Abhishek Mehra

8572347103 | me.abhishek18@gmail.com | LinkedIn | Github |

PROFESSIONAL EXPERIENCE

UMass Chan Medical School   **October 2024 – Present**
Software Engineer   **Massachusetts**

- Spearheaded the deployment of a secure, serverless Amazon Redshift architecture using Terraform IaC, enforcing VPC isolation, KMS encryption, and granular IAM policies to achieve 100% compliance with enterprise security standards.
- Architected and optimized an OMOP CDM-aligned Redshift data warehouse, integrating EHR and research datasets from 5 studies, empowering 10+ researchers with faster insights.
- Engineered a fault-tolerant, production-grade ETL pipeline leveraging Apache Airflow (MWAA), AWS Glue, and Redshift, incorporating dynamic schema evolution, automated alerting (SNS), and processing 1TB of clinical research data daily.
- Designed and implemented a PII detection framework using AWS Comprehend and regex-based logic, ensuring HIPAA compliance and mitigating data breach risks across sensitive datasets.
- Orchestrated multi-study workflows with Apache Airflow (MWAA) to monitor study-specific S3 buckets, transform data using Python, and load into Amazon Redshift schemas, improving data ingestion efficiency across 5+ studies by 30%.
- Led the development and deployment of a Patient Summary Report Generator using Next.js, TypeScript, and SQL Server on AWS EC2, streamlining data retrieval for doctors and enhancing patient care through rapid, comprehensive insights.

Software Engineer Co-Op   **July 2023 – December 2023**

- Led the design and implementation of an event-driven notification system using AWS Lambda, SNS, and Redshift, delivering critical updates to 50,000+ research participants across 5+ studies .
- Spearheaded the development of a Streamlit-based business intelligence (BI) dashboard for 5 clinical studies, enabling real-time monitoring of participant engagement metrics and driving data-informed recruitment optimizations.
- Directed the creation of a Python SDK with comprehensive unit tests and CI/CD integration, automating data extraction and transformation from REST APIs, optimizing pipeline workflows, and saving 10 man-hours weekly.

Torm Shipping   **August 2018 – September 2021**
Software Engineer   **India**

- Architected and implemented a scalable ETL pipeline using Apache Spark, AWS EMR, and Redshift, processing vessel telemetry data daily to deliver actionable insights into fuel efficiency and operational performance.
- Established a CI/CD-driven testing framework with AWS ECR and Docker, achieving 95% ETL test coverage and reducing deployment errors by 70%, significantly enhancing pipeline reliability.
- Developed an innovative automated pipeline integrating OCR (AWS Textract), NLP (Huggingface), machine learning (scikit-learn) to extract, structure data from Material Safety Data Sheets into CSVs, improving data processing efficiency by 40%.
- Designed and implemented ETL pipelines to aggregate and preprocess data from diverse Excel sources, and creating an interactive Tableau inventory dashboard that reduced stockouts by 5%.

TECHNICAL SKILLS

- **Databases & Programming Languages:** Python, NoSQL, PostgreSQL, MySQL, MongoDB, SSIS, SQL Server
- **Cloud:** AWS (S3, Lambda, DynamoDB, Step Functions, Athena, QuickSight, Glue, EMR, Redshift, CDK, EC2, SNS, MWAA)
- **Development Tools & Methodologies:** Git, Terraform, Jenkins, Tableau, Power BI, DBT, PySpark, Apache Airflow, Apache Spark, Docker, AWS Textract, Next.js, TypeScript, Streamlit, Huggingface, scikit-learn, Tesseract
- **Practices:** Agile, Continuous Integration/Continuous Deployment (CI/CD), API Development

EDUCATION

Northeastern University
Masters in Data Analytics
Birla Institute of Technology
Bachelor of Technology in Engineering