

**ООО «ИТ СТРОЙИНТЕЛ»**  
ИНН 0272934208 · ОГРН 1250200037746

# **ОПИСАНИЕ АРХИТЕКТУРЫ И ТЕХНОЛОГИЧЕСКОГО СТЕКА**

Программа для ЭВМ «СтройИнтел» (StroyIntel)

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

## Содержание

1. Назначение .....	3
2. Общая архитектура .....	3
3. Прикладной слой .....	3
4. Слой искусственного интеллекта .....	3
5. Слой данных и платформенные сервисы .....	4
6. Мультиарендность и изоляция данных .....	4
7. Безопасность .....	4
8. Технологический стек (сводно) .....	4
9. Размещение .....	5

## 1. Назначение

Документ описывает архитектуру и технологический стек программного обеспечения «СтройИнтел».

## 2. Общая архитектура

Платформа построена по сервис-ориентированному принципу и состоит из трёх слоёв: прикладного (микросервисы), слоя искусственного интеллекта и слоя данных с платформенными сервисами. Единая точка входа — API-шлюз, который выполняет аутентификацию, маршрутизацию запросов и изоляцию данных по арендаторам.

## 3. Прикладной слой

Прикладная логика реализована набором микросервисов (Node.js), управляемых менеджером процессов PM2, и сервисом обработки информационных моделей (Python). Каждый микросервис отвечает за свою предметную область: проекты, документы, сметы, контроль качества, исполнительную документацию, уведомления, экспорт, аудит и другие.

Сервис	Назначение
API-шлюз	Единая точка входа, JWT-аутентификация, изоляция арендаторов
Базовый сервис	Проекты, задачи, команда, подрядчики, склад, аналитика
Сервис документооборота	Хранение и обработка документов, анализ проектов
Сервис сметных расчётов	Сметы на базе ГЭСН/ФСНБ-2022
Сервис BIM/IFC	Обработка и анализ информационных моделей
Сервис контроля качества	Дефекты, инспекции, журналы, предписания
Сервис ИПСД	Исполнительная документация, нормативные справочники
Сервис аудита	Журналирование действий и событий безопасности
Сервис чата	Взаимодействие пользователя с ИИ-агентами

## 4. Слой искусственного интеллекта

Слой ИИ построен вокруг локально размещённой большой языковой модели и инструментов поиска по знаниям:

- сервер инференса vLLM с моделью Qwen3.6-27B (квантизация INT4, лицензия Apache 2.0), развёрнутой на сервере в РФ;
- оркестратор интеллектуальных агентов, маршрутизирующий запросы между 18 специализированными агентами;
- сервисы поиска по нормативной базе (RAG, LightRAG) и эмбеддингов (bge-m3);
- векторное хранилище (Qdrant) и сервис долговременной памяти (mem0);
- модуль компьютерного зрения для анализа чертежей.

Модель работает локально (self-hosted), без обращения к зарубежным сервисам в процессе работы.

## 5. Слой данных и платформенные сервисы

Слой данных реализован на наборе контейнеризованных сервисов: реляционная СУБД PostgreSQL, файловое хранилище, сервисы аутентификации, доступа к данным и реального времени.

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

## 6. Мультиарендность и изоляция данных

Платформа является мультиарендной. Данные каждой организации изолированы. Изоляция реализована на уровне API-шлюза: при каждом запросе из токена доступа извлекается идентификатор арендатора, который принудительно применяется при выборке данных. Прямой доступ к прикладным сервисам в обход шлюза на сервере закрыт.

## 7. Безопасность

- аутентификация по токенам (JWT) с проверкой на уровне шлюза;
- изоляция данных арендаторов на уровне шлюза;
- ограничение доступа к служебным сервисам пределами сервера;
- журналирование действий и событий безопасности;
- обработка и хранение данных на территории Российской Федерации.

## 8. Технологический стек (сводно)

Компонент	Технология / версия
Операционная система	ALT Server 11.1, ядро Linux 6.12 (ОС из реестра РФ)
Серверная среда	Node.js 24.14, Python 3.12.7
СУБД	PostgreSQL 15.8
Контейнеризация / процессы	Docker 29.2, PM2 6.0
Инференс ИИ	vLLM 0.19.1
Языковая модель	Qwen3.6-27B (INT4 AutoRound), Apache 2.0
Поиск по знаниям	RAG, LightRAG v1.5.1, bge-m3, Qdrant v1.13.4, mem0
Обработка моделей	IfcOpenShell, ezdxf
Клиент	Веб-приложение (SPA), доступ по HTTPS

## 9. Размещение

Серверная часть, данные и языковая модель размещаются на инфраструктуре в Российской Федерации. Платформа функционирует на российской операционной системе ALT Server.