



Informe Final de Práctica Empresarial TIGO-UNE

Juan David Sánchez Jaramillo

Informe de práctica para optar al título de Ingeniero de Sistemas

Tutor

Deisy Loaiza Berrío, Ingeniera de Sistemas

Universidad de Antioquia
Facultad de Ingeniería
Ingeniería de Sistemas
Medellín, Antioquia, Colombia
2021

Cita	(Sánchez, 2021)
Referencia	Sánchez Jaramillo, J. (2021). <i>Informe Final de Práctica Empresarial TIGO-UNE</i> [Trabajo de grado profesional]. Universidad de Antioquia, Medellín, Colombia.
Estilo APA 7 (2020)	



Centro de Documentación Ingeniería (CENDOI)

Repositorio Institucional: <http://bibliotecadigital.udea.edu.co>

Universidad de Antioquia - www.udea.edu.co

Rector: John Jairo Arboleda Céspedes.

Decano/Director: Jesús Francisco Vargas Bonilla.

Jefe departamento: Diego José Luis Botía Valderrama.

El contenido de esta obra corresponde al derecho de expresión de los autores y no compromete el pensamiento institucional de la Universidad de Antioquia ni desata su responsabilidad frente a terceros. Los autores asumen la responsabilidad por los derechos de autor y conexos.

Dedicatoria

A Dios, por haber cuidado siempre de mi familia y por permitirme cumplir mis objetivos e iluminar siempre mi camino. A mis padres Rosa y Horacio, por haber sido mi apoyo incondicional, por siempre darme una razón para esforzarme, porque cuanto soy se los debo y siempre se los deberé a ustedes, los adoro con cada fibra de mi ser.

Tabla de contenido

Resumen	7
Abstract	8
Introducción	9
1 Objetivos	10
1.1 Objetivo general	10
1.2 Objetivos específicos.....	10
2 Marco teórico	11
3 Metodología	13
4 Resultados	14
5 Conclusiones	18
Referencias	20

Lista de figuras

Figura 1	Interfaz gráfica principal del proyecto.....	14
Figura 2	Interfaz gráfica del proceso Centro de Mando Integrado	15
Figura 3	Interfaz gráfica del proceso Informe de Seguimiento Comercial	16
Figura 4	Interfaz gráfica del proceso Informes Sharepoint.....	16
Figura 5	Interfaz gráfica del proceso Activos de Salesforce.....	17
Figura 6	Interfaz gráfica del proceso Informe Cuentas con Saldo.....	18

Siglas, acrónimos y abreviaturas

ETL	Extract, transform and load
SSIS	SQL Server Integration Services
API	Application programming interface
GUI	Graphic user interface
SQL	Structured query language
BI	Business intelligence
IBM	International Business Machines Corporation
ODBC	Open database connectivity

Resumen

En los procesos que se tenían asignados dentro de la compañía se identificó la posibilidad de generar un grado de automatización que contribuyera a la liberación de tiempo para realizar otras actividades asignadas en la empresa, y disminuir la probabilidad de errores en la ejecución de los mismos. Así, se desarrolló un aplicativo gráfico el cual tiene la posibilidad de agregar o remover módulos, y que en el momento reúne 5 módulos relacionados con actividades de extracción, transformación y carga de información (ETLs), estos módulos además son capaces de funcionar independientemente. Se consiguió mediante este aplicativo una reducción en tiempos de ejecución que rondaba entre 3-4 horas, y que actualmente tiene una duración de aproximadamente 30 minutos. El aplicativo fue desarrollado en el lenguaje de programación Python, el cual es uno de los aceptados dentro de la compañía, incluyó en su última versión previa a la finalización de las prácticas una lista de librerías para manejo del sistema operativo, conexión a bases de datos, manipulación de documentos principalmente de extensión .xlsx, creación de gráficos, automatización en la navegación de páginas web y creación de GUIs. Además, se realizó un manual para el uso del aplicativo, en el cual se detallan las librerías que deben ser instaladas, los pasos a seguir para la ejecución de cada proceso, y demás detalles de la solución desarrollada.

