A picture containing text, indoor, wall, clock

Description automatically generated

DATA WAREHOUSE REPoRTS AUTOMATION – IMPLEMENTATION DOCUMENT

Table of Content

[1. Project Objective 2](#_Toc112681379)

[2. SOFTWARE REQUIREMENTS 3](#_Toc112681380)

[3. PYTHON PACKAGE REQUIREMENTS 3](#_Toc112681381)

[4. FOLDER STRUCTURE FOR PROGRAM 5](#_Toc112681382)

[5. RUNNING PYTHON FILE 6](#_Toc112681383)

[6. PROGRAM EXECUTION 8](#_Toc112681384)

[7. READING SOURCE DATA 9](#_Toc112681385)

[8. ENTITY RELATIONSHIP DIAGRAM 10](#_Toc112681386)

[9. REPORT GENERATION 11](#_Toc112681387)

[10. REPORT DELIVERY (E-MAIL FUNCTIONALITY) 11](#_Toc112681388)

[11. PROCESS COMPLETION 11](#_Toc112681389)

[12. ADDING/DELETING A CLIENT, REPORT RECIPIENT, ETC. 12](#_Toc112681390)

[13. REPORT LOGIC 15](#_Toc112681391)

[14. MISCELLANEOUS 16](#_Toc112681392)

1. Project Objective

To build a custom Data Warehouse to meet Greenberg Traurig LLP (“GT”)’s client reporting needs. Specifically, to meet the current requirement of running weekly Open Cases Report for Attorneys, and weekly Case Status Report (CSR) for end-clients.

The solution involves a code-based program using Python and SQL that would automate the process of importing data, preparing reports, and delivering reports on/at scheduled days/times, to designated stakeholders.

1. SOFTWARE REQUIREMENTS

|  |  |
| --- | --- |
| Software Requirements | |
| **Operating System** | Windows 10 64 bit |
| Database | MS SQL SERVER 2019 |
| Database Management Tool | SQL Server Management Studio 18 (SSMS)  https://www.microsoft.com/en-in/sql-server/sql-server-downloads?rtc=1 |
| Python | Python – Version 3.9.13 |
| VS Code | (Version 1.53) and above |
| Chromelessdriver | <https://chromedriver.chromium.org/downloads>  (Download driver based on chrome browser version) |

1. PYTHON PACKAGE REQUIREMENTS

|  |  |
| --- | --- |
| List of required External python packages | |
| Chardet 3.0.4 | For character detection while reading source excel |
| Openpyxl-0.10 | for excel formatting |
| Pandas-1.4.3 | To retrieve and manipulate source file data |
| Pyodbc-0.32 | To establish connection between Python code and DBMS |
| XlsxWriter-.0.3 | to write data from data frames to external excel files |

1. FOLDER STRUCTURE FOR PROGRAM

Below is the folder structure to house to program files, source files, and processed reports.

* + GT Immigration – Operational Reports
    - Processed Reports
    - Source Data

Graphical user interface, application, Word

Description automatically generated

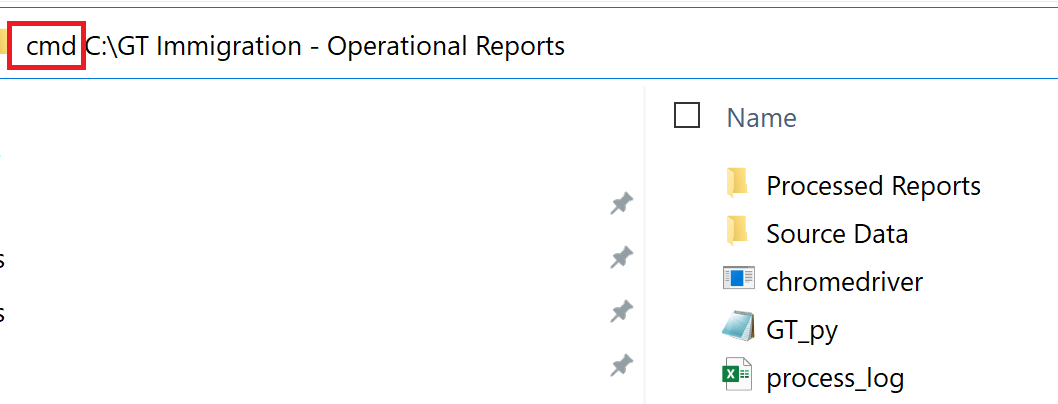
* **Source Data:** The source Beneficiary and Case Data files would be uploaded to this folder.
* **Processed Reports:** Once processed, the processed reports would be emailed to the designated recipient, and also saved to this folder for access, as needed.
* **GT.py:** This is the executable Python file that we have to run it in the command prompt.
* **Run\_application.bat:** This is the batch file that is either double-clicked or mapped to any task scheduler to invoke the actual python script **GT.py**

