

# Shivansh Singh

+60 162650096 | +91 9454960857 | ShivanshSingh1008@gmail.com | LinkedIn

## PROFILE

Data Engineer with 5+ years of experience designing and building scalable data pipelines, ETL workflows, and data warehouse solutions. Skilled in **Python, SQL, Spark, Hadoop, Hive, ODI** with a strong focus on data quality, reliability, and performance. Adept at transforming raw data into meaningful insights to support analytics and decision-making.

## TECHNICAL SKILLS

**Programming Languages:** Python, SQL  
**Big Data Technologies:** Spark, PySpark, Spark SQL, YARN, Hadoop, Hive, Impala, HDFS  
**Data Engineering Tools:** ODI(Oracle Data Integrator), Data Modelling, ETL/ELT data Pipeline  
**Cloud computing:** Azure ,ADF(Azure data Factory), Databricks, ADLS, Bigquery  
**Orchestration:** Zena  
**Certification:** Microsoft certified Azure Data Engineer associate

## EXPERIENCE

- Maybank** June 2024 - Present  
*Senior Data Engineer* Kuala Lumpur, Malaysia
  - Designed, built and maintained end-to-end ETL pipeline in Pyspark on Hadoop ecosystem (HDFS,hive) an ingesting 1M+ record/day into the International data warehouse(IDW) for Basel-3 reform.
  - Developed ETL pipelines to ingest raw CSV data, apply transformations using **PySpark**, and store the output in optimized **Parquet** format for efficient querying and downstream processing.
  - Implemented Spark optimization techniques such as **caching, multithreading, and broadcast** joins, resulting in a 20% decrease in processing time for handling a daily load of around **2 Million records**.
  - Conducted in-depth data analysis using **Hive, Impala**, and **Spark SQL**, providing SIT/UAT fixes and ensuring smooth operations in the production environment.
- Vodafone** June 2022 – May 2024  
*Data Engineer* Pune, India
  - Built modular **ADF** pipelines for data ingestion and then used **Databricks** notebooks to perform complex data cleansing and enrichment, improving maintainability and accelerating downstream reporting.
  - Migrated legacy ETL processes from Oracle Data Integrator (**ODI**) to PySpark, improving maintainability and reducing processing time by 40%.
  - Leveraged **Spark performance tuning** to achieve a 20% reduction in ETL job runtime on high-volume production pipeline.
  - Worked on data ingestion pipeline to ingest flat files in the **Data Lake**.
- Tata Consultancy Services** July 2019 - June 2022  
*ETL Developer* Bangalore, India
  - Implemented the ETL process using **Oracle Data Integrator (ODI)**, configuring repositories, creating ODI models, and designing projects with interfaces, packages, procedures, variables, and knowledge modules for data processing on **Hadoop** and **HDFS**.
  - Streamlined and simplified a complex ODI-based pipeline by removing redundant steps, eliminating hard coded components, and optimizing performance for downstream **Hive** queries—resulting in a more efficient and maintainable process.
  - Collaborated with stakeholders to define and document user requirements, subsequently designing and executing unit test cases and test scripts.

## EDUCATION

**Rajiv Gandhi Proudhyogiki Vishwavidyalaya** Bhopal, India  
*Bachelor of Engineering in Electronics and Communication* July 2015 - June 2019