

# RAKESH CHOUDHURY

📍 Boston, MA | ✉ rakeshchoudhury6@gmail.com | ☎ 857.210.8264 | 🌐 <https://linkedin.com/in/rakesh--choudhury>  
📁 [github.com/rakesh-choudhury](https://github.com/rakesh-choudhury) | 🌐 <https://rakesh-choudhury.github.io>

## EDUCATION

**Northeastern University,**

*Master of Science in Information Systems*

**September 2019 - May 2021**

**Kalinga Institute of Industrial Technology,**

*Bachelor of Technology, Electronics and Telecommunication*

**May 2011 - May 2015**

## TECHNICAL SKILLS

**Programming skills:**

SQL, Python, Linux

**Data Science Libraries:**

Numpy, Pandas, Tensorflow, Pytorch, Matplotlib, Plotly(DASH), Seaborn

**Business Intelligence tools:**

Tableau, Power BI, Snowflake, Flourish

**Cloud:**

AWS (EC2, S3, RDS, Redshift), Heroku, Azure DevOps

**Tools:**

Jira, Git, Alteryx (ETL), Talend (ETL), Snowflake

**Databases:**

SQL Server, Mongo DB, PostgreSQL, MySQL, Oracle

## WORK EXPERIENCE

**ZS Associates (Data Engineer)**

**August 2021 - June 2023**

- Reduced manual data download and export through automated data pipeline using sFTP and **AWS S3** as source and destination for data drop and store, saving the client approximately **\$7,000**.
- Automated **ETL** pipeline with **Informatica**: API data to **Snowflake**. Saved **\$5k** monthly on manual data transfers.
- Developed base tables in **Snowflake** for **15+** business use cases, fulfilling target KPIs and metrics for PowerBI dashboard reporting.
- Built an upgraded ETL pipeline using **SparkSQL**, **PySpark**, **AWS EMR**, and **Azure DevOps**, replacing traditional **Hadoop HDFS**, resulting in a 40% increase in processing efficiency and scalability.

**Serole Technologies (Data Engineer)**

**April 2018 - November 2018**

- Achieved **0.81** F1 score in random forest model for customer classification.
- Developed **ETL** pipelines pulling **100 GB** data for analysis while meeting audit requirements.
- Reduced policy ad spend by **\$5k** each through A/B testing analysis.
- Increased revenue by **100%** through PowerBI visualization enabling data-driven decisions for customer acquisition and retention.

**Accenture (Data Engineer)**

**June 2015 - January 2018**

- Automated **SQL** Query replication to **10+** data marts using Python.
- Led team in designing star schema, reducing query time by 50%.
- Parallelized **ETL** with **SCD Type 2**, loading data **9x** faster for Tableau reports.
- Improved database and query performance by nearly **3x** with **SQL indexing** and query optimization.
- Won Ace award, saving highest-paying client **\$160k annually**.

## ACADEMIC PROJECTS

**IMDB Data Warehousing Project (Amazon S3, PowerBI, Data Model, Alteryx, Dimensional model, BigQuery)**

- Built Talend **ETL** jobs to cleanse and load **20GB** of data from **AWS S3** to **MySQL** Data Warehouse, boosting data quality and improving business insights.
- Integrated facts and dimensions into **MySQL** DB with optimized **ETL** workflows, reducing memory usage by up to **3GB** for improved database performance.
- Developed interactive Power BI dashboards with **100+** attributes, measures, and calculations, delivering actionable insights and driving data-informed decisions.

**Airport Database Management System (Tableau, RDBMS, Business Intelligence, AWS RDS, MSSQL, Talend)**

- Engineered **Talend** data pipeline that loaded dimensional data into **SQL Server** from **Amazon RDS**, reducing memory usage by **1.5 GB**.
- Utilized **SQL** query joins, functions, and procedures to create detailed reports from **4 GB** of data, streamlining business operations.
- Integrated disparate data sources into SQL views for a Tableau dashboard with **15+** parameters, offering deep insights into Airport operations.

**Brand Promotion on YouTube Channels (NLP, Logistic regression, k-means, AWS EC2, Flask, Statistical Methods)**

- Engineered YouTube channel classification model using **Scikit-Learn's K-means** and **Logistic Regression** on **4 GB** dataset.
- Developed trend metric with **NLP** sentiment analysis and finance parameter, achieving **225%** ROI for brand influencing.
- Deployed **Streamlit** web app with integrated **Python** code on **AWS EC2** instance for efficient data analysis and visualization.