Rushikesh Kulkarni

3.1 Years as Data Engineer at Yash Technologies

Contact info

9545032501

rushikeshkulkarni2501@gmail.com

India, Pune, B-504 Riddhi siddhi Towers,Charlholi Phata,Pune

Education

- **Savitribai Phule Pune University** 2019 - 2022
  India, Pune

Skills

- SQL
- Python
- Big Data
- Data Warehousing
- Data Mining
- ETL
- Apache Hadoop
- Apache Spark
- NoSQL
- Data Modeling
- Cloud Computing
- Data Visualization
- Machine Learning
- Data Analysis
- Business Intelligence
- Pyspark
- Data Bricks
- Azure Data Factory
- Azure Data Lake

Professional summary

3.1 Years as Data Engineer at Yash Technologies From Oct 2022 to tilll date

Data Engineer with 3.1+ years of experience in Data Engineering, Big Data technologies and Data Analysis, . Proficient in Python, PySpark,AWS,AWS Glue,Lambda,Stepfunction,s3,blob storage,AWS SNS,SQS,SES,Redshift,Snowflake,,Data Validation,Data Migration,Data Modeling,Databricks SQL,Hadoop,Data Mideling, Hive , 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

- Data Engineer January 2024 - Now
  UHC, United States

  Project Name:

  AWS Insurance Data Pipeline – Broker Analytics Platform

  Technical Skills:

  AWS S3, AWS Lambda, API Gateway, AWS Glue, Glue Crawler, Databricks, PySpark, Delta Lake, Amazon Redshift, SQL, ETL Pipelines, Data Quality, Schema Evolution, Data Modeling, KPI Reporting

  Project Overview:

  This project involved building a scalable AWS-based data pipeline for an insurance organization to

Snowflake

Data Migration

Data Validation

Data Modeling

Data Cleaning

Hobbies

Reading

Bike riding

Traveling

Languages

English Hindi Marathi

Personal info

Date of birth:
25 June 2001

Place of birth:
Latur

Nationality:
Indian

automate and standardize broker data processing. The pipeline ingested CSV broker data from internal APIs into Amazon S3, automated ingestion through AWS Lambda and API Gateway, and used AWS Glue Crawlers to catalog schema and maintain metadata.

Databricks ETL pipelines transformed data across Bronze, Silver, and Gold layers, ensuring standardization, validation, deduplication, and enrichment. The Gold layer produced analytical data marts that supported broker onboarding analysis, lifecycle monitoring, regional trends, and KPI reporting. Curated Gold datasets were then loaded into Amazon Redshift using AWS Glue jobs for analytics and dashboard consumption.

The system improved reporting reliability, reduced manual intervention, and delivered consistent, high-quality broker data for business decision-making.

Outcomes:

- Improved data quality and consistency by automating standardization in Silver and Gold layers
- Reduced reporting delays through automated ingestion and validation workflows
- Enabled real-time broker lifecycle insights using curated Gold-level data marts
- Increased accuracy of KPIs such as active brokers and regional performance metrics
- Enhanced business decision-making through Redshift-based dashboards and analytics
- Strengthened data governance by using Glue Crawlers and schema evolution support

Challenges:

- Inconsistent and incomplete broker data received from multiple internal systems
- Frequent schema drift across incoming CSV files, requiring dynamic schema handling
- Lack of standard formatting across email, phone numbers, and address fields
- Processing delays due to manual validations befor

Data Engineer October 2022 - December 2023

Cigna Group, United States

Project Title: Azure Healthcare Data Platform

Tech Stack:

Azure Data Lake Storage (ADLS), Azure Functions, Azure Databricks, Azure Data Factory, PySpark, Azure Monitor, Python SDK, Azure CLI

1) Project Architecture

- Raw Customer, HCP (Health Care Professional), and interaction data landed in **Azure Data Lake Storage (ADLS)** in CSV/JSON format.
- CRM & in-house applications pushed files using API-based scheduled ETL jobs and middleware tools.
- **Azure Functions** triggered automatically on new file creation events, performing metadata extraction and file-level checks.
- **Azure Databricks (PySpark)** processed large-scale customer & HCP interaction datasets.
- **Spark SQL** used to aggregate interaction duration by **HCP, specialty, and territory**.
- Business validation framework ensured data completeness, reference checks, and format standardization.
- **Azure Data Factory Pipelines** orchestrated the complete workflow from ingestion processing validation publishing.
- Final curated, enriched datasets were sent back to CRM systems to power campaign intelligence and personalized outreach.

2) Project Goal

To build a **scalable, secure, and centralized healthcare data platform** that integrates customer, HCP interaction, and sales activity data from multiple disjoint systems, standardizes it using Databricks-based Spark pipelines, enforces data quality, and publishes enrichment-ready datasets to CRM systems.

This enabled:

- Personalized engagement campaigns
- Faster decision-making
- Improved HCP experience
- Stronger analytics and medical insights

3) Problem Statement

Healthcare organizations had fragmented data spread across multiple systems with **inconsistent formats, missing values, and no unified structure**.

This created delays in:

- Understanding HCP engagement trends
- Evaluating campaign performance

- Running analytics due to poor-quality and delayed data

4) Role & Responsibilities

- Developed **data validation & rule-check scripts** to enforce business logic.
- Implemented **deduplication, referential integrity**, and format standardization.
- Bui