

# Evaluación de SuperMicro SMExa con Enfoque en la Revolucionaria Tecnología AMD ROCm

Jorge Andrey Garcia

*Escuela de ingeniería en sistemas e informática*  
*Universidad Industrial del Santander*  
Bucaramanga, Colombia  
jorge2180115@correo.uis.edu.co

Jhoan Sebastian Garcia Reyes

*Escuela de ingeniería en sistemas e informática*  
*Universidad Industrial del Santander*  
Bucaramanga, Colombia  
jhoan2202045@correo.uis.edu.co

Carlos Esteban García Ortiz

*Escuela de ingeniería en sistemas e informática*  
*Universidad Industrial del Santander*  
Bucaramanga, Colombia  
carlos2210074@correo.uis.edu.co

Luis Andres Gonzalez Corzo

*Escuela de ingeniería en sistemas e informática*  
*Universidad Industrial del Santander*  
Bucaramanga, Colombia  
luis2201493@correo.uis.edu.co

Juan Nicolas Garcia Vega

*Escuela de ingeniería en sistemas e informática*  
*Universidad Industrial del Santander*  
Bucaramanga, Colombia  
juan2210050@correo.uis.edu.co

## Resumen

El objetivo de este artículo es presentar una propuesta para un servidor de alto rendimiento que utiliza componentes AMD junto a una placa AS-2025HS-TNR de SuperMicro para aplicaciones de cómputo de alto desempeño, análisis de datos e inteligencia artificial. Se describen los requisitos funcionales clave del servidor y se explican las ventajas de la solución propuesta en términos de diversificación tecnológica, integración de GPU, versatilidad y escalabilidad. Se analizan aspectos de implementación como eficiencia energética, refrigeración, costos y tolerancia a fallos. Se evalúan métricas de rendimiento, impacto ambiental y comparación técnica y de costos frente a alternativas del mercado. Se concluye resaltando el potencial de innovación de la propuesta y la importancia de superar obstáculos técnicos y adaptarse efectivamente a las nuevas tecnologías.

## Palabras Clave

Servidor, AMD, Rendimiento, Arquitectura, Inteligencia Artificial, Eficiencia, Tecnología, HPC, ROCm, Redes, Componentes, Soporte técnico, estándares, calidad de servicio.

## Abstract

The main objective of this paper is to show an proposal for a high performance server that uses AMD specs with a AS-2025HS-TNR plate of SuperMicro to high performance computing applications, Data analysis and Artificial intelligent. We describe the principal functional requirements of the server and some explications of the advantages of this solution in terms of the technological diversification, GPU integration, versatility and scalability. We analyze the implementation aspects as energy efficiency, the cooler system, expenses and fail tolerance. We evaluate performance metrics, environment impact, technique comparison and the costs compared to market alternatives. We conclude highlighting the innovation potential of the proposal and the importance of grow off the technical issues and adapt efficiently to the new technologies.

## Index Terms

Server, AMD, Performance, Architecture, Artificial Intelligence, Efficiency, Technology, HPC (High-Performance Computing), ROCm, Networks, Components, Technical Support, Standards, Service Quality

## I. INTRODUCCIÓN

En vista del avance tecnológico que se ha presentado en los últimos años, el grupo de investigación **SC3UIS** de la Universidad Industrial de Santander se ha visto en la necesidad de requerir a nuevas tecnologías para plantear un cambio a futuro de sus servidores de super cómputo, para ello se ha decidido llevar a cabo el diseño e implementación de un servidor más novedoso y con un enfoque de mercado diferente en el cual poder realizar cálculos científicos, simulaciones, machine learning, entre otras tareas que requieran un alto poder de cómputo. Este proyecto será llevado a cabo de la mano de la empresa SuperMicro junto con el grupo SC3UIS con el nombre de SMExa.

Este reto será abordado desde una perspectiva diferente a la tradicional forma de ensamblar servidores, en este caso, se hará uso de la mas alta tecnología proporcionada en la actualidad por el distribuidor de componentes **AMD**, desde sus CPUs CPUs AMD EPYC Genoa hasta la GPU AMD Instinct MI210.

La propuesta se ve enfrentada a un desafío particular: la novedad y desconocimiento de estas tecnologías de vanguardia, desde la falta de experiencia por parte del equipo técnico hasta las pocas pruebas de rendimiento y/o compatibilidad con software especializado en el área de investigación y super cómputo. Ante este desafío, el grupo SC3UIS busca un análisis profundo de diversos factores que pueden afectar la implementación y el uso de este nuevo equipo, para ello, se buscó un equipo con conocimiento en el tema de la arquitectura de computadores, comprendiendo tanto software como hardware y se les pidió

realizar el análisis de especificaciones técnicas, medidas de rendimiento, precios de mercado, soporte técnico, limitaciones, entre otras características esenciales que se requiere conocer en profundidad las posibilidades y limitaciones del proyecto SMExa.

