



8237941314



surajhajare5795@gmail.com



Pune, Rajhans Apartment, Flat No:
310, Shivajinagar Pune

EDUCATION

Bachelor of engineering

Savitribai Phule Pune University/ PES
Modern College Of Engineering Pune,
India, Pune

2013 - 2017

SKILLS

- SQL
- Python
- Data Mining
- ETL
- NoSQL
- Cloud Computing
- Data Visualization
- Data Analysis
- Data Modeling
- Apache Spark
- Data Warehousing
- AWS
- Data Migration
- Data Validation
- Data Governance
- AWS Glue
- Redshift
- Snowflake
- AWS Lambda
- ETL
- Azure Data Factory
- s3 | RDS | Redshift |
- Pyspark
- Data bricks
- Pandas
- Big Data
- Business Intelligence

HOBBIES

- Reading
- Drawing
- Swimming

SURAJ DHANAJI HAJARE

DATA ENGINEER AT TECHBORN IT SOLUTIONS

CLIENT NAME: CAPGEMINI

PROFESSIONAL SUMMARY

Data Engineer with 4.3+ years of experience in Data Engineering, Big Data technologies and Data Analysis, Proficient in Python, PySpark, Databricks SQL, Azure Cloud with expertise in building and managing scalable data pipelines using Azure Data Factory. Skilled in Data Mining, Data Preparation, Data modeling and ETL processes, ensuring high-quality data flow for analytical and business needs. Experienced in handling large datasets, implementing Machine Learning algorithms, and working with cloud platforms for data storage and processing. Strong knowledge of Proof of Concepts (PoC) and gap analysis to drive data-driven solutions for enterprise applications.

EXPERIENCE

- January 2023 - Now

Data Engineer

Unum [Insurance Domain] LIC Insurance Tax Benefits(Healthcare) / India, Mumbai

• **Technologies: Python, SQL, ETL, Databricks, PySpark, AWS Cloud Services, Git**

The **Insurance Data Lake** project centralizes and processes insurance data from various sources like Oracle, MySQL, AWS S3, and files Using **AWS Glue, Databricks and Redshift**, it ensures data quality, transformation, and enrichment for analytics. The pipeline ingests, standardizes, and structures data into **Raw, Silver, and Golden Zones** for optimized querying and reporting. Business insights include **fraud detection, claim processing efficiency, and customer engagement analytics**. The solution is **scalable, secure, and compliant** with regulations like **GDPR, HIPAA, and CCPA**, ensuring accurate, real-time decision-making.

- September 2021 - December 2022

Data Engineer

/ AISE Datalake

Technologies Skills : AWS S3, AWS Glue, Amazon Athena, AWS Lambda, RDS/Redshift, PowerBI, PySpark, SharePoint API, Python, SQL, Spark, ETL, AWS Services like S3, lambda, glue, Redshift, IAM

Insurance Data Lake – High-Level Design (HLD)

The **Insurance Data Lake** project streamlines **policyholder data, claims processing, fraud detection, and customer insights** by integrating data from multiple sources like **S3, MySQL, Oracle and APIs**. The ETL pipeline follows a **multi-layered approach (Raw → Silver → Gold)** **AWS Glue, Redshift, and Databricks** to ensure high-quality, structured data for analytics.

Technical Skills & Stack

AWARDS

- Employee of the Month April 2024

LANGUAGES

- English
- Hindi
- Marathi

- **Data Engineering:** Python, SQL, ETL, AWS Glue, Apache Spark
- **Cloud G Storage:** Amazon S3, Redshift, Athena
- **Data Processing:** Databricks, Delta Lake, Kafka
- **Orchestration G Monitoring:** AWS Step Functions, AWS Glue Workflow, AWS CloudWatch
- **BI & Reporting:** Power BI, Tableau, AWS QuickSight
- **Security & Compliance:** IAM roles, Encryption, GDPR, HIPAA

1. This scalable architecture **enhances fraud detection, claim efficiency, and business intelligence**, enabling real-time decision-making for insurers.

-
- December 2020 - August 2021

Data Engineer

HDFC /

Project Name: Integrated Loan Management/CIBIL ML Pipeline

Client: HDFC, India

System/Component: AWS Services

Technology Stack:

- **Programming:** Python
- **Machine Learning:** Loan predictions, Aadhar masking
- **Cloud Platforms:** AWS (Lambda, Glue, S3, Step Functions, CloudWatch, ECS), Azure (Databricks, PySpark, Azure Data Factory)

Role: Data Engineer

Description: HDFC, a leading banking firm, aims to automate loan and CIBIL check management processes using AWS and Azure services. The project involves:

1. Automating call routing and loan management tasks.
2. Using machine learning algorithms for loan prediction (auto and personal loans) based on historical data.

Responsibilities:

1. Pipeline Development:

- Built an automated pipeline leveraging AWS Lambda for serverless execution and Step Functions for workflow orchestration.
- Integrated Glue for ETL processes and used S3 for data storage and retrieval.
- Monitored the pipeline using CloudWatch.

2. Data Integration and Transformation:

- Managed data ingestion pipelines using Databricks and PySpark for processing large datasets.
- Integrated Azure Data Factory for workflow management across Azure services.

3. Database Management:

- Loaded and managed data in RDS for relational queries.
- Integrated Redshift for analytics and large-scale data storage.

4. Monitoring and Optimization:

- Ensured the efficiency of data pipelines.
- Troubleshoot and optimized workflows using ECS and monitoring services.