SUSMITHA GOPAVARAPU

+1(203) 606-2944 ◇ East Brunswick, NJ

sususmithasansa@gmail.com ◇ [https://www.linkedin.com/in/susmitha- gopavarapu](https://www.linkedin.com/in/susmitha-gopavarapu) ◇ https://github.com/Susmigop

EDUCATION

Master's in Computer Science, University Of New Haven   01/2022 - 08/2023

SKILLS

Programming Languages:	Python, SQL, Scala,Java, HTML,CSS
Data Processing/ETL:	Apache Spark, Apache Kafka, Apache Airflow
Database Systems:	MySQL,SQL Server,DB2,Oracle(Relational),MongoDB(NoSQL),Talend, Snowflake
Big Data Technologies:	Hadoop ecosystem (HDFS, MapReduce, YARN), Apache Hive
Cloud Platforms :	AWS-Redshift, S3, AWS Glue, EMR, Kinesis, FireHose,Lambda,IAM,
	Azure- Azure Data Factory, Databricks,Blob,Data Lake Storage, Azure Synapse
	GCP-Big Query, Pub/sub, Google Cloud Storage,Cloud Composer
Data Visualization/BI :	Tableau, Pandas, Matplotlib, Seaborn,Power BI
Additional Skills :	Business Requirements,Data Analysis Data Analytics,Data Science,
	Data Warehousing, ETL,RDBMS,Code Review,Creative Problem Solving,

EXPERIENCE

Data Engineer   04/2023 - 04/2024
Aaratech Inc   *Jersey City, NJ*

- Led the implementation of a scalable data warehousing solution on Microsoft Azure utilizing Azure Synapse Analytics, enhancing query performance and maintained a 15%.
- Designed Azure Blob Storage solutions for storing and managing unstructured data, resulting in a 20% reduction in storage costs through optimization techniques.
- Implemented Apache spark as a Batch data streaming platform, achieving a 40% improvement in data ingestion rates and reducing latency by 25%, enabling timely insights for business stakeholders.
- Created and deployed ETL pipelines utilizing Azure Data Factory which increased data accuracy by 19%.

Data Engineer   11/2020 - 01/2022
Wipro Technologies LTD   *Chennai, India*

- Implemented AWS Redshift data warehousing solution to streamline data storage and retrieval processes, enhancing organizational efficiency and facilitating informed decision-making
- Integrated CI/CD pipelines with version control systems like Git, SVN, and Bitbucket to enable seamless code versioning, collaboration and automated triggering of build and deployment processes.
- Managed relational databases such as PostgreSQL, MySQL, and SQL Server, optimizing queries for performance and achieving a 25% reduction in query execution time.
- Architected and deployed real-time data streaming solutions on AWS using Amazon Kinesis services, monitored performance metrics using CloudWatch, and optimized processing efficiency through AWS Lambda functions.

Data analyst   04/2018 - 09/2020
Gebbs Technologies   *Hyderabad, India*

- Proficient in leveraging the Power Platform, particularly Power BI and Power Automate, for creating insightful visualizations and automating data workflows to drive business insights.
- Automated report generation using AWS Data Pipeline and Amazon S3, reducing reporting time by 50% and ensuring timely delivery of critical insights to stakeholders.
- Leveraged Python and pandas to process and analyze large datasets, and got a 23% reduction in data processing time.

- Conducted sentiment analysis on customer feedback using Amazon Comprehend, leading to a 20% improvement in customer satisfaction scores and actionable insights for product improvements.

PROJECTS

Automated ETL Pipeline with Databricks and Delta Lake Integration

- Built ETL pipeline with Databricks, extracting from Irisdata.csv, transforming via PySpark, and loading into Delta Lake and then Employed Spark SQL to enhance readability, creating iris transformed Delta table.
- Orchestrated monthly execution with error handling and data validation for robustness.

E-commerce Returns Management System

- Analyzed online purchase return behavior to optimize business strategies, then cleaned and preprocessed data, handling missing values (average: 3.5%) and correcting attribute values.
- Integrated PyCaret for model building, achieving 82.5% accuracy with algorithms like XGBoost and Random Forest, and conducted chi-squared tests, evaluating model accuracy (average: 81.2%) for predictive performance.

Used Car Price Prediction Model

- Performed EDA on car sales data, revealing market trends such as company market share and annual purchases and Employed various regression models, achieving an average R-squared score of approximately 78% on training and 74% on testing with Random Forest.
- Used Python libraries including NumPy, Pandas, Matplotlib, Seaborn, and scikit-learn for data manipulation, visualization, and modeling, enhancing model performance.

CERTIFICATIONS

- Certified as AWS Certified Solutions Architect - Associate
- Certified as Data Engineer by Udemy.
- Worked as Freelancer in computer science in chegg.