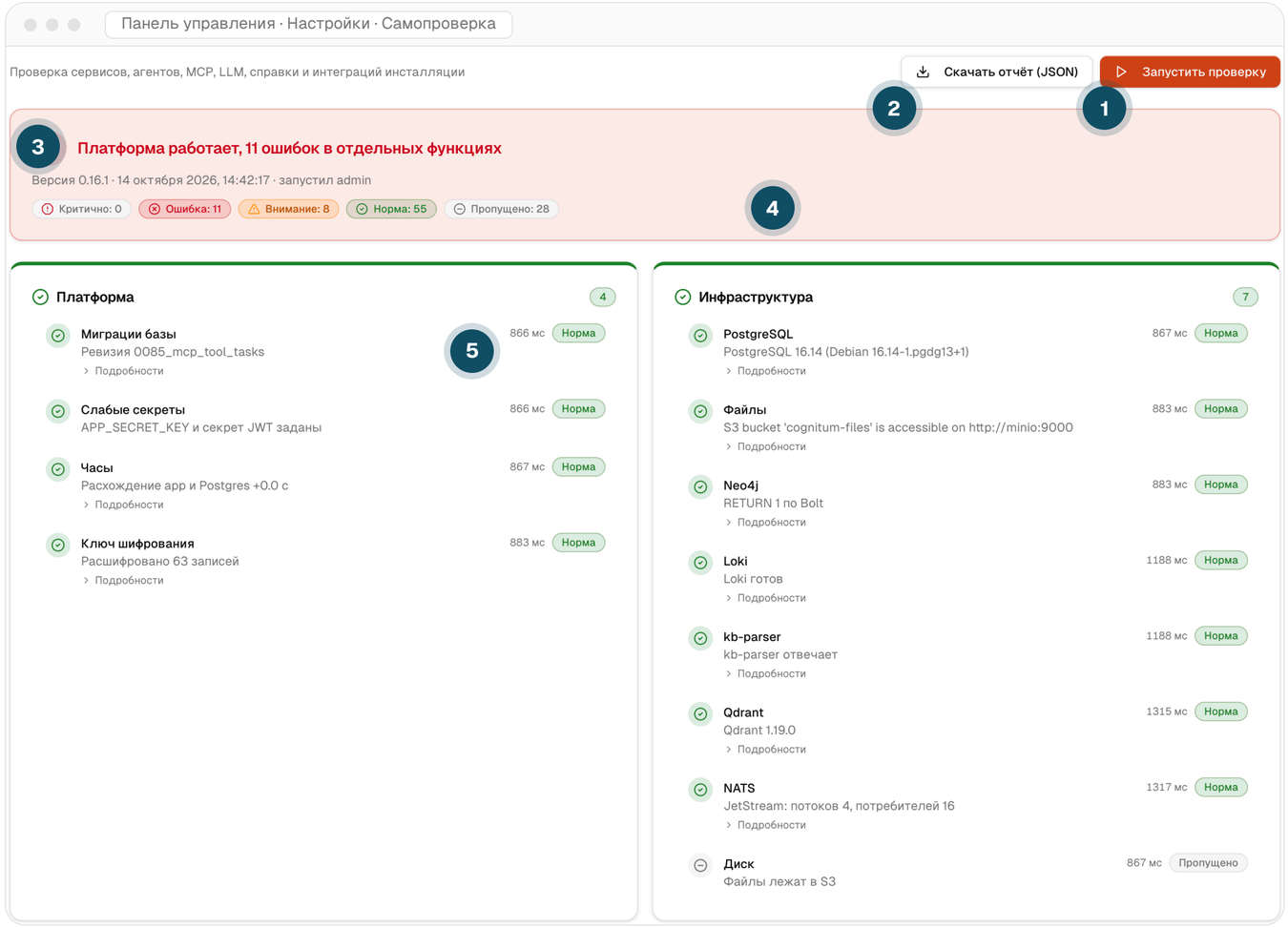
Инструменты хода — в отдельном сервисе через NATS

app вызывает «Офис» и «Сайты» по шине; файлы сервис забирает и возвращает через публичный адрес платформы. Тяжёлые LibreOffice и Chromium не живут в app.



Инсталляция целиком — за две минуты

«Администрирование → Настройки → Самопроверка» (/admin/diagnostics). Проверки идут параллельно, итог — словами и по статусам.