Palabras clave: Práctica académica, Automatización de procesos, ETL, Python

Abstract

In the processes that were assigned within the company, the possibility of generating a degree of automation that would contribute to freeing up time to carry out other activities assigned in the company and reduce the probability of errors in their execution was identified. Thus, a graphical application which has the possibility of adding or removing modules was developed, and which currently gathers 5 modules related to information extraction, transformation and loading activities (ETLs), these modules are also capable of operating independently. Through this application, a reduction in execution times was achieved, which was around 3-4 hours, and currently lasts approximately 30 minutes. The application was developed in the Python programming language, which is one of those accepted within the company, it included a list of libraries to manage the operating system, connection to databases, manipulation of documents mainly with .xlsx extension, creation of graphics, automation in the navigation of web pages and creation of GUIs in its latest version prior to finishing off the practices. In addition, a manual was made for the use of the application, in which the libraries that must be installed, the steps to follow for the execution of each process, and other details of the developed solution are detailed.

Keywords: Academic Practice, Process Automation, ETL, Python

Introducción

La organización invertía un tiempo considerable en procesos manuales o semiautomáticos que exigían que quien ejecutara dichos procesos no pudiera realizar otras acciones, dado que exigían en casi todo momento estar pendiente de poder ejecutar un siguiente paso dentro del proceso. Los procesos en cuestión habían sido heredados entre practicantes y tenían una curva de aprendizaje moderada para lograr ejecutarlos correctamente. Estos procesos, que debían ser ejecutados en su mayoría diariamente, al ser estudiados durante algún tiempo, se consideró que era posible reducir el tiempo que se les dedicaba, y además disminuir la probabilidad de que se presenten errores debido a la interacción humana. Además del tiempo de ejecución de los mismos se evidenció que se podían solucionar otros inconvenientes como era el no cambio de variables al actualizar cierto tipo de datos, dejar credenciales de conexiones en archivos o, lo que es peor aún, la pérdida de conocimiento o dificultad para recuperar la correcta ejecución de ciertos procesos, debido a su poca o nula documentación y a la ausencia de un practicante anterior con conocimientos al respecto, lo cual además hacía que estos procesos tuvieran un flujo difícil de corroborar al momento de buscar si hubo algún tipo de error. Así, los procedimientos que se lograron automatizar, total o de manera casi total ayudaron a solventar la mayoría de estos inconvenientes. Los procesos en su mayoría estaban relacionados con extraer información de bases de datos o ciertas páginas de la organización, transformarla para cumplir con las necesidades de otras áreas y cargarla de nuevo a una base de datos o llevar a cabo ciertas funciones con la misma, esto es conocido como ETL por sus siglas en inglés. Para el momento de finalización de la práctica, solo dos procesos seguían estando compuestos por ETLs desarrolladas en el software de Microsoft para integración de servicios, esto debido principalmente a que dependían mayormente de actualizaciones que hacía uno de los miembros con más experiencia en el área dentro de la organización, por lo cual en caso de realizar modificaciones a bases de datos, facilitaba el proceso de adaptarlas de nuevo, igualmente, aquéllos procesos que presentaban limitaciones al ser ejecutados en este software fueron implementados en su totalidad en Python, lo cual ayudó a sortear las limitaciones del software de Microsoft. El aplicativo finalmente permitió que sin salir del mismo se pueda llevar a cabo algunos procesos con solo presionar uno o más botones, o en caso tal, se llevó a cabo la implementación para que un botón abra la ETL en el software de Microsoft para poder realizar parte de la ejecución y el resto del proceso se realiza desde el aplicativo.

1 Objetivos

1.1 Objetivo general

Automatizar la ejecución de procesos internos de la organización relacionados con ETLs.