## II. DESARROLLO

### 1) **Requerimientos Funcionales para un Servidor con AMD EPYC™ 9554 Series y AMD Instinct MI210:**

- a) **Potencia de Cómputo:** El servidor debe aprovechar los 64 núcleos a 128 hilos de la CPU para ejecutar simultáneamente aplicaciones intensivas en cómputo como simulaciones y modelos de inteligencia artificial (IA).
- b) **Integración de GPU para cómputo de alto Rendimiento:** La plataforma debe ser capaz de aprovechar la arquitectura CDNA2 de la GPU AMD Instinct MI210 para tareas específicas de cómputo de alto rendimiento (HPC) y cargas de trabajo de inteligencia artificial.
- c) **Rendimiento de memoria y ancho de banda:** El servidor debe admitir y optimizar el rendimiento de la memoria DDR5 y HBM2e, proporcionando un ancho de banda suficiente para cargar y manipular grandes conjuntos de datos.
- d) **Almacenamiento de gran capacidad y velocidad:** Se requiere un sistema de almacenamiento con capacidad de 300 TB y un rendimiento de lectura/escritura rápido para manejar grandes cantidades de datos utilizados en aplicaciones de análisis y simulaciones.
- e) **Red de alta velocidad y baja latencia:** El Servidor debe integrar una red Infiniband de 100Gb para facilitar la transferencia rápida de datos entre la estación de trabajo y otros dispositivos de la red, crucial para aplicaciones HPC y de análisis de datos.
- f) **Soporte para Frameworks de IA y bibliotecas HPC:** La plataforma debe ser compatible con frameworks populares de inteligencia artificial como TensorFlow y PyTorch, así como bibliotecas HPC para facilitar el desarrollo y ejecución eficiente de aplicaciones.
- g) **Optimización de cargas de trabajo paralelas:** El servidor debe ser capaz de ejecutar eficientemente tareas en paralelo, optimizando la utilización de los múltiples núcleos y unidades de cómputo disponibles.
- h) **Seguridad y confidencialidad de datos:** Se requieren medidas de seguridad robustas para proteger datos sensibles, cumpliendo con estándares de seguridad de la industria, especialmente en aplicaciones relacionadas con inteligencia artificial.
- i) **Flexibilidad de configuración:** La plataforma debe admitir configuraciones flexibles, incluyendo opciones de configuración de CPU y GPU según las necesidades específicas de la aplicación.
- j) **Escalabilidad y adaptabilidad:** El servidor debe ser escalable para permitir futuras expansiones y actualizaciones de hardware sin comprometer el rendimiento general.

Estos requerimientos funcionales proporcionan la base para diseñar y desarrollar un servidor optimizado para aplicaciones de HPC, simulaciones, análisis de datos e inteligencia artificial, aprovechando al máximo las capacidades de la CPU AMD EPYC™ 9004 Series y la GPU AMD Instinct MI210.

### 2) **Originalidad y novedad de la solución propuesta:**

La propuesta de integrar la plataforma SMExa, basada en servidores SuperMicro con procesadores AMD EPYC Genoa y GPU AMD Instinct, junto con la programación utilizando ROCm, presenta varias características originales y novedosas:

- **Diversificación Tecnológica:** La introducción de hardware basado en tecnología AMD en lugar de la tradicional combinación de Intel CPUs y NVIDIA GPUs implica una diversificación significativa. Esto puede proporcionar una perspectiva fresca y nuevas oportunidades para optimizar el rendimiento en cargas de trabajo.
- **Enfoque Integrado:** Al utilizar la plataforma ROCm para programar tanto las CPUs como las GPUs AMD, se está adoptando un enfoque integrado que puede facilitar la programación y la optimización, permitiendo un mejor rendimiento y eficiencia en comparación con soluciones que requieren diferentes entornos de programación para CPUs y GPUs.
- **Aplicación en Diversos Dominios:** La capacidad de la plataforma para abordar tanto las aplicaciones de HPC tradicionales como las necesidades de analítica de datos e Inteligencia Artificial sugiere una versatilidad que puede ser crucial en entornos de investigación y desarrollo.

