

incentro

Google Cloud

10 años de GKE

Cómo una década de evolución está
moldeando el futuro de la tecnología.



Gracias por tu interés en este eBook

Como partner de Google Cloud, en Incentro trabajamos cada día con equipos que quieren modernizar su plataforma o llevar IA a producción sin perder control de costes. Este eBook te ayudará a:

- Entender cuándo GKE encaja y qué aporta frente a otras alternativas.
- Elegir el modelo operativo (Autopilot o Standard) según tu contexto.
- Aterrizar buenas prácticas de seguridad y gobierno desde el día uno.
- Estimar el impacto en costes y en eficiencia operativa.
- Planificar los primeros pasos: migración, modernización o nuevos casos de IA.

Si tienes preguntas o quieres comentar tus necesidades específicas, no dudes en ponerte en [contacto con nosotros](#).

Atentamente,
Incentro Cloud España



Google Kubernetes Engine siempre ha definido el mañana

En tecnología, una década es una eternidad. En los últimos diez años, la forma de crear y desplegar aplicaciones se ha redefinido por completo. Desde el auge de los microservicios y DevOps hasta la era de la IA nativa en cloud y el edge computing, Google Kubernetes Engine (GKE) ha estado ahí: creciendo y adaptándose contigo para cada nuevo reto. Desde Google Cloud, no solo han impulsado Kubernetes desde el principio; también han seguido siendo su principal contribuidor, dando forma a la transformación desde el primer día.

Hoy, al mirar atrás a una década de innovación, está claro que GKE no es solo la base, sino el futuro del desarrollo de aplicaciones y la ingeniería de plataformas potenciadas por IA. Con GKE abriendo camino, Kubernetes es mucho más que una solución: **es tu superpoder de IA.**

Kubernetes (K8s)

Kubernetes es un sistema open source para desplegar, escalar y gestionar aplicaciones en contenedores, en cualquier entorno. Lo desarrolló originalmente Google y se publicó como open source en 2014; en 2016 fue aceptado por la Cloud Native Computing Foundation (CNCF).

Google Kubernetes Engine (GKE)

GKE es un servicio gestionado que usa la infraestructura de Google para desplegar y operar aplicaciones en contenedores a escala. Lanzado en 2015, fue el primer servicio de Kubernetes totalmente gestionado.



Nuestra experiencia nace de la práctica

¿Dónde es mejor ejecutar Kubernetes que con su creador y mayor contribuidor open source?

