High-Level Design (HLD)

Project: Automated ETL System for Banking Market Capitalization Data

Objective: To create an automated system that extracts, transforms, and loads data on the largest banks' market capitalization from a specified URL. The system will convert the data to various currencies and store it locally in both CSV and database formats.

Components:

1. **Data Extraction:**
    - **Source:** Web page containing market capitalization data of the largest banks.
    - **Method:** Use Python's requests and BeautifulSoup libraries to scrape and extract the relevant data.
2. **Data Transformation:**
    - **Operations:**
        - Convert market capitalization data from USD to GBP, EUR, and INR using provided exchange rates.
        - Round the converted values to two decimal places.
3. **Data Loading:**
    - **CSV File:** Save the transformed data to a local CSV file.
    - **Database:** Store the data in an SQLite database.
4. **Logging:**
    - **Purpose:** Track the progress and status of the ETL process.
    - **Method:** Write log messages to a log file with timestamps.
5. **Query Execution:**
    - **Purpose:** Run specific queries on the database to retrieve information.
    - **Method:** Use SQL queries to fetch and display data from the database.
6. **Automation:**
    - **Purpose:** Ensure the ETL process can be executed quarterly.
    - **Method:** Create a script that can be run periodically.

High-Level Workflow:

1. Extract data from the specified URL.
2. Transform the data to include market capitalization in GBP, EUR, and INR.
3. Load the transformed data into a CSV file and SQLite database.
4. Log the progress of each stage.
5. Run specific queries on the database and display the results.