

Maruthi Chadalapaka

chadalapakam@gmail.com | (571) 668-2378 | [LinkedIn](#) | [GitHub](#) | [Medium](#)

OBJECTIVE

Experienced and accomplished Data Engineer with 5+ years of expertise in data modeling, MLOps, Machine Learning, and developing robust pipelines in AWS and Azure. Skilled in designing and deploying efficient data solutions using cutting-edge technologies to generate actionable insights and support data-driven decision-making. Demonstrated success in delivering projects, collaborating across teams, and optimizing data infrastructure for smooth data flow. Seeking opportunities to apply data engineering skills in a role focused on Data Governance, integrating data sources, implementing data quality measures, and driving business outcomes with data-driven strategies.

WORK EXPERIENCE

Graduate Assistant (Azure Data Engineer)

MI

Grand Valley State University

Aug 2022 - Apr 2023

- Developed Python-based AI cognitive analysis software applications utilizing statistical methods.
- Created an ETL pipeline with Azure Data Factory and Azure Blob Storage to convert NoSQL databases into RDBMS, facilitating data flows.
- Implemented data governance and security solutions through Azure Key Vault.
- Established data integration connections with APIs and cloud services.
- Enhanced code quality and decreased application downtime by 35% by implementing automated testing, MLOps, and CI/CD processes for software development.

Student Data Assistant

MI

Grand Valley State University

Aug 2021 - Aug 2022

- Engaged with stakeholders to comprehend data needs, conducted extensive analysis to generate tailored insights and solutions.
- Automated manual data processing tasks through Airflow DAGs, saving 15 hours per week for the data engineering team.
- Utilized PySpark and Azure Databricks to partition and bucket data, resulting in 80% reduction of data processing time.
- Coordinated with teams to comprehend data needs, and developed ETL workflows using Azure Synapse Analytics to meet project deadlines and objectives.

Data Engineer (Real Time Analyst)

India

Amazon

Jan 2020 - Jul 2021

- Developed and optimized ETL pipelines using S3 buckets and AWS Glue to reduce data processing time by 30%.
- Implemented AWS services like Lambda and EC2 to automate data ingestion and storage, leading to a 50% increase in data accessibility.
- Conducted in-depth data analysis to improve data-driven decision-making and resource allocation.

Machine Learning Associate (ML Data Engineer)

India

Amazon

Nov 2018 - Dec 2019

- Developed and implemented supervised machine learning algorithms for inventory classification and predictions.
- Collaborated with cross-functional teams to identify data quality issues and institute data cleansing procedures, reducing data discrepancies by 25% and enhancing the overall accuracy of business intelligence dashboards.
- Developed advanced data cleaning techniques, dimensionality reduction, and k-fold cross-validation for model accuracy assessment with 86% accuracy using AWS Sage Maker

EDUCATION

Grand Valley State University
Masters, Data Science and Analytics, GPA 3.53

MI
Apr 2023

MIST
Bachelor of Technology, Electronics and Communication, GPA 3.50

India
Jun 2016

SKILLS

Azure Data Factory • Azure SQL Database • ETL/ELT • Data Modeling • Azure Data Lake • Apache Spark • AWS S3 • Azure Synapse Analytics • AWS Glue • AWS Lambda • AWS Sage Maker • Python • R • Tableau • Power BI • Data Warehousing • Azure Databricks • DAX • SSIS • SSRS • NoSQL • Data Governance • CI/CD (Continuous Integration/Continuous Deployment) • Apache Airflow

PROJECTS

- **ETL Project in Azure:**

Description: I led an ETL project to manage and process large datasets efficiently. Leveraging Azure Blob Storage and Azure Data Lake as secure and scalable data storage solutions, I designed data workflows and orchestrated data processing using Azure Data Factory.

- **API Integration Pipeline using AWS:**

Description: I have spearheaded development of data pipeline to connect and ingest data from multiple API sources. Utilized AWS S3 to efficiently manage incoming data and AWS Glue for data transformation and created dashboard in Tableau. Integrated Python scripts to handle API requests and process data.

- **Real-time Data Streaming and Processing with Apache Kafka and Apache Spark**

Description: I have designed and implemented a real-time data streaming and processing pipeline using Apache Kafka and Apache Spark. Integrated various data sources into Kafka topics for seamless ingestion and utilized Spark Streaming to process and analyze data in real-time. The project resulted in a significant reduction in data latency, enabling timely insights and actionable decision-making.

- **Regression Analysis and ML Deployment:**

Description: Designed and deployed machine learning pipelines on AWS, including model training, evaluation, and deployment. Utilized AWS services such as Amazon SageMaker, Elastic Kubernetes Service (EKS), CloudWatch, and Lambda for efficient model management, scalability, and automated model retraining. Implemented MLOps practices to establish an automated and continuous model deployment process.

ACCOMPLISHMENTS

- Recognized with the "**Gold Award**" at Amazon for creating an impactful Attendance Dashboard. This dashboard played a crucial role in tracking day-level shrinkage percentage, enabling the operation team to efficiently manage a workforce of 2000 Associates.
- Acknowledged with the prestigious "**Orange Cap**" for my exceptional contributions to the Invent and Simplify initiative one of the tenants of Amazon Leadership Principles.

CERTIFICATIONS

- Received Tableau Analyst Certification from TABLEAU
- Certified AWS Cloud Practitioner
- Azure Data Engineer Certification