

ООО «ИТ СТРОЙИНТЕЛ»
ИНН 0272934208 · ОГРН 1250200037746

ИНСТРУКЦИЯ ПО УСТАНОВКЕ И РАЗВЁРТЫВАНИЮ

Программа для ЭВМ «СтройИнтел» (StroyIntel)
серверная часть · для системного администратора

Свидетельство о госрегистрации ПО № 2026611149 от 19.01.2026
Республика Башкортостан, 2026

Содержание

1. Назначение	3
2. Архитектура развёртывания	3
3. Требования к серверу	3
4. Состав системного программного обеспечения	3
5. Порядок развёртывания	3
5.1. Подготовка операционной системы	3
5.2. Развёртывание слоя данных и платформенных сервисов	4
5.3. Развёртывание слоя искусственного интеллекта	4
5.4. Развёртывание прикладных микросервисов	4
6. Сетевые порты сервисов	4
7. Проверка работоспособности	5
8. Сопровождение	5

1. Назначение

Документ описывает порядок развёртывания серверной части программного обеспечения «СтройИнтел» для системного администратора правообладателя. Платформа предоставляется конечным пользователям по модели SaaS; установка клиентского ПО не требуется.

2. Архитектура развёртывания

Серверная часть размещается на выделенном сервере под управлением российской операционной системы и состоит из трёх взаимодействующих слоёв:

- прикладной слой — микросервисы платформы (Node.js) под управлением менеджера процессов PM2 и сервис обработки моделей (Python);
- слой искусственного интеллекта — сервер инференса языковой модели (vLLM) и сервисы поиска по знаниям;
- слой данных и платформенных сервисов — реляционная СУБД, хранилище файлов, сервисы аутентификации и доступа (контейнеры).

3. Требования к серверу

Компонент	Требование (эталонная конфигурация)
Операционная система	ALT Server 11.1 (Mendelevium), ядро Linux 6.12 — ОС из реестра российского ПО
Процессор	не менее 8 физических ядер (AMD Ryzen 7 7700X)
Оперативная память	не менее 64 ГБ
Графический ускоритель	не менее 32 ГБ видеопамати (NVIDIA RTX 5090) — для локального инференса
Дисковая подсистема	SSD/NVMe не менее 100 ГБ

4. Состав системного программного обеспечения

Компонент	Версия
Среда исполнения JavaScript	Node.js 24.14
Интерпретатор Python	Python 3.12.7
Контейнеризация	Docker 29.2
Менеджер процессов	PM2 6.0
Сервер инференса ИИ	vLLM 0.19.1
Языковая модель	Qwen3.6-27B (INT4 AutoRound), лицензия Apache 2.0

5. Порядок развёртывания

5.1. Подготовка операционной системы

1. Установите ОС ALT Server 11.1 и примените обновления.
2. Установите среды исполнения Node.js 24 и Python 3.12, систему контейнеризации Docker и менеджер процессов PM2.

3. Подготовьте драйверы графического ускорителя для инференса языковой модели.

5.2. Развёртывание слоя данных и платформенных сервисов

Запустите в контейнерах инфраструктурные сервисы платформы:

Назначение	Компонент	Роль
СУБД платформы (PostgreSQL)	supabase/postgres 15.8	Хранение данных платформы и арендаторов
Сервис аутентификации	supabase/gotrue	Аутентификация и управление сессиями
REST-доступ к данным	postgrest	Программный доступ к БД по REST
Сервис реального времени	supabase/realtime	Подписки на изменения данных
Файловое хранилище	supabase/storage	Хранение загруженных файлов и документов
API-шлюз платформы данных	kong	Маршрутизация запросов к платформенным сервисам
Обработка изображений	imgproxy	Преобразование изображений
Пул соединений	supavisor	Управление пулом подключений к СУБД
Векторное хранилище	Qdrant v1.13.4	Хранение векторных представлений для поиска
Сервис LightRAG	LightRAG v1.5.1	Индексация и поиск по нормативной базе
СУБД для RAG	PostgreSQL 16.6	Хранение графа знаний и индексов RAG
Автоматизация процессов	n8n	Внутренние сценарии автоматизации

5.3. Развёртывание слоя искусственного интеллекта

1. Запустите сервер инференса vLLM с локально размещённой моделью Qwen3.6-27B.
2. Запустите сервисы поиска по знаниям и эмбеддингов (RAG, LightRAG, bge-m3), векторное хранилище и сервис долговременной памяти.

5.4. Развёртывание прикладных микросервисов

1. Запустите микросервисы платформы под управлением PM2.
2. Запустите сервис обработки моделей (BIM).
3. Проверьте, что все сервисы перешли в рабочее состояние.

6. Сетевые порты сервисов

Перечень основных сервисов и используемых внутренних портов. Служебные порты доступны только в пределах сервера; внешний доступ пользователей осуществляется исключительно через единый API-шлюз по защищённому соединению.

Сервис	Порт
API-шлюз (единая точка входа, JWT, изоляция арендаторов)	3010

Сервис	Порт
Базовый сервис (проекты, задачи, команда, склад, аналитика)	3011
Сервис реального времени (BFF, GraphQL/WebSocket)	3012
Сервис документооборота	3013
Сервис уведомлений	3014
Сервис экспорта	3015
Сервис мессенджеров (каналы)	3016
Сервис исполнительной документации (ИПСД)	3017
Сервис контроля качества	3018
Сервис чата с агентами	3019
Сервис аудита (журналирование)	3020
Сервис сметных расчётов	3022
Сервис BIM/IFC	3202
Сервер инференса ИИ (vLLM)	30000
Оркестратор ИИ-агентов	8001
Сервис базы знаний (RAG)	8002
Голосовые функции (STT/TTS)	8005
Сервис эмбедингов (bge-m3)	8090
LightRAG	9621
Долговременная память (mem0)	8020
Векторное хранилище (Qdrant)	6335

7. Проверка работоспособности

После развёртывания убедитесь в работоспособности:

- все прикладные микросервисы находятся в рабочем состоянии;
- сервер инференса ИИ отвечает на запросы;
- платформенные контейнеры запущены без ошибок;
- веб-интерфейс открывается через API-шлюз по защищённому соединению.

8. Сопровождение

Обновления и сопровождение серверной части выполняются правообладателем. Резервное копирование данных производится на стороне правообладателя на инфраструктуре в Российской Федерации.