ARSEN KENESBAYEV
Data Engineer | SQL & Python

- +7 707 446 6793
- Kazakhstan, Almaty
- arsen801777@gmail.com
- Telegram

Summary

Data Engineer specializing in SQL, Python, and Automated Workflows. Proven experience architecting RAG ecosystems, designing normalized banking databases, and building high-performance ETL pipelines. Skilled in optimizing infrastructure with Docker and n8n, and integrating Vector Databases for AI solutions. Creator of an interactive portfolio demonstrating end-to-end engineering capabilities.

Skills

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

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

Infrastructure & Automation: n8n, Docker, Linux, Zabbix, Grafana, Git/GitHub

Web & Tools: Flask, FastAPI, Waitress, Ngrok, Selenium, HTML/CSS/JavaScript

Projects

Browser-Integrated RAG Ecosystem & Infrastructure
09/2025 - 11/2025

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

- Engineered a browser extension to inject RAG context into Google AI Studio; handled complex MIME-type validation and JSON parsing to trick the API into recognizing injected data as native files.
- Implemented recursive chain summarization and a "Thought Cleaning" trigger that strips internal CoT logs, reducing token usage by ~70% and preventing context window overflow.
- Built a local data pipeline where Telegram commands trigger data loading from Google Drive into a Dockerized Qdrant instance via a custom automation script.

Movie recommendation system website
11/2024 - 12/2024

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

- Built a movie recommender with content-based filtering (TF-IDF + cosine similarity) and an Akinator-style assistant.
- Processed 5,000+ TMDB movies, included adaptive filtering, and integrated a team portfolio.
- Deployed via Waitress + Ngrok with user authentication and dark/light mode.

Online Bank database using Apex Oracle SQL
10/2023 - 12/2023

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

- Designed a Kaspi-style database with normalized schemas for customers, cards, loans, and transactions.
- Wrote 30+ optimized queries (joins, subqueries, window funcs) and added triggers/indexes for speed.
- Generated mock data via Mockaroo + Java; built and presented using Miro & Canva.

Education

Suleyman Demirel University
Kazakhstan, Kaskelen
09/2022 - 05/2026

Bachelor's degree, Computer Science

Certifications:		Languages:	
	Google Data Analytics(link)		Kazakh(Native)
	PostgreSQL(Michigan University)		Russian(Fluent)
	HCIA Datacom - Huawei		English(Advanced, IELTS 7.5)
	Selenium on Python - Stepik		
	Python Backend - Skylearn(TechOrda)		

Experience

KazNetCom
Almaty, Kazakhstan
08/2025 - 09/2025

DevOps Engineer (Internship)

- Managed a Dockerized Frigate NVR system on Linux, implementing NVIDIA CUDA hardware acceleration to optimize server performance by minimizing CPU load.
- Built a Zabbix & Grafana monitoring stack with custom dashboards for real-time tracking of network infrastructure health and critical service availability.
- Automated network outage notifications by integrating Zabbix with Telegram for instant alerts on base station status changes.
- Engineered a custom GIS platform (OpenStreetMap/Yandex Maps) with an integrated draw.io editor for interactive visualization and management of network diagrams.

Financial Regulatory Agency of Kazakhstan
Almaty, Kazakhstan
06/2025 - 08/2025

Automation & Data Engineer (Internship)

- Designed and implemented two end-to-end systems on the n8n platform: a data pipeline for automated report generation (integrating GLPI/Bitrix24) and a RAG system for an AI assistant.
- Developed a full RAG cycle, from data ingestion (web scraping) and text vectorization (Embeddings) to storage and retrieval from a PostgreSQL/PGVector database.
- Executed a complex integration with GLPI and Bitrix24, which involved reverse-engineering APIs, data enrichment, and extracting structured information from .docx files using an LLM (Ollama).
- Engineered fault-tolerant asynchronous workflows by implementing a distributed locking mechanism in PostgreSQL to resolve race conditions.
- Architected and implemented a PULL-based data retrieval model to bypass network security restrictions that blocked standard PUSH-based webhooks.