

## AGRADECIMIENTOS

Aprovecho totalmente esta oportunidad para así expresar mis más sinceros agradecimientos las personas y organizaciones que aportaron en gran medida a mi desarrollo profesional como estudiante y poder demostrar lo que he aprendido en la institución.

Primeramente, quiero agradecer a mis padres y mis hermanas por siempre confiar en mi y nunca dudar de mis capacidades, que me dieron aquellas palabras de aliento cuando quería dejarlo todo, por su apoyo incondicional durante toda mi vida y estancia en la institución. GRACIAS enormemente por ser los pioneros de llevarme hasta este punto de mi vida profesional, que siempre fueron mi motivo clave para recorrer este camino.

Agradezco a la institución, Instituto Tecnológico de Nuevo Laredo, por permitirme esta oportunidad de crecer profesionalmente y brindarme todas las herramientas para mi desarrollo profesional, por indicarnos este requisito que nos ayudara a potencializar mis habilidades obtenidas en mi estancia en la carrera. De igual manera, agradezco de corazón a TODOS mis docentes de la carrera de Ingeniería en Sistemas Computacionales, por siempre brindarnos sus experiencias, temas y consejos para ser alguien mejor en la vida y profesionalmente.

Agradezco a TECNOLOGISTICA ADUANAL S.A. DE C.V., por confiar en mí y brindarme un espacio más en su Trayectoria, de unirme a su equipo lleno de personas altamente capacitadas y por darme la oportunidad de demostrar y cumplir mis responsabilidades requeridas para mi carrera profesional.

Finalmente, quiero agradecer a Todos mis amigos de dentro y fuera de la carrera, que siempre me demostraron su gran apoyo, que siempre ha sido de los más importantes en mi vida como estudiante, por siempre estar cuando más lo necesitaba, por motivarme a nunca dejar de escalar, por su gran comprensión durante todo este proceso, siempre los llevare conmigo.

GRACIAS

## RESUMEN

Este documento describe y demuestra, el desarrollo completo y los resultados del proyecto asignado “EXTRACCION DE DATOS EN FACTURAS PDF PARA LA CREACION DE LAYOUT”, que ayudara al personal de Tecnológica Aduanal S.A de C.V para algunas de sus sucursales, acelerar la producción en el departamento de Trafico.

Este proyecto, desarrollado con la ayuda de tecnologías de Adobe, Visual Studio Code, Javascript, FireBird y Node JS para su funcionamiento total.

La problemática que se presenta, radicaba en la lentitud en la que los layouts eran creados en Excel manualmente, provocando asi retrasos, falta de tiempo y/o quejas por parte de los clientes/proveedores.

Destacando cual es la principal problemática en una de sus áreas, tomando así en cuenta las peticiones por parte del personal a quien va dirigido este proyecto. Dando así como objetivo principal, acelerar la creación de Layouts con datos importantes de Facturas (creadas manualmente en Excel) a ser generados automáticamente, llenados con los datos requeridos. Siendo así, se aplico una interfaz donde mediante una lista desplegable el operador puede seleccionar el proveedor al que se le extraerán los datos de su factura, una vez seleccionado, se le permite cargar el archivo PDF de la factura correspondiente, una vez cargada, el sistema se encargara de realizar la tarea esperada.

Se utilizo Postman para probar las peticiones del API que nos ayudara a extraer datos de un PDF, para así obtener un json, y así tengamos la posibilidad de recorrerlo y consultar ciertos datos como Números de parte, Fracción y Claves de proveedor, que nos darán la información necesaria para el Layout final.

Al realizar esto, se aplican algoritmos para consumir el API de manera automática y también para mostrar los datos necesarios. Tomando en cuenta que cada proveedor maneja diferentes formatos de Facturas, debido a esto, se aplicaron diferentes tipos de recorrido para el JSON, para mostrar correctamente la información de diferentes facturas.

Al realizar pruebas, concluyendo cada una de las etapas del desarrollo, se instancio mediante el servicio de IIS para la prueba y error, y corrección de problemas. Siendo así, obteniendo resultados rápidos y eficientes, acelerando el tiempo de ejecución de cada una de las operaciones en el departamento de Trafico.

## INDICE

### **Tabla de Contenido**

<b>1.</b>	<b>INTRODUCCION.....</b>	<b>6</b>
<b>2.</b>	<b>PROBLEMA A RESOLVER.....</b>	<b>7</b>
<b>2.1.</b>	<b>Situación Actual: .....</b>	<b>7</b>
<b>3.</b>	<b>OBJETIVOS.....</b>	<b>8</b>
<b>3.1.</b>	<b>Planeación:.....</b>	<b>8</b>
<b>3.2.</b>	<b>Objetivo General: .....</b>	<b>9</b>
<b>3.3.</b>	<b>Objetivos Específicos: .....</b>	<b>9</b>
<b>4.</b>	<b>JUSTIFICACION. ....</b>	<b>10</b>
<b>5.</b>	<b>METODOLOGIA DE GESTION DE PROYECTOS Y MARCO TEORICO.....</b>	<b>11</b>
<b>5.1.</b>	<b>Fundamentos Científicos. ....</b>	<b>11</b>
<b>5.1.1.1.</b>	<b>Generación de documentos.....</b>	<b>11</b>
<b>5.1.2.</b>	<b>Tecnologías de la Información y Comunicación (TIC). ....</b>	<b>12</b>
<b>5.2</b>	<b>Procedimiento de Cascada. ....</b>	<b>12</b>
<b>5.2.1.</b>	<b>Requisitos. ....</b>	<b>13</b>
<b>5.2.2.</b>	<b>Diseño. ....</b>	<b>13</b>
<b>5.2.3.</b>	<b>Implementación. ....</b>	<b>14</b>
<b>5.2.4.</b>	<b>Verificación. ....</b>	<b>14</b>
<b>5.2.5.</b>	<b>Mantenimiento.....</b>	<b>14</b>
<b>5.3.</b>	<b>Tecnologías utilizadas. ....</b>	<b>15</b>
<b>5.3.1.</b>	<b>HTML y CSS3: .....</b>	<b>15</b>
<b>5.3.2.</b>	<b>Lenguaje de Programación: Javascript.....</b>	<b>15</b>
<b>5.3.3.</b>	<b>Node JS y NPM: .....</b>	<b>16</b>
<b>5.3.4.</b>	<b>API de servicios de Adobe PDF:.....</b>	<b>17</b>
<b>5.3.5.</b>	<b>FireBird y DBeaver: .....</b>	<b>17</b>
<b>5.3.5.1.</b>	<b>SISTEMAS CASA (Uso de su base de datos): .....</b>	<b>18</b>
<b>5.3.6.</b>	<b>Pruebas hacia el API:.....</b>	<b>19</b>
<b>5.3.7.</b>	<b>IDE: Visual Studio Code. ....</b>	<b>19</b>
<b>5.4.</b>	<b>Metodología TDD:.....</b>	<b>20</b>
<b>5.5.</b>	<b>Modelo Cliente-Servidor: .....</b>	<b>21</b>
<b>5.6.</b>	<b>Automatización para Generar un Documento: .....</b>	<b>21</b>

<b>6. PROCEDIMIENTO DE CADA UNA DE LAS ACTIVIDADES PARA EL DESARROLLO DEL PROYECTO .....</b>	<b>21</b>
<b>6.1. 1ra ETAPA: REQUISITOS.....</b>	<b>21</b>
<b>6.1.1. Platicas con la empresa para recopilar y documentar los requisitos del sistema o software, para definir claramente lo que se realizara.....</b>	<b>21</b>
<b>6.1.2. Reunión para definir requisitos de hardware que se van a utilizar para el sistema del proyecto. ....</b>	<b>22</b>
<b>6.2. 2da Etapa: DISEÑO. ....</b>	<b>22</b>
<b>6.2.1. Diseño de la interfaz de la pantalla donde se cargara un archivo PDF mediante HTML:... </b>	<b>23</b>
<b>6.3. 3ra Etapa: IMPLEMENTACION. ....</b>	<b>24</b>
<b>6.3.1. Implementación de API de Adobe para extraer información de un PDF: .....</b>	<b>24</b>
<b>6.3.1.1. Funcionamiento de API en Postman.....</b>	<b>26</b>
<b>6.3.1.2. Funcionamiento del API Adobe Services en Javascript (Visual Studio Code) .....</b>	<b>32</b>
<b>6.3.2. RECORRIDO DEL JSON Y EXTRACCION DE SUS DATOS.....</b>	<b>38</b>
<b>6.3.2.1. Recorrido de Ejemplo. ....</b>	<b>39</b>
<b>6.3.2.1.1 Recorrido de un dato Fijo:.....</b>	<b>40</b>
<b>6.3.2.1.2. Recorrido Dinámico (extraer información de una columna):.....</b>	<b>41</b>
<b>6.3.2.1.3. Extracción de datos que serán consultados a la base de datos: .....</b>	<b>43</b>
<b>6.3.3. APLICACIÓN DE FORMATOS DE DIVERSAS FACTURAS.....</b>	<b>46</b>
<b>6.3.3.1. Consultando el JSON, las posiciones de los datos de cada proveedor.....</b>	<b>49</b>
<b>6.4. 4ta Etapa: VERIFICACION.....</b>	<b>51</b>
<b>6.4.2. Despliegue. ....</b>	<b>52</b>
<b>6.4.2.1. Descarga del Layout en Excel.....</b>	<b>54</b>
<b>6.5. 5ta Etapa: DESARROLLO.....</b>	<b>55</b>
<b>7. RESULTADOS.....</b>	<b>56</b>
<b>8. CONCLUSIONES Y RECOMENDACIONES.....</b>	<b>58</b>
<b>9. COMPETENCIAS APLICADAS Y/O DESARROLLADAS.....</b>	<b>59</b>
<b>10. FUENTES DE INFORMACION. ....</b>	<b>61</b>
<b>11. ANEXO.....</b>	<b>63</b>
<b>11.1. Glosario.....</b>	<b>63</b>

## 1. INTRODUCCION

En las empresas aduaneras, siempre cuentan con una constante aceleración, sobre todo en la zona fronteriza donde se encuentra, sus operaciones son constantes y eficientes, aunque tienden a requerir más estrategias y/o tecnologías para una mejor eficiencia. Sin embargo, hay áreas en las que sus procesos mayormente operan sin usar tecnologías que aceleren su proceso, provocando así algunos retrasos en algunas operaciones.

Siendo así, surge la necesidad de aplicar una automatización que ayude a abordar esta problemática. En el área de Trafico, donde la operación incluye realizar Layouts con información importante de diversas facturas, que incluyen: Numero de parte, Numero de fracción, descripción del producto, cantidad, fecha de factura, Numero de factura, Clave del proveedor, precios, entre otros.

Se ha decidido abordar el problema con la ayuda de un sistema que con la ayuda de Tecnologías como: Javascript, NodeJS, Firebird, HTML y CSS, nos brindaran las herramientas para que pueda ser aplicado y optimizar este proceso. Sin embargo, también se cuenta con la Tecnología de Adobe Services PDF, que consiste en un API impulsado con Inteligencia artificial que nos ayuda a extraer TODO el contenido de un PDF, aportando Texto, Imágenes, Tablas, etc.

Este sistema aportara una gran ayuda en cuestión de Tiempo, lo cual, ahorrara más tiempo de operación, es decir, las operaciones que se realizaban manualmente, se demoraban minutos, y una vez aplicada esta tecnología, se realizaran en Segundos. Siendo así, una operación más acelerada, evitando quejas por parte de los clientes y retrasos.

## 2. PROBLEMA A RESOLVER.

### 2.1. Situación Actual:

La empresa requiere acelerar el proceso de operación, lo cual, actualmente, en el área de operación, los encargados llenan layouts, con datos de la factura que ellos reciben en formato PDF. A base de ello, crean dicho layout en un archivo Excel, el cual, en una tabla con todos los datos de la factura, son rellenados manualmente, lo que provoca retrasos o quejas por parte del cliente.



Figura 1 Procesos Actuales

1. Ineficiencia por operaciones manuales.

- El trabajo de creación de archivos Excel, de aproximadamente entre 15 o 300 renglones, con información demasiado importante y minuciosa, es propenso a retrasos.

2. Ausencia de automatización.

- A falta de una automatización que se encargue de llenar un archivo de Excel, tiende a contener información errónea, causado por el “error humano”, teniendo así problemas al momento de exportar o presentar Layouts.

3. Falta de escalabilidad.

- La metodología realizada manualmente, no son suficientemente escalables para abordar tiempos de operación en la organización, limitándose a adaptarse a las demandas crecientes de gestión de personal.

4. Falta de Integración:

- La falta de integración entre diferentes sistemas y aplicaciones dificulta la fluidez de información y procesos, limitando la eficiencia operativa y la capacidad de respuesta a cambios y necesidades emergentes.

### 3. OBJETIVOS

#### 3.1. Planeación:

Se pretende que esta aplicación sea desarrollada como un sitio web para la propia empresa, la cual mostrara 3 componentes para interactuar:

- **INPUT:** Se encargará de subir un archivo en PDF.
- **Lista de proveedores:** Es una lista que servirá básicamente para seleccionar el proveedor de la factura a la que se le extraerán los datos.
- **Descargar Excel:** Que servirá para descargar el archivo Excel ya generado (después de cargar el PDF).

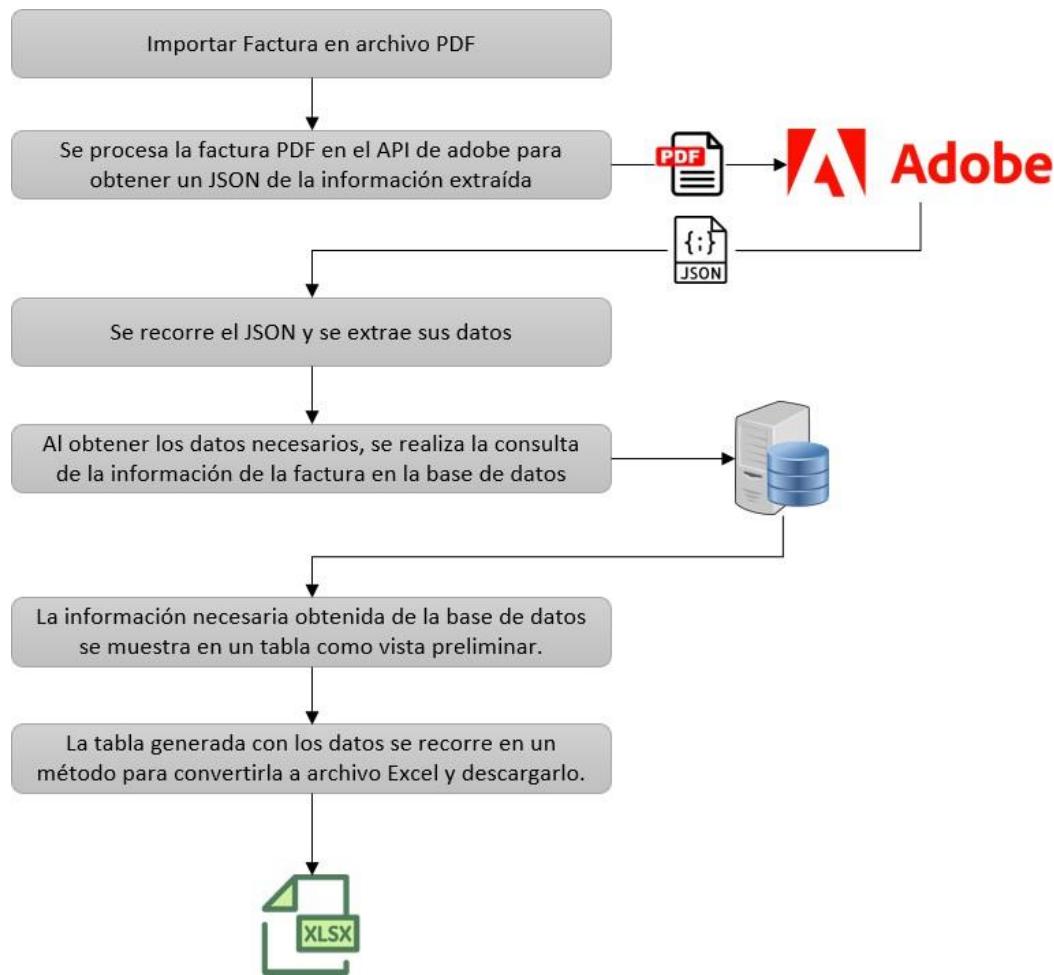


Figura 2 Diagrama de procesos planificados

### 3.2. Objetivo General:

Acelerar el proceso de captura manual, que dicha acción no se necesite realizar, desarrollando una aplicación que nos permita cargar solamente un archivo PDF de la factura y de manera automática se convierte a un layout en Excel, lo que se hacía manualmente, ahora se hace de manera automatizada, ahora se hará en segundos.

### 3.3. Objetivos Específicos:

- Conocer el proceso actual, cual es su problema principal, dificultades y procedimientos que ralentizan el trabajo de los ejecutivos.

