

Shardul Chavan

shardulc36@gmail.com / (857)313-5138 / Boston, MA / [linkedin/shardulchavan36](https://www.linkedin.com/in/shardulchavan36/) / [github/shardulchavan](https://github.com/shardulchavan)

Data Engineer with 2 years of experience architecting data warehousing solutions, optimizing data preprocessing pipelines, and performing Semiconductor Data Analytics. Skilled in building regression and classification models and transforming raw data into actionable insights to drive data-informed decision-making.

EDUCATION

Northeastern University, Master of Science in Information Systems, **GPA: 3.72/4.00** September 2022 - December 2024
Relevant Coursework: Big Data Engineering and Intelligence Analytics, Data Science Engineering

Mumbai University, Bachelor of Computer Science July 2019 - June 2022

PROFESSIONAL EXPERIENCE

Data Engineer - Analytics Intern, Skyworks Solutions Inc., Boston, MA January 2024 - June 2024

Tech Stack: Python, SQL, Docker, Azure SQL Server, Databricks, PowerBI

- Constructed 25+ Airflow data pipelines to extract, transform, and load (ETL) data from Azure Data Lake, SQL Server, and flat files into SQL Server Data Warehouse
- Elevated historical data tracking by 30% through deploying Slowly Changing Dimensions (SCD) and advanced transformations leveraging Databricks SQL and PySpark APIs
- Collaborated with engineering teams to create dimensional data models and **PowerBI dashboards** for product analysis, KPIs and prototype builds, maximizing efficiency across 3+ teams in Business unit
- Engineered pipeline to transform product test data into insights, saving RF engineers 5+ hours of manual work weekly
- Implemented validation checks, and retry mechanisms enhancing fault tolerance and data accuracy by 20% at scale
- Managed Agile workflows and task tracking using Kanban, ensuring effective project execution

Graduate Teaching Assistant, Northeastern University, Boston, MA September 2023 - December 2023

Tech Stack: Python, Pandas, TensorFlow, Numpy, HDFS, Predictive Analytics & Validation

- Facilitated weekly lab sessions for 40+ students, teaching **data science techniques** and A/B tests, while providing support during office hours, proctoring exams, and grading assignments
- Benchmarked **GenAI** models (LLaMA, GPT-3), using BLEU and ROUGE scores to improve model accuracy by 5%
- Led ML pipeline using **PySpark** on **Hadoop** to optimize data ingestion, perform EDA, apply PCA for dimensionality reduction, and automate feature selection with correlation analysis and chi-squared tests
- Built and fine-tuned classification models, including **KNN, Random Forest, and XGBoost**, enhancing predictive accuracy by 25% and strengthening risk assessment and pricing strategies for Prudential Life Insurance

Data Analyst Intern, Accion Labs, India January 2023 - July 2023

Tech Stack: JavaScript, Snowflake, MySQL, Natural Language Processing

- Designed ELT scripts to process 10GB+ daily data from **Snowflake** to **MySQL**, leveraging advanced **SQL** (CTEs, window functions) to streamline data access for analytics dashboards
- Developed **REST API** integrations with large language model APIs into ServiceNow, boosting virtual chatbot performance by 20%, and architected vector databases to enhance NLU in knowledge management

SKILLS

Programming Languages	Python, Java, SQL, PL/SQL, R, Scala, VBA, C#, shell scripting (UNIX/Linux)
Data Analysis	Data Visualization (Tableau, PowerBI, Matplotlib, Seaborn), Data Mining
Data & Big Data Tools	Apache Spark, Hadoop, Kafka, Hive, Dynamo dB, NoSQL, MongoDB, Redis, Jira, Presto, dbt
Cloud & DevOps	AWS, GCP, Azure, Docker, Kubernetes, Jenkins, Terraform

PROJECTS

AWS-Based YouTube Data Analytics Pipeline (EC2, IAM, Athena, Redshift) June 2024 - August 2024

- Architected scalable ELT pipeline to process large YouTube data, using AWS S3 for storage and AWS **Glue** for cataloging, reducing data preparation time by 30%
- Automated AWS **Lambda** workflows and refined SQL transformations in Athena, increasing data processing efficiency, and integrated QuickSight dashboards, improving insights into YouTube trends and performance for 1,000+ videos

Snowflake NLP Query & Analytics System (NLP, LLMs, Git Actions)  September 2023 - October 2023

- Built AI-powered SQL generation app using **LangChain**, transforming user queries into **ad-hoc SQL** statements, and providing real-time analytics on **NFL sponsorship data** from Snowflake databases
- Integrated 4+ databases into **Snowflake** using UDFs, views, and stored procedures, and implemented **CI/CD** pipelines with Git Actions to streamline development and ensure seamless deployment on **GCP**