Vidhya Suram

(551) 697-8371 | vidhyareddy131@gmail.com | https://linkedin.com/in/vidhya131 | https://github.com/vidhya131

EDUCATION

Binghamton University, State University of New York, Thomas J. Watson College of Engineering and Applied Science
Master of Science in Computer Science, Artificial Intelligence Track *Expected May 2027*
Cumulative GPA: 4.0/4.0
Keshav Memorial Institute of Technology, Hyderabad, India
Bachelor of Technology in Information Technology *June 2018 – June 2022*
Cumulative GPA: 3.8/4.0

TECHNICAL SKILLS

Languages/Database: Python, Spark, Linux, Scala, Java, Hive, C++, JavaScript, MySQL, PostgreSQL, Snowflake, MongoDB.
Software: Git, GitHub, Bitbucket, VS Code, Jupyter, Pycharm,Putty, Confluence, Jenkins, Docker, Jira, Splunk, Grafana, Kibana.
Cloud: AWS(IAM, S3, EC2, Athena, AWS Lambda, EMR, Glue, QuickSight, Step Functions, VPC, Redshift, CloudWatch, EventBridge, Elastic Search, ECS, DynamoDB, SNS, SQS, Sagemaker), Azure(Databricks, Synapse, ADF).
Frameworks: LangChain, LlamaIndex, Flask, PyTorch, FastAPI, RAG, Aiohttp, Pytest, TensorFlow, Scikit-learn, Keras.

PROFESSIONAL EXPERIENCE

Experian, *Software Engineer* |Hyderabad, India| **Project: Ascend-OPs** *Nov 2021 – Aug 2025*

- Implemented high-performance **Flask REST APIs** secured with **JWT-based authentication and role-based access control (RBAC)**, improving frontend data delivery and reducing **API latency by 10%**, while owning **Oncall support** and **incident resolution** to ensure system reliability.
- Implemented **real-time asynchronous Python Application** using Asyncio**, Aiohttp, Nginx, and WSGI** to handle client requests**, aggregate data from multiple sources,** perform **model inference and scoring,** and store results in **S3 for A/B testing** and **performance benchmarking;** increased **Pytest unit test** coverage from **32% to 96%.**
- Designed & deployed **a Python Flask microservice** on **AWS ECS** using **Jenkins CI/CD**, integrating **Kafka** with **ETL pipelines** to enable real-time failure detection, error analysis, and automated restart, significantly reducing manual operational efforts.
- Migrated **50+ Python Aiohttp workflows** from **mainframe** to **AWS Cloud**, **reducing latency by 20%**.
- Reduced **AWS costs** by implementing **S3 lifecycle policies**, **Spot Instances**, and **auto-scaling** for ETL pipelines.
- Monitored usage and expenses using **AWS Cost Explorer and CloudWatch**, proactively identifying underutilized resources.
- Built scalable data pipelines to test models deployed on the **Ascend-OPS** platform, handling pre and post-processing of **SageMaker** model outputs using **PySpark** on **AWS EMR**, with **Lambda** and **S3** for storage across multiple testing use cases.
- Migrated **60+ Python** models from Ascend-OPs 1.0 to 2.0, optimizing performance and scalability, and built **AWS QuickSight dashboards for monitoring the realtime transactions** while **documenting** the tech stack to onboard new team members.
- **Collaborated with cross-functional teams** to design and implement **scalable workflows** for **pre and post-processing for execution of risk/fraud/score models** on **large-scale consumer datasets**, actively contributing to code and design reviews.
- Designed and orchestrated **large-scale ETL** pipelines using **PySpark, Airflow, EMR, Kafka, and AWS Glue, Lambda, S3, Redshift, Event Bridge** to process and transform dataframes with billions of consumer records.
- Built complex **PySpark scripts to transform data**, executed on **AWS EMR** and orchestrated using **Apache Airflow**.
- Partnered with data scientists to productionize **ML models** by converting research notebooks into automated, fault-tolerant pipelines, building complex **PySpark** scripts for data transformation executed on **EMR** and orchestrated using **Airflow**.
- Built end-to-end workflows to aggregate client transaction streams and **deliver daily summary reports**, and designed **PowerBI** and **Grafana** dashboards integrating **DynamoDB** and **RDS** audit tables to track and analyze client-specific transaction trends, enabling **score and attribute-based billing**. Tech**: PySpark, DynamoDB, RDS, S3, Glue, SNS.**
- Used advanced AI agents (**ChatGPT, Cursor, Gemini, Microsoft Copilot**) to assist with **SQL query design, Python development, refactoring, test generation, and troubleshooting** in data engineering pipelines.
- Designed and optimized complex SQL queries leveraging **CTEs, window functions, advanced joins, GROUP BY/HAVING, and UNION/EXCEPT** to analyze and report on **high-volume datasets at scale**.

PROJECT EXPERIENCE

Linked Post Generator, *https://github.com/vidhya131/linkedin-post-generator.git* *Nov 2025 – December 2025*
*Technologies: **GenAI, Python, Streamlit, LangChain, Groq, llama-3.2-90b-text-preview***

- Built a GenAI LinkedIn post generator that learns an influencer's writing style using topic extraction and few-shot prompting.
- Designed two-stage pipeline with Streamlit and Groq-hosted LLM to generate style-consistent posts by topic, language, length.

ACHIEVEMENTS

- Recognized at Experian with 7 Spot Awards and a 5/5 Exceptional Performer rating for strong delivery and ownership.
- **Data Structures & Algorithms:** Solved 200+ challenges on LeetCode(link) and 150+ on GeeksforGeeks(link).