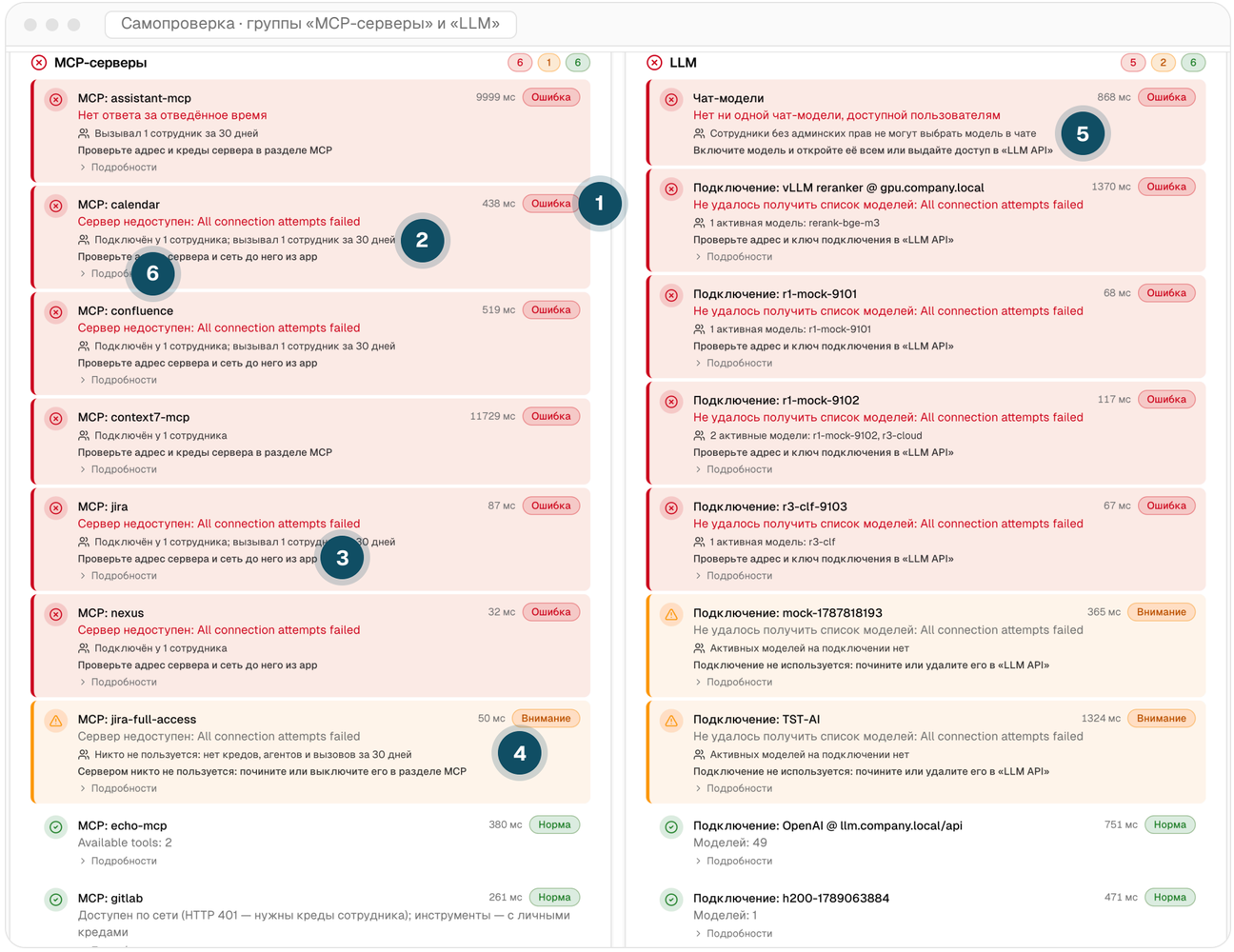
1. **«Запустить проверку»**
   Право `settings.manage`; прогоны хранятся — `0084_diagnostics_runs`
2. **«Скачать отчёт (JSON)»**
   Без паролей и ключей — к обращению в поддержку
3. **Итог и версия**
   Версия берётся из `APP_VERSION`
4. **Счётчики статусов**
   Сколько строк в каждом статусе
5. **Миграции базы**
   Норма — ревизия `0085_mcp_tool_tasks`

Критично	Платформа не работает целиком: миграции, ключ шифрования, PostgreSQL, NATS, Qdrant, хранилище
Ошибка	Не работает часть функций — что проверить, написано в строке
Внимание	Работает, но есть риск или деградация
Норма	Проверка прошла
Пропущено	Компонент не включён — это норма

Скриншот со стенда разработки: ошибки — от выключенных тестовых серверов; имя администратора заменено

Каждая строка: что сломалось, кого затронуло, что делать

Важность считается по использованию: сервер, у которого есть пользователи, — «Ошибка», тот же сбой у неиспользуемого — «Внимание».



