



## PROFILE SUMMARY

Experienced Data Engineer with expertise in building and optimizing data pipelines, Azure Data Factory, and Azure Synapse Analytics to deliver business-critical solutions. Proficient in designing ETL workflows, implementing dynamic components, and preparing fact and dimension tables for analytics. Additionally, has a background in quality assurance, process improvement, and cross-functional collaboration in manufacturing environments. Known for problem-solving and delivering results in diverse industries.

# RAHUL SAJWAN

## DATA ENGINEER

Dehradun, INDIA

7 Years 11 Months of  
experience



(+91) 8410529357



rahul.sajwan.10@gmail.com

## WORK EXPERIENCE

### Data Engineer

Globaltek Prosolutions

Feb 2023 - Present

- Participating in the scrum calls to get the requirements and provide the updates for assigned tasks.
- Prepared the Azure data factory pipelines based on business requirements using different activities like copy activity, stored procedure activity, look up activity, if condition, for each and so on.
- Implemented the data flow logic to do transformation in azure data factory.
- Implemented the dynamic data set components to work for different objects.
- Trigger the pipelines based on schedule intervals.
- Prepared the external tables, dimension and fact tables in azure synapse based on user requirements.
- Prepared the business logic using procedures in azure synapse analytics based on user requirements.
- Involved in bug fixing.

### QA Engineer

Shiv Shakti Prints and Fabrics  
Pvt Ltd

Mar 2017 - Jan 2023

Ensured quality standards compliances in the supplier's workplace by regular auditing. Verification of the manufacturing and quality plans for new product platforms from suppliers by working with cross- functional teams to understand suppliers' process. Handle in- process Quality Issue, Check sheet development Handling control of records and documents, Analyse Proto Inspection DATA, Improvement Project etc.

## KEY SKILLS

- Python
- SQL
- Spark
- Hive
- Azure Data Factory
- Data Bricks
- Azure Synapse
- Azure Data Lake
- Spark Streaming
- Kafka Streams
- Kafka

## CERTIFICATION

- secured an All India Rank (AIR) of 6078 in GATE 2015
- Cloud Data Engineer

## LANGUAGES

English  
Hindi

## HOBBIES

Taking long walk

## SOCIAL LINKS

[www.linkedin.com/in/rahul-sajwan-898b7186](https://www.linkedin.com/in/rahul-sajwan-898b7186)

## EDUCATION

2013

**B.Tech/B.E. - electronics and communication engineering**

Uttaranchal Institute of Technology,  
Dehradun

## PROJECTS

### 5G Network Performance Monitoring and Customer Experience Analytics

15 Months

This project focuses on building a scalable ETL pipeline and analytics solution for monitoring Verizon's 5G network performance and improving customer experience. The solution collects real-time data from network monitoring tools, customer feedback systems, and usage logs. It processes and transforms the data to provide actionable insights into network quality, customer satisfaction, and potential service disruptions.

### Mule & Shells (FMAML)

3 Months

The solution was to build multiple Rules and AI model Financial Monitoring and Anti Money laundering based on various risk scenarios. The strategic solution is expected to trigger alerts based on a model-based system on all high risk rated FM trade population feed these into ECM (Enterprise Case Management System) for further Analysis on the transactions.

Responsibilities

- As a Data Engineer, worked on the End-to-End development and testing of data pipelines and deployed it into the production environment.
- Building features for model output using SQL and pyspark
- Implementation of new transformations and development of the pyspark wrapper, used some optimisation techniques of spark.
- Build data pipeline using development of the Control- M. Tested each pipeline into QA environment and deployed into the production.
- Contribution in production environment setup. Activities like AD group creation, security matrix implementation

### Loan Risk Assessment Model for Lending Club

3 Months

Identifying loan default probabilities. Optimizing lending decisions by analysing borrower profiles. Enhancing profitability by mitigating risks.

- Extracted raw data from multiple sources (e.g., databases, CSV files) into a distributed data storage environment like HDFS.
- Performed data cleaning to handle missing, duplicate, or inconsistent data.
- Applied PySpark transformations, such as `withColumn` , `regexp_replace` , and `sha2` , to normalize and enrich the data.
- Built end-to-end ETL pipelines for loading, transforming, and saving cleaned datasets into raw and curated storage layers.