



SWATHI THOKALA

Kansas, US (Open to Relocation)

📞 660-580-5129 ✉ swathireddythokala16@gmail.com  [linkedin](#)  [github](#)

EDUCATION

University of Central Missouri

Master of Science in Computer Science

Jan 2023 – May 2024

Gpa – 4.0 / 4.0

Jawaharlal Nehru Technological University

Bachelor of Technology in Computer Science

Aug 2016 – Sep 2020

Gpa – 8.27 / 10

EXPERIENCE

Tata Consultancy Services

Software Engineer (Data)

Feb 2021 – Dec 2022

Hyderabad, India

Developed a data ingestion pipeline to integrate daily batch data from Vendor Server to IBM DB2 data warehouse.

- Designed and implemented a batch **data ingestion pipeline** with **Java** to integrate end-of-day retail sales data from **CSV / Flat** files into **DB2 database** achieving monthly cost savings of \$15k as a result of automation.
- Created and deployed **Python** and **Unix** scripts to aggregate results from the pipeline and implemented an automated email system to provide **daily report** on the files processing status eliminating need for manual monitoring.
- Leveraged advanced **SQL** techniques such as **joins, CTEs, and query optimization** to migrate data from **DB2** to vendor server and orchestrated the pipeline using **cron** jobs.
- Collaborated with clients and 2 cross-functional teams to apply business rules on the data and created design plans.

Re-engineered an ETL pipeline for XML files, cutting pipeline processing time from 40 to 1 minute.

- Developed an ETL pipeline replacing **SSIS** and **C#** with **Java** and **Xpath Selectors** to process Sales Orders XML files.
- Optimized memory footprint of the pipeline by 75% with **stream processing** and reduced the server load.
- Utilized **Git** for **continuous integration and continuous deployment (CI/CD)** processes and experienced with **Agile/Scrum** methodologies as well as the **Waterfall** software development life cycle (SDLC).
- Solved 5+ **production level issues** and provided the resolutions for **high priority incidents** and created the **documentations** for **Root Cause Analysis**.

ACADEMIC PROJECT EXPERIENCE

Youtube Trend Analysis Pipeline | *Python, Airflow, Docker, AWS S3, PostgreSQL* | [See Article](#)

May 2024 – Jun 2024

- Created a fully automated ETL pipeline by following the **Medallion Architecture** with **Pyspark** and **Airflow** to identify the daily top 5 Youtube trending videos from 30+ countries based on views and likes.
- Implemented pipeline that extracted data from the **YouTube Data API** and transformed it using PySpark.
- Designed the database using a **Star Schema** with 3 dimension tables and 1 fact table. Utilized **GitHub** for **CI/CD** and containerized the ETL pipeline with **Docker**.
- Developed a materialized view that refreshes daily at 12 AM Midnight and built a **Tableau** Dashboard to visualize data insights.
- [View Tableau Dashboard](#)

CERTIFICATIONS

- **AWS Certified Data Engineer Associate** – [See Certificate](#)
- **SQL Intermediate Hackerrank Certificate** – [See Certificate](#)

SKILLS

- **Languages:** Java, C#, Python, SQL
- **Database Systems:** MySQL, PostgreSQL, NoSql(MongoDb, Neo4j)
- **Cloud:** AWS
- **Tools and Frameworks :** Apache Airflow, Apache Camel, Docker, Git, Tableau, Jupyter Notebook
- **Big Data :** Apache Spark(PySpark) , ETL, Data pipeline, Data Warehousing(Star, Snowflake)

ACHIEVEMENTS

- **Academic Excellence Award** for ranking first among 10,000 students in the university.
- Inducted into **Phi Kappa Phi**, a National Honor Society that honors high academic achievement.