Instituto Tecnológico de Costa Rica

Escuela de Ingeniería en Computadores

Programa de Licenciatura en Ingeniería en Computadores



Optimización y Migración del Proyecto “Sistema de procesamiento de Big Data para el agendamiento de citas” a AWS Lambda

Informe Final de Proyecto de Graduación

Avance 3

Autor: José Agustín Venegas Vega

Coordinador del Proceso: Jennier Sonalo Cordero

Profesor Asesor: Marco Hernández Vásquez

# Hoja de Aprobación

# Dedicatoria

Dedico este proyecto a todas aquellas personas que influyeron en mi desarrollo personal y profesional, amigos, compañeros de estudio, compañeros de trabajo y por supuesto, profesores, los cuales han sido una constante indispensable en muchos aspectos de mi trayectoria. De una forma especial, este proyecto es dedicado a mi familia, a mi padre, a mi madre, a mi hermano y a mi hermana, siendo estos los pilares de mi vida, y piezas fundamentales que componen lo que soy hoy en día.

# Agradecimientos

Agradezco a todas aquellas personas que me dieron su mano, que confiaron en mí, que me apoyaron cuando lo necesité, quienes me hicieron crecer y me permitieron aprender y ser cada día una mejor persona. Quisiera agradecer al Instituto Tecnológico de Costa Rica en general, pero enfatizando, por supuesto, en todo el cuerpo administrativo y ejecutivo de la Escuela de Ingeniería en Computadores. Y, reiterando a los mencionados en la dedicatoria, amigos, profesores, y principalmente, mi familia.

# Resumen

Este proyecto aborda la optimización y migración del sistema de procesamiento de Big Data para el agendamiento de citas médicas a AWS Lambda. Se busca mejorar la escalabilidad, eficiencia y adaptabilidad del sistema existente, integrando algoritmos de aprendizaje automático e inteligencia artificial. La implementación en AWS Lambda también permitirá una gestión de recursos más eficiente y una reducción de costos operativos, al tiempo que pretenderá mejorar la calidad del servicio en las instituciones de salud de Costa Rica a largo plazo.

# Abstract

This project addresses the optimization and migration of the Big Data processing system for medical appointment scheduling to AWS Lambda. It aims to improve the scalability, efficiency, and adaptability of the existing system by integrating machine learning algorithms and artificial intelligence. Implementation in AWS Lambda allows for more efficient resource management and reduced operational costs while enhancing the quality of service in Costa Rican health institutions.

Índice General

[Hoja de Aprobación 2](#_Toc163314792)

[Dedicatoria 3](#_Toc163314793)

[Agradecimientos 4](#_Toc163314794)

[Resumen 5](#_Toc163314795)

[Abstract 6](#_Toc163314796)

[Índice de Tablas 9](#_Toc163314797)

[Índice de Figuras 10](#_Toc163314798)

[Siglas y Acrónimos 11](#_Toc163314799)

[Introducción 11](#_Toc163314800)

[Antecedentes del Proyecto 11](#_Toc163314801)

[Descripción de la Institución 11](#_Toc163314802)

[Área de Conocimiento que se desarrollará en el Proyecto 12](#_Toc163314803)

[Trabajos Similares Encontrados 12](#_Toc163314804)

[Planteamiento del Problema 13](#_Toc163314805)

[Contexto del Problema 13](#_Toc163314806)

[Justificación del Problema 14](#_Toc163314807)

[Enunciado del Problema 14](#_Toc163314808)

[Objetivos del Proyecto 15](#_Toc163314809)

[Objetivo General 15](#_Toc163314810)

[Objetivos Específicos 15](#_Toc163314811)

[Alcances, Entregables y Limitaciones del Proyecto 15](#_Toc163314812)

[Alcances 15](#_Toc163314813)

[Entregables 16](#_Toc163314814)

[Limitaciones 16](#_Toc163314815)

[Marco de Referencia Teórico 17](#_Toc163314816)

[Gestión de Citas Médicas en Costa Rica 17](#_Toc163314817)

[Big Data 17](#_Toc163314818)

[Procesamiento de Big Data en Salud 18](#_Toc163314819)

[Amazon Web Services (AWS) 19](#_Toc163314820)

[AWS Lambda 20](#_Toc163314821)

[AWS S3 21](#_Toc163314822)

[AWS API Gateway 22](#_Toc163314823)

[Inteligencia Artificial 23](#_Toc163314824)

[Inteligencia Artificial en Salud 24](#_Toc163314825)

[Aspectos Legales y Éticos de la Inteligencia Artificial en la Salud 24](#_Toc163314826)

[Seguridad de la Información en Salud 25](#_Toc163314827)

[Marco Metodológico 27](#_Toc163314828)

[Diagnóstico Inicial 27](#_Toc163314829)

[Diseño del Sistema 27](#_Toc163314830)

[Selección de Tecnologías 28](#_Toc163314831)

[Planificación 29](#_Toc163314832)

[Estrategias de trabajo 29](#_Toc163314833)

[Actividades y Fases del proyecto 30](#_Toc163314834)

[Implementación basada en tareas 31](#_Toc163314835)

[Secuencia de desarrollo 31](#_Toc163314836)

[Monitoreo y Evaluación 32](#_Toc163314837)

[Documentación 32](#_Toc163314838)

[Descripción del Trabajo Realizado 33](#_Toc163314839)

[Descripción del Proceso de Solución 33](#_Toc163314840)

[Análisis de Resultados Obtenidos 33](#_Toc163314841)

[Conclusiones y Recomendaciones 34](#_Toc163314842)

[Apéndices y Anexos 35](#_Toc163314843)

[Referencias 36](#_Toc163314844)

# Índice de Tablas

[Tabla 1. Tareas planeadas para implementación. 31](#_Toc163314351)

# Índice de Figuras

[Figura 1. Diagrama Usuario – AWS 23](#_Toc163311968)

[Figura 2. Relación entre módulos. 28](#_Toc163311969)

# Siglas y Acrónimos

* AWS: Amazon Web Services
* Lambda: AWS Lambda (Servicio de cómputo sin servidor de AWS)
* S3: Amazon S3 es un servicio de almacenamiento en la nube altamente escalable ofrecido por AWS.
* IA: Inteligencia Artificial
* ML: Machine Learning (Aprendizaje automático)
* BD: Big Data
* API: Application Programming Interface (Interfaz de Programación de Aplicaciones)
* SQL: Structured Query Language (Lenguaje de Consulta Estructurada)
* CSV: Comma-Separated Values (Valores Separados por Comas)

# Introducción

La Escuela de Ingeniería en Computadores está avanzando en un proyecto innovador para el desarrollo y mejora de un sistema de gestión de citas médicas, utilizando tecnologías de Big Data. Este proyecto tiene como objetivo principal mejorar la eficiencia y efectividad en la asignación de citas médicas en instituciones de salud en Costa Rica, aprovechando la capacidad de procesamiento y análisis de grandes volúmenes de datos. Se busca crear un sistema que no solo maneje eficientemente la información existente, sino que también proporcione análisis predictivos y recomendaciones basadas en patrones de datos.

Uno de los propósitos que este proyecto pretende lograr, es la mejora social en términos de Salud, ya que un sistema más justo y preciso aumentará significativamente la satisfacción, así como también disminuirá el descontento.

Este documento detalla el contexto, los antecedentes, las tecnologías empleadas, y todo trabajo a realizar con el fin de alcanzar los objetivos trazados.

## Antecedentes del Proyecto

### Descripción de la Institución

El Instituto Tecnológico de Costa Rica (ITCR), una institución educativa líder en el país se destaca en el campo de la ingeniería y la tecnología. Con un fuerte enfoque en la excelencia académica y la investigación aplicada. El ITCR ofrece programas avanzados en diversas áreas de la tecnología, preparando a sus estudiantes para enfrentar desafíos contemporáneos. El ITCR demuestra su compromiso con la innovación tecnológica y el impacto social positivo en Costa Rica. El presente proyecto será desarrollado específicamente bajo la Escuela de Ingeniería en Computadores.