- Analizar y aplicar las tecnologías bases para el desarrollo de este proyecto o sistema, entre ellos: Node JS para el desarrollo Backend, FireBird y DBeaver para el acceso a la base de datos y CSS3 para el Diseño Frontend. Ademas de contar con un servicio externo propio de Adobe Inc, que nos brindara el servicio de extracción de texto plano de archivos PDF
- Diseño del procedimiento Backend (Funcionamiento del API Adobe) que se encargara de extraer la información de un PDF de manera automatizada.
- Diseño del algoritmo que se encargara mostrar la información adecuada y requerida por los ejecutivos de la empresa.
- Aplicar diversos formatos para que el sistema acepte cualquier tipo de formato de Factura.
- Realización de pruebas por los ejecutivos y atentar a cualquier problema que podría surgir.
- Optimización constante del sistema, es vulnerable que con el paso del tiempo nuevos proveedores se presenten con un nuevo formato de Facturas y estas serán aplicadas al sistema.

#### **4. JUSTIFICACION.**

Ya debido al procedimiento actual de ejecución en el Departamento de Trafico, el proyecto denominado “Extracción de datos en Facturas PDF para la creación de Layout”, en la empresa Tecnológica Aduanal S.A de C.V. tiene como gran objetivo la gestión de Tiempos del proceso en dicho departamento.

Dicho sistema/proyecto tiene como puntos importantes a desarrollar:

- Reducir el tiempo de captura de datos.

Este sistema completamente automatizado, para que pase a ser hecho completamente manual a ser solamente hecho de manera automática y en segundos.

- Reducir la vulnerabilidad de quejas por parte de los clientes o proveedores.  
La generación de archivos Excel creados a base de una carga de una Factura en formato PDF de manera automática son una manera optima y fundamental para la comodidad de la empresa y el cliente.
  
- Mejora y optimización de procesos de la empresa.  
Este sistema es una manera de que la empresa pueda demostrar una buena capacidad de respuesta ante sus clientes, ya que adaptando las tecnologías y automatizaciones de procesos mejoran su capacidad y competitividad.

## 5. METODOLOGIA DE GESTION DE PROYECTOS Y MARCO TEORICO.

### 5.1. Fundamentos Científicos.

#### 5.1.1 Gestión de Documentos (DM).

La gestión de documentos (DM) es el proceso sistemático de capturar, rastrear y almacenar documentos electrónicos, incluidos PDFs, archivos de procesamiento de texto e imágenes digitales.

##### 5.1.1.1 Generación de documentos

La generación de documentos es la creación automatizada de documentos utilizando plantillas predefinidas y datos de entrada. Este proceso optimiza la producción de varios documentos, como contratos, facturas e informes, al extraer información relevante de bases de datos u otras fuentes y completar las plantillas con estos datos. La generación de documentos asegura consistencia, reduce el riesgo de errores y acelera significativamente el proceso de creación de documentos, convirtiéndola en una herramienta esencial para

las empresas que buscan mejorar la eficiencia y la precisión en sus flujos de trabajo de documentación.

### **5.1.2. Tecnologías de la Información y Comunicación (TIC).**

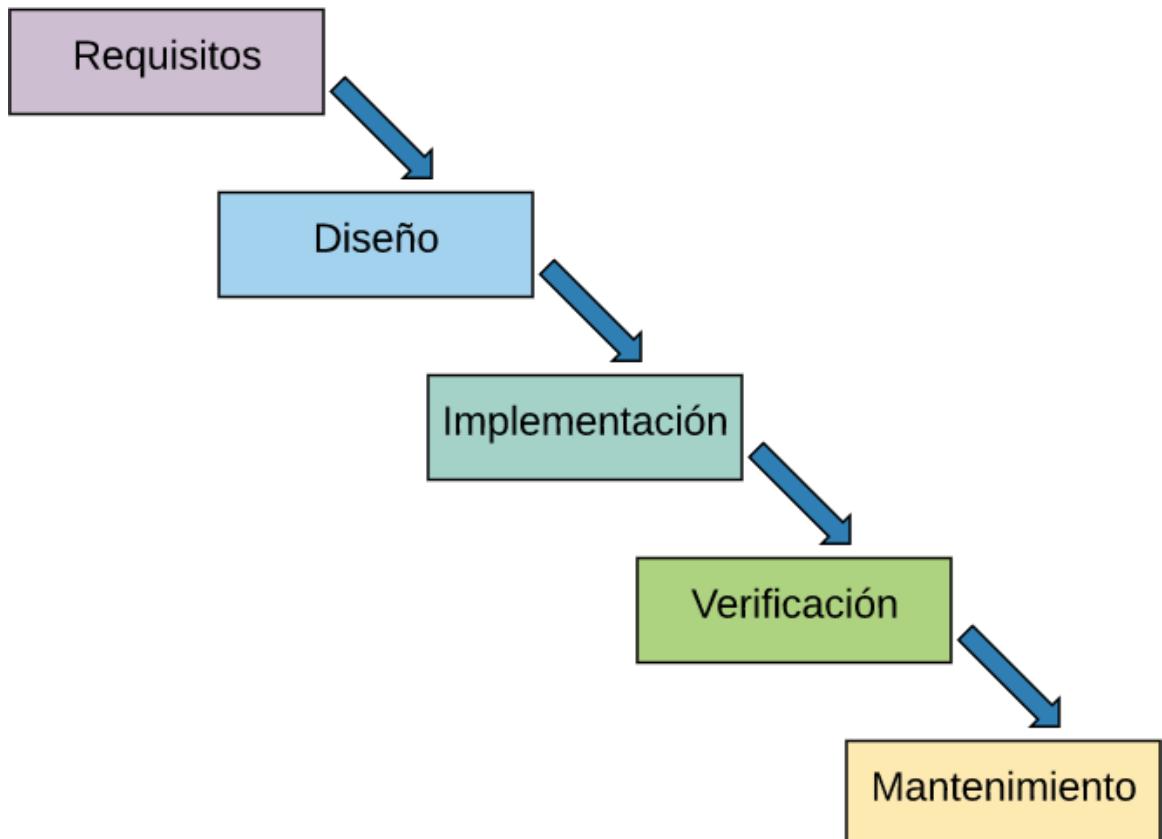
Las TICs son todas las herramientas tecnológicas que nos ayudan a agilizar diferentes tareas cotidianas, desde el uso de dispositivos móviles personales hasta la automatización de procesos empresariales con diversos softwares.

En el caso de las empresas, las TICs sirven para automatizar procesos administrativos, mejorar la comunicación y colaboración entre trabajadores, y sobre todo, para analizar grandes cantidades de datos y así tomar decisiones informadas en las organizaciones.

## **5.2 Procedimiento de Cascada.**

El procedimiento de Cascada o Waterfall (en Ingles), es una metodología para gestión de proyectos que se divide en distintas fases. Cada fase comienza una vez concluyendo la anterior.

Este enfoque para la gestión de este proyecto, nos ayudara de manera óptima la realización de cada etapa de este proyecto, respetando el orden de cada una.



*Figura 3Metodología de cascada*

### 5.2.1. Requisitos.

En esta primera etapa, se identifican los problemas que se presentan y que hay que resolver, se documenta el procedimiento que actualmente la empresa maneja, además, que es lo que específicamente requiere la empresa para resolverlo, en este caso acelerar el proceso que actualmente se realiza manualmente.

### 5.2.2. Diseño.

Creación de la interfaz mediante Lenguaje de Marcado de Hipertexto (HTML) y CSS3, que le permitirá al ejecutivo, Seleccionar proveedor al que desea extraer datos de su

factura, cargar la factura y una vez realizado el trabajo, descargar el Archivo Excel rellenado automáticamente.

### 5.2.3. Implementación.

Para esta etapa intermedia, se utilizaron tecnologías y servicios:

- Para el lado Frontend: Para este ámbito se utilizó HTML y CSS3, para su diseño y estilo de la interfaz.
- Backend: Para la automatización del proceso de layouts, se utilizó XAMPP con Javascript y NODE JS para el servidor que manejara las peticiones, AXIOS para realizar peticiones hacia un servicio externo y hacia la base de datos de la Empresa. En el ámbito de la base de datos, se utilizó FireBird como motor y DBeaver como el gestor, para principalmente realizar las consultas.
- Servicio Externo: Para un servicio externo o API, se utilizó el API de Adobe para extraer información de un PDF, que nos sirve de ayuda para manipular el texto que el mismo nos regresa en archivo JSON.

### 5.2.4. Verificación.

Esta etapa, se realizan pruebas para garantizar la funcionalidad y eficacia del sistema. Se realizan pruebas por parte de los ejecutivos para asegurar la eficacia y rapidez del proceso que anteriormente era manual. Cumpliendo así con las necesidades de la empresa.

### 5.2.5. Mantenimiento.

En esta etapa final es donde se hace corroboración del código del sistema, optimización de funciones e interfaz, para así instanciarlo en el servidor de operaciones de la Empresa. Sin embargo, el sistema estará en constante evolución, puesto que, durante

cierto tiempo, la empresa recibe proveedores o clientes nuevos, los cuales reciben nuevas Facturas que implica a aplicar un nuevo formato al sistema mediante código y así el mismo, pueda procesar cualquier tipo de Factura que la empresa tome en consideración y sea un sistema robusto.

### 5.3. Tecnologías utilizadas.

En este sistema, durante su desarrollo, se destaca el uso de algunas de las siguientes Tecnologías:

#### 5.3.1. HTML y CSS3:



Figura 4 Logotipo de HTML y CSS3

HTML es el lenguaje con el que se “escribe” la mayoría de las páginas web a las que accedemos diariamente desde Internet. Estas se desarrollan y después, los usuarios acceden a ella a través de un navegador, que las interpreta dicho lenguaje para mostrarlo de forma clara y ordenada.

CSS, es el lenguaje de marcado que determina el aspecto con el que se visualiza el contenido de un sitio web. Esto pasa por elementos como el color, estilos de letra o imágenes, etc. Su código se conjuga con el de HTML para crear páginas completas.

#### 5.3.2. Lenguaje de Programación: Javascript.



## JavaScript

*Figura 5 Logotipo de Javascript*

Javascript es un lenguaje de programación de alto nivel, interpretado y orientado a objetos, utilizado principalmente en desarrollo web, pensando para agregar potencial de interacción y dinamismo a las páginas web.

### 5.3.3. Node JS y NPM:



*Figura 6 Logotipo de Node JS*

Es una plataforma de código abierto que ha cambiado la manera en que los desarrolladores construyen aplicaciones del lado del servidor. Hace referencia a un entorno Javascript para la capa del servidor.



*Figura 7 Logotipo de NPM*

También se utilizó esta herramienta administradora de paquetes que nos sirve para utilizar dependencias en Javascript, o más bien, librerías para desarrollar este proyecto y nos ayudan a tener un desarrollo mas optimo y limpio.

#### **5.3.4. API de servicios de Adobe PDF:**



*Figura 8 Logotipo de Adobe*

La API de servicios PDF es una colección de múltiples servicios capaces de resolver rápidamente desafíos específicos y impulsando flujos de trabajo de documentos de varios pasos mediante SDK para Node.js, Java y .Net. Con él, ganamos acceso a servicios básicos de PDF, como crear, asegurar, comprimir, convertir, combinar y división de archivos PDF, así como servicios más avanzados, como la generación de documentos y la extracción de PDF.

#### **5.3.5. FireBird y DBeaver:**



Figura 9 Logotipo de Firebird Database Server

Para la realización de consultas a la base de datos se utilizo mediante Firebird, el cual es un sistema de gestión de base de datos relacionales SQL, gratuito y de código abierto.



Figura 10 Logotipo de DBeaver

Se utilizo esta herramienta de base de datos universal multiplataforma gratuita para desarrolladores, administradores de bases de datos, analistas y todos los que trabajan con datos. Es compatible con todas las bases de datos SQL populares como MySQL, MariaDB, PostgreSQL, SQLite, Apache Family entre otros.

#### 5.3.5.1. SISTEMAS CASA (Uso de su base de datos):

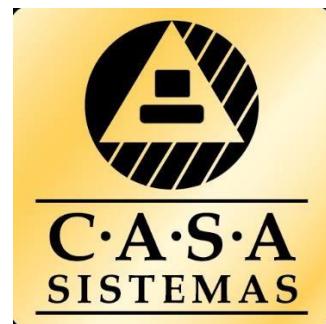


Figura 11 Logotipo de Sistemas CASA

**Propósito:** Brindar soluciones tecnológicas que mejoren la productividad de la operación aduanal y el comercio exterior de México y América Latina.

Ser el proveedor #1 de soluciones tecnológicas basadas en el modelo PaaS (conjunto de servicios basados en la nube que permite crear aplicaciones donde no hay necesidad de preocuparse por la configuración y el mantenimiento de servidores).

Sistemas casa tiene una base de datos muy robusta y completa, con gran numero de tablas. En este caso usaremos 2 de ellas que se etiquetan como **CTRAC\_FRACC** y **CTRAC\_FRACPAR**.

### 5.3.6. Pruebas hacia el API:



Figura 12 Logotipo de Postman

Sirve para múltiples tareas, tales como:

- Testear colecciones o catalogos de APIs, tanto frontend como backend.
- Organizar en carpetas, funcionalidades y módulos de los servicios web.
- Permite el ciclo de vida de nuestra API.
- Generar documentación de nuestras APIs.
- Trabajar con entornos.

Se utilizo esta confiable herramienta que ayudo en realizar pruebas de el Servicio de Adobe API, donde se realizaron peticiones hacia las diferentes rutas que esta maneja.

### 5.3.7. IDE: Visual Studio Code.

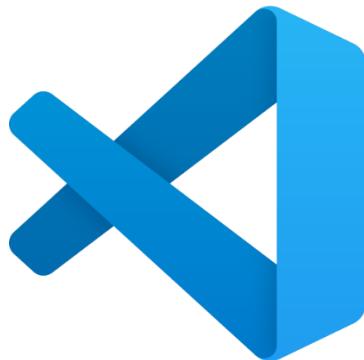


Figura 13 Logotipo de Visual Studio Code

Para el desarrollo de este proyecto se utilizo en su totalidad este IDE, y claro, utilizando unas extensiones que nos funcionan de guía para monitorear los componentes del sistema en desarrollo.

#### 5.4. Metodología TDD:

Es una metodología de desarrollo cuyo objetivo es crear primero las pruebas y luego escribir el software. Sus siglas en inglés son: Test Driven Development y en español significa: Desarrollo guiado por pruebas.

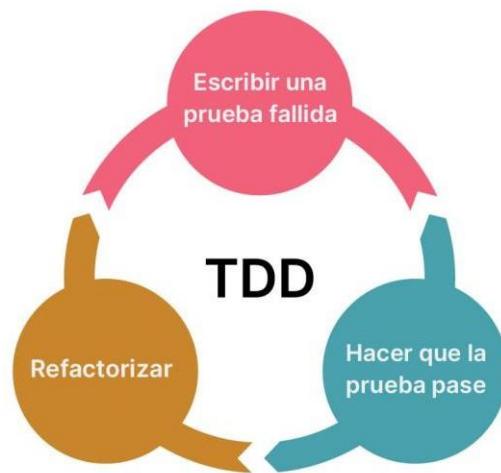


Figura 14 Representación gráfica de la metodología TDD

Esta metodología se utilizó con la finalidad de realizar pruebas al automatizar el funcionamiento del API de Adobe, Consulta a la base de datos y finalmente al crear el resultado final, asegurando su funcionamiento total, así después proceder a desarrollar el software.

### **5.5. Modelo Cliente-Servidor:**

El proyecto contiene el modelo Cliente-Servidor, ya debido a que se consume solamente un API, donde cada documento que se carga en este, el API se encarga de realizar de extraer su información para posteriormente proceder al resto de las funciones. Esto hace que el sistema sea un sitio más dinámico.

### **5.6. Automatización para Generar un Documento:**

Este proyecto contiene una función especial que generar un archivo Excel, el cual al momento de ser generado, este se descarga. Su contenido es básicamente la información que se extrajo del documento PDF y lo que se consulto desde la base de datos gracias a los datos que se consiguen del PDF, siendo este el Layout que los ejecutivos buscaban obtener.

## **6. PROCEDIMIENTO DE CADA UNA DE LAS ACTIVIDADES PARA EL DESARROLLO DEL PROYECTO.**

Como bien se incluyó en la Metodología De Gestión de Proyectos se utilizó la metodología de Cascada para Gestionar el desarrollo de cada etapa del proyecto, es por eso que es importante describir cada una de ellas, a continuación:

### **6.1. 1ra ETAPA: REQUISITOS.**

Semana 1: 5 de agosto

**6.1.1. Platicas con la empresa para recopilar y documentar los requisitos del sistema o software, para definir claramente lo que se realizará.**

En esta etapa o fase inicial, se realizan las pláticas con los ejecutivos de la empresa y asesor externo sobre lo que exactamente se necesita mejorar en el área en la cual se utilizará este proyecto, definiendo en sí, las dificultades que ellos tienen en su área.

Una vez definiendo la problemática, se realiza una visión clara y concreta sobre cómo se usará el sistema y que alcances tendrá. Además de solicitar un permiso especial a Sistemas CASA para tener un acceso exclusivo a su base de datos, que será necesario.

### **6.1.2. Reunión para definir requisitos de hardware que se van a utilizar para el sistema del proyecto.**

Aquí se realizó una plática con el Asesor externo de este proyecto, el hardware necesario para desarrollarlo, lo cual es lo siguiente:

- Equipo PC o Laptop: Windows 10 mínimo.
- Mínimo 8gb de RAM.
- Memoria de estado Solido (SSD) mínimo 256gb.

Y posteriormente se eligieron los programas o softwares que nos ayudaran a desarrollar este proyecto, los cuales consisten en:

- IDE para Desarrollo o codificación: VISUAL STUDIO CODE.
- Motor para la base de datos: Firebird y Su gestor DBeaver.
- Lenguaje de programación Javascript.
- Node JS para utilizar librerías o dependencias.
- Navegador EDGE o Chrome para inspeccionar y ver funcionar el sistema.

