

Pragathi Gopishetty

pragathigopishetty@gmail.com | 4102949828 | Baltimore, MD(Open for Relocation) |
linkedin.com/in/pragathigopishetty/

SUMMARY

Data Engineer with over 3 years of hands-on experience designing, implementing, and optimizing end-to-end data solutions. Proven expertise in data processing, ETL pipeline development, and data modeling. Proficient in leveraging big data technologies, cloud platforms, and machine learning to drive actionable insights.

EXPERIENCE

Research Assistant - Data Engineer, University of Maryland Baltimore County *Aug 2022 - May 2023*

- Designed and implemented a real-time data pipeline, integrating 2M+ raw records from 10+ data sources using Kafka and PySpark, achieving a 20% reduction in data processing time.
- Leveraged Apache Spark for large-scale data processing, cleaning, and Spark jobs, resulting in the creation of user profiles based on historical interactions.
- Defined an efficient data model using RDBMS and data modeling tools, leading to a 30% enhancement in data retrieval speeds. Deployed a scalable data storage solution with AWS Redshift, reducing storage costs by 15%.
- Applied advanced Machine Learning techniques to develop a real-time recommendation engine using Spark Streaming. Hosted a website on EC2 to provide users with personalized recommendations.

Data Engineer, Flujo Technologies *Dec 2019 - Aug 2021*

- Led the successful migration from Oracle to S3 using AWS DMS resulting in an annual cost savings of \$650,000.
- Extracted insights from raw data using ETL processes with AWS Glue, Athena, Lambda and Redshift, resulting in a 20% reduced user attrition.
- Leveraged Databricks and PySpark to architect robust data processing pipelines, optimizing queries with PySQL to improve efficiency. Enhanced CI/CD pipelines for Databricks, automating deployment processes seamlessly through git version control.
- Designed a scalable MySQL multi tenant database architecture, allowing for efficient data management and analysis. Built RESTful APIs using Python Flask with Boto3 SDK to streamline data extraction from S3 buckets.
- Developed SSIS Packages optimizing ETL workflows with SQL server, reducing data processing time by 30% and enhancing loading phase accuracy by 25% and automated monitoring with Power BI, reducing manual efforts by 80%.
- Collaborated with stakeholders to define business objectives, analytical, problem-solving with attention to detail. Improved data integration, analysis, and test automation for customer data across 20+ clients.
- Analysed user experience by fine-tuning stored procedures and SQL queries for efficient data retrieval from approximately 4 million records accessed by 5000+ users worldwide.
- Orchestrated comprehensive infrastructure deployment utilizing Docker to build images, containerized with ECR and manifested using EKS. Showcased AWS infrastructure seamlessly using CloudFormation.

PROJECTS

Maryland Crashes Analysis based on weather

Conducted thorough data cleaning and analysis on Maryland crash data using PySpark on Databricks. Identified rain and snow as key contributors to accidents, leading to targeted safety interventions.

Payment mode prediction for Chicago Taxi Trips

Analyzed Python and SQL for data analysis of Chicago taxi trips, applying Machine learning algorithms to predict payment modes. Incorporated S3 and SageMaker to enhance model performance, visualized via QuickSight.

EDUCATION

Master's of Professional Studies, Data Science

Aug 2021 - May 2023

University of Maryland Baltimore County

Bachelor of Technology, Computer Science and Engineering

Jun 2016 - Apr 2020

MLR Institute of Technology

SKILLS

Programming Languages: Python, SQL, R

Databases: MySQL, SQL Server, PostgreSQL

Business Intelligence & Reporting: Tableau, Power BI, QuickSight

Project Management & Collaboration : Jira, Agile Methodology, Git

Big Data & Cloud: AWS(EC2, S3, Lambda, Glue, Athena, SNS, RDS, EMR), Databricks(Spark, PySQL, PySpark)

CERTIFICATIONS

AWS Certified Cloud Practitioner