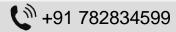
# PREET PATIDAR

## DATA ENGINEER



✓ preetpatidar115@gmail.com





## **EDUCATION**

### MEDI-CAPS UNIVERSITY **INDORE**

JUL 2022 - PRESENT **B.TECH IN COMPUTER SCIENCE ENGINEERING** 

## TECHNICAL SKILLS

PROGRAMMING LANGUAGES C++, C,Python

BACKEND FRAMEWORKS Django

**DATABASES** MySQL, Mongodb, Cassandra

BIG DATA FRAMEWORKS Hadoop, Spark, Hive, Kafka, BigQuery

DASHBOARDING TOOL Quicksight

SCHEDULERS & AWS Airflow, S3, EMR, Glue, Lambda, Cloudwatch , Dynamo DB, Athena

#### PROJECTS-

#### 1. Quality Movie Data Analysis

- Built an end-to-end data analysis pipeline to extract insights from public movie datasets (IMDB, TMDB, Box Office Mojo).
- Performed deep cleaning, transformation, and schema alignment using PySpark and AWS Glue for scalable ETL.
- Designed a star schema on Amazon Redshift for efficient querying, with Facts capturing revenue and ratings, and Dimensions covering genre, actors, and release timelines.
- Leveraged AWS Athena and QuickSight to enable business teams to self-serve analytics dashboards for content acquisition and trend forecasting.
- Optimized long-running gueries using partitioning and distribution strategies: reduced report latency by 70%.
  - 2. Real-Time Data Ingest (MySQL  $\rightarrow$  Kafka  $\rightarrow$  Spark  $\rightarrow$  Snowflake)
- Architected a real-time data pipeline to decouple production MySQL from reporting workloads.
- Implemented Kafka producers to stream CDC (Change Data Capture) events using Debezium, and Spark Structured Streaming jobs to process high-volume transactional data with low latency.
  - Transformed and cleansed data in Spark before loading into Snowflake using the Snowflake Connector for Apache Spark.
- Ensured schema evolution and data integrity with custom metadata tracking and alerting logic.
- Achieved sub-minute end-to-end latency, allowing stakeholders to access near real-time metrics via Snowflake dashboards.

#### ACHIEVEMENTS

- JEE Mains 2022 Cleared with 82.3 Percentile
- Solved 250+ Questions on LeetCode
- 🗆 4★ C++ Coder on HackerRank