An Than

thanthanhan9101@gmail.com
https://linkedin.com/in/thanthanhan
Berlin, Brandenburg

SUMMARY

Results-driven Senior Data Engineer with over 6 years of experience in designing and optimizing data flows for scalable machine learning systems. Demonstrated success in building incremental batch data processing, with a strong track record of delivering cost-effective ML solutions across AWS and GCP environments. Proven ability to enhance operational efficiency through data systems that contribute to significant cost reductions and improved performance.

SKILLS AND TECH

- **AI & Machine Learning:** Embedding, Transformers, LLM, RAG Architecture, Feature Engineering.
- **MLOps:** Hugging Face, Vertex AI, SageMaker, MLflow, Model Monitoring, Great Expectations.
- **Data Engineering:** Data Lake, Airflow, Dataflow, BigQuery, MySQL, Apache Hudi, Prometheus, Grafana, DBT.
- **Cloud & DevOps:** AWS, GCP, Terraform, GitHub Actions, Docker, Lambda/Cloud Function, Cost Optimization.
- **Soft skills:** Results-Oriented, Data Strategy, Leadership & Initiative, Problem-solving, Team Enablement.

EXPERIENCE

Just Eat Takeaway   **Senior Data Engineer**   **Oct 2022 — Present**

- Led design of data flow system for 4B products with 100M daily updates, implementing hash-key approach and incremental processing that reduced ML costs by 400x and extended full-load cycles from monthly to quarterly.
- Engineered incremental ML inference pipeline with country-specific flexibility and prevent redundant ML processing during upstream reloads, reducing costs by 91% and accelerating data delivery from weekly to daily.
- Designed denormalized item-level fact table for ML deliverables, replacing series of inefficient model tables, reducing downstream costs by 68% and enabling 60% immediate inference reuse.
- Led BigQuery optimization initiatives, achieving 30% cost reduction while improving query performance.
- Led initiatives to migrate ML pipelines from SageMaker/EMR Spark to VertexAI/Dataflow, consolidating platforms, reducing operational cost, eliminating cross-cloud data transfer costs.
- Transitioned to DBT for SQL ETL, improving development and testing efficiency by 40% over Airflow operator.
- Created LLM experimentation modules with LiteLLM, MLflow and LangFuse, enabling DS team to rapidly prototype classification and NER models with integrated prompt tracing and cost tracking for PoC deployment.
- Mentored for 10+ engineers on VertexAI and BigQuery analytics through Google incubator program.

Just Eat Takeaway   **Data Engineer**   **Jul 2020 — Oct 2022**

- Standardized AWS Terraform infrastructure with reusable modules ensuring consistency across environments, reducing provisioning time by 75%.
- Transitioned ML workloads from EC2 to SageMaker serverless, cutting deployment time by 70%.
- Implemented end-to-end MLOps with Terraform, containerized testing, and automated CI/CD via GitHub Actions.
- Led migration from MySQL to Apache Hudi, improving upserts and incremental processing for large-scale data.

Just Eat Takeaway   **Data Scientist**   **Apr 2019 — Jul 2020**

- Developed ML-driven alcohol detection system (94% accuracy), ensuring compliance across multiple regions.
- Designed structured food taxonomy, enhancing search accuracy and promotions.

Delivery Hero   **Data Scientist**   **Nov 2018 — Apr 2019**

- Built AI-driven food classification models, optimizing search and recommendation performance.

EDUCATION

EURECOM & Telecom ParisTech   **Jan 2016 — Jan 2018**

MSc - Data Science, GPA: 4.0/4.0

University of Science   **Jan 2012 — Jan 2016**

BSc - Computer Science, GPA: 3.8/4.0