

# Vikee Sonawane

## Data Engineer

+91 7303597047 | [vikeesonawane99@gmail.com](mailto:vikeesonawane99@gmail.com) | Thane, India | [Linkedin](#)

### SUMMARY

Result-driven Data Engineer with 2.5 years of experience in designing and optimizing efficient ETL pipelines using PySpark, Databricks, SQL, Python and Azure Cloud services. Proven ability to enhance data processing speed by 2x and improve data accuracy by 20%. Skilled in performance tuning and cloud-based data engineering solutions.

### SKILLS

- **Programming Languages:** Python, SQL
- **Big Data Technologies:** Spark (DataFrame, SparkSQL), Delta Lake, Hadoop, HDFS, Data Modeling
- **Azure Cloud Technologies:** Databricks, Data Factory, Data Lake, SQL DB, Key Vault
- **Version Control:** GitHub
- **Database:** PostgreSQL
- **Operating System:** Linux

### PROFESSIONAL EXPERIENCE

#### DXC Technology

Oct 2022 – Present

#### Data Engineer

Mumbai, India

#### Project – Leading Retail Company

- Optimized Spark workloads in Databricks by implementing Partitioning, Bucketing, Z-ordering, and Delta caching, leading to a 2x reduction in processing time.
- Performed data quality checks and validations in ADF and Azure Databricks, increasing data accuracy by 20%.
- Designed, developed and optimized end-to-end data pipelines using Azure Data Factory (ADF) to integrate data from diverse sources.
- Implemented a Medallion Architecture in a Lakehouse environment, improving data quality and structure through Bronze, Silver, and Gold layers, reducing data retrieval time by 15%.
- Integrated Azure Key Vault with various Azure services to securely store and access secrets and keys, ensuring compliance with industry standards and enhancing data security.
- Collaborated with stakeholders for requirement gathering to provide tailored ETL solutions.
- Incorporated Delta Lake with Azure Data Lake to utilize ACID transactions, enforce schema, and enable time travel features, ensuring consistent and reliable data.

#### Project – Leading Financial Service Company

- Developed a solution to calculate the Risk score for loan applicants, improving credit risk assessments by 20%.
- Created and optimized SQL queries to enhance database performance and reduce execution time.
- Conducted performance tuning on PySpark using Partitioning, Caching, and Adaptive Query Execution (AQE) techniques optimizing queries and improving execution time by 30%, leading to faster data processing.
- Thoroughly cleaned and preprocessed large financial datasets from clients, ensuring high precision and reliability.
- Utilized Parquet file format to optimize data storage and retrieval, reducing storage costs by 25%.

### EDUCATION

#### D.G.Ruparel College

July 2019 – June 2022

#### Bachelor of Science – Physics

Mumbai, India

### ACHIEVEMENTS

- **Champ Award** – Recognized for optimizing PySpark jobs, reducing data processing time by 30%.
- **Emerald Award** - Awarded for exceptional technical contributions, continuous learning, and problem-solving skills.
- **Top Rated Employee** - Received an exceptional impact rating for consistently performing well.

### CERTIFICATIONS

- Databricks Fundamentals
- Azure Data Fundamentals (DP900)
- Azure Fundamentals (AZ900)
- AWS Cloud Practitioner (CLF-C02)