

Hamza Ahmed

Location: New York, NY

Social: [LinkedIn](#) | [GitHub](#) | [Medium](#)

Phone: (910) 495-3304

Email: hamza-94@live.com

SUMMARY

Senior Data Engineer with 7+ years of experience building data platforms, real-time pipelines, and cloud infrastructure across AWS, GCP, Azure, Snowflake, and Databricks. Over the last few years, my work has shifted toward automating data workflows and building AI and agentic systems on top of production data platforms. I've owned systems end to end, from ingestion and orchestration through analytics and deployment.

EDUCATION

North Carolina State University - Raleigh, NC

B.S. Physics, Minor in Mathematics

August 2015 - May 2019

TECHNICAL SKILLS

Languages: Python, SQL, Go, Bash, TypeScript, Rust

Data Platforms & Storage: Snowflake, Databricks, BigQuery, Redshift, PostgreSQL, MongoDB, ClickHouse, Delta Lake

Data Engineering & Orchestration: Spark/PySpark, dbt, Airflow/MWAA, Fivetran, Airbyte, AWS Glue, Dataflow, Kafka, Pub/Sub

Cloud & Infrastructure: AWS, GCP, Azure, Terraform, Docker, Kubernetes, GitHub Actions, CircleCI

AI / ML & Agentic Systems: LangChain, LlamaIndex, OpenAI API, RAG, FAISS, BERT, MCP, Databricks Genie/GenieAgent, TensorFlow, Keras

Analytics & Integrations: Sigma, Tableau, Looker Studio, Salesforce, Customer.io, Segment, Contentful, Sailthru

RECENT EXPERIENCE

Pidwin LLC

January 2024 - August 2026

Remote - Founding Data Engineer

- Built the core data platform for real-time and batch ingestion from MongoDB into S3 and Databricks, using Kafka for streaming and Pub/Sub notifications to trigger retrieval and downstream processing of delivered data; processed **millions of records per day**.
- Cut pipeline infrastructure costs by **96%** by moving streaming workloads to EC2 and improving S3 compression and storage.
- Built an **agentic workflow** in Databricks using SearXNG and **GPT-OSS 120B** to research new games, generate SEO-ready content, and identify relevant users for targeted launches.
- Built an **agentic VIP support workflow** using Python, Telethon, PostgreSQL, S3, and Databricks. It identifies customer issues, creates dedicated chats, routes them to the right support agents, and uses LLMs for summaries and personalization, cutting manual response time by **80%**.
- Built a crypto fraud detection system using blockchain intelligence APIs, deep learning models, and anomaly detection for AML/KYC, reducing fraud by **70%**.
- Built production AI tooling in Databricks, including a real-time moderation service that reduced abuse incidents by **87%**, plus Genie/GenieAgent, Customer.io, and Sigma **MCP integrations** for stakeholder workflows.
- Built **Gross Gaming Revenue and Enterprise Value** reporting in Python/SQL and Sigma and standardized shared metric definitions across the business.
- Integrated Python and Databricks backend services with Pidwin's **TypeScript application stack** and managed AWS/Databricks deployments with Terraform and GitHub Actions.

PREVIOUS EXPERIENCE

Hitachi (Method)

December 2021 - January 2024

Remote - Senior Software Data Engineer

- Built data solutions across AWS, GCP, Azure, Snowflake, and Databricks for clients including **Google, Nike, Epic Games, and Hitachi Energy**.
- Managed Kafka pipelines in Databricks for near-real-time ingestion of **100K+ records per day**.
- Cut AWS Athena costs by **30%** and improved query speed by **45%** by tuning S3 partitioning, bucketing, and compression.
- Built Python and AWS Glue validation jobs for **25K+ XML files per day**, catching transformation issues before downstream loads.
- Built a **natural-language data querying tool** using LangChain, DIN-SQL, BERT, and FAISS.
- Built and maintained dbt pipelines and Airflow/MWAA DAGs; automated API ingestion into BigQuery with Python, Go, and Cloud Functions, helping increase BigQuery usage by **20%** and tracking speed by **60%**.
- Built incremental **SQL Server-to-Snowflake** loads using Change Tracking and automated staging and production setup with Terraform, GitHub Actions, and Cloud Build.

Liquid Agency

January 2020 - December 2021

Remote - Founding Data Engineer

- Owned and documented the agency's **data engineering strategy and infrastructure**.
- Built Fivetran and dbt pipelines that cut processing time by **60%** and improved data accuracy by **40%**.
- Built Python scrapers for **250+ web pages** to collect campaign, content, and competitive data.
- Built a **Snowflake warehouse** for Google Analytics and social media data using Fivetran and Cloud Functions.
- Built **Tableau dashboards** for campaign and social performance reporting.

NCSU Astrophysics

August 2017 - October 2019

Raleigh, NC - Researcher / Data Scientist

- Built mathematical models around **Einstein Field Equations** using Keras and TensorFlow.
- Pulled and processed observatory data in **Azure Synapse** for modeling and analysis.
- Solved Morris-Thorne wormhole and geodesic equations using **Python and MATLAB**.
- Compared theoretical models with observational data as part of wormhole research.

SELECTED PROJECTS & VENTURES

MartianBee - Founder / Engineer

2026 - Present

Rust, DataFusion, Delta Lake, Apache Arrow, PostgreSQL

- Building a Rust-based analytics engine where dashboard definitions drive ingestion, transformation, and caching plans so small businesses can get warehouse-grade analytics without a dedicated data team.
- Built the execution layer with DataFusion, Apache Arrow, Delta Lake, and PostgreSQL, with AI-assisted query optimization and text-to-SQL for non-technical users.

Electronic Drumkit Calibration System

- Converted an acoustic drum kit into an electronic kit using piezoelectric sensors, Python, and PostgreSQL.
- Built an auto-calibration module that adjusts sensor thresholds, filters noise, and maps hit intensity to velocity values.