1. RUNNING PYTHON FILE

The program is designed to run automatically via a task scheduler app (to be configured by GT) designated to run on defined dates and at defined times, by triggering the “Run\_application” batch file.

Alternatively, the program can also be run manually, as needed. Below are the steps for running the program manually.

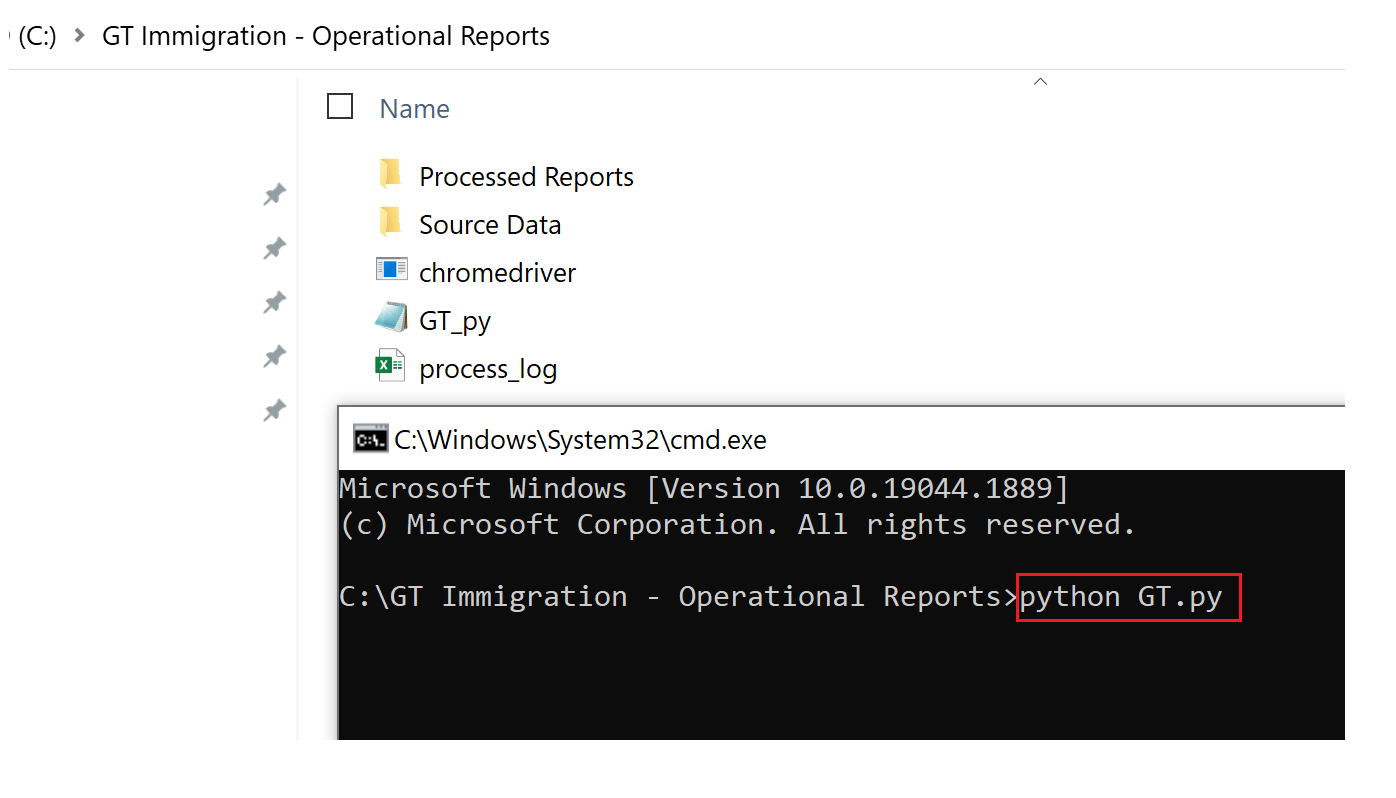
* **Step 1**: Go to the respective folder and type “**cmd**” in the address bar and click enter. It will open a command prompt window. The command prompt window can also be opened by



