Naineel Soyantar

Ahmedabad, Indianaineelsoyantar@gmail.com | naineel.in | LinkedIn | GitHub

Data Engineer / Analytics Engineer

Summary

Data Engineer with nearly 3 years of experience building production ingestion pipelines, lakehouse transformations, and analytics models for logistics, sales, and marketing. Hands-on with Python, SQL, PySpark, Databricks, dbt, Delta Lake, Airflow, and AWS. Background covers API and email ingestion, dimensional modeling, data quality, Terraform, and CI/CD.

Technical Skills

Data: Python, SQL, Databricks, dbt, Apache Spark (PySpark), Delta Lake, Apache Iceberg, Kafka, Power BI
Pipelines: Airflow, Temporal, Fivetran, dimensional modeling, data quality
Cloud: AWS S3, IAM, SNS, ECS Fargate, Terraform, Docker, Kubernetes, GitLab CI/CD

Experience

ZURU Tech Ahmedabad, India
Data Engineer 1 Jan 2026 to Present

- Co-built Spine, a Python ingestion framework that cut source onboarding from **5 days to 1 day**. Configured Delta Lake, Parquet, and Apache Iceberg exports on AWS; reduced peak memory use by **50%** by storing API-response batches before creating DataFrames.
- Co-migrated Python notebooks into **300+** dbt SQL models on Databricks. Consolidated jobs into one deployment with reusable macros and explicit model dependencies.
- Provisioned Amazon Managed Workflows for Apache Airflow (MWAA) through Terraform to orchestrate **25+** production ingestion and dbt pipelines. Built a shared notification layer for these Airflow DAGs that routes pipeline alerts to Microsoft Teams, email, or Amazon SNS, with default routing and per-DAG overrides.
- Own daily logistics pipelines across eight country schemas for inventory, order, inbound, and outbound reporting.
- Delivered Logistics Cost Analysis on Databricks across **10.5M+** outbound records and **900K+** cost records. Designed freight and storage allocation models for product, order, and billing-period reporting, replacing calculations embedded in Power BI. Added validation checks and allocation-exception reporting to make cost calculations traceable.
- Designed a Databricks on-time, in-full (OTIF) scorecard across **132K+** orders, consolidating warehouse feeds into country-specific fulfillment metrics. Estimated delivery dates using a transit reference covering 98 shipping routes.
- Added source tracking and data-quality indicators to distinguish fulfillment failures from missing or stale evidence and prevent inflated warehouse scores.
- Developed reusable email ingestion with Microsoft Graph, Temporal, and ECS Fargate to replace Zapier-triggered notebooks. Validated attachments and retained originals in S3 for audit and replay.
- Led a cross-account AWS and Databricks migration without downtime using Terraform and Delta Lake backfills.

Junior Software Engineer Jan 2024 to Dec 2025

- Built AI agents to standardize Chinese customer names and addresses, verified through Exa search with manual review. Automated SAP stock-shortage and shipping-delay alerts, using Gemini to structure warehouse replies.
- Configured Fivetran ingestion for Meta, TikTok, and Google ads into S3, exposed through Databricks external locations. Provisioned cross-account Databricks and S3 access with reusable Terraform modules.
- Filled Fivetran's Meta/TikTok reach-and-frequency gaps with asynchronous Python ingestion for daily refreshes and backfills. Added seven-day rolling and cumulative reports across **3.2M+** Meta/TikTok records.

Projects

IoT Lakehouse Pipeline GitHub | Built a local Kafka and Spark Structured Streaming pipeline for synthetic device data. Added Delta Lake tables, duplicate-safe writes, invalid-record quarantine, and DuckDB quality reports.

Spine Open Source | Extended the framework with paginated API handling, multiple authentication methods, and connectors for logistics and advertising sources.

Cyclops GitHub | Added ReplicaSet views, live log streaming, and Prometheus cache metrics to an open-source Kubernetes UI.

Stock Analysis API GitHub | Implemented Piotroski F-Score in Go to evaluate financial strength across nine accounting criteria. Received public recognition from the project creator during Hacktoberfest 2024.

Awards and Certifications

Databricks Building Intelligent Apps with Data + AI Hackathon May 2026
Placed **2nd** among 39 APAC teams and won a $2,500 team prize for Kāinga, an AI housing assistant.
Developed chat and agent experiences with Lakebase, Genie, and Mosaic AI Agent Framework.

AWS Certified Cloud Practitioner Jan 2026 to Jan 2029

Education

L.D. College of Engineering Oct 2020 to May 2024
Bachelor of Engineering in Computer Engineering | CGPA: 9.11/10.0 | Ahmedabad, India