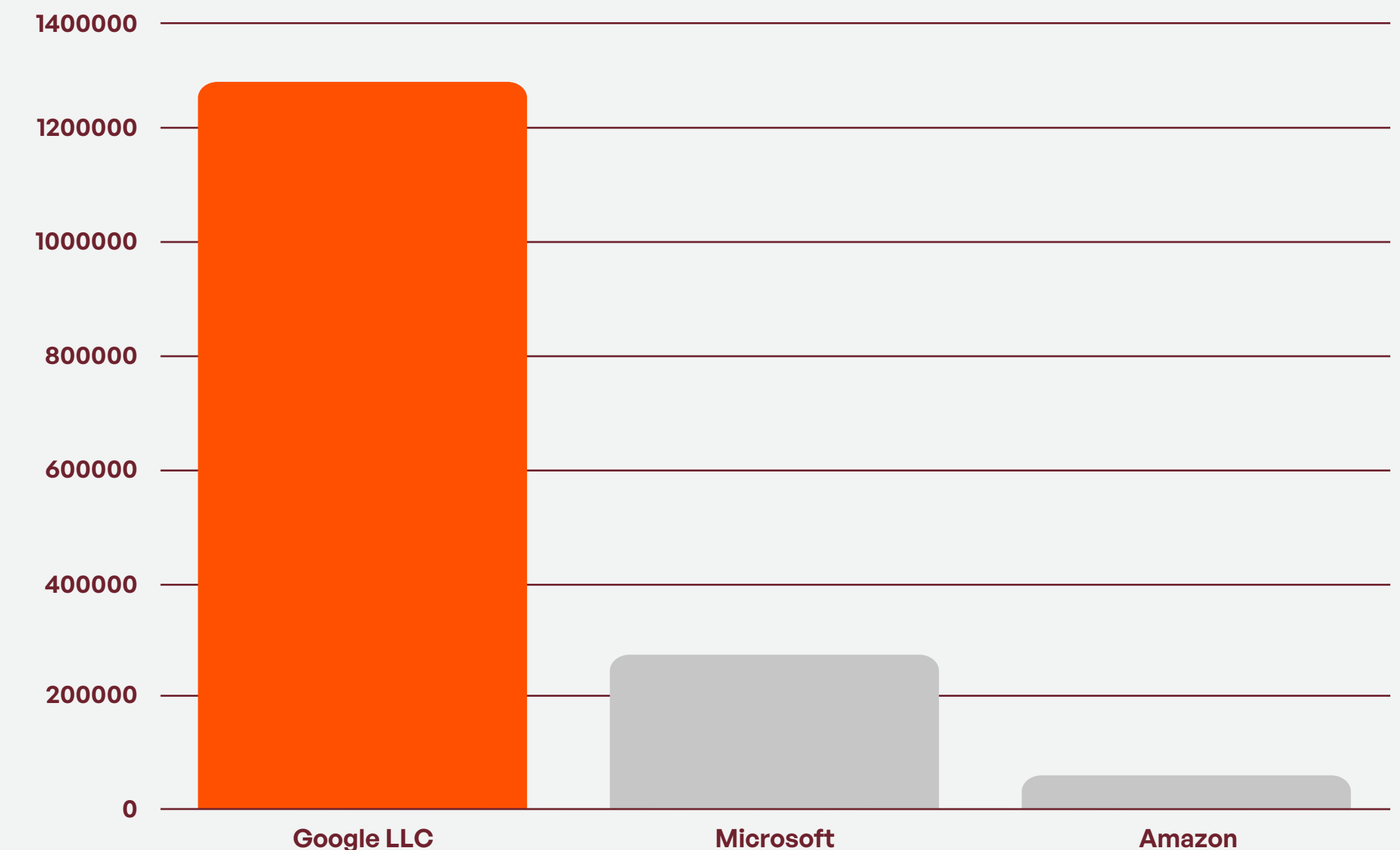
En Google llevan trabajando con contenedores desde principios de los 2000, mucho antes de que fuesen estándar en la industria. Desde Gmail y YouTube hasta Search, todo funciona en contenedores, lo que les permite desarrollar rápido, desplegar con eficiencia e innovar a escala. Con miles de millones de contenedores lanzados cada semana, Google es su mejor cliente y comparten una década de experiencia para que tu organización aproveche el poder de la innovación en contenedores.

**1 millón de
contribuciones**

Google es el mayor contribuidor a Kubernetes con más de un millón de contribuciones.¹

¹ Google Cloud (2023). [A decade of Kubernetes leadership: why Google Cloud should be your choice for Kubernetes.](#)

Contribuciones a Kubernetes por los principales proveedores cloud



[Kubernetes - DevStats](#)

¿Qué opinan nuestros clientes?

Contextual AI

Saber más →

Hotspring

Saber más →

De los primeros pasos al éxito futuro

Un vistazo a uno de los primeros clientes de GKE y a cómo han crecido en la última década.



Acerca de Signify

Antes conocida como Philips Lighting, Signify es líder mundial en iluminación para profesionales, consumidores e Internet de las Cosas. Sus productos aportan valor al negocio y transforman la vida en hogares, edificios y espacios públicos. Opera más de 153 millones de puntos de luz conectados a nivel mundial bajo marcas como Philips Hue.

Sus retos

Con un negocio repartido por todo el mundo, el equipo necesitaba escalar rápido y con eficiencia cuando distintas regiones disparaban el tráfico. También necesitaban innovar a escala sin perder el control del gasto.

Su estrategia

Como cliente de Google Cloud a largo plazo, Philips Hue adoptó Kubernetes muy pronto, apenas un año después de su publicación open source en 2014. Kubernetes y GKE se convirtieron en la base de su infraestructura, escalando para gestionar patrones de iluminación naturales en todo el mundo y manteniendo los costes bajo control.

Resultados

Durante la última década, con GKE como base, la plataforma de Philips Hue ha escalado enormemente su infraestructura para apoyar su expansión. GKE permitió absorber el crecimiento de tráfico garantizando la fiabilidad y el rendimiento de sus servicios a nivel global.

La próxima década

Para seguir impulsando el crecimiento de su plataforma, Philips Hue quiere aprovechar GKE para nuevas cargas, incluida ingeniería de plataformas e IA. Están estableciendo una práctica de platform engineering para estandarizar su infraestructura, armonizar prácticas entre equipos y liberar tiempo para el desarrollo de aplicaciones. También explorarán con GKE la IA para supervisión multiclúster, transformando un gran volumen de señales técnicas en diagnósticos accionables que ayuden a localizar rápido el origen de incidencias.

