

# Priyank Thakur

Jersey City, NJ | [priyankthakur3@gmail.com](mailto:priyankthakur3@gmail.com) | 551-331-7841 | [LinkedIn](#) | [GitHub](#) | [priyankthakur3.vercel.app](https://priyankthakur3.vercel.app)

## EDUCATION

### Stevens Institute of Technology, Hoboken, NJ

Master's in Computer Engineering

Dec 2023

GPA: 3.70/4

Coursework: Web Programing, Data Acquisition Modeling and Analysis, Engineering Python, Deep Learning

### Mumbai University, Mumbai, India

Bachelor of Engineering in Computer Engineering

May 2018

GPA: 7.02/10

## EXPERIENCE

### Teradata Corporation, Mumbai, India

Aug 2019 – Jan 2022

#### Data Engineer

- Designed & owned end-to-end ETL Pipeline using SQL & Java Scripts to clean, ingest & transform 8M+ daily sales records from AWS S3 into Data Warehouse utilizing Airflow for job orchestration
- Loaded data into HDFS using Sqoop and created queries in Hive, Impala and Spark for loading, manipulating, and cleaning data. Monitored query execution on YARN using Ambari
- Conducted code reviews for ETL scripts, implemented best practices & data quality rules using AWS Glue Studio for null & duplicate rows, contributed to existing regression & performance test cases
- Acted as Leader for team of 4 members, build & maintained SOP's for addressing common & new occurring data issues
- Collaborated with sales team during system design phase, offering expertise on performance optimization, which played a pivotal role in securing 3-year contract extension

### Teradata Corporation, Mumbai, India

Aug 2018 – Jul 2019

#### Junior Technical Associate

- Reduced multiple ETL pipeline runtime by 40% leveraging Teradata TASM (Teradata Active System Management) for resource & session management, applied filters & throttles to avoid inefficient ad hoc queries
- Improved reporting queries by 20% implementing Join Indexes for pre-computation of complex joins for high-accessed rows eliminating need for Full table Scan
- Delivered dashboards for Performance KPI overview, enhancing Data Warehouse domain-specific usage analytics such as query response, data ingestion rate, table/views access
- Built Shell Scripts in Linux for Automating task like daily archive job & daily system performance report
- Enforced Data Security using Role-based access control mechanism & Instated 'Auto-Access' SQL procedure for access control, aligning with Data Governance guidelines

## PROJECTS

### Sentiment Analysis on Trending Twitter Topics (Python, Apache Spark, Airflow, React.js, Express.js, MongoDB)

- Used Airflow to orchestrate ETL pipeline for sentiment analysis on trending topics tweets utilizing Vader Model
- Used Twitter API & PySpark for scraping popular tweets & store transformed dataset into MongoDB
- Created comprehensive Trend Analysis dashboard using React.js, providing visual insights into trends & patterns

### Friends Dine (Java, JavaScript, MongoDB, H2, Spring Boot, Google Firebase, React.js)

- Developed Full stack web application with RESTful backend using Spring Boot allowing users to collaborative choose restaurant among friends through voting based system
- Implemented voting-based prioritization system using Boyer-Moore Majority Voting algorithm
- Utilized H2 database for caching previous Yelp API responses & Google Firebase for push notification about finalized restaurant choices post voting phase

### Campus Compass (JavaScript, Express.js, turf.js, MongoDB, HandleBars)

- Collaborated within 4-member team throughout Software Development Life Cycle (SDLC) to design, develop, debug and optimize a JavaScript web application, fostering seamless student navigation to halls
- Built backend service using express.js overseeing 10K user session management, authentication & authorization
- Utilized Leaflet, turf.js, tota11y & jQuery for client-side data validation, render & ensure application is accessibility compliant

## SKILLS

### Programing Languages

Java, Python, SQL, Shell Script, JavaScript, HTML, CSS

### Frameworks

Express.js, React.js, Next.js, Node.js, Apache Spark, Hadoop, Tailwind CSS

### Database

Teradata Database, Oracle, MySQL, SQL Server, MongoDB, Snowflake

### Cloud Technologies

GCP, AWS, S3, Lambda, EC2, Glue

### Tools

PowerBI, Tableau, Git, Docker, Airflow, Visual Studio Code, IntelliJ IDE, Atom

### Certifications

Teradata Vantage Certified Associate, Teradata Vantage Certified Developer