Project: Automated ETL System for Banking Market Capitalization Data

1. Data Extraction:

Function: extract(url, table_att)

- **Inputs:**
  - url: URL of the web page to scrape data from.
  - table_att: List of initial table attributes.
- **Outputs:**
  - DataFrame with extracted data.
- **Steps:**
  1.1. Send an HTTP GET request to the URL.
  1.2. Parse the HTML content using BeautifulSoup.
  1.3. Locate the relevant table and rows containing data.
  1.4. Extract and clean data from the table.
  1.5. Append the data to a DataFrame.

2. Data Transformation:

Function: transform(df_)

- **Inputs:**
  - df_: DataFrame with extracted data.
- **Outputs:**
  - Transformed DataFrame with additional currency columns.
- **Steps:**
  1.1. Read the exchange rates from the CSV file.
  1.2. Convert the market capitalization values from USD to GBP, EUR, and INR using the exchange rates.
  1.3. Round the converted values to two decimal places.
  1.4. Append the new columns to the DataFrame.

3. Data Loading:

Function: load_to_csv(df_, file_path)

- **Inputs:**
  - df_: Transformed DataFrame.
  - file_path: Path to save the CSV file.

- **Outputs:**
  - None (saves DataFrame to CSV file).
- **Steps:**
  1. Save the DataFrame to a CSV file using pandas.

Function: load_to_db(df_)

- **Inputs:**
  - df_: Transformed DataFrame.
- **Outputs:**
  - None (saves DataFrame to database).
- **Steps:**
    1. Save the DataFrame to an SQLite database using pandas.

4. Logging:

Function: log_progress(message)

- **Inputs:**
  - message: Log message string.
- **Outputs:**
  - None (writes log to file).
- **Steps:**
    1. Get the current timestamp.
    2. Append the timestamp and message to the log file.

5. Query Execution:

Function: run_query(query_statement, conn_)

- **Inputs:**
  - query_statement: SQL query string.
  - conn_: SQLite database connection.
- **Outputs:**
  - None (prints query results).
- **Steps:**
    1. Execute the SQL query using pandas.
    2. Print the query statement and results.

Main Workflow:

1. **Initialize Logging:**
    1.1. Log the start of the ETL process.
2. **Extract Data:**
    2.1. Call the extract function with the URL and table attributes.
    2.2. Log the completion of data extraction.
3. **Transform Data:**
    3.1. Call the transform function with the extracted DataFrame.
    3.2. Log the completion of data transformation.
4. **Load Data:**
    4.1. Call the load_to_csv function to save the DataFrame to a CSV file.
    4.2. Call the load_to_db function to save the DataFrame to a database.
    4.3. Log the completion of data loading.
5. **Run Queries:**
    5.1. Execute predefined queries using the run_query function.
    5.2. Log the execution of queries.
6. **Close Connection:**
    6.1. Close the SQLite database connection.