

Desarrollo de módulos para la automatización de la extracción de certificados de los portales web de Colpensiones, Armada Nacional, Ministerio de Defensa, Fiduprevisora, Ejército Nacional y Fopep para el cliente Bayport de la empresa Xirius



Informe Final
Modalidad: Práctica Profesional

Juan Camilo Anacona Salazar
104617011656

Director: PhD. Hugo Armando Ordoñez Erazo
Co director: PhD(C). Manuel Alejandro Pastrana Pardo
Asesor de la empresa: Ing. Julián Solarte

Universidad del Cauca
Facultad de Ingeniería Electrónica y de Telecomunicaciones
Departamento de Sistemas
Ingeniería de Sistemas

TABLA DE CONTENIDO

1.	INTRODUCCIÓN.....	6
1.1	PLANTEAMIENTO DEL PROBLEMA.....	6
1.2	METODOLOGÍA	9
1.3	APORTES.....	10
1.4	OBJETIVOS.....	11
1.4.1	OBJETIVO GENERAL.....	11
1.4.2	OBJETIVOS ESPECÍFICOS	11
1.5	ORGANIZACIÓN DEL DOCUMENTO	11
2.	MARCO TEÓRICO.....	12
2.1	RPA.....	12
2.2	GOOGLE CLOUD FUNCTIONS.....	14
2.3	NODE.JS.....	14
2.3.1	PUPPETEER.....	15
2.3.2	AXIOS.....	15
2.3.3	CHEERIO	15
2.4	DOCKER.....	15
2.5	MARCOS AGILES PARA DESARROLLO DE AGILES.....	17
3.	ANÁLISIS DEL PROYECTO	19
3.1	Historias de usuario.....	19
3.1.1	Revisión del proceso para la extracción del certificado de pensión en el portal web Colpensiones.....	20
3.1.2	Revisión del proceso para la extracción de los certificados de ingresos y tiempos de servicio en el portal web Armada Nacional	22
3.1.3	Revisión del proceso para la extracción de los certificados de cupón de pago y descuentos en portal web Fopep	23
3.1.4	Revisión del proceso para la extracción de los certificados de mesada pensional y comprobante de nómina en el portal web Fiduprevisora	24
3.1.5	Revisión del proceso para la extracción del certificado de colilla de pago en el portal web Mindefensa	27
3.1.6	Revisión del proceso para la extracción de los certificados de tiempos de servicio e ingresos en el portal del Ministerio de defensa usando QR	27

3.1.7	Revisión del proceso para la extracción de los certificados de ingresos y tiempos de servicio en el portal web Ejército Nacional	28
3.2	Diseño de la solución	29
3.2.1	Diseño de los mecanismos para la obtención de certificados	31
4.	PROGRAMACIÓN DE ACTIVIDADES	32
4.1	PRIMER MES DE LA PRÁCTICA PROFESIONAL.....	32
4.2	SEGUNDO MES DE LA PRÁCTICA PROFESIONAL	34
4.3	TERCER MES DE LA PRÁCTICA PROFESIONAL	36
4.4	CUARTO MES DE LA PRÁCTICA PROFESIONAL	37
4.5	QUINTO MES DE LA PRÁCTICA PROFESIONAL	39
5.	DESARROLLO DE LOS MÓDULOS	41
5.1	Módulo para el portal web de Colpensiones.....	42
5.2	Módulo para el portal web de la Armada.....	44
5.3	Modulo para el portal web de Fopep	46
5.4	Módulo para el portal de Fiduprevisora Fomag	47
5.5	Módulo para el portal del Ministerio de Defensa.....	49
5.6	Módulo para el portal del Ministerio de Defensa Versión QR	50
5.7	Módulo para el portal del Ejército Nacional	51
5.8	Creación y despliegue del ambiente de producción.....	52
5.8.1	Despliegue de la API en Cloud Run.....	53
5.8.2	Configuración de la base de datos.....	56
5.8.3	Despliegue de los módulos de Automatización	57
5.8.4	Configuración de los Bucket	59
6.	LECCIONES APRENDIDAS.....	60
7.	CONCLUSIONES	62
8.	REFERENCIAS BIBLIOGRÁFICAS	63

LISTA DE FIGURAS

Figura 1. Diagrama del proceso de solicitud de crédito de Bayport.....	7
Figura 2. Roles en RPA tomado de [9].....	13
Figura 3. Arquitectura de una Máquina Virtual.....	16
Figura 4. Arquitectura de un Contenedor.....	17
Figura 5. Marco de trabajo Scrum tomado de [23].....	19
Figura 6. Inicio de sesión en el portal de Colpensiones	20
Figura 7. Menú de certificados en el portal de Colpensiones	21
Figura 8. Formulario para los datos del certificado del portal de Colpensiones.....	21
Figura 9. Formulario para la descarga del certificado.....	22
Figura 10. Formulario para obtener los certificados de la Armada.....	23
Figura 11. Formulario para obtener los certificados de Fopep	24
Figura 12. Ventana de confirmación para obtener los certificados de Fopep.....	24
Figura 13. Inicio de sesión en el portal de Fiduprevisora	25
Figura 14. Menú de opciones del portal de Fiduprevisora	25
Figura 15. Formulario para solicitar el certificado de mesada pensional Fiduprevisora	26
Figura 16. Formulario para solicitar el certificado de comprobante de nómina Fiduprevisora	26
Figura 17. Pestaña con el certificado en Fiduprevisora.....	26
Figura 18. Formulario para la solicitud del certificado Mindefensa	27
Figura 19. Datos dentro del portal al abrir un código QR Mindefensa.....	28
Figura 20. Datos dentro del portal al abrir un código QR Mindefensa.....	28
Figura 21. Formulario para obtener el certificado del Ejercito Nacional.....	29
Figura 22. Vista de despliegue.....	30
Figura 23. Estructura de respuesta de los módulos.....	42
Figura 24. Diagrama de secuencia - Obtención de certificado Colpensiones	44
Figura 25. Captcha de la Armada	45
Figura 26. Diagrama de secuencia - Obtención de certificado Armada	46
Figura 27. Diagrama de secuencia - Obtención de certificado Fopep.....	47
Figura 28. Diagrama de secuencia - Obtención de certificado Fiduprevisora	49
Figura 29. Diagrama de secuencia - Obtención de certificado Ministerio de Defensa	50
Figura 30. Diagrama de secuencia - Obtención de certificado Ministerio de Defensa versión QR	51
Figura 31. Captcha del portal web de Ejercito Nacional.....	52
Figura 32. Configuración región del servicio Cloud Run	54
Figura 33. Configuración asignación de CPU en el servicio de Cloud Run	55
Figura 34. Configuración de las variables de entorno y conexión a la base de datos.....	55
Figura 35. Configuración nombre y contraseña de la instancia de Base de Datos	56
Figura 36. Configuración de la Región de la Base de Datos	56
Figura 37. Configuración de la máquina y cantidad de almacenamiento de la Base de Datos ..	57
Figura 38. Configuración de la región y el activador de las Cloud Functions	58
Figura 39. Configuración memoria y tiempo de ejecución Cloud Functions	58
Figura 40. Implementación de la lógica de programación en las Cloud Functions.....	59
Figura 41. Configuración de Buckets	59

LISTA DE TABLAS

Tabla 1 - Mecanismos para la obtención de certificados	31
Tabla 2 - Primer mes de práctica: actividades referentes a las reuniones en el mes.....	33
Tabla 3 - Primer mes de práctica: actividades referentes a la capacitación del practicante.....	33
Tabla 4 - Primer mes de práctica: actividades referentes al análisis del proyecto.....	33
Tabla 5 - Segundo mes de práctica: actividades referentes a las reuniones en el mes.....	34
Tabla 6 - Segundo mes de práctica: actividades referentes a la construcción de los módulos de Colpensiones y la Armada	34
Tabla 7 - Tercer mes de práctica: actividades referentes a las reuniones en el mes	36
Tabla 8 - Tercer mes de práctica: actividades referentes a la construcción de los módulos de Fopep, Fiduprevisora y Mindefensa	36
Tabla 9 - Cuarto mes de práctica: actividades referentes a las reuniones en el mes.....	37
Tabla 10 - Cuarto mes de práctica: actividades referentes a la construcción de los módulos de Mindefensa QR y Ejercito Nacional.....	38
Tabla 11 - Quinto mes de practica: actividades referentes a las reuniones en el mes.....	39
Tabla 12 - Cuarto mes de práctica: actividades referentes a ajustar los módulos de Mindefensa QR y Ejercito Nacional además del despliegue en producción	39
Tabla 13 - Resumen del desarrollo de los módulos	40
Tabla 14. Servicios para la resolución de captchas	45

1. INTRODUCCIÓN

1.1 PLANTEAMIENTO DEL PROBLEMA

En la actualidad, la tecnología ha transformado la manera en que las personas viven, trabajan y se relacionan, al igual que sucede con las empresas e instituciones y su relación con sus clientes, principalmente establecida por medio digital de acuerdo con [1]. Esta transformación digital ha ocasionado que para que una empresa sea competitiva, no solo dependa de la calidad de su producto o servicio, sino que también de como utiliza las tecnologías, ya que su uso adecuado permite agregar valor a sus procesos de comunicación, comercialización y en general con toda la cadena de valor como sugiere [2]. Un ejemplo de ello es la posibilidad de solicitar una tarjeta de crédito sin necesidad de acudir presencialmente al banco, realizando él envío de los documentos requeridos (extractos laborales, historias créditos, entre otros) por medio del portal web del banco, para que un empleado analice la información, complete formularios y finalice el trámite. Este tipo de interacciones implican una gran cantidad de trabajo manual, lo que involucra tiempo y esfuerzo para las empresas, que puede ser utilizado en tareas de mayor importancia. Por lo anterior, es posible hacer uso de la tecnología para automatizar labores simples a través de RPA ¹o automatización robótica de procesos, que permite mediante software replicar actividades manuales, repetitivas y basadas en reglas que no requieren la toma de decisiones complejas y que son comunes en los procesos que involucran intervención humana [3].

Bayport es una empresa que ofrece créditos de libranza a pensionados y empleados del sector público en Colombia. Para tramitar un crédito, el solicitante debe presentar a Bayport unos documentos requeridos por el proceso, donde destacan los certificados de ingresos o la afiliación laboral; que son requeridos por la empresa para determinar si se puede acceder o no al crédito. Durante el proceso de solicitud, Bayport valida la veracidad de la información accediendo de manera manual a los portales web que gestionan la información de pensión o afiliación laboral del solicitante de crédito, de esta manera pueden establecer si los certificados presentados por el solicitante son válidos o por el contrario han sido adulterados. Este es un proceso que se repite cada vez que se realiza una solicitud de crédito, lo que conlleva a un gran esfuerzo de tiempo y dinero para Bayport. La **Figura 1** especifica el proceso para la solicitud de crédito mediante un Modelo de Procesos de Negocio (BPMN).

¹ RPA: Robotic Process Automation

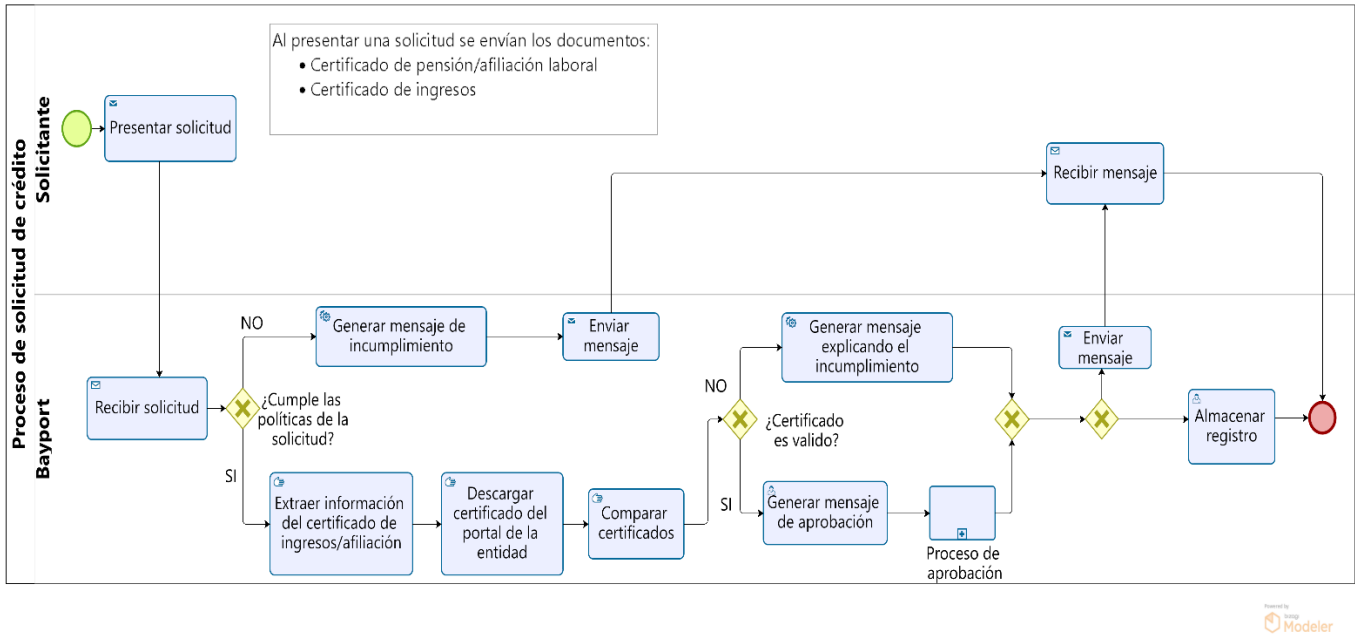


Figura 1. Diagrama del proceso de solicitud de crédito de Bayport

Con el fin de optimizar el proceso manual, Bayport optó por contratar los servicios de Xirius Tech S.A.S, empresa dedicada al desarrollo de software para que le brindará una solución de automatización que permitiera: extraer la información de un certificado, obtenerlo del portal web correspondiente y validar su veracidad del tal forma que un empleado no tenga que llenar un formulario con la información de los tres anteriores procesos sino más bien que al subir el certificado del solicitante a su aplicativo, los formularios con la información de los procesos se llenen de manera automática. Teniendo en cuenta todo lo anterior Xirius dividió el problema en tres 3 etapas que son:

1. Extracción de la información contenida en el certificado proporcionado por el solicitante mediante reconocimiento óptico de caracteres (OCR)².
2. Obtención del certificado de manera automática, cuya finalidad es hacer uso de módulos de automatización para obtener el certificado proporcionado por el solicitante a través de su respectivo portal web.
3. Validación, usando la información extraída de los certificados en las etapas uno y dos al compararlos.

² OCR (Optical Character Recognition): es una tecnología que permite convertir diferentes tipos de documentos, tales como documentos escaneados, archivos de PDF o imágenes en datos: <https://www.ibm.com/blog/optical-character-recognition/>

Xirius encontró a través de OCR la manera de dar cumplimiento a la primera fase (extracción). Sin embargo, se necesita dar solución a la segunda fase (obtención de certificados). Este proyecto se enfoca en brindar una solución que automatice el proceso de obtención de los certificados (desprendibles de pago o afiliación) desde los portales web de la empresa a la que pertenece cada solicitante de crédito mediante servicios basados en la nube. La obtención del certificado de manera automática se hará aplicando técnicas de:

- Bots³: De acuerdo con [4] son programas informáticos constituidos por un conjunto de reglas que se ejecutan en internet de manera repetitiva efectuando tareas recurrentes.
- Web Scraping: De acuerdo con [5] el objetivo principal del Web Scraping es extraer información de uno o varios sitios web y procesarlo en estructuras simples como hojas de cálculo, base de datos o JSON. El proceso de Web Scraping se divide en tres etapas según [6] que son:
 - Fetching: Se accede al sitio web deseado que contiene los datos relevantes. Esto se hace a través del protocolo HTTP, un protocolo de Internet que se utiliza para enviar solicitudes y recibir respuestas de un servidor web.
 - Extracción: Una vez que se obtiene el documento HTML, se extraen los datos de interés, esto mediante librerías que permiten interactuar con el HTML.
 - Transformación: Una vez hecha la extracción se transforman los datos a una versión estructurada.
- Clientes REST: Los servicios web como REST permiten la interoperación de sistemas distribuidos haciendo posible que sean usados a través de la infraestructura proporcionada por internet [7]; el cliente REST hace posible gestionar el uso de estos servicios.

Se decide adoptar el desarrollo de automatizaciones robóticas, buscando una mejora en la productividad y eficiencia, puesto que ya existen casos en los que se registran los beneficios de la aplicación de la automatización de procesos. Por ejemplo, Telefónica O2, el cual es un proveedor de servicios de telecomunicaciones británico, en abril de 2015 implementó más de 160 robots capaces de procesar entre 400 mil y 500 mil transacciones por mes, brindando apoyo al área de oficinas administrativas, lo que generó un retorno de la inversión a tres años de entre el 650 y 800 por ciento como indica [8].

³ Bot: Aféresis de “robots”

Por lo anterior, la empresa Xirius Tech ha decidido incorporar a un practicante que lleve a cabo la implementación de módulos que brinden las funcionalidades para automatizar la obtención de certificados haciendo uso de tecnologías como Node.js y los servicios proporcionados por Google Cloud, específicamente Google Cloud Functions, siguiendo la metodología de la empresa la cual toma principios del marco de trabajo Scrum como reuniones de seguimiento y planeación.

Finalmente, es importante destacar que la realización de esta práctica profesional permitirá al estudiante de ingeniería de sistemas reforzar y desarrollar habilidades además de ampliar sus conocimientos relacionados con el desarrollo de software, desarrollo en la nube, automatización robótica de procesos y como brindar soluciones a problemas en un ámbito real.

1.2 METODOLOGÍA

Para el desarrollo de la práctica profesional se hizo uso de una metodología proporcionada por la empresa, que agrupa las actividades en tres etapas: 1) Análisis, 2) Desarrollo de software y 3) Documentación. A continuación, se exponen dichas etapas las cuales se llevaron a cabo durante la práctica profesional para cumplir con los objetivos planteados.

- Etapa de análisis:

Para dar inicio con el desarrollo de las funcionalidades requeridas fue necesario iniciar conociendo la solución planteada por Bayport, su arquitectura, la comunicación entre sus componentes, esto para tener un mejor contexto sobre cómo funcionará todo el sistema. Por otra parte, también fue necesario analizar las historias de usuario proporcionadas por Xirius, donde se encuentran los portales web y las instrucciones que siguen los trabajadores de Bayport para obtener los certificados. Para esto, la empresa Xirius brindó apoyo al practicante guiándolo en las tecnologías que fueron utilizadas; Node.js, JavaScript, Puppeteer, Axios y los servicios de Google Cloud.

