

Tanmay Singh

Delhi, India | tanmay.work10@gmail.com | [LinkedIn](#) | [Github](#) | [Portfolio](#)

Data Engineer with experience designing and scaling ETL pipelines, relational schemas, and cloud data infrastructure for AI-driven products and enterprise financial systems. Skilled in building reliable, high-throughput data platforms across Airflow, dbt, BigQuery, PostgreSQL, Supabase, and AWS, with a track record of cutting pipeline runtime by 40% and processing 10M+ records monthly at production scale.

EDUCATION

Arizona State University — Master of Science in Data Science, Analytics and Engineering
Manipal University, Jaipur — Bachelor of Technology in Computer Science and Engineering

EXPERIENCE

Vault IQ

Dec 2025 – May 2026

Data Engineer (AI and Backend Systems)

Remote

- Designed and implemented relational schemas and storage for user, brand, product, and social data in Supabase (PostgreSQL), giving 4+ product modules one structured data flow
- Worked full stack across the data layer, backend services, and product UI of an AI-powered brand development platform integrating OpenAI, Gemini, and Flux Pro
- Sped up the brand-generation pipeline that orchestrates multi-model LLM and image-generation calls for content and product visualization
- Built a photoshoot agent that stages generated product images in marketing-ready settings such as hero shots, using reasoning-based prompt orchestration while preserving product fidelity
- Built and deployed a production chatbot with a custom knowledge-base integration layer for contextual querying and FAQ automation
- Built the membership dashboard backend serving brand portfolio, product catalog, and revenue metrics in real time, integrating 6+ external services and APIs

Capgemini Pvt Ltd

Oct 2020 – Mar 2023

Data Engineer – HSBC UK (Application Data Management)

United Kingdom

- Designed and orchestrated scalable ETL pipelines in Apache Airflow (DAG-based workflows) for batch processing of HSBC's customer credit data systems
- Managed and processed 10M+ customer records monthly across Google BigQuery and PostgreSQL environments, ensuring data consistency, integrity, and availability
- Developed transformation and validation workflows in Python and PySpark, with data quality checks aligned to business-critical credit risk models
- Optimized BigQuery queries and storage structures and implemented SCD logic, reducing batch pipeline runtime by 40% and meeting ETL SLAs
- Built monitoring into pipeline runs to catch data anomalies pre-emptively, improving the reliability of production credit data systems

PROJECTS

Mettle

In development

Gamified workout tracker: workouts earn XP, benchmarks rank you E to S

React Native · TypeScript · Supabase

- Designed an offline-first data model: every set is written to on-device SQLite (Drizzle ORM), then synced to Supabase Postgres on sign-in, so logging works with no gym signal
- Secured the cloud schema with row-level security on every table, Google and email-code sign-in, and full account deletion
- One TypeScript XP engine also runs in a Supabase Edge Function, recomputing leaderboard XP server-side so it can't be faked

CheckYourAQI

Live

Air-quality forecasts 1–3 days ahead for 70 Delhi NCR stations

Airflow · dbt · AWS · LightGBM

- Airflow on EC2 ingests station readings (OpenAQ), weather and air-quality forecasts (Open-Meteo), and satellite fire detections (NASA FIRMS) into an S3 bronze layer
- Modeled a bronze → silver → gold Iceberg lakehouse with 27 dbt models on Athena and 44 data tests, including one that fails the build if any feature post-dates its prediction
- Serves a public dashboard and FastAPI on Lambda from versioned Parquet snapshots; retrains weekly on GitHub Actions and runs unattended for ~\$21/month with failure alerts

CheckYourData

Complete

Upload a dataset; an AI agent proposes data-quality checks to review and run

FastAPI · Claude API · Postgres

- Profiles each uploaded CSV's schema and stores datasets, check configs, run results, and history in Postgres, tracking drift between runs
- An agent proposes checks from the schema and a sample; per-check validation rejects malformed suggestions and retries
- Caches suggestions by a schema hash (~4 s → ~40 ms on repeat); ships as one Docker service with uploads in Supabase Storage

SKILLS

Languages: Python, SQL, PySpark, TypeScript

Data Platforms & Warehousing: PostgreSQL, Google BigQuery, Supabase, Amazon S3, Apache Iceberg, Athena, Data Modeling, Schema Design

Pipelines & Orchestration: Apache Airflow (DAG-based ETL), dbt, AWS Glue, Kinesis, Step Functions, Batch & Streaming, Data Quality Checks

Cloud & Tools: AWS (EC2, Lambda), Google Cloud, Docker, GitHub Actions, Git, FastAPI, REST APIs, LLM/Generative AI Integration