FERDIAN IFKARSYAH

+62895 605184009 ⋄ Jakarta, Indonesia

ferdian.ifkarsyah@gmail.com ⋄ linkedin.com/in/ifkarsyah ⋄ ifkarsyah.github.io ⋄ github.com/Ifkarsyah

EXPERIENCES

StraitsX   Jakarta, Indonesia
Senior Engineer, Data Infrastructure   *May 2024 - Present*

- Replaced expensive batch **RDS** snapshot ingestion pipeline with **real-time CDC → BigQuery** across 800+ **PostgreSQL/MySQ**L tables, improving data latency while reducing monthly spend by $8,000.
- Upgraded **Airflow** from Composer 2 to Composer 3, modernizing underlying Python runtimes, and system packages while maintaining backward compatibility for more than +100 DAGs.
- Led cost observability and optimization initiative by implementing a monitoring dashboard and review cycles, reducing **BigQuery** slot usage by 527 hours/day and saving $1.5K/month by partitioning large tables.
- Deployed **Apache Spark** on GCP to enable **DBT Python models**, previously impossible due to BigQuery's SQL-only limitation—unlocking more flexible, scalable transformation logic for the data team.

Xendit   Remote
L4 Data Infrastructure Engineer   *Oct 2022 - March 2024*

- Achieved $7,400/month cost reduction by analyzing network traffic with **Datadog** and migrating **Trino** on **Kubernetes** from a multizone to an active-standby architecture.
- Improved clickstream analysis speed by migrating from batch to real-time processing with **Kafka**, implementing **JavaScript** event tracker, transforming events via **Spark Streaming**, and delivering insights to **Databricks**.
- Created **Databricks** notebook templates with integrated **MLflow** tracking, enabling data scientists to rapidly prototype and monitor experiments with standardized logging, metrics, and artifact storage.
- Created a custom Trino logging plugin in **Java** to track query ownership and identifying high-cost users.

L3 Data Engineer   *Oct 2021 - Sep 2022*

- Developed an extensible, config-driven ingestion framework using the Strategy pattern, orchestrated via **FastAPI** as the backend layer, enabling data consumers to trigger pipelines through simple **YAML** configurations. Integrated 10+ sources including **PostgreSQL**, **MongoDB**, **Jira**, **Zendesk**, **CleverTap**, and **Facebook Ads**.
- Reduced manual **COGS** calculation from 3 days to less than one hour and guaranteed rates auditability by implementing **Slowly Changing Dimension (SCD)** logic within scalable **DBT** models.
- Implemented Infrastructure as Code (**IaaC**) principles with **Terraform** to automate provisioning of production resources (e.g., Spark clusters, AWS IAM roles), ensuring auditability and environment parity.

L2 Junior Data Engineer   *Mar 2021 - Sep 2021*

- Consolidated 2,000+ lines of user-contributed **PySpark** code from **Databricks** notebooks into shared-library.
- Developed a CI/CD validation for **DBT** and **Looker** to enforce partitioning and prevent inefficient deployments.

EDUCATION

- **Bandung Institute of Technology**, B.S. in Computer Science   Aug 2017 - Oct 2021
  Lab Assistant: KU1102 - Introduction to Information Technology (2019)

ACHIEVEMENTS

- **[National] First Place, Tokopedia Hackathon DevCamp 2021** (4000+ contestants)
- **[Regional] Finalist (Top 50), ACM-ICPC Asia Jakarta Regional 2019**.

PROJECTS

- **Ifkarsyah's Garden** Digital garden on data infrastructure, platform, and backend systems.
- **Room Finder ITB** Developed a high-usage public tool (5000+ students) for navigating campus classes.