- Principales actividades:

- Conocer el funcionamiento de la solución.
- Conocer las interacciones entre los componentes.
- Analizar las historias de usuario.
- Capacitarse en las tecnologías a usar.

- Etapa de desarrollo de software:

Una vez cumplida la etapa de análisis, el practicante inició la fase de desarrollo de los requisitos de cada uno de los módulos para la automatización solicitados. Esta etapa se llevó a cabo bajo la metodología de la empresa Xirius, la cual está basada en el marco de trabajo ágil Scrum. Por lo tanto, se tomaron algunos de sus elementos, como el *Sprint review*, reunión con el equipo para analizar lo que se realizó durante el Sprint, esto

teniendo en cuenta que un Sprint para la empresa Xirius está definido en un tamaño de una semana.

- Principales actividades:
 - Construcción del módulo de Colpensiones
 - Construcción del módulo de la Armada Nacional
 - Construcción del módulo de Fopep
 - Construcción del módulo de Fiduprevisora
 - Construcción del módulo de Ministerio de defensa
 - Construcción del módulo de Ministerio de defensa QR
 - Construcción del módulo de Ejército Nacional

- Etapa de documentación:

El avance del desarrollo fue reportado en la herramienta Jira, utilizada por la empresa Xirius para la gestión de proyectos, donde se evidencia el trabajo realizado durante el transcurso de la práctica. También se llevaron a cabo reuniones periódicas con el tutor de la universidad y el asesor de la empresa, evidenciando el cumplimiento de los objetivos planteados. Por último, se elaborará un documento como monografía donde se detallará todo lo realizado durante la práctica profesional.

1.3 APORTES

El desarrollo de la práctica profesional dentro de la empresa Xirius Tech generó aportes valiosos, principalmente en las habilidades técnicas obtenidas, tales como el consumo de servicios mediante una API REST, la separación e interacción de los componentes para que en conjunto brinden una solución y el manejo de los servicios ofrecidos por Google Cloud para el despliegue de aplicaciones, con todo lo que ello implica. Esto teniendo como base el conocimiento en programación, algoritmia manejo de base de datos y demás, brindado en la formación del programa académico. Además, la empresa puso empeño en que se adquirieran habilidades para brindar soluciones basadas en la nube.

Desde el punto de vista académico, este trabajo aspira dar a conocer la calidad y entrega de los estudiantes formados en el programa de ingeniería de sistemas de la universidad de Cauca, no solamente por el desarrollo de este proyecto, sino también porque Xirius Tech SAS es una empresa creada por egresados del programa de sistemas, lo que da a conocer a nivel regional y nacional la calidad de los ingenieros que se forman en esta universidad.

Por último, en el ámbito empresarial, el producto software generado en esta práctica beneficiará la productividad de la empresa Bayport, agilizando el proceso de solicitud de crédito, lo que genera un impacto positivo tanto en los usuarios como en los empleados de Bayport. Por ejemplo, los aspirantes a un crédito pueden obtener respuesta del estado de su trámite en mucho menos tiempo que antes, y los empleados se liberan de la carga de trabajo tedioso y pueden enfocar su tiempo y esfuerzo en procesos más complicados y de mayor importancia.

1.4 OBJETIVOS

1.4.1 OBJETIVO GENERAL

Desarrollar módulos de automatización que permitan obtener el certificado laboral, de colilla de pago, de pensión, de tiempos y haberes (civiles y uniformados), de cupón de pago y de mesada pensional, mediante servicios basados en la nube (Cloud Functions) con la finalidad de aumentar la productividad de la empresa Bayport eliminando el trabajo manual y repetitivo.

1.4.2 OBJETIVOS ESPECÍFICOS

- Diseñar los mecanismos de web scraping, bots o clientes REST teniendo en cuenta las características de los portales web de la Armada Nacional, Colpensiones, Fopep, Fiduprevisora fomag, Ministerio de defensa y Ejército Nacional, analizando su comunicación, estructura y complejidad al interactuar.
- Implementar los diferentes módulos de automatización teniendo en cuenta el diseño realizado haciendo uso de Google Cloud Functions para poder obtener los certificados de los portales web.
- Evaluar los módulos implementados a través de casos de prueba realizados por el equipo de pruebas de Bayport y entregados a Xirius para determinar que los módulos cumplan con el comportamiento esperado.

1.5 ORGANIZACIÓN DEL DOCUMENTO

A continuación, se describe de manera general el contenido y organización del presente informe final:

CAPÍTULO 1: INTRODUCCIÓN: Hace referencia al presente capítulo donde se describe el planteamiento del problema, los aportes generados y los objetivos a cumplir.

CAPÍTULO 2: MARCO TEÓRICO: En este capítulo se describen los conceptos teóricos más relevantes que se emplearon para la realización del trabajo de grado.

CAPÍTULO 3: ANÁLISIS DEL PROYECTO: En este capítulo se describen las características técnicas del aplicativo, tales como, el diseño de cada módulo, los servicios de terceros entre otros, así mismo una caracterización del sistema identificando las arquitecturas aplicadas.

CAPÍTULO 4: PROGRAMACIÓN DE ACTIVIDADES: En este capítulo se detalla las actividades para satisfacer los requerimientos y el tiempo de su elaboración.

CAPÍTULO 5: DESARROLLO DE LOS MÓDULOS: En este capítulo se detalla el proceso para llevar a cabo los objetivos planteados y las funcionalidades creadas.

CAPÍTULO 8: LECCIONES APRENDIDAS: En este capítulo se describen las lecciones aprendidas a partir del desarrollo de esta práctica profesional.

CAPÍTULO 7: CONCLUSIONES: En este capítulo se describen las conclusiones generadas a partir del desarrollo de la práctica profesional

CAPÍTULO 8: BIBLIOGRAFÍA: En el último capítulo del presente documento se listan las referencias bibliográficas de los artículos, libros y demás recursos utilizados para el desarrollo del trabajo de grado.

2. MARCO TEÓRICO

2.1 RPA

La automatización robótica de procesos (RPA por sus siglas en inglés) es una tecnología que permite automatizar tareas repetitivas dentro de una organización mediante la construcción de robots. En este caso la palabra robot no hace referencia a una máquina como las usadas en un proceso de ensamblaje en masa, sino en una pieza de software cuya finalidad es realizar actividades mecánicas y bien definidas, como, por ejemplo, enviar un correo o llenar un formulario esto con la finalidad de ayudar al empleado eliminando la carga operativa de las tareas reiterativas para que pueda brindar su atención a tareas de mayor complejidad como sugiere [8].

Existen dos tipos de automatización robótica de procesos:

- **Atendida:** Es aquella cuyos robots requieren de la activación humana para poderse ejecutar y es común que un empleado intervenga en momentos específicos, permitiendo que mientras el robot se ejecuta el empleado pueda realizar otras actividades.
- **Desatendida:** Es aquella cuyos robots se ejecutan sin intervención humana y son programados para ejecutarse a ciertas horas. Por ejemplo, ejecutarse a las 15:45 PM todos los días para crear un reporte; por lo general se usan en tareas administrativas según [9].

Esta tecnología es versátil por lo tanto puede ser usada para automatizar varias tareas en diferentes sectores, como son el sector bancario, las auditorías o los seguros. Las áreas dentro de una empresa en las que se puede usar por lo general son:

- Ventas
- Recursos humanos
- Área operacional
- Servicio al cliente

RPA ayuda a las empresas creando una fuerza de trabajo virtual que brinda la capacidad de ganar una ventaja competitiva en la satisfacción del cliente y la agilidad empresarial, estos robots ofrecen una gran cantidad de beneficios que incluyen:

- **Precisión:** RPA puede mejorar significativamente la precisión de las operaciones comerciales, ya que son impermeables al error humano.
- **Consistencia:** RPA puede automatizar cualquier proceso comercial que sea repetible y basado en reglas.
- **Reducción de costos:** Los bots RPA cuestan mucho menos que un empleado de tiempo completo. Al implementar bots para automatizar procesos comerciales simples, RPA puede reducir significativamente los costos operativos.
- **Cumplimiento mejorado:** Los bots RPA ejecutan procesos según las instrucciones que han sido configuradas para seguir y proporcionan un registro para cada paso. Esta naturaleza controlada de los bots aumenta la transparencia y elimina el fraude.
- **Informes mejorados:** RPA genera una cantidad significativa de datos que permite a las organizaciones analizar e identificar ineficiencias en los procesos. Esta información operativa proporcionada por los bots permite a las organizaciones optimizar los procesos comerciales existentes.
- **Calidad mejorada:** RPA permite a las empresas optimizar y estandarizar sus procesos, lo que reduce los errores de datos. Esta reducción de errores conduce a datos de alta calidad que permiten un análisis más fiable de acuerdo con [10].

Para que una empresa pueda integrar RPA dentro de sus procesos de negocio se necesita adaptar tanto su estructura tecnológica como de personal, en términos de la organización aparecen cinco roles según [9]. (Figura 2)

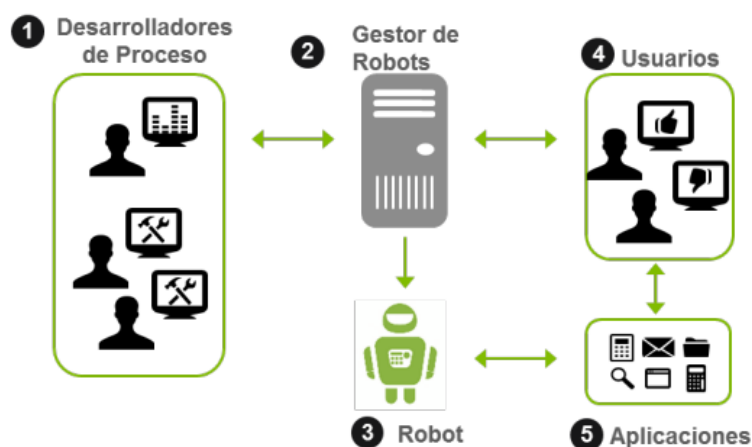


Figura 2. Roles en RPA tomado de [9]

Los roles presentados son a detalle:

- I. **Desarrolladores de proceso:** Son quienes conocen las tareas que debe desarrollar el RPA.
- II. **Gestor de robots:** Quien asigna y monitorea los robots.

- III. Robot: Es el software que interactúa con las aplicaciones del negocio.
- IV. Usuarios: Aquellos que resuelven las incidencias que el robot reporta.
- V. Aplicaciones: Es la plataforma o aplicación mediante la cual el robot interactúa con el usuario.

2.2 GOOGLE CLOUD FUNCTIONS

Como se describe en la documentación [10], Google Cloud Functions es un entorno de ejecución sin servidores, para los cuales Google se encarga en su totalidad de gestionar la infraestructura, configurar software, actualizar frameworks o actualizar el sistema operativo, de esta manera los desarrolladores no deben preocuparse por administrar servidor sino únicamente por crear el código, otra característica importante es que cuentan con una capa conectora que permite interactuar con la mayoría de servicios en la nube, como el almacenamiento en Cloud Storage, o responder a mensajes si se trabaja con Pub/Sub⁴. Por otra parte, debido a su naturaleza las Cloud Functions son herramientas idóneas para la creación de APIs⁵ ligeras, estas pueden ser invocadas a través de HTTP y entre sus otros usos están, el procesamiento de datos, al poder reaccionar a eventos en Cloud Storage, como lo son la subida o edición de archivos y procesar ya sea imágenes o codificar video, así mismo son de gran uso en IoT al poder manejar cientos de miles de mensajes provenientes de dispositivos haciendo uso del Pub/Sub.

2.3 NODE.JS

Node.js, también llamado Node es un entorno de ejecución para JavaScript que se ejecuta del lado del servidor de acuerdo con [11], está basado en el motor de código abierto V8⁶ de Google, el cual está escrito en C++ y soporta principalmente JavaScript en el navegador.

Gracias a Node.js JavaScript se convirtió en el lenguaje de programación más popular en la Web de acuerdo con [12], que JavaScript junto con Node.js hayan tenido tanto éxito es debido, entre otras razones al equilibrio entre la productividad del desarrollador y el rendimiento de la aplicación, permitiendo a los desarrolladores crear rápidamente aplicaciones listas para salir a producción utilizando miles de módulos disponibles a través del repositorio de paquetes NPM según [13].

⁴ Pub/Sub: Es un patrón de mensajería utilizado en sistemas distribuidos que permite la comunicación asíncrona entre componentes.

⁵ API (Application Programming Interface): Las API son mecanismos que permiten a dos componentes de software comunicarse entre sí mediante un conjunto de definiciones y protocolos.

⁶ Información obtenida de la documentación oficial de V8: <https://v8.dev/>

2.3.1 PUPPETEER

Puppeteer es una librería de Node.js para automatizar el navegador la cual provee una API que permite controlar Google Chrome o Chromium. Puppeteer permite automatizar tareas repetitivas como dar click en un botón, llenar formularios, tomar capturas de pantalla o generar PDFs de acuerdo con [14].

Con Puppeteer los desarrolladores pueden generar scripts de automatización de tareas simulando la interacción que ocurre entre una persona y el navegador, esto la hace una librería ideal para automatizar pruebas, realizar WebScraping y en general cualquier proceso que requiera de automatizar tareas dentro de un navegador web.

2.3.2 AXIOS

Axios es un cliente HTTP para Node.js usado para realizar peticiones y gestionar las respuestas en aplicaciones web y móviles. Proporciona una API de alto nivel para enviar y recibir los datos de peticiones HTTP además de que ofrece la posibilidad de cancelar las peticiones, transformar las respuestas, hacer autenticaciones o manejar errores lo que lo convierte en una opción popular a la hora de manejar peticiones según [15].

2.3.3 CHEERIO

Cheerio es una librería de Node.js para analizar y manipular HTML, con esta librería es posible manipular el HTML como si se trabajara con un documento DOM en el navegador web como menciona [16]. Cheerio proporciona una API para la manipulación lo que permite que de manera sencilla se puedan extraer títulos, párrafos imágenes y demás de una página web; estas características hacen que Cheerio sea útil para la automatización de tareas que necesite de WebScraping lo que ahorra recursos.

2.4 DOCKER

De acuerdo con [17] docker es una plataforma de código abierto en cual se ejecutan aplicaciones y facilita los procesos de desarrollo y distribución del software. Las aplicaciones construidas en docker se empaquetan junto con todas sus dependencias en lo que se denomina contenedor, estas aplicaciones virtualizadas pueden trabajar en diferentes entornos sin ninguna alteración, estos contenedores pueden desplegarse fácilmente en entornos basados en la nube. La forma en la que Docker está diseñado es para brindar un entorno donde el código pueda ejecutarse de manera eficiente.

La virtualización es un concepto antiguo, que se usa bastante en la computación en nube, donde se hace uso de IaaS⁷ para la construcción de sistemas, sin embargo, una máquina virtual y un contenedor son dos conceptos diferentes, la figura muestra la arquitectura de una máquina virtual.

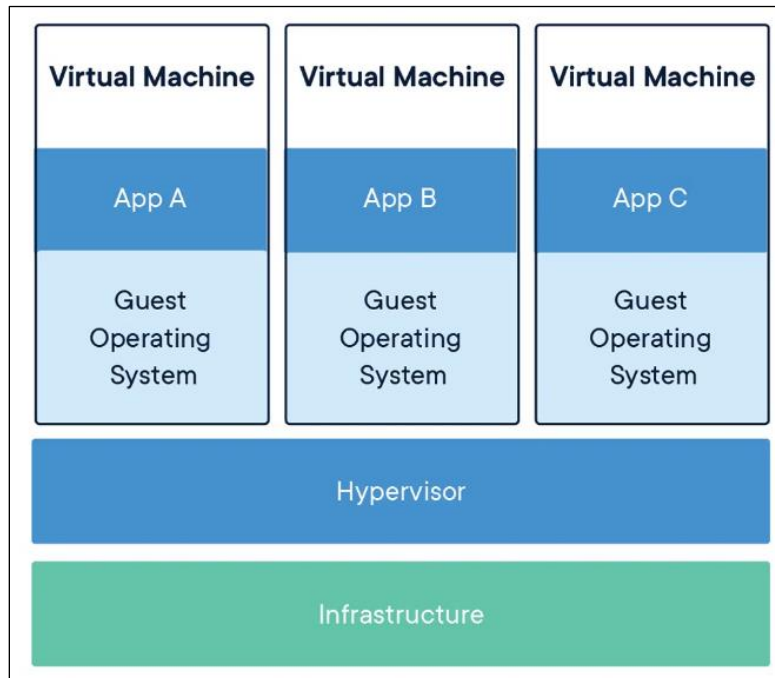


Figura 3. Arquitectura de una Máquina Virtual

Una máquina virtual (VM) es una abstracción del hardware físico que convierte un servidor en muchos servidores. El Hypervisor permite ejecutar varias máquinas virtuales dentro de una sola máquina, cada VM incluye una copia completa de un sistema operativo, la aplicación junto con sus dependencias según [18].

⁷ IaaS: es un modelo de servicio en la nube que ofrece recursos de infraestructura bajo demanda, como computación, almacenamiento, redes y virtualización, a empresas y particulares a través de la nube.

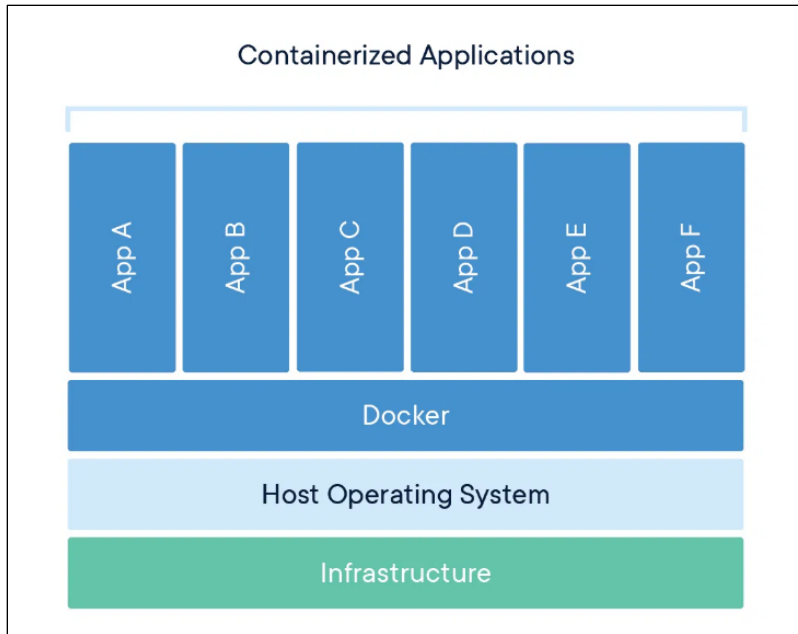


Figura 4. Arquitectura de un Contenedor

Un contenedor es una atracción en la capa de aplicación que empaqueta el código y las dependencias, múltiples contenedores pueden ejecutarse en la misma maquina y compartir el sistema operativo con otros contenedores. Cada contenedor se ejecuta como un proceso aislado [18].

