





SACHIN LODDIYA KARTHIK

☎ 517-243-9938 ✉ sachinlkece@gmail.com  [Sachin](#)  [GitHub](#)  [Portfolio](#)  [Open to relocation](#)

SUMMARY

Data Engineer with 3+ years of experience designing and optimizing scalable data pipelines that improved processing efficiency by up to 40%. Expert in **PySpark, SQL, Azure Databricks, and Synapse**, with proven expertise in transforming complex raw data into actionable business intelligence for enterprise-scale solutions.

TECHNICAL SKILLS

Data Engineering: Apache Spark, Apache Kafka, Apache Airflow, Databricks, Docker

Cloud Platforms: Azure (Data Factory, Synapse Analytics, DevOps), AWS (EC2, S3, Glue, Athena, Redshift), GCP (BigQuery, Pub/Sub, Dataflow)

Programming: Python, R, T-SQL, Java, C, C++, HTML, CSS, JavaScript

Databases: Azure SQL, Oracle SQL, MySQL, MS SQL Server, PostgreSQL

Visualization: Power BI, Tableau

ML & Libraries: Pandas, NumPy, Scikit-Learn, TensorFlow, PyTorch, OpenCV

Statistical Analysis: ANOVA, Hypothesis Testing, Regression Analysis

Tools: PyCharm, VS Code, Jupyter Notebook, R Studio, WinSCP, PuTTY, MS Office

EXPERIENCE

Data Engineer — WMU, USA

Jan 2025 – Apr 2025

- Designed an automated class scheduling solution using **Google OR-Tools** and **constraint programming**, reducing manual scheduling time by **85%** and optimizing 200+ course assignments across 15 departments.
- Created and deployed a **Streamlit**-based UI to process 50+ CSVs with real-time validation, reducing data input errors by **90%** and eliminating over **40 hours/week** of manual effort.

Data Engineer — Green Expectations LLC, USA

Jan 2024 – Apr 2024

- Built backend infrastructure for an AI-powered sustainability platform using a rule-based **House Sustainability Calculator**, processing **1,000+ data points/user** and improving recommendation accuracy by **35%**.
- Implemented secure user authentication with persistent session storage and optimized data flows for an **NLP-based AI chatbot**, reducing latency by **40%** and increasing engagement by **25%**.

Data Engineer — Accenture, India

Jul 2021 – Jul 2023

- Engineered ETL pipelines using **Azure Data Factory** to reformat supply chain data into **Parquet**, improving processing efficiency by **40%**.
- Processed raw supply chain data in **Azure Databricks** from **Data Lake Gen2**, boosting accuracy by **25%** and halving transformation times.
- Developed external tables in **Azure SQL** and interactive **Power BI dashboards**, improving data access and increasing operational efficiency by **25%**.
- Integrated ADF triggers, Databricks notebooks, and CI/CD pipelines with **Azure DevOps**, accelerating delivery by **30%**.

Data Engineer — Claritrics India Pvt Ltd, India

Nov 2020 – May 2021

- Architected an ETL pipeline using **Azure Data Factory** and **Databricks** to integrate OCR-extracted data, improving document processing accuracy by **30%**.
- Automated recurring workflows and enhanced reliability, reducing processing time by **40%**.
- Enabled quicker deployments by **30%** with modular coding practices and containerized environments.

EDUCATION

Master of Science in Data Science — Western Michigan University

Aug 2023 - Apr 2025

GPA: 3.86/4.0

Kalamazoo, MI

- Relevant Coursework:** Machine Learning, RDBMS, Azure Databricks & Spark (PySpark / SQL), Applied Linear Models, Big Data Analysis, Google Cloud Big Data & ML Fundamentals

PROJECTS

Netflix Data Engineering Project (Azure Data Factory, Databricks, PySpark, Power BI)

Jul 2024 - Aug 2024

- Built a scalable and automated end-to-end data pipeline for Netflix datasets using **Azure Data Factory**, **Databricks** with **PySpark**, **Delta Live Tables**, and **Synapse**, enabling efficient incremental data ingestion, processing, and storage within a multi-layered architecture.
- Created interactive **Power BI** dashboards to provide key stakeholders with actionable insights, supporting data-driven business decisions and enhancing reporting capabilities.

F1 Data Analytics Project (Azure Synapse, Databricks, PySpark, SparkSQL)

May 2024 - Jun 2024

- Engineered end-to-end **ETL pipelines** in Azure Synapse to integrate Ergast API data, automating execution with triggers and refining workflows via **PySpark** and **SparkSQL**.
- Implemented **Azure Monitor** for pipeline metrics tracking and established robust schema enforcement and monitoring protocols to guarantee data quality and reliability.