

# Shanmuganathan S

[shanmuganathans33s@gmail.com](mailto:shanmuganathans33s@gmail.com) | [LinkedIn](#) | [GitHub](#) | Chennai, India | +91 7339242953

## Professional Summary

Results-driven IT professional with 3.10 years of experience, including data engineering and automation. Key achievements include developing ETL pipelines for a banking and finance client and automating internal processes. Technical expertise encompasses data pipeline development, big data technologies (Hadoop, Spark, Airflow), Python & MySQL. A proactive and collaborative team member with good communication skills.

## Skills

- **Programming Languages:** Python, C#
- **Packages & Framework:** Pandas, NumPy, PySpark, Matplotlib, Flask & Streamlit
- **Database:** SQL, MySQL
- **Data Technologies:** Apache Spark, Apache Airflow, SQL, Data Modeling, Data Warehousing
- **Development Practices:** Agile, Continuous Integration/Continuous Deployment (CI/CD), Microservices, REST APIs
- **Cloud Platforms:** Azure, AWS, Snowflake & databricks
- **Tools & Other:** Git, Pycharm, VS Code, Jupyter Notebook & Postman

## Experience

### Infosys Ltd

Data Engineer

Chennai, India

May 2021 – Present

#### Project 1 - Real Time Data migration on lateral recruitment process

- Designed and implemented a robust ETL process to automate the lateral hiring process within Infosys.
- Built **PySpark pipelines** from scratch to ingest candidate details from candidate application form to recruitment portal, every day via Airflow – processing around **700GB data in under 20 minutes with 96% success rate** – creating daily, weekly, monthly level data with incremental loads.
- Created integrated reports and dashboards using SQL and Python to provide real-time insights into the hiring process on infosys recruitment platform.

#### Project 2 – Data Analysis on Housing Finance Data

- Analyzed LIC HFL data to identify key customer trends and predict potential loan defaults.
- Developed a scalable data pipeline using Apache Spark and PySpark for distributed data processing and ETL.
- Implemented workflow orchestration and scheduling with Airflow, automating data ingestion and processing around **1TB data in under 30 minutes with 98% success rate** and stores final transformations to MySQL Server.
- Created dynamic, interactive visualizations with Streamlit and Plotly, enabling stakeholders to explore data insights. Deployed the Streamlit application on Virtual Machine.

## Achievements

- Awarded **1<sup>st</sup> Place** in **Infosys Quality PACE Competition 2024**.
- **2<sup>nd</sup> Runner Up** in **Digital Hackathon 2024** Conducted on Infosys Ltd.

## Awards

- **Ramp Award** for outstanding performance at Infosys Ltd., May 2023.
- **Spot Award** for outstanding performance at Infosys Ltd., Sep 2022.
- **Rising star** award at Infosys., Apr 2022.

## Certifications

Microsoft Azure Certifications – Azure Data fundamentals(DP-900).

Infosys Certification – Infosys Certified Python Programmer,

Infosys Certified Database and SQL Professional,

Infosys Certified PySpark Professional,

Infosys Certified MySQL Associate.

## Education

Kongu Engineering College, Erode, Tamil Nadu, India

Bachelor of Engineering, Electronics and Communication Engineering

July 2016 - June 2020