2.5 MARCOS AGILES PARA DESARROLLO DE AGILES

Los marcos ágiles han cobrado relevancia en la industria del desarrollo de software y es que se caracterizan principalmente por: adaptarse a los cambios, presentar entregas de manera incremental y una activa participación del cliente, por otra parte, éstas metodologías han demostrado ser útiles en proyectos cuyos equipos de trabajo son pequeños, rondando entre siete y nueve personas como indica [19].

En el desarrollo ágil de software las interacciones y el personal son más importantes que el proceso y las herramientas, un software funcional es más importante que una documentación completa, la comunicación con el cliente es más importante que el contrato y responder a los cambios es más importante que ceñirse a un plan, esto genera como consecuencia que el proceso de desarrollo sea tolerante a los cambios de requisitos y la respuesta a estos sea más rápida de acuerdo con [20].

Uno de los marcos ágiles de trabajo de mayor popularidad es scrum, el cual es un marco para proyectos de software y desarrollo de aplicaciones, donde la atención se centra en el desarrollo de productos de manera flexible y el equipo de desarrollo trabaja como una unidad para lograr objetivos comunes como indica [20].

Como indica [21] Scrum requiere de un equipo con los roles de:

- Scrum Master: es el responsable de hacer que los miembros del equipo conozcan y sigan los principios del marco de trabajo.
- Product Owner: es el responsable de maximizar el valor y entregar el mejor producto posible.
- Developers: son las personas del equipo Scrum que se comprometen a crear nuevas características utilizables que permitan llegar al objetivo del producto en cada Sprint, los Sprints son el eje central de Scrum y consisten en unidades de trabajo que se necesitan para alcanzar un requerimiento.

Utilizando el ciclo de desarrollo expuesto por [22] se presenta a continuación un empalme entre las etapas del proceso, los artefactos y eventos que se encontrarían en una metodología basada en scrum. Aunque cabe aclarar que, al ser un marco de trabajo las empresas pueden optar por incluir o no incluir algunos de los elementos:

Análisis y planeación:

- Artefactos:
 - Product Backlog: Lista con los detalles de características requeridas para la construcción del proyecto
 - Sprint Backlog: Lista con las características asignadas para ser completadas durante cada sprint.
 - Tablero Scrum: Herramienta para visualizar el flujo de trabajo y facilitar la sincronización diaria del equipo durante el proyecto.
- Eventos de Scrum:
 - Planning meeting: Reunión con el equipo para determinar qué elementos del backlog se realizarán y preparar el próximo sprint.

Diseño:

- Artefactos:
 - Documentos DAS (Desarrollo Adaptativo de Software) : Vistas de arquitectura de software con sus diagramas UML y atributos de calidad observables y no observables.

Construcción y calidad:

- Artefacto
 - Código fuente.
 - Scripts de base de datos.
- Eventos de SCRUM.
 - Daily Scrum: Reunión diaria de aproximadamente 15 minutos donde el equipo expone lo realizado del sprint.

Despliegue:

- Artefactos:
 - Unidad desplegable
 - Manual de usuario
 - Manual de despliegue
- Eventos de scrum
 - Sprint review: reunión con el equipo para analizar lo que se realizó durante el sprint.
 - Sprint retrospective: reunión con el equipo que se lleva a cabo al final de un sprint y se utiliza para analizar qué salió bien durante el sprint anterior y qué se puede mejorar para el próximo sprint.

La **Figura 5**, expone de manera grafica cómo funciona Scrum junto con todos los eventos y artefactos que lo componen.

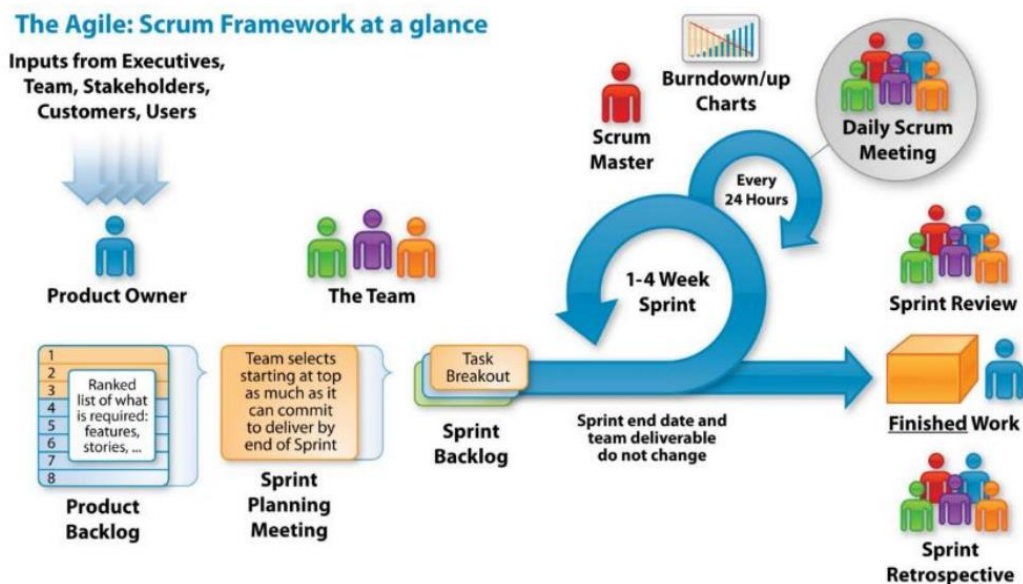


Figura 5. Marco de trabajo Scrum tomado de [23]

3. ANÁLISIS DEL PROYECTO

3.1 Historias de usuario

Como primer paso para iniciar con este proyecto se realizó el proceso de interpretación de las historias de usuario, las cuales fueron proporcionadas por el área de requisitos al practicante, esto con el fin de dejar en claro todos los requerimientos, así como los pasos que sigue un trabajador para obtener un certificado y así definir estrategias para automatizar este proceso, cabe mencionar que los datos solicitados por los portales son proporcionados por el solicitante al iniciar con la solicitud de crédito.

Además de conocer los pasos que realiza un trabajador para obtener un certificado, se hizo análisis de las peticiones que realizan los portales web a través de las herramientas del navegador con el fin de encontrar el uso de APIs que facilitara la obtención del certificado, así como características que pudieran influir en el desarrollo de los módulos como el uso de desafíos captcha.

3.1.1 Revisión del proceso para la extracción del certificado de pensión en el portal web Colpensiones

Para el primer módulo, el portal web de Colpensiones(https://www.colpensiones.gov.co:8070/zonat/Inicio_zona_trans.aspx) presenta un formulario destinado al inicio de sesión dentro de la plataforma (**Figura 6**), donde se requieren los campos de tipo de identificación, número de identificación y contraseña, un factor que cabe resaltar es la latencia e inestabilidad del sitio ya que no siempre está disponible y puede llegar a tener unos tiempos de espera elevados.



Figura 6. Inicio de sesión en el portal de Colpensiones

Posterior a ello el proceso continúa desplazándose por un menú lateral en donde se encuentra la opción de certificado de pensión como se muestra a continuación en la **Figura 7**.

Inicio > blanca bustillo

Certificados ▾

- Certificado de Afiliación
- Certificado de Deducidos y Devengados
- Certificado de EPS
- Certificado de Indemnización
- Certificado de Ingreso y retenciones por rentas
- Certificado de no pensión
- Certificado de Pensión
- Certificado de Vinculación BEPS
- Envía tus certificados de pensionado a tu correo electrónico

Consultas ▾

En esta vista encontrará las últimas diez transacciones realizadas en los trámites.

No.	Trámite	Fecha de operación	Estado	Acción
1	Certificado de Pensión	2023/03/14 04:42 PM	Terminado	👁
2	Certificado de Pensión	2023/03/14 04:41 PM	Terminado	👁
3	Certificado de Pensión	2023/03/07 09:03 AM	Terminado	👁
4	Certificado de Pensión	2023/03/07 09:01 AM	Terminado	👁
5	Certificado de Pensión	2023/02/08 11:13 AM	Abandonado	
6	Certificado de Pensión	2023/01/25 02:56 PM	Terminado	👁
7	Certificado de Pensión	2023/01/25 02:55 PM	Terminado	👁
8	Certificado de Pensión	2022/12/14 10:38 AM	Terminado	👁
9	Certificado de Pensión	2022/12/14 01:19 PM	Terminado	👁
10	Certificado de Pensión	2022/12/14 01:12 PM	Terminado	👁

Figura 7. Menú de certificados en el portal de Colpensiones

Después de seleccionar la opción de certificado de pensión se genera un formulario que solicita los datos con los cuales le certificado será generado, como indica la **Figura 8**.

Inicio > Mis trámites > Certificado de Pensión

Trámites y servicios

Certificados ▾

- Certificado de Afiliación
- Certificado de Deducidos y Devengados
- Certificado de EPS
- Certificado de Indemnización
- Certificado de ingreso y retenciones por rentas
- Certificado de no pensión
- Certificado de Pensión
- Certificado de Vinculación BEPS
- Envía tus certificados de pensionado a tu correo electrónico

Certificado de Pensión

Información básica

Tipo de documento *

Número de identificación *

Período *

Requiere valores *

[Continuar](#)

Figura 8. Formulario para los datos del certificado del portal de Colpensiones

Finalmente, de muestra el ultimo formulario que ya contiene la opción para descargar este certificado en el ordenador, así como indica la **Figura 9**.

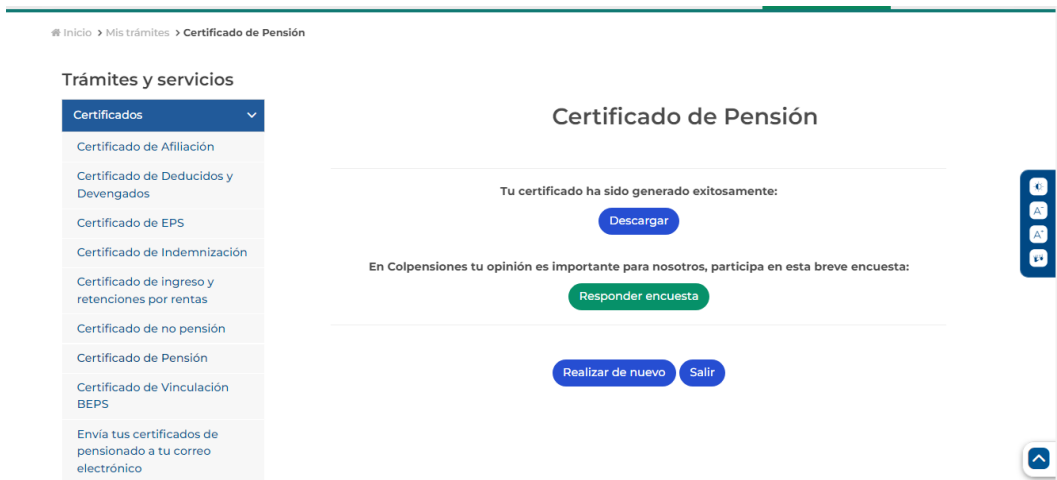


Figura 9. Formulario para la descarga del certificado

Este portal web no presenta ningún tipo de seguridad tipo captcha ni tampoco detección contra bots, sin embargo, los tiempos de carga son elevados llegando a necesitar hasta 6 minutos para lograr descargar el certificado.

3.1.2 Revisión del proceso para la extracción de los certificados de ingresos y tiempos de servicio en el portal web Armada Nacional

En el segundo módulo, el portal de la Armada Nacional (https://www.armada.mil.co/arc_certificates/validation), la manera de obtener los certificados, tanto de ingresos como del tiempo de servicio prestado, son la misma, esto debido a el código de verificación, el cual es diferente dependiendo del tipo de certificado, de ahí en adelante, en el portal web solo se necesita llenar el formulario con los datos de la persona y dar click en “validar” para obtener el certificado como muestra la **Figura 10**.

▼ Información de la persona que hace la consulta

Nombres *

Registre en el campo Nombres la información correspondiente de la persona q realiza la validación.

Correo electrónico *

Registre en el campo Correo Electrónico el email de la persona que realiza la validación.

▼ Información de la certificación que se desea validar

Cédula *

Por favor escriba la cédula que aparece en la certificación.

Código militar *

Por favor escriba el código militar que aparece en la certificación.


Código de verificación *

Por favor escriba el código de verificación que aparece en la certificación.

Si no puede visualizar el captcha por favor haga clic [AQUI](#).

CAPTCHA

Esta pregunta se hace para comprobar que usted, es una persona real e impedir el envío automatizado de mensajes.



Generar una nueva imagen

¿Qué código hay en la imagen? *

Por favor digite los caracteres mostrados en la imagen para continuar el proceso.

Figura 10. Formulario para obtener los certificados de la Armada

Este portal web presenta un captcha alfanumérico, no presenta detección contra bots y el certificado no se genera, sino que se envía al correo puesto en el formulario, el correo no presenta validación por lo que a cualquier correo que se ponga en el formulario se hará el envío del certificado.

3.1.3 Revisión del proceso para la extracción de los certificados de cupón de pago y descuentos en portal web Fopep

Para el tercer módulo, el portal de Fopep (<https://www.fopep.gov.co/validacion-de-certificados/>), el formulario para obtener los certificados no se requiere de los datos de la persona, sino del tipo de certificado y el código de verificación, después de ingresar los datos se da click en “verificar” (**Figura 11**),, esto hace que se genere una ventana con la opción de descarga (**Figura 12**), la cual genera el certificado para ser guardado en el ordenador como se muestra a continuación.

Actualizado el 05-04-2023 / 10:50:37 am

Directorio de Entidades Boletín Informativo Notas Saludables RSS

Consortio FOPEP

Imprimir Modo lectura Iniciar sesión

Inicio Acerca del FOPEP Entidades Sustituidas Trámites de Pensionados Atención al Pensionado Noticias

Estás en > Validación de Certificados

Servicios en Línea

- Contacto PQRSD
- Certificados y Cupones de Pago
- Otros servicios en línea

Validación de Certificados

Esta consulta es utilizada por quien desea comprobar la autenticidad del certificado. Por favor seleccione el tipo de certificado que desea consultar y digite el código de verificación; a continuación de clic en el botón "Verificar":

Tipo de certificado Código

Seleccione tipo de certif [] [] [Verificar]

Figura 11. Formulario para obtener los certificados de Fopep

×

El código es válido para un documento expedido por el Consorcio FOPEP, por favor verifique el citado documento a continuación:

Verificación de certificado [Icono de documento]

Figura 12. Ventana de confirmación para obtener los certificados de Fopep

Este portal web no presenta ningún tipo de seguridad tipo captcha ni tampoco detección contra bots, además se evidencia en las peticiones del portal web que se está haciendo el uso de un API.

3.1.4 Revisión del proceso para la extracción de los certificados de mesada pensional y comprobante de nómina en el portal web Fiduprevisora

En el cuarto módulo, el portal web de Fiduprevisora Fomag (<https://www.fomag.gov.co/fomagzp/login/GestLogin/login>), tiene como pasos para obtener los certificados, primero, iniciar sesión seleccionando de una lista desplegable el tipo de documento (cedular, cedula de extranjería, pasaporte, Nit, registro civil, tarjeta de identidad, numero único de identificación), el número de identificación y una clave como se muestra a continuación en la **Figura 13**.

Iniciar sesión

Si su clave se encuentra bloqueada, por favor dirigir su solicitud de desbloqueo a nuestras líneas telefónicas en Bogotá (60+1) 7562424, a nivel nacional 01 8000 180 510, en nuestra Página Web <https://pags.fiduprevisora.com.co/radicar.php> mediante Oficio Físico el cual podrá remitir a la Calle 72 N° 10-03 Local 114 en la ciudad de Bogotá

Seleccione tipo de documento ▾

Número de documento

Clave

[RECOMENDACIONES GENERALES](#)

[VALIDAR CERTIFICADO DE GENERACIÓN](#)

Figura 13. Inicio de sesión en el portal de Fiduprevisora

Una vez se inicia sesión en el portal web, se despliega un menú lateral con todas las opciones para realizar, donde se debe de seleccionar “Certificados” como se muestra a continuación en la **Figura 14**.

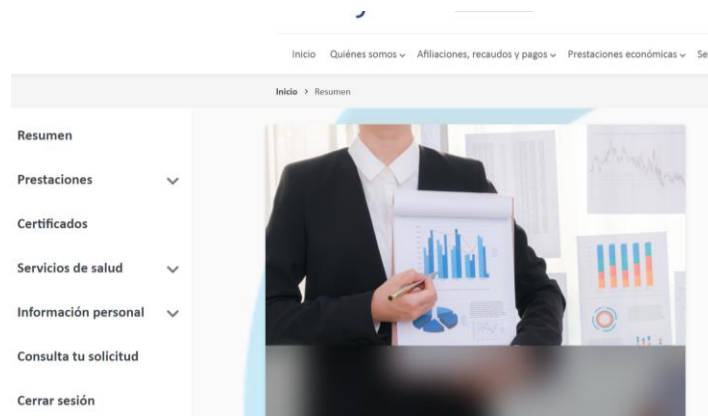


Figura 14. Menú de opciones del portal de Fiduprevisora

Después de seleccionar la opción de “certificados” aparecerá una lista desplegable con todos los tipos de certificados (**Figura 15**), si el certificado es de tipo Mesada Pensional únicamente se debe dar click en “enviar” y esto abrirá una nueva pestaña en el navegador con el certificado.

