

Suraj Kumar Soni

Senior Data Engineer

✉ surajkr1000000@gmail.com ☎ [8210891342](tel:8210891342) 📍 Kolkata, West Bengal

in <https://www.linkedin.com/in/suraj-soni-lkdin/>

Data Engineer with 6+ years of total professional experience, including 5+ years building and scaling cloud native data platforms on Azure using PySpark and Databricks. Skilled in designing high-performance ETL/ELT pipelines, optimizing large-scale distributed data processing, and implementing Medallion architecture for reliable data lakes. Proven track record of improving pipeline performance and reliability, reducing compute costs, and enabling data driven decision making in production environments.

Skills

Languages: Python, SQL

Big Data: Apache Spark, PySpark, Hadoop

Cloud: Azure (Databricks, Azure Data Factory, ADLS Gen2)

Data Warehousing: Snowflake, Delta Lake, PostgreSQL, MySQL, MS SQL Server

Orchestration: Apache Airflow

CI/CD & DevOps: Azure DevOps (Git, CI/CD pipelines)

Data Engineering Concepts: ETL/ELT, Medallion Architecture, Dimensional Modeling, SCD (Type 1 & 2)

Data Quality & Governance: Data Profiling, Validation, Metadata Management

Experience

Tredence Analytics

Kolkata, West Bengal

Sr. Data Engineer

28/03/2024 - Present

- Designed and owned **end-to-end data architecture on Azure (ADF, Databricks)** for large-scale CPG datasets, building **15+ production-grade ETL pipelines** to ensure high data availability and reliability.
- Implemented **Medallion Architecture (Bronze, Silver, Gold)** with optimized PySpark transformations, enabling scalable, analytics-ready datasets for downstream business reporting.
- Improved system performance by **optimizing Spark** workloads using **partitioning, Z-ordering, and data skipping**, reducing query latency and overall compute costs.
- Built **reusable PySpark frameworks** and automated legacy system migration via a Python-based DDL conversion tool, reducing manual effort by **60%** and improving engineering efficiency.
- Developed and maintained **dimension and fact pipelines (SCD Type 1 & 2)**, ensuring data consistency, historical accuracy, and alignment with business requirements through STTM.
- Implemented **CI/CD pipelines using Azure DevOps (Git-based workflows)** for automated deployment, testing, and version control of data pipelines.

R1RCM

Noida

Senior Associate- Data Engineer

10/11/2023 - 27/03/2024

- Built and maintained **scalable data pipelines using Python, SQL, and PySpark** for healthcare datasets, supporting downstream analytics and reporting.
- Orchestrated workflows using **Apache Airflow**, improving pipeline reliability, scheduling, and monitoring.
- Optimized **Snowflake data warehouse performance**, improving query execution time and storage efficiency.
- Processed large-scale patient and financial datasets, enabling accurate categorization and analytics for business operations.

Infosys Limited

Bhubaneswar

Data Engineer

17/02/2022- 08/11/2023

- Designed and implemented **ETL pipelines using Azure Data Factory**, enabling efficient ingestion and transformation of enterprise data.
- Developed **dimensional data models** and optimized schema design to improve reporting performance and scalability.
- Leveraged **Spark and SQL** for large-scale data processing and transformation in cloud environments.

- Ensured reliable **end-to-end data flow** across systems with focus on data quality and consistency.

Takhil Technologies

Jamshedpur

Data Engineer

01/02/2021 - 15/02/2022

- Developed ETL workflows using PySpark and SQL to process structured data and load it into data lake/storage systems for reporting and analysis.
- Assisted in monitoring and troubleshooting data pipelines, ensuring data quality and timely data delivery for business use cases.

Pratham Education Foundation

Simdega

Trainer

01/12/2019 - 12/11/2020

- Worked as **Electrical Trainer** and given training about Assistant Electrician (NSQF Level- 3) to underprivileged students.

Education

Dr. B C Roy Engineering College

Durgapur

Electrical Engineering

08/2015 - 05/2019

Certifications

Databricks Certified Data Engineer Associate

28 July 2024

Databricks

Microsoft Certified: Azure Data Fundamentals

3 Aug 2023

Microsoft

Languages

English

Hindi