

Salayhin Md Sirajus

Staff Data Platform Engineer — Governance & FinOps

Tokyo, Japan · +81-080-6918-4753 · salayhin.lab@gmail.com · linkedin.com/in/salayhin · github.com/salayhin · salayhin.github.io/me

SUMMARY

Staff Data Platform Engineer with 15+ years of experience, from software-engineering roots to sole architect and primary technical decision-maker of a company-wide data platform. Architected and operate a production Apache Iceberg lakehouse on GCP (Scala/Spark, BigLake, dbt) powering self-service analytics, ML, and BI. Drove a platform-wide cost program cutting monthly spend ~60% (¥1.2M → under ¥500K).

Own governance and security for the data platform — PII masking at ingest, GDPR user-data erasure across Iceberg, access control as code, SOC 2 support. Designed a custom MCP server enabling governed, AI-assisted analytics on BigQuery. Previously operated batch and near-real-time telemetry pipelines for 100K+ daily trips at Pathao.

SKILLS

- **Languages:** Python, Scala, SQL, Bash
- **GCP:** BigQuery, BigLake, Cloud Storage, Pub/Sub, Dataproc, Dataflow, Cloud Composer, GKE
- **AWS:** S3, Lambda · S3 Tables (Iceberg), Athena, Glue Catalog, DynamoDB (personal projects)
- **Lakehouse & Modeling:** Apache Iceberg (production), Parquet, Medallion Architecture, Dimensional Modeling
- **ELT & Processing:** dbt Core / dbt Cloud, Apache Airflow, Apache Spark
- **Governance & Security:** PII masking, GDPR erasure, IAM-as-code (Terraform), data lineage & metadata, BigQuery audit logs, data contracts, SOC 2 support
- **FinOps & Observability:** GCP Billing Export, Asset Inventory, Cloud Monitoring, data pipeline observability, cost & data freshness alerting
- **DevOps:** Terraform (Infrastructure as Code), CI/CD (GitHub Actions), Docker, Kubernetes (GKE)
- **AI Tooling:** MCP server design (governed LLM access to data), Claude Code (team enablement, custom skills), AI-assisted code review & engineering
- **BI:** Tableau, Redash

EXPERIENCE

Zeals Co., Ltd — Tokyo, Japan

Apr 2021 – Present

Staff Data Platform Engineer (Nov 2023–Present) · Senior Data Platform Engineer (Nov 2022–Oct 2023) · Senior Data & Infrastructure Engineer, remote (Apr 2021–Oct 2022)

- Sole architect and primary technical decision-maker for the company-wide data platform that analytics, ML, and BI teams self-serve on — owning architecture reviews, tooling decisions, and platform standards across the data organization, and operating the BI layer (Tableau, Redash) as platform-managed services.
- Architected and operate a production Apache Iceberg lakehouse on GCS (BigLake Metastore) with dual-engine access via BigQuery and Spark on Dataproc — chosen for cost, multi-engine reads without duplication, and protobuf schema evolution.
- Built the team's core pipeline framework — a Docker- and script-driven system that provisions and manages ephemeral Dataproc clusters — running Scala/Spark ingestion of ~100 GB/day (peaks ~1 TB) of protobuf event streams from Pub/Sub through landing → bronze → gold medallion layers, with automated Iceberg table maintenance (compaction, snapshot expiry) — orchestrated end-to-end on Cloud Composer (Airflow), the platform's primary orchestrator.
- Drove a platform-wide cost program spanning BigQuery optimization and the Iceberg migration, lowering monthly spend ~60% (¥1.2M → under ¥500K).
- Optimized BigQuery workloads using system tables and billing exports — partitioning, clustering, and gold-layer star-schema modeling (dim_/fact_ tables) — reducing query runtime by up to 90% and scan volume by 80%.
- Led adoption of dbt Cloud with CI/CD, testing, and documentation standards — now the platform's modeling layer, where analytics engineers build and own models — cutting data incidents reaching stakeholders from 2–3/month to under one every two months, and pipeline maintenance from 1–2 days/week to a few hours.
- Own platform governance and security: column-level PII masking at ingest backed by BigQuery policies, GDPR right-to-erasure across Iceberg tables and snapshots, cloud resources and IAM managed as Infrastructure-as-Code via Terraform with PR review (dedicated BigQuery IAM module), end-to-end data lineage and metadata via dbt and GCP-native lineage tracking, and partnered with security engineers on SOC 2 compliance for the analytics layer.