### Área de Conocimiento que se desarrollará en el Proyecto

El proyecto de desarrollo y mejora del sistema de gestión de citas médicas, que se implementará en AWS Lambda, abarca un área de conocimiento que incluye la actualización de modelos de procesamiento de datos y la adaptación a un entorno de nube, complementado con una consideración estratégica del hardware. Este enfoque multidisciplinario se centra en refinar algoritmos para incrementar la eficiencia en el procesamiento y análisis de datos, esencial para una gestión de citas más efectiva y basada en análisis profundos. La actualización de estos modelos no solo busca mejorar la velocidad y precisión, sino también su adaptabilidad al dinámico entorno de las instituciones de salud, respaldado por una infraestructura de hardware optimizada para la nube.

La transición a AWS Lambda representa una oportunidad para aprovechar la flexibilidad y escalabilidad de la nube, lo que implica un rediseño del modelo para alinearlo con este entorno virtualizado. Este cambio facilita una gestión de recursos de hardware más eficiente y adaptativa, asegurando que el sistema pueda responder ágilmente a las variaciones en la carga de trabajo y manejar eficientemente un volumen creciente de datos, sin las limitaciones de un hardware físico tradicional.

Además, la incorporación de técnicas avanzadas de aprendizaje automático e inteligencia artificial mejorará la precisión y efectividad del sistema. Estas tecnologías permitirán una gestión de citas más eficiente y la generación de perspectivas a partir de los patrones y tendencias de datos.

En resumen, el área de conocimiento que se desarrollará en este proyecto integra la mejora de algoritmos de procesamiento de datos, la adaptación a la infraestructura de nube de AWS Lambda y la consideración estratégica del hardware, junto con la integración de inteligencia artificial y aprendizaje automático. Esto orientado a crear un sistema de gestión de citas médicas más eficiente, adaptativo, escalable y alineado con las capacidades de un entorno de computación moderno.

### Trabajos Similares Encontrados

Como principal antecedente al proyecto de desarrollo y mejora del sistema de gestión de citas médicas, tenemos un proyecto inicial que sentó las bases para esta iniciativa. El proyecto “Desarrollo de un sistema de procesamiento de Big Data para la gestión y asignación de citas.” realizado por el estudiante del ITCR José Agustín Venegas Vega, se centró en la gestión de citas médicas, logrando un sistema funcional y con conclusiones altamente factibles de su funcionalidad a gran escala. Este trabajo se limitó en la escalabilidad, la mantenibilidad, y la complejidad de uso, esto debido principalmente a su enfoque de desarrollo local.

Además, se encontró un proyecto con una finalidad similar, desarrollado en la Universidad de Valladolid en el año 2019, el cual se enfoca en el desarrollo de un sistema de Big Data para mejorar la gestión y explotación de datos en un centro de salud, utilizando tecnologías como Apache Hadoop y Apache Spark. Los objetivos de este no corresponden en su totalidad a los objetivos planteados en el presente proyecto, ya que el proyecto “Sistema de Gestión de Citas Médicas para un Centro de Salud” se centra en optimizar el procesamiento de grandes volúmenes de datos para mejorar la toma de decisiones y la calidad del servicio en el ámbito sanitario, diferenciándose, el presente proyecto mejora un sistema con dichas capacidades, a su vez lo migra a la nube y brinda al usuario utilidades como generación de reportes. (Alberto Segovia C, 2019)

## Planteamiento del Problema

En esta sección, se proporcionará una descripción detallada de la estructura y los objetivos del proyecto de desarrollo y mejora del sistema de gestión de citas médicas, que se implementará en AWS Lambda. Se abordarán tanto las metas específicas que se pretenden alcanzar con esta iniciativa como las razones fundamentales que justifican la evolución del proyecto inicial. Del mismo modo, se discutirán las bases que sustentan la investigación académica a realizar, enfocándose en cómo la integración de tecnologías avanzadas y la adaptación a un entorno de nube pueden significativamente mejorar la eficiencia, escalabilidad y adaptabilidad del sistema existente.

### Contexto del Problema

El proyecto original de sistema de gestión de citas médicas, aunque exitoso en su implementación, reveló limitaciones en cuanto a escalabilidad, eficiencia de procesamiento de datos, y adaptabilidad. Estas limitaciones se hicieron evidentes a medida que se intentó aumentar el volumen de datos a almacenar y la cantidad de ejecuciones en procesar. El sistema requería una solución que pudiera manejar de manera más eficiente un gran volumen de datos, adaptarse a las cambiantes necesidades de las instituciones de salud, y ofrecer una gestión de citas médicas más efectiva y basada en datos.

La experiencia adquirida y las lecciones aprendidas del proyecto inicial han sido fundamentales para identificar oportunidades de mejora. Aspectos como la escalabilidad del sistema, la eficiencia en el procesamiento de datos, y la adaptabilidad a cambios y demandas crecientes surgieron como áreas clave para el desarrollo. Este conocimiento ha sido esencial para formular el nuevo proyecto, que no solo busca abordar y perfeccionar estas áreas, sino también integrar tecnologías avanzadas y prácticas de vanguardia, como la migración a AWS Lambda y la implementación de algoritmos más sofisticados de aprendizaje automático e inteligencia artificial.

### Justificación del Problema

La migración y mejora del sistema a AWS Lambda se justifica por la necesidad de abordar estas limitaciones del sistema indicadas en la sección anterior. AWS Lambda ofrece una solución que mejora significativamente la escalabilidad y la eficiencia en el procesamiento de datos, gracias a su capacidad de auto escalado y su modelo de precios basado en el uso real. Esto permite al sistema adaptarse a variaciones en la carga de trabajo sin la necesidad de una infraestructura de hardware costosa y de alto mantenimiento. Adicionalmente, la integración de tecnologías avanzadas como el aprendizaje automático y la inteligencia artificial en AWS Lambda mejorará la capacidad del sistema para analizar datos, ofrecer puntos de vista más profundos y mejorar la toma de decisiones basada en datos. Estas mejoras son fundamentales para satisfacer las necesidades actuales y futuras de las instituciones de salud, proporcionando un servicio más eficiente y adaptativo.

La elección de AWS Lambda como plataforma para este proyecto se justifica por sus múltiples ventajas. Principalmente, AWS Lambda ofrece un modelo de precios basado en el uso real, lo que significa que los costos se alinean directamente con el tiempo de ejecución del código y la cantidad de activaciones del sistema, minimizando así los costos operativos​​. Además, la capacidad de AWS Lambda para escalar automáticamente se ajusta de manera eficiente a las variaciones en la demanda, eliminando la necesidad de administrar y mantener servidores físicos​​. (AWS-Lambda Guide, s.f)

Estas características hacen de AWS Lambda una solución ideal para el proyecto, ya que facilitan un sistema de gestión de citas médicas más adaptable, escalable y económicamente eficiente. Con la integración de tecnologías avanzadas de aprendizaje automático e inteligencia artificial, el sistema no solo mejora la gestión de citas médicas, sino que también ofrece análisis de datos más profundos y precisos, lo que resulta en una toma de decisiones mejor informada.

En resumen, este proyecto busca desarrollar un sistema de gestión de citas médicas que no solo sea más eficiente y adaptable a las necesidades cambiantes, sino que también aproveche las ventajas económicas y técnicas de un entorno de nube moderno como AWS Lambda​​. (AWS-Lambda Guide, s.f)

### Enunciado del Problema

¿Cómo mejorar y escalar un sistema existente de gestión de citas médicas para adaptarse a las crecientes y cambiantes demandas de las instituciones de salud, implementando una solución en AWS Lambda que asegure una mayor eficiencia en el procesamiento de datos y escalabilidad automática, e integrando tecnologías avanzadas como el aprendizaje automático y la inteligencia artificial, para proporcionar una gestión más efectiva y basada en datos y así mejorar la calidad del servicio en múltiples entornos de salud?

## Objetivos del Proyecto

Este proyecto busca rediseñar y optimizar un sistema de gestión de citas médicas, aprovechando las capacidades de AWS Lambda y las tecnologías de aprendizaje automático e inteligencia artificial.

### Objetivo General

Rediseñar el sistema de gestión de citas médicas existente utilizando AWS Lambda, integrando algoritmos de aprendizaje automático e inteligencia artificial, así como la adición de un generador de reportes, con el fin de mejorar la eficiencia, adaptabilidad y calidad del servicio en las instituciones de salud de Costa Rica, mediante el uso de tecnologías en la nube que potencien la eficiencia, modularización y la escalabilidad.