Al definir estos requisitos de software y hardware es muy importante, para obtener el alcance del proyecto, así mismo con el paso de las semanas, se consideran algunas tecnologías extras para la ejecución del proyecto.

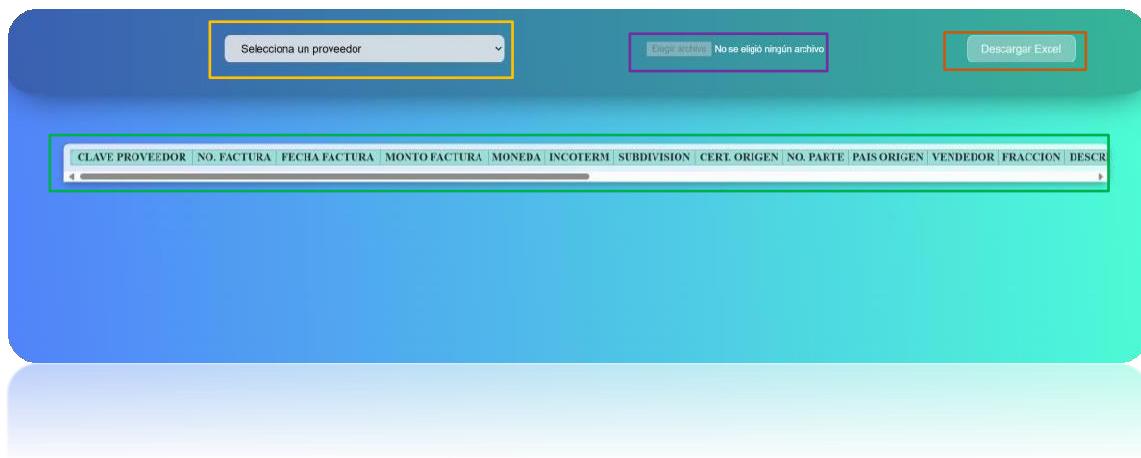
## **6.2. 2da Etapa: DISEÑO.**

Semana 2: 12 de agosto

### **6.2.1. Diseño de la interfaz de la pantalla donde se cargara un archivo PDF mediante HTML:**

Al recopilar toda la información concreta y necesaria, se realizó una interfaz sencilla y legible para que el usuario utilice sin problema, simulando el uso final y sus componentes los cuales los ejecutivos podrán usar e interactuar, para así darles instrucción sobre su uso futuro.

Diseñado con HTML y CSS3, implementando una interfaz simple y con colores vivos difuminados, luciendo de la siguiente manera.



- **Seleccionar un proveedor.**

Consiste en un <select> en html, que nos permite ver los proveedores que el sistema es capaz de extraer sus datos.

- **Elegir archivo.**

Es un elemento de html de tipo <input>, que se encarga de abrir nuestro directorio para seleccionar el archivo PDF que se va cargar.

- **Descargar Excel.**

Al presionar este botón, automáticamente se descargará el archivo Excel ya con los datos extraídos del PDF.

- **Tabla de vista preliminar.**

Esta tabla es la que mostrara al usuario de operación los datos que se trajeron del PDF, siendo así, al momento de descargar el Excel, se mostrara tal cual en la tabla preliminar.

## **6.3. 3ra Etapa: IMPLEMENTACION.**

A continuación, en esta etapa es donde se definirá el algoritmo que se encargará de hacer funcionar el proyecto de “Extracción de datos en facturas PDF para la creación de Layout”, donde se describirá cada una de las fases de desarrollo de funciones que el mismo tendrá.

### **6.3.1. Implementación de API de Adobe para extraer información de un PDF:**

Semana 3: 19 de agosto

Este microservicio impulsado con inteligencia artificial la cual nos permite extraer información de contenido PDF, entre ellos: Texto, imágenes, tablas, escaneados en un archivo JSON estructurado, para identificar y comprender el orden de lectura natural de diferentes elementos, como encabezados, listas y párrafos que abarcan varias columnas o páginas.

Esta herramienta nos ayudara a obtener de alguna manera una facilidad de detectar información en texto plano de una factura en PDF. A continuación, se adjunta el diagrama de peticiones para consumir el API de Adobe Services:

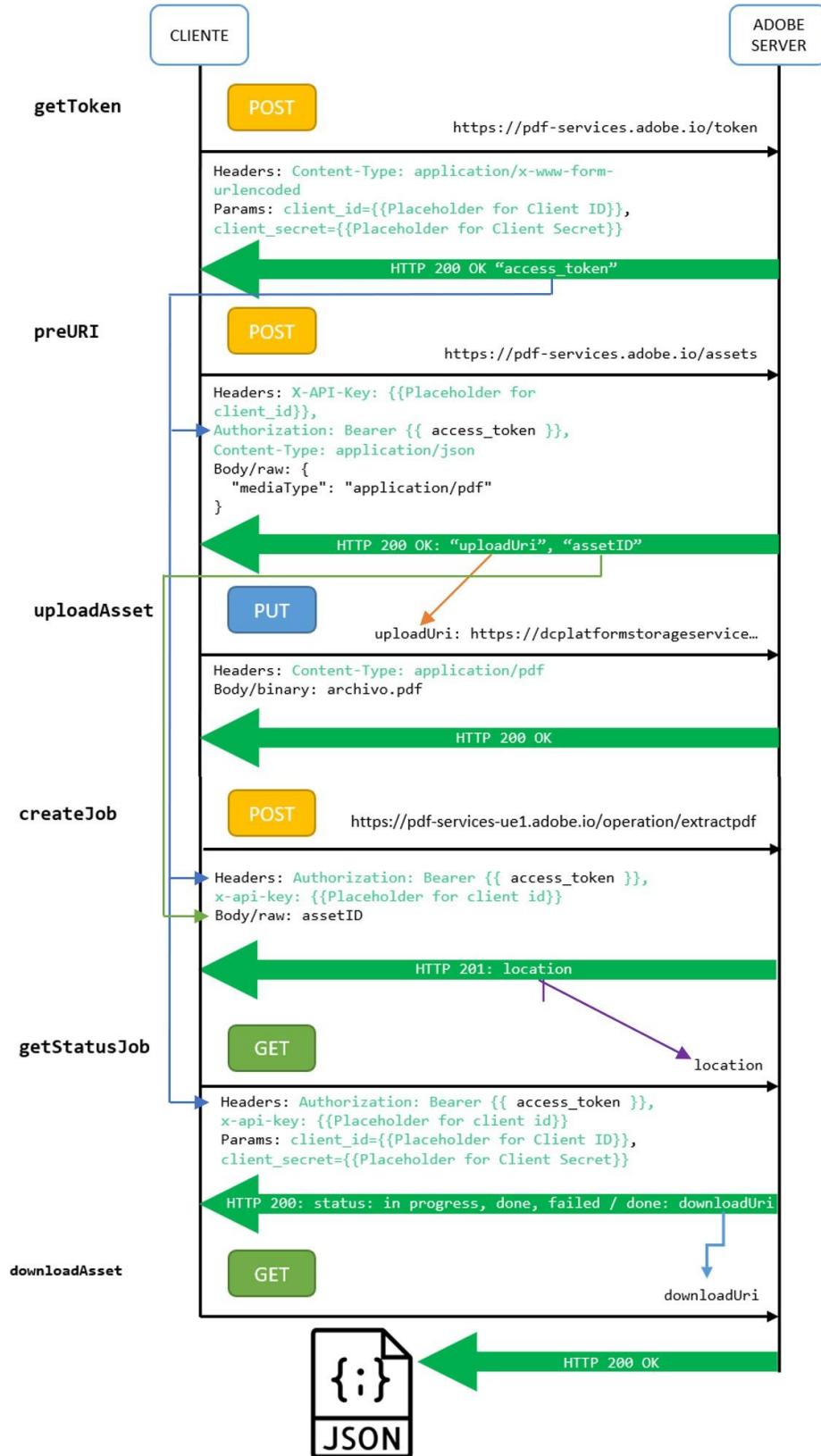


Figura 15 Diagrama de peticiones API de Adobe.

### 6.3.1.1. Funcionamiento de API en Postman.

Para realizar pruebas, se utilizó Postman, siendo así, se respeta con totalidad el flujo de las peticiones para lograr obtener el JSON que contendrá la información del JSON, a continuación, el listado de las rutas en Postman:



Figura 16 Nombres de Rutas De API en Postman

#### POST: getToken:

Para obtener el token de acceso, que es el que nos servirá para realizar peticiones hacia el servidor de Adobe, solo basta con hacer clic en “Send”:

A screenshot of the Postman request configuration for the 'getToken' endpoint. The method is set to 'POST' and the URL is 'https://pdf-services-ue1.adobe.io/token'. The 'Body' tab is selected, showing two form fields: 'client\_id' with value '7d835d6ff9e846d8bb98404609a40b5e' and 'client\_secret' with value 'p8e-L7Zy61b-sUmMBSzmD0I5YF0yErlt\_P...'. A red box highlights the 'Send' button at the top right of the body panel.

Figura 17 Endpoint para generar Token

Una vez realizada la petición, se obtiene como resultado un Json estructurado donde en el elemento “access\_token” se encuentra dicho token:



```

1 {
2   "access_token": "eyJhbGciOiJSUzI1NiIsIng1dSI6ImI1c19uYTe2V5LWF0LTEuY2VvIiwi2lkIjoiaw1zX25hMS1rZXktYXQtMSIsImI0dCI6ImF0In0...  

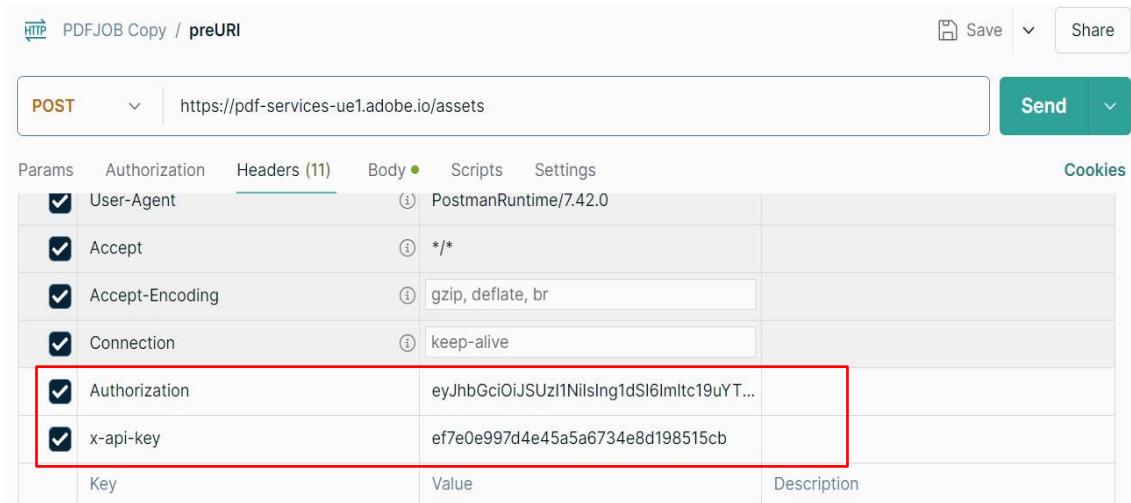
3   "token_type": "bearer",

```

Figura 18 Token de acceso generado

## POST: preURI.

Esta ruta nos dará una URL que servirá para subir el archivo pdf al servidor de Adobe, el cual, se necesita del token de acceso generado anteriormente y un api-key (que se obtiene del sitio web de adobe Developer).



Key	Value	Description
Authorization	eyJhbGciOiJSUzI1NiIsIng1dSI6ImI1c19uYTe2V5LWF0LTEuY2VvIiwi2lkIjoiaw1zX25hMS1rZXktYXQtMSIsImI0dCI6ImF0In0... eyJhbGciOiJSUzI1NiIsIng1dSI6ImI1c19uYTe2V5LWF0LTEuY2VvIiwi2lkIjoiaw1zX25hMS1rZXktYXQtMSIsImI0dCI6ImF0In0...	
x-api-key	ef7e0e997d4e45a5a6734e8d198515cb	

Figura 19 Endpoint para Obtener url de subida de archivo PDF

Una vez realizada la petición, se obtiene un JSON estructurado donde el elemento de la URL donde el archivo será subido esta en “uploadURI”, de igual manera se obtiene también “assetID”:

Body	Cookies	Headers (12)	Test Results
			200 OK · 712 ms · 2.38 KB ·  Save Response
Pretty	Raw	Preview	Visualize
JSON			
1			
2			
3			
4			

*Figura 20 Url para subir PDF al servidor de Adobe*

## PUT: uploadAsset

Esta ruta es la encargada para subir el archivo hacia el servidor de Adobe, así posteriormente, ser procesado.

En ella, el Endpoint, será la URL obtenida previamente, y en el cuerpo, se cargará un archivo pdf, que será la que se extraerá sus datos.

PUT https://dcplatformstorageservice-prod-us-east-1.s3-accelerate.amazonaws.com/7d835d6ff9e846d8bb98404609a40b5e\_5E773462660C2D240A495FDC%40techacct.adobe.com

Send

Params Authorization Headers (10) Body Scripts Settings Cookies

Key	Value	Description
Content-Type	application/pdf	
Content-Length	<calculated when request is sent>	
Host	<calculated when request is sent>	
User-Agent	PostmanRuntime/7.42.0	
Accept	*/*	
Accept-Encoding	gzip, deflate, br	
Connection	keep-alive	
Content-Type	application/pdf	

*Figura 21 Subida de PDF al Servidor de Adobe.*

Carga del archivo PDF en Body:

The screenshot shows a Postman interface with the following details:  
Method: PUT  
URL: https://dcplatformstorageservice-prod-us-east-1.s3-accelerate.amazonaws.com/uploadAsset  
Body tab selected  
File type: binary (FACTURACOMERCIAL 47558.pdf)  
The file name 'FACTURACOMERCIAL 47558.pdf' is highlighted with a red box.

Figura 22 Carga de archivo PDF para subirlo.

Al realizar la operación, se obtendrá una respuesta HTTP 200, lo cual no tendremos algún dato de regreso, ya que su función fue cargar el archivo, así permitiéndonos seguir crear el trabajo.

### POST: createJob.

En este endpoint, se va a crear el trabajo, lo cual se empezara a procesar el archivo pdf, donde se aplica la Inteligencia artificial de adobe que se encarga de extraer los datos del PDF.

Para ello se necesita como parte de los encabezados, el Token de acceso y assetID (generados previamente), y un Api Key (obtenida del sitio web de Adobe Devloper).

The screenshot shows a Postman interface with the following details:  
Method: POST  
URL: https://pdf-services-ue1.adobe.io/operation/extractpdf  
Headers tab selected  
List of headers:

- Content-Type: application/json
- Content-Length: <calculated when request is sent>
- Host: <calculated when request is sent>
- User-Agent: PostmanRuntime/7.42.0
- Accept: \*/\*
- Accept-Encoding: gzip, deflate, br
- Connection: keep-alive
- Authorization: eyJhbGciOiJSUzI1NiIsIng1dSI6ImItc19uYTEta2V5LWF0LTEuY2Vyl...
- x-api-key: ef7e0e997d4e45a5a6734e8d198515cb

The 'Authorization' and 'x-api-key' fields are highlighted with a red box.

Figura 23 Creación del trabajo.

```

1  {
2    "assetID": "urn:aaid:AS:UE1:2f6ab55d-dfe6-4c20-86bb-c314fbe48398",
3    "getCharBounds": false,
4    "includeStyling": false,
5    "elementsToExtract": [
6      ...
7    ]

```

Figura 24 Obtención de Asset ID

Una vez realizada la petición obtendremos como respuesta varios Headers, entre ellos “location” donde se encuentra la URL que necesitamos para obtener el estado del trabajo (a continuación).

Body	Cookies	Headers (13)	Test Results
			201 Created   1397 ms   611 B    Save Response
		Transfer-Encoding	chunked
		Connection	keep-alive
		x-request-id	aFRXXIMnbgMIZ1Fv7hEUKfijF8JNvhZ
		location	<a href="https://pdf-services-ue1.adobe.io/operation/extractpdf/gFRXXIMnbgMIZ1Fv7hEUKfijF8JNvhZ/status">https://pdf-services-ue1.adobe.io/operation/extractpdf/gFRXXIMnbgMIZ1Fv7hEUKfijF8JNvhZ/status</a>
		retry-after	1

Figura 25 Obtención de URL para obtener el estado del trabajo.

### GET: getStatusJob.

Este Endpoint es el encargado de obtener el estado del trabajo y ver su progreso. Para ello se necesita la URL de location (obtenida en createJob), el token de acceso (generado en getToken) y un Api Key (obtenida del sitio de Adobe Developer).

The screenshot shows the Postman interface with a red box highlighting the 'Authorization' and 'x-api-key' fields in the Headers section. The 'Authorization' field contains a long token starting with eyJhbGciOiJSUzI1Ni... and the 'x-api-key' field contains ef7e0e997d4e45a5a6734e8d198515cb.

Figura 26 Petición para Obtener el estado del trabajo.

Una vez realizada la petición se obtiene un JSON donde contiene como elemento “downloadURI”, donde podremos descargar el contenido generado por el API de adobe, entre ellos el JSON que será utilizado.

