

Manish Kumar

+91 ***** | mverma6250@gmail.com

LinkedIn | GitHub

PROFILE

Data Engineer with 5+ years of experience in building large-scale data pipelines, ETL processes, and data warehouse solutions. Utilized technologies like **Python, SQL, Spark, Kubernetes, Databricks, and Kafka** to develop multi-terabyte scalable big data solutions for Fortune 100 Pharmaceutical and Telecom companies.

TECHNICAL SKILLS

Programming Languages: Python, SQL, Scala
Big Data Technologies: Spark, PySpark, Spark SQL, YARN, Kubernetes, Hadoop, Hive, Impala
Streaming Technology: Kafka, Spark structured streaming
Cloud computing: ADF, Databricks, ADLS, Amazon S3, AWS Glue, AWS EMR
Data Engineering Tools: Azure DevOps, Data Modelling, ETL/ELT data Pipeline
Orchestration: Airflow, Kubernetes, CA Autosys
Familiar: FastAPI, Elastic search, Neo4J

EXPERIENCE

Reliance Jio August 2022 - Present
Senior Data Engineer Gurgaon, India

- Led a team of 4 Data Engineers in designing, developing, and optimizing data pipelines and ETL processes.
- Developed an automated invoicing system for dynamic cloud pricing, ensuring accurate monthly billing adjustments and improving invoicing accuracy by 100%.
- Integrated and processed billing data from multiple sources including **AWS, Azure, GCP, and Oracle**, handling various file formats like JSON, CSV, and Parquet.
- Implemented Spark optimization techniques such as **caching, multithreading, and broadcast** joins, resulting in a 20% decrease in processing time for handling a daily load of around **2 Million records**.

Reliance Jio August 2022 - Present
Data Engineer Gurgaon, India

- Created an API service using Python to generate dynamic DAGs in **Apache Airflow**.
- Designed and implemented advanced scheduling capabilities using Airflow for data pipeline orchestration, reducing manual intervention time by 80% and streamlining workflow efficiency.
- Developed a microservice deployed on Kubernetes, integrating FastAPI to push messages into a **Kafka** broker for real-time tracking of processing status (success or failure).

ZS Associates April 2020 - July 2022
Data Engineer Pune, India

- Delivered a project to migrate legacy on-premise processes to the cloud using Big Data technologies (Spark), reducing processing time by 20%.
- Conducted in-depth data analysis using **Hive, Impala, and Spark SQL**, providing SIT/UAT fixes and ensuring smooth operations in the production environment.
- Optimized overall process performance through **Spark performance tuning**, improving job run times by 20% and efficiently managing a **15Terabyte(TB)** dataset containing approximately **12 billion** records.
- Worked on data ingestion pipeline to ingest the flat file in the **Data lake**.

Standav July 2019 - March 2020
Data Analyst Bangalore, India

- Analysed more than 30,000 customer sales data to find the bottlenecks in the business process.
- Worked as Salesforce Admin and got exposure to the sales cloud.

EDUCATION

NIT Trichy Trichy, Tamilnadu
Bachelor of Technology in Materials Engineering July 2015 - June 2019

SDS College Kaler, Bihar
Senior Secondary School April 2012 - Mar 2014