

SHRIVIDYA HEGDE

Senior AI Data Engineer | Apache Airflow Contributor & Champion | 5× Astronomer Certified | LLM / RAG Pipelines | Cloud-Native Data Platforms
shrividyahegde222@gmail.com • Parsippany, NJ • [LinkedIn](#) • [GitHub](#) • [Medium](#) • [Substack](#)

PROFESSIONAL SUMMARY

Senior AI Data Engineer with 10+ years of progressive engineering experience across data platform architecture, ETL/ELT pipeline design, MLOps, and LLM/RAG pipeline development. Apache Airflow open-source Contributor with 7 approved merged PRs and recognized Airflow Champion (Jan 2026); holds 5 Astronomer certifications across Airflow 2 and 3. Deep expertise in Apache Airflow, Astronomer/Astro, PySpark, Snowflake, Databricks, dbt, and cloud-native data orchestration on AWS, GCP, and Azure. Hands-on experience building production-grade AI systems, including LLM-powered RAG pipelines using LangChain, ChromaDB, and OpenAI API grounded in a quality-first, reliability-obsessed engineering mindset. Active open-source contributor, technical author on Medium, and NJ Chapter Lead for Women in Data™.

CORE COMPETENCIES

Data Engineering: Apache Airflow 3, Astronomer/Astro, ETL/ELT Pipelines, Data Orchestration, dbt, Data Modeling, Data Warehousing, Data Lakehouse, Apache Iceberg, Apache Spark, PySpark, Batch processing, ClickHouse, Snowflake, Databricks, Streamlit

AI / ML Engineering: LLM Integration, RAG Pipelines, LangChain, OpenAI API, Agentic AI – Crew AI, Vector Databases (ChromaDB, Weaviate), NLP (NLTK, TextBlob, Vader), Sentiment Analysis, Semantic Search, Transformers, Prompt Engineering, Generative AI, MLOps

Cloud & DevOps: AWS (Certified), GCP, Microsoft Azure, Kubernetes, Docker, Terraform, CI/CD, GitHub Actions, CircleCI, GitLab CI, Azure DevOps, Infrastructure-as-Code

Programming: Python, SQL, PySpark, SQLAlchemy, TypeScript (Basics), Shell, FastAPI, GraphQL

Data Quality & Observability: Data Quality Engineering, Observability, Reliability Engineering, Production Monitoring, Pipeline Testing, Performance Optimization, Astro Observe (OpenLineage + Airflow),

Testing Frameworks: Selenium, Playwright, Pytest, Cypress, Appium, BrowserStack, REST Assured, Postman, Locust

FEATURED PROJECTS

Article Insights NLP Pipeline | Apache Airflow 3 · Python · NLTK · TextBlob · NLP · dbt-style Orchestration ■

Automated end-to-end NLP pipeline orchestrated with Airflow 3, ingesting Medium articles and executing parallel DAG tasks: keyword extraction, extractive summarization, sentiment analysis, named entity recognition (people/orgs/locations), Flesch-Kincaid readability scoring, and auto-generated social media snippets.

- Results output as structured JSON for downstream content strategy and SEO use — demonstrating production-grade data pipeline design with LLM-adjacent NLP tooling.

Real-Time Stock Sentiment Analyzer | Python · Streamlit · TextBlob · yFinance · NewsAPI · Docker · GitHub Actions

- Built a production-ready AI dashboard fetching live stock prices and news headlines, performing sentiment analysis (positive/negative/neutral), and visualizing results interactively via Streamlit.
- Fully Dockerized for deployment with a CI/CD pipeline via GitHub Actions; supports real-time and sample data modes — end-to-end MLOps deployment lifecycle.

PROFESSIONAL EXPERIENCE

Apache Airflow Open-Source Contributor

The Apache Software Foundation | Remote | Mar 2026 – Present

- Active open-source contributor to Apache Airflow with 7+ approved and merged PRs, contributing code improvements, bug fixes, and documentation enhancements to one of the most widely adopted data orchestration frameworks in the data engineering ecosystem.
- Recognized as Apache Airflow Champion (Jan 2026) by Astronomer for outstanding community contributions, PR reviews, issue triage, and advocacy across the Airflow ecosystem.
- Participating in open-source community governance and feature development for Apache Airflow 3, deepening expertise in next-generation DAG authoring, dynamic task mapping, and scheduler improvements.

Apache Spark Open-Source Contribution (PR reviews only):

- Reviewing the PRs submissions in the open source community.