### Objetivos Específicos

1. Generar una lista de las deficiencias del sistema existente de gestión de citas médicas, así determinando las áreas de mejora en escalabilidad, eficiencia y modularización, mediante pruebas exhaustivas del sistema.
2. Adaptar el sistema de gestión de citas médicas existente, utilizando una infraestructura sin servidor para lograr escalabilidad, eficiencia y modularización, proporcionando una opción potente, cómoda y rápida, mediante la utilización de Lambda AWS.
3. Optimizar el sistema de gestión de citas médicas, abordando específicamente las áreas de deficiencia identificadas, así proporcionando un sistema más completo y potente, basándose principalmente en algoritmos de aprendizaje automático e inteligencia artificial.
4. Desarrollar un módulo dentro del sistema que permita a los usuarios la generación de reportes específicos basados en filtros seleccionados, facilitando el análisis detallado y personalizado de la información de citas, pacientes y tendencias de uso, mediante el uso de tecnologías de análisis de datos.

## Alcances, Entregables y Limitaciones del Proyecto

### Alcances

1. Implementación libre de errores en AWS Lambda: Completar con sin errores, la migración y optimización del sistema de gestión de citas médicas en AWS Lambda, logrando un sistema sin servidor, escalable, modularizado y eficiente en términos de costos computacionales.
2. Capacidad de Análisis Predictivo: Implementar con éxito la funcionalidad de análisis predictivo y recomendaciones personalizadas basadas en Inteligencia Artificial, lo que permite un agendamiento más preciso de las citas, haciendo un análisis de datos más detallado.
3. Reportes Basados en Filtros: Desarrollar un módulo de generación de reportes que permita a los usuarios obtener información. Se enfoca en proveer la visualización de los datos de forma gráfica, facilitando el análisis de los datos de citas y pacientes.
4. Mejora en la Eficiencia de Asignación de Citas: Demostrar una mejora significativa en la eficiencia y precisión de la asignación de citas médicas a través de algoritmos de aprendizaje automático, realizando una comparación con el sistema anteriormente desarrollado.

### Entregables

1. Sistema migrado a Lambda AWS, completamente funcional.
2. Interfaz de usuario para consultas e ingreso de datos.
3. Sistema de generación de reportes integrado.

### Limitaciones

1. Dependencia de la Conectividad a Internet: La conectividad a internet se presenta como una necesidad a partir de la migración, debido a la naturaleza basada en la nube del sistema.
2. Limitaciones de Presupuesto: Restricciones presupuestarias podrían limitar la escala de implementación o las características del sistema, debido que, al presentarse como un sistema en la nube, su eficiencia y calidad depende de la velocidad y potencia de los servicios contratados, involucrando aspectos como almacenamiento de memoria, unidades de procesamiento, etc.
3. Resistencia al Cambio: Puede haber resistencia al cambio por parte de los usuarios o instituciones, lo que afectaría la adopción y eficacia del sistema a largo plazo.
4. Limitaciones Tecnológicas: A pesar de las capacidades de AWS Lambda, puede haber limitaciones en cuanto a la adición de procesamientos complejos o pesados, los cuales lleven al límite el sistema o inclusive no permitan su implementación parcialmente.
5. Variabilidad en el Rendimiento de IA: Los algoritmos de inteligencia artificial pueden tener variaciones en su rendimiento dependiendo de aspectos como la cantidad de datos y la naturaleza de estos, lo cual podría significar una limitante de peso en términos de resultados.

# Marco de Referencia Teórico

## Gestión de Citas Médicas en Costa Rica

La gestión de citas médicas es un proceso fundamental en la organización de la atención sanitaria. Este proceso implica la asignación eficiente de horarios para pacientes y profesionales de la salud, con el objetivo de optimizar los recursos disponibles y reducir los tiempos de espera.

Este proceso es un tanto complicado, el cual se ve regido por un sistema de gestión de colas, es decir, una persona que agende una cita médica tendrá su cita antes que todos aquellos que hayan agendado con posterioridad a dicha persona. En Costa Rica, el proceso más cómodo y moderno para agendar una cita médica es mediante la aplicación EDUS, el cual puede ser accedido mediante un teléfono móvil. Este requiere un proceso de registro, el cual requiere los datos personales de la persona en cuestión, específicamente: correo electrónico, tipo de identificación y número de identificación, así como también, domicilio, establecimiento de salud de comodidad o más cercano, y número telefónico.

Una vez creado el usuario, se debe seleccionar el servicio médico requerido, así como también, la especialidad, médico de preferencia y fecha a convenir en caso de haber disponibilidad. Además, las fechas, horarios y médicos disponibles varían diariamente a partir de las seis de la mañana en punto, y es altamente recomendado el ingreso a tempranas horas debido a la alta demanda. (Araya F, 2023)

El procedimiento actual de asignación de citas médicas no puede modificarse en el presente proyecto. Sin embargo, este proyecto tiene como objetivo permitir una toma de decisiones más detallada y analizada que la que se obtiene mediante el método de asignación por cola. Además, busca proporcionar reportes detallados basados en la lista de usuarios inscritos. Estos reportes permiten a los usuarios visualizar de manera gráfica la información almacenada, lo que les ayudará a evaluar las decisiones tomadas, ya sea por el método actual o por el método que se desarrollará en el proyecto.

## Big Data

El Big Data es un conjunto de tecnologías diseñadas para recopilar, analizar y gestionar los altos volúmenes de datos generados por los usuarios. Estas tecnologías han revolucionado la industria moderna en diversos ámbitos, permitiendo identificar patrones y comportamientos que pueden ser útiles para sectores específicos. Cuando navegamos por Internet, generamos una gran cantidad de datos que las páginas web, aplicaciones, o cualquier entorno virtual recopilan, como nuestros intereses, búsquedas, uso de aplicaciones, entre otros. Estos datos, aunque anonimizados, representan el comportamiento de los usuarios en Internet.

El Big Data aprovecha esta enorme cantidad de datos para analizarlos y obtener información útil. Este no se limita a los datos de navegación, sino que también incluye los datos generados por los usuarios en diferentes aplicaciones o servicios. Cada empresa de servicios puede tener grandes cantidades de datos generados por los usuarios, que luego utilizan herramientas de Big Data para analizar y procesar en busca de beneficios. Los usos del Big Data son casi infinitos, ya que permite a las empresas conocer mejor a sus usuarios y tomar decisiones informadas basadas en las tendencias y comportamientos detectados. Por ejemplo, las redes sociales pueden predecir cambios sociales o nuevas tendencias antes de que se vuelvan populares, y los servicios de streaming pueden ofrecer contenido basado en el interés general de los usuarios. (Fernandez Y, 2021)

Para mejorar el entendimiento general del Big Data en el área de la medicina, se destaca que el rol de ésta en dicha área radica en la cantidad masiva de datos que posee cada entidad de la salud en cada país del mundo, en este caso específico, Costa Rica. Por ejemplo, en Costa Rica la Caja Costarricense de seguro social contaba con 4 190 244 personas afiliadas (Delfinocr. s.f). Dicha cantidad se supone ha crecido en los últimos años debido al evidente aumento de la población, por lo cual el Big Data nos permite tomar esa masiva cantidad de datos, y tomar decisiones, en este caso para favorecer la eficiencia en los sistemas de salud costarricenses.

## Procesamiento de Big Data en Salud

El procesamiento de Big Data en el sector de la salud está revolucionando la forma en que se gestionan los datos y se toman decisiones clínicas. La gran cantidad de datos disponible hoy en día se utilizan en instalaciones médicas en todo el mundo para mejorar la atención al paciente y la gestión de la salud pública. El análisis de Big Data en el ámbito de la salud presenta desafíos únicos debido a la gran cantidad de datos y a la sensibilidad de estos. Es por eso por lo que se utilizan técnicas y herramientas de Big Data Analytics para analizar y extraer información de Big Data, lo que permite predecir tendencias futuras y crear modelos predictivos en el campo de la medicina.

