HAREENA CHOWDARY POLAVARAM

(734)-489-2831 | hareena.cp@gmail.com | Ann Arbor, MI | **Ready to Relocate**

https://www.linkedin.com/in/hareena-cp/ | https://hareenacp.vercel.app https://github.com/HareenaChowdary

SKILLS

Cloud Platforms: AWS (S3, EC2, Redshift, Glue, CDK), Azure (ADF, Databricks, Data Lake, Synapse), BigQuery
Big Data & ETL Tools: PySpark, SparkSQL, Delta Lake, Kafka, Airflow, Azure Data Factory
DevOps & CI/CD: Azure DevOps, Jenkins, Docker, Kubernetes, Terraform
Languages & Scripting: Python, SQL, R, Shell Scripting, HTML, CSS, JavaScript
Databases: Snowflake, PostgreSQL, MySQL, SQL Server, DynamoDB, Redshift, Synapse
Visualization & Tools: Power BI, Tableau, Excel VBA, Git, Jupyter, VS Code, JIRA, Splunk

EXPERIENCE

Data Engineer | ***Tata Consultancy Services (Client: Aviva), India*** ***Aug 2021 – Jul 2023***

- Engineered 20+ scalable, production-grade ETL pipelines using Azure Data Factory and Databricks, enabling ingestion of structured and semi-structured data from distributed sources and reducing manual effort by 40%.
- Designed optimized Azure Data Lake storage with dynamic partitioning, improving data accessibility and cutting storage overhead by 15%.
- Processed 500M+ transactional records using PySpark and SparkSQL, enhancing pipeline efficiency and reducing compute resource consumption by 25%.
- Refactored SQL logic powering executive reporting layers, reducing latency from 10 minutes to <2 minutes for real-time insights.
- Built modular validation/enrichment components for end-to-end pipelines, increasing pipeline robustness and data consistency.
- Led the cloud migration from on-prem to Azure, improving data infrastructure reliability and reducing maintenance overhead by 40%.
- Orchestrated CI/CD deployment pipelines with Azure DevOps, accelerating release cycles by 35% and improving deployment consistency.

AI Intern | ***National Instruments (Cognibot), Remote*** ***May 2020 – Jun 2020***

- Constructed and deployed ML models for predictive maintenance using 100K+ IoT industrial telemetry points from sensors and devices; achieved >85% anomaly detection accuracy across test environments.
- Integrated real-time edge analytics to trigger alerts with sub-2s latency, improving responsiveness to mechanical faults.
- Enhanced model pipelines through hyperparameter tuning and feature engineering, increasing overall F1 score by 18%.

Software Developer & CS Lab Assistant | ***Sri Venkateswara University, India*** ***Jun 2019 – Mar 2020***

- Automated lab infrastructure setup using Bash and Python, reducing environment provisioning time by 30%.
- Created internal web-based utilities for lab admin tasks using HTML, CSS, SQL, and shell scripting.
- Mentored 200+ students on debugging, OS-level scripting, and software tools within diverse lab setups.

EDUCATION

Western Michigan University *Kalamazoo, MI* | ***Aug 2023 – Apr 2025***
Master of Science in Computer Science *GPA: 3.78/4*

Sri Venkateswara University *Tirupati, India* | ***Jul 2017 – Jul 2021***
Bachelor of Technology in Computer Science *GPA: 3.6/4*

Coursework: Machine Learning, AI, Big Data Analysis using Python, Algorithms, Advanced Databases Concepts, Statistics

PROJECTS

COVID-19 Data Engineering Pipeline

- Designed and deployed an Azure-based data engineering pipeline to ingest and transform COVID-19 data from ECDC and Eurostat using Azure Data Factory, HDInsight, and Databricks.
- Enabled real-time Power BI dashboards for trend analysis across 30+ countries and supported ML workflows to predict virus spread using curated datasets.

F1 Race Analytics Platform

- Created a scalable F1 race analytics lakehouse by ingesting Ergast API data into Azure Data Lake, transforming it with PySpark in Databricks, and storing with ACID compliance in Delta Lake.
- Automated 20+ ETL workflows via Azure Data Factory and delivered Power BI dashboards covering 1,000+ races and 70+ circuits for race insights and historical comparisons.

Azure End-to-End Data Engineering Project

- Developed a production-grade Azure data platform with Data Factory, Databricks, Delta Lake, and Unity Catalog; implemented SCD Type 2 logic and automated CI/CD pipelines via GitHub.
- Modeled a star schema in the gold layer for business analytics, enforced enterprise data governance using Unity Catalog, and visualized 15+ KPIs through Power BI dashboards.

IPL Match Winner Prediction (R + Shiny)

- Architected an interactive R Shiny dashboard to forecast IPL match outcomes using ensemble models (Random Forest, XGBoost); derived predictive features like venue strength, toss bias, and head-to-head performance.
- Conducted comprehensive EDA using ggplot2 and deployed a real-time prediction interface enabling users to explore dynamic match scenarios and outcomes.