

Syed Vaseem

New Jersey | +1-551-208-3752 | syevasee@gmail.com | www.linkedin.com/in/vaseem-syed-977777224

Education

Pace University

Master of Science

Jan 2023 - Dec 2024

Jawaharlal Nehru Technological University

Bachelor of Technology

Jul 2016 - May 2019

Relevant Experience

Amazon

Data Engineer

Aug 2019 - Jun 2022

Hyderabad, India

- Designed, optimized, and maintained scalable data pipelines using Python and Apache Spark for ETL processes, incorporating best practices in data integrity, testing, and documentation to support comprehensive data solutions in a high-volume e-commerce environment.
- Managed large-scale data warehousing on Snowflake with integrated SQL queries, reinforcing data modeling and schema design best practices to consistently improve query performance and overall system efficiency.
- Conducted comprehensive data modeling and implemented distributed systems strategies to create robust and scalable data structures, supporting batch processing and enhancing data reliability in Big Data environments.
- Performed thorough data cleansing and transformation using Pandas and PySpark, enabling deep dive investigations into data discrepancies and ensuring high accuracy for detailed analytics.
- Developed and deployed Python-based data validation frameworks across multiple data sources, upholding rigorous standards in data integrity, validation, and documentation to support data-driven decision making.

Customer Churn Prediction and Analysis

Customer Churn Prediction and Analysis

Telecommunications Company

- Developed a machine learning pipeline to predict customer churn for a telecommunications company. The project involved data collection, cleaning, and transformation, followed by model development using machine learning algorithms to enable targeted retention strategies.
- Improved model accuracy to 85% and recall to 87%, leading to targeted retention efforts that decreased monthly churn by 10%

Real-Time Fraud Detection System

- Developed a fraud detection pipeline leveraging AWS Kinesis for real-time data streaming and AWS Sagemaker for ML model deployment, achieving early fraud detection.
- Reduced fraud detection time by 60%, resulting in enhanced data integrity and customer trust.

Real-Time Data Pipeline for E-Commerce Analytics

E-Commerce Platform

- Designed and implemented a scalable real-time data pipeline to process and analyze clickstream data from an e-commerce platform. Used Apache Kafka for data ingestion, Spark Streaming for real-time processing, and stored transformed data in AWS Redshift for business intelligence and reporting.
- Reduced reporting latency from 24 hours to under 5 minutes, enabling near real-time marketing decisions and improving customer targeting by 15%.

Technical Skills & Projects

- **Programming Languages:** Python, Java, SQL, JavaScript, R
- **Cloud Platforms:** AWS, Azure DevOps
- **Data Tools:** Tableau, Azure Data Factory, SSIS, ETL Pipelines, Docker, Git
- **Databases:** MySQL, MongoDB, DynamoDB, Neo4j, Azure Databricks
- **Libraries & Frameworks:** Pandas, NumPy, Scikit-Learn, PySpark, SQLAlchemy