1.2 Objetivos específicos

- Definir el alcance de las soluciones, en búsqueda de una configuración que posibilite tanto un manejo de la misma sin mucha complejidad, como la posible futura extensibilidad que pueda requerir en caso de buscar añadir más procesos.
- Identificar las herramientas/librerías que faciliten el desarrollo y que cumplan además con los lineamientos de la organización.
- Implementar las diferentes soluciones a cada uno de los procesos, priorizando los que actualmente implican más interacción humana.
- Implementar una interfaz que agrupe las diferentes soluciones, de forma que permita la ejecución de todos los procesos.
- Verificar el correcto funcionamiento tanto individualmente como en conjunto de cada una de las soluciones implementadas.

2 Marco teórico

El concepto de ETL, introducido en 1970 se define como un proceso de integración de datos compuesto por tres pasos, los cuales son extraer, transformar y cargar datos de una o múltiples fuentes dentro de un solo almacén o base de datos con la finalidad de que la misma pueda ser analizada o utilizada para procesos relacionados con la inteligencia del negocio o BI por sus siglas en inglés (International Business Machines Corporation [IBM], s.f.).

En la actualidad, existen múltiples herramientas desarrolladas en diferentes lenguajes, algunas inclusive con una interfaz gráfica muy desarrollada que facilita el diseño de ETLs, una de las más conocidas en la industria es SQL Server Integration Services (SSIS) la cual es la plataforma para la construcción de soluciones de integración de información con alto rendimiento de Microsoft. SSIS entre otros servicios ofrece facilidad en el diseño de ETLs mediante su interfaz gráfica, herramientas para ejecutar instrucciones SQL, carga de archivos, envío de mensajes, entre otros (Microsoft Corporation [Microsoft], s.f.).

El problema principal que presentaba SSIS eran las limitaciones del software para la ejecución de algunas instrucciones SQL, especialmente cuando las mismas son de un tamaño considerablemente grande, además de la imposibilidad de ejecutar los diferentes paquetes con ETLs desde un mismo aplicativo/algoritmo, dado que el ecosistema que tiene Microsoft con esta plataforma se podría decir es bastante cerrado y no existen protocolos definidos para acceder a esta plataforma mediante lenguajes de programación y las herramientas que brinda Microsoft para la plataforma es sumamente limitada aún en procesos simples como es el cambio de variables al momento de la ejecución de una ETL. Así, se encontró que era necesario buscar una nueva alternativa capaz de sortear las limitaciones que se presentaban para ciertos procesos.

Debido al ecosistema interno de la compañía, y a ciertas restricciones existentes, un desarrollo en el lenguaje de programación Python sería posible. Python es un lenguaje de programación que permite trabajar rápidamente e integrar sistemas de un modo más efectivo (Python Software Foundation [Python], s.f.). Además, sería una elección ideal dado que dicho lenguaje está siendo adoptado rápidamente en la compañía por su extenso uso en la ciencia de

datos, su fácil curva de aprendizaje, y dado que posee una gran cantidad de librerías que facilitan muchos de los procesos que constantemente se deben realizar, además del rápido desarrollo de aplicativos.

Entre las librerías que se utilizaron en Python para llevar a cabo la automatización de estos procesos se encuentra Selenium, la cual se define como una herramienta para automatización de aplicaciones web con propósito de realización de pruebas, la cual no se limita solo a esto sino que puede y debería ser utilizada también para la automatización de tareas “monótonas” (Selenium Project [Selenium], s.f.). También se integró la librería Pywin32, la cual provee acceso a muchas de las APIs de Windows (Hammond, s.f.), permitiendo automatizar el uso de aplicaciones como Excel. También se integró Pyodbc, el cual es un módulo de código abierto que simplifica el acceso a bases de datos haciendo uso del estándar ODBC [Kleehammer, s.f.]. También se añadió al desarrollo la librería Plotly, la cual tiene como función principal la realización de gráficos para facilitar la visualización de datos (Plotly Python Open Source Graphing Library [Plotly], s.f.). Finalmente, la librería Tkinter definida como una parte integral de Python que ha proveído un set de herramientas para aplicaciones gráficas robustas y de plataforma independiente y que además presenta facilidad para realizar modificaciones y se encuentra ampliamente documentada (Python Tkinter [Tkinter], s.f.) fue la seleccionada para la creación de la interfaz gráfica.