Senior Software Engineer - Astro Observe(Data Platform)

Astronomer | New York, NY (Remote) | Mar 2022 – Nov 2025

- Joined as SDET and rapidly expanded scope to design and maintain cloud-native data pipeline frameworks for Astro (Astronomer's managed Apache Airflow platform) using Python, Airflow DAGs, and cloud SDKs (boto3, azure-sdk, gcp-sdk, astro-sdk), increasing test execution throughput by about 30%.
- Built and optimized ETL/ELT orchestration workflows with Apache Airflow and Kubernetes across GCP, Azure, and AWS environments, reducing data delivery timelines by at least 20%.
- Engineered RAG pipelines using LangChain and ChromaDB (vector database) for an AI recommendation system that summarized test logs for release gating, generated recommended fixes, surfaced potential bugs, and routed findings to developer channels via Slack Webhooks — boosting developer productivity and resolution speed by roughly 40%.
- Developed automated cloud resource lifecycle management scripts using GCP CLI and AWS CLI, reducing infrastructure costs by approximately 60%.
- Integrated data quality checks and observability hooks into CI/CD pipelines (CircleCI, GitHub Actions), ensuring pipeline reliability and data integrity at enterprise scale.
- Earned Apache Airflow Champion recognition and 5 Astronomer certifications (Airflow 2 and Airflow 3) during tenure.

Senior Test Automation Engineer (Core Payments Data)

Commonwealth Bank | Bangalore, India | May 2021 – Feb 2022

- Led strategic automation upliftment for Core Payments data systems, achieving 95% automation coverage and reducing testing costs by 15%.
- Designed cost-efficient, reusable automation frameworks across microservices environments enabling faster, more reliable data pipeline validation.
- Received Best Partner Award for outstanding automation framework delivery.

Consultant – Automation Specialist

KPMG India | Bengaluru, India | May 2020 – May 2021

- Developed and enhanced automation frameworks across 4+ project teams, reducing manual QA cycles by 25% and improving automation adoption by 30%.
- Conducted PoC on SAP S/4 HANA applications using Eggplant Functional & DAI, delivering data readiness and governance recommendations.
- Extended mobile data validation across Android and iOS using BrowserStack and Appium.

IT Analyst – Automation Engineer

Tata Consultancy Services (TCS) | Sydney, Australia & Bangalore, India | Jun 2016 – Apr 2020

- Built unified automation frameworks supporting CI/CD across Banking & Finance data platforms, reducing manual testing efforts by 50% and improving deployment frequency by 35%.
- Engineered Selenium WebDriver, Java, and Python automation solutions across Agile and Waterfall delivery models.
- Won Service Commitment Award and On the Spot Award for critical project delivery.

Software Engineer / ASE Trainee

Tata Consultancy Services (TCS) | Bengaluru, India | Jun 2014 – Jun 2016

- Built Java automation tools for SQL and PL/SQL data operations, reducing manual database intervention by 60%; trained in Informatica, Business Intelligence (BI), and PL/SQL, forming a foundational Data Engineering skillset.
- Won On the Spot Award ×3 and Special Initiative Award (1st Prize, National ILP Debate Competition).

CERTIFICATIONS

- **Astronomer Champions Program – Apache Airflow** — Astronomer — Jan 2026
- **AWS Certified Cloud Practitioner** — Amazon Web Services — Jun 2025
- **Generative AI Fundamentals** — Databricks Academy — Mar 2025
- **Python Data Associate** — DataCamp
- **Data Engineer Associate** — DataCamp
- **AI Fundamentals Certified** — DataCamp

COMMUNITY & THOUGHT LEADERSHIP

Women in Data™ – NJ Chapter Lead | Pro Member, Data Engineering Track

- Leading the New Jersey chapter, organizing events and fostering a local community of data professionals.
- Participating in advanced data bootcamps covering SQL, Python, PySpark, feature engineering, applied ML, and end-to-end data workflows.

Apache Airflow Championship:

- Spoke at [Apache Airflow NYC MeetUp](#) on "Airflow + AI : Orchestrating the Intelligent Stack"
- Session accepted at Apache Airflow Summit 2026
- Answered various queries from airflow users globally at the user-troubleshooting slack channels in Airflow slack
- Recognised as one of the top 5 Airflow Champions Globally in the Q2 Champions meet.
- Conducted sessions on Airflow + dbt in Dbt labs - Women in Data
- Conducted hands-on Airflow Workshops for DE Club, Women in Data

Volunteering:

- Member and Volunteer of Women Applying AI
- Speaker at [Future World Conversations](#)
- Writer and Editor for 5 publications under [PlainEnglish.io](#)
- Editor for publication "dbt for Analytics"
- Expert Content Writer for Data Engineering Byte on Substack

Technical Writing — [Medium](#) & [Substack](#)

- RAG Without Vectors: How LLMs Are Learning to Navigate Documents Like Humans
- The Hidden Tax: How Kubernetes Is Quietly Draining Your Cloud Budget and Ways to Stop It — DevSecOps Community
- What's Actually Costing You in Apache Airflow®?
- Before You Deploy an AI Agent, Read This
- LangChain's Deep Agents Are Actually Worth Your Time

Portfolio Website(Airflow themed):[Shrividya Hegde](#)

EDUCATION

Bachelor of Engineering – Computer Science & Engineering

Malnad College of Engineering (9.06 CGPA) 2010 – 2014