### 3) **Eficiencia energética e implementación:**

El servidor AS-2025HS-TNR trabaja con fuentes de poder desde 1200 hasta 2600W; sin embargo, debido a los componentes necesarios para este caso particular, la demanda energética del mismo es de 2600W, una potencia bastante elevada según los estándares actuales. A pesar de su alto consumo, el servidor no requiere de un enfriamiento muy potente gracias a la tecnología más novedosa del mercado (utilizada en este servidor), que puede funcionar correctamente a temperaturas de 25 a 30° centígrados, superando el promedio de los servidores del mercado de 21° a 25°C. Este cambio

permite reducir el consumo energético de la refrigeración hasta en un 4 %. Además, al ser un servidor de 2U de tamaño, es fácilmente ampliable y se refrigera de manera mas eficiente que otros servidores.

Estas optimizaciones en eficiencia energética, temperatura y tamaño del servidor hacen posible su refrigeración mediante un sistema CRAC (Computer Room Air Conditioners) sin necesidad aparente de cambiar a tecnologías más costosas, como la refrigeración por inmersión. para la implementación del sistema CRAC, se debe considerar la temperatura ideal a la que deberá permanecer el servidor, la redundancia en el sistema, la eficiencia energética y/o las normativas pertinentes. Finalmente, el sistema de refrigeración requiere de tareas especiales que debería realizar el personal técnico designado, como el control de humedad para evitar o controlar la carga estática en la sala, el monitoreo y gestión de alarmas en todo momento para prevenir cualquier daño al equipo y el mantenimiento tanto del servidor como del sistema de refrigeración.

#### 4) Tecnología Utilizada para el Servidor:

El servidor propuesto incorpora tecnologías avanzadas y potentes para abordar aplicaciones de alto rendimiento en áreas como HPC, simulaciones, análisis de datos e inteligencia artificial. A continuación, se detallan las tecnologías clave utilizadas:

##### a) Procesador AMD EPYC™ 9004 Series:

- Arquitectura Zen 4: El procesador utiliza la arquitectura Zen 4 de AMD, que proporciona un rendimiento excepcional en términos de cálculos de precisión, paralelismo y eficiencia energética.
- 64 Núcleos y 128 Hilos: Con 64 núcleos y 128 hilos, la CPU ofrece una capacidad de procesamiento masiva para abordar tareas intensivas en cómputo.

##### b) GPU AMD Instinct MI210:

- Arquitectura CDNA2: La GPU cuenta con la arquitectura CDNA2, optimizada para cargas de trabajo de cómputo de alto rendimiento y aplicaciones de inteligencia artificial.
- 6656 Procesadores de Transmisión: La alta cantidad de procesadores de transmisión permite un procesamiento paralelo eficiente en tareas de cálculo intensivo.

##### c) Memoria de Alto Rendimiento:

- DDR5: La plataforma utiliza memoria DDR5 para un acceso rápido a los datos, mejorando el rendimiento general del sistema.
- HBM2e: La GPU está equipada con memoria HBM2e, proporcionando un ancho de banda excepcional para manipular grandes conjuntos de datos de manera eficiente.

##### d) Almacenamiento Masivo y Rápido:

- 300 TB de Disco Duro: El servidor cuenta con un sistema de almacenamiento de 300 TB para manejar grandes volúmenes de datos, garantizando la disponibilidad de información para aplicaciones de análisis y simulaciones.

##### e) Red de Alta Velocidad:

- Infiniband 100Gb: La red Infiniband de 100Gb permite una transferencia rápida de datos entre el servidor y otros dispositivos en la red, esencial para aplicaciones HPC y análisis de datos.

##### f) Soporte para Últimos Estándares:

- PCIe 5.0: La plataforma incluye soporte para PCIe 5.0  $\times 128$ , asegurando una conectividad de alta velocidad con otros dispositivos periféricos.

##### g) Seguridad Integral:

- Medidas de Seguridad: Se implementan medidas de seguridad robustas para proteger datos sensibles y cumplir con estándares de seguridad de la industria.

##### h) Configuración y Escalabilidad:

- Flexibilidad: La plataforma ofrece opciones de configuración flexibles para adaptarse a diversas necesidades de aplicación.
- Escalabilidad: Diseñada para permitir futuras expansiones y actualizaciones de hardware, asegurando la adaptabilidad a cambios en los requisitos de rendimiento.

En conjunto, este servidor representa un entorno de computación avanzado y altamente especializado, aprovechando las tecnologías más recientes para impulsar aplicaciones exigentes en términos de rendimiento y capacidad de procesamiento.