En el sector de la salud, el procesamiento de Big Data permite analizar grandes conjuntos de datos de cantidades importantes de pacientes, identificar grupos y correlaciones entre los datos, y desarrollar modelos predictivos utilizando técnicas de minería de datos. Esto ayuda a mejorar la toma de decisiones clínicas, optimizar los tratamientos y mejorar la eficiencia de los servicios de salud. La implementación de sistemas de Big Data en el sector de la salud permitirá el desarrollo de la medicina personalizada y precisa, basada en información personalizada entregada en tiempo real y debidamente adaptada. Esto conducirá a una mejor comprensión de los factores que influyen en la salud a nivel individual, del sistema de salud y de la sociedad en su conjunto, y a enfoques mejorados para detectar problemas de seguridad con medicamentos y dispositivos. (Batko K, & Ślęzak A, 2022)

## Amazon Web Services (AWS)

Cómo herramienta vital para el desarrollo de esta iniciativa se cuenta con AWS, el cual es una plataforma de computación en la nube completa y en constante evolución proporcionada por Amazon. Incluye una mezcla de servicios de infraestructura como servicio (IaaS), plataforma como servicio (PaaS) y software como servicio empaquetado (SaaS). AWS ofrece herramientas como potencia de cálculo, almacenamiento de bases de datos y servicios de entrega de contenido. AWS ofrece muchas herramientas y productos diferentes para empresas y desarrolladores de software en 245 países y territorios. Agencias gubernamentales, instituciones educativas, organizaciones sin fines de lucro y organizaciones privadas utilizan los servicios de AWS. (Kirvan P. Barney N. & Gillis. A., 2024)

AWS presenta muchas conveniencias que lo hacen destacar, enumeradas a continuación: (RootStack, 2022)

* “Garantiza la disponibilidad de servicios en distintas zonas, evitando interrupciones por fallas eléctricas o eventualidades en los centros de datos.”
* “La infraestructura de AWS está disponible en varias regiones, permitiendo distribuir servicios a través de distintos países en minutos.”
* “Permite pagar únicamente por los recursos utilizados, lo que se traduce en costos operativos ajustados a la demanda real.”
* “Ofrece planes de pago flexibles basados en la cantidad y tiempo de uso de los servicios, lo que permite optimizar los costos.”
* “Ajusta la potencia de los servicios automáticamente según las variaciones de la demanda, sin que el usuario tenga que preocuparse por ello.”
* “Ofrece un modelo de responsabilidad compartida y servicios para detectar y mitigar riesgos de seguridad.”
* “Permite establecer límites de gasto y ofrece análisis de utilización para planificar de forma óptima los presupuestos.”
* “Algunos servicios son administrados de forma independiente, eliminando la necesidad de provisión por parte del usuario.”
* “A través de su Marketplace, ofrece servicios de terceros compatibles y seguros para integrar con la configuración.”
* “Permite provisionar recursos rápidamente y clonar configuraciones, simplificando el proceso de implementación y escalado.”

El desarrollo comprenderá la utilización de al menos 3 herramientas proveídas por esta compañía (Para mejorar el entendimiento del contraste entre dichas herramientas, ver figura 1), las cuales son las siguientes:

### AWS Lambda

AWS Lambda es un servicio de computación sin servidor ofrecido por Amazon Web Services (AWS) que permite ejecutar código o funciones sin la necesidad de administrar servidores. Con Lambda, los desarrolladores pueden cargar su código y definir cuándo y cómo deben ejecutarse las funciones. Esto elimina la necesidad de preocuparse por la infraestructura subyacente, como los servidores, el almacenamiento o la escalabilidad, ya que AWS se encarga de todo eso.

Lambda se basa en el concepto de "serverless computing" o computación sin servidor, donde las funciones individuales se ejecutan de forma independiente en respuesta a eventos específicos, como una solicitud HTTP, un cambio en una base de datos o una carga de archivos. Esto permite a las empresas escalar automáticamente según la demanda y pagar solo por el tiempo de ejecución y los recursos utilizados, lo que puede resultar en costos más bajos y una mayor eficiencia en comparación con los modelos tradicionales de infraestructura. (Correa A., 2024)

Desde su adopción, AWS Lambda se ha convertido en el arma secreta de empresas muy grandes, como Netflix, el cual le dio uso a esta tecnología para diseñar y autogestionar una arquitectura impulsada por reglas. Para perseguir la excelencia, ahorrar tiempo y aumentar la eficiencia, Lambda reemplaza procedimientos obsoletos. Simplifica procesos, especialmente en la manipulación de datos, decisiones de respaldo y controles de seguridad. También automatiza medidas de seguridad y garantiza el cumplimiento durante la configuración de instancias y la respuesta a accesos no autorizados. (Bacancy Technology, 2024)

#### AWS LAMBDA vs Azure Functions

El principal competidor de AWS Lambda es su contraparte en Azure, Azure Functions, la comparación entre estos dos sistemas será detallada a continuación: (Shilkov M, 2019)

* AWS Lambda fue pionero en el modelo de aplicación Function as a Service en 2014, mientras que Azure Functions es la respuesta de Microsoft en su plataforma de nube.
* Ambos servicios ofrecen un enfoque sin servidor que elimina la necesidad de administrar servidores, lo que permite una mayor escalabilidad y eficiencia en el uso de recursos.
* AWS Lambda tiene un modelo de programación simple donde una función recibe un objeto JSON como entrada y devuelve otro JSON como salida.
* Azure Functions utiliza triggers y bindings, lo que permite una mayor flexibilidad en la integración con otros servicios de Azure y en la manipulación de datos.
* AWS Lambda admite JavaScript, Java, Python, Go, C#, F#, PowerShell y Ruby.
* Azure Functions admite JavaScript, Java, Python, C#, F# y PowerShell (en versión preliminar), pero no soporta Go ni Ruby.
* Ambos servicios ofrecen una alta escalabilidad y disponibilidad, pero AWS Lambda tiene una ventaja en la gestión de picos de carga y grandes volúmenes de eventos.
* El modelo de precios de AWS Lambda se basa en el tiempo de ejecución y los recursos utilizados, con opciones de pago por uso.
* Azure Functions también ofrece un modelo de precios similar, pero con opciones de hospedaje adicionales que pueden afectar el costo final.
* Tanto AWS Lambda como Azure Functions ofrecen integración con otros servicios de su respectiva nube, lo que facilita la construcción de aplicaciones complejas.

Como se evidenció anteriormente, ambos tienen pros muy evidentes, y al mismo tiempo son equivalentes en muchos aspectos, en este caso, AWS Lambda resulta una opción más atractiva debido a aspectos de, costos, estructuras de datos utilizadas, ambiente y enfoque.

### AWS S3

AWS S3 (Amazon Simple Storage Service) es un servicio de almacenamiento de objetos en la nube ofrecido por Amazon Web Services (AWS). S3 está diseñado para almacenar y recuperar grandes cantidades de datos de forma segura y escalable. (AWS-S3, s.f.)

A continuación, se detallarán una serie de ventajas que destacan un gran beneficio de la utilización de AWS S3: (Apser, 2018)

* **Facilidad de uso:** Amazon S3 ofrece una amplia variedad de opciones de transferencia de datos, lo que facilita su uso tanto para profesionales como para usuarios menos técnicos. Las API de S3 son confiables y fáciles de manejar, lo que simplifica el almacenamiento y la transferencia de datos.
* **Seguridad:** La seguridad es una prioridad en Amazon S3. Ofrece privacidad y seguridad de extremo a extremo, garantizando la integridad y confidencialidad de los datos almacenados.
* **Disponibilidad:** Amazon S3 garantiza que los datos estarán disponibles el 99,99% del tiempo. Además, ofrece un reembolso del 25% en caso de una disminución en la disponibilidad por debajo del 99%.
* **Costos:** Amazon S3 es una opción rentable para empresas de todos los tamaños, ya que solo se paga por el uso que se hace de los servicios de almacenamiento y transferencia. Los precios de Amazon S3 han disminuido con el tiempo, lo que lo convierte en una opción económica y accesible.
* **Escalabilidad:** Con Amazon S3, es posible ampliar o reducir rápidamente el volumen de datos almacenados y transferidos según sea necesario, sin necesidad de realizar inversiones en infraestructura tecnológica adicional.
* **Facilidad de interacción con aplicaciones de terceros:** Amazon S3 se integra fácilmente con aplicaciones de terceros, lo que permite a los desarrolladores conectar los datos almacenados en S3 con otras aplicaciones móviles o basadas en la web de manera eficiente.
* **Facilidad para distribuir contenido:** Amazon S3 no solo es un servicio de almacenamiento en la nube, sino que también puede utilizarse para crear una red de distribución de contenido, lo que mejora el rendimiento de las aplicaciones al reducir el tiempo de carga y espera.

