

ROSHAN FAREED N

Aspiring Data Engineer | PySpark · Databricks · SQL · Azure

✉ roshanfareed53@gmail.com ☎ +91 86084 00363

📍 Alagankulam, Ramanathapuram, Tamil Nadu - 632512 🌐 [linkedin.com/in/roshanfareed53](https://www.linkedin.com/in/roshanfareed53)

🐙 github.com/ROSHANFAREED 🔗 <https://ROSHANFAREED.github.io>

Professional Summary

Aspiring Data Engineer with hands-on experience in building **end-to-end Azure data pipelines** using **Azure Data Factory, Databricks (PySpark), ADLS Gen2, Azure SQL, and Power BI**. Strong in ETL design, data transformation, cloud storage optimization, and analytics-ready data modeling. Proven ability to design production-style pipelines with security-aware practices and scalable architecture. Foundational understanding of Machine Learning concepts (no production implementation)

Core Skills

Data Engineering & Cloud (Azure):

Azure Data Factory (ADF), Azure Databricks (PySpark, SparkSQL), ADLS Gen2, Azure SQL Database, Azure Synapse (exposure), Microsoft Fabric (exposure)

Programming & Data Processing:

Python (Pandas, NumPy), PySpark, SQL, Spark SQL

Version Control & Dev Practices:

Git, GitHub, Basic CI/CD concepts, Secure coding practices

Databases & Data Modeling:

MySQL, SQL Server, Joins, Subqueries, Window Functions, Stored Procedures, Aggregations

Business Intelligence:

Power BI (DAX, Data Modeling, Dashboards), Tableau, Advanced Excel

Modern Data Stack:

dbt (in progress), Snowflake (in progress)

Projects (GitHub Portfolio)

COVID-19 Azure Data Engineering Pipeline (Flagship Project), Tech: Azure Data Factory, Databricks (PySpark), ADLS Gen2, Azure SQL, Power BI, GitHub

- Designed and implemented an **end-to-end Azure Data Engineering pipeline** ingesting COVID-19 datasets from public APIs and blob storage using **Azure Data Factory**.
- Stored raw and curated datasets in **ADLS Gen2 (Raw & Clean zones)** following data lake best practices.
- Transformed datasets using **ADF Mapping Data Flows** and **Databricks (PySpark)** for population-level analytics.
- Loaded curated data into **Azure SQL Database** to support downstream BI and analytics workloads.
- Built **Power BI dashboards** on Azure SQL to analyze cases, hospital admissions, testing trends, and population distribution.
- Implemented **secure coding practices** by externalizing secrets and preparing the pipeline for CI/CD deployment.

Retail Sales Data Engineering Pipeline using Databricks (Medallion

11/2025 – 12/2025

Architecture, *Tech: Databricks, PySpark, Delta Lake, Power BI, CSV/JSON, SparkSQL*

- Built an end-to-end ETL pipeline using Bronze–Silver–Gold layers to process raw retail data.
- Performed data cleaning, schema enforcement, null handling, and multi-format date parsing using PySpark.
- Created Gold-level aggregates including **Total Sales by Category, Order Trends, AOV, and Avg Price per Item.**
- Designed an interactive analytics dashboard visualizing customer behavior and revenue insights.
- Implemented transformation logic using Delta Lake for optimized storage and incremental updates

E-commerce Online Retail Analytics, *MySQL, SQL Server, Power BI*

- Implemented advanced DAX measures and drillthroughs, improving dashboard performance by 25%.
- Delivered insights on purchase patterns and profitability drivers, leading to strategic recommendations.

Professional Experience

Quality Inspector, *Pro Candour Automotive Pvt Ltd, Chennai*

2019 – 2021

- Conducted end-to-end inspections ensuring 100% compliance with quality standards.
- Developed Excel-based reporting system that reduced manual reporting time by 40%.Built visual dashboards for weekly review meetings, enabling management to make quicker decisions.
- Collaborated with production and engineering teams, reducing defect rates by 15% through corrective actions.

Education

B.E. Mechanical Engineering, *AMS College of Engineering, Anna University*

2019

CGPA: 6.56/10

Key Achievements

- Completed **end-to-end data engineering and analytics projects**, demonstrating strong understanding of ETL pipelines, PySpark transformations, SQL, and dashboard development.
- Automated reporting workflows using Excel & Python, reducing manual effort and improving accuracy.Improved quality control in automotive operations by **reducing defect rates by 15%** through data-driven insights.

Certifications

Google Data Analytics Professional Certificate – Coursera
(Completed)

Microsoft Power BI Analyst Certificate – Coursera
(Completed)

Microsoft PL-300 (Power BI Data Analyst)
– In Progress

Availability:

Immediate Joiner | Open to Relocation