

Manish Kumar

mverma6250@gmail.com | 979*****

SKILLS

PROGRAMMING

Python • MySQL

Basic:

Scala

BIG DATA

Spark • PySpark • Hive

Impala • HDFS • YARN

Spark SQL

ORCHESTRATION

Airflow • Kubernetes • CA Autosys

FAMILIAR

Elastic search • Neo4j

Azure Devops • Fast API

Azure Blob • AWS S3

EDUCATION

NIT TRICHY

MATERIALS ENGINEERING

June 2019 | Trichy, TamilNadu

Cum. GPA:8.16/10

SDS COLLEGE

SENIOR SECONDARY SCHOOL

March 2014 | Kaler, Bihar

LINKS

Github:// [manisnitt](#)

LinkedIn:// [manish-kumar](#)

YouTube:// [mverma6250](#)

COURSEWORK

Python Programming

Spark

Hadoop

MySQL

Cloud computing

Datawarehouse Modelling

Data Structure and Algorithm

EXPERIENCE

RELIANCE JIO | DATA ENGINEER

Aug 2022 - Present | Gurgaon, India

Mentoring a team and driving project development through agile CI/CD pipelines.

AUTOMATED DAG SCHEDULAR

- Created an API service to generate dynamic DAGs in **Apache Airflow**.
- Implemented scheduling capabilities, allowing for the efficient orchestration of data pipelines, optimizing data processing workflows.

BILLING LICENCE

- Developed an automated invoicing system, enabling dynamic pricing adjustments on cloud service provider invoices.
- Integrated and processed billing data from multiple sources including AWS, Azure, GCP, and Oracle, handling **various file formats**.
- 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 **800,000 records**.

ZS ASSOCIATES | DATA ENGINEER

Apr 2020 - Present | Pune, India

Led a DevOps workstream independently and managed a team of 3 members.

MHCDM

- Delivered a project to migrate legacy on-premise processes to the cloud using Big Data technologies(**Spark**)
- Actively involved in detailed data analysis using **HIVE, Impala, Spark SQL, SIT** and UAT fixes in the production environment.
- Worked on **Spark performance tuning** for overall performance optimization of the process.

INGESTION

- Worked on data ingestion pipeline to ingest the flat-file in the **Data lake**.
- Automated the flat-file loading from **AWS S3** to hive table with the help of an Orchestration tool like CA Autosys.

STANDAV | ASSOCIATE CONSULTANT

Jan 2019 - Mar 2020 | Bengaluru, India

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