Developing ETL Pipelines: Case Study

ETL stands for extract, transform and load. For developing ETL pipelines, we need to extract data from various sources, transform it into a usable format, and load it into a data warehouse or database for analysis.

Your task is to develop a data ETL pipeline for the Fashion MNIST dataset. The dataset contains 70,000 grayscale images of size 28×28 pixels, categorized into ten clothing categories like t-shirts, dresses, sneakers, etc. The pipeline should extract the dataset from the data source, and perform the necessary transformations such as scaling, normalization and feature engineering.

Load the preprocessed data into an SQLite database for storage and easy retrieval.