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, Lambda, Redshift, Glue), Azure (Azure Data Factory, Databricks, Data Lake, Synapse), GCP (BigQuery)
Big Data & ETL Tools: PySpark, SparkSQL, Delta Lake, Kafka, Azure Data Factory, Informatica PowerCenter, Informatica BDM
DevOps & CI/CD: Azure DevOps, Jenkins, Docker, Kubernetes, Terraform
Languages & Scripting: Python, SQL, HTML, CSS, JavaScript
Databases: MySQL, SQL Server, PostgreSQL, Oracle, DynamoDB
Visualization & Tools: Power BI, Excel VBA, Git, Jupyter, VS Code, JIRA, Grafana, Splunk

EXPERIENCE

Data Engineer | ***Tata Consultancy Services, India*** ***Aug 2021 – Jul 2023***

- Built and maintained scalable ETL pipelines using Informatica PowerCenter, Informatica BDM, Azure Data Factory, and Databricks to ingest, transform, and load multi-source data for enterprise reporting.
- Optimized SQL queries and transformations in Oracle and Databricks, reducing execution time by ~25% and improving pipeline efficiency.
- Developed complex mappings and datasets using transformations such as Lookup, Joiner, Aggregator, Router, Rank, and Source Qualifier to meet analytical requirements.
- Validated and cleansed data through automated checks in ADF and SQL, achieving 99% data accuracy across environments.
- Migrated ETL mappings, datasets, and workflows across development, QA, and production environments, ensuring smooth deployments with minimal downtime.
- Translated business requirements into technical STTM documents and mapping logic, enhancing traceability and reducing rework by 15%.
- Resolved production defects and data issues reported by clients, maintaining 100% SLA compliance and supporting data quality initiatives.
- Tracked project tasks, progress, and issue resolution using JIRA and Azure DevOps while collaborating with cross-functional teams on cloud migration and data transformation projects.

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%.

EDUCATION

Western Michigan University *Kalamazoo, MI* | ***Aug 2023 – Apr 2025***
Master of Science in Computer Science *GPA: 3.7/4*

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

Azure Databricks End-to-End Data Engineering Pipeline

- Built an end-to-end data pipeline on Azure (Data Lake, Data Factory, Databricks, PySpark, Delta Lake) to ingest, transform, and standardize raw racing datasets into bronze, silver, and gold layers for analytics.
- Implemented incremental ingestion, schema validation, and optimized Delta/Parquet storage, enabling scalable, reusable, and BI-ready datasets for downstream reporting in Power BI/Synapse.

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.

AdventureWorks Data Engineering Project using Azure Stack

- Designed and deployed a data engineering pipeline on Azure (Data Lake, Data Factory, Databricks, Synapse, Power BI) to ingest, transform, and model AdventureWorks sales and customer data for advanced analytics.
- Implemented ETL with PySpark and Delta Lake to build curated data layers, applied data quality checks and performance optimizations, and delivered BI-ready datasets for interactive reporting in Power BI.

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 Data Engineering Project

- 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.