Graphical user interface, text, application

Description automatically generated

* **Step 2**: Then, type the following in the command prompt window to run the python app file (GT.py):



1. PROGRAM EXECUTION

Once executed, the program -

* **Step 1**: Reads the “**Source Files**”**.**

Monday Reports Source Files

* Reports Automation\_Beneficiary Data\_INSZoom\_mmddyyyy
* Reports Automation\_Case Data\_INSZoom\_mmddyyyy
* Reports Automation\_Open Cases\_INSZoom\_mmddyyyy

Tuesday-Thursday Reports Source Files

* Reports Automation\_Beneficiary Data\_INSZoom\_mmddyyyy
* Reports Automation\_Case Data\_INSZoom\_mmddyyyy
* **Step 2**: Ingests/imports the data into the GT Data Warehouse.
* **Step 3**: Generates the predefined daily reports.

The program performs two different cycles depending on the day the reports are generated (see details below).

Monday -

**Cycle 1**

* Client Name\_Status Report\_mmddyyyy
* Client Name\_Document Expiration Report\_mmddyyyy

**Cycle 2**

* Open Cases\_Attorney Name\_mmddyyyy

Tuesday-Thursday -

**Cycle 1**

* Client Name\_Document Expiration Report\_mmddyyyy
* Open Cases\_Attorney Name\_mmddyyyy
* **Step 4**: Emails the report(s) to the designated recipient. Refer to “**Client Details**” Table in the Database for details.
* **Step 5**: Saves the reports in the “**Processed Reports**” folder, for further access.
* **Step 6**: Deletes the data in the Database.

1. READING SOURCE DATA

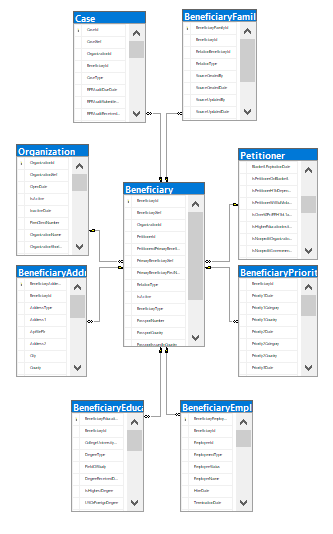
The Source File name is critical for the program to recognize the appropriate file.

* Reports Automation\_Beneficiary Data\_INSZoom\_mmddyyyy
* Reports Automation\_Case Data\_INSZoom\_mmddyyyy
* Reports Automation\_Open Cases\_INSZoom\_mmddyyyy

Following are steps taken by the program for reading the source data:

* **Step 1:** The code searches for the “Beneficiary Data” file in the “Source Data” folder that contains a current date (mmddyyyy) in the file name.
* **Step 2:** The code utilizes pandas framework to read the content of the file and store within as data frames.
* **Step 3:** The code reads the number of rows the dataframe has and loops repetitive for ‘n’ number of rows the dataframe holds.
* **Step 4:** During each looping, the code reads the data of all the columns for each row and assigns value to the respective columns as defined by the mappings and pushes/adds the data to the Beneficiary table as denoted in the respective SQL code.
* **Step 5:** The loop ends once feeding data of all the rows in dataframe that is extracted from the source beneficiary file.
* **Step 6:** Once the Beneficiary data is inserted int the DBMS, the process is repeated to ingest the Case Data.