Figura 15. Formulario para solicitar el certificado de mesada pensional Fiduprevisora

Pero si el certificado es de tipo comprobante de nómina, se debe de ingresar la fecha desada (**Figura 16**) y finalmete dar click en “enviar” de igual manera se abra una nueva pestaña con el certificado.

Figura 16. Formulario para solicitar el certificado de comprobante de nómina Fiduprevisora

La pestaña que se abre luego de dar click en “enviar” es la de una URL de la misma Fiduprevisora y en sus parámetros se encuentran los identificadores del certificado y de la persona (**Figura 17**).



Figura 17. Pestaña con el certificado en Fiduprevisora

Este portal web no presenta ningún tipo de seguridad tipo captcha ni tampoco detección contra bots, además se evidencia en las peticiones del portal que se está haciendo uso de un API.

3.1.5 Revisión del proceso para la extracción del certificado de colilla de pago en el portal web Mindefensa

Para el quinto módulo, en el portal web de Mindefensa (<https://wasedeelectronicapro.azurewebsites.net/mindefensa.html#/verify-document>), se debe de ingresar un código de verificación, resolver un recaptcha y finalmente dar click en “verificar” esto genera el certificado en el recuadro que se encuentra debajo del botón (Figura 18).



The image shows a web form titled "Verificación de documentos" from the "MINISTERIO DE DEFENSA NACIONAL". At the top, there is a navigation menu with items: Inicio, Ministro, Ministerio, Políticas, planes y proyectos, Gestión y desarrollo, Contrataciones, Centro de Prensa, and Servicios al Ciudadano. Below the menu, there are five empty text input fields for document verification. To the left of these fields is a reCAPTCHA widget with the text "I'm not a robot" and a checkbox. Below the input fields is a blue button labeled "Verificar". At the bottom of the form is a large, empty rectangular box intended for the certificate output.

Figura 18. Formulario para la solicitud del certificado Mindefensa

Este portal web presenta un desafío tipo recaptcha donde se deben seleccionar imágenes para poder obtener el certificado, pero no presenta ningún tipo de detección contra bots.

3.1.6 Revisión del proceso para la extracción de los certificados de tiempos de servicio e ingresos en el portal del Ministerio de defensa usando QR

El sexto módulo automatiza al portal web de Mindefensa, sin embargo, a diferencia del anterior solo se puede acceder mediante un QR que se encuentra dentro de uno de los certificados que el solicitando de crédito presenta al inicio del proceso para solicitar un

crédito, al acceder al link en el QR no aparece un certificado en formato PDF, sino que aparece los datos de la persona embebidos dentro del HTML (**Figura 19 y Figura 20**).

Ministerio de Defensa Nacional

Los datos asociados al código QR consultado, corresponden a:

[Nombre del afiliado]

Número documento: 11111111

Total de tiempo en la fuerza (AA-MM--DD): 19-0-25

Fecha de generación: 20-04-2022

Fecha de vencimiento: 20-05-2022

Figura 19. Datos dentro del portal al abrir un código QR Mindefensa

Ministerio de Defensa Nacional

Los datos asociados al código QR consultado, corresponden a:

[Nombre del Afiliado]

Número documento: [Numero de documento del afiliado]

Mes del certificado (AAAA-MM): 2023-5

TOTAL DEVENGADO: \$3,322,151

TOTAL DESCUENTOS: \$872,002

NETO A PAGAR: \$2,450,149

Fecha de generación: 23-05-2023

Fecha de vencimiento: 22-06-2023

DEVENGADO	PORC	VALOR	DESCUENTO	COD.	INICIO	TERMINO	VALOR
SUEL_BASICO	0	\$ 1.624.000,00	PREVISORASASUB	981K	202305	202305	\$ 17.311,00
SEGVDSUBS	0.0	\$ 17.311,00	COOSERPARK	9960	202301	202802	\$ 27.511,00
SUBFAMILIAR	20.0	\$ 324.800,00	SISTEMA SALUD FFMM	9101	202305	202305	\$ 78.000,00
BONORDPUFF	25.0	\$ 406.000,00	APORTE CAJA DE RETIRO FFMM	9105	202305	202305	\$ 99.500,00
PRSOLVOL	58.5	\$ 950.040,00	CAJA VIVIENDA MILITAR AHORRO	9103	202305	202305	\$ 113.680,00
			BANCO CORBANCA	952Q	201704	202603	\$ 536.000,00
TOTAL DEVENGADOS		\$ 3.322.151	TOTAL DESCUENTOS				\$ 872.002

Figura 20. Datos dentro del portal al abrir un código QR Mindefensa

Este portal web no presenta ningún tipo de seguridad tipo captcha ni tampoco detección contra bots.

3.1.7 Revisión del proceso para la extracción de los certificados de ingresos y tiempos de servicio en el portal web Ejército Nacional

El último módulo para el portal de Ejército nacional (<https://www.mindefensa.gov.co/SedeElectronica/verifyDocs.jsp#no-back-button>) necesita de dos partes de las siete partes que conforman un código de verificación para esta entidad como se ve debajo del código de barras de la **Figura 21**, las dos partes

necesarias son aleatorias es decir en un momento el portal web puede solicitar las dos primeras partes del código de verificación pero si se recarga el portal puede necesitar la última y la primera, por ultimo cabe resaltar que este portal contiene un captcha de números y letras

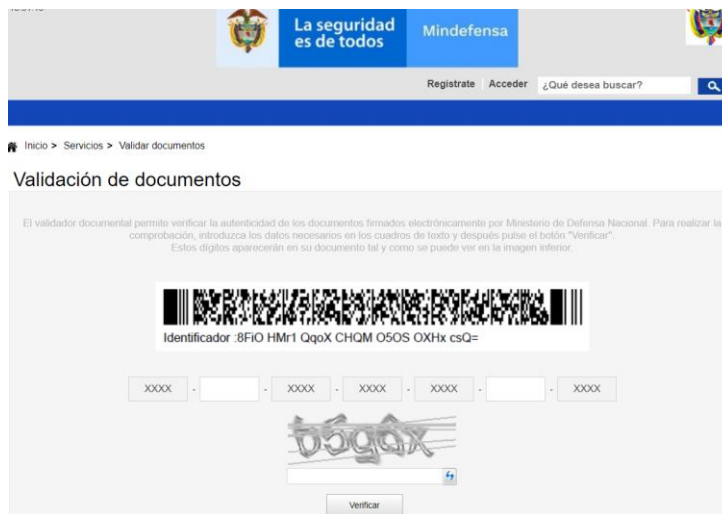


Figura 21. Formulario para obtener el certificado del Ejército Nacional

Este portal web presenta las barreras más altas en seguridad de todos los portales vistos, como ya se dijo las partes del código que se deben ingresar varían cada vez que se carga el portal web, presenta un captcha alfanumérico distorsionado y además es capaz de detectar comportamientos anormales, o si el navegador está siendo controlado por un Bot y bloquear la IP.

3.2 Diseño de la solución

Para brindar una solución que redujera el trabajo repetitivo en el proceso de solicitud de crédito, Xirius decidió crear una API de servicios REST que permita hacer la extracción de información, la obtención de certificados y su validación. Esta solución hace uso de características de Google Cloud Platform (GCP) que facilitan el desarrollo como son: el almacenamiento en Buckets, la detección mediante OCR y la capacidad de ejecutar código sin servidores mediante el uso de Google Cloud Functions (GCF). A continuación, mediante un diagrama de componentes se presenta la solución general propuesta por Xirius, donde el practicante es el encargado de construir los módulos mediante Cloud Functions.

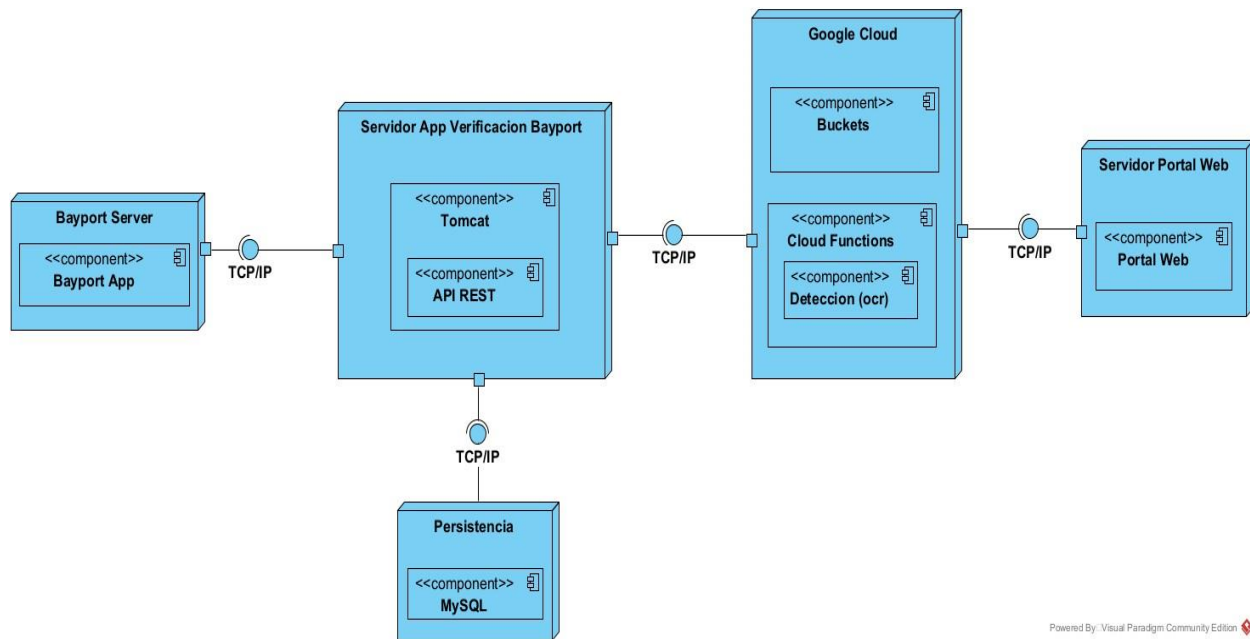


Figura 22. Vista de despliegue

En la **Figura 22** se aprecia que Bayport dispone de un servidor (nodo más a la izquierda del diagrama de despliegue) donde se encuentra alojada la aplicación BayportApp. Esta aplicación se comunica por TCP/IP hacia el puerto P (por cuestiones de confidencialidad no se puede exponer) donde tendrá acceso al servidor de aplicaciones TOMCAT 9.0.38 en el que se encuentra el API REST que permite la validación de certificados. La información transaccional del API es alojada en una base de datos MySQL 8.0, alojada en otro servidor. Esta aplicación se comunica con los servicios de GCP de almacenamiento a través de Buckets, detección de texto mediante Vision AI para la extracción de los datos de un certificado y Cloud Functions que finalmente se conectarán con el servidor externo de cada entidad donde deba descargar los certificados.

A continuación, se aborda el proceso de análisis de la documentación entregada, como las historias de usuario, certificados a obtener y los datos requeridos para generar los certificados dentro de los portales web; esto con el fin de determinar qué tipo de mecanismos implementar para obtener los certificados, así mismo poder determinar que tecnología usar durante la fase de desarrollo de los módulos.

Al inicio de la práctica profesional, la empresa dedicada al desarrollo de software Xirius Tech, proporcionó al practicante una inducción en la cual se le presentó el proyecto en el que trabajaría durante su práctica profesional. Este proyecto es la realización de una solución de automatización para la parte inicial del proceso de solicitud de crédito en la empresa Bayport. El practicante tiene la tarea de desarrollar módulos que permitan obtener los certificados de los portales web de manera automatizada.

Para poder iniciar el desarrollo y cumplir con los requerimientos del cliente, el primer paso es indicar la información disponible en forma de historias de usuario y proporcionar el acceso al entorno de desarrollo en la nube así mismo como el acceso al repositorio donde

se encuentra el código de la solución. De esta manera, el practicante está listo para el proceso de inducción y comprensión de las tecnologías que deberá usar y la arquitectura de la solución. Esto se logra mediante capacitaciones y ejemplos para ampliar el conocimiento respecto a las tecnologías que permitan la automatización robótica de procesos. Cabe resaltar que además de las capacitaciones, el equipo de desarrollo también brindó apoyo en lo que respecta a las prácticas de desarrollo de la empresa, es decir, cómo reportar el estado de una tarea o que hacer en caso de presentar inconvenientes, así como las reglas que se siguen a nivel interno para el nombrado de versiones en git.

3.2.1 Diseño de los mecanismos para la obtención de certificados

Del anterior análisis se obtuvo la información necesaria para el diseño de los mecanismos que implementará cada módulo para automatizar la obtención de los certificados, a continuación, la **Tabla 1** resume dichos mecanismos:

Tabla 1 - Mecanismos para la obtención de certificados

Portal Web	Descripción del portal	Mecanismo
Colpensiones	El portal no hace uso de un API por lo que se decidió optar por la creación un Bot que imitará los pasos que realiza un trabajador para obtener el certificado.	Creación de un Bot mediante Puppeteer
Armada Nacional	El portal no hace uso de un API y no es posible obtener a través de peticiones la imagen con el captcha, por lo que se decide crear un Bot para obtener la imagen del navegador e imitar los pasos que realiza un trabajador.	Creación de un Bot mediante Puppeteer
Fopep	El portal si cuenta con un API pues se evidenciaba en sus peticiones el intercambio de mensajes JSON y de igual manera fue posible obtener el certificado usando Postman. Por lo anterior se opta por Axios para realizar los pasos de obtención a través de peticiones.	Uso de un cliente REST
Fiduprevisora Fomag	Este portal se comporta igual que el caso anterior de Fopep. Por lo anterior se opta por Axios para realizar los pasos de obtención a través de peticiones.	Cliente REST

Ministerio de Defensa	El portal no hace uso de un API, además de que contiene un desafío reCaptcha por lo que se decidió crear un Bot para resolver el desafío e imitar los pasos que realiza un trabajador.	Creación de un Bot mediante Puppeteer
Ministerio de Defensa versión QR	El portal no cuenta con un certificado para descargar. Sin embargo, si cuenta con la información embebida dentro de su HTML por lo que se hace necesario hace web scraping.	Web Scraping mediante Cheerio
Ejército Nacional	El portal no hace uso de un API y no es posible obtener a través de peticiones la imagen con el captcha por lo que se decide crear un Bot para obtener la imagen del navegador e imitar los pasos que realiza un trabajador.	Creación de un Bot mediante Puppeteer

4. PROGRAMACIÓN DE ACTIVIDADES

A continuación, se presenta la documentación referente a los requerimientos de cada módulo de automatización. Los requerimientos se dividen en actividades programadas para ser desarrolladas en una semana, cada actividad cuenta con una clasificación de prioridad teniendo en cuenta las fechas establecidas con el cliente, el orden de entrega de los módulos es: Colpensiones, Armada Nacional, Fopep, Fiduprevisora, Ministerio de defensa, Ministerio de defensa QR y Ejército Nacional.

El desarrollo de las actividades de la práctica se hizo bajo la metodología de la empresa Xirius, basada en Scrum y donde cada Sprint equivale a una semana, las actividades que no eran culminadas dentro del Sprint pasaban con una prioridad más alta al siguiente, esto se hizo así durante los 5 meses que duró el desarrollo. Enseguida se detallan las funcionalidades solicitadas por la empresa y los requerimientos asociados a estas.

4.1 PRIMER MES DE LA PRÁCTICA PROFESIONAL

El primer mes de la práctica profesional tuvo como objetivo la familiarización con el proyecto, el análisis de los requisitos de cada módulo, las capacitaciones para adquirir conocimientos en automatización junto con entrenamientos con Node.js y Puppeteer, además de indicaciones de como analizar cada portal web para diseñar la mejor estrategia de automatizar el proceso. Estas actividades fueron:

Tabla 2 - Primer mes de práctica: actividades referentes a las reuniones en el mes

Actividad	Prioridad	Duración prevista	Duración Real
Reunión de seguimiento semanal con el asesor	Alta	2	2
Reunión de seguimiento semanal con todo el equipo de Xirius	Alta	3	3

Durante todos los meses se programaron reuniones de seguimiento semanales, tanto con el asesor de la empresa como con todo el equipo de desarrollo de Xirius.

Tabla 3 - Primer mes de práctica: actividades referentes a la capacitación del practicante

Actividad	Prioridad	Duración prevista	Duración Real
Inducción y socialización de todo el proyecto	Alta	3	2
Capacitaciones para entender como trabajar en Google Cloud	Alta	8	8
Capacitaciones y entrenamiento en las tecnologías a usar	Alta	35	30
Capacitaciones para comprender las diferentes estrategias para automatizar un proceso	Alta	16	16

Estas actividades dieron como resultado el análisis del proyecto plasmado en el capítulo tres del presente informe.

Tabla 4 - Primer mes de práctica: actividades referentes al análisis del proyecto

Actividad	Prioridad	Duración prevista	Duración Real
Analizar el portal web de Colpensiones, las peticiones que hace y si presenta algún tipo de restricciones.	Alta	4	9
Diseñar la estrategia para automatizar la obtención del certificado de Colpensiones	Alta	8	8
Analizar el portal web de la Armada, las peticiones que hace y si presenta algún tipo de restricciones.	Alta	5	4
Diseñar la estrategia para automatizar la obtención de los certificados de la Armada	Alta	8	9
Analizar el portal web de Fopep, las peticiones que hace y si presenta algún tipo de restricciones.	Media	5	4

Analizar el portal web de Fiduprevisora, las peticiones que hace y si presenta algún tipo de restricciones.	Media	5	4
Analizar el portal web del Ministerio de defensa, las peticiones que hace y si presenta algún tipo de restricciones.	Media	5	4
Analizar el portal web del Ministerio de defensa QR, las peticiones que hace y si presenta algún tipo de restricciones.	Media	5	3
Analizar el portal web del Ejército nacional, las peticiones que hace y si presenta algún tipo de restricciones.	Media	5	10
Investigar como integrar el servicio de Cloud Storage con Cloud Functions	Alta	3	2

Al finalizar el primer mes se realizaron un total de 16 actividades, que tomaron un tiempo de 118 horas, el tiempo estimado fue de 120 horas por que la estimación tuvo un desfase del 1,7 % menos de lo planeado lo cual no resulta problemático y es entendible ya que el primer mes se dedicó principalmente al análisis y diseño sin pasar aún al desarrollo de software.