The screenshot shows the Postman response body with a red box highlighting the 'downloadURI' field. The value is a URL: https://dcplatformstorageprod-us-east-1.s3.amazonaws.com/7d835dff9e846dbb94846694a0b5e... . The entire JSON response is as follows:

```

5   "type": "application/json",
6   "size": 1228012
7 }
8 "downloadURI": "https://dcplatformstorageprod-us-east-1.s3.amazonaws.com/7d835dff9e846dbb94846694a0b5e... "
9 X-Amz-Security-Token=I0oJb3jZ21uX2VjEEAaCXVzIwNh30tMSJcMEQC1HHPVzhZNPD5fwRB1o2jy6EWijy8WAoEhJRDXEVA1BFyQN8WwYL=1y0ufm%2FGvqGGwtC1tBXnFuNPY9eA9y
ghBQjZs2F%2F2P%2F2P%2F2P%2F2P%2F8BEAAu0D0yMDuMz0NTk4M1iMx0pFnN1nP6v0KvEL%2B7rYd4eqZD7vUFzaubRkuP05VjAAxaXwm59gyT2H28atpFcgC9Np4HGUS%2FaWhNjY
xhAI1vb#6Ta2PIAa1m8o7g%2FQkmS4s0fUDBp%2Fb16GyET3FWf164NT71OH5asNP%2B1b0f7ysHt%2BBJLkwvo%wtg9z70j041Vm5Rev1957fV6e22a6gE42IG21Ek6e6iz5fTh3P2XbwYFhm
Tj8%2FgSYHB%2Fn69%2F0616a0oRnnTP%2B1UhssxkCY8Vgnf0f0fAVV128pEsgaTHt+j5x78HQo0fwLXXKK%2B44PHigJ14KHCPhgt9CE92x2z60158BqrXZ03zPmN3DfrHsUxpI0n8URK0609yq
CaRo1XNuToQ2YRjokew894zux15%2Bu6xwH9L71By711o8IvygHh6tVm196x0n2B2wt2u6Ny5h3B1v80aHp44D3V1Y%2BvzCofkwo1nToja5eBvkFChk1g1fx7naA2180Z51wB8Ffh1uIn
Hzhi50g%0o3L8k%20xB0yByzvauNEW1Ubzb7b1sh17%2BH5tw5Ke1gsJ2zr%2P%2Fnh%2Fe3%2F0mmEEB0xnp03AeYuVGITYL%2F0NDKY%2FPu985pK1q1m4xa198G%2B6ykoYk%2B
%2F1KUvK0b%2fRIdcJ39TLkf2xg0ykT7V1df8v11Ui0pV07%2BgJ5Rc2Fw623SN8x4T2pK1cpYXYB075EZQ7M%STV63CjC160L%2F2MXFnSVwsuhAh42zWq10614U11%2F133062M%29KTxz2975nx
Uzecpxtx%2F6RWpqo%0K1t8u3uayz3L41t0g%52F0%2BSn1wtL1RwWf66CjgzVa2PChugY6nAGLdagS9uSE1345cb9%228XLj0XuTzYQavnc0%2F0w88VMu5fCh5a6j2qj0%2FaHFn9o%2FmqAFx
QoCsJvBsnlLqB5Oe5IRySiyB1OFBahoCjSD0LwOyS34RPofXbpcpQNNhvgV1hvfvQv6Eh0r1%2Fk1wvEJNeALDevHcd426aVGKbVfAg16EoxzjEyvcVn1fJ0xq44ze2Fo9nqb77Wc%3D6
X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Date=20241123T171316Z&X-Amz-SignedHeaders=host&X-Amz-Expires=3600s
X-Amz-Credential=ASIAWd2NTEVPABM715Zk%2F20241123%2Fus-east-1%2F5%2Fam4_request&
X-Amz-Signature=823ge95d96734e963a6684e6be8c6e4e49e0bc2dea666ea2a8929c41fb4de0e",
10 "assetID": "urn:aaid:AS:UE1:e8025ce1-3a29-41fa-8ba9-2ff6511da010"
11 "resource": f

```

Figura 27 Obtención de URL para descargar el JSON.

## GET: downloadAsset.

Para finalizar el proceso de extracción, solamente el downloadUri que se obtuvo previamente, se coloca como endpoint, para así obtener el JSON esperado.

Figura 28 Petición hacia la URL downloadUri para obtener el JSON

```

1  {
2    "version": {
3      "json_export": "210",
4      "page_segmentation": "55",
5      "schema": "1.1.0",
6      "structure": "1.1125.0",
7      "table_structure": "5"
8    },
9    "extended_metadata": {
10      "ID_instance": "5C 9A BD 33 61 BB B2 11 0A 00 86 AB 32 4D C3 AC",
11      "ID_permanent": "91 60 17 AB 0C 13 08 03 0B E3 25 E8 D9 5A 81 18",
12      "has_acroform": false,
13      "has_embedded_files": false,
14      "is_XFA": false,
15      "is_certified": false,
16      "is_encrypted": false,
17      "is_digitaly_signed": false,
18      "language": "es",
19      "page_count": 6,
20      "pdf_version": "1.6",
21      "pdfa_compliance_level": "",
22      "pdfua_compliance_level": ""
23    },
24    "elements": [
25      {
26        "Bounds": [
...

```

Figura 29 Obtención del JSON esperado.

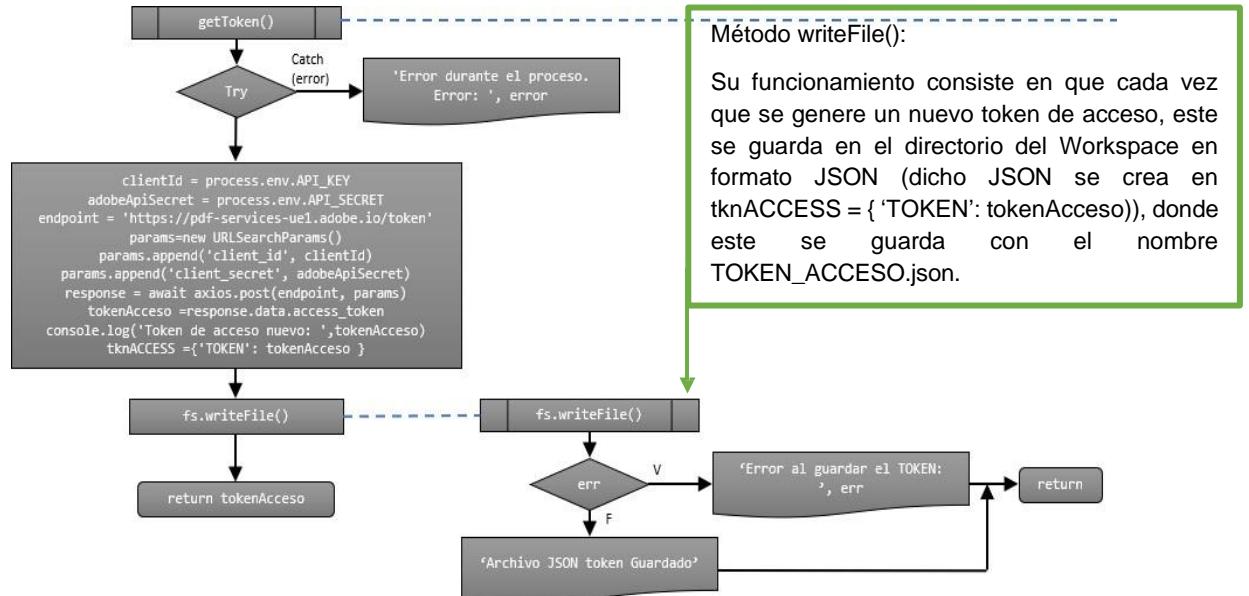
Realizada la petición finalmente obtenemos el JSON, que a continuación será utilizado para el proceso del Layout.

### 6.3.1.2. Funcionamiento del API Adobe Services en Javascript (Visual Studio Code)

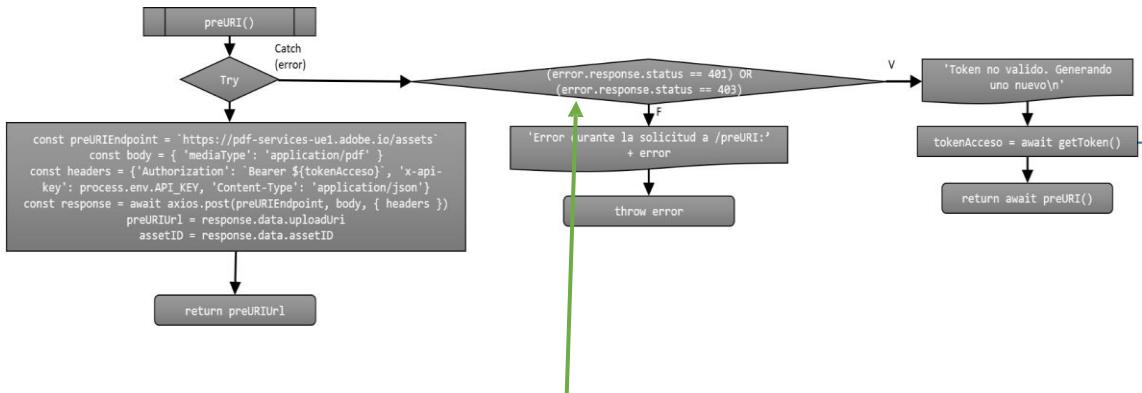
Para hacer que el api sea consumida mediante un servidor Express, se utilizo Node.js en el lado del servidor, donde serán ejecutadas las rutas de dicha API. Se debe seguir el mismo

flujo que el diagrama mostrado anteriormente, a continuación, Se muestran los diagramas de flujo de cada uno de los métodos o rutas (Endpoints).

**Figura 30. Diagrama: `getToken()` (obtener token de Acceso).**



*Figura 30 Diagrama `preURI()` (obtener url para cargar el archivo).*



*Figura 31 Diagrama uploadAsset(uploadUri, fileBuffer).*

Verificación de que el token de acceso sea valido y vigente:

En caso de que se muestre un error HTTP 401 o 403 (Acceso restringido o no autorizado), se pida otro Token nuevo a `getToken()`.

De lo contrario, esta ruta se ejecutará sin problema.

(carga del archivo pdf al servidor de Adobe (mediante la URL generada en el endpoint anterior: preURI()).

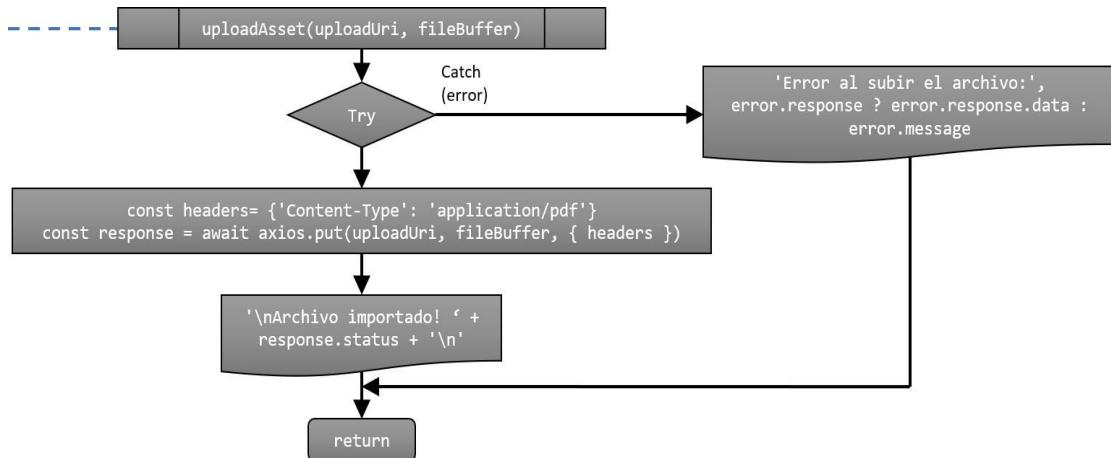


Figura 32 Upload Asset

- **uploadSamplePDF():** Este método se encarga de cargar el archivo PDF importado al servidor, guardándolo en la carpeta '/uploads', en el directorio del Backend), dicho archivo se lee y se guarda su dirección en la constante **fileBuffer**.

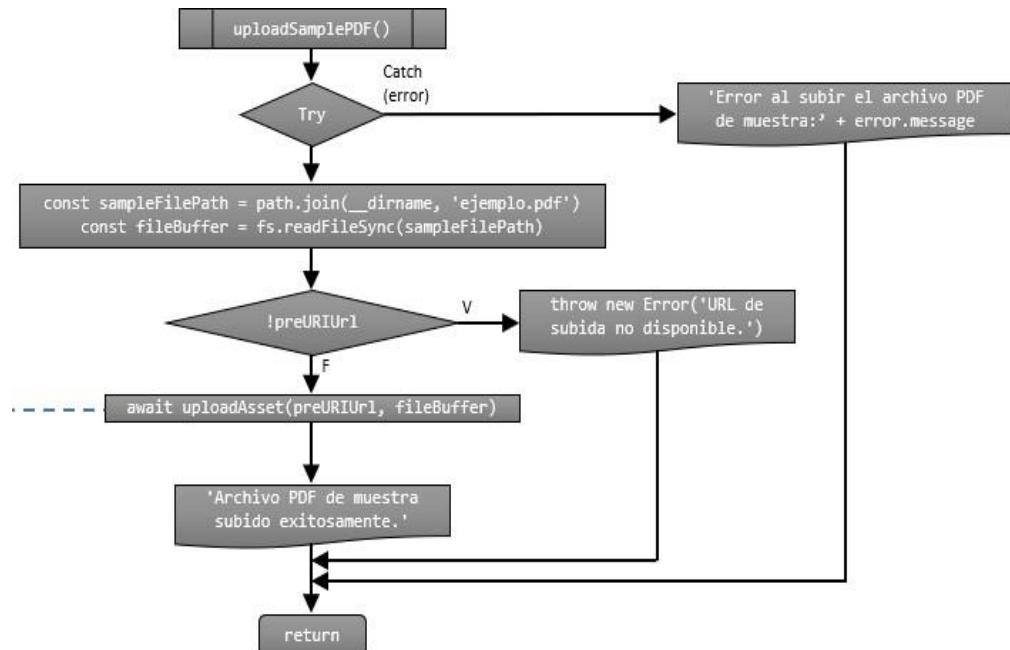


Figura 33 Diagrama uploadSamplePDF()

### *createJob()*

Este método se encarga de hacer la petición a Adobe, para que se extraigan sus datos, donde sus credenciales para realizarla son: el token de Acceso, api\_key (se obtiene desde la pagina de Adobe developer con una cuenta) y un ‘**assetID**’ (obtenido desde preURI). Una vez realizada la petición, obtenemos una URL (locationURL o jobLocation).

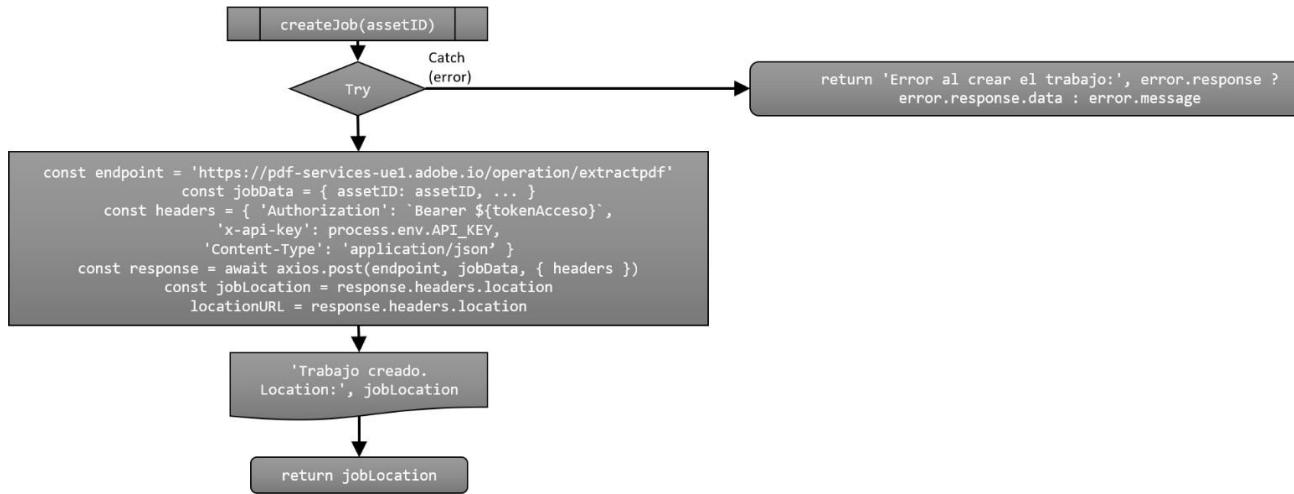


Figura 34 Diagrama `createJob()`

### *getStatusJob(locationURL)*

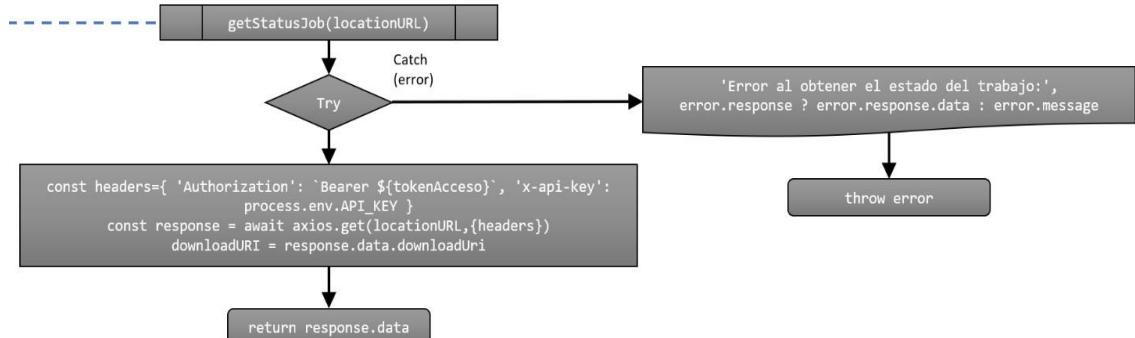


Figura 35 Diagrama `getStatusJob(locationURL)`

Este método se encarga de realizar la petición a adobe para obtener el estado del trabajo, si ya esta terminado, en proceso o fallido, requiriendo como credenciales, el **tokenAcceso**, **api\_key** y **locationURL** (obtenido desde `createJob()`). Una vez terminado el proceso, se obtendrá un **downloadURI** (url de descarga del json)

- **pollJobStatus(locationURL):**

Este método se encarga de realizar el sondeo del estado del trabajo, nos va mostrando la depuración del trabajo, si esta en proceso, o fallo, durante cierto tiempo. Como credencial solamente requiere de locationURL, y el estado del trabajo.

