**SHUBHAM WADEKAR**

[youremail@gmail.com](mailto:youremail@gmail.com) | +91-000000000 | https://www.yourlinkedinprofilelink.com

**EXPERIENCE**

**Data Engineer | Company Name | India | December 2021 – Present**

* Designed and implemented a comprehensive data pipeline integrating diverse data sources and leveraging technologies such as Apache Spark, Hive, Kafka, MySQL, HBase.
* Improved data processing efficiency by 30% through Spark-Hive and Spark-HBase integrations, reducing latency in real-time and batch processing scenarios.
* Automated data transfers using Sqoop and Airflow, resulting in a 2x reduction in data transfer time between Mysql and Hive.
* Developed Spark Streaming applications to enforce fraud detection rules, reducing false positives by 15% and enhancing the system's accuracy.
* Achieved a 40% increase in Kafka data ingestion speed, optimizing the streaming pipeline, and reducing overall processing time by 25% across various data pipeline components.
* Led performance tuning initiatives that optimized queries to yield a 5x improvement in data processing speed.
* Achieved and maintained an impressive 99.8% data pipeline up-time by leveraging Spark and Python to ingest streaming and transactional data from 8 primary sources; ensured seamless data flow and minimized disruptions

**Data Engineer | Company Name | India | December 2021 – Present**

* Used Airflow to build ETL solutions that helped improve conversion rates by 16%
* Orchestrated the development and continuous maintenance of a robust data pipeline, leveraging Google BigQuery, Google Cloud Storage, and Python to ingest and transform data from 5-6 disparate sources; achieved a 30% reduction in data ingestion time and enhanced data integrity.
* Communicated with business departments to understand needs and requests to build 20+ data pipelines for analyzing technical issues.
* Led the migration from Oracle and SQL to BigQuery resulting in an annual cost savings of $678,000 and an increase in performance of 14%
* Developed scalable data pipelines and ETL processes, using Python to process large datasets with a daily volume of 10 TB and improving ingestion and speed by 67%
* Analysed large datasets with SQL to identify trends and draw insights, resulting in a 29% increase in operational efficiency.
* Integrated Data Proc for Spark job submission, achieving an average data loading speed of 1 GB per minute into BigQuery Data Warehouse. This optimization led to a 10% improvement in overall data loading performance.

**TECHNICAL SKILLS**

**Big Data Technologies:** Hadoop, Spark, Hive, Sqoop, PySpark, HBase, Airflow

**Cloud Technologies**: Google Cloud Platform - BigQuery, Data Fusion, Cloud Composer, Data Proc, Cloud Storage, Cloud SQL, Pub/Sub, Compute Engine, Cloud Spanner

**Languages:** Python, SQL

**Databases and Tools:** Oracle, MySQL, HBase

**Scheduling:** Airflow

**Platforms:** Windows, Linux.

**EDUCATION**

**Engineering college Name,** Bachelor of Technology in Mechanical Engineering, Pune, India | **August 2017 – August 2020**

**[Diploma College Name,](https://www.gpamravati.ac.in/)** [Diploma in Mechanical Engineering, Amravati, India |](https://www.gpamravati.ac.in/) **[August 2014 - August 2017](https://www.gpamravati.ac.in/)**

**CERTIFICATIONS**

**Associate Cloud Engineer**, Google | **September 2023 – September 2026**