System Architecture

Source & Intake Layer

Job postings are discovered manually from high-signal job boards and company career pages. A local Streamlit application serves as an intake interface, allowing users to paste job descriptions and minimal metadata (URL, company, title, location, source).

Parsing & Normalisation Layer

Python parsing logic processes the pasted job descriptions to infer seniority, work mode, role family, and extract technical and soft skills using a curated skill dictionary and aliases. Raw descriptions are preserved to allow reprocessing as extraction logic improves.

Data Layer

PostgreSQL stores raw job data, normalised entities, and analytical outputs. The relational model supports aggregation, filtering, and trend analysis across roles, locations, and companies.

Application & Delivery Layer

A FastAPI service exposes insight-driven endpoints, while a Streamlit dashboard enables interactive exploration and live demonstrations. Generated reports translate analytical results into business-facing outputs.

Observability & Reliability

Basic logging, validation, and data quality checks ensure ingestion reliability and transparency without introducing unnecessary infrastructure complexity.