Shripad Sunil Mathekar
shripadmathekar@gmail.com | 640-229-1683 | LinkedIn

EDUCATION

Syracuse University   **New York, USA**
Master of Science in Computer Science, **GPA: 3.7/4.0**   **(August 2023 – May 2025)**
Relevant Coursework: Design and Analysis of Algorithms, Object Oriented Design, Intro to Machine Learning

University of Pune (SPPU)-Pune, India (August 2014- May 2018)   **Maharashtra, India**
Bachelor of Engineering in Electronics and Telecommunication, **Distinction**   **(August 2014 – May 2018)**
Relevant Coursework: Data structures, Object Oriented Programming.

TECHNICAL SKILLS

Languages: Python, PySpark, Java, C, SQL
Data Frameworks: Hadoop**,** Spark, Hive, Pig
Databases: PostgreSQL, MySQL, MongoDB, Redshift, Snowflake
Cloud & Tools: AWS, Airflow, Git**,** Bit-Bucket, Design Patterns, Dbeaver, SQL Workbench, PoweBI

PROFESSIONAL EXPERIENCE

Software Developer at **KPMG**   **June 2021 – February 2022**
Technologies: Python, Spark, AWS, SQL.
1. Datawarehouse Implementation | Vanguard:
- End-to-End implementation of data warehouse in AWS cloud to handle the 360-degree view of Vanguard's financial advisors handling over 5 TB of data.
- Engineered Python and Spark-based data pipelines for efficient ETL, processing over 1 million records daily.
- Architected a YAML-based framework for data cleansing and ETL standardization, resulting in a 30% reduction in data processing time.
- Designed, deployed and scheduled AWS Glue jobs using AWS workflow.
- Conducted rigorous data validation checks on over 50 data attributes to ensure data accuracy.

2. Master Data Management | Vanguard: **March 2022 - July 2023**
- Master Data Management (MDM), focused on enhancing Redshift data warehouse content with the most current Vanguard's financial advisor's information obtained from 4 external sources.
- Designed and implemented rules to match the advisors from multiple sources with the target Redshift data.
- Developed Python scripts to enforce data quality standards, resulting in a 25% improvement in overall data integrity.
- Designed and implemented Spark-based data pipelines, based on matching and survivorship rules.
- Leveraged AWS Glue for automating data ingestion and transformation, achieving a 30% increase in data throughput.

Software Developer at **Infosys**   **November 2018 - June 2021**
Technologies: Python, Spark, AWS, SQL, Hive
1. Cloud Migration | Ralph Lauren:
- Developed data pipelines using Spark to perform ETL as per the Informatica workflow logic
- Designed, developed and scheduled Airflow DAGs to automate pipelines.
- Performed file format conversion using AWS Athena
- Conducted performance tuning and optimization of SQL queries, achieving a 30% improvement in query response time.

2. Assortment Optimizer Automation | Mondelez:
- Automated the Assortment Optimizer module for a retail client using Big Data tools (Spark, Hive, Oozie) to streamline data processing and improve decision-making.
- Developed shell scripts for efficient data transfer to HDFS.
- Developed business logic and performed ETL using Spark.
- Executed unit test cases on developed pipelines and performed data validation on target data.

PROJECTS

Flight Arrival Prediction Model - Syracuse University
- Developed a Random Forest Classifier to predict flight arrival times at Syracuse Airport using historical flight and weather data.
- Enhanced model robustness through advanced data preprocessing, feature engineering, and cross-validation techniques, optimizing predictive reliability.
- Achieved 81% accuracy, enabling delay management and offering data-driven insights for improved operational planning.

CERTIFICATIONS

- Google Cloud Certified Associate Cloud Engineer(4cNDXE)