

# Maruthi Chadalapaka

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

## 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 AI technologies to generate actionable insights and support data-driven decision-making. Seeking opportunities to apply data engineering skills in a role focused on data governance, integrating data sources, data integrity, implementing data quality measures, and driving business outcomes with data-driven strategies.

## WORK EXPERIENCE

---

### Graduate Assistant

MI

Grand Valley State University

Aug 2022 - Apr 2023

- Developed Python-based AI cognitive analysis software applications with PyCharm IDE by understanding the ethical aspects of AI.
- Created an end-to-end 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.
- Optimized Azure virtual machines and storage resources, resulting in a 40% improvement in application performance and reduced monthly infrastructure expenses by 15%.
- 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

- Collaborated with stakeholders and researchers 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.
- Prepared PySpark and Azure Databricks to partition and bucket data by utilizing DBFS(Data Bricks File System) Delta Lake, resulting in 80% reduction of data processing time.
- Coordinated with teams and developed ETL workflows using Azure Synapse Analytics and Azure Machine Learning.

### Data Engineer

India

Amazon

Jan 2020 - Jul 2021

- Designed and optimized ETL pipelines using S3 buckets and AWS Glue to reduce data processing time by 30%.
- Implemented large scale data models utilizing 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 using Python to improve data-driven decision-making.
- Developed a comprehensive metadata schema, resulting in a 25% reduction in data retrieval time and enhanced data accessibility for the team.

### Machine Learning Associate

India

Amazon

Nov 2018 - Dec 2019

- Optimized S3 resource utilization and reduced processing time for large-scale data analysis by 30% to develop the cost-effective robotic warehouses.
- Conducted performance tuning on AWS EMR clusters and Apache Spark jobs with Python and Hadoop File System, optimizing resource utilization and reducing processing time for large-scale data analysis by 30%.

- Developed and deployed end-to-end ML solutions using AWS SageMaker, resulting in a 25% reduction in model development time and improved predictive accuracy by 15%.
- Employed AWS Lambda functions to automate data ingestion and model retraining, leading to a 30% decrease in manual intervention and faster model iteration cycles.

## EDUCATION

---

Grand Valley State University	MI
Masters, Data Science and Analytics, GPA 3.53	Apr 2023

MIST	India
Bachelor of Technology, Electronics and Communication, GPA 3.50	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 • SQL • DAX • SSIS • SSRS • NoSQL • Data Governance • CI/CD • HDFS • HADOOP • Apache Kafka • DBFS • Meta Data • ETL Pipeline • Fraud Detection • Agile Methodology • Azure Machine Learning

## 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.
- **End to End Data Pipeline using Spark and Databricks:**  
Description: I have designed and implemented a real-time data streaming and processing pipeline for customer interaction. Integrated various data sources into AWS for seamless ingestion. 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

---

- 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