1. ENTITY RELATIONSHIP DIAGRAM

****

1. REPORT GENERATION

Once the data ingestion process is completed, depending on the report run day, the designated reports are generated. The report templates listed below are included for reference.

1. Case Status Report\*
2. Document Expiration Report
3. Attorney Open Cases Report

\* This report contains the following tabs:

* Open Cases -
* Filed & Pending Cases
* Approved & Closed Cases
* Cases on Hold
* Priority Date Report
* Active Beneficiary List

1. REPORT DELIVERY (E-MAIL FUNCTIONALITY)

Once the designated reports have been processed, in addition to saving a copy of the reports to the “**Processed Reports**” folder, the program will also e-mail a copy of the report(s) to the designated recipient(s)\* pursuant to the recipient’s email address listed in the “**ClientDetails**” Tale of the Database (see “**ClientDetails**” Table in the Database for details).

\* Both Primary and Secondary Recipients.

GT would have to configure the e-mail address from where the e-mails will be generated.

1. PROCESS COMPLETION

Subject to timely availability of stakeholders’ time and attention, and access to relevant information and documentation needed to complete the assessment, ImmiLytics will deliver an assessment report in approximately three (3) weeks from the time all of the required information and documentation is provided to ImmiLytics, and deploy the Data Warehouse in five (5) weeks from the date of completion of the assessment.

Text

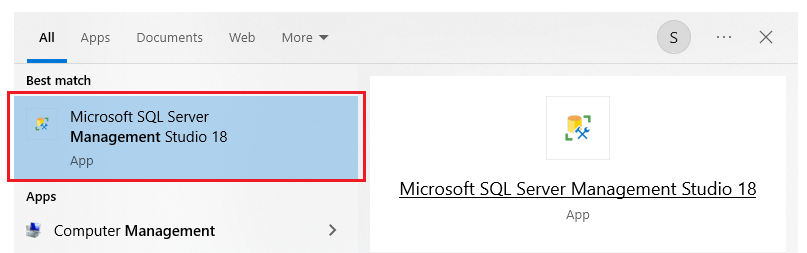
Description automatically generated

1. ADDING/DELETING A CLIENT, REPORT RECIPIENT, ETC.

The addition or deletion of a Client and/or a Report Recipient is controlled from the “ClientDetails” Table in the Database. Please see the table for details.

Following are steps taken for adding or deleting a Client, Attorney, Paralegal, etc.:

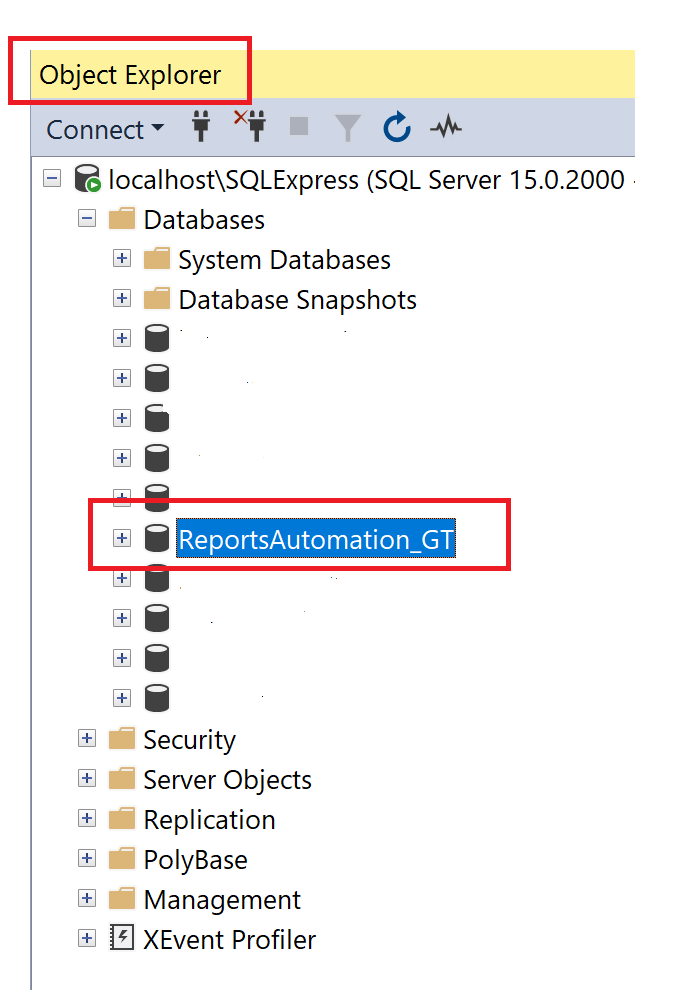
* **Step 1:** Open SQL Management Studio 18 from the “Start” menu.



