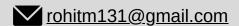
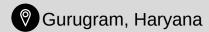
## ROHIT MITTAL

## DATA ENGINEER







## **EDUCATION**

# Thapar Institute of Engineering and Technology

JUL 2018 - JUL 2022 BACHELOR OF ENGINEERING CGPA 8.29/10

### TECHNICAL SKILLS

PROGRAMMING LANGUAGES:

Python, Java, SQL

DATABASES:

MySQL, Cassandra

DATA WAREHOUSES:

Hive, AWS Redshift

MESSAGE QUEUE

Confluent Kafka

DISTRIBUTED COMPUTATION

FRAMEWORK:

Apache Hadoop, Apache Spark

DASHBOARDING AND

MONITORING TOOLS:

AWS Quicksight

**WORKFLOW MANAGEMENT:** 

Airflow

AWS SERVICES:

S3, Lambda, DynamoDB,

Redshift, Athena, Glue, SNS,

SQS, Kinesis, StepFunction

#### **WORK EXPERIENCE**

#### **DELOITTE INDIA - ANALYST**

JAN 2023 - PRESENT

**Reinforcement Learning Through Human Feedback** 

- Worked as a Java annotator in various streams such as SOF, CW Quality Study, Reward Modeling, LeetCode, Code Quality for a project based on RLHF.
- Contributed to the refinement of a **no-code generative AI** platform for a large American MNC.

#### **SNAPDEAL - DATA ANALYST INTERN**

JAN 2022 - JUL 2022

- Designed and implemented the data and reporting infrastructure from scratch using SQL, providing real-time insights into product performance and user funnels, improving the data accuracy by 30%.
- Partnered directly with the product team to create and manage the different **metrics** across the app using SQL and Google Sheets, leading to a 25% increase in actionable insights.
- Developed comprehensive root cause analysis reports to address problems with customer conversion, uncovering key insights that boosted conversion rates by 15%.
- Identified procedural areas of improvement through customer data analysis.

#### **PROJECTS**

#### **Movie Data Analysis**

- Developed an **end-to-end data pipeline** leveraging AWS services: uploaded data to **S3**, utilized **AWS Glue Crawler** and **Glue Catalog** for metadata creation.
- Implemented data transformation using AWS Glue Catalog Data Quality and Glue Low Code ETL, ensuring high-quality data processing.
- Automated data loading to AWS Redshift, enhancing data accessibility and query performance for analytical purposes.
- Set up alerts and notifications using Amazon EventBridge and SNS topics, providing real-time
  updates on data pipeline events and anomalies.

#### **Daily Airline Data Ingestion**

- Automated daily data ingestion pipeline for flight data from S3 to Redshift, ensuring timely and efficient data updates.
- Configure Amazon EventBridge to detect new files in S3 and trigger AWS Step Functions, orchestrating the entire ETL process.
- Utilized AWS Glue Crawler and Glue ETL jobs to process and transform data, **maintaining accurate** and **up-to-date metadata** and data quality.
- Implemented a notification system using SNS topics to provide real-time alerts on job success
  or failure, enhancing monitoring and operational efficiency.

#### **ECom Store Data Processing**

- Developed and managed two data streams, Orders and Payments, using pub-sub topics for real-time data distribution.
- Created a denormalized table in Cassandra combining attributes from both the Orders and the Payments tables, enhancing data integration and retrieval.
- Implemented Python consumer code to **process and merge data streams**, ensuring **efficient data consumption** and **transformation**.
- Handled data integrity issues by routing unmatched payments to a Dead Letter Queue (DLQ) topic, maintaining data accuracy and reliability.

#### **ACHIEVEMENTS**

- Recognized as a Gold Batch member and Super Annotator due to consistently delivering high-quality work for a span of 6 months.
- Recognized as a star performer during my internship for creating daily user data reports, which
  led to the efficient execution of sales and helped increase customer engagement.