Aditya Narayan Daw

14/1 Kali Kumar Mukherjee Lane, Howrah, West Bengal, India
(+91) 7003879581
aditya.n.daw@gmail.com | GitHub | LinkedIn

Data enthusiast with a passion for understanding and leveraging the vast and ever-evolving field of big data and data engineering.Seeking to contribute to this exciting field as a valuable member of a team.

PROFESSIONAL EXPERIENCE

TATA CONSULTANCY SERVICES, **Kolkata, West Bengal**
Big Data Developer *October 2021–Present*

- Developed fault-tolerant frameworks to ingest huge volumes of data from various sources into the Hadoop environment which was 30% more efficient than the previous approach.
- Performed validation checks on ingested data, such as removing extra new lines and rejecting data based on metadata, before storing it in Hive tables for further processing which reduced the manual filtering of the data, which reduced the downstream processing time by 70%.
- Built a micro-batch streaming pipeline using Kafka and Spark to ingest data in Hive after flattening nested data into a structured format which helped downstream users to perform analytics on the nested data efficiently.
- Helped to migrate a large banking client from traditional Apache Pig based application to Spark based application successfully which not only resulted in reducing the processing time of the data by 20% also helped the client to be more flexible with the modern changes in technology.
- Worked also in an environment as a sole developer of a business team, who communicates with the business users and other teams to work on new POCs and to validate and perform requirements.

TECHNICAL SKILLS

SQL, NoSQL, Spark, Pyspark, Hadoop, Kafka, Talend, Scala, Python, Amazon EMR

EDUCATION

University Institute of Technology *2017-2021*
Instrumentation Engineering,
Honors: cum 8.48 CGPA

Howrah Zilla School *2015-2017*
Higher Secondary Education, Achieved 80 percent marks at final examination

PROJECTS

- **Spark Share-Point Connector**
  Built a Spark V2 connector to ingest data from SharePoint 2019 lists into Spark for further analytics on the data. The connector currently supports text, numbers, and dates. It is still under development.
- **Flattening Logic In Spark**
  Created a flattening logic in Spark that can dynamically flatten semi-structured data in structured format for further processing. The flattening logic currently supports JSON data format. It can handle nested data and arrays.