3 Metodología

El proyecto fue dividido inicialmente en tres etapas, las cuales en orden eran etapa de análisis y diseño, etapa de desarrollo y etapa de pruebas. Debido a cambios sobre la necesidad de continuar algunos de los procesos inicialmente definidos, y la adición de otros procesos más durante la práctica, se llevó a cabo las tres etapas definidas inicialmente, pero con cada proceso de manera individual, en lugar de realizar cada fase con todos los procesos en simultáneo.

En la etapa de análisis y diseño se definió requisitos y limitaciones para el proceso que se estuviera automatizando en cada momento, se realizó una búsqueda de las librerías que podrían intervenir en diferentes partes del proceso y se seleccionó las más adecuadas, además se definió también la estructura de directorios, forma en que se realizaría la comunicación entre scripts, entre otros.

Durante la etapa de desarrolló se llevó a cabo la implementación de la solución definida en la etapa de análisis, además de desarrollar cada solución, se agregó código en un script que se definió como solución para unir los diferentes módulos que iban a irse desarrollando. Para cada proceso hubo una etapa en la que se hicieron pruebas, consiguiendo así solucionar algunos de los errores que se generaban al realizar la ejecución de los módulos.

Finalmente, se realizaron cambios a la estructura como se había definido el módulo diseñado para unir las diferentes soluciones, esto en búsqueda de facilitar su uso y realizar un aislamiento de responsabilidades para cada solución.

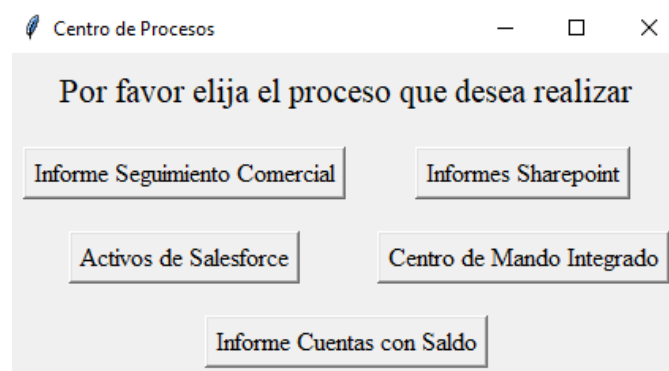
4 Resultados

Se entregó una solución que automatiza en gran medida los procesos que se recibieron al inicio de la práctica, y que permite, tras un cambio de parámetros ser adaptada a un nuevo practicante, la cual además, consigue llevar procesos que en algunos casos podían tardar cada uno alrededor de 30-40 minutos, a que estos se realicen casi que sin supervisión y en un tiempo inferior a 5 minutos en su mayoría. Además, se realizó entrega de un manual de usuario para la correcta utilización de la solución y se llevó a cabo reuniones virtuales con la persona encargada de reemplazar la ejecución de los procesos para facilitar el uso del aplicativo.

La interfaz gráfica (**Figura 1**) del aplicativo cuenta, al momento de finalizar la práctica, con 5 módulos, 3 planteados desde el análisis y diseño, y 2 que se agregaron durante el desarrollo, esto debido a que en el transcurso del proyecto se removió dos de los procesos que se habían planteado inicialmente, y se agregó otros dos procesos nuevos.

