Shayan Rahmatullah

Linkedin | www.shayan.us | GitHub | me@shayan.us | 973-885-2849

SKILLS

Programming Languages: SQL, Java, Python, Pandas, Spark, HTML, CSS, PHP, C#, MongoDB
Data Engineering Tools: ETL, Data Mining, Data Modeling, Data Warehousing, Airflow, DBT, SSMS, Databricks, Snowflake, AWS - EC2, S3, Glue, Sagemaker, Tableau, Git
Education: Rutgers University **Major:** B.S. in Computer Science **Graduation:** May 2024

PROJECTS

Machine Learning Content Sorter *Jan 2025*

- Developed an automated file sorting system using Python and AWS services, categorizing files into predefined folders based on their names and contents. Integrated AWS SageMaker for machine learning-based classification of file content, enhancing sorting accuracy and efficiency.
- Leveraged AWS Glue and SageMaker for data cleansing, feature extraction, and model inference during the transformation phase, resulting in a 40% improvement in sorting precision. Integrated version control (Git) for collaborative development and continuous updates.

Store Analysis *Jan 2025*

- Optimized data models using Snowflake by migrating department and store data from AWS S3 buckets. Utilized DBT's snapshot for version tracking, ensuring accurate data lineage and consistency across environments.
- Designed and implemented automated data workflows with Apache Airflow, orchestrating ETL processes and scheduling data pipeline tasks for regular updates. Integrated Snowflake and AWS S3 with Airflow for seamless data ingestion, transformation, and loading, ensuring timely and accurate data processing.
- Conducted comprehensive data analysis on store and department performance metrics using Snowflake and SQL, uncovering key business insights on seasonal trends and inventory efficiency.
- Created interactive Tableau dashboards to visualize sales patterns, enabling data-driven decision-making for stakeholders.
- Applied data profiling techniques to ensure accuracy in raw data ingested from AWS S3, improving data quality across the pipeline.

Real-Time Flight Data Streaming and Processing *Dec 2024*

- Developed and deployed a real-time data streaming pipeline to process flight data by pushing data into AWS Kinesis Data Streams using an existing online API. Enabled seamless ingestion, transformation, and storage of flight schedules, status updates, and passenger information into Snowflake for analytics.
- Leveraged AWS Lambda and API Gateway for efficient event-driven processing and integration of flight data sources. Utilized DBT for data transformation and version control, ensuring data quality and consistency across more than 100 tables. Built a scalable architecture capable of handling high-frequency flight data streams, optimizing data delivery and ensuring near-real-time updates for operational and business intelligence purposes.
- Performed real-time analytics on flight data to monitor delays, optimize schedules, and improve operational efficiency.
- Used and DBT to perform time series and trend analysis on flight arrivals and departures, driving improvements in airline punctuality KPIs.
- Delivered business intelligence reports highlighting peak flight hours, passenger loads, and regional demand patterns.

Online Travel Reservation System *Apr 2024*

- Designed and optimized a MySQL database for managing flight reservations, customer data, and booking history, ensuring data integrity through structured schema design and foreign key constraints. Implemented efficient SQL queries for retrieving, updating, and managing flight availability.
- Developed a Java and Python-based data pipeline using JDBC and Python libraries (e.g., Pandas, SQLAlchemy) to handle real-time user transactions, enabling seamless database interaction and reducing data latency by 30%.
- Analyzed user booking trends and customer behavior using SQL queries and Pandas dataframes, offering insights for targeted marketing and route planning.
- Created detailed statistical summaries on reservation frequency, peak seasons, and user demographics to support business strategy development.

PROFESSIONAL EXPERIENCE

Co-founder, WhatsTheWave.co *Aug 2023 – Present*

- Led a team of 2 students to develop a real-time activity planning application, boosting friend meetups by 40% through real-time data streaming and scheduling optimizations.
- Engineered a scalable backend using Apache Kafka, processing 100K+ data points daily for dynamic scheduling updates.
- Integrated student class and exam schedules, enhancing social activity coordination by reducing scheduling conflicts by 35%.
- Analyzed user behavior data to determine peak engagement times, guiding app feature enhancements that increased user interaction by 25%.

- Built a real-time dashboard using Python and SQL to visualize trends in event scheduling and attendance across campus, improving product-market fit.

Software Engineer Intern, CHEF International *May 2023 – July 2023*

- Performed in-depth analysis of webpage dead links, achieving a 50% recovery rate, improving website navigation and user retention.
- Redeveloped the company website using WordPress, SQL, and PHP, resulting in a 20% faster load time and a 15% increase in page views.
- Introduced AI autocomplete to the patient portal system, reducing form errors by 30%, enhancing data accuracy during patient registration.
- Implemented a Continuous Integration/Continuous Delivery (CI/CD) system with version history, cutting deployment errors by 40% and minimizing downtime.
- Conducted A/B testing on website design elements and call-to-action buttons, improving user conversion rates by 18%.
- Used SQL and Excel to generate weekly reports tracking user behavior, bounce rates, and top-performing web pages.

Freelancer, shayan.us *Jun 2019 – Present*

- Delivered 15+ client projects, showcasing expertise in client communication and problem-solving.
- Developed data-driven software solutions using Python and SQL, automating ETL processes and reducing manual effort by 60%.
- Built scalable data architectures and integrated cloud services (AWS, Azure), improving system efficiency by 45% and reducing infrastructure costs by 25%.
- Designed and deployed custom analytics dashboards using Python (Plotly, Dash) and SQL, helping clients track KPIs like sales performance, user retention, and churn rate.
- Applied clustering and regression techniques on customer datasets to deliver insights on purchase behavior and product performance.
- Built ETL pipelines that extracted web traffic logs and CRM data for marketing campaign effectiveness analysis, supporting clients in ROI optimization.