

NAFIUL HADI SAPUTRA

Pati, Central Java, Indonesia | +62 813-2686-8657 | nafiulhadisaputraa@gmail.com

LinkedIn: [linkedin.com/in/nafiul-hadi-saputra-bb7037439](https://www.linkedin.com/in/nafiul-hadi-saputra-bb7037439) | GitHub: github.com/NafiulSaputra | Website: Nafiul.com

PROFESSIONAL SUMMARY

Data Engineer whose pipelines run on schedule, recover on their own, and raise a flag when they can't. I built a daily ELT pipeline (PostgreSQL, dbt, Airflow) that unifies 4 operational areas into 1 analytical warehouse for a metalworking manufacturer. Every run has 2x automatic retries, 30-minute task timeouts, and Telegram failure alerts, and the pipeline was stress-tested with a synthetic dataset containing ~2.5% deliberately dirty records.

Behind it are 5+ years in data operations and governance: 50 institutions and 2,000+ users consolidated into one verifiable source of truth, and 50% less processing time after automated validation across 30 units. I have also modeled usage data from 20 countries in dbt and BigQuery (served via FastAPI) and built an AI-graded homework platform for a ~120-student school client using Gemini.

I treat data quality as a feature, not a cleanup task.

Open to Junior / Associate Data Engineer roles · Remote-friendly · UTC+7

TECHNICAL SKILLS

Data Engineering: Python, SQL, dbt, Google BigQuery, Apache, Docker, FastAPI

Cloud & Tools: GCP, Git/GitHub, VS Code, REST APIs, phpMyAdmin, DBeaver

TECHNICAL PROJECT

Global Telecommunications Data Pipeline | Python, SQL, dbt, BigQuery, FastAPI 2026

- Architected an end-to-end ELT pipeline integrating complex usage data spanning 20 countries and 4 core entities (destinations, plans, customers, sessions).
- Built a modular transformation layer using dbt to process 90 days of session history for 2,000+ users, converting raw inputs into a clean, analysis-ready BigQuery data warehouse.
- Developed a robust REST API utilizing FastAPI to expose real-time eSIM connectivity data and usage insights for internal services.
- Implemented automated data validation, connection status checks, and deduplication protocols to ensure absolute data integrity across source tables.

Sales Data Pipeline Automation | Python, SQL, Git 2026

- Engineered an automated Python-based ETL pipeline to extract, clean, and process recurring business sales records.
- Loaded transformed transactional data into a relational SQL database for structured storage and scalable querying.
- Authored complex SQL queries to generate aggregate metrics and automated reporting summaries, utilizing Git for continuous version control.

PROFESSIONAL EXPERIENCE

Open Source Creator & Maintainer – pipeline-lint (Data Pipeline Linter) Oct 2026 – Present

Open Source · Published on PyPI · github.com/NafiulSaputra/pipeline-lint

Created and maintain an open source linter that helps data engineers in industry catch the mistakes AI-generated pipeline code makes in SQL, PySpark/Databricks, and Airflow before they reach production. I designed the rule set and CLI behavior; AI tools wrote much of the code, and I specified, reviewed, and tested it.

- Defined 7 lint rules for v0.1 (DE001–DE007) targeting failure modes such as non-idempotent writes and duplicate data on rerun, problems that silently corrupt warehouse tables when schedulers retry jobs.
- Shipped a CLI that fails CI on violations, with a `--fail-on` error option so teams decide how strict the merge gate is.
- Released v0.1.0 to PyPI through GitHub Actions with Trusted Publishing (no long-lived API tokens stored), under the Apache-2.0 license.
- Planned roadmap: an incremental-watermark rule in v0.2 and a VS Code extension that flags issues while code is being written.

Stack: Python · SQL · PySpark · Databricks · Airflow · GitHub Actions · PyPI · Git

Freelance Data Engineer – AI-Graded Homework Tracking Platform (SPASI) Aug 2026 – Present

Independent Project

Sole architect and engineer of a homework-tracking and AI-grading platform for a school client (~120 students) under a 5-year renewable contract. I designed the schema, data flow, security model, and infrastructure; AI tools wrote much of the code, and I specified, reviewed, tested, and corrected it.