4.2 SEGUNDO MES DE LA PRÁCTICA PROFESIONAL

En el segundo mes de la práctica profesional, las actividades se enfocaron hacia el desarrollo de los módulos para los portales web de Colpensiones y la Armada Nacional además del despliegue en el ambiente de pruebas, las actividades fueron:

Tabla 5 - Segundo mes de práctica: actividades referentes a las reuniones en el mes

Actividad	Prioridad	Duración prevista	Duración Real
Reunión de seguimiento semanal con el asesor	Alta	2	2
Reunión de seguimiento semanal con todo el equipo de Xirus	Alta	3	3

En las reuniones de seguimiento de este mes se validaron los compromisos y actividades realizadas además de hacer revisiones en el avance de los módulos para los portales web de Colpensiones y la Armada nacional.

Tabla 6 - Segundo mes de práctica: actividades referentes a la construcción de los módulos de Colpensiones y la Armada

Actividad	Prioridad	Duración prevista	Duración Real
-----------	-----------	-------------------	---------------

Definir la estructura de respuesta que emitirán los módulos	Alta	1	1
Construir el módulo de automatización para Colpensiones	Alta	40	42
Integrar al módulo de Colpensiones la capacidad de interactuar con Cloud Storage	Media	1	1
Desplegar en el ambiente de pruebas el módulo de Colpensiones	Alta	2	1
Investigar y elegir un servicio que permita saltar captchas	Alta	4	6
Investigar tecnologías que me permitan revisar la bandeja de entrada de un correo	Alta	2	2
Construir el módulo de automatización para la Armada	Media	40	42
Integrar el servicio para saltar captchas al módulo de la Armada	Alta	2	2
Integrar al módulo de la Armada la capacidad de interactuar con Cloud Storage	Media	1	1
Crear un módulo para revisar la bandeja de entrada de un correo y obtener la información del email	Alta	4	6
Integrar al módulo de la armada la interacción con el módulo para revisar correos	Alta	2	3
Desplegar en el ambiente de pruebas el módulo de la Armada	Media	2	1
Adicionar a la lógica de los módulos el guardar como archivo el resultado de la operación al obtener los certificados	Media	2	2
Diseñar la estrategia para automatizar la obtención de los certificados de Fopep	Media	5	5
Diseñar la estrategia para automatizar la obtención de los certificados de Fiduprevisora	Media	5	5
Diseñar la estrategia para automatizar la obtención del certificado de Ministerio de defensa	Baja	5	5
Realizar pruebas en los diferentes escenarios posibles	Media	1	1
Crear un directorio para almacenar el código de los módulos en el repositorio y subir los cambios a GitHub	Baja	1	1

Al terminar el segundo mes se realizaron 20 actividades, en un tiempo de 127 horas excediendo en un 5.8 % las 120 horas planeadas para el mes, esto debido a que los módulos presentaban particularidades que debían manejarse como se mencionó en el

capítulo de análisis del proyecto. El portal web de Colpensiones presenta un tiempo de espera elevado y la Armada contiene un captcha, en el capítulo de Desarrollo se detalla cómo se manejaron estas situaciones.

4.3 TERCER MES DE LA PRÁCTICA PROFESIONAL

En el tercer mes de la práctica profesional las actividades se enfocaron hacia el desarrollo de los módulos para los portales web de Fopep, Fiduprevisora y Ministerio de defensa además del despliegue en el ambiente de pruebas, las actividades fueron:

Tabla 7 - Tercer mes de práctica: actividades referentes a las reuniones en el mes

Actividad	Prioridad	Duración prevista	Duración Real
Reunión de seguimiento semanal con el asesor	Alta	1	1
Reunión de seguimiento semanal con todo el equipo de Xirius	Alta	3	3
Reunión de entrega a Bayport de los primeros certificados soportados por la aplicación (Colpensiones, Armada, Fopep)	Alta	1	1

En las reuniones de seguimiento de este mes se validan los compromisos y actividades realizadas además de hacer revisiones en el avance de los módulos para el portal web de Fopep, Fiduprevisora y Ministerio de defensa. Por otra parte, se empezó con las reuniones con el cliente para la entrega de los documentos de la aplicación que hacían uso de los módulos de automatización para su validación.

Tabla 8 - Tercer mes de práctica: actividades referentes a la construcción de los módulos de Fopep, Fiduprevisora y Mindefensa

Actividad	Prioridad	Duración prevista	Duración Real
Construir el módulo de automatización de Fopep	Alta	30	30
Integrar al módulo de Fopep la capacidad de interactuar con Cloud Storage	Alta	1	1
Desplegar en el ambiente de pruebas el módulo de Fopep	Alta	2	2
Adicionar a los módulos el registro de cada paso que realizan	Baja	2	2
Realizar pruebas en los diferentes escenarios posibles para Fopep	Media	3	3
Añadir los parámetros de cada módulo a la documentación de la solución	Baja	1	1

Construir el módulo de automatización de Fiduprevisora	Alta	32	32
Integrar al módulo de Fiduprevisora la capacidad de interactuar con Cloud Storage	Alta	1	1
Desplegar en el ambiente de pruebas el módulo de Fiduprevisora	Alta	2	2
Realizar pruebas en los diferentes escenarios posibles para Fiduprevisora	Media	3	3
Construir el módulo de automatización de Mindefensa	Alta	40	40
Integrar el servicio para saltar captchas al módulo de la Mindefensa	Alta	1	1
Integrar al módulo de Fopep la capacidad de interactuar con Cloud Storage	Alta	1	1
Desplegar en el ambiente de pruebas el módulo de Mindefensa	Alta	2	2
Realizar pruebas en los diferentes escenarios posibles para Mindefensa	Media	3	4
Diseñar la estrategia para automatizar la obtención del certificado de Ministerio de defensa versión QR	Media	5	5
Diseñar la estrategia para automatizar la obtención del certificado del Ejército Nacional	Media	6	7

Al terminar el tercer mes, se realizaron 20 actividades en un tiempo de 141 horas, excediendo en un 0.7 % las 140 horas planeadas para el mes, la planeación fue más precisa teniendo en cuenta que ya se tenía la experiencia de desarrollo del segundo mes.

4.4 CUARTO MES DE LA PRÁCTICA PROFESIONAL

En el cuarto mes de la práctica profesional las actividades se enfocaron hacia el desarrollo de los módulos para los portales web de Mindefensa versión QR y Ejército Nacional además del despliegue en el ambiente de pruebas, las actividades fueron:

Tabla 9 - Cuarto mes de práctica: actividades referentes a las reuniones en el mes

Actividad	Prioridad	Duración prevista	Duración Real
Reunión de seguimiento semanal con el asesor	Alta	1	1
Reunión de seguimiento semanal con todo el equipo de Xirius	Alta	3	3
Reunión de entrega a Bayport de los siguientes certificados soportados por la	Alta	1	1

aplicación (Fiduprevisora, Mindefensa, Mindefensa QR)			
---	--	--	--

En las reuniones de seguimiento de este mes de este mes se validan los compromisos y actividades realizadas además de hacer revisiones en el avance de los módulos para los portales web de Mindefensa versión QR y Ejército Nacional. Por otra parte, continuaron las reuniones con el cliente para la entrega de los documentos de la aplicación que hacían uso de los módulos de automatización para su validación.

Tabla 10 - Cuarto mes de práctica: actividades referentes a la construcción de los módulos de Mindefensa QR y Ejército Nacional

Actividad	Prioridad	Duración prevista	Duración Real
Dividir el módulo de Mindefensa QR en dos para cada tipo de certificado	Alta	1	1
Construir el módulo de automatización para Mindefensa QR Colilla pago	Alta	20	20
Construir el módulo de automatización para Mindefensa QR certificado de tiempo de servicio	Alta	20	20
Integrar al módulo de Mindefensa QR Colilla pago la capacidad de interactuar con Cloud Storage	Alta	1	1
Integrar al módulo de Mindefensa QR certificado de tiempo de servicio la capacidad de interactuar con Cloud Storage	Alta	1	1
Desplegar en el ambiente de pruebas el módulo de Mindefensa QR Colilla pago	Alta	2	2
Desplegar en el ambiente de pruebas el módulo de Mindefensa QR certificado de tiempo de servicio	Alta	2	2
Realizar pruebas en los diferentes escenarios posibles para Mindefensa QR Colilla pago	Media	4	4
Realizar pruebas en los diferentes escenarios posibles para Mindefensa QR certificado de tiempo de servicio	Media	4	4
Construir el módulo de automatización para el Ejército Nacional	Alta	45	60
Integrar el servicio para saltar captchas al módulo del Ejército Nacional	Alta	1	1
Desplegar en el ambiente de pruebas el módulo del Ejército Nacional	Alta	2	2
Realizar pruebas en los diferentes escenarios posibles para el Ejército Nacional	Media	4	6

Al terminar el cuarto mes, se realizaron 16 actividades en un tiempo de 129 horas, excediendo en un 15.1 % las 112 horas planeadas para el mes, la razón de es que como se mencionó en el análisis en portal web del Ejército Nacional presenta los mecanismos de detección de bots más elevados de entre todos los portales web.

4.5 QUINTO MES DE LA PRÁCTICA PROFESIONAL

En el quinto mes de la práctica profesional las actividades se enfocaron a realizar los últimos ajustes de los módulos pendientes y pasar al despliegue en el ambiente de producción las actividades fueron:

Tabla 11 - Quinto mes de práctica: actividades referentes a las reuniones en el mes

Actividad	Prioridad	Duración prevista	Duración Real
Reunión de seguimiento semanal con el asesor	Alta	1	1
Reunión de seguimiento semanal con todo el equipo de Xirius	Alta	3	3
Reunión de entrega final de la solución a Bayport	Alta	2	2

En las reuniones de seguimiento de este mes se validan los compromisos y actividades realizadas además de hacer revisiones en ajustar el módulo del Ejército Nacional. Por otra parte, se hizo entrega del ambiente de producción para que Bayport integra la API a su sistema.

Tabla 12 - Cuarto mes de práctica: actividades referentes a ajustar los módulos de Mindefensa QR y Ejército Nacional además del despliegue en producción

Actividad	Prioridad	Duración prevista	Duración Real
Investigar sobre técnicas para rotar la IP de una Cloud Function	Alta	8	8
Investigar servicios de proxy que afecten lo menos posible la velocidad del módulo del Ejército Nacional	Alta	6	6
Implementar en el módulo del Ejército Nacional la rotación de IP	Alta	5	8
Modificar la lógica para obtener un PDF de mayor resolución en el módulo del Ejército Nacional	Alta	8	8
Desplegar en el ambiente de pruebas el módulo del Ejército nacional	Alta	1	1

Modificar el formato de los datos de los módulos de Mindefensa QR	Alta	1	1
Desplegar los módulos de Mindefensa QR en el ambiente de ambiente de pruebas	Alta	1	1
Recibir capacitaciones en Google Cloud	Media	30	30
Crear el ambiente de producción del proyecto	Alta	18	20
Desplegar los módulos de automatización construidos en el ambiente de producción	Alta	3	3
Realizar pruebas de todas las fases de la solución (extracción, obtención, validación)	Media	10	10
Realizar un manual para la creación de ambientes en Google Cloud	Baja	30	30

Al terminar el quinto mes, se realizaron 16 actividades en un tiempo de 132 horas, excediendo en un 3.9 % las 127 horas planeadas para el mes, en este mes el practicante recibió capacitaciones extra para adquirir conocimientos en cómo crear ambientes de desarrollo con Google Cloud. Al terminar el quinto mes los módulos construidos quedaron de la siguiente manera.

Tabla 13 - Resumen del desarrollo de los módulos

Módulo	Completado
Módulo de automatización de Colpensiones	100 %
Módulo de automatización de la Armada Nacional	100 %
Módulo de automatización de Fopep	100 %
Módulo de automatización de Fiduprevisora Fomag	100 %
Módulo de automatización del Ministerio de defensa	100 %
Módulo de automatización de Ministerio de defensa QR colilla de pago	100 %
Módulo de automatización de Ministerio de defensa QR certificado de tiempo de servicio	100 %
Módulo de automatización del Ejército Nacional	90%

Como se ve en la Tabla 13, el módulo del ejército no se completó. En el capítulo siguiente se detalla la estrategia diseñada para automatizar el portal web del Ejército Nacional y por qué no es posible automatizar este tipo de portales.

5. DESARROLLO DE LOS MÓDULOS

En este capítulo se presenta el proceso de implementación de los módulos que automatizarán la obtención de certificados y se documenta el proceso de desarrollo necesario para cumplir con los requisitos de la empresa Bayport.

Cada módulo pasa por cuatro fases: implementación, pruebas, ajustes y despliegue. Una vez implementado un módulo, se procede con las pruebas creadas por los encargados de calidad en Xirius, basadas en las historias de usuario. Si se encuentra alguna inconsistencia, se realizan los respectivos ajustes. Posteriormente, se realiza el despliegue en el ambiente de test para que Bayport realice más pruebas y de encontrar inconsistencias presentar un reporte para realizar los ajustes y volver a desplegar. Este proceso se repite hasta que el módulo sea aprobado por Bayport. Al finalizar el proyecto, el sistema es pasado al ambiente de producción para finalmente entrar en funcionamiento junto los sistemas de Bayport.

Para el versionamiento del proyecto, se utilizó Git con un repositorio central en GitHub, esto permitió la integración continua de cambios en el código y facilitó la realización de pruebas. Los cambios realizados eran revisados por otros desarrolladores de Xirius antes de pasar al repositorio central. Cuando un cambio era aprobado gracias GitHub Actions⁸, la aplicación se desplegaba de manera automática en el ambiente de test para entregarle a Bayport los nuevos cambios.

Como se mencionó en el capítulo de análisis, la solución brindada por Xirius es un API REST donde se plasman las etapas planteadas en la introducción, para ello se hace uso de cuatro endpoints⁹ que permiten automatizar todo el proceso de la solicitud de un crédito en Bayport, los endpoints son los siguientes:

1. Solicitar enlace de subida: A través de este endpoint, se proporciona un enlace para subir los certificados proporcionados por el solicitante a una carpeta dentro de un bucket en la nube de Google. La subida de un archivo dispara automáticamente la extracción de texto del certificado mediante OCR y este texto extraído es guardado como archivo JSON en la misma carpeta del certificado.
2. Extraer Información: Mediante este endpoint, se buscan los campos a extraer del certificado y se formatea dicha información según las necesidades de Bayport.
3. Obtener Certificado: Este endpoint hace uso de los de los módulos de automatización y recibe los parámetros que cada módulo necesite según el caso, como contraseñas o fechas. Cada certificado obtenido se guarda en una carpeta dentro del mismo bucket usado para almacenar los certificados proporcionados por el solicitante lo que desencadena también el proceso de extracción por OCR.

⁸ GitHub Actions: plataforma de integración y despliegue continuos (IC/DC) que permite automatizar las tareas de compilación, pruebas y despliegue

⁹ Endpoint: Es una ubicación digital en la que una API recibe solicitudes sobre un recurso específico en su servidor

4. Validar Información: A través de este endpoint, se compara la información extraída del certificado proporcionado por el solicitante contra la información extraída del certificado obtenido por los módulos de automatización y se retorna cualquier diferencia encontrada.

Los módulos desarrollados siguen la misma lógica general en su funcionamiento, excepto en el caso del Ministerio de Defensa con la versión QR, como se mostró en el análisis para ese caso no hay un certificado, sino que la información esta embebida en el HTML de sitio. Para el resto de módulos, se intenta descargar el certificado del respectivo portal web y guardarlo en un bucket. En caso de presentar errores, se realiza un nuevo intento de descarga del certificado, cada módulo cuenta con una cantidad N de intentos, teniendo en cuenta que un portal web puede caerse, dejar de responder o se necesita reintentar un captcha. Todo lo ocurrido se registra en un JSON con la estructura de la **Figura 23**, el cual se utiliza como respuesta a la petición. Todas las posibles excepciones fueron mapeadas para cada módulo con un mensaje de error comprensible para el usuario, éste JSON también se guarda como un archivo junto con el certificado, si fue posible obtenerlo.

```
{  
  "status": "(OK,FAIL)",  
  "errors": [],  
  "data": "DATOS OPCIONALES DEL CERTIFICADO"  
}
```

Figura 23. Estructura de respuesta de los módulos

Solo el módulo del ministerio de defensa versión QR usa el atributo data, puesto que ahí va la información que se extrae del HTML, para el resto de módulos este atributo no se usa.

5.1 Módulo para el portal web de Colpensiones

Este módulo tiene el objetivo de obtener el certificado llamado “certificado de pensión”. Durante el análisis, se evidenció a través de las peticiones del portal que no se estaba haciendo uso de un API. Es por esto que se creó un Bot con Puppeteer para automatizar el proceso en Chromium, los parámetros que necesita el módulo son: el tipo de documento del solicitante, el número de documento, la contraseña del sitio y la fecha del certificado. Puppeteer además de automatizar el navegador, permite interceptar las peticiones, esta característica cobra importancia debido a que posibilita disminuir la latencia y tiempos de espera elevados que tiene el portal web. Esto se logró abortando las peticiones de estilos, imágenes y demás archivos multimedia que realiza el sitio, dando como resultado una mejora significativa los tiempos de carga del portal.

La primera acción que se automatizo con Puppeteer es el inicio de sesión, ingresando el tipo de documento, el número de documento y la contraseña en el formulario, si las credenciales son incorrectas se captura el mensaje de error que aparece en el portal web y se guarda dentro del arreglo del atributo "errors" del JSON anteriormente mencionado y se fija el atributo "status" en "FAIL", caso contrario se prosigue accediendo directamente a la URL:

<https://sede.colpensiones.gov.co/loader.php?IServicio=Se&ITipo=Process&IFuncion=start&id=3>

No se navega por el menú de opciones, sino que se accede directamente al formulario de solicitud del certificado, gracias a que el sitio almacena en cookies las credenciales una vez se inicia sesión. En el formulario desplegado al acceder a la URL (**Figura 8** del tercer capítulo), se ingresa la fecha del certificado deseado y se acciona el botón de continuar lo cual redirige a la interfaz de la **Figura 9** del tercer capítulo, en este punto se hace uso de la petición:

https://sede.colpensiones.gov.co/loader.php?IServicio=Se&ITipo=Process&IFuncion=start&id=3&id_node=12&single=1

Pero a diferencia la anterior URL esta no se hace desde el mismo navegador sino a través de Axios puesto que se debe controlar la respuesta a la petición que es el archivo PDF, aquí cobra importancia otra característica de Puppeteer ya que permite extraer las cookies del navegador necesarias para que la petición funcione. Finalmente se guarda el archivo usando la librería de Storage proporcionada por Google para el manejo de archivos con Node.js en las Cloud Functions así mismo se guarda el JSON resultante cuyo status al final de una operación exitosa es "OK".