#### 5) Costo de las tecnologías utilizadas:

- a) AMD EPYC Genoa 9554 64C/128T 3.1G 256M: Este procesador tiene un costo de venta de 5,295 dólares, según Amazon. Y como son 2 daría un total de 10,590 dólares



Figura 1. AMD EPYC Genoa

- b) GPU AMD Instinct MI210: Esta GPU en 2022 estaba listada en tiendas de Japón en 16,000 dólares, pero se pudo encontrar en una tienda virtual en 9,280 dólares

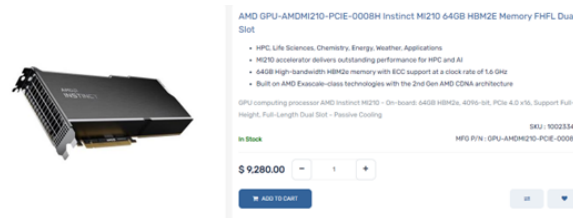


Figura 2. AMD Instinct MI210

- c) 384 GB de RAM: Para estas memorias RAM se encontró una marca reconocida (Corsair) que ofrece bastantes opciones en cuanto a tamaño de la memoria, usando 2 módulos de memoria:

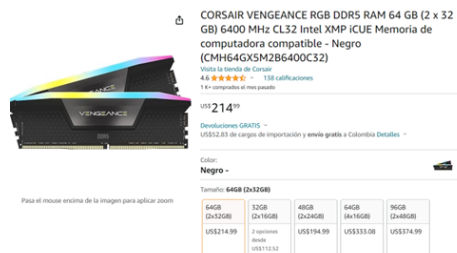


Figura 3. RAM Corsair de 64 GB

Para suplir el requerimiento de 384 GB de RAM podemos comprar:

- 4 opciones de 94 GB( $2 * 48GB$ )= 1,499.96 dólares.
- 6 opciones de 64 GB( $2 * 32GB$ )= 1,289.94 dólares.

Que son las mejores opciones si se quiere menos cantidad de módulos o un menor precio, respectivamente.

- d) 300 TB de disco duro: Para suplir este requerimiento se encontró este disco con 22TB el más grande que se pudo conseguir con un precio de 419.99 dólares, y se necesitan mínimo 14 (308GB) que costarían 5,879.86 dólares.



Figura 4. HDD Western Digital de 22TB

- e) Infiniband 100Gb: Para este requerimiento se encontró el siguiente adaptador con un precio de 2,286 dólares, con 2 puertos:

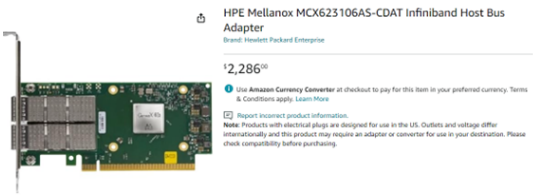


Figura 5. Inifiband Mellanox

Y este con un puerto, con un precio de 619 dólares:

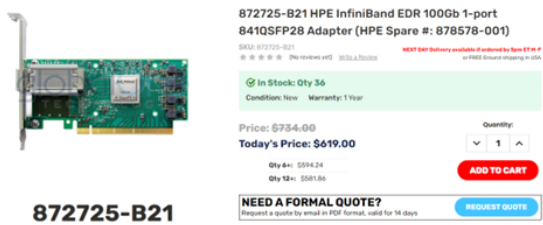


Figura 6. Infiniband EDR

Todo esto da un total de: 29,535.82 dólares usando la opción de RAM de 4 opciones de 94 GB y el adaptador de 2 puertos, ya que esto daría un precio máximo.

- 6) **Precio en el mercado:** El precio estaría dado únicamente por el costo de la máquina, ya que al ser una máquina de prueba no se tiene en cuenta precio de instalación, ya que este proceso está contenido en las actividades que normalmente se hacen en SC3.
- Precio en el mercado estimado:


Formal Quotation Request	
	
A+ Server 2025HS-TNR My System November 30th, 11:53 pm EST	
Basebone	Supermicro Hyper Server 2025HS-TNR - 2U - 12x NVMe/SATA/SAS - 2x M.2 - 1x AIOM - 1600W (1+1) Redundant
Processor	2x AMD EPYC™ 9554 Processor 64-core 3.10GHz 250MB Cache (360W)
Memory	12x 32GB PC5-38400 4800MHz DDR5 ECC RDIMM
Hard Drive	12x 22TB SATA 6.0Gb/s 7200RPM - 3.5" - Ultrastar™ DC HC370 (3.12e/HKs)
Network Adapter	NVIDIA® ConnectX®-6 100-Gigabit HDR100 InfiniBand Adapter - PCIe 4.0 x16 - 2x QSFP56
I/O Modules - Networking	Supermicro AIOM OCP 3.0 - 2x 10Gb RJ45 - Intel i350 - PCI-E 2.1 x4 - AOCAG-12M
Server Management	Supermicro Update Manager (SUM) (OOB Management Package), Included
Operating System	No Operating System
Warranty	3 Year Depot Warranty (Return for Repair)
Configured Price: \$30,138.00	
Quantity: 1 x \$30,138.00 = \$30,138.00	