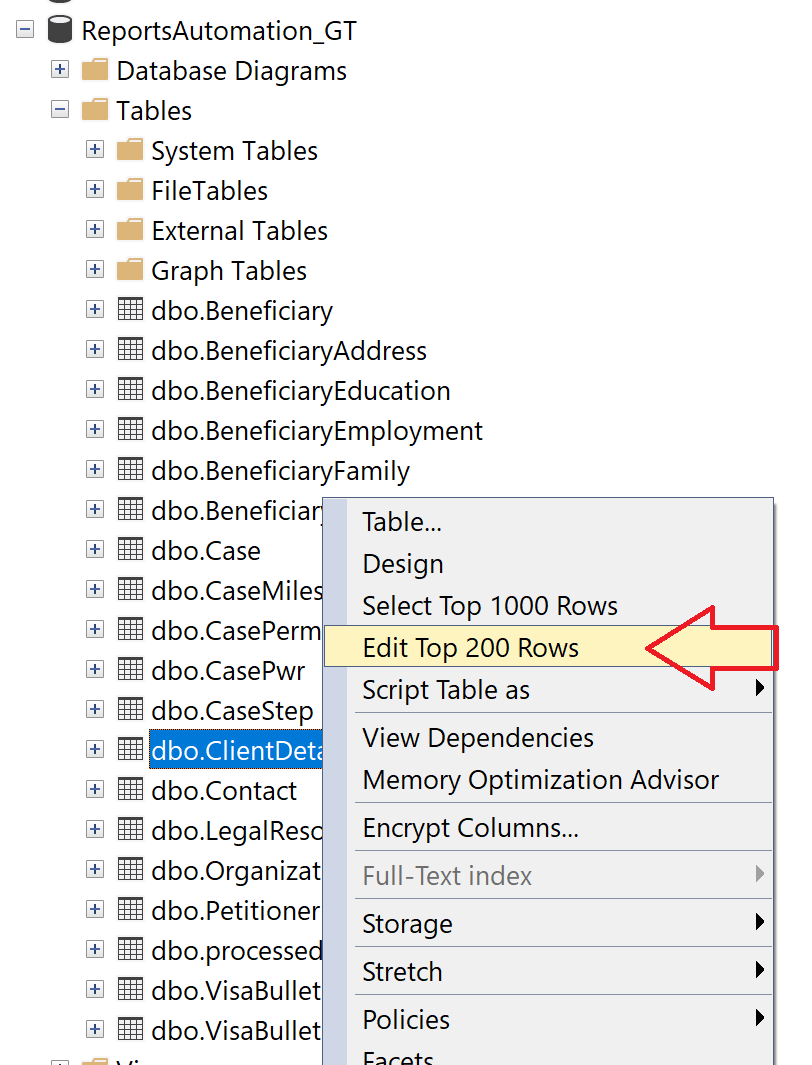
Graphical user interface, application, Word

