

# RAHUL SOHANDANI

+1 201-423-0534 | [rsohandani@gmail.com](mailto:rsohandani@gmail.com) | [LinkedIn](#) | [Github](#)

## PROFESSIONAL SUMMARY

Accomplished Data Engineer with 3 years of experience, architecting large-scale data pipelines, real-time streaming infrastructures, and cloud-native solutions across **AWS, Spark, and Kubernetes** ecosystems. Specializing in designing and scaling distributed data solutions, automating real-time processing workflows, and **optimizing data pipeline performance and data warehouse architectures** to enhance scalability and reliability. Adept at applying critical thinking and problem-solving to complex data models, and dedicated to engineering robust backend platforms that transform data into actionable insights, empower business intelligence, accelerate decision-making, and drive analytics innovation.

## SKILLS

- **Languages:** Python, Java, Scala, SQL, Node.js, React.js, TypeScript
- **Big Data & Tools:** Apache Spark, Kafka, Airflow, Hadoop, Hive, Databricks
- **Cloud & DevOps:** AWS, Google Cloud Platform, Docker, Kubernetes, Prometheus, Grafana, Git, CI/CD
- **Frameworks & Libraries:** Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, LangChain, Transformers, MCP
- **Databases:** Snowflake, MongoDB (NoSQL), PostgreSQL, MySQL, Firebase, ArangoDB
- **Visualization & Platforms:** Tableau, Streamlit, REST API, Power BI, Django, Spring Boot, Flask

## WORK EXPERIENCE

### Stevens Institute of Technology

May 2024 - Jul 2024

#### Graduate Research Assistant

Hoboken, NJ

- Built geospatial data processing workflows using GeoPandas and Shapely, **optimizing 100 GB+ of GBFS, GIS, and Census data**, enabling faster equity-focused mobility policy decisions.
- Engineered large-scale ETL pipelines using **PySpark on GCP Dataproc** to ingest, process, and analyze 500 M+ geospatial data records, accelerating analytics for cloud-driven urban mobility insights.
- Designed scalable data architecture strategies, including **data modeling, partitioning, and efficient storage** formats within Dataproc pipelines, improving processing speed, and supporting downstream geospatial analytics.

### LTIMindtree

Jul 2021 - Aug 2023

#### Senior Software Engineer

Bangalore, India

- Developed **real-time ETL pipelines** leveraging **Apache Spark Structured Streaming, SQL, and Kafka**, reducing event stream processing time by 40%.
- Implemented IT Service Management (ITSM) ticket and Enterprise Service Management (ESM) alert convergence solutions, **improving data processing time by 90%**.
- Led the **migration of Spark and Kafka clusters** from **Azure Cloud HDInsight to an on-premise environment**, achieving a **70% reduction in operational costs** while enhancing data processing capabilities for analytics tasks.
- Configured **Prometheus and Grafana monitoring**, leading to real-time issue detection and enhanced system reliability.
- Architected a **predictive analytics project on time-series data**, achieving a 94.3% model accuracy to improve business forecasting.
- Orchestrated **end-to-end Kubernetes deployment of containerized machine learning applications**, enhancing system scalability, resilience, and uptime.

### Netwin Infosolutions

May 2019 - Jul 2020

#### Software Trainee

Nashik, India

- Built **automated data ingestion pipelines** to preprocess 10,000+ facial images, preparing structured datasets for MobileNet model training on AWS SageMaker.
- Engineered end-to-end **model training and deployment workflows on SageMaker**, achieving 91% accuracy in customer interest detection for targeted marketing optimization.
- Designed and provisioned **cloud infrastructure using Terraform**, automating deployment of 10+ data services and reducing manual setup time by hours to support scalable, reproducible ML workflows.
- Implemented **CI/CD pipelines to automate data ingestion and infrastructure provisioning processes**, reducing manual intervention, increasing deployment speed, and supporting agile data development practices.

## EDUCATION

Stevens Institute of Technology, Hoboken, NJ | MS - Computer Science | GPA 4.0/4.0

Aug 2023 - May 2025

Relevant Coursework: Big Data Technologies, Data Mining, Natural Language Processing, Data Structures

University of Mumbai, Mumbai, India | BEng - Computer Engineering | GPA 3.58/4.0

Jul 2017 - May 2021

## PUBLICATIONS AND ACHIEVEMENTS

- AWS Solution Architect Associate Certification | Amazon Web Services | [Certification Link](#)
- Identification of food oils and quality | IEMAICLOUD 2021 | SPRINGER | [Publication Link](#)