

# RONIT ROY

857-396-6089 | Brookline, MA, 02445 | [ronitr300701@gmail.com](mailto:ronitr300701@gmail.com) | [LinkedIn](#) | [Github](#) | [Portfolio](#)

## SUMMARY

Data Engineer with experience in building scalable ETL pipelines, real-time data systems, and cloud-native analytics solutions. Skilled in Python, SQL, Airflow, Spark, and AWS. Strong background in data modeling, machine learning workflows, and delivering insights through automation and dashboards.

## EDUCATION

### Boston University

*Masters of Science in Applied Data Analytics*

Boston, MA

*Sept. 2023 – Dec. 2024*

- Coursework: Advanced Machine Learning, Advanced Database Management, Data Mining

### SRM Institute of Science and Technology Chennai

*Bachelor of Technology in Computer Science*

Chennai, India

*Jul. 2019 – Apr. 2023*

- Coursework: Probability and Statistics, Data Structure and Algorithms, Data Visualization

## EXPERIENCE

### GLOB S Research Lab

*Research Assistant*

Boston, Massachusetts

*Oct. 2023 - May. 2024*

- Designed and deployed scalable ETL pipelines using Airflow and PostgreSQL, improving data processing speed by 25%, enabling faster research insights.
- Built and containerized NLP pipelines (NER) with Python and Docker, reaching 95% entity extraction accuracy, enhancing healthcare document mining.
- Reduced average query execution time by 30% through SQL query optimization and indexing strategies.
- Delivered actionable visualizations to stakeholders via dashboards, facilitating data-driven decision-making.

### HighRadius

*Machine Learning Intern*

Bangalore, India

*Jan 2022 – Apr. 2022*

- Engineered CNN models for financial anomaly detection, increasing fraud detection accuracy by 30%, reducing risk in billing systems.
- Created automated ETL workflows integrating Python and Snowflake, streamlining model training and data ingestion.
- Developed and deployed Flask APIs in Docker containers for seamless integration with front-end applications.
- Implemented CI/CD pipelines using Jenkins, reducing model deployment time by 15% and improving development efficiency.

## PROJECTS

### Spotify Data Pipeline

*Python, Snowflake, AWS, Airflow*

- Designed an end-to-end ETL pipeline processing over 1M records daily, storing structured data in Snowflake.
- Scheduled DAGs with Airflow to automate transformations, reducing processing time by 40%.
- Integrated AWS Lambda & Snowpipe for real-time data ingestion and fault-tolerant stream handling.

### Customer Data Lake Design

*Python, Spark, AWS S3*

- Built a robust data lake architecture handling 10+ TB of historical retail data using Spark and S3.
- Leveraged partitioning and compression to cut storage costs and access latency by 35%.

### E-commerce Recommendation System

*Python, Flask, TensorFlow, PostgreSQL*

- Developed a collaborative filtering-based recommendation system using TensorFlow, improving user engagement by 20%.
- Deployed API endpoints using Flask for seamless integration with front-end systems.
- Optimized PostgreSQL queries for real-time recommendation generation.

### Advanced ETL for Retail Analytics

*Airflow, AWS, Kafka, Snowflake, Redshift*

- Constructed a dynamic ETL pipeline handling 500K+ records/hour using Apache Kafka and Snowflake.
- Enabled near real-time analytics through AWS Redshift warehouse optimization.
- Implemented automated error handling in Airflow to ensure data consistency.

## TECHNICAL SKILLS

**Programming & Scripting:** Python, SQL, R, Java, Shell Scripting, PL/SQL

**Data Engineering & Big Data Tools:** ETL Pipelines, Apache Airflow, Apache Spark, Kafka, Hadoop, Flink, DBT

**Databases & Warehousing:** PostgreSQL, MongoDB, MySQL, Snowflake, Big Query, Redshift

**Cloud Platforms:** AWS (S3, Lambda, Snowpipe, Redshift), GCP, Azure

**DevOps & Automation:** Docker, Kubernetes, Jenkins, CI/CD, Git

**Machine Learning & Analytics:** Scikit-learn, TensorFlow, PyTorch, NLP(NER), Data Modeling(Kimball, OBT), Performance Tuning

**Visualization & Reporting:** Tableau, Power BI, AWS QuickSight

**Soft Skills:** Problem Solving, Team Collaboration, Clear Communication, Agile Methodologies