Figura 7. Estimación de precio del servidor

Esta estimación se hizo con una página (www.thinkmate.com) en la que se hizo una construcción estimada del servidor, con ciertos cambios en sus componentes, buscando la mayor similitud al proyecto propuesto, con un precio de 30,138 dólares.

7) **Límites, detección y tolerancia a fallas, y problemas posibles de la solución propuesta:**

Los límites de la solución propuesta incluyen:

- Incompatibilidad de códigos: La migración de aplicaciones desarrolladas para arquitecturas Intel-NVIDIA a la nueva plataforma AMD puede presentar desafíos en términos de compatibilidad y optimización. Será crucial revisar y ajustar los códigos existentes para aprovechar al máximo el rendimiento de los nuevos componentes.

- **Curva de aprendizaje:** El equipo de desarrollo de SC3UIS puede enfrentar una curva de aprendizaje lenta al adaptarse a la programación con ROCm y a las peculiaridades de la arquitectura AMD. Se requerirá tiempo y recursos para familiarizarse y aprovechar todas las capacidades ofrecidas.
- **Disponibilidad de soporte técnico:** Dado que la comunidad de desarrollo está acostumbrada a trabajar con tecnologías Intel y NVIDIA, la disponibilidad de soporte técnico específico para la plataforma AMD puede ser un desafío. Se debe establecer una comunicación efectiva con los recursos de soporte de AMD para abordar cualquier problema técnico.

En cuanto a la detección y tolerancia a fallas, se recomienda implementar estrategias que incluyan:

- **Monitorización continua:** Establecer sistemas de monitorización en tiempo real para supervisar el rendimiento de la plataforma y detectar posibles anomalías.
- **Respaldos y redundancias:** Implementar estrategias de respaldo y redundancia para garantizar la continuidad del trabajo en caso de fallas en algún componente crítico.
- **Procedimientos de recuperación:** Desarrollar procedimientos claros y rápidos para la recuperación después de una falla, minimizando el tiempo de inactividad.

La plataforma ROCm tiene algunas limitaciones. Por ejemplo, la versión 5.7 de ROCm tiene un soporte limitado para configuraciones multi-GPU. Además, aunque la mayoría de las versiones de ROCm funcionan con kernels de Linux, se ha observado que ROCm v4.1 no funciona con kernels de Linux para dispositivos Vega20 de 7 nm.

En cuanto a la detección y tolerancia a fallas, la comunidad de desarrolladores de ROCm es activa y puede proporcionar asistencia en la resolución de problemas.

#### 8) Métricas propuestas para medir el rendimiento de acuerdo con los requerimientos:

- **Rendimiento de cómputo:** Esta métrica mide la cantidad de cálculos que la GPU puede realizar por segundo. Para la GPU AMD Instinct MI210, el rendimiento de cómputo de precisión media máxima (FP16) es de 181 TFLOPs.
- **Uso de la memoria:** Esta métrica mide la cantidad de memoria utilizada por la GPU durante la ejecución de un programa. La GPU AMD Instinct MI210 tiene 64 GB de memoria HBM2e.
- **Ancho de banda de la memoria:** Esta métrica mide la cantidad de datos que la GPU puede leer o escribir en su memoria por segundo. Para la GPU AMD Instinct MI210, el ancho de banda de memoria es de hasta 1638.4 GB/s.
- **Tiempo de ejecución:** Esta métrica mide el tiempo que tarda un programa en ejecutarse en la GPU.

Estas métricas son recopiladas y analizadas utilizando la herramienta ROCm Profiler, que proporciona una visión detallada del rendimiento de la GPU.

