


Vipin Sahu

✉ vipinsahu027@gmail.com | **in** linkedin.com/in/vipinsahu027 |  github.com/vipinsahu027 | ☎ +91-9425335649

PROFILE SUMMARY

Data Engineer with over 3 years of experience designing and optimizing large-scale data pipelines, ETL processes, and data warehousing solutions. Proficient in **Python, Spark/PySpark, SQL, and Azure, Databricks**, delivering scalable big data solutions for the BFSI sector, improving performance by **15%**.

TECHNICAL SKILLS

Programming Languages: Python, C/C++, SQL

Big Data Technologies: Apache Spark, Spark Streaming, PySpark, Spark SQL, Hadoop

Data Engineering Tools: ETL/ELT Pipelines, Data Modeling, Azure Data Factory, Databricks

Cloud Platforms: Azure

Version Control: GIT

EXPERIENCE

LTIMindtree

March. 2023 – Present

Data Engineer

- Designed and optimized ETL pipelines for **BFSI data**, integrating sources like **Azure**, and **file formats (CSV, Parquet)**, improving data processing efficiency.
- Applied **Spark optimization** techniques, including **Adaptive Query Execution (AQE)**, **Caching**, **Salting**, and **Broadcast Joins**, **reducing processing time by 15%** and compute costs.
- Collaborated with business teams to align pipeline requirements, ensuring accurate reporting for financial analytics.

Mindtree

December 2021 – February 2023

ETL Developer

- Developed **Python** and **SQL** scripts to reconcile source and target data for **BFSI** projects, **increasing processing speed by 10%**.
- Engineered **SQL Table-Valued Functions** to replicate source table structures, streamlining data validation and **Optimized ETL workflows**, **reducing data latency** for reporting by automating data movement to unified target tables.

PROJECTS

Sales DataMart | *Python, PySpark, SQL, AWS S3, Docker*

- Automated daily and monthly sales report generation**, processing over 10 million records for salesperson incentives.
- Implemented **Spark optimization** technique such as **AQE**, **Caching**, **salting**, **Broadcast** join.
- Writing data back to S3 after transformation in Parquet format partitioned on required column.

EDUCATION

ITM Group of Institutions, Gwalior

July 2017 - July 2021

B.Tech. in Computer Science and Engineering

CGPA: 8.1/10

AWARDS

Shooting Star Award: Recognized for improving Apache Spark data pipeline performance by 15%, reducing query latency and compute costs.