Es por esto mencionado anteriormente que se destaca AWS S3 como una opción muy sólida y segura para el almacenamiento de los datos, tomando en cuenta la sensibilidad e importancia de estos. Y presenta un sistema de costos muy favorable al usuario. Al mismo tiempo brindaría acceso rápido y sencillo a los datos almacenados.

### AWS API Gateway

AWS API Gateway es un servicio totalmente gestionado que facilita la creación, publicación, mantenimiento, supervisión y protección de APIs en la nube. La adición de AWS API Gateway en el proyecto resulta más una necesidad que una posibilidad, debido a que representa al medio por el cual se podrá realizar la comunicación entre el usuario y el sistema en la nube, siendo este el intermediario que servirá como comunicador entre ambas partes, mediante un API en la nube debidamente protegido y configurable, como se mencionó anteriormente. Además, siendo este el servicio de API de la plataforma que se utiliza la cual es Amazon Web Services.

A diagram of a software application

Description automatically generated

Figura 1. Diagrama Usuario – AWS

## Inteligencia Artificial

La inteligencia artificial (IA) es un campo de la informática que se enfoca en desarrollar sistemas y tecnologías que imitan la función cognitiva humana. Estos sistemas utilizan máquinas, procesadores y software para realizar tareas de procesamiento y análisis de datos, así como para razonar, aprender, tomar decisiones y resolver problemas.

La IA se puede aplicar en una amplia variedad de áreas y sectores, incluyendo la robótica, las ciencias de la computación, las finanzas, la salud, los sistemas de transporte autónomos, los videojuegos y las comunicaciones. Por ejemplo, la IA se utiliza para crear asistentes virtuales, motores de búsqueda, sistemas de reconocimiento de voz y rostro, robots, drones y vehículos autónomos. Esta se clasifica en varios tipos, como máquinas reactivas, con memoria limitada, teoría de la mente y autoconciencia, cada uno con diferentes capacidades y aplicaciones. Estos sistemas pueden manejar grandes cantidades de datos y realizar acciones complejas con rapidez, lo que los hace ideales para tareas que requieren percepción del entorno, toma de decisiones y acciones específicas. (Ferrovial, 2022)

### Inteligencia Artificial en Salud

La inteligencia artificial en el cuidado de la salud está transformando la forma en que se brinda atención médica al aprovechar grandes conjuntos de datos y algoritmos para descubrir patrones que los humanos no podrían encontrar por sí mismos. Algunos beneficios clave de la IA en el cuidado de la salud incluyen: (IBM, s.f)

* Experiencias centradas en el usuario: La IA puede ayudar a encontrar insights de manera más rápida y precisa, lo que resulta en una mayor satisfacción tanto internamente como con los pacientes.
* Mejora de la eficiencia operativa: Al analizar los patrones de datos, la IA puede ayudar a las organizaciones de salud a utilizar mejor sus recursos, aumentando la eficiencia y mejorando el rendimiento de los flujos de trabajo.
* Conexión de datos dispares: La IA puede ayudar a conectar datos de salud fragmentados y en varios formatos para obtener una imagen más unificada de los pacientes.

Un caso de uso común de la IA en el cuidado de la salud es el procesamiento de lenguaje natural, que permite a los algoritmos identificar patrones en el lenguaje utilizado en registros médicos y otros documentos de salud. Esto puede ayudar a los profesionales a encontrar información relevante más rápidamente y tomar decisiones informadas sobre el cuidado de los pacientes.

Además, la IA se utiliza en imágenes médicas para ayudar a los radiólogos y cardiólogos a identificar casos críticos, realizar diagnósticos más precisos y evitar errores potenciales. La IA también puede ayudar a garantizar la equidad en el cuidado de la salud al reducir el sesgo en la investigación y promover la diversidad y la transparencia en los datos.

A pesar de los beneficios, la adopción de la IA en el cuidado de la salud enfrenta desafíos, como cumplir con los requisitos regulatorios y superar los problemas de confianza en los resultados de machine learning. Sin embargo, la IA está mejorando las operaciones, agilizando los flujos de trabajo y ayudando a los profesionales a encontrar respuestas rápidamente, lo que lleva a mejores experiencias para los pacientes y una atención médica más efectiva. (IBM, s.f)

En el presente caso, la Inteligencia Artificial pretende solucionar la problemática de la decisión, a la hora de agendar una cita médica, basándose en los datos de los pacientes, principalmente en las condiciones de salud relevantes a la seguridad y salud personal.

### Aspectos Legales y Éticos de la Inteligencia Artificial en la Salud

A continuación, se listan una serie de aspectos éticos relacionados a la Inteligencia Artificial en la Salud. (Organización Mundial de la Salud, 2021)

* Proteger la autonomía humana: La IA no debe socavar la capacidad humana de tomar decisiones. Las personas deben mantener el control sobre los sistemas de atención de la salud y las decisiones médicas.
* Promover el bienestar y la seguridad de las personas y el interés público: Las tecnologías de IA no deben dañar a las personas y deben cumplir con normas de seguridad, precisión y eficacia.
* Garantizar la transparencia, la claridad y la inteligibilidad: Las tecnologías de IA deben ser comprensibles para desarrolladores, profesionales de la salud, pacientes, usuarios y reguladores. Deben ser explicables y transparentes en su diseño y funcionamiento.
* Promover la responsabilidad y la rendición de cuentas: Debe haber una especificación clara de las tareas que los sistemas de IA pueden realizar y las condiciones en las que pueden alcanzar el rendimiento deseado. Las partes interesadas deben asegurarse de que las tareas se realicen adecuadamente y por personas debidamente capacitadas.
* Garantizar la inclusividad y la equidad: La IA debe ser concebida de manera que fomente la utilización y el acceso equitativos, independientemente de características como la edad, el sexo, el género, los ingresos, la raza, el origen étnico, la orientación sexual o la capacidad. Debe minimizar los sesgos y las disparidades de poder.
* Promover una IA con capacidad de respuesta y sostenible: Los sistemas de IA deben ser evaluados continuamente en situaciones reales para determinar si responden adecuadamente y de acuerdo con las expectativas. También deben ser coherentes con la promoción de la sostenibilidad en los sistemas de salud y el medio ambiente.

## Seguridad de la Información en Salud

La seguridad de la información en el ámbito de la salud es fundamental para preservar la confidencialidad y la privacidad de los datos sensibles de los pacientes. La colaboración y la creación de mecanismos adecuados son necesarios para garantizar la seguridad de la información personal en entornos de salud digital, al mismo tiempo que se promueve el acceso y la transparencia en la información y el conocimiento.

Las tecnologías digitales pueden contribuir al acceso de grupos vulnerables en materia de salud, pero es necesario desarrollar capital humano y una infraestructura que permitan utilizar estas tecnologías de forma inclusiva, ética y segura. Es importante establecer mecanismos de confianza en el entorno digital de la salud pública, considerando dimensiones éticas, técnicas y legales. (Organización Panamericana de la Salud, 2023)

Se evidencia entonces no solo que la importancia en conservar el íntegros los datos en el área de la salud es crucial, sino que en términos virtuales y en lo que al internet se refiere, la seguridad de un sistema debe ser y será tan robusta como así se configure, por lo cual, se considera una fortaleza construir la base de la información en AWS, debido a la solidés de la plataforma en términos de seguridad y las posibilidades de configuración que brinda.

Para efectos del proyecto, la seguridad se respalda y refuerza basándose en las herramientas a utilizar, por ejemplo, los datos se verían contenidos y procesados dentro de la plataforma Amazon Web Services, la cual no únicamente es segura por sí sola, sino que provee herramientas y configuraciones que fortalezcan este aspecto, como se mencionó debidamente en la sección Amazon Web Services (AWS). Herramientas como S3, y API Gateway son las encargadas de proteger la información para efectos prácticos del presente proyecto.

