

VARSHIL R. MEHTA

New York City, NY • varshilmehta99@gmail.com • (201) 755-4662 • [LinkedIn](#) • [GitHub](#) • [Portfolio](#)

SUMMARY

Data Engineer with expertise in designing and optimizing real-time data pipelines for customer experience analytics, transforming raw data into actionable business metrics to drive personalized user engagement and operational efficiency. Skilled in building scalable data architectures and cloud-native solutions on AWS (S3, Redshift, Lambda, Aurora) and big data platforms (Presto, Flink, Spark). Experienced in leveraging Generative AI, advanced analytics, and machine learning to enhance customer insights, recommendation systems, and sales forecasting. Strong background in automating ETL workflows using Informatica, optimizing data infrastructure for self-service modelling.

WORK EXPERIENCE

Metropolitan Transportation Authority (MTA) | New York, NY | [Data Analyst](#) Jun 2024 – Dec 2024

- Engineered automated Power BI dashboards to surface key operational insights, improving data-driven decision-making by 40%.
- Redefined ETL workflows to migrate legacy datasets to PostgreSQL, increasing query performance and data accessibility by 30%.
- Automated data validation using Python and SQL, ensuring 99% accuracy and minimizing manual intervention.

Powerweave Software Services Pvt. Ltd. | Mumbai, India | [Data Engineer](#) Feb 2023 – May 2023

- Designed and deployed an ETL pipeline using PySpark, cutting data processing time by 50% and enhancing real-time analytics.
- Refactored MySQL data models and indexing strategies, enabling scalable reporting and reducing query latency by 60%.
- Integrated AWS S3 to enhance data storage and retrieval efficiency, reducing latency by 40% and improving system resilience.

Deloitte – USI | Mumbai, India | [Associate Analyst](#) May 2021 – Dec 2022

- Designed adaptable data architectures supporting 70+ business units, ensuring seamless integration of evolving business processes and reporting needs.
- Led a data quality initiative, eliminating duplicate records and increasing sales conversion rates by 25%.
- Enhanced SQL-based data transformation workflows, reducing data processing time by 30%.

YLP (Your Learning Platform) | Mumbai, India | [Data Analyst](#) Jul 2020 – Apr 2021

- Conducted sales performance analysis, increasing revenue by 20% through actionable insights.
- Automated SQL-based reporting, reducing manual intervention by 50% and improving data accuracy.
- Created interactive business intelligence reports, enhancing reporting efficiency by 30%.

TECHNICAL SKILLS

Languages: Python (PySpark, Pandas), SQL (PL/SQL, SparkSQL, HiveQL, DDL), R, SAS, Scala

Data Visualization & Analytics: Tableau, Power BI, Git, MS Excel (Advanced), Business Intelligence

AI/ML: Generative AI (LLMs, Synthetic Data), Clustering, Classification, NLP, LSTM, Scikit-Learn, Recommendation Systems

Big Data & Streaming: Presto, Flink, Apache Spark, Kafka, Airflow, Hadoop, Hive, Databricks

Cloud & Infrastructure: AWS (S3, Lambda, Redshift, Aurora, DynamoDB), Docker, Kubernetes, Terraform

Data Engineering & ETL: Informatica, SSIS, DataStage, ODI, BODI, Snowflake, PostgreSQL, MySQL, MongoDB, Cassandra

Data Modeling & Analytics: A/B testing, adaptive schema design, evolving data structures, user behavior analytics

Soft Skills: Effective verbal and written communication, Problem-solving, critical thinking and Troubleshooting

ACADEMIC PROJECTS

eCommerce Recommendation System (AWS, Spark, Python, Clickstream Data)

- Built real-time recommendation engine leveraging customer behavior data, increasing user engagement by 35%.
- Processed 5M+ user interactions daily using Spark Streaming, reducing pipeline latency by 50%.
- Implemented personalized product recommendations, increasing average order value (AOV) by 20%.

Football Data Analytics | Azure End to End Data Engineering Project ([Apache Airflow](#), [Azure](#), [Tableau](#))

- Developed a scalable pipeline analyzing audience engagement trends in football, enabling interactive insights for sports analysts.
- Migrated and processed data in Azure Data Lake and Azure Synapse, cutting query execution time by 50%.
- Built Tableau dashboards to visualize football statistics, improving analyst decision-making.

Olympics Data Analytics (SQL, Pandas, AWS Athena)

- Predicted medal counts (R^2 : 0.82) using GDP & historical data, aiding 5+ national teams. Scraped athlete data, enriching datasets by 40% for sponsorship insights.

COVID-19 End-to-End Data Engineering Pipeline (AWS, Python, Spark, Kafka)

- Designed a real-time data pipeline to process and analyze COVID-19 cases globally, integrating API data ingestion, ETL, and visualization.
- Utilized Kafka & Spark Streaming to handle live case updates, enabling real-time analytics with a 50% faster refresh rate. Built interactive Tableau dashboards for trend analysis, assisting policymakers in resource

EDUCATION

Pace University – School of Computer Science

MASTER OF SCIENCE IN DATA SCIENCE

S.V.K.M Usha Pravin Gandhi College of Science, University of Mumbai

BACHELOR OF SCIENCE IN INFORMATION TECHNOLOGY

New York, USA

Dec 2024

Mumbai, India

Oct 2020