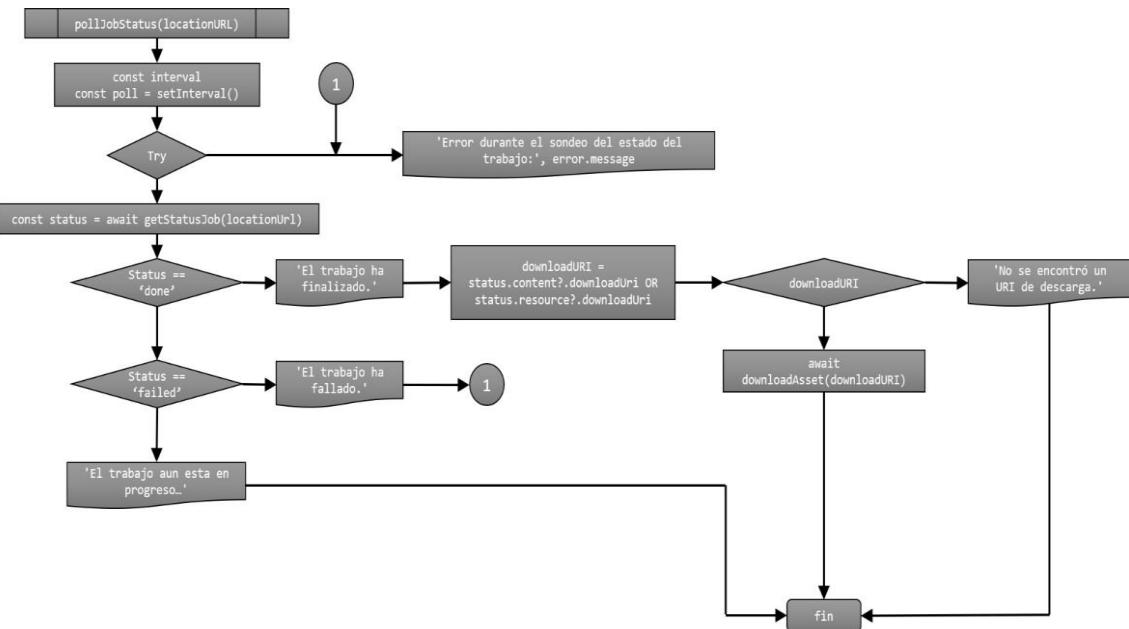


Figura 36 Diagrama `pollJobStatus(locationURL)`

### **downloadAsset(downloadURI):**

Este método es el encargado de realizar la descarga del JSON esperado, esto realizando una petición GET a **downloadURI** (obtenido desde `getStatusJob()`).

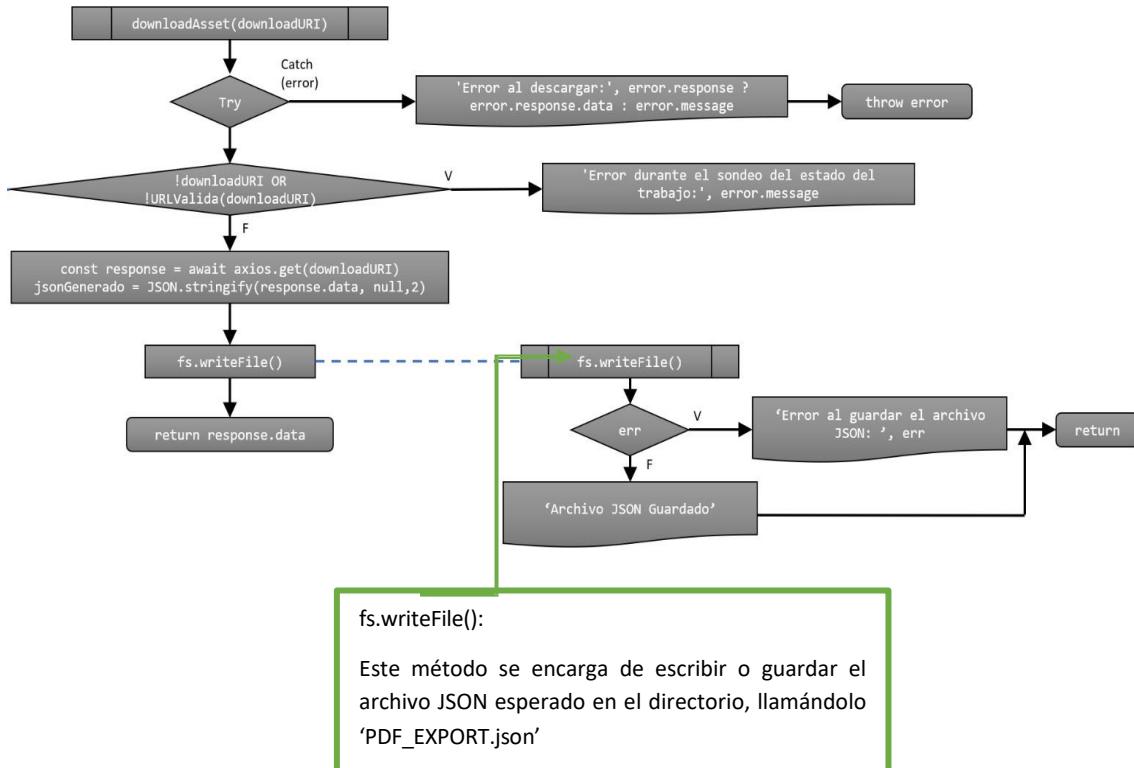


Figura 37 Diagrama downloadAsset(downloadURI)

### URLValida(string):

Este método se encarga simplemente de verificar si el URL es valido o no.

Donde el parámetro recibido (string) es la URL.

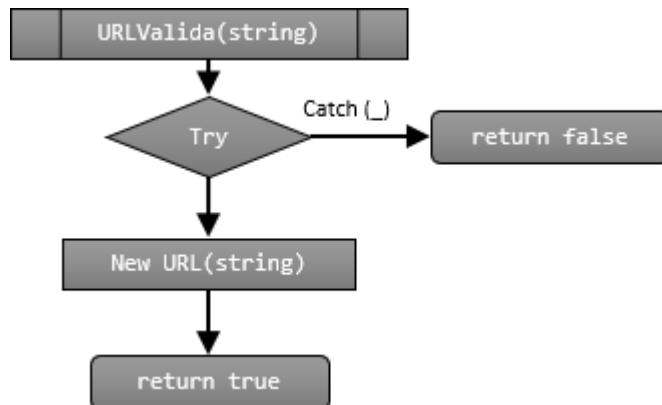


Figura 38 Diagrama URLValida(string)

### 6.3.2. RECORRIDO DEL JSON Y EXTRACCION DE SUS DATOS.

Semana 4 a semana 7: del 26 de agosto al 20 de septiembre.

Al terminar el trabajo del servicio de Adobe, teniendo el JSON, posteriormente se realiza el recorrido para extraer su contenido. En este caso necesitaremos de los datos **Path** y **Text**, donde Path contiene la posición del contenido (ya sea un texto o una tabla), y Text que básicamente es el texto plano que contiene el PDF.

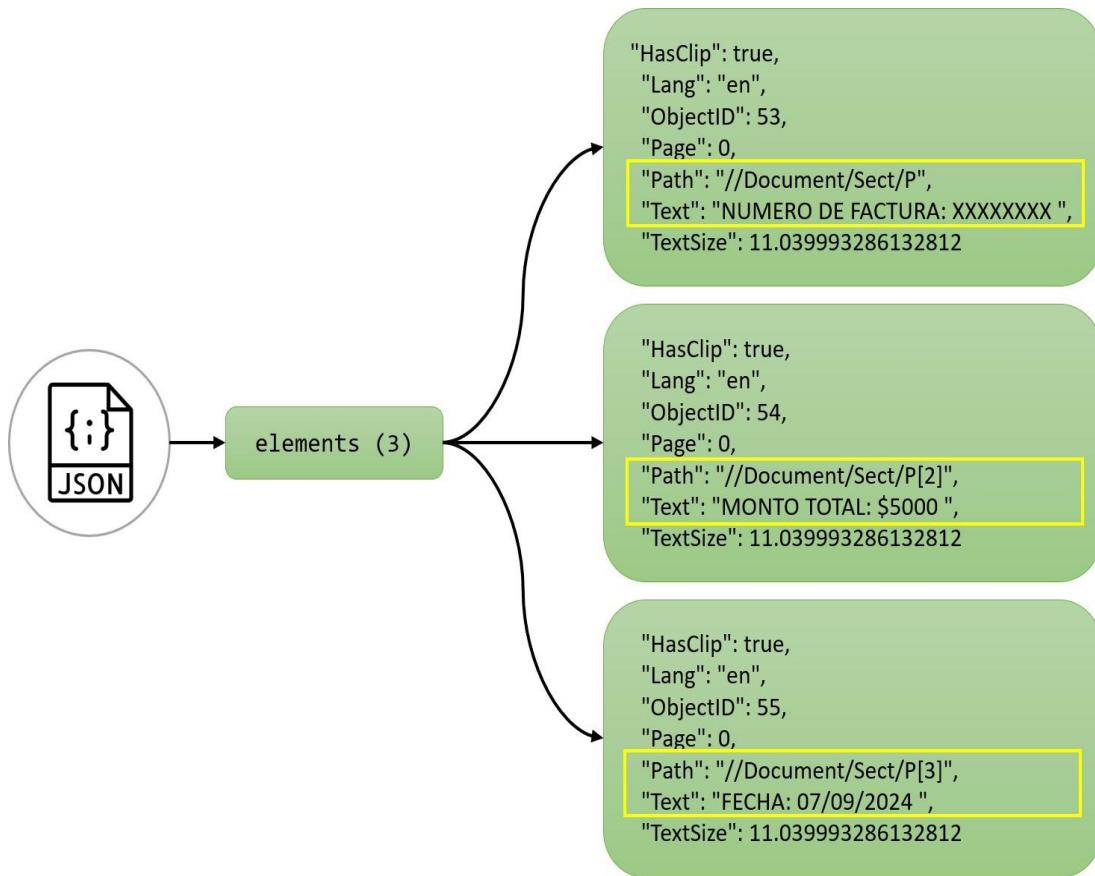


Figura 39 Diagrama del JSON y sus datos a extraer.

#### NOTA:

Considerando que se aplicara un formato de recorrido diferente para cada Factura PDF de cada proveedor, lo cual algunos datos de este ejemplo NO siempre serán los mismos. Esto debido a que los clientes los cuales la empresa recibe, las facturas son totalmente diferentes.

### 6.3.2.1. Recorrido de Ejemplo.

A continuación, se adjunta un ejemplo de recorrido y como es que el sistema recorre los datos y los muestra en una tabla como vista preliminar del Layout Final. En este caso es el proveedor llamado “Editorial Juventud”, que su respectiva factura luce de esta forma:

Editorial Juventud s. a.  Calle Provenza, 101 08029 Barcelona - España C.I.F.: A-08009243 editorialjuventud@retemail.es www.editorialjuventud.es		N.º de control de FEDECALI	Pág.: 1	
<b>Destinatario IRONPORT GROUP SA DE CV</b>				
		Factura N.º: <b>47558</b>	Nº Bultos: <b>2</b>	
		Mod. Exportac.: <b>Depósito</b>	Peso Bruto: <b>798,00Kg</b>	
		Aduana: <b>Barcelona</b>	Peso Neto: <b>796,00Kg</b>	
		Cod. Exportador: <b>594-37</b>	Marca:	
		Pos. Estadística.: <b>49.019900</b>	Prov. origen: <b>España</b>	
		Forma de Envío: <b>Agencia</b>		
Total Euros: <b>2.583,04</b>	Cambio: <b>1 Eur/ Eur</b>	Forma de Cobro: <b>Transferencia</b>		
Total Divisas: <b>2.583,04</b>	Valor Estadístico: <b>2.583,04</b>	Cond. <b>Fabricante</b>		
Divisa: <b>Euros</b>	Valor CIF <b>2.583,04</b>	Vinculo	<b>No existe</b>	
Firma y Sello del Titular				
<b>Barcelona, 26/07/2023</b>				
Cant.	ISBN	Autor / Título	Precio	Total
10	9788426147677	Donaldson, Julia	Las gafas del Topo	0,95
10	9788426147691	Donaldson, Julia	La banda del Tejón	0,95
20	9788426147721	Blyton, Enid	Los Cinco y la perrita perdida	1,05
10	9788426147943	Pasinski, Mikolaj	Van Dog	2,21
10	9788426147837	Spires, Ashley	Binky bajo presión	1,25
15	9788426147790	Montanari, Eva	Un dia en el parque	1,25
5	9788426148018	Kim , Minu	Caracol	1,54
5	9788426147813	Besora , Ramon	Los iris	1,63
20	9788426147745	Blyton, Enid	Tim encuentra un tesoro	1,05
10	9788426147929	Skowronkska, Anna	Viento	2,11
10	9788426148056	Ardite , Marta	Verde	1,54
10	9788426148070	Ardite , Marta	Azul	1,54
50	9788426148179	Dussutour, Alice	Nacer niña	2,21
10	9788426148193	Escoffier, Michaël	;Fuego! ;Fuego!	1,34
10	9788426148261	Ardite , Marta	Amarillo	1,54
10	9788426148285	Ardite , Marta	Rojo	1,54
3	9788426147967	Gravel, Elise	Las lecciones del profesor Tarambana. Lec	1,15
10	9788426148148	Ruillier, Jérôme	El pequeño mundo de Nur	1,44
10	9788426148452	Morstad, Julie	El tiempo es una flor	2,21
20	9788426148346	Kaid-Salah Ferrón, Shedad	Una ciudad en Marte	1,91
10	9788426148094	Clanton, Ben	Una ciudad en Marte	1,91
5	9788426147981	Bensard, Éva	Narvalicorno y Medu	1,34
5	9788426148247	Vo, Nancy	El gran libro de los récords del arte	2,49
3	9788426148216	Fernández García , ChristEl	Mamas	1,63
10	9788426148308	Ardite , Marta	monje del turbante blanco	2,60
15	9788426148223	Straßer, Susanne	Naranja	1,54
3	9788426148360	Tabora, Carla; Faria, RosGaudi,	¡La sopa está lista!	1,73
15	9788426148322	Barrecheguren, Pablo	un verano inolvidable	1,54
10	9788426148124	Pérez de Pesloñán, Lucile	La genética	2,20
10	9788426148414	Sánchez, Kiko	El mundo de Max	2,11
		Túneles		2,40
<b>Mercancía destinada a la exportación - Exenta de IVA</b>				

Figura 40 Factura PDF de prueba.

En ella podemos observar cada uno de sus datos que se van a consultar, tales como: Número de Factura y El número de parte. Para el número de Factura accederemos a su contenido con la ayuda de dichos elementos (path y text).

### Recorrido:

Realizados con Visual Studio Code en lenguaje Javascript.

#### 6.3.2.1.1 Recorrido de un dato Fijo:

```
///- FORMATO --  
//PROVEEDOR EDITORIAL JUVENTUD 'VER106'  
//NUMERO DE FACTURA  
jsonData.elements.forEach((unElemento) =>{  
    let text = unElemento.Text || '';  
    let path = unElemento.Path || '';  
    if(path.includes('//Document/Table/TR[2]/TD[2]/P')){  
        numFactura = text.match(regex);  
    }  
});
```

Figura 41 Algoritmo para la extracción de un dato fijo.

Donde se verifica si el **path** contiene cierto contenido, en ese caso la posición del número de la Factura, donde se ubica en la segunda columna y segunda fila de la tabla (**TR[2]/TD[2]/P**). Donde **numFactura**, recogerá el texto de esa posición. Hay que tomar en cuenta que la posición del **numFactura** se encuentre fijo en el formato de la factura, en caso contrario, podría mostrar valores erróneos si el documento cambia de diseño, en estos casos, se identifica la nueva posición del texto de **numFactura**.

Y así se realizará lo mismo sucesivamente con datos que tal vez no se encuentren en la base de datos, como la *Fecha de la Factura, monto, valores de mercancía, moneda, incoterm*, entre otros datos que se requieran para el Layout final.

