



Oğuz Erdoğan

Senior Data & AI Engineer

Building AI, Data and Kubernetes Platforms at Scale

Çankaya, Ankara
+90 554 524 01 64
mail@oguzerdogan.com
www.oguzerdogan.com

ABOUT

Senior Data Engineer, **Google Cloud Professional Data Engineer & Cloud Architect** certified, building scalable **AI, data, and Kubernetes platforms at scale**.

I design and operate **LLM-based products, agentic RAG systems, and GitOps-driven lakehouse architectures** processing **TB-scale data workloads**.

My focus is **cost-efficient analytics, platform reliability, and production-grade engineering** across AI, data, and infrastructure.

EXPERIENCES

Nesine – Senior Data & AI Engineer

09/2023 –

Present

Data Engineering

- Designed and implemented an **end-to-end Medallion Lakehouse architecture** using **Spark, Iceberg, Trino, dbt, Airflow, and Cosmos**, fully automated with **GitOps practices**.
- Implemented **enterprise data governance and lineage** using **DataHub** and **Ranger**.
- Led data platform evolution **Ambari → Cloudera → Kubernetes**, owning architecture and migration strategy.
- Processed and optimized **TB-scale data workloads** for core analytics and business domains.
- Delivered **100% cost optimization on GA4** through complete data model redesign and query optimization with **BigQuery**

Data Stack: Spark · Iceberg · Flink · Trino · dbt · Airflow · DataHub · Ranger · Cosmos · Kubernetes · GitOps · BigQuery

AI Engineering

- Deployed and optimized **local LLMs, embeddings, rerankers, and vision models** on **H100 GPU infrastructure**, following production best practices.
- Deployed and operated **OpenWebUI as a self-hosted local LLM platform**, enabling secure, internal model access for **300+ users**, including updates and ongoing system maintenance.
- Designed and implemented **self RAG** for customer support, including retrieval pipelines, tool routing, and model orchestration.
- Built **agentic QA solutions** to:
 - Generate test cases from **Jira & TestRail**.
 - Provide semantic search over existing test scenarios,
 - Execute agent-based test planning and validation workflows.
- Operated **Langfuse** for LLM observability, tracing, and evaluation.

AI Stack: OpenAI · LangChain · LangGraph · Langfuse · FastAPI · TypeScript · Grafana · Local LLMs · H100 GPU · OpenWebUI · RAG · Vector Retrieval

Platform Engineering

- Built and managed a **Kubernetes platform** hosting **20+ internal services** across data, AI, and analytics.
- Standardized deployments with **GitOps (ArgoCD) + Helm**, ensuring auditable, automated releases.
- Enforced secrets/security with **Kubeseal** and managed registry governance via **Harbor**.
- Implemented cluster-wide observability: **Grafana dashboards** and **ELK stack (Elasticsearch, Logstash, Kibana)** for logs and metrics.
- Managed **database backup, restore, and migration** operations for platform services.
- Enabled developer self-service using **Coder** (cloud-style workspaces).
- Deployed and operated key platform services on Kubernetes:

Platform Stack: Kubernetes · Helm · ArgoCD · Kubeseal · Harbor · Grafana · ELK · Coder · Airflow · Trino · Spark · Flink · Iceberg · dbt · Superset · DataHub · Ranger · MLflow · OpenWebUI · Langfuse · MinIO

Oredata – Data Engineer

09/2022 – 09/2023

- Certified **Google Cloud Professional Data Engineer** and **Google Cloud Professional Cloud Architect**.
- Designed and implemented a **lakehouse architecture** on Google Cloud using **BigQuery, Airflow, dbt, and Cosmos**.
- Built scalable **data lake and warehouse solutions** supporting analytics and operational reporting.
- Developed **batch and streaming data pipelines** using managed GCP services.
- Collaborated with analytics and business teams to **design, model, and optimize analytical datasets**.
- Improved **data quality, documentation, and test coverage** for reliable data pipelines.
- Evaluated and integrated tools to improve platform performance and processing efficiency.

Tech Stack: Google Cloud (BigQuery, GCS, Dataflow, Pub/Sub, Cloud Functions, IAM) · Python · SQL · Airflow · dbt · Docker · DataHub · Linux

Miul – Data Engineer Team Lead

05 / 2022 – 09/2022

- Led the **Data Engineering learning path and project delivery**, mentoring engineers and guiding technical implementations.
- Designed and productionized **data ingestion pipelines** from APIs, event streams, and databases.
- Built and orchestrated data workflows using **Airflow, Kafka, and distributed processing tools**.
- Delivered hands-on **data engineering workshops and technical training**, supporting real project development.
- Provided architectural guidance and code reviews, aligning implementations with scalability and data reliability goals.

Tech Stack: SQL, Python, GCP, Airflow, Docker, Linux, Hive, Sqoop, Kafka, PySpark, Git

Missafir – Data Team Lead

11/2021 – 07/2022

- Led a **5-person data team**, driving analytics, automation, and pricing tech initiatives.
- Owned **dynamic pricing and revenue optimization** for **500+ properties**, powering business-critical decisions.
- Designed and deployed **data pipelines and automation workflows** for pricing, monitoring, and competitive analysis.
- Built internal tooling to support **real-time insights, performance tracking, and alerting**.
- Acted as the data partner for **Product, CRM, Sales, and Finance**, translating business needs into scalable data solutions.
- Standardized engineering practices including **code reviews, documentation, and pipeline reliability improvements**.

Missafir – Data Scientist

01/2021 – 11/2021

- Built and automated **data ingestion pipelines** using APIs, web scraping, and internal microservices.
- Designed and implemented a **Data Warehouse from scratch**, enabling centralized analytics and reporting.
- Developed **dynamic pricing models, customer segmentation, and forecasting solutions** to support revenue growth.
- Implemented **dashboarding and reporting systems** to drive data-informed decisions.
- Automated alerts and internal notifications (Slack, scheduled jobs, rule-based triggers).
- Deployed and orchestrated data jobs on **Linux servers and cloud environments**.

Tech Stack: SQL, Python, Linux, Docker, Git, DataStudio, Selenium, BeautifulSoup

Veri Bilimi Okulu – Mentor

09/2020 – 09/2022

- Mentoring over 200+ participants about data science, career path etc.
- Teaching data science topics, code reviews
- Writing articles about data science & data engineering
- Worked closely with Miul data science team to create and evaluate models for bespoke industry problems using a combination of machine learning.

EDUCATION

BSc, Statistics — Anadolu University (2019)