- Designed an event-driven pipeline (FastAPI → Redis → stateless Python workers) with a Dead Letter Queue for failed AI gradings and idempotent PostgreSQL upserts, so resubmissions never create duplicate records.
- Integrated Gemini multimodal grading of photo and Word submissions against teacher-defined rubrics, constraining output to a fixed JSON schema and quarantining malformed or hallucinated responses to the DLQ with automatic alerts.
- Encoded school policy as business logic: Haversine geofencing flags homework submitted from inside school grounds, and event-time vs ingestion-time checks accept late uploads but apply an automatic 20% penalty flag instead of rejecting them.
- Designed a three-layer student registration flow (roster verification by name, attendance no., and NISN with rate limiting; single-use activation code; OTP to the registered phone) with an audit log of attempts; teachers sign in via Google OAuth with a 14-day idle session.
- Designed privacy controls for minors' data: student identities hashed to UUIDs with a secret pepper, phone numbers encrypted at rest, and raw GPS coordinates deleted within 24 hours after conversion to a location status.
- Built a micro-batch reverse-ETL sync that writes grades to per-class Google Sheets (tolerant of API rate limits and outages) so teachers can review and override scores without database access; architecture planned to scale toward 10,000 students.
- Directed a full rewrite of AI-generated API and worker code to match my final specification (including a corrected geofence rule), backed by 16 unit tests covering geofence and security logic.
- Defined GCP infrastructure as code with Terraform and automated tests with GitHub Actions; deployment is manually triggered.

Stack: Python · FastAPI · PostgreSQL · Redis · Docker · Google Gemini API · Google Sheets API · Terraform · GCP · GitHub Actions · Tailwind · JavaScript · Git

Fishery Bookkeeping App (Client Project) | Data Architect & Developer

Jul 2026 – Present

- Designed and built an offline-first Android app (Flutter/Dart) that replaces paper bookkeeping for a fish-procurement business buying from 100–200 GT vessels, handling ~1 ton of fish per week (~Rp50 million, ≈US\$3K, in weekly transactions).
- Modeled the transaction data (vessel, catch category, weight, price per kg) and automated price categorization and margin calculation at the point of entry, so totals and margins are computed instantly.
- Chose a local-first design (HiveDB) so operators can record transactions with no connectivity at the harbor; all records persist on-device.
- Automated reporting with Excel (.xlsx) ledger export and PDF receipt printing for thermal printers, replacing hand-written ledgers.

Stack: Flutter · Dart · HiveDB · Excel (.xlsx) export · PDF · Git/GitHub

Dinas Pendidikan Kab. Pati (Gunungwungkal) | Data Operations Specialist - Data Pipeline & Governance

Jan 2024 – Present

- Designed a unified institutional data framework consolidating fragmented datasets across 50 regional institutions and 2,000+ users into a single, verifiable source of truth.
- Developed data automation scripts to streamline collection, validation, and processing workflows, significantly reducing manual bottlenecks and data entry errors.
- Integrated offline operational data streams with official government web portals to enable centralized reporting and cross-platform synchronization.
- Managed asset tracking systems for inventory valued at over \$120,000, enforcing strict data validation rules and governance policies for audit readiness.
- Coordinated with IT staff, institutional operators, and department leadership across 50 regional institutions to align data collection standards and resolve integration issues.

Stack: Python · VBA · EXCEL · CSS · SIMDA

Dinas Pendidikan Kab. Pati (Pucakwangi) | Data Operations Specialist - Data Pipeline & Governance

Jul 2021 – Dec 2023

- Established foundational data governance protocols to structure and process institutional and financial records across 30 operational units.
- Built automated ingestion and data cleaning workflows to transform fragmented fiscal records into reliable datasets for budgetary tracking and compliance.
- Implemented strict data validation rules that minimized reporting discrepancies and decreased overall processing time by 50%.

- Collaborated with treasurers, finance staff, and unit leadership across 30 operational units to validate fiscal data accuracy and align on governance requirements.

Stack: Python · VBA · EXCEL · CSS · SIMDA

PDAM Tirta Bening Kab. Pati | Web Systems Developer (Intern)

Aug 2018 – Oct 2018

- Optimized backend database structures to improve processing efficiency by 70%, enabling stable concurrent access for up to 10,000 users without server overload.
- Diagnosed and resolved critical technical outages on legacy platforms to swiftly restore public-facing digital services.

Stack: Python · HTML · VS Code · CSS · JavaScript · Joomla

EDUCATION & CERTIFICATIONS

Universitas PGRI Semarang | Bachelor's Degree in Information Engineering

2015 – 2020

- Game Developer
- Website Developer
- Data Engineer
- Mathematic
- physical science
- Cumulative GPA: 3.18/4.0

HackerRank | Python (Basic) Certificate

HackerRank | SQL (Intermediate) Certificate

HackerRank | Rest API (Intermediate) Certificate