

Shailen Sharma

Mahwah, NJ 07430 • (845) 825-9786 • shailen.sharma@yale.edu

Education

Yale University

9/2022 – Present

- *B.A. in Applied Mathematics*
- Relevant Coursework: Intermediate Machine Learning (in progress), Computational Vision (in progress)

Skills

- **Languages:** Python, SQL, Unix (CLI & scripting)
- **Libraries & Tools:** Pandas, PySpark, boto3
- **AWS Stack:** Glue, S3, Redshift, Athena
- **Core Proficiency:** Scalable automated ETL design, LLM agent tuning, and production data integration

Experience

Data Engineering Intern

Jersey City, NJ

6/2025 – 8/2025

SageSure

- Automated ETL monitoring by developing a Python regex parser that **processes job-run emails**, extracting job IDs and run statuses to generate structured operational logs
- Tuned an Amazon Q **LLM agent** via JSON configuration to translate natural language into complex SQL over a near real-time replica of production table, allowing internal teams to explore live metrics safely
- Engineered an **S3-to-Athena ETL pipeline** in PySpark with ingestion, partitioning, and schema management
- Configured **dynamic partitioning**, metadata syncing, and schema evolution for low-latency automated data access to transformed tables with no manual intervention
- Built Redshift ingestion logic in AWS Glue to auto process Excel drops via Pyspark for transforming DataFrames, logging file metadata, and **populating production tables**

Infantry Team Leader

United States Marine Corps

Global Deployments: Norway, Sweden, Scotland

Aug 2017 – Mar 2018

Stateside Operations: Various US Installations

5/2016 – 5/2020

- Deployed above the **Arctic Circle** for NATO exercises under live-fire, night, and extreme cold-weather
- Assisted platoon medic in treating hypothermia casualties, performing **patient care** and evacuations during Arctic mountain exercises
- Led junior Marines in combat readiness, running casualty drills and unit training while recovering from reconstructive surgeries

Projects

cuDF CSV Reader | Python, cuDF, WSL2 (Linux)

- Built a Python tool to dynamically ingest large CSVs and Parquet files into cuDF with error handling, file-type switching, and automatic file/path detection
- Enabled CSV to Parquet conversion for fast read of 10GB+ files in **GPU accelerated pipelines** on Linux

IMDB Data Transformation | Python, SQLite3

- Built a data pipeline to ingest and clean 500K+ IMDb records from three sources into a SQLite3 database
- Developed a **Python classification system** to **segment** movie titles for genre-based comparison

Volunteering

Rescue Diver | Mahwah Fire Department

- Public safety diver supporting underwater rescue operations with local, regional, and state agencies