NOTA: Se debe extraer solamente el dato que se necesita, en este caso, se extrae un texto, el cual contiene “Fecha de factura: dd/mm/aaa” el cual se aplican técnicas o Expresiones regulares para extraer solamente la fecha.

```

//FECHA DE FACTURA
let fechaFactura = '';
jsonData.elements.forEach((unElemento)=>{
    let text = unElemento.Text;
    let path = unElemento.Path;
    if(path.includes('//Document/H1')){
        fechaFactura = text.match(regexFecha) ? text.match(regexFecha)[0] : '';
    }
});

```

*Figura 42 Algoritmo para extraer un dato Fijo (2).*

### 6.3.2.1.2. Recorrido Dinámico (extraer información de una columna):

Para lograr obtener los datos de cada fila de una Columna, se necesita realizar el recorrido, verificando la posición de la información necesaria, de una manera dinámica, es decir, que el índice de la posición de esa fila, vaya incrementando automáticamente (indirectamente).

Para ello es necesario obtener el Path de esa posición de la columna, para este ejemplo es:

//Document/Table[2]/TR[1]/TD[6]/P

Donde TR es la posición de la Fila, que posteriormente incrementara su índice (1), conforme se va recorriendo el JSON y TD es la posición de la columna la cual necesitamos obtener sus datos, esta última se mantiene fija en este recorrido, ya que es la necesaria para obtener cierta información (lo mismo sucede con los Números de Parte y demás información necesaria que se encuentre en la tabla de la factura.) El recorrido se mostraría de la siguiente manera:

Path: //Document/Table[2]/TR[1]/TD[6]/P	Text: 9.52
Path: //Document/Table[2]/TR[2]/TD[6]/P	Text: 9.52
Path: //Document/Table[2]/TR[3]/TD[6]/P	Text: 20.96
Path: //Document/Table[2]/TR[4]/TD[6]/P	Text: 22.12
Path: //Document/Table[2]/TR[5]/TD[6]/P	Text: 12.50

## Información recorrida en la factura:

Cant.	ISBN	Autor / Título	Precio	Total
10	9788426147677	Donaldson, Julia	Las gafas del Topo	0,95
10	9788426147691	Donaldson, Julia	La banda del Tejón	0,95
20	9788426147721	Blyton, Enid	Los Cinco y la perrita perdida	1,05
10	9788426147943	Pasinski, Mikolaj	Van Dog	2,21
10	9788426147837	Spires, Ashley	Binky bajo presión	1,25
				12,50

Figura 43 Columna recorrida para prueba.

A continuación, se muestra la manera en que se recorre el json y se verifica la posición para obtener dicha información en Visual Studio Code:

```
//VALOR DE LA MERCANCIA
jsonData.elements.forEach((unElemento)=>{
    let text = unElemento.Text || '';
    let path = unElemento.Path || '';
    if(path.includes("//Document/Table[") && path.includes(']/TR[') && path.includes(']/TD[6]/P')){
        text = text.replace(",",".");
        valoresMercancia.push(text);
    }
});
```

Figura 44 Verificación de la posición de renglones de una tabla.

Donde se obtiene en una variable el texto y el **path**, una vez almacenándolos, se verifica si el **Path** contiene el **path** (“//Document/Table[“, “]/TR[“ y “]/TD[6]/P”), transformándolo así en un recorrido dinámico, porque así se tomara en cuenta solamente la información que se encuentre en esa posición (**path**), indicándolo en la columna necesaria.

Finalmente, obteniendo el texto de esa posición, se almacena en un Arreglo para posteriormente mostrarlos en la vista preliminar del Layout final.

También se aplica este método para otros datos de alguna columna que se requieran mostrar.

### 6.3.2.1.3. Extracción de datos que serán consultados a la base de datos:

Una vez realizado el trabajo de la extracción de los datos del PDF y guardados en un JSON, sus recorridos, a continuación, con la ayuda de algunos datos extraídos de dicho JSON, nos sirven para consultarlos en la base de datos de Sistemas CASA en caso de que algunos datos no se encuentren en el PDF de la factura, pero si necesarios para mostrar en el Layout final.

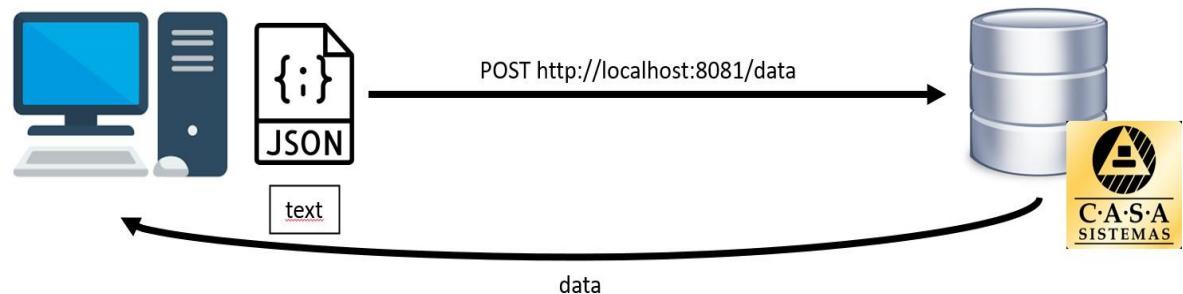


Figura 45 Estructura de las consultas a la base de datos de Sistemas CASA.

Considerando que por medio de un dato como **Número de parte**, **fracción**, **Clave de Proveedor**, podemos realizar la consulta a más datos, según lo que se necesite de la factura previamente cargada.

#### Tablas que contienen la información que se requiere.

Sistemas casa tiene una base de datos muy robusta y completa, con gran numero de tablas. En este caso usaremos 2 de ellas que se identifican como **CTRAC\_FRACC** y **CTRAC\_FRACPAR**.

#### Conexión a la base de datos en el *serverIndex.js*:

Para ello se almacenan las credenciales hacia la base de datos de **SISTEMAS CASA**:

```

const bddCredenciales = {
    host: '192.168.10.69',
    port: 3050,
    database: 'C:/CASAWIN/CSAAWIN/Datos/CASA.GDB',
    user: 'Admin',
    password: 'admin',
    pageSize: 4096
};

```

Figura 46 Credenciales para acceder a la base de datos.

En el servidor (serverIndex.js) también es necesario agregar una ruta la cual nos servirá para realizar la consulta hacia la base de datos, donde esta misma recibirá datos desde el Frontend, donde se recolectan los datos. A continuación, se muestra dicha ruta en Visual Studio Code:

```

const numPartSet = new Set();
//CONSULTA
app.post('/data', (req, res) => {
  const { text, cveProveedor } = req.body;

  if(cveProveedor == 'VER106'){
    if (!/^d+$/ .test(text)) {
      return res.status(400).send('El numero de parte no es válido');
    }
  }else if(cveProveedor == 'ZSI'){
    if (!/\b[A-Z0-9]{6,10}\b/.test(text)) {
      return res.status(400).send('El numero de parte no es válido');
    }
  }else if(cveProveedor == 'TON23'){
    if (!/^d+(-d+)?$/.test(text) && !/^A-Z0-9-[6,25]$/.test(text)) {
      return res.status(400).send('El numero de parte no es válido');
    }
    console.log("TON23 verificado");
  }
  firebird.attach(bddCredenciales, (err, db) => {
    if (err) {
      console.log(err);
      return res.status(500).send('No se pudo conectar a la base de datos');
    }

    console.log("Conexion establecida a la base de datos");
    //Consulta para obtener todos los registros donde CVE_PROV sea 'VER106'
    db.query("SELECT fpar.CVE_PROV, fpar.DES_MERC, fpar.NUM_PART, fracc.NUM_FRACC FROM CTRAC_FRACPAR fpar JOIN CTRAC_FRAC fracc ON fpar.ID
    if (err) {
      return res.status(500).send('Error al consultar');
    }

    db.detach();
    const filteredResults = result.filter((row) => {
      if (numPartSet.has(row.NUM_PART)) {
        return false;
      } else {
        numPartSet.add(row.NUM_PART);
        return true;
      }
    });
    //Si hay resultados, los enviamos como JSON
    if (filteredResults.length > 0) {
      res.json(filteredResults);
    } else {
      console.log("Fallo")
      res.json([]);
    }

    /*if (result.length === 0) {
      return res.json([{ NUM_PART: text, DES_MERC: 'No encontrado', CVE_PROV: cveProveedor, NUM_FRACC: '' }]);
    }
    res.json(result);*/
  });
});

```

Figura 47 Ruta del servidor que realiza la consulta a Sistemas CASA.

¿Dónde se reciben los datos a consultar?

- El tipo de numero de parte que se va a consultar (números, Números y letras, Números con guion Medio (-))
- Que los números de parte consultados, no se dupliquen a la respuesta

Conexión a la base de datos mediante Firebird.attach.

Query/Consulta que se encargara de consultar los datos que se necesitan para el Layout final. Donde en ella se reciben los datos necesarios extraídos del JSON y que, además consultados, obtenemos más información requerida.

```
[cveProveedor, text], (err, result) => {
```

Donde se recibe la clave del proveedor y numero de Parte (almacenado en **text**).

Por lo tanto, la consulta/query se realiza de la siguiente manera:

```
SELECT fpar.CVE_PROV, fpar.DES_MERC,
       fpar.NUM_PART, fracc.NUM_FRACC
  FROM CTRAC_FRACPAR fpar
 JOIN CTRAC_FRACC fracc
    ON fpar.ID_FRACC = fracc.ID_FRACC
   WHERE fpar.CVE_PROV = ? AND NUM_PART = ?
```

Donde se consulta:

- Clave del proveedor (CVE\_PROV)
- Descripción de la mercancía (DESC\_MERC)
- Numero de parte (NUM\_PART)
- Numero de fracción (NUM\_FRACC)

### **6.3.3. APLICACIÓN DE FORMATOS DE DIVERSAS FACTURAS.**

Semana 8 a semana 15: del 23 de septiembre al 15 de noviembre

Teniendo ya concluidas las etapas anteriores, a continuación, esta etapa, consiste en conocer la estructura de cada Factura de proveedores diferentes, esto se debe a que no todas las Facturas que se reciben en el área de Trafico en Tecnológica Aduanal, tendrán la misma manera de extraer sus datos que de otra de otro proveedor.

Esta etapa se dio la tarea de consultar manualmente el JSON proporcionado por el API de adobe, la posición de los datos fijos (que cada proveedor los tiene en una misma ubicación) y la posición de la tabla que contiene los productos de la factura (de igual manera, cada proveedor la tiene en una sola ubicación) para así ubicar sus columnas y extraer los datos necesarios para la consulta a la base de datos.

A continuación, se adjuntan 2 ejemplos de facturas de proveedores diferentes, donde se señala algunos de los datos que serán extraídos y/o consultados.

 <p><b>Editorial Juventud s. a.</b>            Calle Provenza, 101            08029 Barcelona - España            C.I.F.: A-08009243            editorialjuventud@retemail.es  <a href="http://www.editorialjuventud.es">www.editorialjuventud.es</a></p>	N.º de control de FEDECALI	Pág.: 1		
<b>Destinatario IRONPORT GROUP SA DE CV</b>  Calle Nelson N° Ext 23 N° Int 103 Colonia Vallarta Norte Guadalajara, Jalisco C.P. 44690	Factura N.º: <b>47558</b> Mod. Exportac.: <b>Deposito</b> Aduana: <b>Barcelona</b> Cod. Exportador: <b>594-37</b> Pos. Estadística.: <b>49.019900</b> Forma de Envío: <b>Agencia</b>	Nº Bultos: <b>2</b> Peso Bruto: <b>798,00Kg</b> Peso Neto: <b>796,00Kg</b> Marca: Prov. origen: <b>España</b>		
Total Euros: <b>2.583,04</b> Cambio: <b>1 Eur/ Eur</b> Total Divisas: <b>2.583,04</b> Valor Estadístico: <b>2.583,04</b> Cond. <b>Fabricante</b> Divisa: <b>Euros</b> Valor CIF <b>2.583,04</b> Vinculo <b>No existe</b>	Forma de Cobro: <b>Transferencia</b>			
Firma y Sello del Titular				
<b>Barcelona, 26/07/2023</b>				
Cant.	ISBN	Autor / Título	Precio	Total
10	9788426147677	Donaldson, Julia Las gafas del Topo	0,95	9,52
10	9788426147691	Donaldson, Julia La banda del Tejón	0,95	9,52
20	9788426147721	Blyton, Enid Los Cinco y la perrita perdida	1,05	20,96
10	9788426147943	Pasiński, Mikolaj Van Dog	2,21	22,12
10	9788426147837	Spires, Ashley Binky bajo presión	1,25	12,50
15	9788426147790	Montanari, Eva Un día en el parque	1,25	18,75
5	9788426148016	Kim , Minu Caracol	1,54	7,69
5	9788426147813	Besora , Ramon Los iris	1,63	8,18
20	9788426147745	Blyton, Enid Tim encuentra un tesoro	1,05	20,96
10	9788426147929	Skowronska, Anna Viento	2,11	21,15
10	9788426148056	Ardite , Marta Verde	1,54	15,38
10	9788426148070	Ardite , Marta Azul	1,54	15,38
50	9788426148179	Dussutour, Alice Nacer niña	2,21	110,60
10	9788426148193	Escoffier, Michaël ¡Fuego! ¡Fuego!	1,34	13,37
10	9788426148261	Ardite , Marta Amarillo	1,54	15,38
10	9788426148285	Ardite , Marta Rojo	1,54	15,38
3	9788426147967	Gravel, Elise Las lecciones del profesor Tarambana. Lec	1,15	3,46
10	9788426148148	Ruillier, Jérôme El pequeño mundo de Nur	1,44	14,42
10	9788426148452	Morstad, Julie El tiempo es una flor	2,21	22,12
20	9788426148346	Kaid-Salah Ferrón, Sheddúa Una ciudad en Marte	1,91	38,26
10	9788426148094	Clanton, Ben Narvalicorno y Medu	1,34	13,37
5	9788426147981	Bensard, Éva El gran libro de los récords del arte	2,49	12,45
5	9788426148247	Vo, Nancy Mamas	1,63	8,18
3	9788426148216	Fernández García , ChristEl monje del turbante blanco	2,60	7,79
10	9788426148308	Ardite , Marta Naranja	1,54	15,38
15	9788426148223	Straßer, Susanne ¡La sopa está lista!	1,73	25,97
3	9788426148360	Tabora, Carla; Faria, RosGaudi, un verano inolvidable	1,54	4,61
15	9788426148322	Barrecheguren, Pablo La genética	2,20	33,03
10	9788426148124	de Pesloüan, Lucile El mundo de Max	2,11	21,15
10	9788426148414	Sánchez, Kiko Túneles	2,40	24,04
<b>Mercancía destinada a la exportación - Exenta de IVA</b>				

Figura 48 Ejemplo de factura I

**NINGBO DYNACO HYDRAULIC CO., LTD.**  
 HUILONG INDUSTRIAL ZONE 39, YIWU, JINHUA, ZHEJIANG PROVINCE, CHINA  
 TAX ID: 91330212704848544D

**Commercial Invoice**

To: Javier Jacobo Mier. JAMJ751112EI6  
 Attn: Carretera Guadalajara-Chapala km 20, no 78, Jardines de la Calera.  
 Tlajomulco de Zúñiga, Jalisco. C.P. 45678. México.

INVOICE NO.:	DH231008-8
CUSTOMER P/O:	
INV DATE:	2024-01-07

Item No.	Description	Part No.	Quantity	PRICE (USD)	AMOUNT (USD)
1	P350 PUMP	P350B542LEAB1298OHAB10-1	10	430.00	4300.00
2	P350 PUMP	P350B542JFAB1725PKAB17-1	1	430.00	430.00
3	MOTOR LIP SEAL	391-2883-115	5	5.40	27.00
4	MOTOR THRUST PLATE	391-2185-068	12	14.50	174.00
5	NYLON BACKUP SEAL	391-2885-075	12	1.20	14.40
6	THRUST PLATE SIDE SEAL	391-2885-073	24	1.20	28.80
7	THRUST PLATE END SEAL	391-2885-074	24	1.20	28.80
8	HOUSING GASKET	391-2884-021	12	1.20	14.40
9	GEAR HOUSING	323-8210-100	2	29.50	59.00
10	GEAR HOUSING	323-8212-100	2	29.50	59.00
11	SHAFT BRG. SNAP RING	391-2686-065	10	1.90	19.00
12	SEAL RETAINER	391-3383-020	10	1.10	11.00
13	INPUT SHAFT BRG	391-0381-077	10	5.60	56.00
14	PUMP LIP SEAL	391-2883-103	10	1.76	17.60
15	PRESSURE PLUG	391-2286-004	10	0.76	7.60
16	SHAFT BUSHING	391-0482-308	60	2.30	138.00
17	MH350 THRUST PLATE	391-2185-057	40	9.18	367.20
18	MHP350 BACKUP SEAL	391-2882-139	40	0.88	35.20
19	MHP350 CHANNEL	391-2882-140	40	0.88	35.20
20	HOUSING DOWELL PIN	391-2082-032	20	0.39	7.80
21	MHP350 GEAR HOUSING	323-8110-100	5	23.50	117.50
22	MHP350 GEAR HOUSING	323-8112-100	10	23.50	235.00
23	MH350 GEAR SET	323-2810-000	4	41.50	166.00
24	SHAFT & GEAR	323-2912-240	3	41.50	124.50
25	SHAFT & GEAR	323-2915-250	1	41.50	41.50
26	SHAFT & GEAR	323-2917-250	1	41.50	41.50
27	GEAR HOUSING	323-8115-100	1	29.50	29.50
28	GEAR HOUSING	323-8117-100	1	29.50	29.50
29	P350 PUMP 3239120351	P350L542CSEB12-98CCEB11-1	2	428.00	856.00
Total:	CIF		382 PCS		7471.00

**Payment Term:** 30% T/T payment in advance, 70% T/T payment before shipment.

**Bank Information:**

SWIFT/BIC Code: CHASHKHH

Account Number: 63007938672

Account Name: YIWU Pengju Trade Company

Bank Name: JPMorgan Chase Bank N.A., Hong Kong Branch

Bank Address: 18/F, 20/F, 22-29/F, CHATER HOUSE, 8 CONNAUGHT ROAD CENTRAL, HONG KONG

Bank Code: 007

Please include the following message to receiver when making a payment:

[Invoice Number] [Product]



Figura 49 Ejemplo de factura 2

### 6.3.3.1. Consultando el JSON, las posiciones de los datos de cada proveedor.

Tomando en cuenta previamente, los proveedores y sus facturas, contienen información que es propia, en el caso de los datos como Clave del proveedor, divisa, país de origen, entre otros datos más, por ejemplo, en la **Figura 47** la divisa es en euros (EUR) y en la **Figura 48** en dólares estadounidenses (USD) ambos siendo datos fijos, por lo cual, dichos datos serán siempre mostrados en su Layout sin tener que consultarlos ni extraerlos de la Factura PDF.

