# **User Manual**

## **Overview**

This project provides an ETL pipeline (Extract–Transform–Load) that ingests financial transactions into Postgres using PySpark, enriches them with fraud detection flags, and serves results through a Streamlit dashboard.

The system is fully containerized with Docker for one‑click setup. Alternatively, you can run it locally with Python.

## **Quick Start (Recommended: Docker)**

1. Install Docker Desktop(https://www.docker.com/products/docker-desktop) (Windows/macOS) or Docker Engine (Linux).

2. Download this project from GitHub:

3. Start the database:

docker compose up -d db

4. Run the ETL pipeline (loads data into Postgres):

docker compose run --rm etl

5. Start the dashboard:

docker compose up -d app

6. Open your browser at: \*\*http://localhost:8501\*\*

## **Option A: Run with Docker (full detail)**

### **Step 1 — Install prerequisites**

- Install Docker Desktop (Windows/macOS) or Docker Engine (Linux).

- Ensure `docker` and `docker compose` commands work.

### **Step 2 — Get the project**

Download the code from GitHub and extract/unzip it, or clone it with Git.

### **Step 3 — Navigate into project folder**

cd my-finance-fraud-etl/my-finance-fraud-etl

### **Step 4 — Start services**

- Launch Postgres only:

docker compose up -d db

- Run the ETL pipeline (this loads data and creates fraud flags):

docker compose run --rm etl

- Start the Streamlit app:

docker compose up -d app

### **Step 5 — Open dashboard**

Go to: [http://localhost:8501](http://localhost:8501)

## **Option B: Run locally (without Docker)**

This requires more setup.

### **Step 1 — Install dependencies**

- Python 3.11+

- Java 17

- Postgres (running locally, create database `etl\_db`)

### **Step 2 — Install Python packages**

pip install -r requirements.txt

### **Step 3 — Set environment variables**

export POSTGRES\_HOST=localhost

export POSTGRES\_PORT=5432

export POSTGRES\_DB=etl\_db

export POSTGRES\_USER=etl\_user

export POSTGRES\_PASSWORD=etl\_password

### **Step 4 — Run ETL pipeline**

python etl/etl\_pipeline.py

### **Step 5 — Run dashboard**

streamlit run app/streamlit\_app.py

## **Troubleshooting**

- If `localhost:8501` doesn’t load, check the container logs:

docker compose logs app

- If Postgres connection fails, confirm DB is healthy:

docker compose ps

- If `pyspark` errors about Java, ensure \*\*Java 17\*\* is installed.

## **Deliverables for Users**

- ETL pipeline (Python + PySpark)

- Postgres database with tables: `transactions`, `monthly\_summary`, `fraud\_flags`

- Streamlet dashboard for fraud monitoring