A continuación, el siguiente diagrama de secuencia detalla la interacción entre el Bot, y el portal web.

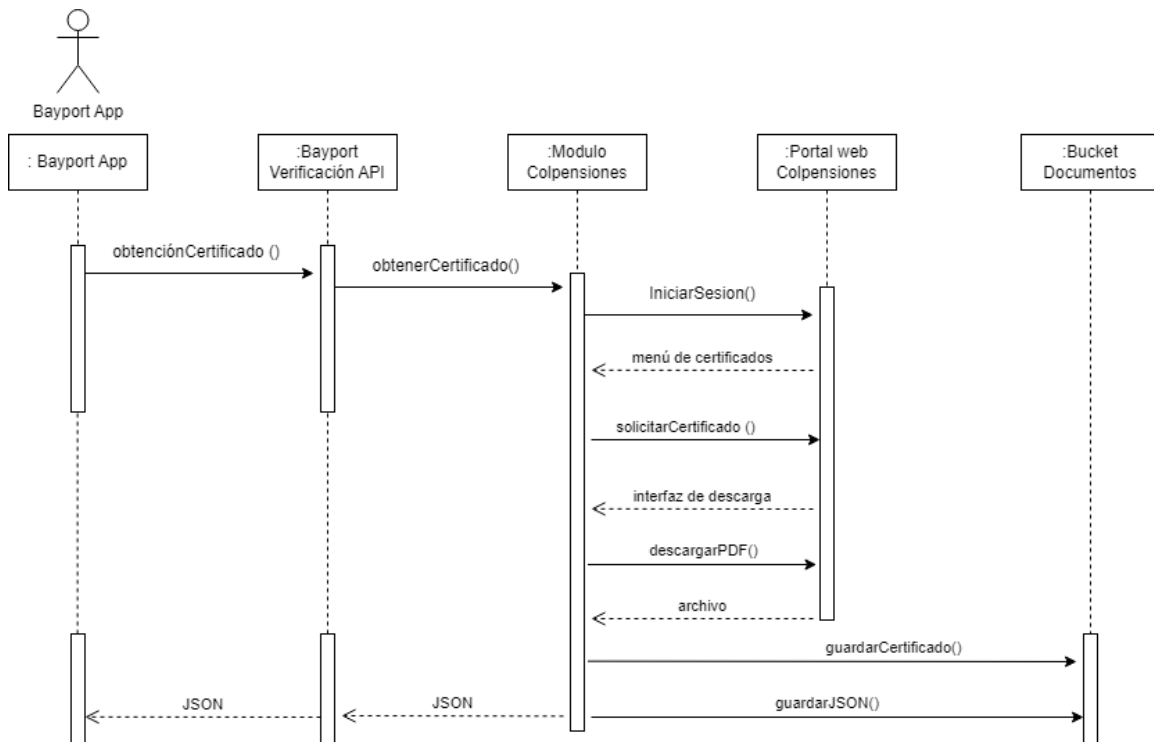


Figura 24. Diagrama de secuencia - Obtención de certificado Colpensiones

5.2 Módulo para el portal web de la Armada

El objetivo de este módulo es descargar los certificados de “colilla de pago”, “tiempos y haberes uniformados” y “tiempos y uniformados civiles”, durante el análisis se evidenció que el portal hacía uso de un desafío tipo captcha. Es por esto que se creó un Bot Puppeteer para automatizar el proceso en Chromium, los parámetros enviados desde el Backend *App Verificación Bayport* al módulo son: el nombre del solicitante, el número de documento, el código militar y un código de verificación. Las particularidades de este portal son el desafío captcha y que el certificado no se descarga desde el mismo portal, sino que es enviado al correo ingresado en el formulario mostrado en el análisis.

Dado que la pagina cuenta con un desafío captcha había que encontrar la manera de que un Bot pudiera superarlo y para ello se hizo una breve investigación en busca de un servicio que permitiera resolver estos desafíos. La **Tabla 14** contiene los servicios que se tuvieron en cuenta durante la investigación, los métodos que usan, los tipos de captcha que resuelvan, así como el tiempo que le toma al servicio dar una respuesta y las limitaciones que presentaban, todo esto obtenido de la misma documentación proporcionada por cada servicio.

Tabla 14. Servicios para la resolución de captchas

Servicio	Método	Tipos de Captcha	Tiempo de respuesta	Limitación
<u>EndCaptcha</u>	Basado en Humanos	Texto, Imagen Captcha	7s	No resuelve problemas matemáticos ni con demasiados caracteres
<u>2Captcha</u>	Basado en Humanos	reCaptcha, Imagen Captcha	8s~48 s	No presenta limitantes importantes, resuelve gran diversidad de captchas
<u>BestCaptchaSolver</u>	Basado en Humanos	reCaptcha, Imagen Captcha	14s ~ 46s	No presenta limitantes importantes
<u>AntiCaptcha</u>	Basado en humanos	Imagen, reCaptcha	7s	El tamaño del captcha debe ser entre 100 y 500000 bytes

Existían diferentes opciones muchos servicios hacían uso de OCR pero para disminuir el error se optó por aquellos servicios donde la respuesta la diera una persona, al final se escogió el servicio de *2Captcha*, por ser de los más rápidos y soportar la mayor diversidad de captchas, el funcionamiento es mediante una API que recibe la imagen del captcha en base 64 y retorna la respuesta en forma de cadena de texto o bien recibe para el caso del recaptcha, recibe la referencia lo resuelve y automáticamente se ve reflejado en el portal web.

El captcha presente en la armada es una imagen levemente distorsionada con números y letras (**Figura 25**), y debido a la capacidad de Puppeteer de hacer pantallazos, dentro del módulo se obtiene la imagen del captcha, se codifica en base 64 y se envía a la API de 2Captcha para obtener respuesta.



Figura 25. Captcha de la Armada

La segunda particularidad, el certificado no se obtiene directamente del sitio, sino que es enviado al correo electrónico puesto en el formulario (**Figura 10** del tercer capítulo) se resolvió creando un módulo aparte encargado de revisar correos, buscar por el código de verificación y descargar el certificado adjunto, el correo no es un parámetro puesto que se usa uno por defecto para todos los certificados.

A continuación, el siguiente diagrama de secuencia detalla la interacción entre el Bot, y el portal web.

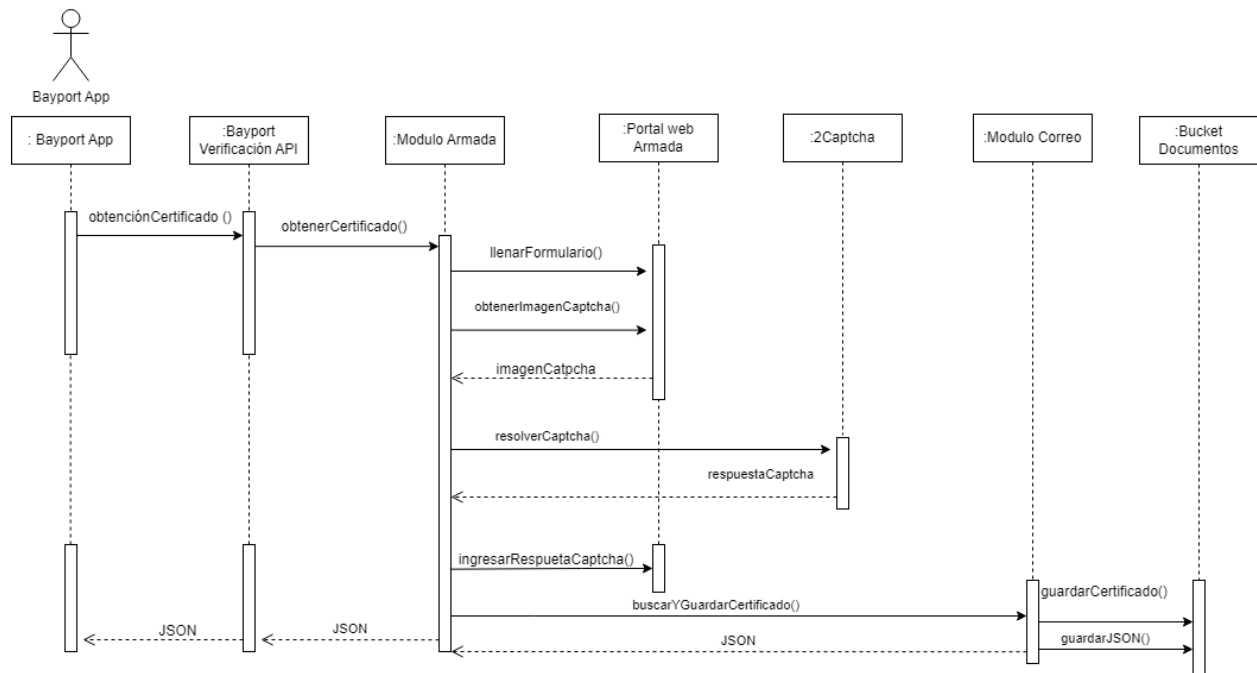


Figura 26. Diagrama de secuencia - Obtención de certificado Armada

5.3 Módulo para el portal web de Fopec

Este módulo es el encargado de descargar los certificados de “cupón de pago” y “descuentos en cola”, durante el análisis se evidenció a través de las peticiones del portal el uso de un API, el cual necesitaba como parámetros el tipo de documento del solicitante, el número de documento y un código de verificación. Es por esto que se hizo uso de Axios para automatizar el proceso de obtención a través del consumo del API sin necesidad de instanciar un navegador.

Para obtener el certificado primero se debe de comprobar que el usuario y el código de verificación existan, enviando el tipo de documento y el código de verificación del solicitante como parámetros en la siguiente URL:

[https://www.fopec.gov.co/sistema/services/Importadores_serv/validacion_de_certificado/?codigo=\[codigoVerificacion\] &tipo_documento=\[tipoDocumento\]](https://www.fopec.gov.co/sistema/services/Importadores_serv/validacion_de_certificado/?codigo=[codigoVerificacion] &tipo_documento=[tipoDocumento])

como respuesta esta petición Fopep retorna un JSON con el estado de la solicitud, mediante un dato booleano y un mensaje, si la solicitud fue exitosa se procede a enviar como parámetros en la URL la cédula junto con el código de verificación a la siguiente URL para obtener el certificado:

[https://www.fopep.gov.co/sistema/dashboard/GestDashboard/regenerar_certificado/\[numeroIdentificacion\] / \[codigoVerificacion\]](https://www.fopep.gov.co/sistema/dashboard/GestDashboard/regenerar_certificado/[numeroIdentificacion] / [codigoVerificacion])

Si la respuesta no fue exitosa se captura el mensaje que viene en la respuesta de la primera solicitud, el cual ya especifica lo sucedido, por ejemplo “No se encuentra registrado el código de verificación”, este tipo de mensajes se agregan al arreglo del atributo “errors” del JSON.

A continuación, el siguiente diagrama de secuencia detalla la interacción entre el Bot, y el portal web.

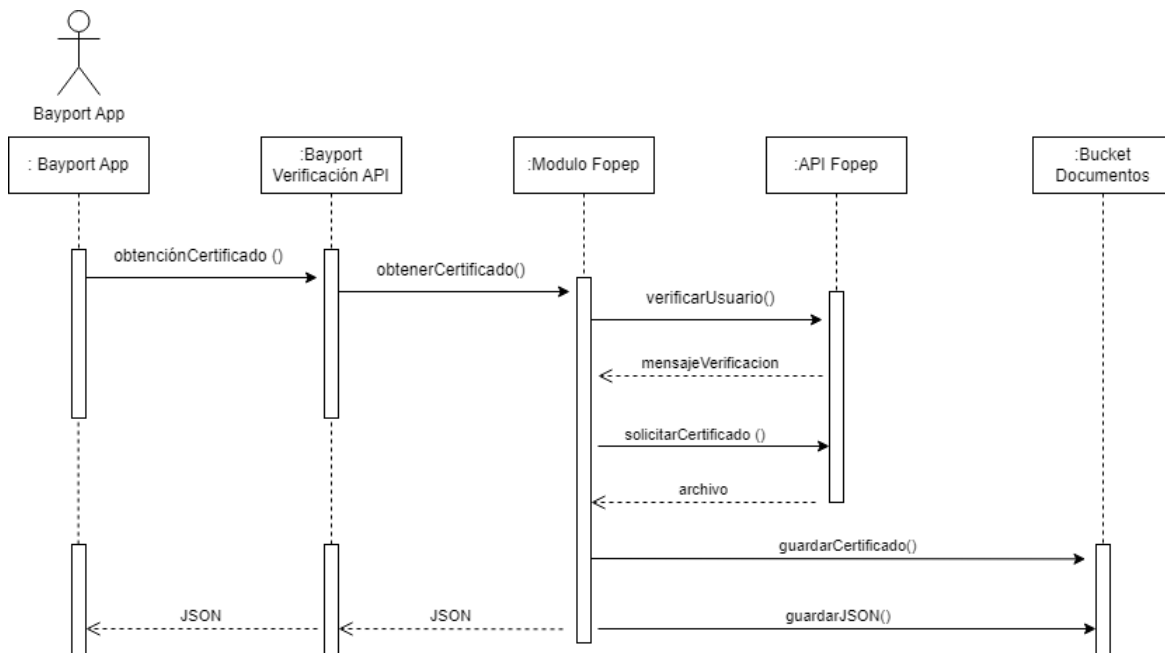


Figura 27. Diagrama de secuencia - Obtención de certificado Fopep

5.4 Módulo para el portal de Fiduprevisora Fomag

De este portal web el objetivo es descargar los certificados de “mesada pensional” y “comprobante de nómina”, durante el análisis se evidenció a través de las peticiones del portal que hacía uso de un API, el cual necesitaba como parámetros el tipo de documento del solicitante, el número de documento, la contraseña para iniciar sesión en el portal y el tipo de certificado, como parámetros adicionales solo para el caso del certificado “comprobante de nómina” se necesita el año y mes, referentes al certificado que se desea

descargar. Es por esto que se hizo uso de Axios para automatizar el proceso de obtención sin necesidad de instanciar un navegador.

Como primer paso se inicia sesión a través de la URL <https://www.fomag.gov.co/fomagzp/services/Login/acceder> cuyos parámetros son: el tipo de documento, el número de documento y la contraseña, estos van en el cuerpo de la petición.

En los headers de la respuesta de esta petición viene la cookie necesaria para poder realizar las peticiones de descarga del certificado.

Cada certificado cuenta con un número de comprobante el cual se obtiene, para el certificado de “mesada pensional” de la URL:

<https://www.fomag.gov.co/fomagzp/services/certificados/consultarCertificados>

y para el certificado de “comprobante de nómina” de la URL:

<https://www.fomag.gov.co/fomagzp/services/certificados/generarCertificado>

en el cuerpo de esta petición se envía el año y mes del certificado.

Finalmente, las peticiones para obtener el archivo son, para “mesada pensional”:

[https://www.fomag.gov.co/fomagzp/certificados/GestCertificado/verPdfCerMesaPensional?resolucion= \[numeroComprobante\] &numeroDocumento= \[numeroDocumento\]](https://www.fomag.gov.co/fomagzp/certificados/GestCertificado/verPdfCerMesaPensional?resolucion=[numeroComprobante]&numeroDocumento=[numeroDocumento])

y para “comprobante de nómina”

[https://www.fomag.gov.co/fomagzp/certificados/GestCertificado/verPdfCer?comprobante= \[numeroComprobante\]](https://www.fomag.gov.co/fomagzp/certificados/GestCertificado/verPdfCer?comprobante=[numeroComprobante])

Al haber un API los mensajes de respuesta traen a su vez mensajes que indican el éxito o fracaso de la solicitud. En caso de presentarse fallas estos mensajes son capturados y agregados al atributo “errors” del JSON.

A continuación, el siguiente diagrama de secuencia detalla la interacción entre el Bot, y el portal web.

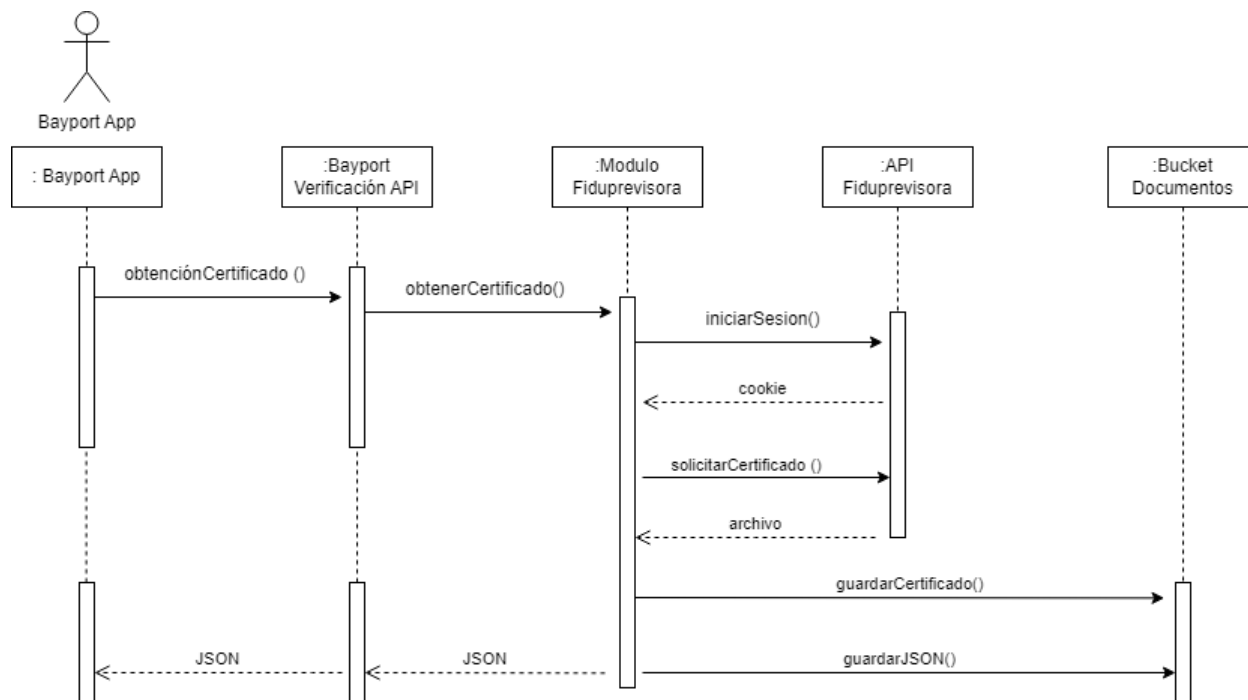


Figura 28. Diagrama de secuencia - Obtención de certificado Fidupervisora

5.5 Módulo para el portal del Ministerio de Defensa

Este módulo tiene el objetivo de descargar el certificado de “colilla de pago”, durante el análisis se evidenció que el portal hacia uso de un desafío tipo captcha más específicamente un recaptcha. Es por esto que se hizo uso de Puppeteer para automatizar el proceso en Chromium, como parámetro se necesita del código de verificación presente en el certificado presentado por el solicitante.

