Hareena Chowdary Polavaram

734-489-2831 hareena.cp@gmail.com /hareena-chowdary /HareenaChowdary Kalamazoo, MI

TECHNICAL SKILLS

Data Engineering and Cloud Technologies: Apache Spark, Apache Kafka, Airflow, Azure (Data Factory, Databricks, Synapse Analytics, Data Lake Storage, Devops, Kubernetes), AWS (EC2, S3, Redshidt), Docker, ETL Processes

Data Visualization: Power BI (DAX, Power Query), Tableau, Excel (Advanced, VBA, Pivot Tables), Statistics (ANOVA, Hypothesis Testing, Regression Analysis), Machine Learning modules

Database: MySQL, PostgreSQL, Azure SQL, Oracle SQL, MS SQL, Azure Data Studio, BigQuery

Languages & Libraries: Python (Pandas, NumPy, Matplotlib, PyTorch, TensorFlow, Scikit-learn, Seaborn, OpenCV), R, T-SQL, C, Java, HTML, CSS, Javascript

Tools & Platforms: Jupyter Notebooks, R Studio, Git(Github, Gitlab, Git Extensions, Bigbucket), Pycharm, VS Code, Winscp, Putty, JIRA, Confluence, MS Office

EXPERIENCE

Data Engineer, Tata Consultancy Services (Aviva Client) — Aug 2021 - Jul 2023

- Designed and streamlined ETL workflows in Azure Data Factory (ADF) and Databricks, enhancing data movement, transformation, and integration across structured and unstructured sources.
- Optimized data pipelines using PySpark, reducing processing time by 30% and improving model performance through efficient parallel execution.
- Developed scalable data lakes in Azure Data Lake Storage (ADLS), ensuring 99.9% uptime while cutting storage expenses by 15% through optimized partitioning and lifecycle policies.
- Processed large-scale insurance datasets (claims, medical records, and third-party APIs) with PySpark and SQL, ensuring compliance, accuracy, and consistency in critical reporting.
- Built automation frameworks for data validation, transformation, and enrichment, increasing data reliability by 20% and minimizing manual intervention.
- Restructured SQL queries to boost performance, reducing execution time by 25% and cutting report generation from 10 minutes to under 2 minutes.
- Led the transition of legacy data processes to Azure, improving scalability, reliability, and operational efficiency, while reducing maintenance efforts by 40%.

EDUCATION

Master of Science in Computer Science — Aug 2023 - Apr 2025

Western Michigan University, Kalamazoo, Michigan — GPA - 3.78/4

Relevant Coursework: *Machine Learning, RDBMS, Azure Databricks and Spark (PySpark / SQL), Big Data Analysis, R*

Bachelor of Technology in Computer Science — Jul 2017 - Jul 2021

Sri Venkateshwara University, Tirupati, India — GPA - 7.47/10

PROJECTS

COVID-19 Analysis (Azure Data Factory, Azure Storage, Databricks, Kafka, DevOps) — **Jan 2025 - Feb 2025**

- Developed end-to-end data pipelines in ADF, integrating multiple sources, including real-time Kafka streams, to facilitate faster COVID-19 insights.
- Refined data transformation logic, reducing pipeline development efforts by 25% while improving overall processing speed for batch and streaming workloads.
- Established an automated CI/CD process using Azure DevOps, integrating Kafka for real-time data handling and ADF for orchestration, cutting release cycles by 35% and downtime by 20%.

F1 Data Analytics (Azure Databricks, Delta Lake, Pyspark, SparkSQL, PowerBI) — **Nov 2024 - Dec 2024**

- Transformed raw Formula 1 race data from 8+ sources (CSV, JSON, API) into structured, queryable datasets using Azure Databricks & Delta Lake, ensuring seamless integration for analytics and reporting.
- Curated historical and real-time race data spanning 70+ years, enabling predictive insights, driver/team performance rankings, and Power BI dashboards for in-depth analysis.
- Streamlined data ingestion and transformation workflows with incremental data handling & automated scheduling, achieving 99.9% data accuracy and compliance with GDPR & time-travel capabilities

CERTIFICATIONS & PUBLICATIONS

- Foundations: Data, Data, Everywhere - Google
- Data Analytics Essentials, Introduction to Data Science - Cisco
- Databases and SQL for Data Science with Python, ETL and Data Pipelines with Shell, Airflow and Kafka - IBM
- "De-Authentication Attacks on Rogue UAVs", ICISS (IEEE)