

Instituto Tecnológico de Costa Rica

Área Académica 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

Capítulo 1

Autor: José Agustín Venegas Vega

Coordinador del Proceso: Jennier Sonalo Cordero

Profesor Asesor: Marco Hernández Vázquez

## **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 del Área Académica 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
Dedicatoria .....	3
Agradecimientos .....	3
Resumen .....	3
Abstract .....	3
Siglas y Acrónimos .....	7
Introducción .....	7
Antecedentes del Proyecto .....	8
Descripción de la Institución .....	8
Área de Conocimiento que se desarrollará en el Proyecto .....	8
Trabajos Similares Encontrados .....	8
Planteamiento del Problema .....	9
Contexto del Problema .....	9
Justificación del Problema .....	9
Enunciado del Problema .....	10
Objetivos del Proyecto .....	10
Objetivo General .....	10
Objetivos Específicos .....	10
Alcances, Entregables y Limitaciones del Proyecto .....	11
Alcances .....	11
Entregables .....	11
Limitaciones .....	11
Marco de Referencia Teórico .....	12
Marco Metodológico .....	13
Descripción del Trabajo Realizado .....	14
Descripción del Proceso de Solución .....	14
Análisis de Resultados Obtenidos .....	14
Conclusiones y Recomendaciones .....	15
Apéndices y Anexos .....	17
Referencias .....	18

## Índice de Tablas

## Índice de Figuras

## 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

El Área Académica 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.

Hasta la fecha, se han alcanzado hitos significativos, incluyendo el desarrollo de una robusta base de datos y la implementación de herramientas de procesamiento de datos como NumPy, Pandas y SciPy. Además, se ha considerado la integración de PySpark para un procesamiento distribuido de datos de gran envergadura.

En la siguiente fase del proyecto, se contempla la migración del sistema a AWS Lambda, lo que permitirá una mayor escalabilidad, optimización de recursos y una mejora en la eficiencia operativa del sistema. Esta transición representa un paso crucial hacia la modernización y adaptabilidad del sistema en un entorno de computación en la nube.

Este documento detalla el proyecto en su estado actual, describiendo el contexto y los antecedentes, las tecnologías empleadas, y las metodologías previstas para su evolución y mejora. También se abordan los planes para la implementación de AWS Lambda, destacando cómo esta migración potenciará el rendimiento y la escalabilidad del sistema, alineándose con los objetivos a largo plazo del proyecto, asimismo su contraste y mejora mediante la utilización de inteligencia artificial.

## 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 el Área Académica 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, abarcará 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 centrará 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. Todo ello 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 original, centrado también en la gestión de citas médicas, logró con éxito la implementación de un sistema funcional que mostró resultados esperanzadores y potencialmente funcionales a gran escala. Sin embargo, a pesar de su éxito, el proyecto demostró tener varias áreas susceptibles de mejora.

Además, se encontró un proyecto con la misma finalidad, desarrollado en la Universidad de Valladolid, el cual se enfoca en el desarrollo de un sistema Big Data para mejorar la gestión y explotación de datos en un centro de salud, utilizando tecnologías como Apache Hadoop y Apache



Spark. El objetivo es optimizar el procesamiento de grandes volúmenes de datos para mejorar la toma de decisiones y la calidad del servicio en el ámbito sanitario. [1]

Con respecto a la utilización de AWS Lambda, se buscó un antecedente de gran peso, que incentive el uso de estas tecnologías emergentes y demuestre de forma sencilla su posible impacto. Desde su adopción, AWS Lambda se ha convertido en el arma secreta de Netflix 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. [2]

## 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. 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 alta 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. [3]

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. [4]

### *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. Los objetivos están cuidadosamente formulados para abordar cómo se puede lograr una gestión más eficiente y adaptativa de citas médicas, respondiendo a por qué son necesarias estas mejoras y para qué beneficiarán a las instituciones de salud.

### *Objetivo General*

Implementar en Lambda AWS el sistema de gestión de citas médicas existente, 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 adaptabilidad, mediante pruebas exhaustivas del sistema.

2. Adaptar el sistema de gestión de citas médicas, utilizando una infraestructura sin servidor para lograr escalabilidad automática 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 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 Exitosa en AWS Lambda: Completar con éxito 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 eficaz de las citas.
3. Reportes Detallados: Desarrollar un módulo de generación de reportes que permita a los usuarios obtener información detallada, facilitando un análisis más profundo de los datos de citas y la efectividad del sistema.
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: El sistema depende en gran medida de una conexión a internet estable, dado que AWS Lambda es una solución basada en la nube.
2. Limitaciones de Presupuesto: Restricciones presupuestarias podrían limitar la escala de implementación o las características del sistema.
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 personalización y la integración con sistemas existentes en las instituciones de salud.
5. Variabilidad en el Rendimiento de IA: Los algoritmos de inteligencia artificial pueden tener variaciones en su rendimiento, especialmente en las etapas iniciales después de su implementación.

## **Marco de Referencia Teórico**

## **Marco Metodológico**

## **Descripción del Trabajo Realizado**

**Descripción del Proceso de Solución**

**Análisis de Resultados Obtenidos**

## **Conclusiones y Recomendaciones**





## **Apéndices y Anexos**

## Referencias

- [1] Alberto Segovia C. 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] Bacancy Technology. (s.f.). 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.>
- [3] "Beneficios de Lambda - Información general acerca de AWS Lambda" - <https://docs.aws.amazon.com/lambda/latest/dg/welcome.html>
- [4] "7 Benefits of AWS Lambda for Cloud Computing" - Check Point Software - <https://www.checkpoint.com/cyber-hub/cloud-security/what-is-aws-lambda-and-how-does-it-work/>