+1.150 %

en transacciones
y comandos en GKE

Antes

200 millones de
transacciones diarias

18 millones de comandos diario

1 clúster de GKE

Ahora

3.500 millones de
transacciones diarias

100 millones de comandos diarios

7 clústeres de GKE

“

Las mejoras constantes de GKE en los últimos 10 años han cambiado profundamente la forma en que diseñamos, desplegamos y evolucionamos nuestros servicios. Dedicamos menos tiempo a gestionar la infraestructura y podemos centrarnos en lo que de verdad importa: la calidad de la experiencia de usuario y la velocidad de innovación.”



Leon Bouwmeester
Director of Engineering
& Head of Hue Platform, Signify

¿Quieres profundizar en la historia de éxito de Signify?

Léela aquí →

Celebrando los éxitos, juntos

10 años de hitos que están definiendo la evolución del mañana.

Con el objetivo de ayudar a las empresas a innovar más rápido, mantenerse ágiles y manejar cargas cada vez más complejas, Google Cloud ha incorporado funciones clave en la última década que eliminan complejidad operativa y habilitan la innovación en IA.



2015

“Elegimos Kubernetes para exprimir al máximo nuestra infraestructura de aplicaciones, y decidimos migrar a Google Container Engine desde otro proveedor cloud para sacar lo mejor de Kubernetes. Nuestra infraestructura en Container Engine funciona con aproximadamente el 40 % del despliegue original en el otro proveedor cloud, y los descuentos por uso sostenido y la tarificación por minuto de Google han supuesto más ahorro.”

Jay Allen, CTO, Porch

21 de marzo

GKE admite hasta 1.000 nodos

29 de septiembre

Pokémon GO cobra vida con GKE

26 de agosto

Tras 10 meses en alfa/beta, Google Kubernetes Engine (inicialmente Google Container Engine) se convierte en el primer servicio de K8s totalmente gestionado

9-11 de noviembre

Google organiza la primera KubeCon en San Francisco

12 de noviembre

GKE admite hasta 250 nodos

2016

“Jamás he participado en algo ni remotamente comparable al crecimiento que experimentó Niantic, cliente de Google Cloud, con el lanzamiento de Pokémon GO.”

Luke Stone, Director de Customer Reliability Engineering, Google Cloud

2017

← 1 de mayo

Google lanza el Kubernetes Podcast
(sigue en emisión)

← 24 de mayo

La integración con containerd en Kubernetes
llega a disponibilidad general

← 24 de julio

GKE se ejecuta en Google Cloud
y en entornos on-premise

3 de mayo

GKE admite hasta 5.000 nodos

13 de noviembre →

Google Container Engine pasa a llamarse
oficialmente Google Kubernetes Engine

2018

“Necesitábamos una plataforma coherente para desplegar y gestionar contenedores tanto en local como en la nube. Dado que Kubernetes se ha convertido en el estándar del sector, fue natural para nosotros adoptar Kubernetes Engine en GCP para reducir el riesgo y el coste de nuestros despliegues.”

Dinesh Keswan, Director Global de Tecnología de HSBC

2019

“Nuestros clientes quieren desarrollar y desplegar sus aplicaciones en cualquier lugar, en local, en la nube pública o en múltiples nubes públicas, de forma fluida y segura. Nos entusiasma hacerlo posible integrando con Anthos las tecnologías de centro de datos, redes y seguridad de Cisco.”

Kip Compton, Vicepresidente sénior de Plataformas y Soluciones Cloud, Cisco

9 de abril

Google anuncia Anthos

15 de enero

Phoenix Labs impulsa su juego Dauntless con GKE

22 de abril

Anthos llega al multicloud

25 de junio

Bayer Crop Science escala con GKE hasta 15.000 nodos

2020

“Contenerizar servidores no es muy habitual en la industria del videojuego, sobre todo en los títulos de mayor tamaño. Google Cloud lideró este esfuerzo con su liderazgo y experiencia tecnológica, y su plataforma nos dio la flexibilidad para usar Kubernetes como servicio en producción.”

