ANSHUMAN SHARMA

Irvine, California | (949) 279-7105 | sharma.anshuman97@gmail.com | Linkedin | Portfolio

EDUCATION

University of California Irvine | Master of Science in Data Science | GPA: 3.61/4.00 — Sep 2022- Dec 2023
Honor: Program Ambassador MDS

Amity University | B.Tech Computer Science and Engineering | GPA: 3.88/4.00 — Aug 2016- Jul 2020
Honor: Recipient of Dr. Ashok K Chauhan Merit Scholarship

SKILLS

Backend:	Python, Flask, Django.
Frontend:	ReactJs, Redux, Javascript, HTML, CSS.
Data Pipelines:	Apache Airflow, Spark, Kafka, Data Warehousing (Redshift, Snowflake, SAP BW), Cassandra, DBT, Hadoop Talend, Grafana, Prometheus, ELK Stack (Elastic Seach, Logstash, Kibana), SQL.
Miscellaneous:	Docker, Kubernetes, Amazon AWS, Microsoft Azure, Databricks, Power BI, Looker, Tableau, A/B Testing, Git.

EXPERIENCE

City of Irvine | Machine Learning Engineer for Capstone Project | Irvine, CA — Sep 2023 - Dec 2023

- Developed an LLM-powered chatbot using **Flask, React,** and **Llama2**, achieving a 97% reduction in time to find contract details.
- Optimized a 1M+ document dataset with the **Llama Index** and fine-tuned the **Llama2** model via prompt engineering.
- **Dockerized** the application for consistency and portability, and orchestrated its deployment using **Kubernetes**.
- Architected a **CI/CD** pipeline with monitoring and auto-retraining for long-term Chatbot efficacy.

Dell Technologies | Data Engineer Intern | Round Rock, TX — Jun 2023 - Aug 2023

- Uncovered manual reporting inefficiencies via 50+ stakeholder interviews. Derived actionable KPIs for performance improvement.
- Streamlined reporting for Marketing team (FMMs), saving over 5,000 hours annually, and an additional 10,000 hours in the PAN ISG initiative.
- Developed and deployed a data integration framework that consolidated data from 6+ sources into a warehouse (**Azure Synapse Analytics**).
- Spearheaded the integration of **Apache Airflow**, enabling automated and scalable data orchestration for the data platform.

Accenture | Data Engineer | Bangalore, KA — Oct 2020 - Jul 2022

Electrolux E-commerce Search Optimization

- Developed a hybrid recommendation system using collaborative filtering and content-based filtering**,** leading to a 12% conversion lift.
- Built real-time product search pipelines in **Python** with **Apache Spark** for efficient processing, enabling up-to-date search results and improved user experience.
- Built data pipelines on **AWS Glue** to ingest & transform millions of daily customer events (purchases, demographics, browsing), enabling real-time recommendations & search filtering.
- Implemented **Airflow** logging, achieving 20% faster troubleshooting and error identification for data pipelines.

Data Modeling and Optimization

- Designed dimensional data models in **SAP BW** on **HANA**, integrating customer & sales data for efficient analysis.
- Optimized data pipelines, achieving a 20% reduction in North American data transfer delays.
- Led a team of 5 to achieve a 98% on-time delivery rate for critical BI reports, exceeding SLA benchmarks.
- Owned the data pipeline lifecycle, ensuring rigorous testing, zero-downtime deployments, and timely delivery of critical BI reports.

Samsung Research Institute (SRI-D) | Data Engineer Intern | Noida, UP — Feb 2020 - Apr 2020

- Automated 90% of manual TV testing workload with **Robot Framework**, saving an estimated 200 man-hours per week.
- Built a real-time data ingestion system using **Apache Kafka** to capture and process test logs, enabling the generation of reports with **Grafana**.

PROJECTS

Chicago Transit Authority (CTA) | Real-Time Train Monitoring

- Designed and implemented a real-time train monitoring system using **Apache Kafka** (Kafka Connect & REST Proxy) for data ingestion and processing, including caching static data (routes, stations) with **Redis**.
- Developed real-time train location and status visualizations for improved passenger experience.

Sparkify Music Streaming App | ETL Pipeline and Data Warehouse

- Built a scalable Apache Airflow ETL pipeline with **DBT** modeling for Sparkify music data. Processed raw events in **Cassandra** and loaded them into **Redshift** data warehouse.

Match Fit | AI-powered talent acquisition system

- Developed a talent acquisition system using Retriever-Augmented Generation (RAG) with **Pinecone** vector database and **Llama 2**, enabling recruiters to efficiently search a vast resume corpus based on desired skills.