1 Статус и время
Сколько шла проверка

2 Кого затронуло
Подключения и вызовы за 30 дней

3 Что сделать
Куда смотреть: адрес, сеть из app, креды

4 Никто не пользуется
Сбой — только «Внимание»

5 Последствие для людей
Не «список пуст», а кто что не может

6 «Подробности»
Сырые данные проверки

АГЕНТЫ И СЕРВИСЫ — ПОСЛЕ ОБНОВЛЕНИЯ

echo-agent: echo-job — Внимание
echo-agent выполнил задачу, но синхронный вызов не дождался ответа: Job execution timeout after 20 seconds
Повторите проверку; если повторяется — смотрите нагрузку app и NATS

nocode-harness — 867 мс — Норма
Heartbeat 0 с назад
Подробности

nocode-tools — 1187 мс — Норма
nocode-tools отвечает

echo-agent: heartbeat — Норма
Heartbeat 0 с назад
Подробности

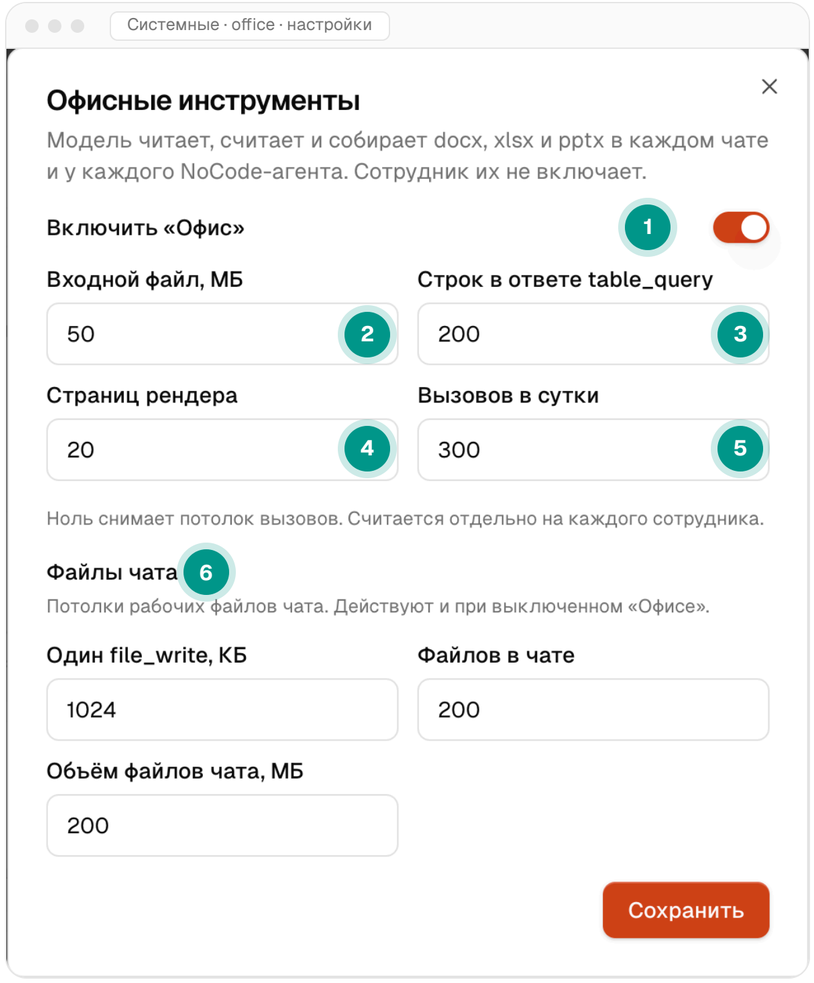
echo-job «не дождался ответа» — известное
Задача выполнена, ответ потерян в гонке синхронного вызова (BUG-28). Повторите проверку

