Rihan Rashid Bagwan

+91-9130901749 | rihanbagwan98@gmail.com | LinkedIn | GitHub | Mumbai

Professional Summary

Software Engineer with 3+ years of experience in Python development, data engineering, and backend systems. Skilled in building ETL/ELT pipelines using Python, SQL, PySpark, and Apache Airflow. Experienced in data processing, ClickHouse, workflow orchestration, and performance optimization. Hands-on with Docker, Linux, shell scripting, and CI/CD for deploying reliable data workloads.

Technical Skills

Data Engineering: ETL/ELT, Data Pipelines, Incremental Loading, Batch Processing, Data Validation, Data Modeling

Processing & Orchestration: Python, SQL, PySpark, Apache Airflow, Celery, DuckDB, Pandas, Parquet

Databases: ClickHouse, MySQL, MongoDB, Elasticsearch, Redis

DevOps & Cloud: Linux, Docker, Kubernetes (basic), Git, CI/CD, AWS, Azure

Experience

Software Development Engineer **Nov 2025 – Present**

Skillmine Technology Consulting Pvt Ltd.

- Built Python ETL/ELT pipelines from databases, REST APIs, and CSV/XLSX/PDF files into ClickHouse, processing 15–20 GB (about 5M records) daily across 12+ pipelines.
- Added schema inference, validation, and normalization, reducing rejected records by about 30%.
- Built async ingestion with Celery and DuckDB, with retries and data recovery that cut failed loads by around 40%.
- PySpark used for large-file batch transformation (CSV to partitioned Parquet) and Airflow to orchestrate 8 DAGs.
- Deployed and troubleshooting data services on Docker (system, log analysis), fixing memory bottlenecks that caused job failures.

Software Development Engineer **Jan. 2024 – Nov 2025**

Reliance Jio Platforms Limited, Navi Mumbai

- Automated data ingestion and synchronization using Python, SQL, and REST APIs, reducing manual effort by approximately 40% (10 hours per week).
- Optimized SQL queries using joins, CTEs, and stored procedures, reducing query latency by 25%.
- Datasets processed and transformed using Python, Pandas, and NumPy to support data analysis and downstream applications.
- Implemented data validation and reconciliation checks to improve data accuracy and maintain consistency across integrated data sources

Software Development Engineer **Aug. 2023 – Jan. 2024**

Dimensionless Technologies, Navi Mumbai

- Integrated GPT-3.5 Turbo into automation workflows, improving task efficiency by 35%.
- Pre-processed data for LLM fine-tuning and embeddings-based search pipelines in Python.
- Built real-time backend services with WebSockets and asyncio, cutting latency by 10%.

Projects

Batch Data Pipeline *Python, PySpark, Airflow, Clickhouse, Docker*

- Built an Airflow DAG to ingest approximately 1M rows daily, process and transform data using PySpark, and load partitioned Parquet datasets into Clickhouse for analytical querying.
- Added validation and incremental loading, cutting runtime by about 50%.
- Implemented logging & error-handling mechanisms to track pipeline failures, simplify troubleshooting, and ensure reliable data processing.

Education

GH Raisoni College of Engineering and Management, Pune **Jun. 2019 – Jun. 2023**

B.Tech in Computer Science