ANAS SHAIKH

Data Engineer

9321974601 | anas19257@gmail.com | Maharashtra – Mumbai 400051

CAREER SUMMARY

Detail-oriented and aspiring Data Engineer with hands-on experience in building automated data pipelines, cloud-based ETL workflows, and interactive dashboards. Skilled in Python, SQL, AWS, Databricks, and Prefect, with projects demonstrating end-to-end solutions from real estate data pipelines to real-time patient monitoring systems. Strong analytical mindset with a passion for transforming raw data into meaningful business insights.

TECHNICAL SKILLS

Python | SQL | pandas | Playwright | Web Scraping | API Integration | Psycopg | Databricks | Apache Spark | Delta Lake | Prefect (workflow orchestration) | Airflow | AWS S3 | AWS Lambda | AWS RDS (PostgreSQL) | AWS API Gateway | AWS SNS | AWS SQS | AWS EC2 | AWS Event Bridge | Serverless Architecture | ETL Pipelines | Data Cleaning & Transformation | Data Modeling | Query Optimization | Event-driven Processing | Workflow Scheduling | Real-time Data Processing | Streamlit | Data Visualization | Dashboard Development | KPI Reporting | System Design | Power BI.

PROJECTS

REAL ESTATE DATA PIPELINE & DASHBOARD (99ACRES WEB SCRAPING → AWS → DATABRICKS → STREAMLIT)

- Designed and implemented an automated data pipeline to extract real estate listings from 99acres using python and Playwright.
- Ingested scraped data into AWS S3 and orchestrated workflows using prefect for scheduling and reliability on an EC2.
- Built a processing pipeline in Databricks to clean, transform, and enrich daily ingested data, storing processed outputs back into S3.
- Configured AWS lambda (triggered by S3 events) to load processed data into amazon RDS (Postgres SQL), while maintaining a tracking table for data consistency.
- Developed an interactive Streamlit web application connected to RDS for data exploration, filtering, and visualization (bar charts, scatter plots, box plots, KPI, dashboard RDS).

Impact: Achieved a fully automated ETL + analytics system, reducing manual intervention and enabling real-time insights into housing trends.

Tech stack: Python, SQL, Playwright, pandas, AWS (S3, lambda, RDS,), Databricks, prefect, Streamlit.

PATIENT MONITORING & ALERT SYSTEM (SEVERLESS REAL-TIME PIPELINE)

- Designed a real-time health monitoring pipeline where a Python-based script generated mock patient vitals (heartbeat, SpO_2, blood pressure, respiration rate) and sent them to AWS Lambda through API Gateway POST requests.
- Implemented threshold-based logic in Lambda to automatically detect critical health conditions (e.g., heartbeat > 140, SpO_2 < 93) and trigger emergency alerts.
- Integrated AWS SNS to send real-time Gmail notifications to medical staff, enabling immediate intervention.
- Leveraged serverless architecture for cost efficiency, scalability, and low-latency processing without manual server management.

Impact: Demonstrated ability to design event-driven, automated health monitoring systems that can reduce emergency response time in real-world scenarios.

Tech Stack: Python, AWS Lambda, AWS API Gateway, AWS SNS, JSON, Event-driven architecture.

EDUCATION

BACHELOR'S DEGREE | Computer Science | Rizvi College of Arts and Science 2020 – 2023

COURSES

Data Engineer – IBM | Data Analyst – Google | Machine Learning – Google

SOCIALS

LinkedIn: www.linkedin.com/in/anas-shaikh-59b551269 | **GitHub: https://github.com/Anas484/Data-Engineer**