Aditya Narayan Daw

Big Data Engineer

Howrah, West Bengal
7003879581
aditya.n.daw@gmail.com

SKILLS

SQL, Spark, Pyspark, Big Data, NoSQL, Hadoop, Kafka, Scala, Python, Amazon Q, LangChain, Unix

EXPERIENCE

TATA CONSULTANCY SERVICES, Kolkata - *Big Data Engineer*
OCT 2021 - PRESENT ***Part of CTI CATE EAP Team in CITI***

- 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.
- 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.

EDUCATION

University Institute of Technology, Burdwan, West Bengal - *Instrumentation Engineering*

JULY 2017 - AUG 2021

Completed graduated successfully with 8.48 CGPA.

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.