(+91) 9944088680
Chennai, India
nkarthick989@gmail.com

Karthick N

Azure Data Engineer

github.com/KarthickN2000
linkedin.com/in/karthick989

Professional Summary

Azure Data Engineer with 3 years of hands-on experience in building scalable, high-performance data pipelines for both batch and streaming workloads. Proficient in developing end-to-end ETL/ELT pipelines using Python, PySpark (Spark 3.x), SQL, and Databricks. Experienced in implementing the Medallion Architecture (Bronze, Silver, Gold layers) using Delta Lake, focusing on data quality, transformation, and validation across structured and semi-structured data sources. Skilled in using tools like Great Expectations and pytest for testing and validation, and working with cloud platforms (Azure, AWS) for data lake design and CI/CD using Azure Pipelines and Git. Exposure to orchestration tools like Apache Airflow and Azure Data Factory, along with familiarity in working with Kafka, HDFS, and NoSQL systems.

Skills

- **Big Data & Processing:** Apache Spark, PySpark , Python , Azure Databricks
- **Workflow Orchestration:** Apache Airflow, Azure Data Factory,Azure Synapse analytics
- **Cloud Platforms:** Azure Data Factory, AWS S3,„GCP
- **Version Control:** Git, GitHub
- **Operating Systems:** Linux
- **Data Engineering Concepts:** ETL/ELT Pipelines, OLAP
- **BI Tools:** Power BI
- **Database:** MYSQL,PostgreSQL
- **Tools & IDEs:** VS Code, Jupyter, PyCharm

Technical Experience

Data Engineer(Azure) **SEP 2022 — Present**
Tata Consultancy Services *Chennai, India*

- Provided system-level support and maintenance for enterprise applications and infrastructure.
- Automated repetitive manual tasks using Python and Shell scripting to improve efficiency.
- Worked in Linux environments for server-side support, file handling, and cron job scheduling.
- Used SQL, MySQL for log analysis, data extraction, and reporting in internal tools.
- Gained hands-on familiarity with Git, basic cloud operations, and scripting — building a foundation for data engineering.
- Pursuing a career transition into Data Engineering through real-world project development in PySpark, Airflow, and Databricks.
- Built real-time streaming POC using Kafka and Spark Structured Streaming for log analysis.
- Familiar with Flink and streaming architecture patterns for high-throughput systems.

Projects

ADF-Driven Delta Lake Pipeline with PySpark Transformations **July 2025**

- Tools: PySpark, Python, SQL, Azure Databricks, Azure DataLake, Delta Lake, Azure Data Factory,Azure Synapse Analytics.
- Built a cloud data pipeline in Azure Data Factory (ADF) to orchestrate data movement from Azure Blob Storage to Azure Databricks. Implemented Medallion architecture (Bronze, Silver, Gold) using Delta Lake for structured data preprocessing. Used to perform data cleaning, transformations, and aggregation with PySpark. Configured ADPySpark. Configured to automate the workflow. Project demonstrates end-to-end integration workflow. Project for scalable data engineering solutions. (Bronze/Silver/Gold).

E-Commerce Data Engineering Pipeline **July 2025**

- Tools: PySpark, Databricks, AWS S3, Delta Lake, Airflow, Python, SQL.
- Built an end-to-end ETL pipeline using PySpark and Databricks following Medallion architecture (Bronze/Silver/Gold).
- Ingested raw data from AWS S3, applied cleaning, transformations, and schema enforcement.
- Developed KPIs such as customer segmentation, product popularity, and basket analysis.
- Applied Python OOP for reusable transformation logic and simulated Airflow DAGs for orchestration.
- Used Delta Lake optimizations (partitioning, compaction) to improve query performance.

Education

Bachelor of Engineering in Mechanical, *Thiagarajar College of Engineering Grade 9.45* MAY 2022