Md Fakrul Islam Taraque

Data Engineer
98fakrulislam@gmail.com | +880 1939 787 248 | GitHub | Dhaka, Bangladesh

Summary

Data Engineer specializing in the design and architecture of cost-efficient, scalable data infrastructure on GCP. Re-architected production pipelines to cut infrastructure cost by **96%** ($514 to $20/month) while raising freshness from daily batch to **15-minute micro-batch** and enforcing **340+ automated data-quality tests** across assets serving **~144M rows/day**. Backs platform work with **Go** and **FastAPI** backend development for pipeline services and scripting, and represents engineering in client-facing **PoCs, demos, and technical negotiation**. 3.5+ years at Vivasoft; strong **SQL** and **Python** across BigQuery, dbt, Airflow, and Cloud Run.

Professional Experience

Vivasoft Limited Dhaka, Bangladesh
Data Engineer II Jun 2026 - Present
Technical lead: solution design, requirements analysis, stakeholder management, internal product development.

- Own the **technical section of client PRDs** with product managers through the pre-sales/negotiation phase, converting requirements into feasible pipeline scope, data-source, and effort assumptions that gate delivery commitments before they are signed.
- Run **feasibility assessments** and build **MVPs and demos**, presenting the technical case to prospective clients as engineering's counterpart to the CTO, product managers, and C-level boards.
- Built and presented a **geospatial decision-intelligence demo** for City Bank (branch-expansion opportunity scoring from multi-source government data) during the negotiation phase, fielding requirements and technical challenge directly from the CEO/CXO board.
- Drove an **in-house product** from brainstorming to rollout, replacing a 3-person, 3-day/month manual reporting workflow with an automated pipeline (GA4, GSC API, PostgreSQL, dbt, Superset).
- Partner with growth and product managers on **feature-usage and roadmap planning**; produced **MSSQL** correlation reports for Pi-HR (HR and payroll software) assessing new-feature impact on the relational schema and translating product needs into data requirements.

Data Engineer I Mar 2024 - Jun 2026
Product: Weselect - digital talent acquisition platform.

- Reduced **infrastructure cost by 96%** ($514 to $20/month) by re-architecting a monolithic Cloud Composer (Airflow) ETL system into containerized **Cloud Run** jobs orchestrated with Cloud Scheduler.
- Built and enforced **340+ automated data-quality tests** across 30 incremental **dbt** models, guaranteeing reliability and provenance of assets serving **~144M rows/day**.
- Migrated dbt Cloud SaaS to **self-hosted dbt Core** on Cloud Run, eliminating $2,400/year in SaaS cost and confining sensitive transforms inside a **private VPC**.
- Engineered an event-driven **observability and anomaly-alerting framework** (Cloud Logging Sinks, Pub/Sub, Cloud Functions) with real-time **Slack** alerts and centralized audit logging into BigQuery.
- Built a containerized **CI/CD framework** (Docker, Bash, Artifact Registry) automating build, test, and deployment of serverless data workloads while preserving execution history.
- Implemented **Direct VPC Egress + Cloud NAT** for static outbound IPs, enabling firewall-compliant access to restricted MSSQL and MySQL databases from serverless workloads.
- Developed backend APIs in **Go and FastAPI** serving ingestion, metadata processing, validation, and orchestration for distributed data pipelines.
- Scaled data freshness from daily batch to **15-minute micro-batch** intervals with no increase in operational overhead, enabling near-real-time downstream analytics.
- Built **Power BI** dashboards for Sybill (loading, cleaning, transformation, report continuity across schema changes) and optimized **BigQuery** views for Xperanti (intelligent IoT solutions and connectivity) powering **Looker Studio** BI integration.
- Review **AI-generated code** (GitHub Copilot, Claude, and others) in production pull requests weekly, enforcing the same CI/CD and VPC/firewall gates regardless of authorship.

Software Engineer I, Backend Aug 2023 - Feb 2024
Product: Financfy - cloud-based accounting software.

- Developed and integrated scalable **RESTful APIs** (**Go**/Gin) for financial transaction tracking within a **microservice architecture**.
- Managed data persistence and caching strategies using **PostgreSQL** and **Redis** within containerized (**Docker**) backend services.

Software Engineer Intern Nov 2022 - Aug 2023
Product: Wellteam - science-backed corporate wellness and longevity programs.

- Increased product value by **30%** by designing and shipping a corporate-offer feature end to end.
- Improved product speed and load distribution by **20%** by consolidating several microservices and configuring the microservice **API gateway**.
- Grew e-mail marketing reach by **20%** by implementing a newsletter feature on a dedicated microservice.
- Improved marketing-analytics coverage by **60%** by building an e-mail tracking service on a dedicated microservice.

Technical Skills

Languages: Python, SQL, Go, Bash
Data & Warehouse: BigQuery, dbt (Core), Apache Airflow, Cloud Composer, DuckDB, PostgreSQL, MySQL, MSSQL, MongoDB, DynamoDB, Weaviate (VectorDB)
Cloud & Infrastructure: GCP (BigQuery, Cloud Run, Cloud Storage, Pub/Sub, Cloud Functions, Cloud Scheduler, Cloud Build, Artifact Registry, IAM, VPC, Secret Manager), AWS (S3, EC2, MWAA)
Pipelines & Quality: ETL/ELT design, data modeling, incremental modeling, MQTT, RabbitMQ, Redis, SODA, Great Expectations, CI/CD, Docker, Git
BI & Serving: Apache Superset, Looker Studio, Power BI, Streamlit, Pandas, REST APIs (Go/Gin, FastAPI)

Projects

AgriStack - Agro-IoT Telemetry Pipeline | Python, MQTT (HiveMQ), PostgreSQL (Supabase), Streamlit, Render [GitHub]
End-to-end data lifecycle for precision-livestock environmental monitoring (ammonia, humidity, temperature) for a 1,000-bird family poultry farm, built entirely on free-tier infrastructure: Python telemetry generators to a **HiveMQ** MQTT broker for edge transport, a containerized subscriber service on Render, **Supabase (PostgreSQL)** time-series persistence, and a **Streamlit** real-time dashboard. Sensor telemetry is currently simulated pending hardware delivery; the ingestion, storage, and serving path runs end-to-end.

Education

B.Sc. in Computer Science & Engineering, Rajshahi University of Engineering & Technology (RUET) 2023
CGPA 3.18/4.00 (First Class). *Relevant coursework:* Database Systems, Operating Systems, Computer Networks, Data Mining, Parallel & Distributed Processing.

Certifications & Achievements

- **HackerRank:** SQL Gold Badge, Problem Solving Silver Badge, SQL (Basic) Certificate, SQL (Intermediate) Certificate. **LeetCode:** Top SQL 50.
- **700+** competitive-programming problems solved; 20+ inter/intra-university contests.

Leadership & Service

- **Webmaster**, Radio Ruet (University Radio Club) Apr 2021 - Sep 2022
- **Team Lead**, Amar iSchool (promoted from Asst. Team Lead) May 2020 - Feb 2021