# Harsh Fadiya

DATA ENGINEER

## Personal Profile

+919662514036
fadia.harsh@gmail.com

## Links

LinkedIn

Profile page

Github

Medium Blog

## Skills

Python, PySpark, MySQL, Databricks

Cloud (Azure, AWS, BQ), Airflow, Azure DevOps, Git

Data Structures, Algorithms, OOP

Confluence, Agile, SDLC, Deployments

## Hobbies

- Music
- Trekking
- Swimming
- Reading

## Languages

English

Hindi

Gujarati

# Career Highlights

As an experienced Databricks and Microsoft Azure certified Data Engineer with a background in Software Engineering, I thrive on challenging opportunities that allow me to apply my expertise in **Python, Spark, SQL, Databricks** and adequate experience in **Azure/AWS**. My cloud enthusiasm drives me to seek continuous learning and deliver exceptional results, making me a valuable asset in achieving organizational objectives. Let's embrace new challenges together!

# Professional experience

### Data Engineer [Contractor], Databricks

11/2023 - Present

- **Metadata-Driven DLT Pipeline:** Involved in building DLT with Python, Spark, MySQL, Databricks for batch and streaming data; integrated Medallion with robust Data Quality, Unit Testing.
- **Excel Parsing Python Framework:** Engineered using Python OOP, Pandas for efficient data extraction and manipulation from Excel files.

### Data Engineer, Koantek Cloud and AI Services, Pune

04/2022 - Present

- Trained in **Databricks, Apache Spark** and built data pipelines for Data Ingestion, Transformation and Load processes (_ETL/ELT_).
- Proficient in **Python, Spark, MySQL, Databricks**.
- Delivered end to end data pipelines with Medallion, Data Quality and optimizations using **Azure Data services**.
- Experience of **consulting Use-cases** to provide the best solutions to the client for Data orchestration needs and optimization of the data-pipelines.
- Worked on AWS Databricks **E2 migration** upgrades with best practices implementations.
- Migrated ~130 SQL from GCP BQ to Databricks with performance optimizations and Airflow integrations.

### Technology Analyst, Infosys Ltd, Pune

01/2019 - 03/2022

- Developed, deployed, and maintained web-applications, proactively resolving the issues and enhancing performance.
- Automated manual tasks using **Python, Pandas, and SQL scripts** to reduce manual efforts significantly.
- Involved in the **migration** of legacy applications to Azure cloud platform and addressed critical production issues for Swiss Bank Project(_UBS_).
- Gained expertise in Python, OOP, data structures, SQL, and UNIX concepts.

# Education

### Master of Technology (Data Science and Engineering), BITS, Pilani (WILP)

08/2022 - Present

Grade: Currently pursuing

### Bachelor of Engineering (Computer Engineering), Gujarat Technological University

07/2014 - 06/2018

G.H. Patel College of Engineering, V.V. Nagar

Grade: 8.43 / 10.0, Distinction

## S.S.C (X) & H.S.C (XII Science), Gujarat Board

06/2011 - 06/2014

SSC : Grade: 94.6%, Distinction

HSC (Science) : Grade: 87.33%, Distinction

# Internship Training

## Trainee, Infosys Ltd

07/2018 - 12/2018

- Extensively trained in Python, OOP, Data structures, Advance SQL and UNIX concepts as part of Generic & Stream Training.

# Professional certifications

## Associate Developer for Apache Spark 3.0, Databricks

## Data Engineer Associate, Databricks

## Azure Data & AI Fundamentals, Microsoft Azure

# Projects

## Azure Databricks Pipeline Development

**Tools:** Azure, Databricks, Pyspark, MySQL

- Integrated Medallion architecture, executed Data Quality checks using Great Expectations, and implemented CDC/CDF, SCD type 2, and other Delta Lake features for efficient data processing.

## Databricks DLT Pipeline Framework

**Tools:** Pyspark, Python, MySQL, suppoted Libs

- Involved in building meta-data driven DLT pipeline framwork for loading both batch and streaming data and performed unit testing for data validation.

## GCP to Databricks Migration

**Tools** : GCP Big Query, Databricks, Airflow

- **Migrated BigQuery to Databricks:** Converted queries, modified functions for Databricks compatibility, validated syntax, optimized performance, and implemented Airflow DAGs for query result validation.

## E2 Migration Project

**Tools** : AWS, E2 Databricks, EC2 console, Pyspark, MySQL

- Provided best practices and suggestions to the client to perform the full E2 migration of Databricks artifacts on all the environments.

# Achievements

- High performer of the batch of 1000+ candidates during Infosys foundational trainings.
- Honored as the 'EMPLOYEE OF THE MONTH' for Jul'23 by KOANTEK for the consistent and exceptional performance.
- Currently working under direct project from Databricks.