

## Contact

6232997604 (Mobile)  
shrivastavaeshan110@gmail.com

www.linkedin.com/in/eshan17  
(LinkedIn)  
github.com/Eshaan1799 (Other)

## Top Skills

Data Lakes  
Apache Airflow  
Hadoop

# Eshaan Shrivastava

Consultant(AI and Data)@Deloitte. Skilled in designing, implementing, and maintaining scalable data pipelines. Hadoop| Spark| Pyspark | Python| Scala |SQL|AWS|Databricks  
Bhopal, Madhya Pradesh, India

## Summary

Data Engineer with 4+ years of experience in crafting and fine-tuning large scale data pipelines to provide robust and reliable data solutions. Committed to collaborative teamwork for delivering synergized results.

Key Skills - Python, Scala, Apache Spark, Pyspark, AWS Glue ,AWS Redshift, AWS Cloudwatch,AWSS3,Hadoop,Git,AWS Lambda ,SQL, NOSQL, Airflow,Databricks

---

## Experience

### Deloitte

#### Senior Consultant

August 2024 - Present (10 months)

Bengaluru, Karnataka, India

- Optmized an existing data pipeline , trimming down execution time from 20 mins to under a minute via elimination of loops and parallelizing code segments via Pyspark.

- Created data pipelines in Azure Databricks reading data from client end ,performing a series of transformation and storing data in database

### Tata Consultancy Services

#### System Engineer

January 2021 - April 2024 (3 years 4 months)

Pune, Maharashtra, India

- Implemented robust, scalable data pipelines sourcing data from diverse channels including Kinesis DataStream and processing data files ranging from 5mb to 10 Gb consolidating a daily data processing volume of about 50 Gb.

- Automated the execution of data pipelines using S3 event notifications and AWS Lambda to trigger only when the file is available. This reduced the overall operational time by 30 %.
- Accelerated the rate of data retrieval from database by 25% using query optimization and creation of indexes.
- Analysed AWS cloud watch metrics to optimally schedule resource heavy spark-jobs decreasing their monthly failure count by 100.
- Used AWS Glue to develop ETL pipeline to transfer data from S3 Bucket to data storage systems.
- Implemented and managed a suite of Spark scripts to detect and resolve occasional data issues, ensuring seamless data flow, minimizing workflow interruptions, and saving the team an average of 5 hours per week.
- Migrated SQS queues from standard to FIFO. It eliminated false triggering of spark jobs and their subsequent failure, cutting down the monthly incident count by 65.
- Employed various performance tuning methods including Adaptive Query Execution(AQE) in Spark, yielding a 15% monthly reduction in resource consumption.
- Used AWS S3 as Data Lake for data pipelines, facilitating the seamless monthly processing of 1.5 Terabytes of data.
- Coded Spark jobs with Pyspark, seamlessly integrating DataFrame and RDD API to unlock the full potential of distributed and parallel processing of big data in Apache Spark.
- Utilized Airflow's automation and scheduling capabilities to ensure timely execution of spark jobs, achieving a 95% on-time completion rate.
- Employed AWS Redshift as the primary data warehouse ensuring high availability of data to facilitate efficient data analytics for business use cases.

Lakshmi Narain College of Technology, Kalchuri Nagar, Raisen  
Road, Post Klua, Bhopal-462021

Bachelor of Engineering - BE, Information Technology · (June 2016 - June  
2020)