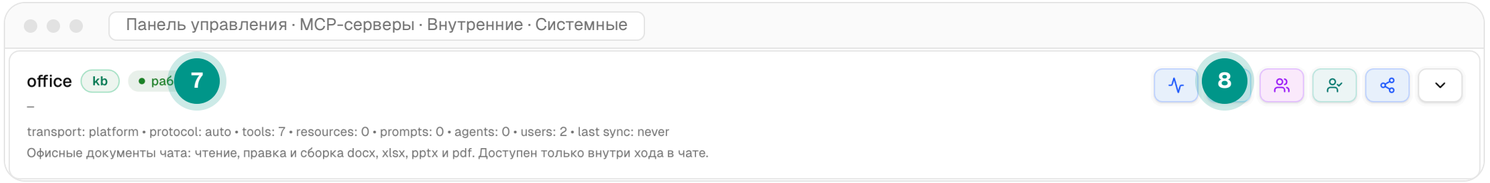
Панель управления · «Самопроверка»

Скриншоты со стенда разработки; адреса хостов заменены

«Офис» — системный сервер с настройками в админке

«Сервисы → MCP-серверы → Внутренние → Системные»: `office`, `web-reader` и `workspace-files` доступны только внутри хода. Сотрудник их не включает.





1. **Выключатель группы** — Выключили — инструментов нет ни в чате, ни у агентов
2. **Размер входного файла** — Больше — модель получит отказ с причиной
3. **Строк в ответе `table_query`** — Полный результат — файлом
4. **Страниц рендера** — Потолок `office_render`
5. **Квота в сутки** — На сотрудника; 0 — без потолка
6. **Файлы чата** — Действуют и при выключенном «Офисе»
7. **Статус и число tools** — `office` — 7 инструментов
8. **Настройки группы** — Тот же диалог, что слева

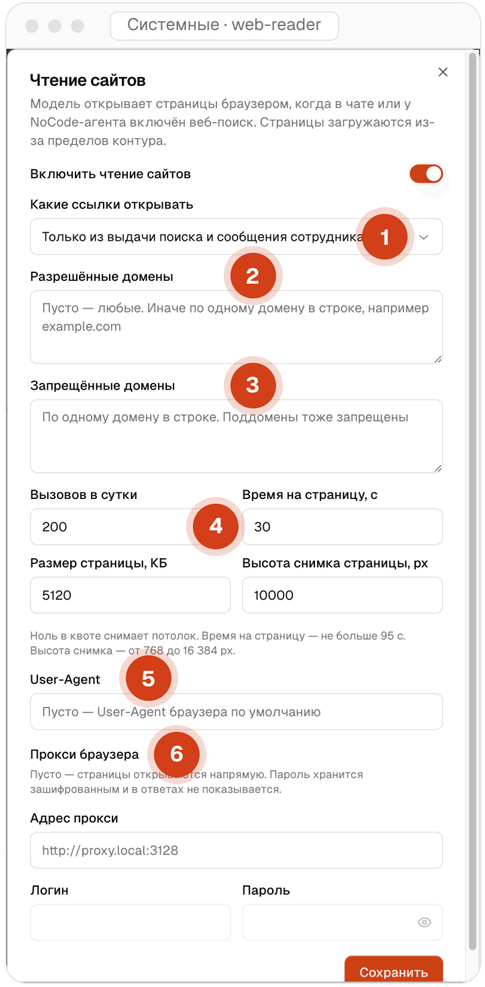
ПЕРЕМЕННАЯ .ENV	ПО УМОЛЧАНИЮ	ЧТО ЗАДАЁТ
`TOOLS_WORKERS` · `TOOLS_WORKER_MAX_TASKS` · `TOOLS_WORKER_MEM_MB`	2 · 50 · 1536	воркеры nocode-tools, их перезапуск и память
`TOOLS_OFFICE_TIMEOUT_S` · `TOOLS_MAX_TIMEOUT_S`	180 · 300	время вызова «Офиса» и потолок сервиса
`TOOLS_MAX_INPUT_MB` · `TOOLS_MAX_UNZIPPED_MB`	50 · 300	вход и его распакованный объём
`TOOLS_OFFICE_DAILY_QUOTA` · `TOOLS_MAX_OUTPUT_FILES`	300 · 20	квота по умолчанию, выходов за вызов

Панель управления · MCP-серверы · «Офисные инструменты»

Скриншоты со стенда; значения — по умолчанию; значения из админки важнее .env

Что браузер платформы может открыть

Страницы загружаются из-за пределов контура. Администратор решает, какие ссылки открывать, куда нельзя, сколько ждать и через какой прокси ходить.