Al igual que con el módulo de la Armada, se hace uso del servicio 2Captcha para resolver el desafío, sin embargo, a diferencia de la armada un recaptcha cuenta con una referencia la cual puede consultarse desde otro sitio para resolverse, es por ello que en este caso solo se envía la referencia y se espera a que el servicio retorne si fue capaz de resolver el recaptcha o no y se prosigue con el proceso.

En este portal web el PDF desplegado se encuentra codificado en base 64 dentro de un selector HTML, por lo que después de resolver el captcha se obtiene la cadena con el PDF codificado y se guarda dentro del bucket.

En caso de no encontrar el código de verificación el portal muestra una alerta indicando que el código ingresado no está registrado, el mensaje de la alerta es capturado y agregado al arreglo del atributo “errors” del JSON.

A continuación, el siguiente diagrama de secuencia detalla la interacción entre el Bot, y el portal web.

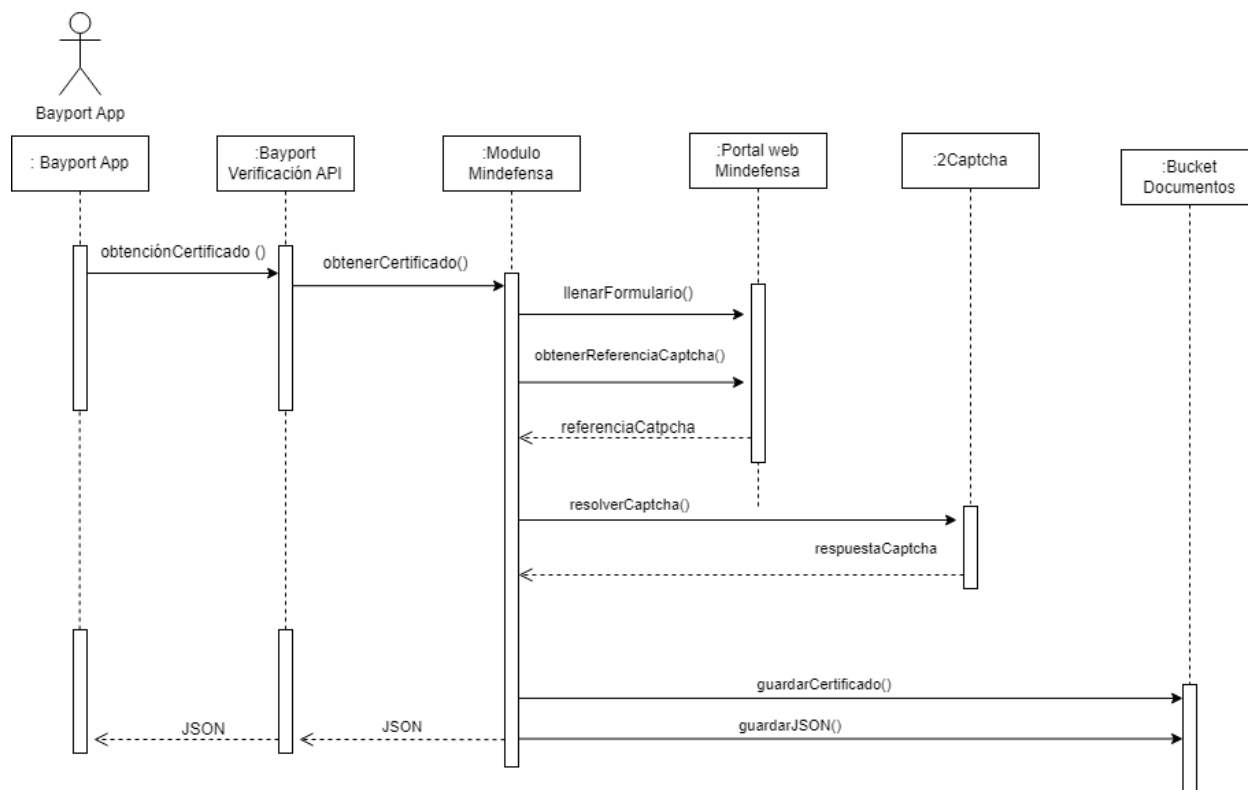


Figura 29. Diagrama de secuencia - Obtención de certificado Ministerio de Defensa

5.6 Módulo para el portal del Ministerio de Defensa Versión QR

Este módulo debe de obtener la información de los certificados de “colilla de pago” y “tiempos y haberes”, que esta embebida dentro del HTML del sitio, en el documento presentado por el solicitante existe un código QR que almacena el link donde se halla la información, en la fase de extracción este link es obtenido y posteriormente enviado como parámetro.

Para extraer la información embebida se obtiene el HTML mediante una petición al portal con Axios, luego se procede a hacer web scraping mediante Cheerio, que permite interactuar con el HTML del sitio, esta información se formatea de la misma forma que se hace al extraer la información del certificado proporcionado por el solicitante con el endpoint de extracción.

Dado que la estructura del sitio para cada certificado es considerablemente diferente y considerando que en un futuro la estructura puede cambiar y con el fin de hacer el mantenimiento más sencillo, se optó por dividir el módulo en dos modulos más pequeños los cuales sean los encargados de obtener la información para cada los dos tipos de certificado, las diferencias se hayan únicamente en la lógica de programación de los módulos, pero a nivel general ambos siguen la misma secuencia para obtener la información.

La información extraída es almacenada en el atributo “data” del JSON que se usa como respuesta y que a su vez se guarda en un bucket.

A continuación, el siguiente diagrama de secuencia detalla la interacción entre el Bot, y el portal web.

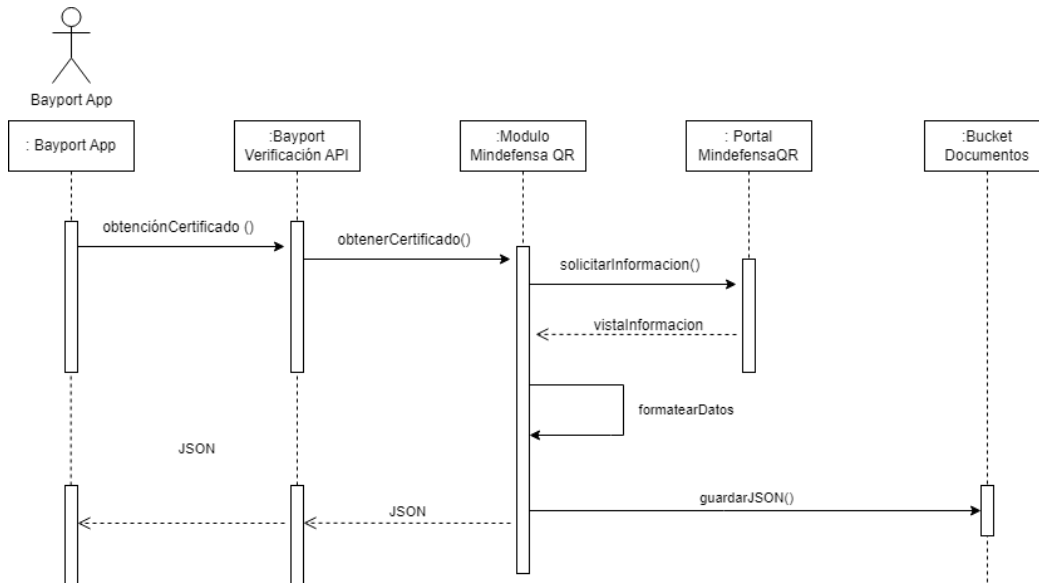


Figura 30. Diagrama de secuencia - Obtención de certificado Ministerio de Defensa versión QR

5.7 Módulo para el portal del Ejército Nacional

Este módulo, a diferencia de los anteriores mencionados, no llegó a ser desplegado en el ambiente de producción. A continuación, se detallan los inconvenientes que se presentaron en el proceso de desarrollo.

Como se mencionó en el análisis, este portal es capaz de detectar Bots y evitar que puedan interactuar con él, esta detección se realiza de dos maneras: la primera es revisando las banderas que se activan en el navegador cuando un software lo está controlando, y la segunda es detectando la cantidad de peticiones que hace una misma dirección IP.

Al crear un Bot con Puppeteer, las banderas de Chromium que indican que esa siendo controlado por software se activan, dando como resultado que no se pueda acceder al portal. Aun así, esto no resultó siendo un problema, ya que existen librerías para Node.js que hacen que Chromium no active estas banderas, lo que permite que el portal web no detecte al Bot y permita interactuar con él. Sin embargo, entraron en juego diversos factores que contribuyeron a la detección por tráfico sospechoso, lo que desemboca en el bloqueo de la dirección IP desde la cual se están haciendo las peticiones.

Otra característica a tener en cuenta del portal es que cuenta con un captcha de letras y números distorsionados (**Figura 31**), pero a diferencia del captcha del portal web de la armada, este es más ilegible, haciendo que muchas veces la respuesta al captcha proporcionada por el servicio 2Captcha sea incorrecta. Esto en sí mismo no es problema puesto que en estos casos el Bot puede reintentar hasta obtener una respuesta correcta. Sin embargo, después del despliegue en el ambiente de test se evidenció que en el día el módulo puede llegar a ser usado con una intensidad elevada. La combinación de reintentar varias veces un captcha, junto con el hecho de que el módulo se usa muchas veces en un mismo día, hace que el Bot sea detectado por tráfico sospechoso y la IP de la Cloud Function quede bloqueada.

La cantidad de veces que se necesitara el módulo en el día y la efectividad del servicio que resuelve los captchas no pueden ser controladas. Para este problema, se intentó dar una solución rotando las IP desde la cual salían las peticiones al portal del ejército, se realizaron dos soluciones.

La primera consistía en hacer uso de la red Tor¹⁰ para que las peticiones salieran desde un nodo de la red diferente cada vez que se usaba el módulo, la consecuencia no deseada fue que los tiempos de espera fueran muy altos, debido a que los nodos podían estar al otro lado del mundo. Por tanto, esta solución fue descartada poco tiempo después de su implementación, ya no tiene sentido esperar varios minutos en obtener un certificado si manualmente se puede hacer en menor tiempo.

La segunda solución consistió en utilizar un servicio de proxy que permitiera seleccionar la ubicación de las máquinas desde donde saldrán las peticiones. En un principio parecía que el problema había quedado solucionado, sin embargo, había días en los que el uso del módulo era intenso, lo que generaba que el servicio de proxy utilizara con mucha frecuencia el mismo nodo para enviar las peticiones lo que a su vez hacía que se detectara el Bot y la dirección IP quedara bloqueada por horas, teniendo en cuenta esto se decidió por no pasar al ambiente de producción y dejarlo en un estado de deprecated.



Figura 31. Captcha del portal web de Ejército Nacional

5.8 Creación y despliegue del ambiente de producción

Si bien los objetivos del presente informe se plantearon solo en relación a los módulos, el desarrollo de la práctica fue más allá de estos. Como se menciona en las últimas actividades del último mes, el practicante fue el responsable de crear y desplegar toda la

¹⁰ Red Tor: Red que implementa un enrutado de cebolla en donde los datos no van directamente del ordenador al servidor web, sino que pasan a través de diferentes nodos en un ruta más o menos aleatoria.

solución en el ambiente de producción, para ello se le capacitó en Google Cloud, que es el servicio que usa Xirius para desplegar sus proyectos.

En el capítulo 3 se mostró que la solución está compuesta de un servidor donde se aloja la API, el componente de las Cloud Functions donde se encuentran los módulos y un componente de persistencia donde se guarda la información extraída de los certificados, la fecha de subida al sistema y demás datos. Para desplegar el servidor y teniendo en cuenta el conocimiento adquirido en las capacitaciones existían dos opciones: utilizar una máquina virtual o utilizar la plataforma Cloud Run.

En una máquina virtual el desarrollador tiene que configurar el sistema operativo de la máquina, configurar el servicio, balanceo, las conexiones y cualquier característica que necesite la aplicación. Sin embargo, Cloud Run permite ejecutar la aplicación mediante contenedores, lo que delega la gestión de infraestructura al proveedor de la nube. Esto significa que el servicio puede escalar automáticamente de manera horizontal, si una máquina llega a su límite, y contar con el balanceo de carga [24]. El servicio siempre estará disponible, sin embargo, como desventaja se tiene el aumento en los costos. Como restricción esta que la aplicación no puede depender de archivos que se encuentren en el servidor.

La API de este proyecto no necesita archivos del servidor y con el fin de reducir costos se hace uso de la plataforma de Cloud Run.

La persistencia de la aplicación hace uso de una base de datos relacional que tiene como gestor para administrar la base de datos a MySQL, como proveedor del servicio también se usa Google Cloud.

5.8.1 Despliegue de la API en Cloud Run

Para desplegar el API en Cloud Run se necesita de crear una imagen ¹¹del contenedor, para ello se hace uso de Docker, ejecutando el siguiente comando donde se indica el nombre de la imagen.

```
docker build -t [IMAGE-NAME]
```

La plataforma de Cloud Run necesita de la imagen, para ello cuenta con repositorio, así que la imagen creada debe ser subida a dicho repositorio mediante el siguiente comando, donde se indica el identificador del proyecto en la nube y el nombre de la imagen previamente creada.

¹¹ Una imagen de contenedor en Docker, es un paquete de software ligero, independiente y ejecutable que incluye todo lo necesario para ejecutar una aplicación

```
docker push gcr.io/[PROJECT-ID]/[IMAGE-NAME]
```

Ya con esto desde la misma interfaz del portal web de Google Cloud se despliega la aplicación. En el apartado de Cloud Run se crea un servicio, donde se puede configurar a través de la interfaz mostrada en la **Figura 32**.

Cada servicio expone un extremo único y ajusta de forma automática la escala de la infraestructura subyacente para controlar las solicitudes entrantes. No se puede cambiar el nombre del servicio ni la región más adelante.

Implementar una revisión desde una imagen de contenedor

URL de la imagen del contenedor SELECCIONAR **1**

[REALIZAR PRUEBAS CON UN CONTENEDOR DE MUESTRA](#)
Debe detectar las solicitudes HTTP en \$PORT y no depender del estado local. [¿Cómo se compila un contenedor?](#)

Implementar de forma continua revisiones nuevas desde un repositorio de código fuente

Nombre del servicio *

Región *
us-central1 (Iowa) **2**

[¿Cómo se selecciona la región?](#)

Figura 32. Configuración región del servicio Cloud Run

1. La imagen del contenedor: al realizar los dos comandos previamente mencionados aparecerá el nombre de la imagen para ser seleccionada.
2. La región: es importante tener en cuenta este valor ya que puede afectar la latencia cuando se consume la API, si el servicio será consumido desde Colombia no es recomendable que los servidores estén en Europa porque se elevarían los tiempos de espera.

Otra configuración importante es la asignación de CPU puesto que esto influye en el costo, para el caso de esta solución la CPU solo se asigna cuando llega una petición a la API esto con el fin de disminuir costos, como muestra la **Figura 33**.

Precios y asignación de CPU ?

La CPU solo se asigna durante el procesamiento de la solicitud
Se te cobra por solicitud y solo cuando la instancia de contenedor procesa una solicitud.

La CPU siempre está asignada
Se te cobra por todo el ciclo de vida de la instancia de contenedor.

Ajuste de escala automático ?

Número mínimo de instancias * Número máximo de instancias

Configúralo en 1 para reducir los inicios en frío. [Más información](#)

Interna
Permite el tráfico de VPC y ciertos servicios de Google Cloud en el proyecto, la VPC compartida, el balanceador de cargas HTTP(S) interno y el tráfico que permiten los Controles del servicio de VPC. [Más información](#)

Pública
Permite el acceso directo a tu servicio desde Internet

3

Figura 33. Configuración asignación de CPU en el servicio de Cloud Run

Por último, la plataforma permite configurar variables (4) de entorno como contraseñas desde la misma interfaz gráfica, así como la conexión de la base de datos (5), si se está usando el servicio de base de datos que también provee Google como es el caso de esta aplicación, como lo muestra la **Figura 34**.

Variables del entorno

Nombre 1 Valor 1

[+ AGREGAR VARIABLE](#)

Secretos ?

[CREA UNA REFERENCIA A UN SECRET](#)

Conexiones de Cloud SQL ?

Instancia de Cloud SQL 1

! Se requiere conexión con Cloud SQL.

[+ AGREGAR CONEXIÓN](#)

4

5

Figura 34. Configuración de las variables de entorno y conexión a la base de datos

5.8.2 Configuración de la base de datos

Cloud SQL es un servicio de base de datos ofrecido por Google que soporta a MySQL, PostgreSQL y SQL Server y que garantiza que las bases de datos sean fiables, seguras y escalables para que la actividad empresarial transcurra sin interrupciones, además asegura una disponibilidad superior al 99,95 % en cualquier lugar del mundo [25]. En este proyecto la base de datos se gestiona a través de MySQL y la configuración a través de la interfaz gráfica (**Figura 35**) es la siguiente.

Adicionalmente, se debe de asignar un nombre a la instancia y una contraseña para el usuario raíz o root (1) junto con la versión de la base de datos (2).

Información de la instancia

ID de instancia *

Usa letras minúsculas, números y guiones. Comienza con una letra.

Contraseña *

Establece una contraseña para el usuario raíz. [Más información](#)

Sin contraseña

[POLÍTICA DE CONTRASEÑAS](#)

Versión de la base de datos *

MySQL 8.0

Figura 35. Configuración nombre y contraseña de la instancia de Base de Datos

Adicionalmente, se debe de configurar la región (3) donde estará alojada la instancia y como recomienda Google los datos deben de quedar cerca de los servicios que lo necesitan, por ello es importante que la región sea la misma que donde esta alojada la API, como indica la **Figura 36**.