Figura 1

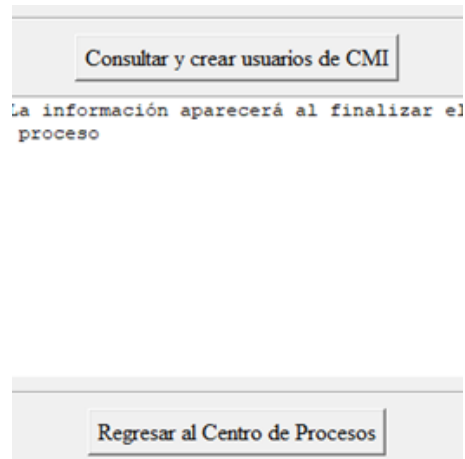
Interfaz gráfica principal del proyecto



Cada proceso presenta una interfaz de usuario diferente dependiendo de las entradas y salidas que tiene el mismo, así, mientras que para un proceso se cuenta con interfaz gráfica en la que solo se debe presionar un botón (figura 2) y esperar a que el mismo finalice para devolver información al respecto, para otro proceso es posible que sea necesario el ingreso de datos específicos como fechas a las que se debe ejecutar un proceso (figura 3).

Figura 2

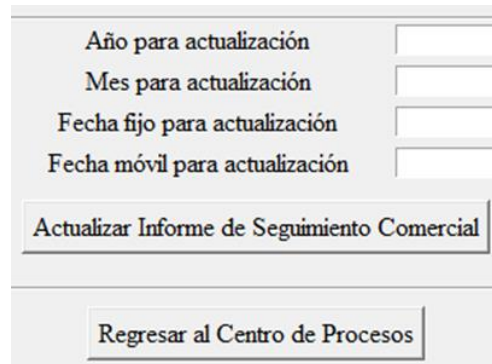
Interfaz gráfica del proceso Centro de Mando Integrado



El proceso de Centro de Mando Integrado cuenta con un botón y un cuadro de texto. Al presionar el botón se ejecuta un módulo encargado de ingresar a través del navegador con ayuda de la librería Selenium donde se encuentran las diferentes órdenes de cambio que hay montadas en el sistema, se busca a través de los campos de cada orden de cambio y se selecciona las que corresponden a creación de usuarios, en caso de coincidir, se guarda la información relacionada al número de la orden de cambio y se procede a realizar una búsqueda de la misma para poder extraer datos relacionados al usuario que se debe crear, como nombre de usuario, correo, entre otros. Posterior a tener toda la información recolectada correspondiente a los usuarios se procede a enviar la información a otro módulo para la creación de los mismos. En este caso se ejecuta una nueva instancia de navegador y se ingresa a un apartado dispuesto para la creación de los usuarios, se busca cada usuario, y en caso de no estar creado en el sistema se registra como un nuevo usuario teniendo en cuenta todos los datos ya recolectados en el módulo anterior. Finalmente cuando se procesa el último usuario, aparece un aviso informando de que se finalizó el proceso, y en el cuadro de texto aparece la información correspondiente a los usuarios que ya existían, a los usuarios nuevos, y la totalidad de los correos procesados para poder informar sobre la creación de los usuarios.

Figura 3

Interfaz gráfica del proceso Informe Seguimiento Comercial

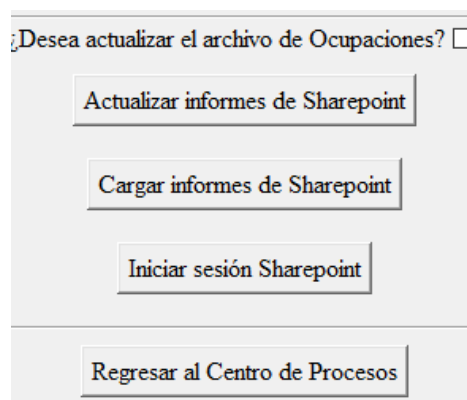


Formulario de actualización del Informe Seguimiento Comercial. Incluye cuatro campos de texto para ingresar la fecha de actualización (Año, Mes, Fecha fija, Fecha móvil) y dos botones: 'Actualizar Informe de Seguimiento Comercial' y 'Regresar al Centro de Procesos'.