# Marco Metodológico

## Diagnóstico Inicial

Inicialmente se recolecta toda aquella información estructural, lógica y de utilidad del sistema “Desarrollo de un sistema de procesamiento de Big Data para la gestión y asignación de citas.”, como archivos de datos contenidos, el formato de estos, los módulos creados, herramientas utilizadas, vulnerabilidades, la ejecución de pruebas y validación de sus tiempos de respuesta, junto con el feedback obtenido de la misma. El listado de todos estos datos recolectados corresponde a áreas de mejora del proyecto, y puntos de vital importancia a abarcar a futuro.

Esto no solo ayuda a empezar con un progreso pasivo, sino también expande y refresca el conocimiento del desarrollador, proveyendo un panorama más claro de los objetivos.

El estudio de todas las opciones disponibles para la implementación de este sistema en la nube se presenta como un punto de inflexión de vital importancia, puesto que define la inclinación, tendencias y direccionamiento de una de las piezas clave del proceso la cual es el desarrollo en la nube y sus complementos.

## Diseño del Sistema

Se debe definir un sistema una vez comprendidos, los objetivos, el sistema existente y las posibles herramientas en la nube que contribuirán a la implementación. Por lo cual se procede a definir módulos, los cuales serán vistos como cajas negras, ya que a este punto solo sabremos qué debe hacer cada uno, pero no cómo ni con qué. Se logran identificar entonces los siguientes apartados: Interfaz de Usuario, generador de reportes, sistema en la nube, unidad de almacenamiento, modelo, y el conector. (Ver Figura 2)

A continuación, se explica brevemente cada uno de ellos:

**Interfaz de Usuario:** Unidad con la cual interactúa el usuario, ya sea para ingresar datos, o utilizar el generador de reportes.

**Generador de reportes:** Encargado de generar representaciones gráficas de la información contenida basándose en filtros añadidos por el usuario.

**Conector:** Encargado de brindar comunicación entre la interfaz de usuario y el sistema en la nube.

**Sistema en la nube:** Corresponde al sistema construido en la nube, encargado de todo el procesamiento lógico interno y desconocido por el usuario.

**Unidad de procesamiento:** Corresponde a la unidad lógica, la cual se encarga de procesar los datos ingresados por el usuario, realizar las consultas necesarias para la obtención y adición de datos.

**Unidad de almacenamiento:** Unidad encargada únicamente de almacenar datos, preferiblemente con una estructura definida que se preste para su sencilla utilización.

**Modelo:** Esta unidad corresponde al modelo generado con inteligencia artificial, el cual recibe una nueva entrada y retorna el resultado correspondiente. Al mismo tiempo, esta unidad permite recalcularse.

A diagram of a system

Description automatically generated

Figura 2. Relación entre módulos.

## Selección de Tecnologías

Basándose en lo investigado y obtenido en el marco teórico, se considera ideal la utilización de Amazon Web Services como plataforma en la nube, ya que esta provee una serie de herramientas ideales para la implementación de este sistema, cada una con una serie de ventajas que son de gran utilidad. Lambda AWS corresponde a la herramienta en la que se desarrolla la unidad de procesamiento, AWS S3 es la herramienta que funciona como unidad de almacenamiento.

El modelo será desarrollado utilizando un algoritmo de inteligencia artificial no definido actualmente, debido a que dicho paso depende de la evolución del sistema, los datos, y la estructura de estos.

## Planificación

En este apartado se define un orden deseado en el cual se debe atacar la idea de solución planteada, tomando en cuenta la complejidad de los módulos, las necesidades del sistema, las herramientas a utilizar y el conocimiento actual, de este modo se plantea la siguiente secuencia de tareas para llevar a cabo el proyecto. Es por esto por lo que se define la siguiente secuencia de tareas (contemplando las ya realizadas con anterioridad):

### Estrategias de trabajo

La estrategia de trabajo se definió basada en la Metodología Ágil (Scrum), la cual contempla los siguientes puntos como algunos de los más relevantes, y a utilizar durante el desarrollo. (Martins J, 2024)

1. Planificación Iterativa:

* Dividir el proyecto en sprints de duración corta, típicamente de 2 a 4 semanas, para permitir una planificación y ejecución flexible y adaptativa.
* Establecer objetivos claros y alcanzables para cada sprint, garantizando avances regulares y medibles.

1. Reuniones Regulares:

* Organizar reuniones diarias de Scrum para revisar el progreso, identificar obstáculos y planificar actividades para el día siguiente.
* Realizar reuniones de revisión al final de cada sprint con el profesor asesor y el coordinador para presentar los avances y recibir retroalimentación.

1. Roles y Responsabilidades Definidos:

* Asignar roles clave como el Scrum Máster y el Product Owner, asegurando una gestión eficiente del proyecto y una comunicación clara con todas las partes interesadas.
* Fomentar la colaboración y participación de todos los miembros del equipo en la toma de decisiones y en la ejecución del proyecto.

1. Adaptabilidad y Revisión Continua:

* Estar preparados para adaptar la planificación y el enfoque del proyecto en respuesta a los comentarios recibidos y a los desafíos identificados durante las reuniones.
* Implementar una revisión continua del progreso y reajustar las prioridades y estrategias según sea necesario para cumplir con los plazos y objetivos del proyecto.

1. Entregables y Retroalimentación:

* Enfocarse en la entrega de componentes funcionales del sistema al final de cada sprint para permitir una evaluación y retroalimentación temprana.
* Utilizar la retroalimentación recibida para mejorar y refinar los entregables en los siguientes sprints.

### Actividades y Fases del proyecto

1. Fase de Inicio y Planificación:

* Análisis de Requisitos: Evaluación detallada del sistema actual e identificación de necesidades y áreas de mejora.
* Definición del Alcance: Establecimiento de metas y objetivos claros para el proyecto.
* Planificación de Sprints: Desarrollo de un cronograma para sprints basados en Scrum.

1. Fase de Diseño y Desarrollo:

* Diseño Técnico: Creación de la arquitectura del sistema mejorado y su integración con AWS Lambda y tecnologías de IA.
* Desarrollo de Prototipos: Construcción de versiones iniciales del sistema para pruebas y validación.
* Pruebas Unitarias y de Integración: Validación de cada componente y su integración efectiva.

1. Fase de Migración y Adaptación:

* Migración a AWS Lambda: Traslado del sistema a la infraestructura de AWS Lambda.
* Adaptación de Datos y Procesos: Asegurar que los datos y procesos existentes sean compatibles con el nuevo sistema.

1. Fase de Implementación y Cambio:

* Implementación Gradual: Despliegue progresivo del sistema en el entorno operativo.
* Gestión del Cambio: Capacitación y apoyo al personal para facilitar la transición al nuevo sistema.

1. Fase de Mejora y Optimización:

* Mejora Continua: Ajustes y mejoras basados en la retroalimentación y el rendimiento del sistema.
* Optimización de IA: Afinamiento de los modelos de IA para mejorar la precisión y eficacia.

1. Fase de Evaluación y Retroalimentación:

* Evaluación del Sistema: Análisis del rendimiento y la eficiencia del sistema mejorado.
* Recolección de Retroalimentación: Obtención de comentarios de usuarios y partes interesadas para futuras mejoras.

1. Fase de Cierre y Documentación:

* Documentación del Proyecto: Preparación de informes finales y documentación técnica.
* Revisión Final y Cierre: Evaluación del cumplimiento de los objetivos del proyecto y formalización del cierre

### Implementación basada en tareas

|  |
| --- |
| Revisión detallada del sistema actual. |
| Identificación de componentes y funcionalidades clave. |
| Recopilación de feedback de usuarios y personal. |
| Análisis de datos de rendimiento y eficiencia. |
| Documentación de deficiencias y áreas de mejora. |
| Configuración inicial y pruebas de AWS Lambda. |
| Migración de datos y funcionalidades al entorno de AWS Lambda. |
| Desarrollo de adaptaciones para la infraestructura sin servidor. |
| Validación de la funcionalidad y rendimiento en AWS Lambda. |
| Ajustes basados en pruebas y feedback. |
| Desarrollo de interfaz de usuario simple y efectiva. |
| Optimización del procesamiento y capacidad de respuesta de datos. |
| Pruebas de rendimiento y ajustes basados en resultados. |
| Implementación de mejoras técnicas y funcionales. |
| Evaluación final de rendimiento y eficiencia mejorados. |
| Diseño del módulo de reportes personalizados. |
| Desarrollo e integración del módulo con el sistema. |
| Pruebas de funcionalidad y precisión del módulo. |
| Ajustes y mejoras basados en pruebas y retroalimentación. |
| Documentación del módulo y capacitación para usuario |

Tabla 1. Tareas planeadas para implementación.

