

# SURIYA RAJAN

Senior Data Engineer | AWS | Apache Spark | Airflow | Python | Scala | Terraform

Bangalore, India (relocating to Warsaw) | +91-7418814170 | [suriyarajan12@gmail.com](mailto:suriyarajan12@gmail.com) | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

Requires EU Blue Card or work permit sponsorship | Available in 30 days

## PROFILE

Senior data engineer with **9 years** on AWS, building and tuning data platforms for media measurement and AdTech. Owned Spark and Airflow pipelines handling **500M+** events a day. At Nielsen, cut EMR compute cost by **40%** by moving to EMR Serverless, and made Spark 3.x jobs **at least 30% faster**. Recent work covers serverless integration, RAG chatbots and BI migration (MicroStrategy to Amazon Quick Suite, Oracle to Apache Iceberg). Has mentored junior engineers, an intern and experienced new joiners, and set pipeline standards across regional teams. AWS Certified Solutions Architect and Astronomer-certified in Apache Airflow.

## SKILLS

- **Data engineering:** Apache Spark 2.x/3.x, PySpark, Scala, Apache Airflow, ETL/ELT, data modelling, data quality, governance, observability, SLA management
- **AWS:** EMR, EMR Serverless, Glue, Lambda, API Gateway, SQS, S3, Athena, CloudWatch, IAM, VPC, EC2, Amazon Bedrock, Amazon Quick Suite
- **Lakehouse and BI:** Apache Iceberg, Oracle, PostgreSQL, MicroStrategy
- **Hadoop ecosystem:** Cloudera, Hive, Oozie, HDFS, Unix
- **GenAI:** RAG, LLMs, FAISS, embeddings
- **Languages:** Python, Scala, SQL, Bash
- **DevOps:** Terraform (IaC since 2019), AWS CloudFormation (working knowledge), GitHub Actions, GitLab CI, Git, Docker (home lab), serverless design
- **Foundational knowledge:** Amazon Kinesis, Databricks, Snowflake, Amazon Redshift, dbt, Kubernetes
- **Domains:** AdTech, media measurement, CTV and OTT ad ratings, retail data onboarding

## EXPERIENCE

**Senior Data Engineer** | Lancesoft, client Warner Bros. Discovery | Remote **04/2026 - Present**

Client title at Warner Bros. Discovery: Data Engineer IV. Sole data engineer on the Market Research GenAI and MicroStrategy migration work, and technical lead for the Nordic middleware.

- Designed a serverless REST middleware on Amazon API Gateway for WBD's Nordic AdTech platform, replacing manual work by business staff. In pilot with **3 external vendors**. Mentored a junior developer on architecture and system design throughout the build.
- Proposed and built a RAG chatbot for the Market Research team over **500+** documents (FAISS, embeddings, 3B LLM). Prototyped on EC2, then moved to Amazon Quick Suite with chat personas and dashboards for production. Approved in UAT, go-live planned for 11/2026.
- Leading the migration of **about 400** MicroStrategy reports to Amazon Quick Suite, moving needed datasets from Oracle to Apache Iceberg. First 50 priority reports due within a month.
- Defined SLA, observability and cost-control standards with the Nordic engineering teams.

**Data Engineer (Member of Technical Staff II)** | Nielsen, DAR CTV | Bangalore **06/2023 - 02/2026**

- Owned the AWS data platform (EMR Serverless, S3, Airflow) behind Nielsen's Connected TV Digital Ad Ratings, covering ingestion, transformation and reporting for **500M+** daily events. Scope: 2 major daily pipelines at 1 to 1.5 TB each, plus 3 to 5 ad hoc pipelines of 50 to 300 GB.
- Cut compute cost by **40%** by moving hourly Spark jobs from 5-node r4.4xlarge EMR clusters (billed for spin-up and spin-down) to EMR Serverless, which charges only for job runtime. Also right-sized executors and tuned dynamic allocation.
- Upgraded **15+** Spark 2.x ETL jobs to Spark 3.x and enabled Adaptive Query Execution and dynamic partition pruning, making runs **at least 30% faster** with lower memory use.
- Redesigned the Airflow DAGs for CTV ingestion with retries and structured alerting, which improved SLA adherence and shortened recovery time after failures.

- Implemented data-quality checks and schema validation so bad upstream data stopped before it reached reporting.
- Mentored **4 junior engineers** and an intern on distributed design, Spark internals and Airflow, and helped experienced new joiners ramp up on the platform and start delivering quickly.

#### **Data Engineer II, Retailer Onboarding** | PDI Technologies | Remote 12/2021 - 06/2023

- Automated retailer onboarding with PySpark ETL pipelines and GitHub Actions CI/CD, cutting the time to onboard a new retailer from about 2 weeks to 2 to 3 days and halving deployment time.
- Migrated terabyte-scale retail datasets from on-premises to AWS (S3, EMR, Glue) with no downtime and no loss of data integrity.
- Re-platformed legacy ingestion for **50+** retailer pipelines onto serverless AWS services, improving fault tolerance and availability.
- Rebuilt Airflow orchestration, cutting manual pipeline intervention by **60%**.
- Added CloudWatch monitoring and structured logging, so the team caught problems earlier and resolved incidents faster.

#### **System Engineer, DAR CTV** | Tata Consultancy Services (Nielsen) | Chennai 08/2019 - 12/2021

- Ran production CTV analytics pipelines for Nielsen Digital Ad Ratings across **10+** streaming channels, keeping to SLAs and data accuracy targets.
- Designed the Airflow framework for the CTV rebuild, and wrote Scala utilities and Git-based CI/CD for JAR build, test and deployment.
- Automated workflows in Airflow, cutting manual intervention by **40%** and reducing processing errors.
- Onboarded the CTV infrastructure to Infrastructure as Code with Terraform in 2019, and kept using it as the standard way to provision since then.

#### **Assistant System Engineer, CORE DAR** | Tata Consultancy Services (Nielsen) | Chennai 08/2017 - 08/2019

- Built data flows between Nielsen's on-premises systems and AWS for Watch Analytics, so measurement data kept moving during the transition.
- Tuned SQL queries and ETL jobs, raising throughput by **25%** and getting downstream reports out sooner.
- Migrated a 10-person team from SVN to Git and set branching rules that reduced merge conflicts.

### PROJECT

---

#### **SathyaMool.ai, AI for Bharat Hackathon** | Hack2skill, prototype stage 06/2026

- Built a serverless AWS platform that uses generative AI to verify Indian property documents. Stack: Amazon Textract, Translate and Bedrock, with Lambda, API Gateway, SQS, DynamoDB, S3 and AWS CDK. Selected for the prototype stage of the national programme.
- [SathyaMool.ai repository on GitHub](#)

### CERTIFICATIONS

---

- AWS Certified AI Practitioner | Amazon Web Services | 01/2026 (valid to 01/2029)
- AWS Certified Solutions Architect - Associate | Amazon Web Services | 12/2024 (valid to 12/2027)
- Astronomer Apache Airflow certifications: Airflow 3 Fundamentals (07/2025), DAG Authoring and Fundamentals (10/2021)

### EDUCATION

---

B.Tech, Electrical and Electronics Engineering | SASTRA University, Tamil Nadu, India | 2017

### LANGUAGES

---

English (C1-C2, fluent), Tamil (native)

I consent to the processing of my personal data by the recruiting company for the purposes of current and future recruitment, in accordance with the GDPR.