

Vipin Sahu

✉ vipinsahu027@gmail.com | [in linkedin.com/in/vipinsahu027](https://www.linkedin.com/in/vipinsahu027) | github.com/vipinsahu027 | ☎ +91-9425335649

PROFILE SUMMARY

Azure-focused Data Engineer with 3+ years of experience building and optimizing scalable data pipelines and ETL workflows using **Python, Spark/PySpark, SQL**, and cloud platforms like **Azure** and **Databricks**, delivering a **15% boost in processing efficiency** for BFSI sector solutions.

TECHNICAL SKILLS

Programming Languages: Python, PySpark, SQL, Bash, C/C++
Big Data: Hadoop, PySpark / Spark (including Streaming), Spark SQL, Data Lake
Data Engineering: ETL / ELT Pipelines, Data Modeling, Data Warehouse
Cloud: Azure - Data Factory, Synapse Analytics, Databricks, Delta Lake
Orchestration : Airflow, Docker, Git
Version Control: Git, Azure DevOps

EXPERIENCE

LTIMindtree March 2023 – Present
Data Engineer

- Designed and optimized **ETL pipelines** processing **BFSI** data using **Azure services** and common file formats (**CSV, Parquet**), improving data processing efficiency by **15%**.
- Implemented **Spark optimizations** such as **Adaptive Query Execution (AQE)**, **caching**, **salting**, and **broadcast joins**, reducing processing time and compute costs.
- Collaborated cross-functionally to align pipelines with **business analytics** requirements ensuring accurate reporting.

Mindtree December 2021 – February 2023
ETL Developer

- Developed **Python** and **SQL** scripts for **source-target data reconciliation** in **BFSI** projects, increasing processing speed by **10%**.
- Engineered **SQL Table-Valued Functions** to replicate source table structures and automated **ETL workflows**, reducing data latency and improving reporting cycles.

PROJECTS

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

- Automated daily and monthly sales report generation**, processing over 1 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.