Simon Beaumont, VP de Tecnología, Phoenix Labs

2021

“GKE Autopilot es la auténtica plataforma serverless de Kubernetes que estábamos esperando. Los desarrolladores pueden centrarse en sus cargas de trabajo y dejar la gestión de la infraestructura subyacente a los SRE de Google.”

Boris Simandoff, VP de Ingeniería, Via Transportation, Inc.

28 de junio

Las GPUs con time-sharing en GKE pasan a estar disponibles de forma general

5 de julio

El mecanismo de actualización blue-green para node pools de GKE pasa a disponibilidad general



13 de julio

GKE permite ejecutar cargas de trabajo Arm con las máquinas virtuales Tau T2A



21 de diciembre

GKE incorpora nuevas opciones de conectividad y aislamiento del plano de control para clústeres

24 de febrero



Lanzamiento de GKE Autopilot

16 de septiembre



Se anuncia Backup para GKE

2022

“Estamos encantados de ejecutar nuestras cargas usando máquinas virtuales T2A basadas en Arm con GKE. Nos impresionó lo sencillo que resultó aprovisionar nodos Arm en un clúster de GKE, tanto de forma independiente como junto a nodos basados en x86. Creemos que esta arquitectura multiprocesador nos ayudará a reducir costes y, al mismo tiempo, ofrecer una mejor experiencia a nuestros clientes.”

Benjamin Rowe, Arquitecto de Cloud y Seguridad, Arcules

2023

“Con la proliferación de contenedores en toda nuestra organización, necesitábamos una forma de aumentar la velocidad de manera segura a medida que escalábamos entre múltiples equipos y cargas de trabajo. Con la nueva edición GKE Enterprise, Google Cloud ofrece una plataforma de contenedores gestionada integral, con seguridad de nivel empresarial y un gobierno optimizado para aplicaciones críticas de negocio.”

Scott Surovich, Principal Engineer & Global Container Engineering Lead, HSBC

16 de marzo



Ulta se asocia con GKE para modernizar aplicaciones

31 de agosto



Soporte de Cloud TPUs en GKE

26 de septiembre



Google es Líder en el Gartner® Magic Quadrant™ 2023 para *Container Management*

6 de noviembre



GKE Enterprise llega a disponibilidad general



18 de marzo

GKE lanza integración con NVIDIA NeMo



13 de noviembre

GKE admite hasta 65.000 nodos

2024

“El nuevo soporte de GKE para clústeres más grandes nos da la escala que necesitamos para acelerar nuestro ritmo de innovación en IA.”

James Bradbury, Head of Compute, Anthropic

Nuestros clientes están definiendo la IA del futuro con GKE

A medida que los modelos de IA escalan y aumentan las cargas, cambian las tendencias y los retos. Con el crecimiento llega la complejidad. Equilibrar coste, rendimiento y seguridad mientras la IA no deja de evolucionar es un desafío. Para afrontarlo, las organizaciones necesitan una plataforma de IA robusta que combine flexibilidad y rendimiento. GKE ofrece aceleradores flexibles y una escala sin igual para las cargas de IA más exigentes y para la inferencia eficiente en costes. Así, los equipos entrenan más rápido, sirven con más inteligencia y reducen costes.

Descubre cómo GKE les está ayudando

Moloco redujo los tiempos de entrenamiento de ML hasta 10 veces ➔

LiveX recortó los costes de atención al cliente hasta en un 85 % ➔



“

En Contextual AI estamos construyendo la próxima generación de **Retrieval Augmented Generation (RAG)**. Nuestros **Contextual Language Models (CLMs)** están optimizados de extremo a extremo para resolver los puntos débiles del RAG 1.0 y ayudar a los clientes a crear flujos de trabajo listos para producción. Para conseguirlo, nos apoyamos en GKE Autopilot, un servicio totalmente gestionado que se encarga de la complejidad de ejecutar nuestra aplicación. Con **GKE Autopilot** podemos escalar fácilmente nuestros pods, optimizar la utilización de recursos y garantizar la seguridad y disponibilidad de nuestros nodos. También aprovechamos los nuevos modelos de facturación que ofrecen GPUs más rentables para las tareas de inferencia, mientras usamos pods de Autopilot para los servicios sin GPU. Estamos muy contentos de usar GKE Autopilot para impulsar los CLMs, ahorrando dinero y mejorando nuestro rendimiento.”