El proceso de Informe Seguimiento Comercial cuenta con 4 campos de texto y un botón. Para realizar este proceso es necesario conocer la fecha a la que están actualizados los datos de pedidos tanto del área fija como móvil, estos datos se ingresan en los campos de texto y se presiona el botón, esto realiza actualizaciones en ciertas tablas de las bases de datos que se tienen y posterior a esto se debe generar la actualización del archivo de seguimiento comercial, en este además de actualizar su contenido a la fecha indicada es necesario cambiar ciertos campos que son los requeridos por ciertos miembros de la compañía para que cada división reciba la información necesaria. Así, mientras que para ciertos grupos genera un archivo de hojas de cálculo con información detallada del mes en curso, para otras divisiones se genera una imagen con datos de interés para la misma, como puede ser la información comercial de la división fija. Finalmente todos los archivos generados se guardan en una carpeta y se envían a la división respectiva.

Figura 4

Interfaz gráfica del proceso Informes Sharepoint

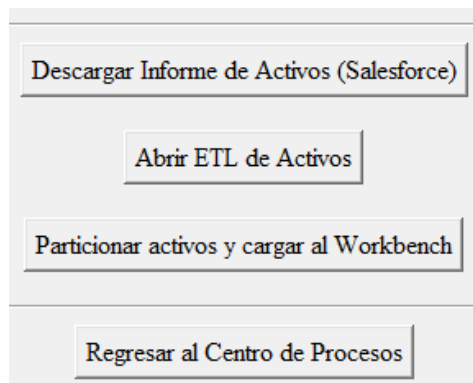


Formulario de actualización de informes de Sharepoint. Incluye un campo de texto con el texto '¿Desea actualizar el archivo de Ocupaciones?' y un botón de selección. Los botones disponibles son: 'Actualizar informes de Sharepoint', 'Cargar informes de Sharepoint', 'Iniciar sesión Sharepoint' y 'Regresar al Centro de Procesos'.

El proceso Informes Sharepoint cuenta con una casilla de verificación y tres botones. La casilla de verificación tiene la funcionalidad de excluir o no el procesamiento de un archivo correspondiente a las ocupaciones de la compañía, dado que el mismo cuenta con muchos datos y es bastante pesado para actualizar, solo se realiza la actualización del mismo una vez cada semana. El proceso comienza al presionar el botón de actualizar informes, el cual toma todos los informes, crea una copia de los mismos y procede a la actualización de los datos mediante consultas específicas para cada informe a las tablas de las bases de datos, cuando finaliza la actualización, y en caso de no haberse presentado un error se reemplazan los archivos anteriores y aparece un aviso en el programa informando de la finalización, posterior a esto se debe iniciar sesión en el navegador y guardar los datos del navegador dado que el inicio de sesión requiere una autenticación vía mensaje de texto al teléfono de quien inicia sesión, para esto el proceso es llevado a cabo mediante el uso de la librería Selenium, y solo es necesario el ingreso de los dígitos de verificación. Posterior a la autenticación, y si se presiona el botón de cargar informes, se ejecutan nuevas instancias para subir cada uno de los informes al repositorio de la compañía.

Figura 5

Interfaz gráfica del proceso Activos de Salesforce

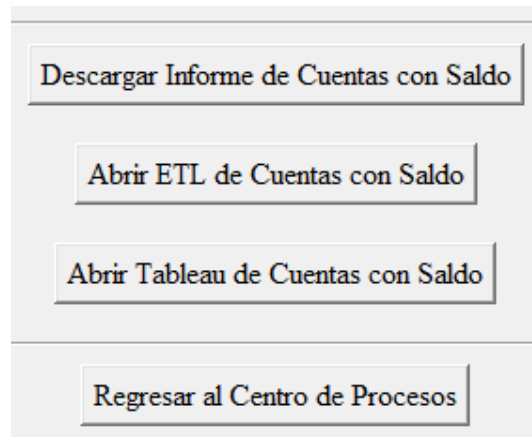


El proceso Activos de Salesforce cuenta con tres botones. El primer botón ejecuta una instancia del navegador e instrucciones para inicio de sesión, y búsqueda de un informe que debe ser actualizado y descargado con datos de ciertos activos de la compañía, el cual cuenta con datos que deben ser filtrados dependiendo de cuáles ya fueron procesados y cuáles aún están pendientes para su procesamiento. Para buscar los datos faltantes se hace uso de una ETL en SSIS la cual compara los datos con una tabla en la que se tienen registros y en caso de no encontrar coincidencias agrega ciertos campos a un archivo para que posteriormente puedan ser cargados al sistema. Dado que el sistema dependiendo de la cantidad de datos cargados genera errores debido a que la capacidad de procesamiento es superada, los archivos generados deben ser particionados en piezas que tengan una cantidad menor de registros, y posteriormente se cargan los archivos nuevamente

