Name: Shubham Awasare

Happiest Minds Email id: shubham.awasare@happiestminds.com

Primary Skills: SQL, Python, Spark, PySpark, ETL, EDA (exploratory data analysis), Data Warehouse, Azure cloud, AWS, Databricks, Data Visualization (Power BI, Tableau), Database Management, Cloud Data Integration with Azure, PostgreSQL, Data Lake, Big Data, Data Pipeline, Azure Data Factory, Delta Lake

Experience Summary:

> Experienced Data Engineer with 3.5+ years of experience, specializing in Data Modeling, Data Warehousing, SQL, building large-scale systems, Azure and AWS ETL processes, improving data transfer workflows, and managing complex data tasks. Having hands-on expertise with Azure data factory and Databricks for seamless data integration, processing and reporting. Skilled in transforming raw data into actionable insights through complex data extraction, transformation, and visualization techniques. Skilled at automating workflows and optimizing data pipelines performance. Adept at aligning data strategies with business goals through cross-functional collaboration, with a focus on industry best practices to drive innovation and efficiency.

Project 1 – Ministry of Commerce and Industry (September 2023 to January 2025)

1. **Description of the Project:**
   - Working with different file formats to perform operations by using pyspark and write data into Data-Lake.
   - Hands on experience in various Big Data application phases like data ingestion, data analytics.
   - Experienced in migrating on-premises databases to Azure Data Lake, Azure Blob Storage, and Azure SQL Database.
   - Mainly use AWS and Azure cloud resources and Databricks components to clean and store Data.
   - Executed transformations and data cleansing activities for improved data quality.
   - Work with internal teams to design, develop and maintain software.
   - I have extensive experience in creating Data pipeline jobs, schedule triggers using Azure data factory.
   - Experienced with the tools in Hadoop Ecosystem including Hive, HDFS, MapReduce, Sqoop, Spark.

2. **Role:** Data Engineer

Project 2 – TCS HOBS - Converge ICT Philippines (March 2022 to June 2023)

Project 3 – Vodafone Idea India

1. **Description of the Project:**
   - Enhanced ETL processes with Azure Data Factory for better efficiency.
   - Developed and implemented pipeline in ADF using link services data set pipeline to Extract Transform and Load data from different data sources to Azure cloud.
   - Executed transformations and data cleansing activities for improved data quality.
   - Demonstrated expertise in gathering and analyzing complex data sets, uncovering actionable insights and opportunities for process optimization.
   - Developed dashboards for executive-level reporting using Power Bi and SQL queries in MYSQL/Oracle

2. **Role:** Data Engineer

Project 4 – Marks and Spencer UK, Procter & Gamble India, Holland and Barrett UK (June 2021 to February 2022)

1. **Description of the Project:**
   - Developed dashboards for executive-level reporting using Power Bi and SQL queries in PostgreSQL.
   - Executed complex SQL queries from PostgreSQL database and data transformations from Apache Spark in Azure Databricks.
   - Worked on data analysis using Pandas and NumPy for extraction and manipulation of the database.
   - Successfully migrated on-premises databases to Azure Data Lake Gen2 and Azure SQL Database.
   - Managed and maintained PostgreSQL database, optimizing performance through query optimization, regular maintenance tasks.

2. **Role:** Data Analyst

Technical Skills:
- **Database Systems:** SQL(Oracle/PostgreSQL/MYSQL), RDBMS, NoSQL, AWS RDS
- **Subject Area:** Data Engineer, ETL, ELT, Data warehouse, Big Data Technologies, Data Visualization tools, Database, Azure Cloud
- **Cloud & Big Data Platforms:**
  1. **Big Data Technologies:** Hadoop, Spark, PySpark, Kafka, HDFS, Map Reduce, Hive, Sqoop, Flume
  2. **Azure Cloud:** Azure Data Lake, ADLS gen2, Azure Data Factory, Synapse, Lakehouse
  3. **Amazon AWS:** AWS Sagemaker, Redshift, EC2, S3, RDS, DynamoDB, Lambda, SNS, IAM
- **Language/Platforms:**
  Python (with Pyspark for distributed data processing)
  SQL (advanced querying, performance tuning, data modeling)