

DAVID LUKIC

Senior Data Engineer | AI/ML Engineer | Cloud Solutions Architect

✉ ldavid797@gmail.com | 📞 +381 65 616 0655 | [linkedin.com/in/davidlukic99](https://www.linkedin.com/in/davidlukic99) | davidlukic99.github.io

PROFESSIONAL SUMMARY

Senior Data Engineer with 5+ years delivering production-grade, cloud-native data platforms and AI-powered solutions. Top Rated Upwork freelancer with 100% Job Success Score. Proven track record architecting end-to-end data ecosystems—from real-time ETL pipelines processing millions of records daily to LLM-powered applications with RAG architectures and MCP integrations. Expert in modern data stack (Airflow, dbt, Spark, Kafka) and AI frameworks (LangChain, LlamaIndex, Vector DBs). Built knowtheprice.com.au from scratch as founder—full-stack product demonstrating entrepreneurial execution. Recognized for delivering 60-70% efficiency gains and quantifiable cost reductions through intelligent automation and data-driven insights.

Core Differentiators: End-to-end AI product development | Production MLOps & LLM deployment | Modern lakehouse architecture (Databricks, Delta Lake) | Real-time streaming (Kafka, Spark) | Cross-functional stakeholder management

PROFESSIONAL EXPERIENCE

Data Engineer / Full-Stack Developer

2022 - Current

BetterCollective | Remote/Hybrid

- Engineered 14 production Django applications integrated with Plotly Dash, eliminating 80% of manual workflows and streamlining operations for 50+ internal users
- Designed and deployed LogAna, an automated log analysis platform with real-time dashboards, achieving 400% improvement in processing speed and 40% increase in data accuracy
- Built high-throughput data pipelines using pandas/Polars processing millions of daily records with 99.9% uptime, supporting critical business intelligence operations
- Developed robust ETL workflows integrating multiple data sources (REST APIs, AWS S3, PostgreSQL, DuckDB) with optimized storage and sub-second query performance
- Implemented enterprise-grade asynchronous web scraping services with proxy rotation, rate limiting, and error handling, scaling to handle large datasets while maintaining system stability
- Managed AWS-hosted PostgreSQL infrastructure implementing Redis caching and strategic indexing for fast, reliable data access patterns

Senior Data Engineer

2024 - 2025

index.dev | Remote

- Architected scalable ETL pipelines using Pydantic models and dbt orchestrated via Airflow, achieving 30% reduction in processing time and 20% boost in operational efficiency for enterprise clients
- Optimized analytical workloads on Postgres and DuckDB through query tuning and indexing strategies, accelerating decision-making by 20% for cross-functional teams
- Enhanced Snowflake Data Warehouse implementation with performance optimization and cost reduction initiatives, improving query response times for high-volume analytics
- Built 10+ reproducible containerized development environments using Docker and Poetry, reducing onboarding time by 40% and standardizing best practices across teams
- Delivered interactive Power BI and Plotly Dash dashboards with advanced features (cross-filtering, callbacks, drill-downs) transforming complex datasets into executive-level insights
- Implemented comprehensive testing (pytest), type safety (mypy/pyright), and code quality frameworks (ruff/black) ensuring production-ready, maintainable codebases

Senior Data Engineer & AI Solutions Architect

2020 - Current

Self-Employed | Remote (Top Rated Upwork Freelancer)

- Founded and built knowtheprice.com.au—end-to-end product featuring real-time price tracking, deployed on Vercel with Supabase backend, demonstrating full-stack and product ownership capabilities
- Architected and deployed production LLM applications using LangChain, LlamaIndex, and RAG pipelines with Pinecone/PGVector, delivering intelligent document Q&A and semantic search solutions
- Built Streamlit and Dash analytics dashboards for global clients, improving data-driven decision-making and saving 20+ hours weekly through accessible visualizations and interactive reports
- Designed end-to-end ETL pipelines for lead generation and marketing analytics, improving data accuracy by 60% and cutting processing time by 70% through intelligent automation
- Implemented monitoring and alerting systems reducing manual oversight by 99%, enabling proactive issue resolution and continuous pipeline reliability
- Delivered custom AI-powered automation workflows integrating GPT-4, Claude, and open-source LLMs, increasing client operational efficiency by 60% across diverse industries

TECHNICAL SKILLS

Programming & Backend

Python (Expert), SQL (Expert), Bash, Django, FastAPI, Flask, Pydantic

AI & GenAI Stack

LangChain, LangGraph, LlamaIndex, MCP (Model Context Protocol), OpenAI API, Anthropic Claude, RAG, Vector DBs (Pinecone, Weaviate, PGVector), Prompt Engineering, DSPy

Data Engineering

Apache Spark (PySpark), Airflow, Dagster, dbt, Kafka, Delta Lake, ETL/ELT, Data Modeling, Data Warehousing

Cloud & Infrastructure

AWS (SageMaker, Lambda, S3, Redshift), Azure, GCP, Databricks, Snowflake, Vercel, Supabase

MLOps & DevOps

Docker, Kubernetes, MLflow, CI/CD (GitHub Actions), Poetry, pytest, mypy, ruff/black, LangFuse, LangSmith

Databases & Storage

PostgreSQL, DuckDB, MongoDB, MariaDB, Redis, S3, Delta Lake

Data Processing & Analytics

pandas, Polars, NumPy, Jupyter, Spark Streaming, Real-time Pipelines

Visualization & BI

Power BI, Plotly Dash, Streamlit, Matplotlib, Seaborn

Specialized Skills

Web Scraping, API Development, Celery Task Queues, Automation, Git, Linux/Unix

KEY PROJECTS & ACHIEVEMENTS

- knowtheprice.com.au:** Founded and developed full-stack price comparison platform using modern tech stack (Vercel, Supabase, Next.js) with real-time data pipelines and user authentication
- LogAna Platform:** Automated log analysis system processing 10M+ daily log entries with Plotly dashboards, reducing manual analysis from 8 hours to 30 minutes (400% improvement)
- RAG-Powered Document Intelligence:** Built production LLM application using LangChain + Pinecone enabling semantic search across 100K+ documents with sub-2-second response times
- Real-time Streaming Pipeline:** Architected Kafka-based event processing system handling 500K+ events/hour with Spark Structured Streaming and Delta Lake storage
- Lakehouse Architecture Migration:** Led Databricks Delta Lake implementation replacing legacy warehouse, reducing storage costs by 40% while improving query performance 3x

LEADERSHIP & SOFT SKILLS

- Stakeholder Management:** Translated complex AI/ML concepts to C-suite executives, securing \$500K+ in project funding and aligning technical roadmaps with business strategy
- Mentorship & Code Review:** Mentored 5+ junior engineers on best practices, implementing structured code review processes that improved team code quality by 35%
- AI Ethics & Governance:** Established GDPR-compliant data handling procedures and bias detection frameworks for responsible AI deployment in production systems
- Cross-functional Collaboration:** Partnered with Product, Marketing, and Operations teams to deliver data solutions driving measurable business outcomes across 20+ projects