

# RAGHAV GUPTA

[raghav9048@gmail.com](mailto:raghav9048@gmail.com) | 412-954-8571 | [linkedin.com/in/raghavgupta007](https://www.linkedin.com/in/raghavgupta007)

## SKILLS

**Technical Skills:** SQL, Python, PL/SQL, Spark, NoSQL, R programming, Scala, Java, HTML, Linux & Powershell scripting  
**Software/Cloud Technologies:** AWS (S3, EC2, RDS, Redshift, Kinesis), Azure, GCP, Terraform, Postgres, Snowflake, PL/SQL Developer, MySQL, Kafka, Hadoop, Hive, Informatica, Tableau, SQL Server, SAS, Airflow, Teradata, Git

## WORK EXPERIENCE

### McKinsey & Company, Inc.

Dallas, US

Data Engineer

December 2022 – April 2024

- Engineered a 40% acceleration in data retrieval speed compared to Redshift, saving over 450 man-hours annually, by developing a self-service data reporting tool using AWS glue job (PySpark), DynamoDB, and Athena. Implemented CI/CD pipelines with GitHub actions and efficiently managed AWS resources with Terraform. Also, created an AWS glue job using python for ingesting approximately 30 million records into an OpenSearch index
- Presented data-based stories to senior stakeholders focusing on cost breakdown analysis, developed analytical solutions for integrating multiple data sources & built multiple tableau dashboards, resulting in subscription of 10+ major clients
- Implemented automation for weekly report generation for 10+ senior stakeholders using a streamlined AWS glue job and developed interactive dashboards, resulting in a 95% reduction in time spent on report generation
- Led the creation of a model using AWS lambdas in python, extracted data from excel & DynamoDB, automated data validation, & fed it to RDS, allowing users to perform advanced analytics on model's output, used in 10+ client studies
- Empowered the enterprise client team to independently execute over 100+ model scenarios by developing complex forecast model using statistical modeling within Kedro framework using ETL principles; integrated automated data quality checks using great expectations module, addressing data accuracy problem in the development of data pipeline using python
- Completed 200+ adhoc complex SQL data extracts using AWS Redshift to support client studies by directly engaging with analysts, contributing in generation of over USD 100,000 in revenue
- Leadership: Mentored and trained 2 engineers on project-specific knowledge, tasks and tools

### American Credit Acceptance

Pittsburgh, US

Database Engineer

September 2022 – November 2022

- Designed and implemented a robust relational database architecture, optimizing query performance by 40%, addressing production issues by debugging stored procedures & enhancing system efficiency through database administration
- Facilitated 60% improvement in data processing time by creating stored procedures to efficiently clean and load data from staging to production tables

### Infosys Limited

Bangalore, India

Systems Engineer

February 2019 - April 2021

- Achieved a 30% improvement in query performance by leading design & implementation of data modeling & architecture. Conducted data analysis, facilitated code reviews, testing and performed SQL performance tuning
- Created dashboards to extract insights from data for senior stakeholders. Applied statistical modeling for analysis & developed a script to automate dashboard population, resulting in a saving of over 5 man-hours per new dataset addition
- Implemented an ETL process to consolidate, clean, transform & load the transaction data using python, resulting in improved data quality & increased platform's accessibility by 70%. Also, built tableau dashboards for 5+ senior leaders

### RoboticWares Pvt Limited

Delhi, India

Process Engineer

June 2018 - January 2019

- Engaged with clients to translate business requirements into technical requirements, designed process flows, and created python functions, resulting in minimizing production issues by 60% on world's first BPM engine for mobility in logistics
- Created python script for data processing & loading; designed tableau dashboards to track metrics related to delivery executives' performance, warehouses' efficiency; presented insights to senior leadership, saving over USD 50,000

### HCL Infosystems Limited

Delhi, India

Intern

June 2017 - July 2017

- Conducted Market-Basket Analysis by applying Apriori Machine Learning algorithm for an internal project

## EDUCATION

### Carnegie Mellon University

Pittsburgh, US

Master of Information Systems Management

May 2021 - August 2022

Courses: Advanced RDBMS, NoSQL Databases, Statistics, Business Intelligence, Distributed Systems

### Guru Gobind Singh Indraprastha University

Delhi, India

Bachelor of Technology - Information Technology

August 2014 - May 2018

## ACADEMIC PROJECTS

Home Credit Default Risk | R programming

January 2022 - February 2022

- Performed EDA & applied statistical methods to get insights into characteristics of credit defaulters