1 Какие ссылки открывать
Только из выдачи поиска и сообщения сотрудника — или любые публичные

2 Разрешённые домены
Пусто — любые; иначе только перечисленные

3 Запрещённые домены
Поддомены тоже запрещены

4 Квота и потолки
Вызовов в сутки, время (≤ 95 с), размер, высота снимка 768–16 384 px

5 User-Agent
Пусто — как у браузера

6 Прокси браузера
Пароль хранится зашифрованным и в ответах не возвращается

Только публичные адреса

Частные, loopback и link-local IP закрыты. Проверяется каждый запрос браузера — и переадресации, и ресурсы страницы.

Когда «Сайты» в ходе

Вместе с веб-поиском или когда в реплике сотрудника есть ссылка. В прямом и агентном режимах и у NoCode-агентов.

Что остаётся в чате

Полный текст — `web/…md`, снимок — `screenshots/…png`. Модели — текст до `TOOLS_MODEL_RESULT_CHARS`.

ПЕРЕМЕННАЯ .ENV	ПО УМОЛЧАНИЮ	ЧТО ЗАДАЁТ
TOOLS_WEB_READER_TIMEOUT_S	120	время вызова группы «Сайты» целиком
TOOLS_WEB_READER_DAILY_QUOTA	200	квота на сотрудника, пока её не задали в админке
TOOLS_MODEL_RESULT_CHARS	8000	сколько текста результата получает модель

Панель управления · MCP-серверы · «Чтение сайтов»

Скриншот со стенда; адрес прокси — пример из подсказки поля

Потолки рабочих файлов и картинок для модели

У файла чата теперь есть путь (`0083_artifact_path`). Квоты общие для файлов модели, выходов «Офиса» и «Сайтов» и картинок из ответов MCP.

ФАЙЛЫ ЧАТА · АДМИНКА ВАЖНЕЕ .ENV

ПЕРЕМЕННАЯ .ENV	ПО УМОЛЧАНИЮ	ЧТО ЗАДАЁТ
`THREAD_FILES_WRITE_MAX_KB`	1024	текст одного `file_write`
`THREAD_FILES_MAX_COUNT`	200	файлов в одном чате
`THREAD_FILES_MAX_TOTAL_MB`	200	объём файлов чата
`THREAD_FILES_READ_MAX_CHARS`	12000	символов за один `file_read`; дальше — со смещением

Источник файла	кто может писать
`attachment` вложение сотрудника	только чтение
`model` файлы, которые записала модель	модель
`tool` выходы «Офиса» и «Сайтов»	модель
`mcp:{сервер}` картинка из ответа внешнего MCP	выход

КАРТИНКИ ДЛЯ МУЛЬТИМОДАЛЬНОЙ МОДЕЛИ

ПЕРЕМЕННАЯ .ENV	ПО УМОЛЧАНИЮ	ЧТО ЗАДАЁТ
`THREAD_VISION_MAX_IMAGES`	~~4~~ 12	картинок в одном запросе
`THREAD_VISION_TOKEN_RESERVE_PER_IMAGE`	~~1024~~ 2048	токенов окна на картинку
`THREAD_VISION_MAX_IMAGE_MB` · `_TOTAL_MB`	8 · 16	одна картинка и все в запросе

Если значения заданы в .env — пересмотрите
В 0.16.1 умолчания подняты: страница 1600×1200 у qwen стоит около 1900 токенов

