

Арсен Кенесбаев

Data Engineer | Python & SQL

- +7 707 446 6793
- Казахстан, Алматы
- arsen801777@gmail.com
- [Telegram](#)



О себе

Data Engineer со специализацией на SQL, Python и автоматизации. Имею опыт проектирования RAG-экосистем, разработки нормализованных банковских БД и построения ETL-пайплайнов. Владею инструментами контейнеризации (Docker) и оркестрации (n8n). Автор интерактивного портфолио, демонстрирующего навыки full-cycle разработки данных.

Навыки

Core Data Engineering: SQL, Python, PostgreSQL, Oracle Database, ETL/ELT, PL/SQL

Data Processing & AI: Pandas, NumPy, Qdrant, PGVector, RAG-системы, Scikit-learn

Инфраструктура и Автоматизация: n8n, Docker, Linux, Zabbix, Grafana, Git/GitHub

Инструменты и Web: Flask, FastAPI, waitress, Ngrok, Selenium, HTML/CSS/JavaScript

Проекты

RAG-экосистема с интеграцией в браузер

09/2025 - 11/2025

Tech Stack: Python, Qdrant (Vector DB), Docker, n8n, JS, Google Drive API, JSON/MIME.

- Разработал браузерное расширение для внедрения RAG-контекста в Google AI Studio; реализовал сложную валидацию MIME-типов и парсинг JSON, чтобы API распознавал внедренные данные как нативные файлы.
- Внедрил рекурсивную суммаризацию (chain summarization) и триггер «Thought Cleaning» для удаления внутренних логов (CoT), что снизило потребление токенов на ~70% и предотвратило переполнение контекстного окна.
- Построил локальный пайплайн данных: команды в Telegram запускают загрузку данных из Google Drive в Docker-контейнер с Qdrant через кастомный скрипт автоматизации.

Веб-сайт системы рекомендаций фильмов

11/2024 - 12/2024

Tech Stack: Flask, TMDB API, Scikit-learn, TF-IDF, Cosine Similarity, Matplotlib, waitress, Ngrok

- Создал систему рекомендаций фильмов на основе контентной фильтрации (TF-IDF + косинусное сходство) и ассистента в стиле Актера.
- Обработал датасет из 5,000+ фильмов TMDB, внедрил адаптивную фильтрацию и интегрировал командное портфолио.
- Развернул приложение через waitress + Ngrok, реализовав аутентификацию пользователей и переключение тем (Dark/Light mode).

База данных онлайн-банкинга (Oracle APEX)

10/2023 - 12/2023

Tech Stack: Oracle APEX, SQL, Java, Miro, Canva

- Спроектировал базу данных (в стиле Kaspi) с нормализованными схемами для клиентов, карт, кредитов и транзакций.
- Написал 30+ оптимизированных запросов (joins, подзапросы, оконные функции), добавил триггеры и индексы для ускорения работы.
- Генерировал тестовые данные с помощью Mockaroo и Java; визуализировал структуру и презентовал проект в Miro и Canva.

Образование

Сулейман Демирель Университет(SDU)

Каскелен, Казахстан

Бакалавр, Компьютерные науки

09/2022 - 05/2026

Сертификаты:



Google Data Analytics([ссылка](#))



PostgreSQL([Michigan University](#))



HCIA Datacom - Huawei



Selenium на Python - Stepik



Python Backend - Skylearn([TechOrda](#))

Языки:

- Казахский (родной)
- Русский (свободное владение)
- Английский (Advanced, ielts 7.5)

Опыт работы

KazNetCom

Алматы, Казахстан

08/2025 - 9/2025

DevOps-инженер (стажировка)

- Развернул систему видеонаблюдения Frigate в Docker, перенеся обработку видео на GPU (NVIDIA CUDA) для максимальной эффективности сервера.
- Создал решение для мониторинга на стеке Zabbix/Grafana с real-time дашбордами для контроля состояния ЦОД, задержек и доступности сервисов.
- Интегрировал Zabbix с Telegram для создания автоматизированной системы оповещений о сбоях и восстановлении работы базовых станций.
- Разработал кастомную GIS-платформу с уникальной интеграцией draw.io, позволяющей инженерам редактировать схемы сети прямо на интерактивной карте.

Агентство РК по регулированию и развитию финансового рынка

Алматы, Казахстан

06/2025 - 08/2025

Инженер по автоматизации и данным (стажировка)

- Спроектировал и внедрил две end-to-end системы на n8n: конвейер генерации отчетов (GLPI/Bitrix24) и RAG-систему для AI-ассистента.
- Разработал полный цикл RAG: от сбора данных (веб-скрейпинг) до векторизации (Embeddings) и хранения в PostgreSQL/PGVector.
- Реализовал сложную интеграцию с GLPI и Bitrix24, включая реверс-инжиниринг API, обогащение данных и извлечение информации из .docx с помощью LLM (Ollama).
- Обеспечил отказоустойчивость асинхронных процессов, внедрив механизм внешних блокировок (distributed locks) в PostgreSQL для решения проблемы состояний гонки (race conditions).
- Спроектировал PULL-архитектуру для обхода сетевых ограничений, блокировавших стандартные PUSH-вебхуки.