

Souvik Dey

souvikdey9510@gmail.com | +91-9093852518 | Github | LinkedIn

PROFILE

- Senior Data Engineer specializing in ETL Pipeline development with all stages of the development cycle. Well-versed in programming languages and frameworks including Google Cloud Platform, Dataproc including cluster initialization action, IAM, GCS, BigQuery, Cloud SQL, Compute Instance, Cloud Shell, Cloud Functions, Hadoop, Hive, Spark, Python, Shell Scripting.

PROFESSIONAL EXPERIENCE

ASSOCIATE DATA ENGINEER | COGNIZANT TECHNOLOGY SOLUTIONS, KOLKATA, WEST BENGAL, INDIA

June 2019 – PRESENT

- Developed and optimized 10+ large-scale data pipelines using PySpark, resulting in an almost 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.
- Involved in implementing data quality check framework through PySpark to monitor data integrity, which decreased data errors, and ensured data compliance with industry regulations.
- Involved in designing Google Cloud POV solutions for client proposals to migrate on-premise workloads, Datalake to Google Cloud leveraging components like PySpark, GCS, Dataproc, and BigQuery. Implemented a working demo of the proposed solution.
- Collaborated with business customers and technical peers to translate different functionalities and business implementation of the system components.
- Engaged in enhancements of the components as per the change in functionalities to maintain data warehouses and data marts to support data analysis and reporting.

INTERN | COGNIZANT TECHNOLOGY SOLUTIONS, KOLKATA, WEST BENGAL, INDIA

January 2019 – April 2019

- Received comprehensive training in various big data technologies like Apache Hadoop, Hive, MapReduce, 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

- Google Cloud Platform (GCP) Certified Professional Architect
- 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

Google Cloud Data Services

Databricks • GCS • DataProc • Big Query • Compute Instance • Cloud SQL

Software Development Methodologies:

Agile

Documentation:

Confluence

Databases:

Hive • SQL Server • Oracle

Programming Languages:

Python • SQL • Shell Scripting

Version Control System:

Gitlab • Rally

Schedulers:

CA Workstation • Autosys