#### 9) Aspectos ambientales y éticos:

Los campos de la tecnología y dispositivos electrónicos han experimentado un crecimiento exponencial en los últimos años. Este aumento ha traído a juego un aspecto que hasta hace poco no se tenía en cuenta al hablar de tecnología: la contaminación. al hablar de contaminación por tecnología, se deben tener en cuenta varios factores, siendo los principales:

- **Impacto de uso:** En 2019, un estudio realizado por Greenpeace reveló que las tecnologías “Big data” y de inteligencia artificial representaron el 7 % del gasto eléctrico mundial. Sin embargo, esto por sí solo no es lo realmente preocupante, a la hora de analizar la contaminación por consumo eléctrico se debe analizar el costo ambiental de producir la cantidad de energía requerida. En este aspecto, la tecnología necesaria para el procesamiento de datos (generalmente servidores) generó un 2 % de las emisiones globales de  $CO_2$  en 2019. Para 2022, estas emisiones representaron el 4 % a nivel mundial, y se prevé que, si continúa el actual ritmo de crecimiento, en 2030 la cifra podría superar el 10 % de las emisiones de  $CO_2$  a nivel global.
- **Impacto del ciclo de vida:** El ciclo de vida de un dispositivo electrónico consta de manufactura, transporte al consumidor, uso y reciclaje. En dispositivos de uso doméstico, la manufactura es el factor más contaminante, responsable de hasta el 70 % del impacto ambiental (en emisiones de  $CO_2$ ) durante la vida útil. Seguido por el uso, que representa aproximadamente el 20 % de la contaminación para dispositivos normales, y, por último, el transporte constituye aproximadamente el 10 % de la contaminación del dispositivo (para servidores, los porcentajes se inclinan mucho más hacia la fase de uso). Actualmente, se está implementando una manera de reducir esta contaminación en el ciclo de vida: la fase de reciclaje. en esta fase, actualmente se logra disminuir cerca del 10 % de la contaminación producida en el resto del ciclo. Sin embargo, se está avanzando en el estudio e implementación de tecnologías más amigables con el medio ambiente. Estas tecnologías se centran en gran parte en la capacidad de reparación y reutilización de dispositivos, tecnología implementada parcialmente en el servidor estudiado en este artículo.

Dicho esto, la creación y posterior uso del servidor para cómputo científico de SC3UIS podría verse obstaculizado a futuro por regulaciones ambientales impuestas al uso de estas tecnologías, como leyes de consumo energético máximo o leyes de reciclaje de aparatos electrónicos.

#### 10) **Calidad técnica de la propuesta presentada:**

La calidad de los componentes y la estructura planteada está bastante justificada para los requerimientos. Sin embargo, vamos a ver, haciendo comparación a propuestas de otras compañías, a ver como se compara la propuesta con el mercado.

- El uso de AMD genova como propuesta, es bastante justificada, debido a los requerimientos y el uso que se le va a dar al servicio. Si miramos la competencia, en terminos de procesadores, podemos mirar las intel xeon. En este caso, podemos compararla con la Intel Xeon CPU Max 9480, con una caché de 112.5 M, 56 núcleos(112 hilos), una frecuencia base/máxima de 1.9/3.5 GHz. A simples rasgos, podemos notar la ventaja de amd sobre ryzen en el número de hilos (que pueden ayudar a trabajo paralelo), frecuencias más altas y un precio menor. También cabe destacar el mayor número de canales soportados, siendo 12 en amd versus los 8 del intel, además de un ancho de banda de memoria más alto.
- Para el caso de la GPU, podemos ver como alternativa a la nvidia A100. Sus características son similares, excepto acaso una potencia en bruto un poco menor. Según un artículo de la CNBC(??) el precio de las nvidia A100 rondan los \$10000 dólares, por lo que podemos ver que los precios son similares. Una de las ventajas de la tarjeta nvidia, es su amplia documentación y la experiencia previa de los administradores en ella. De todos modos, solo por las características, podemos decir que la elección de una u otra no debería cambiar el resultado una vez implementada la plataforma.

Cabe destacar que en materia de GPU, actualmente hay equipos con mayor capacidad y potencia, pero su consumo y precio también son mucho más altos. Además, para cargas donde se requiera un uso intensivo de GPU, se puede recurrir a la incipiente industria de máquinas para entrenamiento en la nube. Podemos ver servicios como nvidia DGX o microsoft azure, que permiten abaratar costes haciendo la capa de entrenamiento en servidores externos (que suele ser la parte más costosa al trabajar en sistemas de inteligencia artificial) y utilizar los modelos entrenados para trabajar en los servidores locales. Con todo aclarado podemos decir que la propuesta es excelente.