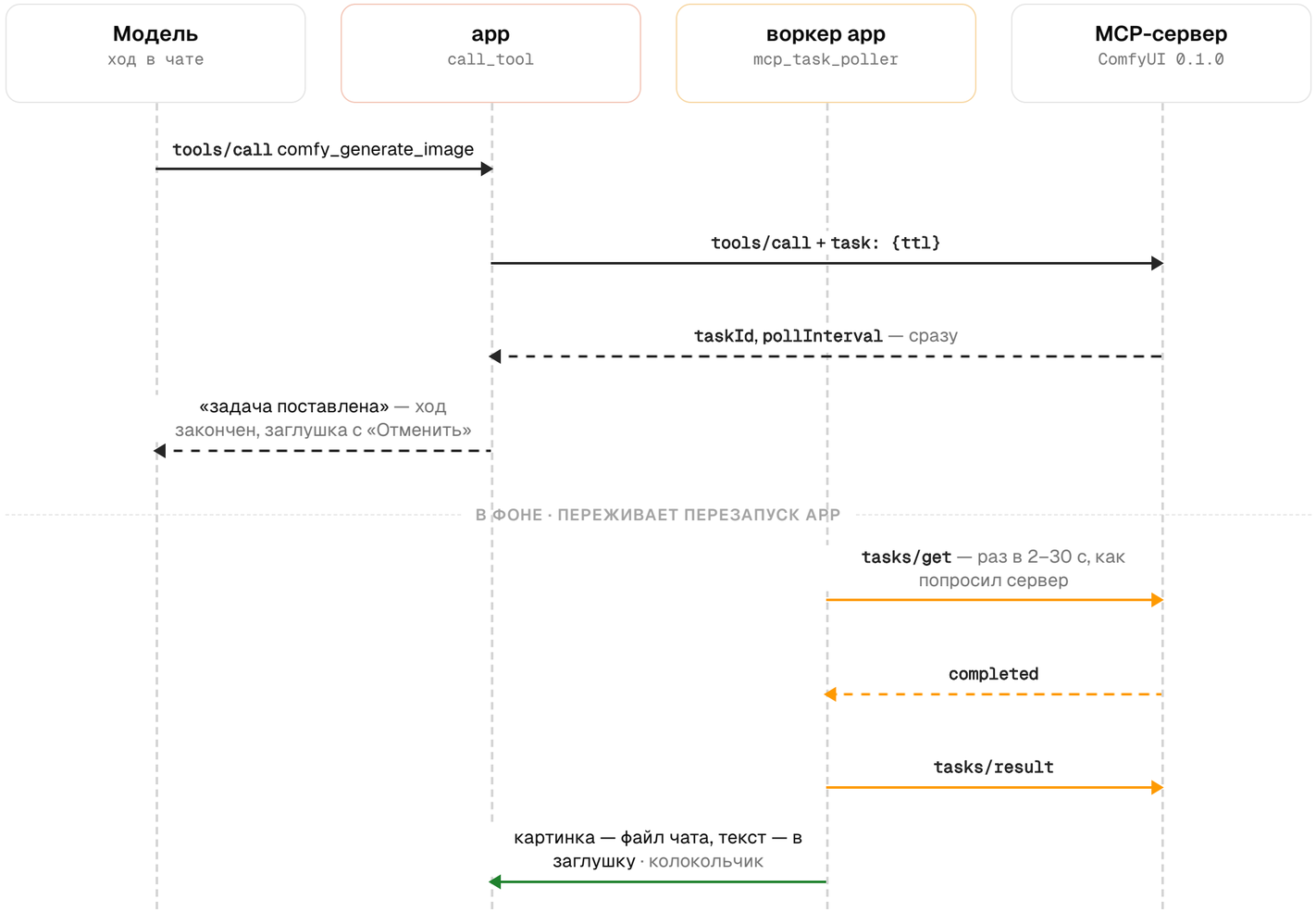
Какие картинки уходят модели
Обязательно — из последней реплики и последнего круга инструментов. Старые — пока запрос не больше половины окна, остальные заменяются текстом с путём

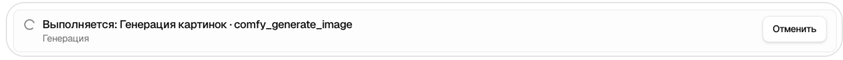
Копия, а не оригинал
Модель получает копию не больше 2048 px по длинной стороне; в чате остаётся исходный файл

Удаление чата
Файлы удаляются из хранилища, если больше нигде не используются; задачи чата отменяются

Долгий инструмент — строка в базе, а не открытый запрос

Если сервер объявил задачи, а у инструмента есть `execution.taskSupport`, app вызывает его задачей и сразу отпускает ход. Опрос ведёт фоновый воркер по таблице `mcp_tool_tasks`.





