Wickes Wong

AI/Data Engineer

linkedin.com/in/wickes-w | wickeswong@gmail.com | wix-doc.com | github.com/wickes1 | 6476360607

Summary

Experienced Data Engineer with over 2 years of dedicated experience in architecting and deploying secure and scalable data pipelines for warehouses and lakes. Proficient in ETL/ELT and cloud platforms, with a keen interest in advancing expertise in AI and LLMs. Currently exploring opportunities within startup environments to leverage emerging technologies and drive innovative data-driven solutions. Proven track record of collaborating effectively within cross-functional teams to deliver high-impact data solutions for informed decision-making.

Skills

Development: Python, JavaScript, TypeScript, Go, Git, Shell Scripting
Database: SQL, PostgreSQL, MySQL, MongoDB, Redis, Supabase, Elasticsearch, Milvus, PostGIS, pgvector
Data and Analytics: Pandas, Matplotlib, PySpark, Airflow, Kafka, Grafana, Tableau, Databricks, dbt
Machine Learning: LLM, RAG, LangChain, LlamaIndex, Transformer, Object Detection, OpenAI API, Chainlit, Ragas
Cloud and Infrastructure: AWS (S3, Lambda, EKS, ECS, Route53, RDS), Azure, GCP, Docker, Kubernetes, Terraform, Helm, Kustomize, Ansible, GitHub Actions, GitLab CI/CD, ArgoCD
Framework: FastAPI, Django, Flask, Next.js, React, Node.js, SQLAlchemy, Fiber, pytest
Concept: CI/CD, RESTful API, MVC, Microservices, OOP, Agile, Scrum, SDLC

Experience

Varadise Limited, Hong Kong | Data Engineer   June 2021 – May 2023

- Led the implementation of an event-driven message distribution system with Kafka and AWS EKS, enabling real-time and cross-channel event notifications. Achieved sub-second message consumption and production for high-performance event delivery. Reduced customization time for output channels and templates from weeks to minutes through Go Templates.
- Implemented scalable ETL pipelines with Python, Databricks, Airflow, and Kubernetes integrated with Kafka and AWS S3, achieving a 20% increase in efficiency in processing 30 GB of data daily. Ensured data quality controls using Pydantic, Great Expectations, and DataHub.
- Built serverless, event-driven ETL processes on AWS Lambda, triggered by events from Kafka and AWS Workmail. Facilitated real-time monitoring of workflows through AWS CloudWatch and Grafana dashboards.
- Developed scalable APIs with FastAPI and gRPC. Deployed APIs on Kubernetes using Helm, offering functionalities such as CRUD, gRPC bridge, OAuth2, and Webhook support.
- Developed internal tools with Streamlit and React-Admin, connecting to the internal database. Integrated LLM for data catalog generation. Accelerated Proof of Concept (PoC) delivery and reduced manual labor work.
- Created interactive dashboards using Grafana for monitoring Kubernetes clusters and deployment resources. Leveraged Kubecost for cost analysis and optimization, resulting in a 30% reduction in infrastructure costs.
- Implemented CI/CD pipelines using GitLab CI and ArgoCD, along with pre-commit, reducing deployment time from week to hours and enhancing Developer Experience (DX).
- Mentored junior data engineers for more than 6 months, fostering a collaborative team environment.
- Applied Agile and Scrum methodologies to streamline the application lifecycle through Jira.

CoolThink@JC, Hong Kong | Ambassador   Sep 2019 - Apr 2021

- Collaborated with educators to enrich young minds with interactive STEM education programs.
- Guided and inspired students through engaging lessons in Scratch and App Inventor for Android.

Projects

Hong Kong Canadians Hackathon@Toronto   github.com/wickes1/hk-can-hackathon-2024

- Led the winning team in designing a solution to address a real-life SME challenge with QuickBooks Desktop, helping the business owner potentially save $1k annually.
- Developed a Java Sync App with QuickBooks Desktop SDK to extract data and inject it into Postgres.
- Developed a responsive web application using Next.js featuring Progressive Web App (PWA).

LLM Exam ChatBot

- Integrated Retrieval-Augmented Generation (RAG) Chatbot with GPT-4, Azure AI Search, and LangChain. Delivered chatbot

API as part of MLOps.
- Conducted data cleansing and preprocessing on exam materials for improved retrieval.
- Optimized prompts for accurate addressing of syllabus-related inquiries.

GIS Dataset Ingestion
- Developed a data pipeline to ingest, transform, and publish 20GB+ geospatial data from internal and public Spatial Data Infrastructures (SDIs) into PostGIS and GeoServer. Enabled OGC Web Services for frontend applications.

Real-time Finger Vein Verification System   github.com/wickes1/FingerVeinApp
- Designed a deep learning-based recognition system using Python and PyTorch, compatible with the Futronic VS80 Finger Vein Scanner, and implemented an innovative enrollment strategy to enhance accuracy and efficiency.

Education

City University of Hong Kong   Sept 2017 – June 2021
BEng Computer and Data Engineering