con una serie de instrucciones en el navegador, esta última parte del proceso es llevada a cabo al presionar el botón de particionar activos y cargar al Workbench.

Figura 6

Interfaz gráfica del proceso Informe Cuentas con Saldo



El proceso Informe Cuentas con Saldo cuenta con tres botones. El primer botón ejecuta una instancia del navegador e instrucciones para inicio de sesión, y genera mediante la selección, remoción y adición de campos una serie de informes que sirven como insumo para generar un informe final de cuentas faltantes de pago. El segundo botón abre una ETL en SSIS que mediante diferentes instrucciones realiza una actualización de datos referentes a cuentas pendientes de pago en las tablas de las bases de datos y realiza respaldos de la información en múltiples bases de datos, además de borrar información obsoleta. Finalmente, con el tercer botón se abre un archivo de Tableau en el cual se actualizan las vistas de los datos en los cuales se pueden ver porcentajes y cantidades de cuentas que están al día o que tienen múltiples pendientes de pago.

5 Conclusiones

- Se cumplió con el objetivo principal del proyecto al haberse creado una solución que automatiza en un alto grado la ejecución de los procesos internos de la organización que guardaban relación con ETLs.
- Se definió el alcance de la solución y se realizó una configuración modular que permite la futura adición de procesos en caso de que en algún momento sea necesario.
- Se hizo uso de librerías, mayormente de código abierto, y ninguna con licenciamientos que exijan algún tipo de retribución monetaria. Además, las librerías seleccionadas tienen en su mayoría buena acogida por parte de la comunidad y reciben mucho soporte, o están respaldadas por grandes compañías.
- Se implementó cada solución, inicialmente, como módulos aislados y se priorizó los que se debían ejecutar de manera más constante y que exigieran mucha interacción humana.
- Se realizó una interfaz gráfica mediante la cual se agruparon los diferentes módulos, desde esta interfaz es posible la ejecución de cada uno de los procesos de manera fácil y conveniente.
- Se verificó que las soluciones, en condiciones normales, funcionan de la manera esperada, tanto individualmente como en conjunto desde la interfaz que las agrupa.

Referencias

- Hammond, M. (s.f.). *Descripción del proyecto PyWin32 [PyWin32 project description]*. <https://bit.ly/3lkF5ZA>
- International Business Machines Corporation [IBM]. (s.f.). *¿Qué es ETL? [What is ETL?]*. <https://ibm.co/2Yp0gAH>
- Kleehammer, M. (s.f.). *Descripción del proyecto PyODBC [PyODBC project description]*. <https://bit.ly/3ozJsSu>
- Microsoft Corporation [Microsoft]. (s.f.). *¿Qué son servicios de Integración SQL Server [What is SQL Server Integration Services]*. <https://bit.ly/3Aj62ku>
- Python Software Foundation [Python]. (s.f.). *Aprende más sobre Python [Learn more about Python]*. <https://www.python.org/>
- Plotly Python Open Source Graphing Library [Plotly]. (s.f.). *Plotly Python librería de gráficos de código abierto [Plotly Python open source graphing library]*. <https://bit.ly/3AdTYkB>
- Python Tkinter [Tkinter]. (s.f.). *Interfaces gráficas de usuario con Tk [Graphical user interfaces with Tk]*. <https://bit.ly/3iYEoDJ>
- Selenium Project [Selenium]. (s.f.). *Acerca de Selenium [About Selenium]*. <https://bit.ly/3myQuUK>