`MCP_TASK_TTL_S=900`
Не закончил за 15 минут — задача снимается, в заглушке «истёк срок»

«Отменить»
Заглушка сразу — «Отменено», `tasks/cancel` отправляет воркер

Сбой опроса
Сеть или 5xx — повтор через 15 с, до TTL

Колокольчик
Только если чат не смотрели последние 30 с

Ограничения
Stateful-сервер после своего перезапуска задачу не найдёт. Каналы и автоматизации ждут ответ целиком — там задач нет

Схема — макет по коду 0.16.1; кадр заглушки — со стенда

Управление — по праву, шлюз — с учёткой владельца токена

Раньше проверить, синхронизировать или вызвать инструмент MCP-сервера из админки мог любой вошедший. Теперь нужны отдельные права в ролях.

ПРАВО	ЧТО ОТКРЫВАЕТ	БЫЛО
`mcp.manage`	Создание и правка MCP-серверов, проверка подключения и синхронизация, включение инструментов, вызов инструмента из админки	любой вошедший
`storages.manage`	Хранилища файлов: создание, правка, удаление, проверка подключения, доступ агентов	любой вошедший
`settings.manage`	«Самопроверка» (новый раздел), а также настройки «Офиса», «Сайтов» и веб-поиска	—

Перед обновлением
Проверьте роли: кто управлял MCP-серверами и хранилищами без этих прав, после обновления получит отказ

Сотрудникам ничего не меняется
Подключать серверы к себе и вызывать инструменты в чате — как раньше

ШЛЮЗ MCP · /API/V1/MCP/*

```
POST /api/v1/mcp/tools/call
Authorization: Bearer «API-токен сотрудника»
project: SALES   # «Поле сотрудника» project
{"server_id": "jira", "tool": "search_issues", …}

200 — вызов с личной учёткой владельца токена
403 MCP_NEEDS_AUTH — учётки нет: подключите сервер
```

1. **Учётка — владельца токена**
   Шлюз берёт его подключение к серверу, как в чате
2. **«Поля сотрудника» — заголовками**
   Имя заголовка = имя поля; остальные заголовки серверу не уходят
3. **Понятный отказ**
   Без учётки — 403 с текстом, что сделать

ЧТО ЕСТЬ В ШЛЮЗЕ

GET /servers · GET /servers/{id}/tools · POST /tools/call · resources · resources/read · prompts · prompts/get · POST /servers/{id}/rpc — Streamable HTTP

0.16.0 → 0.16.1: простой — около минуты

Миграции только добавляют таблицы и колонки и применяются сами. Новые переменные необязательны — кроме APP_VERSION.

1. **Резервная копия базы и конфигурации**

```
pg_dump -Fc · cp .env docker-compose.yml $BACKUP_DIR/
```

2. **Запас на хосте и сеть**

`nocode-tools` — до 2 CPU и 4 ГБ; `PUBLIC_BASE_URL` открывается из контейнеров. Образ app — только amd64

3. **Новые `docker-compose.yml` и `deploy.sh` из тега v0.16.1**

В compose — сервис `nocode-tools` из `NOCODE_HARNESS_IMAGE`, без профиля

4. **В .env — `APP_VERSION=0.16.1` и теги образов**

app, echo-agent, nocode-harness, `KB_PARSER_IMAGE` (docling 0.16.1 — DOC, XLS, PPT)

5. **Запуск**

```
docker compose pull && docker compose up -d --force-recreate
```

6. **Миграции — сами**

`0083` → `0085`; ждать `Alembic migrations applied`, затем `/api/version` и `ps nocode-tools` — healthy

7. **«Самопроверка» и отчёт JSON**

Миграции, nocode-tools, nocode-harness, kb-parser, echo-agent — «Норма»

Известные предупреждения — не мешают работе

- **echo-agent: echo-job — «синхронный вызов не дождался ответа»**
  Задача выполнена, ответ потерян в гонке (BUG-28). Повторите проверку
- **Ошибки агентов — cognitum-echo-agent**
  Демо-агент пишет тестовые ERROR; выключается `enable_background_logs`
- **Индексация баз знаний — «документы с ошибкой»**
  DOC, XLS, PPT, упавшие до обновления, — «Переобработать»
- **«Все агенты» — SDK не 0.16**
  Процессные агенты работают; SDK обновите по плану

Откат

Вернуть прежние `.env` и `docker-compose.yml`, `docker compose pull`, пересоздать базу и `pg_restore` из копии. Данные, появившиеся после обновления, теряются.

Состав поставки

platform 0.16.1 · sdk 0.16.1 · nocode-harness 0.16.1 · echo-agent 0.16.1 · deploy 0.16.1 · docs 0.16.1 · cognitum-docling 0.16.1 · MCP ComfyUI 0.1.0