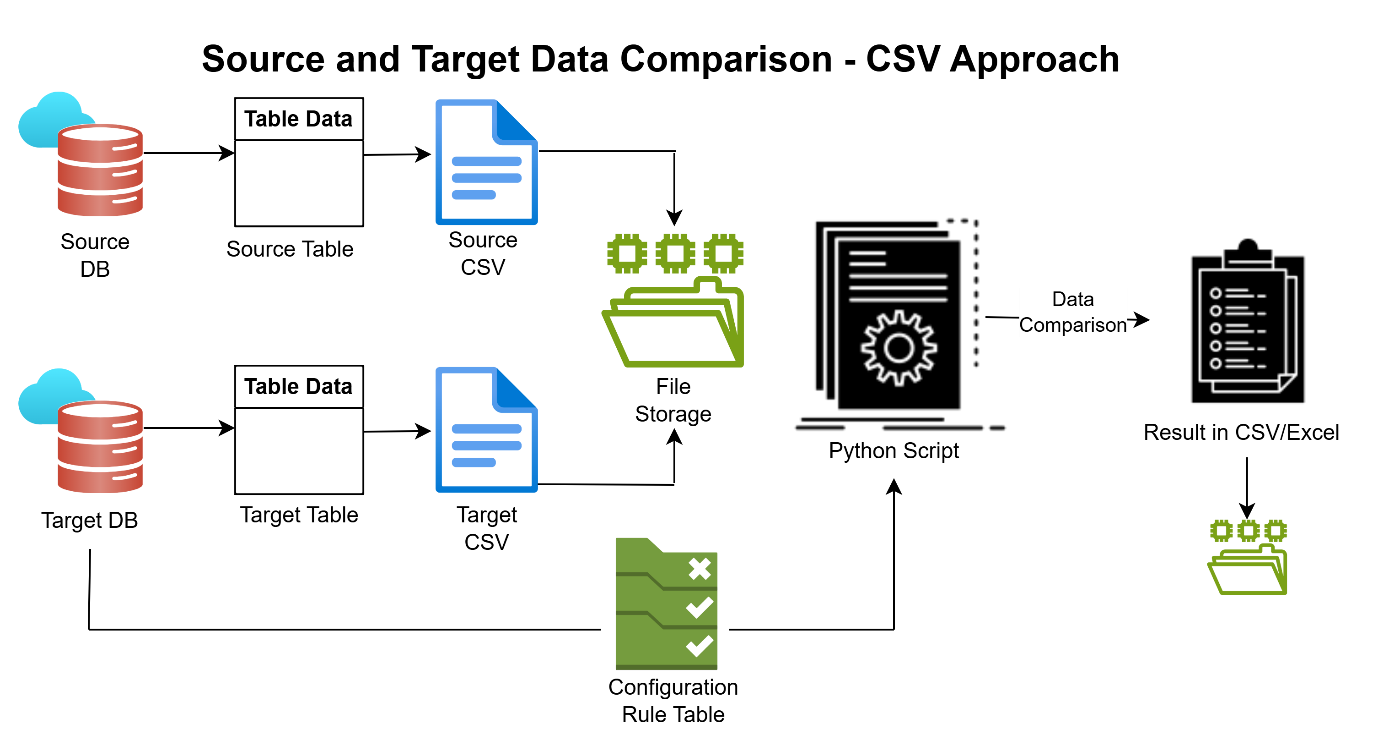
**Documentation for the Provided Code**

**Overview**

This Python script performs a comparison between source and target CSV files based on configurations provided in a configuration file. It includes functionality for:

1. Comparing row counts between source and target files.
2. Performing row-by-row comparisons to identify mismatches.
3. Logging the results in a test execution result CSV file.
4. Handling large datasets using chunk-based processing.
5. Tracking memory usage and execution time.

**WorkFlow Diagram**



**Modules Used**

1. [os](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html):
   * Used for file path operations and checking file existence.
2. [pandas](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html):
   * Used for reading, processing, and writing CSV files.
3. [datetime](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html):
   * Used for generating timestamps for logging and file naming.
4. [time](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html):
   * Used for measuring execution time.
5. [psutil](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html):
   * Used for tracking memory usage.

**Code Components**

**1.**[load\_active\_records(config\_file)](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html" \o ")

**Purpose**:

* Reads the configuration file (config.csv) and filters only the active tasks where the Enable column is set to 'Y'.

**Parameters**:

* [config\_file](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Path to the configuration file.

**Returns**:

* A DataFrame containing active tasks.

**2.**[prepare\_test\_execution\_result](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html" \o ")

**Purpose**:

* Prepares a dictionary containing test execution results.

**Parameters**:

* [test\_execution\_id](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Unique ID for the test execution.
* [test\_case\_id](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): ID of the test case.
* [test\_case\_description](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Description of the test case.
* [table\_name](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Name of the table being compared.
* [execution\_status](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Status of the test case (Passed or Failed).
* [execution\_details](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Additional details about the test case.

**Returns**:

* A dictionary containing the test execution result.

