Rohan Lagare

linkedin | 408-690-7294 | Portfolio site | rohanlagare97@gmail.com | github

SKILLS & CERTIFICATIONS

- **Programming Languages**: **Python**(pandas, NumPy, PySpark, Scikit-learn, Django), **SQL**, JAVA, C++, **shell scripting**.
- **Databases**: SQL (Oracle, Teradata, Snowflake, PostgreSQL), NoSQL (MongoDB).
- **Big Data Ecosystem**: Spark, Kafka, **Airflow**, Hadoop, MapReduce, Hive, Sqoop.
- **Cloud Technologies**: **AWS**(S3, Athena, Glue, EC2, EMR, RDS, Sagemaker, Redshift, IAM), **Azure**, **GCP**.
- **ETL and other tools**: **Ab-initio**, Tableau, IBM work scheduler, Kubernetes, Docker, Jupyter notebook, REST, Git.

WORK EXPERIENCE

SCU Frugal Innovation Hub **Santa Clara, USA**

Software Developer

Project: **Bilingual Math learning application** *January 2024 - present*

- Developed and maintained a Bilingual Math learning application using Flutter, Figma, and Firebase, resulting in increased user engagement and math proficiency for bilingual students.
- Implemented responsive design principles, ensuring seamless usability on smartphones and tablets, and collaborated with designers and educators to enhance user experience.
- Contributed to comprehensive project documentation, including technical specifications and API documentation, to facilitate clarity and transparency for stakeholders and future developers.

Bitwise Solutions (Client: Discover Bank) **Pune, India**

Data Engineer *October 2018 - November 2021*

Project: Model Tower | **AI/ML focused Data warehouse**

- Designed and implemented ETL data pipelines utilizing Ab-initio, AWS and Snowflake, delivering high-quality data inputs essential for AI/ML model training and inference.
- Achieved a 35% reduction in data errors through implementation of comprehensive data quality checks, ensuring compliance with critical business requirements and enhancing overall system reliability.
- Orchestrated ETL workflows using IBM Workload Scheduler to automate and optimize critical data processing tasks.
- Collaborated closely with stakeholders to translate project requirements into robust, timely data pipeline solutions optimized for AI/ML model integration.

Project: Line of Credit | **New product data warehouse**

- Designed and implemented robust data pipelines for a new product data warehouse using Ab-initio, AWS, Teradata, Snowflake, and shell scripting, enabling enhanced data processing capabilities and system scalability.
- Minimized development time and effort by 30% through the implementation of Generic applications for data manipulation across Data Staging (DWH) and Operational Data Store (ODS) layers.
- Devised automated scripts for error handling and SLA notifications, reducing manual activity by 70% and enabling real-time system monitoring and reporting.
- Collaborated in Agile sprints to deliver key features, ensuring alignment with project goals. Utilized data modeling expertise to optimize database design, enhancing data storage efficiency and meeting project deadlines.

Project: Capstone | **Migration from Hadoop**

- Accomplished successful migration of data from Hadoop file systems to Linux servers, meeting project requirements.
- Developed automated scripts to streamline data migration processes, reducing migration time by 80% and ensuring error-free transfer of large datasets from Hadoop to Linux servers.
- Engineered custom data pipelines for extracting and transferring data to Teradata servers, ensuring a smooth and accurate migration process for millions of records.

PROJECTS

Customer segmentation **Santa Clara, USA**

- Accomplished a customer segmentation using Parallel K-means clustering, leveraging parallelization to achieve a 7X speedup compared to serial algorithm, resulting in more efficient customer segmentation.
- Analyzed runtime statistics and identified key insights about relationship between speedup and other runtime variables, utilizing parallelization techniques to optimize customer segmentation performance.

Face Recognition **Santa Clara, USA**

- Developed and trained a face recognition model, achieving 98.7% accuracy.
- Accomplished live webcam face recognition using trained model, facilitating real-time identification.
- Employed advanced techniques for data preprocessing and model training to handle variations in lighting, occlusions, and facial features, resulting in increased model accuracy from 70% to 98.7%.

EDUCATION

Santa Clara University **California, USA**

MS, Computer Science (GPA: 3.59/4.00). *September 2022 - June 2024*

Relevant Coursework: Database Systems, Distributed Systems, Parallel Programming, AI, Machine learning, OS.

PVG's College of Engineering and Technology, University of Pune **Pune, India**

BE, Computer Engineering. *June 2014 - June 2018*