Si bien se documentó en el capítulo **6.3.2. RECORRIDO DEL JSON Y EXTRACCION DE SUS DATOS**, donde se implementaron recorridos dinámicos y fijos, los cuales se encargan de recorrer el JSON que nos proporciona el API de Adobe y posteriormente mostrar los datos en la columna de la tabla que conformara el layout final. Si bien, Cada formato tiene diferentes **Path** o posiciones del contenido.

Ejemplo:

Extraer un dato fijo de dos facturas de proveedores diferentes. La factura que conforma la **Figura 47**, la extracción de su numero de factura tiene como path lo siguiente:

`//Document/Table/TR[2]/TD[2]/P`

Mientras que en la factura que conforma la **Figura 48**, su numero de factura su path contiene lo siguiente:

`//Document/Sect[2]/P[3]/Sub`

Que notando, no son las mismas, es por eso que se implementan diferentes recorridos para cada proveedor.

Para hacer esto posible, para el elemento <select> se obtiene el nombre del Proveedor, lo cual se almacena su nombre en una variable para ser utilizado como parámetro al método que implementara los formatos de recorrido.



Figura 50 Diagrama de carga de datos según el proveedor seleccionado.

## 6.4. 4ta Etapa: VERIFICACION.

### 6.4.1. Prueba de servidor, consumo de API mediante Javascript/NODE.js.

Semana 16 y 17: del 18 de noviembre al 29 de noviembre.

Para la ejecución del servidor previamente implementado en Javascript con NODE js y axios para las peticiones, se ha implementado una depuración mediante consola para verificar que es lo que sucede mientras el API entra en funcionamiento automático.

```
ps C:\xampp\htdocs\PDFtoEXCEL\Backend> node serverIndex.js
Servidor en el puerto: 8081
Petición a /upload
Token leído
Token válido, continuando con las demás peticiones...
Subiendo archivo PDF...

Archivo importado! 200

Archivo PDF de muestra subido exitosamente.
Creando trabajo...
Trabajo creado. Location: https://pdf-services-ue1.adobe.io/operation/extractpdf/qAwC8ybxzXd3xt6Ra4Dxzum02dichRyH/status
Sondeando el estado del trabajo...
El trabajo ha finalizado.
Archivo JSON guardado
Consultando datos...
```

Figura 51 Depuración del API en consola.

Donde va mostrando un mensaje de lo que se va realizando en cada petición, siendo así, una automatización, donde el servidor al ejecutar el API e importar un PDF, este se guarda en el directorio, posteriormente se sube al servidor de Adobe para así procesarlo, al terminar de consumir el API, el JSON proporcionado se guarda en el directorio del proyecto para así ser leído y consultado.

## 6.4.2. Despliegue.

Para verificar el despliegue, a continuación, se realiza la prueba de importar un PDF, ver su layout preliminar y por último, descargar el Excel. Para ello se extraerán los datos de la Factura de la **Figura 49**.

CLAVE PROVEEDOR	NO. FACTURA	FECHA FACTURA	MONTO FACTURA	MONEDA	INCOTERM	SUBDIVISION	CERT. ORIGEN	NO. PARTE	PAÍS ORIGEN	VENDEDOR	FRACCION
TON23	DH231008-8	2024-01-07		USD	CIF			P350B542LEAB12980HAB10-1		84136099	BOMBA H
TON23	DH231008-8	2024-01-07		USD	CIF			P350B542JFAB1725PKAB17-1		84136099	BOMBA H
TON23	DH231008-8	2024-01-07		USD	CIF			391-2885-115		84842001	JUNTA M
TON23	DH231008-8	2024-01-07		USD	CIF			391-2185-068		84129001	PARTES I
TON23	DH231008-8	2024-01-07		USD	CIF			391-2885-075		30269099	EMPAQUE
TON23	DH231008-8	2024-01-07		USD	CIF			391-2885-073		30269099	EMPAQUE
TON23	DH231008-8	2024-01-07		USD	CIF			391-2885-074		30269099	EMPAQUE
TON23	DH231008-8	2024-01-07		USD	CIF			391-2885-021		40169304	EMPAQUE
TON23	DH231008-8	2024-01-07		USD	CIF			323-8210-100		84129001	PARTES I
TON23	DH231008-8	2024-01-07		USD	CIF			323-8212-100		84129001	PARTES I
TON23	DH231008-8	2024-01-07		USD	CIF			391-2686-065		40169304	EMPAQUE
TON23	DH231008-8	2024-01-07		USD	CIF			391-3383-020		84842001	JUNTA M
TON23	DH231008-8	2024-01-07		USD	CIF			391-0381-077		84842099	RODAMIE
TON23	DH231008-8	2024-01-07		USD	CIF			391-2883-103		84842001	JUNTA M
TON23	DH231008-8	2024-01-07		USD	CIF			391-2286-004		73181599	TORNILLO
TON23	DH231008-8	2024-01-07		USD	CIF			391-0482-308		84833004	BUJE
TON23	DH231008-8	2024-01-07		USD	CIF			391-2185-057		84129001	PARTES I
TON23	DH231008-8	2024-01-07		USD	CIF			391-2882-139		30269099	EMPAQUE
TON23	DH231008-8	2024-01-07		USD	CIF			391-2882-140		30269099	EMPAQUE
TON23	DH231008-8	2024-01-07		USD	CIF			391-2085-029		73182103	PASADOR

Figura 52 Layout preliminar con datos extraídos de factura “DH231008-8”.

DESCRIPCION	VALOR MERCANCIA	UMC	CANTIDAD DE UMC	CANTIDAD DE UNIT	REFERENCIA ARANCELARIA	Marca	Submodelo	No. Serie	VALORACION	TIPO DE MERCAN	VINCULACIO
A HIDRAULICA	4300.00	10							1	N	0
A HIDRAULICA	430.00	1							1	N	0
MECANICA DE ESTANQUEIDAD	27.00	5							1	N	0
S PARA MOTOR HIDRAULICO	174.00	12							1	N	0
QUES DE PLASTICO	14.40	12							1	N	0
QUES DE PLASTICO	28.80	24							1	N	0
QUES DE PLASTICO	28.80	24							1	N	0
QUES DE CAUCHO	14.40	12							1	N	0
S PARA MOTOR HIDRAULICO	59.00	2							1	N	0
S PARA MOTOR HIDRAULICO	59.00	2							1	N	0
QUES DE CAUCHO	19.00	10							1	N	0
MECANICA DE ESTANQUEIDAD	11.00	10							1	N	0
MIENTOS	56.00	10							1	N	0
MECANICA DE ESTANQUEIDAD	17.60	10							1	N	0
LLOS	7.60	10							1	N	0
	138.00	60							1	N	0
S PARA MOTOR HIDRAULICO	367.20	40							1	N	0
QUES DE PLASTICO	35.20	40							1	N	0
QUES DE PLASTICO	35.20	40							1	N	0
OD	7.80	20							1	N	0

Figura 53 Layout preliminar con datos extraídos de factura “DH231008-8” #2.

Segunda prueba para otro proveedor diferente, factura “50956”:

DETALLE DE FACTURA											
CLAVE PROVEEDOR	NO. FACTURA	FECHA FACTURA	MONTO FACTURA	MONEDA	INCOTERM	SUBDIVISION	CERT. ORIGEN	NO. PARTE	PAÍS ORIGEN	VENDEDOR	FRACCIÓN
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426147677	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426147691	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426147943	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426147837	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426147790	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426148018	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426147813	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426147929	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426148056	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426148070	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426148179	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426148193	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426148261	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426148285	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426148148	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426148452	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426148346	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426148094	ESP	49019904	LIBROS IMPRESOS E
VER106	50956	07/06/2024	12,662.68	EUR	DAP			9788426147981	ESP	49019904	LIBROS IMPRESOS E

© 2024 Adobe Services. Utiliza el servicio de Adobe Services. Departamento de Sistemas, Tecno Logistica Adunam S.A de C.V. Nuevo Laredo, Tamps.

Figura 54 Layout preliminar con datos extraídos de factura “50956”.

DETALLE DE FACTURA											
DESCRIPCION	VALOR MERCANCIA	UMC	CANTIDAD DE UMC	CANTIDAD DE UMT	REFERENCIA ARANCELARIA	Marca	Submodelo	No. Serie	VALORACION	TIPO DE MERCAN	VINCULACION
N IDIOMA ESPAÑOL	73.30	84							1	N	0
N IDIOMA ESPAÑOL	73.30	77							1	N	0
N IDIOMA ESPAÑOL	203.50	92							1	N	0
N IDIOMA ESPAÑOL	70.00	56							1	N	0
N IDIOMA ESPAÑOL	108.75	87							1	N	0
N IDIOMA ESPAÑOL	55.37	36							1	N	0
N IDIOMA ESPAÑOL	58.86	36							1	N	0
N IDIOMA ESPAÑOL	152.28	72							1	N	0
N IDIOMA ESPAÑOL	110.74	72							1	N	0
N IDIOMA ESPAÑOL	110.74	72							1	N	0
N IDIOMA ESPAÑOL	95.12	44							1	N	0
N IDIOMA ESPAÑOL	68.19	51							1	N	0
N IDIOMA ESPAÑOL	78.44	51							1	N	0
N IDIOMA ESPAÑOL	78.44	51							1	N	0
N IDIOMA ESPAÑOL	57.68	40							1	N	0
N IDIOMA ESPAÑOL	181.38	82							1	N	0
N IDIOMA ESPAÑOL	49.74	26							1	N	0
N IDIOMA ESPAÑOL	61.50	46							1	N	0
N IDIOMA ESPAÑOL	84.86	34							1	N	0

© 2024 Adobe Services. Utiliza el servicio de Adobe Services. Departamento de Sistemas, Tecno Logistica Adunam S.A de C.V. Nuevo Laredo, Tamps.

Figura 55 Layout preliminar con datos extraídos de factura “50956” #2.

### 6.4.2.1. Descarga del Layout en Excel.

Una vez creado y mostrado el layout preliminar, se permite descargar el layout en archivo Excel, teniendo por nombre lo siguiente:

Layout\_(fecha actual)\_(hora actual).xls

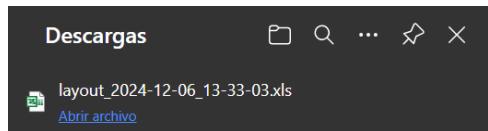


Figura 56 Nombre de descarga de Layout en Excel.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
CLAVE PROVEEDOR	NO. FACTURA	FECHA FACTURA	MONTO	MONEDA	INCOTERM	SUBDIVISI	CERT. ORIGEN	NO. PARTE	PAÍS ORIGEN	VENDEDOR	FRACCIÓN	DESCRIPCIÓN	VALOR MERCANCIA UMC	CANTIDAD DE UMC	CANTIDAD	REFERENCIA	Marca	Submodelo	No. Serie	VALORACION	Tipo de MERCAN	VINCULACIÓN	
TON23	DH21008-8	07/01/2024	USD	OF	P35054216A12950H4810-			84136098 BOMBA HIDRAULICA			4305		10							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	P35054216A12950H4817-			84136098 BOMBA HIDRAULICA			430		1							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-383-115			84120001 JUNTA MECANICA DE ESTANQUEIDAD			27		5							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-318-060			84120001 PARTES PARA MOTOR HIDRAULICO			174		12							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-385-075			84120001 EMPAQUE DE PLASTICO			14.4		12							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-385-073			84120001 EMPAQUE DE PLASTICO			20.8		24							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-385-074			84120001 EMPAQUE DE PLASTICO			28.8		24							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-384-001			84120001 EMPAQUE DE CAUCHO			14.4		12							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	323-210-100			84120001 PARTES PARA MOTOR HIDRAULICO			59		2							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	323-212-100			84120001 PARTES PARA MOTOR HIDRAULICO			59		2							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-384-005			84120001 EMPAQUE DE CAUCHO			19		10							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-385-010			84120001 JUNTA MECANICA DE ESTANQUEIDAD			11		10							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-381-077			84120001 ROGRAMIENTOS			56		10							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-383-103			84120001 JUNTA MECANICA DE ESTANQUEIDAD			17.6		10						1 N	0			
TON23	DH21008-8	07/01/2024	USD	OF	391-238-004			73181598 TORNILLOS			7.6		10							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-0482-308			84833004 BURILS			138		60							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-185-057			84120001 PARTES PARA MOTOR HIDRAULICO			367.2									1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-185-155			84120001 EMPAQUES DE PLASTICO			35.2		40							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	295-2882-140			84120001 EMPAQUES DE PLASTICO			25.25		40							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	391-202-021			73181403 PASADOR			7.8		20							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	323-210-100			84120001 PARTES PARA MOTOR HIDRAULICO			117.5		5							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	323-212-100			84120001 PARTES PARA MOTOR HIDRAULICO			235		10							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	323-210-000			84831008 JUEGO DE FLECHAS			166		4							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	323-212-240			84831008 FLECHAS			124.5		3							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	323-212-250			84831008 FLECHAS			41.5		1							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	323-215-100			84120001 PARTES PARA MOTOR HIDRAULICO			28.5		1							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	323-217-100			84120001 PARTES PARA MOTOR HIDRAULICO			29.5		1							1 N	0		
TON23	DH21008-8	07/01/2024	USD	OF	P350L542CSB12-98CCB11-			84136098 BOMBA HIDRAULICA			856		2							1 N	0		

Figura 57 Layout final, factura "DH231008-8" en archivo Excel.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
CLAVE PROVEEDOR	NO. FACTURA	FECHA FACTURA	MONTO	MONEDA	INCOTERM	SUBDIVISION	CERT. ORIGEN	NO. PARTE	PAÍS ORIGEN	VENDEDOR	FRACCIÓN	DESCRIPCIÓN	VALOR MERCANCIA UMC	CANTIDAD DE UMC	CANTIDAD DE UNIT	REFERENCIA ARANCELARIA	Marca	Submodelo	No. Serie	VALORACION	Tipo de MERCAN	VINCULACIÓN	
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		73.3		84						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		73.3		77						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		209.5		92						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		70		56						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		108.75		87						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		59.37		36						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		58.86		36						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		152.28		72						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		110.74		72						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		110.74		72						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		95.12		44						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		69.19		51						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		78.44		51						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		78.44		51						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		57.48		40						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		181.38		82						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		49.74		26						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		61.5		46						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		84.66		34						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		27.0		17						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		118.12		82						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		31.16		18						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		180.56		15						1 N	0		
VER106	50956	07/06/2024	12.642.68	EUR	DAP			97.788E-12	ESP	84120004 LIBROS IMPRESOS EN IDOMA ESPAÑOL		24.04		82						1 N	0		

Figura 58 Layout final, factura "50956" en archivo Excel.

## 6.5. 5ta Etapa: DESARROLLO.

### 6.5.1. Instancia del sistema en el servidor de operaciones.

Semana 18: 2 de diciembre.

La etapa de verificación concluyó con éxito, posteriormente se procede a la instalación y configuración del sistema en el entorno del área de tráfico para su posterior producción.

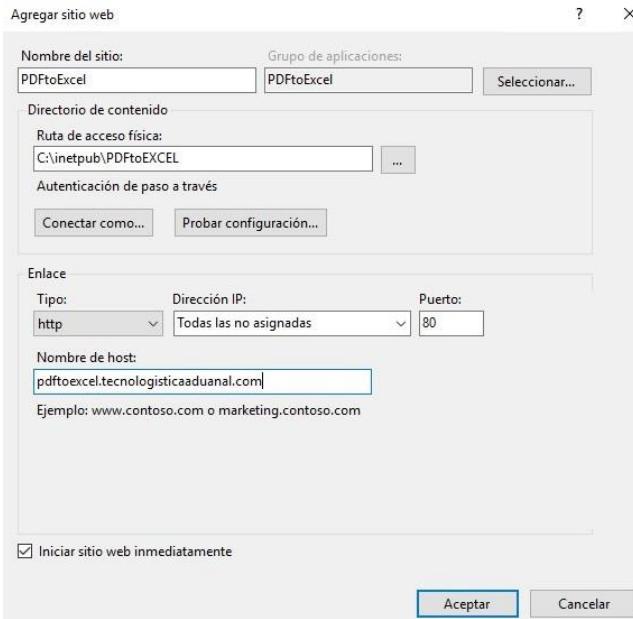


Figura 59 Instancia del sitio al entorno de producción, con nombre de Host.