### Secuencia de desarrollo

Se evidencia un orden, el cual destaca la prioridad de las tareas, no por complejidad, sino por relevancia en el proyecto, de modo que las tareas de investigación, de búsqueda, entendimiento, etc, fueron abarcadas inicialmente, para reducir lo máximo posible el impacto de la curva de aprendizaje en cada apartado, después se abarca el área principal, el cual es el desarrollo en la nube, siendo este un punto vital en este proyecto, seguidamente se trabaja el módulo de los reportes con lo construido anteriormente, y por último se trabaja el modelo, esto debido a que el sistema funciona sin el modelo, pero no al revés, siendo este una mejora más que una necesidad funcional, y el desarrollo del mismo depende de la calidad del sistema.

## Monitoreo y Evaluación

El monitoreo y evaluación se hará principalmente sobre dos aspectos vitales, inicialmente el funcionamiento correcto (sin errores) del sistema, además, la velocidad de respuesta de los módulos ya sea unitariamente o en integración.

## Documentación

La documentación es parte relevante del sistema como tal, contemplado en las tareas, ya que la finalidad del sistema es su utilización, de modo que el sistema será documentado de forma constante, errores, aciertos, logros, avances, todo aquello relevante al mismo será documentado, ya sea para su mantenimiento, y principalmente, para su utilización.

# Descripción del Trabajo Realizado

Seguidamente se detallan aspectos relevantes al trabajo realizado, específicamente abarcando detalles de la implementación y su proceso, los cuales fueron con brevedad mencionados anteriormente, así como también los resultados obtenidos y un análisis correspondiente a los mismos.

## Descripción del Proceso de Solución

El proceso de solución para el rediseño del sistema de procesamiento de Big Data para la gestión y asignación de citas comienza con un diagnóstico inicial. En esta etapa, se recopila información estructural, lógica y útil del sistema, incluyendo archivos de datos, formatos, módulos, herramientas utilizadas, vulnerabilidades, pruebas realizadas y retroalimentación obtenida. Esta información se utiliza para identificar áreas de mejora y puntos importantes a abordar en el proyecto, proporcionando un panorama claro de los objetivos.

Una vez se ha realizado el diagnóstico, se procede al diseño del sistema. En esta fase, se definen los módulos del sistema, que incluyen la Interfaz de Usuario, el Generador de Reportes, el Sistema en la Nube, la Unidad de Almacenamiento, el Modelo y el Conector. Cada uno de estos módulos cumple una función específica en el sistema, como la interacción con el usuario, la generación de reportes, el procesamiento lógico interno y la comunicación entre los diferentes componentes.

La selección de tecnologías es un paso importante en el proceso de solución. En este caso, se elige Amazon Web Services (AWS) como plataforma en la nube, con herramientas como AWS Lambda como unidad de procesamiento y AWS S3 como unidad de almacenamiento. Estas tecnologías se consideran ideales debido a sus ventajas y capacidades para la implementación del sistema. Para efectos de comunicación con el sistema, AWS API Gateway fue la opción a convenir para dicha finalidad.

La planificación del proyecto se realizó considerando la complejidad de los módulos, las necesidades del sistema, las herramientas a utilizar y el conocimiento actual. Se establece una secuencia de tareas para llevar a cabo el proyecto, siguiendo la Metodología Ágil (Scrum) para una planificación y ejecución flexible y adaptativa.

El desarrollo del proyecto dio inicio evaluando completamente el sistema inicial, hasta comprender a profundidad cada mínimo apartado teórico y práctico relevante del mismo. Con el fin de proceder a la implementación en la nube del sistema, teniendo en cuenta cada repercusión positiva o negativa de los cambios a realizar, así como también cumplir con un desarrollo más dinámico y ágil.

La implementación en la nube del sistema correspondió a un proceso sumamente demandante en términos de tiempo y conocimiento requerido, siendo este una barrera importante a superar para completar el objetivo de este proyecto. No obstante, gracias al estudio de tecnologías y del Sistema de Procesamiento de Big Data para la Gestión de Citas Médicas, esta labor no se convirtió en un obstáculo mayor, logrando migrar el sistema a la nube de forma eficiente y en el tiempo planeado.

Una vez la aplicación se encontró en un estado óptimo a nivel de procesamiento, el paso siguiente fue la migración de los datos, los cuales anteriormente se encontraban almacenados en una base de datos relacional, la cual era debidamente cargada mediante un único archivo contenedor de los datos de los pacientes. El nuevo enfoque facilitó este aspecto en términos económicos y de complejidad, ya que, debido a la naturaleza del sistema, una base de datos corresponde a una herramienta innecesariamente compleja, según los requerimientos del sistema, por lo cual se optó por almacenar los datos en un formato a convenir (en este caso JSON), en la herramienta S3 (También proveída por Amazon Web Services), para facilitar la extracción y manipulación de estos, y al mismo tiempo prescindiendo de un coste innecesario.

El siguiente paso, ya construido de forma estructurada y sólida el sistema, corresponde a las pruebas de comunicación. Estas se hicieron mediante un API Gateway (ya mencionado anteriormente) generado en Amazon Web Services, especializado para las necesidades del sistema, el cual inserta un dato en la aplicación, al mismo tiempo es procesado por la misma, y añadido a la unidad de almacenamiento en S3.

Logrado lo anterior, la migración del sistema de forma completa y funcional fue alcanzada. Lo cual dio paso a las mejoras propuestas para el sistema, las cuales corresponden a dos principales mejoras, la adición de un generador de reportes basado en los datos de los pacientes, y, por último, pero no menos importante, el cambio de modelo de cálculo de peligrosidad.

Al tener un tiempo limitado, se decide iniciar con el generador de reportes, ya que es una funcionalidad que no existía con anterioridad, en este apartado se definió qué tipo de reportes se desea generar que aporten valor al usuario, por lo cual, estos se limitarán a gráficos basados en la información contenida de los pacientes. La implementación de este apartado no corresponde a un gran reto, en comparación con las demás tareas planteadas, puesto que el procesamiento lógico necesario para el mismo no representa una complejidad elevada para el conocimiento adquirido a día de hoy, destacando que la interfaz de usuario no es un aspecto a tener en cuenta en cuanto a detalle se refiere, ya que este apartado se centra en la generación de reportes, que puedan proveer de forma eficaz mediante gráficos información valiosa para los usuarios. Para este caso se decidió utilizar Streamlit para lograr un producto funcional de forma veloz en Python, y al mismo tiempo generar gráficos con elegancia y buena definición, por supuesto, la extracción de los datos para le generación de dichos reportes es debidamente realizada mediante consultas al sistema en la nube.

Por último y con abundante tiempo en disposición, se procedió a la implementación del generador del modelo de cálculo utilizando XGBoost, una biblioteca que facilita la implementación de algoritmos de aprendizaje automático en Python, y al mismo tiempo le provee al desarrollador la una amplia gama de posibilidades no disponibles con el modelo implementado en el sistema anterior. Esto presentó un reto importante, principalmente en términos de complejidad computacional y conocimiento requerido, más no en términos de tiempo, debido a que se conocen bien los datos a utilizar y los posibles resultados esperados, lo cual provocó que el tiempo de implementación radique principalmente en el algoritmo a desarrollar, y las debidas pruebas de este.

## Análisis de Resultados Obtenidos

El sistema mostró completitud y eficiencia una vez finalizado, sus componentes muestran un acoplamiento entre módulos sólido. Su comparación con el sistema anterior muestra resultados muy satisfactorios, los cuales serán presentados a continuación.

