

Muhammad Iqbal Gozy — Curriculum Vitae

Data & Analytics Engineer

South Jakarta, Jakarta, Indonesia | iqbal.gozy@gmail.com (mailto:iqbal.gozy@gmail.com) | +62-812-5235-1606

GitHub (<https://github.com/qybbs>) | LinkedIn (<https://linkedin.com/in/iqbalgozy>) | Portfolio (<https://iqbalgozy.com>)

Summary

A data-driven Analytics & Data Engineer with proven expertise in building robust ETL pipelines, data warehouses, and Business Intelligence (BI) mechanisms. Specialized in SQL query optimization, cloud cost intelligence, and Python automation within the Google Cloud Platform (GCP) and AWS ecosystems. Track record of architecting incremental data transformations, orchestrating complex cross-platform synchronizations, and delivering actionable BI dashboards (Looker Studio) to empower business decision-making.

Professional Experience

Data Engineer

AstraPay | Jakarta, Indonesia (Contract via Berijalan) | Oct 2025 – Present *Contributing to the development of AstraPay's data infrastructure, specializing in modern data warehousing, automated ETL pipelines, data quality, and BI reporting.*

Observability & Cost Intelligence (BI): Developed an enterprise BigQuery cost monitoring and Looker Studio usage intelligence platform using GCP Cloud Logging Sinks, Information Schema metadata, and Google Colab APIs, providing actionable BI mechanisms that significantly reduced GCP compute scan expenditures.

Data Warehouse Optimization (SQLX): Architected highly optimized internal data warehouse transformations using **Dataform in Incremental Mode**. Wrote and optimized complex SQL queries to replace legacy full-refresh pipelines, eliminating redundant daily terabyte-scale scans.

Enterprise ETL Automation (Python/Java): Built automated ETL synchronization pipelines integrating BigQuery, GCS, Sharepoint, and MoEngage, successfully automating marketing user-segment syncs and eliminating manual spreadsheet processes (saving **24 hours** per sync cycle).

Resilient Data Pipelines: Designed and developed highly resilient data ingestion and Reverse ETL pipelines featuring watermark-driven delta loading, dynamic schema evolution, and polymorphic type conflict self-healing.

Tech Stack: Python, SQL, BigQuery, Dataform SQLX, Looker Studio, Google Colab Enterprise, Google Cloud Storage (GCS), Datastream, Spreadsheets.

Data Developer / Backend (Internship)

Berijalan Techno Center | Yogyakarta, Indonesia | Jun 2025 – Oct 2025 *Developed data-centric automation tools and advanced backend systems to streamline operations.*

AI Chatbot Pre-processing (Python): Improved query understanding and chatbot scoring accuracy by **40%** by building advanced Python text pre-processing pipelines, manipulating large text datasets to enhance user intent parsing.

Open-Source OCR Data Ingestion: Designed and built a custom document-text-extractor API (<https://github.com/qybbs/document-text-extractor>) using Python OCR libraries and Elasticsearch, creating an automated pipeline for document data ingestion and text searching.

Custom ERP Data Modeling: Refactored the default Chart of Accounts (COA) schema in ERPNext to expand financial tracking dimensions from 2 to 7 segments, handling complex relational data modeling.

Tech Stack: Python, SQL, Elasticsearch, Next.js, Java, ERPNext, Git.

Backend Developer (Internship)

GeekGarden Software House | Yogyakarta, Indonesia | Mar 2025 – Jun 2025 *Engineered core backend systems handling critical relational data for enterprise applications.*

Ministerial Procurement System: Architected REST APIs for a complex ministerial procurement platform, working heavily with PostgreSQL to manage relational data, transactional tender lifecycles, and large-scale document records.

Tech Stack: Node.js, PostgreSQL, SQL, REST API, Git.

Technical Skills

Data Engineering & Manipulation: Python, Advanced SQL, ETL/ELT Pipeline Development, Data Warehousing, Workflow Automation, Big Data Frameworks (Exposure to Spark/Hadoop concepts)

Databases: PostgreSQL, MySQL, SQL Server (Familiar), BigQuery, MongoDB, Elasticsearch

Cloud & BI Tools: Google Cloud Platform (GCP), AWS (Familiar), Looker Studio, Power BI (Familiar), Tableau (Familiar), Dataform, Spreadsheets

Software Engineering: Java, TypeScript, REST APIs, Git, Docker, CI/CD

Featured Projects

document-text-extractor API (Unstructured Data Pipeline)

Lead Developer | Jul 2025 – Present

Developed an asynchronous data extraction pipeline utilizing Python OCR engines (Tesseract/EasyOCR) and Elasticsearch indexing for large-scale document text retrieval.

Stack: Python, OCR, Elasticsearch, Docker, Git.

Education

Bachelor of Information Technology

Universitas Pembangunan Nasional "Veteran" Yogyakarta | Yogyakarta, Indonesia

Aug 2021 – Jul 2025 | **GPA: 3.77 / 4.00**

Certifications & Organizations

Bootcamp Graduate: Fullstack Developer - Bootcamp Berijalan Techno Center Batch 14

Leadership Role: Chairman, HIMATIF (Informatics Student Association) UPN "Veteran" Yogyakarta (Feb 2024 – Mar 2025)