Soumitr Pandey
Miembro del Equipo Técnico,
Contextual AI

“

Elegimos **GKE Autopilot** para nuestra infraestructura de ML porque permite a nuestro equipo centrarse en I+D en lugar de en la gestión de clústeres. Este enfoque no solo automatiza el aprovisionamiento de recursos en todo el clúster regional, también agiliza nuestras operaciones. Las últimas mejoras en Autopilot nos emocionan especialmente: proporcionan un pool de recursos unificado e introducen capacidades de reservas, dándonos más control para cumplir plazos.”

Jon Mason
CEO,
HotSpring

Impulsando el negocio hoy con la IA de mañana

Por qué las principales empresas eligen asociarse con GKE.

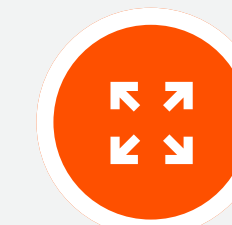
10× de crecimiento interanual

en uso de aceleradores de IA en GKE.²

² Google Cloud (2024). Google is a Leader in Gartner Magic Quadrant for Strategic Cloud Platform Services

Otra década de innovación con GKE

Desde el nacimiento de Kubernetes en 2014, GKE ha evolucionado de aplicaciones sin estado a cargas con estado, cloud híbrido y edge computing. Ahora, GKE está aprovechando plenamente el poder de la IA.



Escalabilidad

Las cargas de IA/ML varían en tamaño y complejidad. A medida que los modelos crecen, necesitas una plataforma capaz de entrenarlos y servirlos. GKE admite **clústeres a gran escala, hasta 65.000 nodos**, esenciales para IA generativa y otras aplicaciones intensivas en recursos.



Flexibilidad

Para IA/ML, la capacidad de trabajar en distintos entornos es clave. Te ayudamos a **entrenar y servir modelos en varias nubes y on-premise**, promoviendo estándares abiertos y evitando el **vendor lock-in**.



Eficiencia

Escalar modelos de IA con eficiencia es un reto. Con soporte para **GPUs, TPUs y CPUs**, GKE permite alta utilización del cómputo, garantizando rendimiento con control de costes y, en modo **Autopilot**, gestiona la infraestructura y reduce la carga operativa para que tu equipo se centre en innovar.



Seguridad

Todas las organizaciones quieren dedicar menos tiempo a preocuparse por la seguridad y obtener mejores resultados. GKE actúa como una extensión de tu equipo de seguridad, aportando **estabilidad, seguridad de nivel enterprise y cumplimiento** para tus aplicaciones cloud nativas.

¿Cómo puede ayudarte GKE?

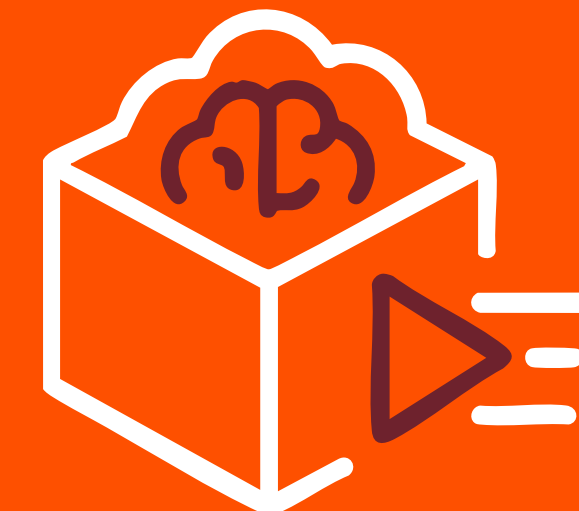
A medida que la IA redefine el futuro del desarrollo de aplicaciones, GKE está aquí para ayudarte a ir un paso por delante.



¿Cómo puedo escalar mis aplicaciones de IA en contenedores de forma rápida y eficiente para atender la demanda variable?



¿Cómo puedo optimizar el coste de ejecutar cargas de IA en contenedores en la nube?



¿Cómo puedo aprovechar los contenedores para acelerar el proceso de CI/CD de aplicaciones de IA?

¿Listo para dar el siguiente paso?

¿Quieres ver cómo Incentro y Google Cloud pueden transformar tu negocio? Nuestro equipo de expertos puede ayudarte a explorar soluciones a tu medida y acompañarte en la implementación.

[Contacta con nosotros hoy](#) para una consultoría personalizada y descubre cómo desbloquear nuevas posibilidades con Google Cloud y GKE.

Estamos aquí para ayudarte en cada paso.

contacto@incentro.com



incentro