

SWETHA R

DATA ENGINEER

+91 7871420012 • swethaselviraj@gmail.com • Puducherry, India
Nationality: Indian • Date of Birth: 23 May 2003 • Languages: English, Tamil

A skilled Data Engineer trained in Python, PySpark, SQL, and Azure Databricks with hands-on experience in building data pipelines and optimizing data workflows. Passionate about leveraging data engineering techniques to deliver actionable insights and support business decision-making, while continuously advancing my technical expertise and career.

EXPERIENCE

Data Engineer

Oct 2024 – Present

Hexaware Technologies Ltd.

- Data Engineer working on data engineering solutions using Azure Synapse Analytics, Apache Spark, PySpark, SQL, and Databricks.
- Re-engineered SAS data steps and PROC transformations into PySpark and Spark SQL in Azure Synapse and Databricks, ensuring functional parity and schema consistency.
- Built parameterized, modular ETL pipelines in Azure Synapse (notebooks and pipelines) integrating Azure Storage, with robust error handling, auditing, and lineage tracking.
- Implemented rigorous data reconciliation — row counts, hash checks, variance thresholds, and business-rule validations — to guarantee zero data loss or mismatch.
- Optimized big-data workloads via partitioning, efficient file formats (Parquet and Delta), broadcast joins, and caching to meet SLA and cost objectives in the cloud.

PROJECTS

Real-Time Social Media Sentiment Analysis

Processed social media data from streaming APIs and CSV files in real time for sentiment analysis and actionable insights — tracking sentiment trends, user engagement, and topic popularity. Built databases for posts and sentiment scores, cleaned data with PySpark, categorized posts as positive / negative / neutral, and built a pipeline with Delta Live Tables for real-time insights. **Python • MSSQL • PySpark • Azure Databricks • Google Colab**

Stock Market Data Analysis

Built a data warehouse to store and analyze stock market data — prices, trading volumes, and indicators. Designed the schema, collected and cleaned data from APIs and CSV files with Python, wrote SQL queries to detect price anomalies, and built real-time PySpark pipelines to analyze market trends and indicators such as moving averages and RSI, helping detect sudden market shifts. **Python • MSSQL • PySpark • Azure Databricks • Google Colab**

SKILLS

LANGUAGES

Python SQL

TOOLS & TECH

PySpark Azure Databricks Azure Synapse
Azure Data Factory Azure Data Lake
Azure DevOps MSSQL Google Colab

DATA ENGINEERING

ETL pipelines, data warehousing, Delta Live Tables, workflow optimization, real-time processing

OPERATING SYSTEMS

Windows

EDUCATION

B.Tech, Electronics & Communication Engineering

Sri Manakula Vinayagar Engineering College, Puducherry • 2024

Higher Secondary (12th)

Amalorpavam Hr. Sec. School, Puducherry • 2020

SSLC (10th)

Immaculate Heart of Mary Girls Hr. Sec. School, Puducherry • 2018

CERTIFICATIONS

- PCEP — Certified Entry-Level Python Programmer
- AZ-900 — Microsoft Azure Fundamentals