**3.**[compare\_row\_counts](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html" \o ")

**Purpose**:

* Compares the total number of rows between the source and target DataFrames.

**Parameters**:

* [results](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): List to store the test execution results.
* [test\_execution\_id](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Unique ID for the test execution.
* [source\_df](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Source DataFrame.
* [target\_df](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Target DataFrame.
* [source\_table](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Name of the source table.
* [target\_table](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Name of the target table.

**Appends**:

* Test execution results to the [results](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) list.

**4.**[compare\_row\_by\_row\_sorted\_chunks\_nested\_for](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html" \o ")

**Purpose**:

* Performs row-by-row comparison between source and target DataFrames in chunks.
* Identifies mismatches and duplicate rows.

**Parameters**:

* [results](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): List to store the test execution results.
* [test\_execution\_id](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Unique ID for the test execution.
* [source\_df](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Source DataFrame.
* [target\_df](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Target DataFrame.
* [source\_table](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Name of the source table.
* [target\_table](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Name of the target table.
* [output\_folder](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Folder to save mismatch and duplicate reports.
* [chunksize](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Number of rows to process in each chunk.
* [id\_columns](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): List of columns to use as keys for comparison.

**Appends**:

* Test execution results to the [results](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html) list.

**Outputs**:

* CSV files for mismatches and duplicates.

**5.**[process\_single\_table\_pair](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html" \o ")

**Purpose**:

* Processes a single source-target table pair by performing row count validation and row-by-row comparison.

**Parameters**:

* [args](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Tuple containing:
  + [test\_execution\_id](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Unique ID for the test execution.
  + [source\_table](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Name of the source table.
  + [target\_table](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Name of the target table.
  + [sort\_columns](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Columns to sort the data by.
  + [source\_file\_path](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Path to the source CSV file.
  + [target\_file\_path](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Path to the target CSV file.
  + [output\_folder](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Folder to save mismatch and duplicate reports.
  + [chunksize](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Number of rows to process in each chunk.

**Returns**:

* A list of test execution results for the table pair.

**6.**[process\_operations](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html" \o ")

**Purpose**:

* Orchestrates the entire comparison process for all active tasks.
* Uses multiprocessing to process multiple table pairs in parallel.

**Parameters**:

* [config\_file](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Path to the configuration file.
* [source\_folder](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Folder containing source CSV files.
* [target\_folder](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Folder containing target CSV files.
* [output\_folder](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Folder to save mismatch and duplicate reports.
* [test\_execution\_file](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html): Path to the test execution report file.

**Steps**:

1. Loads active tasks from the configuration file.
2. Determines the starting TEST\_EXECUTION\_ID from the test execution report file.
3. Creates tasks for each source-target table pair.
4. Uses multiprocessing to process the tasks.
5. Aggregates results from all processes.
6. Writes the aggregated results to the test execution report file.

**7. Main Execution**

**Purpose**:

* Defines file paths and starts the comparison process.

**Steps**:

1. Defines file paths for the configuration file, source folder, target folder, output folder, and test execution report file.
2. Tracks the start time and memory usage.
3. Calls [process\_operations](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html" \o ") to perform the comparison.
4. Logs the total time and memory used.

**Input Files**

1. **Configuration File (config.csv)**:
   * Specifies the source and target tables, columns to sort by, and whether the task is enabled.
   * Example:

Source Tablename,Target Tablename,Sort\_Columns,Enable

source\_table1,target\_table1,id,Y

source\_table2,target\_table2,id,Y

1. **Source and Target CSV Files**:
   * Contain the data to be compared.

**Output Files**

1. **Mismatch Report**:
   * Logs mismatched rows with details of discrepancies.
   * Example: source\_table1\_detailed\_mismatches.csv
2. **Duplicate Report**:
   * Logs duplicate rows in the target dataset.
   * Example: target\_table1\_duplicates.csv
3. **Test Execution Report**:
   * Summarizes all test cases with statuses and execution details.
   * Example: [test\_execution\_result.csv](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html)

**Multiprocessing**

* The script uses Python's [multiprocessing.Pool](vscode-file://vscode-app/c:/Users/10735332/AppData/Local/Programs/Microsoft%20VS%20Code/resources/app/out/vs/code/electron-sandbox/workbench/workbench.html" \o ") to process multiple table pairs in parallel.
* Each process handles a single source-target table pair, performs the comparison, and returns the results.

**Error Handling**

* The script includes error handling for:
  + Missing or malformed files.
  + Mismatched column names or shapes.
  + Duplicate rows in the target dataset.

**Performance Optimization**

1. **Chunk-Based Processing**:
   * Processes data in chunks to reduce memory usage.
2. **Multiprocessing**:
   * Utilizes all available CPU cores for parallel processing.

**How to Run**

1. Place the script and input files in the same directory.
2. Update the file paths in the \_\_main\_\_ section.
3. Run the script:

python Data\_validation\_csv.py

**Future Enhancements**

1. **Cloud Support**:
   * Enable processing of files stored in cloud storage (e.g., AWS S3, Azure Blob).
2. **GUI Interface**:
   * Develop a user-friendly interface for non-technical users.