### III. CONCLUSIONES

- Este servidor es una solución integral que aprovecha las últimas tecnologías para impulsar la innovación y la eficiencia en una variedad de aplicaciones exigentes.
- El dispositivo tiene un alto consumo energético, pero debido a su disposición y tecnología innovadora puede ser refrigerado con tecnología de bajo costo.
- Se destaca la viabilidad de un sistema CRAC para la refrigeración, evitando la necesidad de tecnologías mas costosas y complejas de mantener.
- La estimación de precio en el mercado realizada en [www.thinkmate.com](http://www.thinkmate.com) para una configuración similar es de 30,138 dolares. Esto muestra que la estimación de costos del proyecto está en línea con los precios de mercado.
- Además de los costos de hardware, es fundamental tener en cuenta otros elementos del proyecto, como la integración con la infraestructura existente, la capacitación del personal en cuanto a las nuevas tecnologías y la escalabilidad a largo plazo.
- Aunque la propuesta presenta ventajas significativas, se deben abordar desafíos como la posible incompatibilidad de códigos y la curva de aprendizaje al adoptar ROCm y la arquitectura AMD.
- La detección y tolerancia a fallas son esenciales, destacando la importancia de estrategias como la monitorización continua y los respaldos. Además, se requiere un enfoque proactivo para establecer un soporte técnico efectivo y superar posibles obstáculos.
- Las métricas propuestas para medir el rendimiento, analizadas con la herramienta ROCm Profiler, ofrecen una visión detallada que puede guiar la optimización continua de la plataforma.
- En conjunto, la propuesta muestra promesas de innovación y eficiencia, pero su éxito dependerá de la superación hábil de desafíos y la adaptación efectiva a las nuevas tecnologías.
- El alto impacto del uso de tecnologías Big data e IA puede ocasionar que el dispositivo se vea sujeto a regulaciones ambientales a futuro que interfieran con el uso normal del mismo.
- El hardware utilizado está justificado para el uso que se le quiere dar, pero vale la pena ir mirando soluciones en la nube como nvidia GDX o Azure AI plaraform, que permiten dejar la parte más costosa de manejo de inteligencia artificial, el entrenamiento, a servicios externos, evitando el uso desmesurado de dispositivos como tarjetas gráficas, que se sabe generan una gran parte del uso de energía y temperaturas, que se puede utilizar para otros procesos.