- Shipped a FinOps & observability suite running on GKE (Kubernetes): daily cost collection across all GCP projects with billing-based actuals, expensive-query and cache-miss alerts, BigQuery freshness and quota alerts, Cloud Composer (managed Airflow) pipeline summaries, and a central observability dashboard.
- Designed a custom MCP server for governed, AI-assisted analytics on BigQuery — schema discovery and self-service, plain-language querying of business metrics, restricted to gold tables via a security layer; championed team AI adoption (Claude Code enablement, best practices, custom skills, AI-assisted code review).
- Standardized CI/CD deployment via GitHub Actions across all platform workloads — Spark pipelines and dbt models alike ship through automated pipelines; defined data contracts, quality standards, and SLAs across teams; mentored junior engineers.

Pathao Ltd — Dhaka, Bangladesh

Jun 2018 – May 2021

Senior Data Engineer Level 2

- Designed and operated batch and near-real-time telemetry pipelines (trip lifecycle, GPS location streams, driver/rider actions) for 100K+ daily trips across ride-sharing, courier, and food delivery — spanning multiple cities across Bangladesh and Nepal.
- Implemented data pipelines integrating transaction data from 20+ sources plus streaming data at high volume (~50 GB/day) and velocity (peaks of 10K events/sec) on Spark, Dataflow, GCS, and Airflow.
- Created the metrics framework computing 100+ business KPIs (driver earnings, leaderboards, operational metrics) with under-5-minute end-to-end latency.
- Spearheaded the company's data lake and 10 TB BigQuery data warehouse on Google Cloud Storage, halving data retrieval times.
- Established the analytics ecosystem on Dataproc, Dataflow, BigQuery, and Data Studio, and delivered a scheduled reporting framework generating 10+ financial and operational reports.

Augmedix Bangladesh Ltd. — Dhaka, Bangladesh

Jun 2015 – Mar 2018

Senior Data Engineer

- Core contributor to the design and build of a SQL Server 2016 data warehouse on AWS (S3-backed storage) supporting company-wide business reporting.
- Developed AWS data pipelines (S3, Lambda) consolidating data from third-party APIs, Google Sheets, and MySQL (100+ attributes), and a reporting suite producing 20+ operational reports.
- Delivered backend services serving millions of warehouse records to internal applications.

Nascenia Ltd — Dhaka, Bangladesh

Jun 2013 – May 2015

Software Engineer

- Launched a B2B agricultural marketplace platform with Ruby on Rails and MySQL.
- Developed a social media analytics and post-scheduling tool with Rails and PostgreSQL.

Right Brain Solution Ltd — Dhaka, Bangladesh

Nov 2011 – May 2013

Software Engineer

- Co-authored a high-traffic automotive portal for the Toronto Star.
- Created APIs for social media mobile apps and a multilingual B2C ski-rental e-commerce platform.

Earlier roles: Assistant Programmer, Nanosoft (2011) — PHP/MySQL appointment scheduling · Software Developer, Sitebloom (2010–2011) — CodeIgniter school management system.

LANGUAGES

English — full professional proficiency · Japanese — beginner, actively studying (Genki I)

EDUCATION

IICT, BUET — Post Graduate Diploma in ICT

2019

Jagannath University — B.Sc in Physics

2008

CERTIFICATIONS

- [Data Engineering with Google Cloud Specialization](#) — Google Cloud Training, Coursera (2023)
- [Data Engineering, Big Data, and Machine Learning on GCP Specialization](#) — Google Cloud Training, Coursera (2023)