

Hritik Rakesh Singh

Boston, MA | 8572042509 | hritiksingh01@outlook.com | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Data Engineer with 3+ years of experience building scalable ETL pipelines across AWS and Azure environments. Proven track record of optimizing large-scale datasets (50M+ records/month) and improving pipeline performance, data reliability, and query efficiency. Experienced in enterprise and fast-paced startup environments, collaborating cross-functionally to deliver analytics-ready data solutions.

EDUCATION

Northeastern University

Master of Science, Computer Software Engineering

Jan 2024 - Apr 2026

Boston, MA

University of Mumbai

Bachelor of Engineering, Information Technology

Jun 2017 - May 2021

Mumbai, India

TECHNICAL SKILLS

- **Programming & Frameworks:** Python, Java, Pandas, PySpark, Scikit-learn, SQL, T-SQL, MSSQL, MySQL
- **Cloud Architecture:** AWS, Lambda, Kinesis, S3, Glue, Redshift, CloudWatch, Azure, Azure Functions, Blob Storage, Azure Monitor
- **Modern Data Stack & Warehousing:** Snowflake, Databricks, Lakehouse Architecture, Dimensional Modeling, SCD Type 2, Parquet/Avro
- **Data Engineering & Orchestration:** Apache Airflow, Azure Data Factory, AWS Step Functions, dbt, ETL/ELT Pipelines
- **Visualization & Analytics:** Power BI, Tableau, Alteryx, Talend, MS Excel
- **DevOps & DataOps:** CI/CD Pipelines, Azure DevOps, Git, Docker
- **Coursework:** Database Management and Design, Advanced Data Architectures for BI, Data Science
- **Certifications:** SnowPro Associate – Snowflake Certification, Google Data Analytics Certification, Databricks Fundamentals Accreditation

WORK EXPERIENCE

GymIn Inc

Data Engineer Intern

Sep 2025 - Dec 2025

Delaware, USA

- Architected real-time streaming pipelines leveraging **AWS Lambda, Amazon Kinesis, and Amazon S3**, processing 1M+ high-frequency events per week with less than 5-second latency, enabling operational analytics and scalable cloud-based telemetry processing.
- Developed robust ETL workflows using **AWS Glue** and **PySpark**, transforming raw event data into analytics-ready datasets, increasing data accuracy by 25% while improving reliability and reducing manual intervention across reporting environments.
- Optimized storage architecture and query performance within **Amazon Redshift**, implementing schema redesign and indexing strategies that reduced dashboard execution time by 30%, improving visibility into equipment usage trends for partner gym owners.
- Established monitoring and validation frameworks using **AWS CloudWatch** and SQL-based data quality checks, decreasing inconsistencies by 20% and ensuring reliable operational insights for business and external stakeholders.

LTIMindtree

Data Engineer

Jun 2021 - Dec 2023

Mumbai, India

- Engineered ETL pipelines in **Azure Data Factory** that processed 50M+ records monthly, cutting runtime by 40% and supporting analytics for three business units within a modern cloud architecture
- Led migration of on-premise **SQL Server workloads** to **Azure SQL** and **Azure Data Lake**, boosting query performance by 30% and providing scalable, cost-effective cloud storage and transformation for enterprise reporting
- Delivered 10+ executive dashboards in **Power BI** utilizing advanced **SQL** and **DAX**, driving KPI visibility, operational monitoring, and cross-functional performance insights for senior leadership teams across global business operations.
- Strengthened production reliability by optimizing ServiceNow incident workflows and delivering L2/L3 support for **.NET** and **Java** systems, achieving 99.9% uptime while accelerating issue resolution by 30% and partnering with users across 7 global business units on 4 continents.

Bhaktivedanta Hospital & Research Institute

Data Analyst Intern

Dec 2018 - Jan 2019

Mumbai, India

- Analyzed structured Healthcare datasets (clinical, operational reporting) using advanced **SQL** and **Tableau**, improving reporting efficiency and increasing stakeholder satisfaction by 20% through performance-optimized queries and enhanced dashboards for inpatient and outpatient cancer care monitoring.
- Streamlined internal reporting workflows by redesigning database logic and automating recurring analytics processes, reducing manual data preparation time by 15% while strengthening data accuracy and cross-departmental accessibility.
- Transformed structured healthcare datasets into standardized reporting models using **SQL**, improving data integrity and enabling reliable analytics for inpatient and outpatient clinical operations.

PROJECTS

Vehicle Collision Analysis [Github](#)

- Integrated 3M+ collision records into a centralized warehouse using **Azure Data Factory** and **Talend**, implementing **Parquet** optimization to enhance processing efficiency, reduce storage footprint, and support scalable downstream analytics workloads
- Designed scalable dimensional models with **SCD Type 2** and delivered KPI-driven dashboards in **Power BI** and **Tableau**, enabling historical trend analysis, interactive reporting, and data-driven decision support for stakeholders.

Retail Sales Lakehouse Pipeline [Github](#)

- Built an incremental pipeline ingesting **CDC** streams from an operational database and file feeds from **AWS S3** into **Databricks Delta Lake**, enabling reliable upserts and scalable analytics-ready storage.
- Modeled curated layers with **dbt** into a **One Big Table** and downstream dimensional marts with embedded **data quality checks**, orchestrated end-to-end in **Apache Airflow** for repeatable, monitored production runs.