DNS	Nombre	Tipo	Datos	Marca de tiempo
TL SERVER	_msdc			
	_sites			
	_tcp			
	_udp			
	DomainDnsZones			
	ForestDnsZones			
Zonas de búsqueda directa	(igual que la carpeta princip...)	Host (A)	192.168.1.67	11/03/2024 12:00:00 a. m.
	tsserver	Host (A)	192.168.10.69	static
	tsserver	Host (A)	192.168.1.67	static
	www	Host (A)	162.241.2.161	static
	pdftoexcel	Host (A)	192.168.10.69	static
	(igual que la carpeta princip...)	Inicio de autoridad (SOA)	[294], tsserver.tecnologisticaaduanal.com., hostmaster.tecnologica.com.	static
	(igual que la carpeta princip...)	Servidor de nombres (NS)	tsserver.tecnologisticaaduanal.com.	static
Zonas de búsqueda inversa	1.168.192.in-addr.arpa			
	10.168.192.in-addr.arpa			
	2.241.162.in-addr.arpa			
Puntos de confianza				
Reenviadores condicionales				

Figura 60 Sitio almacenado en el servicio de DNS.

## 7. RESULTADOS.

El proyecto realizado denominado “Extracción de datos en facturas PDF para la creación de layout”, surge en base a los tiempos extensos que los operadores del área de tráfico de Tecnología Aduanal, con el objetivo de acelerar los procesos de operación, ya que la creación de layouts manualmente, eran una tarea considerada minuciosa y tardada.

El sistema no solo acelera los tiempos de operación, sino también evita los retrasos y quejas por parte de los clientes/proveedores, aun así, realizando varias tareas internamente, el sistema mantiene la confiabilidad, la rapidez y la instantaneidad del proceso.

### Necesidad del proyecto.

CLAVE DEL	NO. FACTURA	FECHA DE FACTURA	MONTOS EN	A INCOTERMS	CERT. ORIGEN	NUMERO DE PARTE	PASO ORIGEN	VERIFICACION	FACCIÓN	DESCRIPCION	VALOR DE LA UNIDAD DE MNC	CANTIDAD AD	CANTIDAD	PREFERENCIA ARANCELARIA	MURO	Sistema	No. Serie	VALORACION MERCAN	TIPO DE VINCULACIO
1. VEP008	47558	26/07/2023	\$2594.82	EUR	DAP	378942619761	CHN	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$15.520	6	10	6			1	N	0
2. VEP008	47559	26/07/2023	\$2594.82	EUR	DAP	378942619721	CHN	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$15.520	6	10	6			1	N	0
3. VEP008	47560	26/07/2023	\$2594.82	EUR	DAP	378942619783	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$22.520	6	10	6			1	N	0
4. VEP008	47561	26/07/2023	\$2594.82	EUR	DAP	378942619737	CHN	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$22.520	6	10	6			1	N	0
5. VEP008	47562	26/07/2023	\$2594.82	EUR	DAP	378942619790	CHN	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$18.750	6	15	6			1	N	0
6. VEP008	47563	26/07/2023	\$2594.82	EUR	DAP	378942619409	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$7.690	6	5	6			1	N	0
7. VEP008	47564	26/07/2023	\$2594.82	EUR	DAP	378942619749	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$15.520	6	10	6			1	N	0
8. VEP008	47565	26/07/2023	\$2594.82	EUR	DAP	378942619745	CHN	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$20.360	6	20	6			1	N	0
9. VEP008	47566	26/07/2023	\$2594.82	EUR	DAP	378942619405	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$21.150	6	10	6			1	N	0
10. VEP008	47567	26/07/2023	\$2594.82	EUR	DAP	378942619056	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$15.380	6	10	6			1	N	0
11. VEP008	47568	26/07/2023	\$2594.82	EUR	DAP	378942619407	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$15.380	6	10	6			1	N	0
12. VEP008	47569	26/07/2023	\$2594.82	EUR	DAP	378942619421	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$15.380	6	10	6			1	N	0
13. VEP008	47570	26/07/2023	\$2594.82	EUR	DAP	378942619401	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$15.380	6	10	6			1	N	0
14. VEP008	47571	26/07/2023	\$2594.82	EUR	DAP	378942619420	POL	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$10.000	5	50	6			1	N	0
15. VEP008	47572	26/07/2023	\$2594.82	EUR	DAP	378942619493	TUR	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$13.370	6	10	6			1	N	0
16. VEP008	47573	26/07/2023	\$2594.82	EUR	DAP	378942619491	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$15.380	6	10	6			1	N	0
17. VEP008	47574	26/07/2023	\$2594.82	EUR	DAP	378942619423	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$15.380	6	10	6			1	N	0
18. VEP008	47575	26/07/2023	\$2594.82	EUR	DAP	378942619422	DEU	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$25.970	6	15	6			1	N	0
19. VEP008	47576	26/07/2023	\$2594.82	EUR	DAP	378942619409	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$4.870	6	3	6			1	N	0
20. VEP008	47577	26/07/2023	\$2594.82	EUR	DAP	378942619405	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$14.000	6	10	6			1	N	0
21. VEP008	47578	26/07/2023	\$2594.82	EUR	DAP	378942619452	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$22.120	6	10	6			1	N	0
22. VEP008	47579	26/07/2023	\$2594.82	EUR	DAP	378942619346	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$38.280	6	20	6			1	N	0
23. VEP008	47580	26/07/2023	\$2594.82	EUR	DAP	378942619348	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$13.370	6	10	6			1	N	0
24. VEP008	47581	26/07/2023	\$2594.82	EUR	DAP	378942619479	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$12.450	6	5	6			1	N	0
25. VEP008	47582	26/07/2023	\$2594.82	EUR	DAP	378942619477	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$10.000	5	5	6			1	N	0
26. VEP008	47583	26/07/2023	\$2594.82	EUR	DAP	378942619426	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$7.790	6	3	6			1	N	0
27. VEP008	47584	26/07/2023	\$2594.82	EUR	DAP	378942619308	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$15.380	6	10	6			1	N	0
28. VEP008	47585	26/07/2023	\$2594.82	EUR	DAP	378942619423	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$25.970	6	15	6			1	N	0
29. VEP008	47586	26/07/2023	\$2594.82	EUR	DAP	378942619409	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$4.870	6	3	6			1	N	0
30. VEP008	47587	26/07/2023	\$2594.82	EUR	DAP	378942619452	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$21.150	6	10	6			1	N	0
31. VEP008	47588	26/07/2023	\$2594.82	EUR	DAP	378942619491	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$24.040	6	10	6			1	N	0
32. VEP008	47589	26/07/2023	\$2594.82	EUR	DAP	378942619755	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$13.370	6	10	6			1	N	0
33. VEP008	47590	26/07/2023	\$2594.82	EUR	DAP	378942619725	POL	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$23.070	6	15	6			1	N	0
34. VEP008	47591	26/07/2023	\$2594.82	EUR	DAP	378942619741	CHN	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$10.000	5	5	6			1	N	0
35. VEP008	47592	26/07/2023	\$2594.82	EUR	DAP	378942619743	CHN	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$57.680	6	40	6			1	N	0
36. VEP008	47593	26/07/2023	\$2594.82	EUR	DAP	378942619743	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$18.590	6	10	6			1	N	0
37. VEP008	47594	26/07/2023	\$2594.82	EUR	DAP	378942619772	CHN	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$7.690	6	5	6			1	N	0
38. VEP008	47595	26/07/2023	\$2594.82	EUR	DAP	378942619770	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$21.060	6	10	6			1	N	0
39. VEP008	47596	26/07/2023	\$2594.82	EUR	DAP	378942619734	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$12.100	6	20	6			1	N	0
40. VEP008	47597	26/07/2023	\$2594.82	EUR	DAP	378942619707	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$28.940	6	20	6			1	N	0
41. VEP008	47598	26/07/2023	\$2594.82	EUR	DAP	378942619732	CHN	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$30.290	6	15	6			1	N	0
42. VEP008	47599	26/07/2023	\$2594.82	EUR	DAP	378942619753	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$15.380	6	10	6			1	N	0
43. VEP008	47600	26/07/2023	\$2594.82	EUR	DAP	378942619741	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$20.180	6	10	6			1	N	0
44. VEP008	47601	26/07/2023	\$2594.82	EUR	DAP	378942619783	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$22.230	6	20	6			1	N	0
45. VEP008	47602	26/07/2023	\$2594.82	EUR	DAP	378942619733	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$24.040	6	10	6			1	N	0
46. VEP008	47603	26/07/2023	\$2594.82	EUR	DAP	378942619733	CHN	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$15.380	6	10	6			1	N	0
47. VEP008	47604	26/07/2023	\$2594.82	EUR	DAP	378942619549	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$16.520	6	3	6			1	N	0
48. VEP008	47605	26/07/2023	\$2594.82	EUR	DAP	378942619529	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$8.180	6	5	6			1	N	0
49. VEP008	47606	26/07/2023	\$2594.82	EUR	DAP	378942619705	CHN	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$11.540	6	10	6			1	N	0
50. VEP008	47607	26/07/2023	\$2594.82	EUR	DAP	378942619705	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$16.350	6	10	6			1	N	0
51. VEP008	47608	26/07/2023	\$2594.82	EUR	DAP	378942619718	ESP	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$9.160	6	5	6			1	N	0
52. VEP008	47609	26/07/2023	\$2594.82	EUR	DAP	378942619717	CHN	ESP	4301930400	LIBROS IMPRESOS ENIDOMA	\$11.540	6	10	6			1	N	0
1400 rows - 1400 col - 1400 row																			

Figura 61 Layout creado manualmente.

El sistema de tráfico de algunas sucursales de la empresa, manejan a diario esta operación, por lo cual lo que fue una operación realizada manualmente en un documento de Excel donde el operador capturaba información minuciosa y variable renglón por renglón, donde el operador tenía la probabilidad de cometer errores humanos.

Resultados ante la automatización de la creación de un layout:

- ✓ Reducción de tiempo del empleado en la creación del documento.
- ✓ Llenado automático en lenguaje español consultado en la base de datos.
- ✓ Consulta a la base de datos automática.
- ✓ Aceleración del proceso de operación.

Resultados en reducción de errores humanos:

- ✓ Minimización de errores humanos o “de dedo”, debido a la captura manual de datos.
- ✓ Información precisa y autentica.
- ✓ Consistencia y precisión de la información extraída o consultada.

Impacto.

La realización de este proyecto “Extraccion de datos en facturas PDF para la creación de Layout” obtuvo un impacto significativo en la rapidez y la instantaneidad de la creación de documentos Excel con información de facturas de diversos proveedores:

Eficiencia operativa:

- ✓ El tiempo de operación ha reducido de manera considerable.
- ✓ La generación automatizada de layouts aplicando información diferente según el operador lo necesita, ha mantenido una escalabilidad en cuanto a la efectividad de operación.

Adaptabilidad y escalabilidad.

- ✓ El sistema es considerado para la adaptabilidad ante las necesidades cambiantes en la empresa y la aduana, permitiéndoles adaptar nuevas tecnologías que les ayuden una mayor efectividad ante la competencia.
- ✓ La escalabilidad asegura la creciente demanda un mayor desempeño por parte de los empleados y la aduana, lo que conlleva a una mayor confiabilidad a la empresa.

## **8. CONCLUSIONES Y RECOMENDACIONES.**

El desarrollo e implementación del proyecto “Extracción de datos en facturas PDF para la creación de Layout”, resultó una efectividad y diversos beneficios significativos para el proceso de operación en el área de tráfico de la Aduana. Gracias a su automatización de procesos que aplican diversas herramientas que nos ayudan a crear este proceso de manera rápida, efectiva y automática.

Eficiencia:

Gracias a las tecnologías modernas se concluyó cada una de las etapas con éxito, siendo así totalmente automatizadas, para evitar que el operador tenga que hacer cualquier cosa de manera manual.

Automatización:

Su capacidad de generar Layouts con información capturada en Excel, simplificó totalmente los procesos de los operadores.

Satisfacción del empleado/usuario:

La interacción con la interfaz intuitiva y fácil de usar, ha recibido una buena aceptación por parte de los empleados, aumentando su confiabilidad ante más clientes.

Monitoreo y actualizaciones:

Este sistema tiende a evolucionar en base a los requerimientos por parte de la aduana, por lo tanto, después de su desarrollo, se dispondrá a mejoras y actualizaciones constantes. También se aplicarán formatos de facturas de proveedores nuevos.

Limitaciones:

Ante la efectividad del sistema, funciona efectivamente con facturas que su información sea legible, entre los requerimientos de facturas son los siguientes:

- Facturas digitales no escaneadas.
- Facturas con tablas estructuradas.
- Información registrada en la base de datos.

## **9. COMPETENCIAS APLICADAS Y/O DESARROLLADAS.**

Se aplicaron diversas competencias técnicas y profesionales. Que aportaron una gran cantidad de habilidades y herramientas para hacer posible el funcionamiento del proyecto. A continuación, se adjuntan las competencias mas relevantes:

Competencias técnicas.

Software:

- Uso de principios de programación y estructuras de datos en Javascript.
- Uso de librerías/frameworks/IDE modernas, entre ellas, Visual Studio Code, NODE js, CSS3, AXIOS y PM2.
- Uso de JWT para las llaves o tokens de autenticación para acceder a un microservicio web.

Base de datos:

- Uso de motor Firebird para la consulta a la base de datos.
- Uso de gestor de base de datos DBeaver para realización de consultas.

Generación de documentos:

- Creación de archivos Excel.
- Escritura de documentos PDF importados al sistema mediante file-stream de Node JS.

Competencias profesionales.

Gestión de proyectos:

- Uso de la metodología Waterfall o de Cascada para la planificación de cada una de las etapas del proyecto.

- Seguimiento total del cronograma previamente detallado.
- Cumplimiento de cada etapa del proyecto.

Resolución de problemas:

- Identificación de cada uno de los problemas que se presentan en un proceso actual.
- Implementación de soluciones eficientes, confiables y efectivas para abordar el problema presentado.

Adaptabilidad:

- Incorporación de nuevas funcionalidades y mejoras al sistema, en base a recomendaciones de los empleados de operación.

## 10. FUENTES DE INFORMACION.

### Marco Teórico:

Gestión y generación de documentos.

<https://www.equisoft.com/es/glossary/document-management-document-generation>

### Tecnologías utilizadas:

HTML y CSS

<https://esdima.com/que-es-html-y-css/>

Javascript, Node JS

<https://blog.hubspot.es/website/que-es-javascript>

[https://www.w3schools.com/nodejs/nodejs\\_intro.asp](https://www.w3schools.com/nodejs/nodejs_intro.asp)

API de Servicios de Adobe PDF

<https://developer.adobe.com/document-services/apis/pdf-extract/>

Firebird y DBeaver

<https://dbeaver.io/>

<https://dbdb.io/db/firebird>

Sistemas CASA

[https://sistemascasa.com.mx/nosotros\\_mision.php](https://sistemascasa.com.mx/nosotros_mision.php)

Postman:

<https://openwebinars.net/blog/que-es-postman/>

### Metodología TDD:

[¿Qué es el TDD? | EDteam](#)

**Glosario:**

<https://www.lucushost.com/blog/que-es-node-js/#:~:text=entorno%2C%20%C2%BFverdad%3F-,Node.,cuando%20las%20conexiones%20se%20multiplican.>

<https://aws.amazon.com/es/what-is/api/#:~:text=API%20significa%20%E2%80%9Cinterfaz%20de%20programaci%C3%B3n,de%20servicio%20entre%20dos%20aplicaciones.>

<https://keepcoding.io/fundamentos-programacion/que-es-jwt/>

<https://blog.hubspot.es/website/que-es-css>

<https://visualstudio.microsoft.com/es/>

## 11. ANEXO.

### 11.1. Glosario

Node JS: Entorno de tiempo open source de código abierto, multiplataforma y que se ejecuta del lado del servidor.

API: “Interfaz de programación de aplicaciones”. En el contexto de las API, la palabra aplicación se refiere a cualquier software con una función distinta. La interfaz puede considerarse como un contrato de servicio entre dos aplicaciones. Este contrato define cómo se comunican entre sí mediante solicitudes y respuestas. La documentación de su API contiene información sobre cómo los desarrolladores deben estructurar esas solicitudes y respuestas.

JWT (Json Web Token): JWT es un objeto de JSON (notación de objeto de JavaScript), una herramienta de estándar abierto cuyo objetivo es establecer una transmisión de información entre dos o más campos. A partir de estos, se puede propagar información de forma segura y efectiva, que, además, es verificada, pues se firma de forma virtual. Este conjunto de información toma la referencia de web token, bajo el estándar abierto de JSON.

CSS: Hojas de estilo en cascada (Cascading Style Sheets). Es un lenguaje que maneja el diseño y presentación de las páginas web, es decir, como lucen cuando un usuario las visita.

Visual Studio Code: Es un editor de código fuente ligero pero eficaz que se ejecuta en el escritorio, ideal para desarrollo Web y aplicaciones multiplataforma.