

Yohan Mok

Data pipeline engineer focused on high-traffic real-time processing, database and distributed application optimization, data integrity, cost efficiency, maintainability, and automation.

flaxinger@gmail.com 010-9418-4370 GitHub LinkedIn

KEY EXPERTISE

- / Building and maintaining high-traffic, real-time data pipelines
- / Optimization of databases and distributed applications
- / Decision making based on data integrity, cost efficiency, and maintainability
- / Automation of repetitive and error-prone workloads

EXPERIENCE

AB180

Datapipeline Engineer / 2022.08 - Present

Development and maintenance of Airbridge's real-time data pipelines

2022.08 - 2026.04

- Owned high-volume stream processing systems for Airbridge.
- Operated more than 10 stream processors that handle about 3 billion events per day.
- Maintained attribution processing that decides which advertising event contributed most to a customer app conversion.
- Maintained postback processing that sends customer conversion events to advertising partners in real time.
- Improved data pipeline capabilities across device ID merging, fraud detection, currency conversion, and settings APIs.
- Implemented privacy-control and masking features by stakeholder and iOS tracking level.
- Built Google Ads conversion-result cross-validation reporting and Meta Install Referrer decoding for event processing.
- Implemented frequency-based fraud prevention rules using tumbling windows.
- Implemented Click Injection prevention based on device install time and event time, including related settings APIs.

Operation of ScyllaDB

2023.02 - 2026.04

- Operated self-managed ScyllaDB clusters on Kubernetes.
- Identified why materialized views had grown oversized, reduced disk usage by 30% (about 9TiB), and prevented ScyllaDB cluster scale-out.
- Handled cluster setup, dashboards, and scale-in/down operations.

AWS Seoul region expansion

2025.07 - 2025.12

- Built event pipelines and storage in the Seoul region for financial-sector customers.
- Handled pipeline implementation, ScyllaDB setup, pipeline load testing, and multi-region CI/CD.

ECS to EKS worker platform migration

2025.01 - 2025.03

- Moved the data pipeline execution environment from ECS to EKS.
- Improved scale-in/out speed, operational convenience, and cost visibility.

Cost optimization across data pipeline

2024.06 - 2025.12

- Converted an AMD-only worker cluster to an ARM/AMD multi-architecture setup, reducing EC2 server cost by 20%.
- Parallelized image builds and tests in CI/CD so build speed did not regress.
- Optimized Kafka-based pipeline CPU usage, reducing monthly cost by about \$6k.
- Tuned KEDA scalers so workers reacted quickly to traffic spikes without wasting CPU.

Service maintainability improvements

2022.08 - 2026.03

- Built integrated Grafana dashboards using CloudWatch, Prometheus, and New Relic GraphQL data.
- Managed New Relic ingestion volume with sampling and added Prometheus metrics to core components to improve cost and visibility.
- Built Elasticsearch and Snowflake dashboards so non-engineering teams could monitor key service metrics.
- Implemented multiple backoffice tools for customer-specific pipeline settings.
- Reduced CI time from about 5 minutes to about 1 minute with pytest parallelization and in-memory databases.
- Changed deployment workflow triggers from PR merge to comments to simplify environment deployment and rollback.
- Fully separated a development environment that previously shared production databases.
- Redefined the development and CI/CD workflow around isolated development and staging environments.

Implementation of a declarative QA system

2024.06 - 2024.09

- Designed and developed a declarative QA system for repeatable and predictable system testing.
- Implemented a system that generates combinations of declared events and validates their processing results.
- Improved internal real-time pipeline QA by replacing manually generated event scenarios with reproducible scenario definitions.
- Supported end-to-end assertions against real-time OpenSearch results.

Exem

Backend Engineer / 2020.08 - 2021.03

Development of CloudMoa

2020.08 - 2021.03

- Developed CloudMoa, a Kubernetes cluster monitoring SaaS.
- Managed PromQL queries used in product features.
- Maintained PromQL used to observe Kubernetes cluster status in the product.

PoC of Velero-based cluster backup

2020.08 - 2021.03

- Built a proof of concept for Velero-based Kubernetes cluster backup.
- Validated periodic snapshot storage for Kubernetes cluster resources.

PROJECTS

BaekjoonHub

[GitHub Repository](#)

GitHub

Created and maintained an open-source browser extension that pushes algorithm solutions to GitHub.

- Created and maintained BaekjoonHub, an open-source browser extension that pushes algorithm solutions to GitHub.
- Transferred project management rights in 2023.03.

EDUCATION

Yonsei University

2015.03 - 2022.02

Computer Science / International Commerce