

Aravind Suresh

Data Engineer

aravind.suresh.tech@gmail.com

linkedin.com/in/aravind-suresh8

(+91) 9004470215

www.aravindsuresh.com

SUMMARY

Data Engineer with 2+ years of experience architecting scalable end-to-end data pipelines and building cloud-native data warehouses, ML data systems, and AI-powered infrastructure. Experienced in collaborating with cross-functional teams across the full data lifecycle – from ingestion and orchestration to serving dashboards that drive decisions at enterprise scale. Strong focus on pipeline reliability, data quality, and operational automation.

EXPERIENCE

Data Engineer Specialist

GE Aerospace

Jul 2024 – Present

Bangalore, India

- Architected a cloud data warehouse pipeline on Redshift and S3 integrating database extracts and file-based feeds into a single orchestrated workflow using Dataiku and Talend, with proactive freshness and completeness validation gates, achieving **99.5% reporting accuracy** for enterprise EHS compliance metrics.
- Spearheaded the migration of on-premises data workloads to AWS (Redshift, Lambda, S3), re-architecting legacy EC2-based processes into event-driven serverless pipelines and managing all infrastructure as code via Terraform, reducing operational costs by **~35%** and achieving zero-downtime cutover.
- Engineered end-to-end ML data pipelines for automated safety compliance review processing 5+ GB/day, and led the transition from a multi-channel CNN model to a hybrid **multi-agent orchestration system** combining LLM agents with ML models, reducing manual review cycles by **~70%**.
- Designed 15+ end-to-end analytics pipelines and dashboards on Tableau and Spotfire, serving **500+ users** including senior leadership with near-real-time visibility into critical infrastructure across global operations.

Data and Analytics DT Intern

GE Aerospace

Jan 2024 – Jun 2024

Bangalore, India

- Built an end-to-end automated data lineage solution integrating Databricks Workflows, Airflow DAGs, AWS Redshift, and Hive Metastore, achieving column-level granularity across the full data stack using APIs, SQL parsers, and custom business logic.
- Automated upstream and downstream lineage scanning with results stored in a database and surfaced via dashboards, reducing root cause analysis time from **~2 weeks to ~2 minutes**.

TECHNICAL SKILLS

Languages: Python, SQL, PySpark, Pandas, Java, C++

Data Platforms & Processing: Databricks, Apache Spark, Delta Lake, Snowflake, Kafka, Talend, Dataiku, Tableau, Spotfire

Databases: PostgreSQL, AWS Redshift, MongoDB (NoSQL), MySQL, MS SQL Server

Orchestration & Architecture: Airflow, dbt, ELT/ETL Pipelines, Dimensional Modeling (Kimball), Data Warehouse

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

AI & Agent Frameworks: LangChain, LangGraph, CrewAI, MCP, OpenAI SDK, RAG, Vector Databases, n8n

CERTIFICATIONS

Databricks and Spark Core · dbt Bootcamp · Big Data with Spark and Hadoop · TIBCO Spotfire Certified · AWS Certified Cloud Practitioner

PROJECTS

Signal Extraction | Hackathon Runner-Up

Nov 2025 – Jan 2026

- Architected a multi-stage ELT pipeline on Snowflake combining market price data and news sentiment, using Snowpark, XGBoost, and SHAP for explainable buy/sell signal generation.
- Engineered orchestration via Snowflake Tasks & Streams, feature pipelines for technical indicators (EMA, RSI, MACD) and NLP sentiment signals, with a Streamlit dashboard for signal exploration and model monitoring.

AWARDS & RECOGNITIONS

- Runner Up, Snowflake Dev Premiere League (Signal Extraction) – Jan 2026
- 3× GE Aerospace Impact Award recipient – Q4 Corp DT Excellence, Q3 Bravo, Q1 Encore; recognized for customer-driven delivery and continuous improvement – 2025

EDUCATION

Amrita School of Computing, Coimbatore, India — B.Tech in Computer Science and Engineering

2020 – 2024

LANGUAGES

English (Fluent) · Hindi (Conversational) · German (Basic A1)