Umesh Bhausaheb Bakhale
Data Engineer
Phone: +91 7378356595 | **Email:** bakhaleumesh5@gmail.com

Professional Summary

Experienced Data Engineer with 2+ years of expertise in designing and managing data pipelines and ETL processes. Skilled in database management and cloud platforms, including AWS and Azure. Passionate about transforming raw data into actionable insights to enhance business decision-making and operational efficiency.

Technical Skills

- **Programming Languages:** Python, SQL
- **Big Data Technologies:** Hadoop, Apache Spark, HDFS, Hive, MapReduce
- **ETL Tools:** Azure Data Factory (ADF), Databricks, PySpark, Spark SQL, Delta Tables
- **Cloud Platforms:** Azure (Blob, ADLS Gen2, ADF, Databricks, Synapse Analytics)
- **Database Management:** MySQL
- **Version Control:** GitHub, Azure DevOps
- **Visualization Tools:** Power BI
- **Operating Systems:** Windows, Linux

Work Experience

Data Engineer

Adunhill Technology Private Limited | 3 Years

- Developed optimized big data pipelines using **Azure Data Factory (ADF), Databricks, and Synapse Analytics**.
- Implemented data extraction logic to move data from RDBMS to Azure Databricks Delta tables using APIs.
- Designed and implemented **partitioning, bucketing, bin packing, indexing, caching**, and optimization techniques for Delta tables.
- Created and managed **Azure Databricks Python UDFs** and **Spark SQL UDFs** for data processing.
- Experienced in **Hadoop HDFS, Sqoop, Hive**, and **cloud-based data warehousing**.

Projects

Project 1: Automated Data Engineering Platform for Real-Time Patient Insights and Hospital Management

- **Domain:** Healthcare (FuelXtract)
- **Duration:** 2 years
- **Objective:** Built a data pipeline to extract, transform, and load (ETL) data from **MySQL to Azure Data Lake**, enhancing analytical capabilities. Developed Power BI dashboards for business insights.
- **Tools & Technologies:** Azure SQL, ADLS, Blob Storage, Azure Databricks, ADF, Azure Key Vault, Azure Virtual Machines, PySpark.
- **Key Responsibilities:**
    - Designed and deployed ETL pipelines using **Azure Data Factory and Databricks**.
    - Performed **data profiling, cleansing, and transformation** using PySpark in Databricks.
    - Developed interactive **Power BI reports** for sales and management insights.
    - Optimized **SQL queries** for efficient data extraction from MySQL.
    - Conducted extensive **testing and debugging** to ensure data accuracy.

Project 2: Cloud-Based IoT Data Pipeline for Predictive Insights

- **Domain:** IoT & Oil & Gas Industry (SensorFlowX)
- **Duration:** 1 year
- **Objective:** Developed a **scalable and automated data pipeline** integrating **IoT sensor data, production metrics, and pricing forecasts** for real-time analytics and predictive insights.
- **Tools & Technologies:** Azure Data Factory, ADLS, Azure Synapse Analytics, Azure Databricks, Power BI, Apache Spark, PySpark, Blob Storage, Spark SQL, Azure DevOps, GitHub.
- **Key Responsibilities:**
    - Configured **Azure Data Factory** pipelines for real-time data ingestion.
    - Implemented **Spark Streaming** techniques for enriched real-time data processing.
    - Created **real-time dashboards in Power BI** integrated with Azure Synapse Analytics & Databricks.
    - Automated pipeline scheduling, monitoring, and performance optimization.

Education

Babasaheb Ambedkar Marathwada University

- **Bachelor of Science**
- **Grade:** 72.70% (A Grade)

Maharashtra State Board

- **HSC:** 67.85%
- **SSC:** 84.40%

Languages

- English
- Hindi
- Marathi