Elige la región y la disponibilidad zonal

Para obtener un mejor rendimiento, mantén tus datos cerca de los servicios que los necesitan. La región es permanente, mientras que la zona se puede cambiar en cualquier momento.

Región

us-central1 (Iowa)

Disponibilidad zonal

Zona única

Si se suspende el servicio, no se aplica la conmutación por error. No se recomienda esta opción para la producción.

Varias zonas (con alta disponibilidad)

La conmutación por error automática se aplica a otra zona en la región que seleccionaste. Esta opción se recomienda para las instancias de producción. Aumenta el costo.

[ESPECIFICAR ZONAS](#)

Figura 36. Configuración de la Región de la Base de Datos

Por último, si la base de datos no va a ser usada con una alta exigencia de inserciones y consultas, una maquina con núcleo compartido o ligera es más que suficiente (5) y un puto a tener en cuenta es que la capacidad (5) solo puede aumentarse mas no disminuirse una vez creada la instancia, como muestra la **Figura 37**.

Personaliza tu instancia

También puedes personalizar las opciones de configuración de instancias más adelante

Tipo de máquina ^

Tipo de máquina

Selecciona un ajuste predeterminado o personaliza tus ajustes. Para obtener un mejor rendimiento, elige un tipo de máquina con memoria suficiente para soportar tu tabla más grande.

Núcleo compartido

1 CPU virtuales, 0.614 GB

1 CPU virtuales, 1.7 GB

Almacenamiento ^

Tipo de almacenamiento

La elección es permanente. El tipo de almacenamiento afecta el rendimiento.

SSD (recomendado)
Opción más popular. Ofrece una menor latencia que un HDD, con mayor OPS y capacidad de procesamiento de datos.

HDD
Menor rendimiento que SSD con tarifas de almacenamiento más bajas.

Capacidad de almacenamiento

10 - 65,536 GB. Las capacidades mayores mejoran el rendimiento hasta el límite que impone el tipo de máquina. La capacidad no puede disminuirse más adelante.

10 GB

20 GB

100 GB

200 GB

Personalizado

Habilitar los aumentos de almacenamiento automáticos
Si la opción está habilitada, cuando te acerques al límite de capacidad, el almacenamiento se incrementará de manera progresiva (y permanente) [Más información](#)

Figura 37. Configuración de la máquina y cantidad de almacenamiento de la Base de Datos

5.8.3 Despliegue de los módulos de Automatización

Para desplegar los módulos de automatización se debe asignar el nombre de la Cloud Function y la región (1) que debe ser la misma donde se encuentra la API, además del tipo de activador (2). El activador indica que acción hará que se ejecute la Cloud Function, en este proyecto el tipo de activador es HTTP puesto que internamente el Backend con la API se comunicará a través la URL (3) con cada módulo, como indica la **Figura 38**.

1 Configuración — 2 Código

Conceptos básicos

Entorno
1ª gen.

Nombre de la función *
function-1

Región
us-central1

Activador

HTTP

Tipo de activador
HTTP

URL

https://us-central1-bayport-verificacion-test.cloudfunctions.net/function-1

Autenticación

Permitir invocaciones sin autenticar
Marca esta opción si estás creando una API o un sitio web públicos.

Autenticación obligatoria
Administrar los usuarios autorizados con Cloud IAM.

El HTTPS es obligatorio

GUARDAR CANCELAR

Figura 38. Configuración de la región y el activador de las Cloud Functions

A cada módulo se le asignó una cantidad de memoria y tiempo de espera diferentes dependiendo de la complejidad de las tareas y la latencia del portal web de cual se encargaba (**Figura 39**), la memoria podía ir desde 256 MB en los casos donde un portal web contaba con API hasta 1 GB en los casos donde había que crear un Bot con Puppeteer.

Configuración del entorno de ejecución, la compilación, las conexiones y la seguridad

< TIEMPO DE EJECUCIÓN COMPILACIÓN CONEXIONES >

Memoria asignada *
256 MB

Tiempo de espera *
60 segundos

Ajuste de escala automático

Número mínimo de instancias: 0

Número máximo de instancias: 3000

Figura 39. Configuración memoria y tiempo de ejecución Cloud Functions

Después de configurar la Cloud Function se procedía a implementar la lógica de programación, el mismo portal de Google ofrece un editor en la web o bien es posible acceder desde editores de texto como Visual Studio Code, al guardar los cambios función queda disponible para ser utilizada a través de su URL, como indica la **Figura 40**.

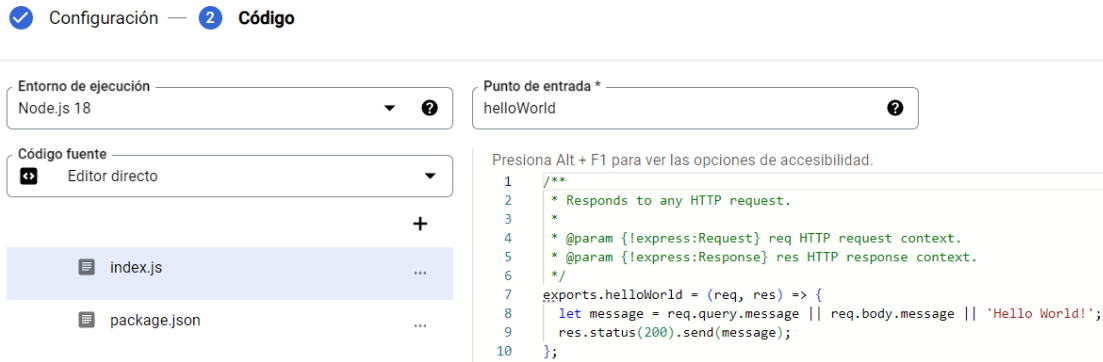


Figura 40. Implementación de la lógica de programación en las Cloud Functions

5.8.4 Configuración de los Bucket

Los Buckets permiten guardar datos no estructurados y se usan para almacenar y controlar el acceso a datos [26] por lo que se hace uso para depositar los certificados proporcionados por el solicitante, así como los obtenidos por los módulos, su configuración es simple, se requiere de un nombre y región, en este se seleccionó Multi-región puesto que se garantiza la mayor disponibilidad de la información mediante redundancia, como se muestra en la **Figura 41**.

The screenshot shows the configuration steps for a new bucket in the Google Cloud console. The first step is 'Asigna un nombre a tu bucket' (Assign a name to your bucket), which requires a globally unique name. A text input field contains the example name 'Ej, 'example', 'example_bucket-1', or 'example.com''. Below the input, a suggestion reads 'Sugerencia: No incluyas información sensible'. The second step is 'Elige dónde almacenar tus datos' (Choose where to store your data), which defines the geographic location. Under 'Tipo de ubicación' (Location type), the 'Multi-region' option is selected, with the subtext 'Máxima disponibilidad en el área más amplia'. A dropdown menu shows 'us (varias regiones en Estados Unidos)'. Other options include 'Dual-region' (High availability and low latency in 2 regions) and 'Region' (Lowest latency within a single region). Both steps have 'CONTINUAR' (CONTINUE) buttons.

Figura 41. Configuración de Buckets

6. LECCIONES APRENDIDAS

Como estudiante de ingeniería de sistemas el realizar una práctica profesional proporciona una experiencia enriquecedora, no solamente en cuanto a conocimiento técnico se refiere sino también a aquellas habilidades necesarias para trabajar en equipo, como lo son el manejo del tiempo, la comunicación clara dentro del equipo o el manejo del estrés. Aunque seis meses no es una gran cantidad tiempo si es lo suficiente para formarse una idea de cómo funciona el panorama laboral dentro del mercado y adquirir la suficiente experiencia para iniciar con pie derecho la vida laboral. Si bien apropiarse de los conocimientos en automatización es una ganancia resultada de esta práctica también lo son los errores e inconvenientes que se presentaron durante la misma porque de ahí también se aprende. En general, la práctica profesional ofrece una valiosa preparación, de la cual se espera encarar de manera exitosa el futuro laboral.

Hay que tener en cuenta que para ciertos casos, no basta únicamente con que los pasos del proceso a automatizar sean de baja complejidad, puesto que el proceso puede tener de manera implícita otros factores que determinen su complejidad, como lo fueron las capas de seguridad presentes en el portal del Ejército, al hacer el análisis de cómo se obtener un certificado de este portal, la atención estuvo únicamente en los pasos a realizar y la seguridad se consideran como características que debían sortearse algún modo sin hacer demasiado hincapié en ellas, así mismo el uso intensivo del módulo no se consideró como un factor que pudiera afectar su funcionamiento puesto que abordó desde un tema de infraestructura que estaba saldado gracias a la escala automática que ofrecen las Cloud Functions sin embargo el hecho de analizar estos factores de manera independiente y superficial (para el caso de las capas de seguridad) tuvieron como consecuencia que el proceso fuera visto como candidato a ser automatizado. Para futuros proyectos, queda el aprendizaje de no solamente analizar los pasos del proceso sino también factores como barreras de seguridad o el uso que se le dará al módulo, este análisis puede facilitarse si se estandariza a través de métricas o heurísticas que terminen qué tan factible es que un proceso puede ser automatizado.

Si bien los módulos desplegados mediante Cloud Function son independientes al backend con la API, es recomendable llevar un registro de los cambios que se realizan para poder llevar un control de las versiones, además como práctica está la de asociar los cambios hechos con las tareas que dieron origen a los cambios ya sean bugs o nuevos requerimientos, muchas herramientas de gestión de proyectos como Jira ya integran la capacidad de relacionar una tarea a un cambio hecho con un sistema de control de versiones como git , esto facilita el mantenimiento a futuro porque se conoce que ocurrió para generar dicho cambio y así no generar más bugs al intentar solucionar otros.

Es importante hacer uso de registros para conocer el estado de las operaciones llevadas a cabo por una aplicación, el uso de registros facilita la detección y corrección de errores

de cara al desarrollador y permite ubicarse más rápido en donde ocurrió el error, entre más descriptivo el registro, más fácil se hace el trabajo de depurar.

7. CONCLUSIONES

A medida que avanza la tecnología, el trabajo para muchas personas cambia, las empresas cada vez desean implementar estrategias que permitan mejorar su desempeño, el RPA hace posible que los trabajadores centren su tiempo y esfuerzo en tareas realmente importantes, así pues, la aplicación Bayport verificación generó valor dentro de la empresa Bayport en donde ahora sus trabajadores realizan el proceso para la solicitud de un crédito en menor tiempo. Los trabajadores necesitaron de capacitaciones, puesto que la manera que realizan su trabajo cambió gracias a la tecnología y es muy probable que siga cambiando entre más avances hallan.

El adquirir conocimientos en Cloud se ha convertido en una necesidad puesto que muchas de las soluciones que se hacen hoy en día tienen en cuenta la infraestructura de proveedores ya sea Amazon, Google o Microsoft y el conocimiento que se tenga influye directamente en los costos de las soluciones, en muchos casos el proveedor ya cuenta con servicios diseñados para permitir una rápida escalabilidad y mantenimientos pero también cuenta con servicios donde la escalabilidad y mantenimiento recaen sobre el desarrollador esto afecta directamente los costos de infraestructura por lo tanto muchas compañías ya apuntan a capacitar o contratar personal que cuente con el manejo de la nube.

La estimación de los requerimientos desarrollados contó con una buena estimación, teniendo en cuenta que el estudiante tuvo un proceso de adaptación los primeros meses al entorno laboral y durante ese tiempo la eficiencia al desarrollar no era la óptima pues muchas veces requirió de consultar y asesorías para resolver dudas, sin embargo, al pasar el tiempo se adquirían habilidades para la resolución de problemas, así como aumentaba el dominio con las tecnologías usadas.

Los servicios proporcionados por los proveedores de la nube van mucho más allá de alquilar una máquina, servicios como Cloud Run o Cloud Functions facilitan el desarrollo de software y ofrecen más alternativas que se pueden adaptar mejor a cada proyecto según las necesidades particulares, ya alternativas para reducir costos, maximizar rendimiento o hacer inteligencia artificial.

Si bien al momento de desarrollar, el conocimiento técnico es un eje fundamental, hay que tener presente que el desarrollo de software no se hace de manera individual sino en equipos cuyos integrantes son de diferentes áreas, por lo que fortalecer y adquirir habilidades interpersonales es de igual manera importante ya que permiten crear un ambiente de trabajo más ameno, donde resalte valores como el respeto, la confianza y la tolerancia así mismo el fortalecer las habilidades en cuanto a comunicación es clave para dar a entender las ideas y entender a los demás, son el conjunto de habilidades técnicas e interpersonales las que permitieron que el desarrollo de la presente práctica se llevará con éxito.

8. REFERENCIAS BIBLIOGRÁFICAS

- [1] J. Ribeiro, R. Lima, T. Eckhardt, and S. Paiva, "Robotic Process Automation and Artificial Intelligence in Industry 4.0 – A Literature review," *Procedia Comput Sci*, vol. 181, pp. 51–58, Jan. 2021, doi: 10.1016/J.PROCS.2021.01.104.
- [2] I. Ángel, B. Parejo, L. David, N. Nuñez, and W. A. Niebles Nuñez, "ANÁLISIS DE LA TRANSFORMACIÓN DIGITAL DE LAS EMPRESAS EN COLOMBIA: DINÁMICAS GLOBALES Y DESAFÍOS ACTUALES ANALYSIS OF THE DIGITAL TRANSFORMATION OF COMPANIES IN COLOMBIA: GLOBAL DYNAMICS AND CURRENT CHALLENGES."
- [3] J. F. Becerra Gaitán, P. Gómez Mogollón, F. S. Rodríguez Montoya, and D. A. Tibavizco, "Implementación de tecnología robótica (RPA) en procesos logísticos. Caso de estudio : organización de servicios petroleros." <https://repository.javeriana.edu.co/handle/10554/45210> (accessed Oct. 12, 2022).
- [4] J. López Túñez, C. Toural Bran, and S. Cacheiro Requeijo, "Uso de bots y algoritmos para automatizar la redacción de noticias: percepción y actitudes de los periodistas en España." <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2018.jul.04/40574> (accessed Oct. 12, 2022).
- [5] R. Diouf, E. N. Sarr, O. Sall, B. Birregah, M. Bousso, and S. N. Mbaye, "Web Scraping: State-of-the-Art and Areas of Application," *Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019*, pp. 6040–6042, Dec. 2019, doi: 10.1109/BIGDATA47090.2019.9005594.
- [6] E. Persson, "Evaluating tools and techniques for web scraping," 2019, Accessed: Oct. 12, 2022. [Online]. Available: <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-271206>
- [7] M. A. Mascheroni and E. Irrazábal, "Framework para la creación y ejecución de pruebas automatizadas sobre servicios REST".
- [8] L. Willcocks and A. Craig, "The Outsourcing Unit Working Research Paper Series Paper 15/02 Robotic Process Automation at Telefónica O2," 2015, Accessed: Oct. 12, 2022. [Online]. Available: www.outsourcingunit.org.
- [9] "Automatización Robótica de Procesos (RPA)," 2017.
- [10] "Cloud Functions overview | Cloud Functions Documentation | Google Cloud." <https://cloud.google.com/functions/docs/concepts/overview> (accessed Apr. 06, 2023).
- [11] S. Tilkov and S. Vinoski, "Node.js: Using JavaScript to build high-performance network programs," *IEEE Internet Comput*, vol. 14, no. 6, pp. 80–83, Nov. 2010, doi: 10.1109/MIC.2010.145.
- [12] S. Tilkov and S. Vinoski, "Node.js: Using JavaScript to build high-performance network programs," *IEEE Internet Comput*, vol. 14, no. 6, pp. 80–83, Nov. 2010, doi: 10.1109/MIC.2010.145.

- [13] H. Sun, C. Humer, D. Bonetta, and W. Binder, "Efficient dynamic analysis for node.js," *CC 2018 - Proceedings of the 27th International Conference on Compiler Construction, Co-located with CGO 2018*, vol. 2018-February, pp. 196–206, Feb. 2018, doi: 10.1145/3178372.3179527.
- [14] "Puppeteer | Puppeteer." <https://pptr.dev/> (accessed Oct. 12, 2022).
- [15] "Axios." <https://axios-http.com/> (accessed Oct. 12, 2022).
- [16] "cheerio." <https://cheerio.js.org/> (accessed Oct. 12, 2022).
- [17] B. B. Rad, H. J. Bhatti, and M. Ahmadi, "An Introduction to Docker and Analysis of its Performance Multidimensional Readiness Index for Electrification of Transportation System View project Smart Grid View project An Introduction to Docker and Analysis of its Performance," *IJCSNS International Journal of Computer Science and Network Security*, vol. 17, no. 3, 2017, Accessed: May 31, 2023. [Online]. Available: <https://www.researchgate.net/publication/318816158>
- [18] "What is a Container? | Docker." <https://www.docker.com/resources/what-container/> (accessed May 31, 2023).
- [19] T. G. C. Hernández, C. E. G. Campo, and C. J. P. Calvache, "Hacia el escalamiento de soluciones ágiles en grandes empresas de software: un mapeo sistemático," *INGE CUC*, vol. 16, no. 2, pp. 180–191, Sep. 2020, doi: 10.17981/INGECUC.16.2.2020.13.
- [20] P. Adi and G. Permana, "Scrum Method Implementation in a Software Development Project Management," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 6, no. 9, 2015, Accessed: Apr. 07, 2023. [Online]. Available: www.ijacsa.thesai.org
- [21] K. Schwaber and J. Sutherland, "2020-Scrum-Guide-Spanish-Latin-South-American".
- [22] R. S. Pressman, *Ingeniería del Software. Un Enfoque Practico*, 7th ed.
- [23] B. G. Sudarsono, "Adopting SCRUM Framework in a Software Development of Payroll Information System," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 3, pp. 2604–2611, Jun. 2020, doi: 10.30534/IJATCSE/2020/17932020.
- [24] "Cloud Run: del contenedor al entorno de producción en cuestión de segundos | Cloud Run | Google Cloud." <https://cloud.google.com/run?hl=es> (accessed May 31, 2023).
- [25] "Cloud SQL para PostgreSQL, MySQL y SQL Server | Cloud SQL: Servicio de bases de datos relacionales | Google Cloud." <https://cloud.google.com/sql?hl=es> (accessed May 31, 2023).
- [26] "Acerca de los buckets de Cloud Storage | Cloud Storage | Google Cloud." <https://cloud.google.com/storage/docs/buckets?hl=es-419> (accessed May 31, 2023).