

Aathif Naseer

Databricks Data Engineer

PROFILE

Data engineer and AI builder with 2 years of internship experience building production ETL pipelines, RAG systems, and full-stack SaaS products. Worked across data processing systems, applied AI, and designed reliable data pipelines.

CONTACT DETAILS

✉ aathif394@gmail.com

🌐 linkedin.com/in/aathifnaseer

🐙 github.com/aathif394

📍 Chennai, India

TECHNICAL SKILLS

Data Engineering:

Databricks, PySpark, Polars, Delta Lake, SQL, ETL/ELT

ML & AI:

LlamaIndex, RAG Pipelines, Vector DBs, pydantic-ai, CrewAI, XGBoost

Backend & APIs:

FastAPI, Django REST, PostgreSQL, Supabase

Frontend:

React, TypeScript, Tailwind CSS, Vite

Cloud & DevOps:

AWS, Cloudflare, Docker, Git, Linux

Languages:

Python, SQL, TypeScript, JavaScript, Bash

CERTIFICATIONS

Databricks Certified Data

Engineer Associate

2026

EXPERIENCE

DATA ENGINEERING INTERN

WizSystems (Remote / New Jersey, USA)

Apr 2025 – Jun 2025

- ◇ Shipped production-grade incremental ETL pipeline using Polars and FastAPI to sync multi-source operational data to AWS S3 on automated schedules.
- ◇ Implemented time-window based extraction, transformation, and load logic with configurable watermarks, structured logging, and diagnostics.
- ◇ Built configurable SQL execution engine that runs scheduled SQL jobs, stores outputs in S3, and exposes results through typed REST APIs.

MACHINE LEARNING INTERN

Ideassion Technology Solutions (Chennai, India)

Jul 2024 – Aug 2024

- ◇ Engineered enterprise RAG pipelines for document Q&A over large PDF corpora using chunking, embeddings, vector search, and reranking.
- ◇ Tuned hybrid retrieval using dense and sparse search to improve answer precision; reduced hallucinations by enhancing context selection.

DATA ENGINEERING INTERN

WizSystems (Remote / New Jersey, USA)

Oct 2022 – Apr 2023

- ◇ Designed declarative JSON configuration format for multi-source ingestion across CSV and relational databases with incremental load support.
- ◇ Built Django REST APIs for transformation configuration, SQL generation, and pipeline orchestration consumed by downstream PySpark workflows.
- ◇ Delivered internal dashboards using Flutter and Supabase to improve visibility into operational data and reduce manual reporting cycles.

PROJECTS

TABLEONE — ENTERPRISE DATA PLATFORM

2024–2025

- ◇ Built full-stack multi-tenant SaaS platform combining dynamic table builder, form engine, dashboard composer, and workflow automation.
- ◇ Engineered tenant isolation via PostgreSQL RLS policies and JWT subdomain routing, with custom RBAC supporting field-level ACL.
- ◇ Developed FastAPI AI agent service using pydantic-ai enabling autonomous data operations over tenant databases.

AGENTIC SUPPLY CHAIN COPILOT

2025

- ◇ Built demand forecasting pipeline using stacked XGBoost + LightGBM ensemble for autoregressive forecasts across a multi-store logistics network.
- ◇ Engineered multi-agent AI system automating procurement scoring, EOQ reorder decisions, and automated supplier selection.

GENBI — TEXT-TO-SQL ANALYTICS PLATFORM

2024

- ◇ Built platform translating natural language into validated SQL with auto-generated charts, reducing ad-hoc query turnaround.
- ◇ Engineered multi-database query engine with live schema introspection, syntax validation, and execution monitoring.

EDUCATION

B.TECH IN ARTIFICIAL INTELLIGENCE & DATA SCIENCE

Crescent University (Chennai, India)

Sep 2022 – May 2026

- ◇ CGPA: 9.5/10