### Interfaz Gráfica

La interfaz gráfica corresponde al módulo encargado de brindarle al usuario acceso a las funcionalidades implementadas, módulo que del todo no existía en el sistema implementado anteriormente, entre las funcionalidades brindadas se encuentran las expuestas en la Figura 3, las cuales son: Sección para añadir nuevo paciente, sección de utilidades de usuario y sección de generación de reportes, respectivamente, cuyos componentes serán mostrados a continuación:

Interfaz de usuario gráfica

Descripción generada automáticamente

Figura 3. Inserción de Pacientes en Interfaz

#### Sección para añadir nuevo paciente

Visualizar el módulo de adición de nuevos pacientes, incluyendo el cálculo respectivo de su peligrosidad (Ver Figura 4).

Interfaz de usuario gráfica, Tabla

Descripción generada automáticamente

Figura 4. Inserción de Pacientes en Interfaz

#### Sección de utilidades de usuario

Regenerar el modelo en caso de que se considere que el mismo está desactualizado o se vea algún beneficio de que el mismo sea calculado nuevamente utilizando los datos más recientes, este proceso es delicado, puesto que interactúa con el módulo de cálculo, por lo cual debe garantizarse que se está de acuerdo con la acción a realizar para poder continuar (Ver Figura 5).

Interfaz de usuario gráfica, Texto, Aplicación, Chat o mensaje de texto

Descripción generada automáticamente

Figura 5. Regenerar el modelo de cálculo.

Restablecer los datos de los pacientes. En caso de que se esté utilizando la herramienta para pruebas, se permite al usuario restablecer los datos de estos al conjunto de datos a su estado original. El mismo solicita permisos y garantía de que se está de acuerdo debido a la delicadez del proceso, esta funcionalidad es adicional y debe utilizarse con suma precaución (Ver Figura 6).

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Figura 6. Restablecer lista de pacientes.

#### Sección de generación de reportes

Módulo de generación de reportes. Este módulo permite acceder a los reportes generados por la aplicación, los cuales muestran un gráfico de la cantidad de pacientes según 3 filtros, los cuales son, padecimiento (Ver Figura 7), edad (Ver Figura 8) y género (Ver Figura 9).

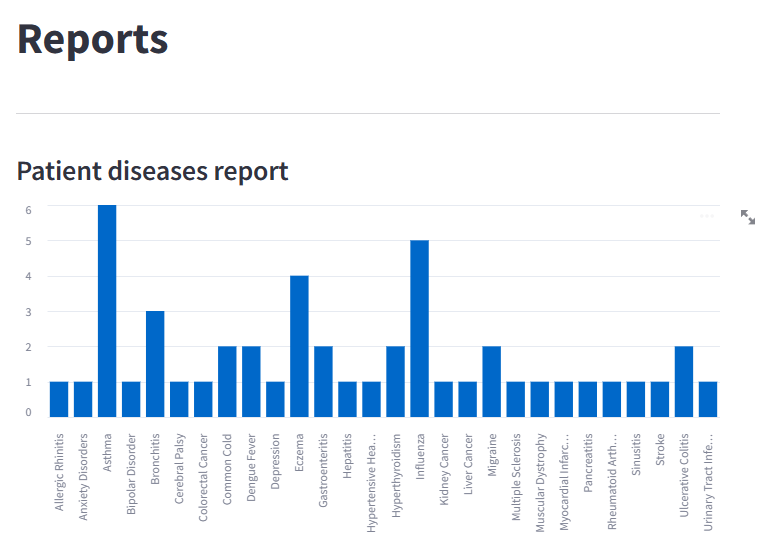


Figura 7. Gráfico pacientes vs padecimiento.

Gráfico, Gráfico de barras, Histograma

Descripción generada automáticamente

Figura 8. Gráfico pacientes vs edad.

Gráfico, Gráfico de barras

Descripción generada automáticamente

Figura 9. Gráfico pacientes vs género.

### Modelo de cálculo de peligrosidad

El cambio del algoritmo de cálculo fue una mejora distinguida en términos de tiempo y calidad de los resultados (Ver figura X), mientras que anteriormente el proceso de cálculo era una caja negra, que no permitía asignar pesos a los datos, el proceso de cálculo actual es personalizable, mantenible y escalable.

## Conclusiones y Recomendaciones

# Apéndices y Anexos

# Referencias

1. Alberto Segovia C. 2019. Sistema de Gestión de Citas Médicas para un Centro de Salud. Repositorio Documental de la Universidad de Valladolid. <https://uvadoc.uva.es/bitstream/handle/10324/36492/TFG-B.1225.pdf?sequence=1>
2. Araya, F. 2023. Edus: Cómo agendar una cita médica desde la aplicación. Telediario Costa Rica. <https://www.telediario.cr/nacional/edus-como-agendar-una-cita-medica-desde-la-aplicacion#:~:text=En%20la%20siguiente%20pantalla%2C%20se%20mostrar%C3%A1%20tu%20nombre,la%20especialidad%20y%20la%20fecha%20que%20te%20convenga>.
3. Apser, E. R. 2018. Qué Es Amazon s3 y cuáles son Sus Principales Beneficios. <https://apser.es/que-es-amazon-s3-y-cuales-son-sus-principales-beneficios/>
4. AWS-API GATEWAY. s.f. API management - amazon API gateway - AWS. (n.d.-a). <https://aws.amazon.com/api-gateway/>
5. AWS Lambda Guide. s.f. "Beneficios de Lambda - Información general acerca de AWS Lambda" - <https://docs.aws.amazon.com/lambda/latest/dg/welcome.html>
6. AWS-S3. s.f. Cloud object storage - amazon S3 - AWS. <https://aws.amazon.com/s3/>
7. Bacancy Technology. 2024. Netflix AWS Migration. Bacancy Technology. Recuperado de <https://www.bacancytechnology.com/blog/netflix-aws-migration#:~:text=Since%20its%20adoption%2C%20AWS%20Lambda,backup%20decisions%2C%20and%20security%20checks>.
8. Batko, K., & Ślęzak, A. 2022. The use of Big Data Analytics in Healthcare. Journal of big data. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8733917/>
9. Correa, A. 2024. Qué es amazon lambda y qué ventajas ofrece a TU Empresa. Codster. <https://codster.io/blog/ventajas-empresas-aws-lambda/>
10. Delfinocr. s.f. App de la CCSS Ahora le muestra cuánto le falta para pensionarse y el monto estimado. <https://delfino.cr/2019/11/aplicacion-de-la-ccss-permite-a-afiliados-al-ivm-visualizar-cuanto-les-falta-para-pensionarse>
11. Ferrovial. 2022. Qué es y para qué sirve la inteligencia artificial (IA). https://www.ferrovial.com/es/recursos/inteligencia-artificial/
12. Fernandez, Y. 2021. Big data: Qué es y para que sirve. Xataka. <https://www.xataka.com/basics/big-data-que-sirve>
13. IBM. s.f. ¿Qué es la inteligencia artificial en el Cuidado de la Salud?. <https://www.ibm.com/mx-es/topics/artificial-intelligence-healthcare>
14. Kirvan, P., Barney, N., & Gillis, A. S. 2024. What is AWS (Amazon Web Services)?: Definition from TechTarget. SearchAWS. <https://www.techtarget.com/searchaws/definition/Amazon-Web-Services>
15. Martins, J. 2024. Scrum: Conceptos Clave y cómo se aplica en la gestión de proyectos [2024] • asana. Asana. <https://asana.com/es/resources/what-is-scrum>
16. Organización Mundial de la Salud. 2021. Resumen - Ética y Gobernanza de la Inteligencia Artificial en el Ámbito de la Salud. [9789240037441-spa.pdf (who.int)](https://iris.who.int/bitstream/handle/10665/350263/9789240037441-spa.pdf?sequence=1)
17. Organización Panamericana de la Salud. 2023. Seguridad de la Información. [OPSEIHIS230016\_spa.pdf (paho.org)](https://iris.paho.org/bitstream/handle/10665.2/57372/OPSEIHIS230016_spa.pdf?sequence=1)
18. Rootstack. 2022. ¿Por qué aws, cuáles son Sus Ventajas? https://rootstack.com/es/blog/por-que-aws-cuales-son-sus-ventajas
19. Shilkov, M. 2019. Aws Lambda vs. Azure functions: 10 major differences. IOD - The Content Engineers. <https://iamondemand.com/blog/aws-lambda-vs-azure-functions-ten-major-differences/>