

VIMALKUMAR G

[Portfolio](#) | Bengaluru, Karnataka | +91 9952663942 | vimalkumargt3996@gmail.com | [LinkedIn](#) | [Github](#)

PROFILE SUMMARY

Results-driven Data Engineer with 2 years and 9 months of hands-on industry experience, available for **Immediate Joining**. Skilled in designing and building scalable ETL pipelines using Big Data technologies such as Spark (PySpark) and Hadoop. Experienced in cloud platforms including Microsoft Azure and Google Cloud Platform (GCP). Proficient in SQL, Python, and database management.

TECHNICAL SKILLS

Languages: SQL, Python

Big Data & Frameworks: Apache Spark (PySpark), Apache Hadoop, Apache Kafka

Cloud Platforms: Azure: Azure Data Factory, Azure Databricks, Azure Data Lake Storage, Azure Synapse Analytics
Google Cloud Platform (GCP): GCS, Dataproc, Pub/Sub

Data Warehousing & Modelling: Snowflake, Big Query

Databases: MySQL, Oracle, PostgreSQL, MongoDB

Tools: Airflow, Big query, Excel, Word

CI/CD & DevOps: Git, Jenkins, Azure DevOps, Harness, Kubernetes, Postman

Project Tools: Jira, Confluence

File Formats: CSV, JSON, Parquet, orc

WORK EXPERIENCE: 2.9 YEARS

DATA ENGINEER at Cognizant Technology Solutions:

April 2022 – Feb 2025

Tech Stack: SQL, Python (Pyspark), Databricks, Microsoft Azure, Airflow, Delta Lake, CI/CD, Devops, Excel

- **Ingested and transformed** data from diverse sources including **MySQL, PostgreSQL, Oracle, MongoDB**, and **Cassandra**, converting structured, semi-structured, and unstructured data into business-ready formats.
- Implemented advanced **optimization** techniques that improved data processing efficiency and reduced execution time by up to **30%**.
- Worked extensively with file formats such as **CSV, JSON, Parquet**, and **ORC**, gaining deep hands-on experience in data handling, transformation, and optimization across various storage and processing systems.
- Leveraged big data frameworks like **Apache Spark** and **Apache Hadoop** (using SQL and PySpark) to build scalable and high-performance data solutions.
- Administered and optimized Relational Database Management Systems (RDBMS), demonstrating strong expertise in **SQL** querying, indexing, and performance tuning.
- Designed, scheduled, and monitored complex **ETL** pipelines using **Apache Airflow**, orchestrating workflows and integrating with processing engines such as Spark and storage systems like Delta Lake.
- Built and optimized ETL workflows in Databricks using PySpark and **Delta Lake**, enabling efficient data handling, versioning, and **ACID** transactions.
- Collaborated **cross-functionally** with data scientists, analysts, and software engineers to gather requirements and deliver robust, scalable, and production-grade data solutions.
- Proficient in **DevOps** practices, with hands-on experience using **Kubernetes**, Jenkins, Git, and Harness. Accessed Kubernetes clusters to monitor pod health, validate deployments, troubleshoot issues, and ensure correct configuration and build integrity.

CERTIFICATIONS:

- **Microsoft** Certified Data Engineer Associate (DP-203)
- **Databricks** Certified Data Engineer Associate

EXTERNAL PROJECTS:

- Designed an end to end to **Automated Event-Triggered Incremental Data Pipeline** using Google Cloud Platform (GCP) Services like Dataproc and Databricks. [Project Link](#)
- Built an end-to-end **Olympics Data Engineering project using Azure Cloud Services** such as Azure Data Factory (ADF), ADLS Gen 2, and Azure Synapse Analytics, Delta Lake using Spark SQL and Pyspark. [Project Link](#)

EDUCATION:

Bachelor of Engineering (BE):

Kongu Engineering College, Erode, Tamil Nadu

April 2018 – April 2022, CGPA (8/10)