## REFERENCIAS

- [1] Supermicro. (2023) Hyper a+ server as -2025hs-tnr. [Online]. Available: <https://www.supermicro.com/en/products/system/hyper/2u/as-2025hs-tnr>
- [2] Plug and Track. (2023) Sistema de alarmas de temperatura para salas de servidores. [Online]. Available: <https://www.plugandtrack.com/es/aplicaciones-casos-de-uso/alarma-de-temperatura-para-salas-de-servidores/>
- [3] CaloryFrio. (2023) Refrigeración en salas de servidores: centros de datos fríos y seguros. [Online]. Available: <https://www.caloryfrio.com/refrigeracion-frio/refrigeracion-salas-servidores-centros-datos-frios-seguros.html>
- [4] N. D. Seguros. (2023) Contaminación digital: la huella ecológica del big data. [Online]. Available: <https://nuestrosdatosseguros.es/contaminacion-digital-la-huella-ecologica-del-big-data>
- [5] AMD. (2023) Ecosistema abierto amd rocm™ — amd. [Online]. Available: <https://www.amd.com/es/graphics/servers-solutions-rocm>
- [6] A. Community. (2023) Rocm - amd community. [Online]. Available: <https://community.amd.com/t5/rocm/ct-p/amd-rocm>
- [7] R. Documentation. (2023) Mi200 performance counters and metrics — rocm 5.7.1 documentation home. [Online]. Available: [https://rocmdocs.amd.com/en/latest/understand/gpu\\_arch/mi200\\_performance\\_counters.html](https://rocmdocs.amd.com/en/latest/understand/gpu_arch/mi200_performance_counters.html)
- [8] —. (2023) Limitations — use rocm on radeon gpus. [Online]. Available: <https://rocm.docs.amd.com/projects/radeon/en/latest/docs/limitations.html>
- [9] TechPowerUp. (2023) Amd radeon instinct mi210 specs — techpowerup gpu database. [Online]. Available: <https://www.techpowerup.com/gpu-specs/radeon-instinct-mi210.c3857>
- [10] AMD. (2023) Amd infinity architecture technology — amd. [Online]. Available: <https://www.amd.com/es/technologies/infinity-architecture>
- [11] —. (2023) Acelerador amd instinct™ mi210 — amd. [Online]. Available: <https://www.amd.com/es/products/server-accelerators/amd-instinct-mi210>
- [12] K. Leswing. (2023) Nvidia's a100 is the \$10,000 chip powering the race for a.i. [Online]. Available: <https://www.cnbc.com/2023/02/23/nvidias-a100-is-the-10000-chip-powering-the-race-for-ai.html>
- [13] M. Azure. (2023) Azure ai platform—artificial intelligence — microsoft azure. [Online]. Available: <https://azure.microsoft.com/en-us/solutions/ai/>
- [14] NVIDIA. (2023) Dgx platform — nvidia. [Online]. Available: <https://www.nvidia.com/en-us/data-center/dgx-platform/>
- [15] —. (2023) Nvidia a100 — nvidia. [Online]. Available: <https://www.nvidia.com/en-us/data-center/a100/>
- [16] AMD. (2023) Amd epyc™ 9554 processors — amd. [Online]. Available: <https://www.amd.com/en/products/cpu/amd-epyc-9554>
- [17] Intel. (2023) Intel xeon cpu max 9480 processor 112.5m cache 1.90 ghz product specifications. [Online]. Available: <https://www.intel.com/content/www/us/en/products/sku/232592/intel-xeon-cpu-max-9480-processor-112-5m-cache-1-90-ghz/specifications.html>
- [18] Amazon. (2023) Amd epyc 9554 processor 64 core 3.1ghz 256mb l3 cache tdp 360w sp5 socket (4th gen, genoa) (100-000000790) (oem tray processor). [Online]. Available: [https://www.amazon.com/AMD-Epyc-9554-Processor-100-000000790/dp/B0BVSF7GYD?language=en\\_US](https://www.amazon.com/AMD-Epyc-9554-Processor-100-000000790/dp/B0BVSF7GYD?language=en_US)
- [19] —. (2023) Corsair vengeance rgb ddr5 ram 64gb (2x32gb) 6400mhz cl32 intel xmp icue compatible computer memory - black (cmh64gx5m2b6400c32). [Online]. Available: [https://www.amazon.com/CORSAIR-VENGEANCE-6400MHZ-Compatible-Computer/dp/B0CCXT8FX2/ref=mp\\_s\\_a\\_1\\_7?keywords=ram%2Bddr5&qid=1701448480&sr=8-7&th=1](https://www.amazon.com/CORSAIR-VENGEANCE-6400MHZ-Compatible-Computer/dp/B0CCXT8FX2/ref=mp_s_a_1_7?keywords=ram%2Bddr5&qid=1701448480&sr=8-7&th=1)
- [20] —. (2023) Western digital 22tb wd red pro nas internal hard drive hdd - 7200 rpm, sata 6 gb/s, cmr, 512 mb cache, 3.5" - wd221kfgx. [Online]. Available: [https://www.amazon.com/Western-Digital-22TB-Internal-Drive/dp/B0B5W1CQ8W/ref=sr\\_1\\_2?crid=214176235KT9Q&keywords=50%2Btb%2Bhdd&qid=1701408004&sprefix=50%2Btb%2Bhdd%2Caps%2C131&sr=8-2&th=1](https://www.amazon.com/Western-Digital-22TB-Internal-Drive/dp/B0B5W1CQ8W/ref=sr_1_2?crid=214176235KT9Q&keywords=50%2Btb%2Bhdd&qid=1701408004&sprefix=50%2Btb%2Bhdd%2Caps%2C131&sr=8-2&th=1)
- [21] —. (2023) Hpe mellanox mcx623106as-cdat infiniband host bus adapter. [Online]. Available: [https://www.amazon.com/Hewlett-Packard-Enterprise-MCX623106AS-CDAT-Infiniband/dp/B0966T9MBH/ref=sr\\_1\\_9?crid=2IHPBR5HTL1XO&keywords=infiniband+adapter&qid=1701408401&sprefix=infiniband+ada%2Caps%2C139&sr=8-9&language=en\\_US](https://www.amazon.com/Hewlett-Packard-Enterprise-MCX623106AS-CDAT-Infiniband/dp/B0966T9MBH/ref=sr_1_9?crid=2IHPBR5HTL1XO&keywords=infiniband+adapter&qid=1701408401&sprefix=infiniband+ada%2Caps%2C139&sr=8-9&language=en_US)
- [22] G. O. Technology. (2023) Hpe infiniband edr 100gb 1-port 841qsfp28 adapter. [Online]. Available: <https://www.globalonetechnology.com/872725-b21.htm>
- [23] Thinkmate. (2023) Thinkmate - server, workstation oem solutions. [Online]. Available: <https://www.thinkmate.com/quotation-request?a=YToxOntzOjI6ImlkIjtpOjY4Njg5MTt9>
- [24] —. (2023) Thinkmate - server, workstation oem solutions. [Online]. Available: <https://www.thinkmate.com/>