

SOHAIL SHAIK

Data Engineer | ETL/ELT | Data Warehousing

+91 96426 39515 | s.suhail9849@gmail.com | GitHub | LinkedIn

PROFESSIONAL SUMMARY

Data Engineer with 3+ years of experience working on data integration pipelines and data warehouse workflows. Experience includes SQL transformations, incremental data loads, data validation, reconciliation, and production troubleshooting using Oracle SQL/PL SQL and Informatica IICS/IDMC.

TECHNICAL SKILLS

Data Integration: Informatica IICS/IDMC, Cloud Data Integration, mappings, taskflows

Programming & Databases: SQL, Oracle, PL/SQL, Python, shell scripting

Data Warehousing: Staging, fact and dimension tables, star schema, SCD Type 1/2

Data Operations: Incremental loads, data validation, reconciliation, production troubleshooting

PROFESSIONAL EXPERIENCE

PrettiFlow

San Francisco, USA (Remote)

Founding Data Engineer

Jan 2026 – Present

- Built ingestion workflows for application data stores and third-party SaaS sources to prepare data for product analytics and reporting.
- Developed incremental data processing using SQL and Python/shell scripts.
- Organized staging and dimensional data structures to support analytics and reporting.
- Added data validation checks and Slack alerts to flag schema changes and data issues.

OmniqAI

Bengaluru, India

Data Engineer – ETL & Data Warehousing

Jan 2023 – DEC 2025

- Developed and maintained Informatica IICS/IDMC mappings and taskflows to integrate data from source systems into a data warehouse.
- Translated business requirements into source-to-target mappings, SQL transformations, and PL/SQL logic.
- Supported staging, fact, and dimension loads, including full and incremental processing and SCD Type 1/2 history tracking.
- Implemented pre- and post-load checks to identify missing, duplicate, or inconsistent records.
- Tuned SQL queries and transformation logic to improve batch processing.
- Investigated failed taskflows and production data issues, tracing problems to source data or transformation logic.

PROJECT

Transaction Data Pipeline & Reconciliation Warehouse

Python, SQL, PostgreSQL, Apache Airflow

- Built a batch pipeline to ingest simulated payment transactions and settlement files into staging tables, then transform them into a warehouse for transaction reporting.
- Implemented incremental loading with a stored watermark and idempotent database upserts to support safe reruns without duplicating previously loaded records.
- Developed SQL reconciliation checks to compare source and warehouse record counts and transaction totals, while capturing duplicate records, invalid transactions, and unmatched settlements for investigation.
- Modeled transaction data using fact and dimension tables and orchestrated scheduled pipeline runs with Apache Airflow, including task dependencies and failure handling.

EDUCATION

Bachelor's Degree in Engineering

2019 – 2022