Project 02: SSIS Package for Loading the Patient Table

Instruction:

1. Need to complete these projects on your own.
2.Try to understand the project by studying it multiple times.
3.Act like you're a Sr Business Intelligent and your manager tasked you with this project.
4.Dont ask any foolish question, only ask relevant questions and be prepared to face further counter questions related to the project.

Objective:

Develop a single SSIS package that extracts patient data from source files from the link, processes and loads the data into a SQL Server **Patient** table previously you created for project 01, and maintains detailed logging of the ETL process. The package must also ensure that the **Patient** table contains no duplicate records.

Source file links:
https://drive.google.com/drive/folders/1gWs8rLBaxYygtCCneHPmKwieri6dTZln?usp=sharing

Download these files into your desktop.

Requirements:

1. **Patient Table Enhancements:**
    - The **Patient** table must include two additional columns:
        - **load_date:** To store the date (and time) when the data was loaded.
        - **extract_filename:** To record the file name from which the data was extracted.
    - The table should enforce no duplicate values (you may use constraints or implement logic in your SSIS package to avoid duplicates).
2. **SSIS Package:**
    - **Single Package, Multiple Files:**
      Build one SSIS package that can process all the source files in a single run.
    - **Dynamic Data Loading:**
      Each record loaded into the **Patient** table should have the appropriate **load_date** and **extract_filename** values captured.

- **Data Quality:**
  The package must ensure that no duplicate records are inserted into the **Patient** table.
- **Logging:**
- Create a **Log** table to record the following details on every run:
  - **previous_row_count:** The number of rows in the **Patient** table before the current load.
  - **current_row_count:** The number of rows in the **Patient** table after the load.
  - **total_rows_loaded:** The number of rows loaded during the current run.
  - **filename:** The name of the file processed.
  - **load_date:** The date and time of the load.
  - **Other details:** Any additional information you find useful

    Ensure that the log table is updated on every SSIS package run, so that an auditor or team member can review the load history.

Flexibility(You can do these things if you want):

- **Schema and Constraints:**
  You are allowed to remove the primary key or any constraints from the **Patient** table if necessary during development. However, remember that the final solution must avoid duplicate records.
- **Data Type Adjustments:**
  You may change data types if needed (for example, to simplify conversions and avoid runtime errors).
- **Staging:**
  Although optional, creating a staging table to pre-load and clean data before loading the final **Patient** table is encouraged to ensure data quality and easier debugging.

Deliverables:

- **SSIS Package File (.dtsx):**
  The complete package that processes the source files, loads the **Patient** table, and writes logs to the **Log** table.
- **SQL Scripts:**
  - Script to create (or modify) the **Patient** table (with additional columns).
  - Script to create the **Log** table.

- **Documentation:**
  A short nice documentation explaining:
    - How the package works.
    - How duplicate records are handled.
    - How the logging is implemented.
    - Any assumptions made or any changes to the schema.