Yatish Sharma

+91-9343711759 | yattishsharma@gmail.com
LinkedIn | **GitHub**

PROFILE

Results-driven Data Engineer with 4 years of experience in building and optimizing large-scale data pipelines, ETL processes, and data warehouse solutions. Utilized technologies like **Python, SQL, Spark and AWS** to develop multi-terabyte scalable big data solutions.

TECHNICAL SKILLS

Programming Languages: Python, SQL
Big Data Technologies: Spark, PySpark, Spark SQL, YARN, Hadoop, Hive, Impala
Cloud computing: Amazon S3, AWS Glue, AWS EMR, AWS Lambda
Data Engineering Tools: Data Modelling, ETL/ELT data Pipeline
Orchestration: Airflow, Autosys
Familiar: MongoDB, Cassandra, MySQL, Unix shell scripting, GIT

EXPERIENCE

Barclays October 2022 - Present
Data Engineer *Pune, India*

- Led a project to migrate the whole architecture from CDH (Cloud Distribution of Hadoop) to BDH i.e. BDH (Barclays Distribution of Hadoop). I have contributed straight from infrastructure set up, models & data migration and ensure data integrity with proper data validation between both the environments. It was a critical migration. It saves **$35 million** to the bank.
- Collaborated with data science and business intelligence teams to design and develop the surveillance models like BABS (Behavior Analytics Based Surveillance) models.
- Worked on developing the pipeline to loads the data from hive tables to **MongoDB** collections using PyMongo driver.
- Developed an automated framework to validate and perform DQ checks on the source data before processing it to enhance the model features by **40%**.
- Implemented Spark optimization techniques such as **caching, multithreading, and broadcast** joins, resulting in a 30% decrease in processing time for handling a daily load of around **2 million records**.
- Conducted in-depth data analysis using **Hive, Impala, and Spark SQL**, providing UAT fixes and ensuring smooth operations in the production environment.

Cognizant Technology Solutions March 2021 – July 2022
Programmer Analyst *Chennai, India*

- Worked on doing EDA (Exploratory Data Analysis) using Python libraries like Numpy, Pandas, Matplotlib and Seaborn to get insights out of the data. Develop easy to understand reports using Microsoft Power BI tool.
- Designed and implemented scheduling capabilities using **Autosys** for data pipeline orchestration, reducing manual intervention time by 70% and streamlining workflow efficiency.

EDUCATION

Lovely Professional University Jalandhar, Punjab
Bachelor of Technology in Computer Science & Engineering *July 2017 - June 2021*

- *Final Grade: 8.12 CGPA*

Shri Krishna Memorial Higher Secondary School Guna, Madhya Pradesh
Senior Secondary School *July 2016 - Mar 2017*

- *Results: 83.4 %*

Languages

- *English*
- *Hindi*