Prachi

prachibaliyan0101@gmail.com | (714) 519-7469 | LinkedIn | Fullerton, CA – 92831

PROFESSIONAL SUMMARY

Experienced AWS & ETL Data Engineer with approx 3 years in the service industry, skilled in Python, AWS, Talend and databases. Proficient in collaborating with key stakeholders to conceptualize and execute solutions that resolve systems architecture-based technical issues. Contributed to the design and development of scalable, robust data engineering frameworks that support business intelligence and decision-making needs.

SKILLSET

Programming Languages: Python, Pyspark, HTML, CSS, C, C++
Frameworks and Tools: AWS, GitHub, JIRA, Talend, DBeaver, Jupyter Notebook
Databases: NoSQL - MongoDB, DynamoDB | SQL - MySQL, Oracle | PostgreSQL

EDUCATION

MS in Computer Engineering| California State University, Fullerton, CA| Aug 2024 – Present |
BS in Computer Science & Engineering| Dr. A.P.J. Abdul Kalam Technical University | Aug 2018 – Jun 2022 |

PROFESSIONAL EXPERIENCE

Data Engineer | **KPMG, India** | April 2024 – August 2024

• Served as a Data Engineer for Maruti Suzuki India Limited, leveraging AWS services including Amazon DynamoDB, Amazon S3, AWS Glue, AWS Lambda, and Amazon EventBridge to design and implement scalable data solutions and optimize data workflows.
• Automated SMS functionality by implementing and configuring AWS SNS through AWS Lambda, eliminating manual sending processes and achieving a 70% reduction in time spent.
• Implemented a data catalog for Maruti Suzuki India Limited by creating an AWS Glue job to extract metadata from raw data and store it in a DynamoDB table, which serves as the backend for a user-facing application that provides essential information to business users.
• Collaborated with the Data Management team to understand project objectives, develop validation scripts, and create algorithms to transform raw data into actionable insights.
• Responsible for building, testing, and maintaining database pipeline architectures, ensuring high-quality data delivery and optimizing SQL queries to enhance performance and efficiency.
• Analyzed and managed large volumes of data using Amazon Redshift for data warehousing, enabling projects to leverage data engineering expertise for effective and scalable solutions.

Data Analyst | **KPMG, India** | July 2022 – March 2024

• Developed an ingestion pipeline using Talend and Python to transfer data from on-premise Oracle servers to AWS S3, extracting data from multiple tables and storing it in Parquet format in the AWS S3 raw layer.
• Streamlined data operations with PySpark and Python scripts in AWS Glue for data wrangling of raw data that is ingested from various systems, both on-premise and IoT, from the raw layer for other downstream data projects.
• Developed and implemented a Python-based AWS Lambda function to automate the triggering of Glue jobs via EventBridge, dynamically fetching job parameters from DynamoDB, processing raw data through cleansing and business logic, and efficiently writing results to the curation layer.
• Facilitated client meetings to thoroughly understand and document business requirements, ensuring effective communication and alignment with project objectives.

Intern | **KPMG, India** | Feb 2022 – June 2022

• Engaged in hands-on training with AWS tools and services, mastering essential functionalities that resulted in a 30% reduction in deployment time for cloud-based applications, streamlining project timelines, and enhancing operational efficiency.
• Designed and implemented a web application hosted on Apache server using AWS EC2, featuring well-defined endpoints for various operations, and built a robust data processing pipeline with AWS Kinesis and AWS Glue to efficiently ingest, process, and transform streaming data.

PROJECTS

EyeTrack: Advanced Eye Detection and Tracking System

• Investigated the feasibility of detecting eye blinking and movement using Python and basic machine learning algorithms, focusing on monitoring recognition patterns as they interact with the digital environment.
• Established a methodology to identify and filter out false and irrelevant features in algorithmic processes, enhancing the accuracy and efficiency of the computations.
• Predicted outcomes using models such as Decision Tree Classifier, Logistic Regression, K-Nearest Neighbors, and Multiple Regression to evaluate and identify the most efficient technique.