Venkata Sai Rishitha Seelam

LinkedIn | **Email**: sairishitha0869@gmail.com | Portfolio | Baltimore, MD

DATA ENIGINEER

Results-driven Data Engineer with **2.5+ years of experience** in building scalable ETL/ELT pipelines, Interactive dashboards, managing cloud data ecosystems, and delivering analytical solutions. Proven expertise in leveraging AWS and Azure cloud services to optimize data pipelines and enable data-driven decision-making. Skilled in **Python**, **SQL**, **Power BI**, **AWS and Azure Cloud services**, **Informatica**, **Snowflake**, **Pyspark** with a track record of delivering actionable insights, reducing process times, and improving data governance. Passionate about building robust, fault-tolerant systems to drive organizational success.

PROFESSIONAL EXPERIENCE

AWS Cloud Data Engineer *Jul 2022 – Jan 2023*

Cognizant Technology Solutions (Healthcare industry) Chennai, India

* Designed and implemented Data solution to manage customer rebates based on their purchase patterns.
* Optimized scalable ETL pipelines to process 5M+ records using AWS services, ensuring data integrity and consistency
* Streamlined data availability using AWS Lambda and Glue, transforming data from four diverse sources into a centralized AWS S3 repository, achieving a 25% reduction in data retrieval time.
* Implemented Snowflake-based data warehousing solutions managing over 3TB of data, improving query performance by 40%.
* Improved data governance policies and metadata management, enhancing compliance with industry standards.
* Enabled 30% faster decision-making by delivering actionable insights using AWS Athena for advanced data analysis.

Data Engineer and Data Analyst *Mar 2021 – Jun 2022*

Cognizant Technology Solutions (Healthcare industry) Chennai, India

* Worked on an end-to-end data project to analyze competitor performance relative to our client, delivering actionable insights and developing a dashboard to identify opportunities for revenue growth.
* Migrated legacy Python ETL workflows to PySpark and Databricks, reducing ETL runtime by 70% (from 20 hours to 1 hour) for 120GB datasets.
* Developed and optimized data pipelines using SQL, Snowflake, and Informatica, ensuring seamless data integration.
* Built 20+ real-time dashboards in Power BI, increasing reporting efficiency by 40% and enhancing data visualization.
* Automated data extraction from APIs and databases, by using python scripts reducing manual efforts by 60%.
* Designed scalable data warehouse models in snowflake for storing and analyzing 2TB+ of data, enabling efficient storage and retrieval.
* Automated reporting workflows, achieving an 80% reduction in manual effort with Python scripts and Power BI.
* Engaged in database and data warehouse modeling and design, creating efficient data relationships and schemas to support robust data storage, retrieval, and analysis using platforms such as SQL.

TECHNICAL SKILLS

* **Programming Languages** : Python, SQL, PySpark, R
* **Cloud Platforms** : AWS (S3, Glue, Lambda, Kinesis, RDS, Redshift), Azure Cloud services
* **Big Data Technologies** : DataBricks, Apache Spark, Spark SQL, Kafka, Hadoop (HDFS, MapReduce)
* **Data Warehousing & ETL** : Snowflake, Datalake, Informatica PowerCenter, Informatica MDM
* **Data Analysis & Visualization** : Power BI, Tableau, Excel, Pandas, NumPy
* **Data Engineering** : ETL Processes, data mining, data cleansing, validation, transformation
* **Others** : Power Automate, Agile/Scrum, Machine Learning & Forecasting

EDUCATION

Master's in Data Science *| GPA: 3.9* *Dec 2022 – Dec 2024*

University of Maryland, Baltimore County Baltimore, Maryland

Relevant Coursework: Data Science, Big Data, Data Visualization, Machine Learning, AI Fundamentals and Virtual Reality

Bachelor's in Electronics & Communication | *GPA: 8.6* *Jul 2017 – Jun 2021*
SASTRA University Thanjavur, India

ACADEMIC PROJECTS

US HOUSEHOLD ENERGY PREDICTION *Oct' 2024 – Dec' 2024*

* Developed a predictive model for U.S. household energy consumption using regression techniques with seasonal adjustments.
* Optimized and deployed the model for scalability and real-world application.
* Designed an interactive web application using HTML, CSS and JavaScript for users to visualize energy consumption predictions

GENE SEQUENCE CLASSIFICATION *Aug' 2023 – Dec' 2023*

* Contributed to the advancement of Splice Junction gene sequence classification through machine learning techniques.
* Preprocessed gene sequence data to prepare it for machine learning algorithms.
* Utilized Python, NumPy, Pandas, and Scikit-Learn for data preprocessing and model development.
* Applied various machine learning algorithms to classify gene sequences.
* Evaluated model performance using metrics such as accuracy, precision, recall, F1-score, and AUC-ROC.

FACIAL EMOTION RECOGNITION *Jan' 2023 – May' 2023*

* Developed a highly accurate machine learning model for detecting and classifying human emotions.
* Utilized advanced machine learning algorithms & Computer Vision to train the model on a diverse dataset of facial expression.
* Achieved high precision and recall in accurately identifying and classifying various human emotions.
* Implemented feature engineering techniques to extract meaningful facial features for emotion recognition.

CERTIFICATIONS

* AWS Certified Data Engineer - Associate (**Amazon Inc**)
* Microsoft Power BI Desktop (Maven Analytics | Udemy)
* Data Analytics Certification (Microsoft | LinkedIn Learnings)

ACHIEVEMENTS

* Achieved a 70% reduction in ETL processing time by optimizing Python workflows into PySpark solutions.
* Contributed to $0.5M in business value through innovative data engineering solutions.
* Improved decision-making speed by 30% with enhanced analytics pipelines.