

Souvik Dey

souvik.dey@tredence.com | +91-9093852518

PROFILE

- Experienced Data Engineer specializing in ETL Pipeline development with all stages of the development cycle. Well-versed in programming languages and frameworks, including Databricks Unified Data Analytics Platform, ADLS, Azure VMs, Azure Active Directory and Service Principal, Azure Container App, Cloud Shell, Cloud Functions, Spark, Python, and Shell Scripting.

PROFESSIONAL EXPERIENCE

SENIOR DATA ENGINEER | KOLKATA, WEST BENGAL, INDIA

June 2019 – PRESENT

- Developed and Optimized over 10+ large-scale data pipelines using PySpark for various client projects, resulting in a nearly 30% improvement in data processing speed.
- Designed and developed a highly efficient data pipeline to ingest, store, and process data from multiple sources, significantly reducing data processing time and improving data accuracy.
- Apply Slowly Changing Dimension (SCD) methodologies and implement Change Data Capture (CDC) pipelines using Pyspark to efficiently manage historical data, track changes over time, and ensure accurate data synchronization.
- Involved in implementing a data quality check framework by leveraging PySpark and Apache Deequ to monitor data integrity, reduce errors, and ensure compliance with security standards and industry regulations.
- Involved in designing Azure POV solutions for client proposals to migrate on-premise workloads and data lakes to Azure using PySpark, ADLS, and Databricks, and implemented a working demo of the solution.
- Collaborated with business stakeholders and cross-functional teams to translate functional requirements into system components, while continuously improving these components to accommodate changes in functionality, ensuring seamless maintenance of data warehouses and data stores for data analysis and reporting.

INTERN | KOLKATA, WEST BENGAL, INDIA

January 2019 – April 2019

- Received comprehensive training in various big data technologies like Distributed Data Processing, Data Modelling, Data Warehousing concepts, etc.
- Developed an end-to-end data analysis project as a Proof of Concept after the training, demonstrating my ability to apply learned concepts.

CERTIFICATION

- Databricks Certified Developer Essential and Lakehouse Platform Essential
- CodeChef SnackDown 2019 Online 1B Round

EDUCATION

FUTURE INSTITUTE OF ENGINEERING AND MANAGEMENT | B.TECH IN COMPUTER SCIENCE AND ENGINEERING

June 2019 | Kolkata, West Bengal • GPA: 7.83/10.0

CANNING DAVID SASSION HIGH SCHOOL | 10TH AND 12TH

- 10th (WBSE), May 2013: 80/100 • 12th (WBCHSE), May 2015: 75/100

SKILLS

Big Data Tools and Technologies:

VSCode • PySpark • Hadoop • HDFS • WinScp • Putty

Azure Cloud Services

Databricks • ADLS • Virtual Machines • Key Vault • Container Apps

Software Development Methodologies:

Agile

Documentation:

Confluence

Databases:

Hive • SQL Server • Oracle

Programming Languages:

Python • SQL • Shell Scripting

Version Control System:

Github • Rally

Schedulers:

CA Workstation • Autosys