Description automatically generated

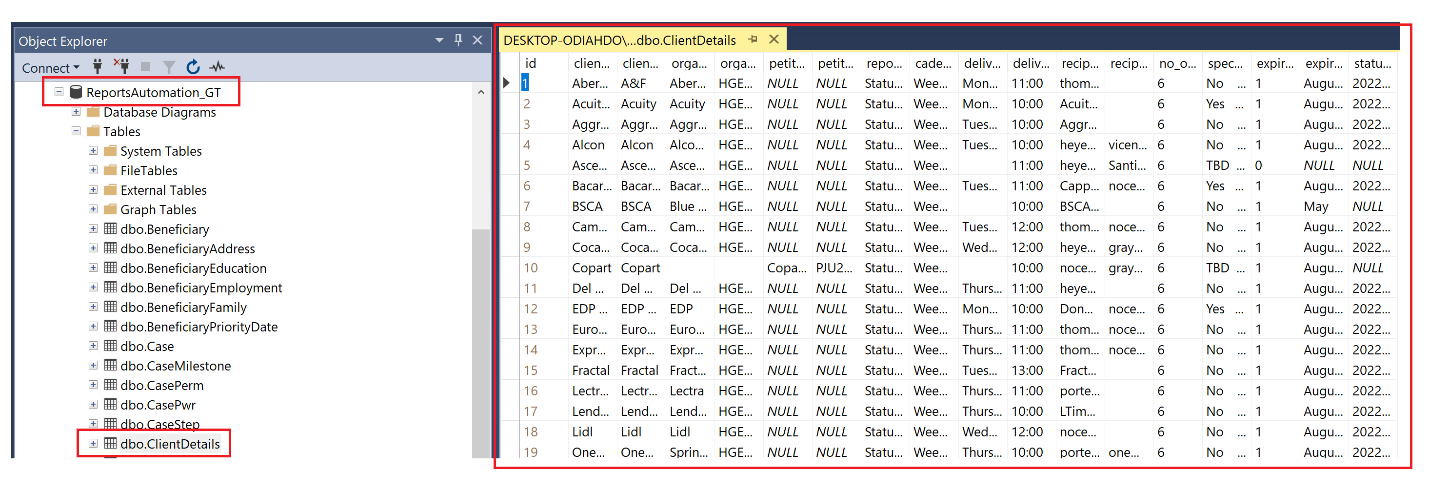
* **Step 2:** From the Object Explorer menu on the left select the Database “**ReportsAutomation\_G**T”.



* **Step 3:** Expand the folder ‘Tables”, find “**dbo.ClientDetails**”, select the table and right-click mouse and click on “**Edit Top 200 Rows**”



* **Step 4:** In the resulting table, update Client Name, Delivery Date, Delivery Time, Recipient E-Mail Address, etc. as needed.



**Note:** In the “recipient\_to” and “recipient\_cc” columns, if multiple e-mails addresses have to entered then separate the e-mail addresses using “;”.

1. REPORT LOGIC

The underlying business logic related to data ingestion, mapping, and report generation can be found in the GT.py file by referring to lines in the code (provided below).

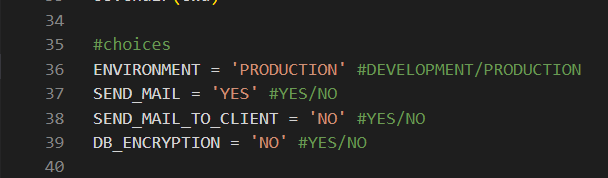
From: Row \_\_

To: Row \_\_

1. MISCELLANEOUS

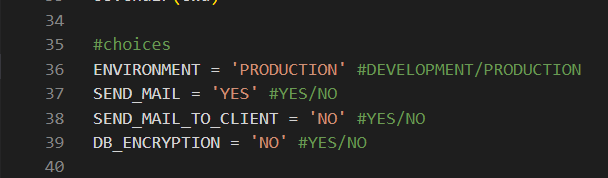
**Settings to switch between “Development” and “Production” environment.**

Include screenshot of relevant sections from code



**Settings to control e-mail delivery to test recipient versus actual recipient.**

Include screenshots from relevant sections from code



Thank you for considering ImmiLytics for the proposal. For any questions, please feel free to contact Shiv CS at [shiv@immilytics.com](about:blank).