

Priyank Thakur

priyankthakur3@gmail.com | +1 551-331-7841 | Jersey City, NJ | [linkedin.com/in/priyankthakur3](https://www.linkedin.com/in/priyankthakur3) | github.com/priyankthakur3

SUMMARY

Data Engineer with 4 years of experience in designing & optimizing data pipelines within Spark & Snowflake. Led initiatives that reduced infrastructure costs by 40% & enhanced data processing speeds by 25%

EXPERIENCE

Data Engineer, JerseySTEM, Florham Park, NJ Feb 2024 - Present

- Streamlined new member ramp-up time by 30% through launching an on-boarding program using Google Classroom, for SQL & Airflow Fundamentals training
- Deployed Data Validation SQL Scripts for null & duplicate record exclusion, email & cross fields validation
- Established Pull Request Review rules on Gitlab & established standardized data engineering processes adopting Agile methodologies
- Reduced infrastructure cost by 30% through Alteryx workflow migration to Kubernetes-managed Airflow container
- Minimized post-release bugs by 30% implementing impact analysis, CI/CD pipeline for Production deployment

Data Engineer, Teradata Corporation, Mumbai India Aug 2019 - Jan 2022

- Contributed in GDPR Compliant Data Lakehouse creation for Inventory & Sales Management integrating SQL Server transactional databases, semi-structured files (XML, JSON) & third-party data feeds
- Decreased infrastructure cost by 40% by migrating legacy POS Sales ELT pipeline from on-perm Teradata to Snowflake SnowPipe, leveraging Kinesis, Lambda, S3 & enforced data lineage for entire data life-cycle
- Optimized Sales ingest Spark SQL job by leveraging broadcast joins & partitioning strategies for large data-sets, reducing shuffle operation for data skewness improving job execution time by 30%
- Adopted Medallion architecture to trim data incidents by 60%, optimizing data models in collaboration with business analysts & deployed fault-tolerant DAG with upsert operations on Product UPC with operator timeouts to meet SLAs
- Secured 3 year contract extension by collaborating with sales team on system design and performance improvement

Junior Technical Associate, Teradata Corporation, Mumbai India Aug 2018 - Jul 2019

- Increased Teradata DWs performance by 25%, through session limit, throttles & workload groups allocation
- Saved 10M CPU cycles through SQL Query tuning using execution plan analysis, indexes updates, & query refactor
- Enforced Data Governance policies by revamping existing Role-based access control rule-sets
- Built SQL procedure for granting temporary access & logging access violation improving audit trails by 30%
- Automated SOX access review & data cleanup emails using PowerShell scripts cutting approval process by 40%
- Developed Tableau Dashboard for reporting object usage by each domain group, access violation & space utilization

PROJECTS

Citibikes Availability Analysis (Python, PySpark, Kafka, Databricks, dbt, Snowflake, Streamlit)

- Developed Trend Analysis for NYC Citibikes availability using Confluent managed Kafka Cluster & Snowflake
- Deployed Producer script to trigger on stations status update & Spark stream for processing queued response
- Utilized Airflow to trigger dbt scripts to populate dimension & fact table, visualized bike availability in streamlit

Sentiment Analysis on Twitter Topics (Python, AWS Glue, AWS Athena, PowerBI, S3)

- Leveraged AWS Glue to build Data Pipelines for performing Sentiment Analysis on trending topics
- Integrated Twitter REST APIs with Glue to apply Vader Model & stored transformed dataset in Parquet Format in S3
- Visualized data in PowerBI using Athena to visualize sentiment variability with Time Series, Heat-map, Bubble charts

Campus Compass (Express.js, React.js, Spring Boot, Leaflet, PostgreSQL)

- Led a 4-member team to design a JavaScript web application for seamless student navigation between campus halls, ensuring WCAG 2.2 compliance and improved accessibility for all students
- Designed database models & routes for endpoints, integrated them using Spring Boot & JPA/Hibernate
- Utilized React Hooks & Redux for state mangement and implemented internationalization using i18n library

EDUCATION

Master of Engineering, Computer Engineering Jan 2022 - Dec 2023
Stevens Institute of Technology

Bachelor of Engineering, Computer Engineering Aug 2014 - May 2018
University of Mumbai

SKILLS

Programing Languages: Python, SQL, Java, JavaScript

Frameworks & Libraries: Spark, Kafka, Flask, PyTest, React, Node.js, Pandas, SQLAlchemy, Matplotlib

Tools: Visual Studio Code, PyCharm, Git, Postman, Alteryx, Tableau, PowerPoint, Word, Excel

Database: MySQL, PostgreSQL, Teradata Vantage, SQL Server, Redis, MongoDB