

Big Data Aplicado

Soluciones Cloud

## Contenido

Soluciones Cloud .....	3
Coste total de propiedad .....	4
Ventajas.....	4
Modelos de Servicio en la Nube: IaaS, PaaS y SaaS .....	6
1. IaaS (Infraestructura como Servicio):.....	6
2. PaaS (Plataforma como Servicio):.....	6
3. SaaS (Software como Servicio):.....	6
Amazon Web Services.....	8
Servicios Fundamentales .....	8
Servicios de Analítica de Datos .....	10
Servicios de Machine Learning.....	10
Servicios de Seguridad .....	10
Azure .....	11
Servicios Fundamentales .....	11
Servicios de Analítica de Datos .....	12
Servicios de Inteligencia Artificial.....	12
Servicios de Seguridad .....	12
Google Cloud Platform .....	13
Servicios Fundamentales .....	13
Servicios de Analítica de Datos .....	14
Servicios de Inteligencia Artificial.....	15
Servicios de Seguridad .....	15
Comparativa y evolución .....	16

## Soluciones Cloud

El *Cloud Computing* permite obtener servicios de computación a través de internet de manera que sólo se pague por los recursos que se usan y en el momento en que se necesitan. Dicho de otro modo, es la entrega bajo demanda de potencia de cómputo, bases de datos, almacenamiento, aplicaciones y otros recursos informáticos, a través de Internet con un sistema de precios de pago por uso.

Los modelos productivos basados en la adquisición de hardware de manera propietaria han quedado atrás, al implicar un proceso largo y costoso de compra de licencias, recursos físicos como oficinas y equipamiento y recursos humanos (tanto técnicos como de seguridad) para su implantación, gestión y mantenimiento.

Se plantea un cambio de perspectiva: la infraestructura se deja de considerar hardware, para verla (y usarla) como software.



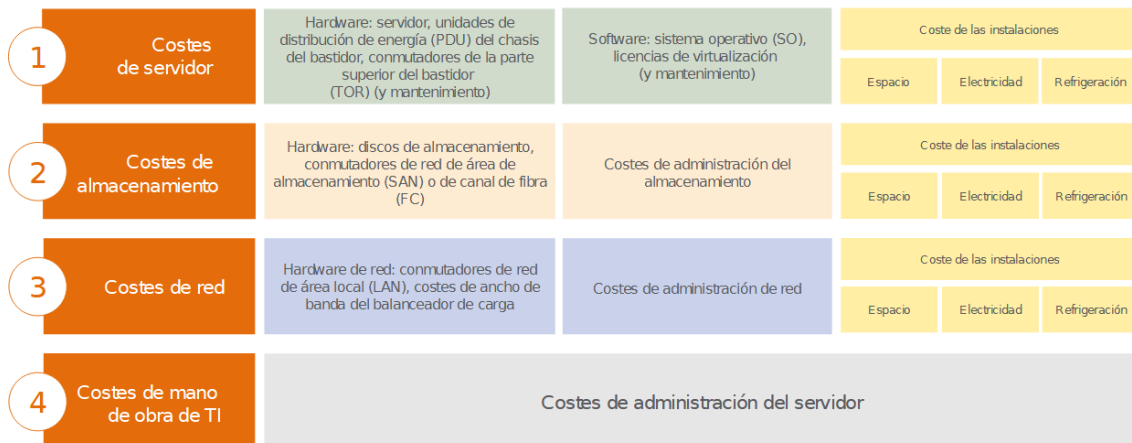
Cloud computing

Según el NIST (*National Institute for Standards and Technology*), la computación en la nube es "un modelo que permite un acceso ubicuo, conveniente y bajo demanda a una red de recursos de computación configurables (por ejemplo, redes, servidores, almacenamiento, aplicaciones y servicios) que se pueden aprovisionar rápidamente y entregar con una interacción mínima con el proveedor del servicio".

## Coste total de propiedad

El **coste total de propiedad (CTO)** es la estimación financiera que ayuda a identificar los costes directos e indirectos de un sistema. Permite **comparar el coste de ejecutar una infraestructura completa o una carga de trabajo específica** en las instalaciones propias frente a hacerlo en la nube.

Los elementos a considerar sobre el coste total de propiedad son:



## Ventajas

Combinando Big Data y Cloud Computing, las empresas pueden:

- **Acceder a grandes volúmenes de datos de forma eficiente.**
- **Procesar información de manera rápida y eficaz.**
- **Aumentar o reducir los recursos, adaptándolos a las demandas de cada proyecto, de forma casi instantánea y optimizando costes (escalabilidad).**
- **Evitar grandes inversiones iniciales y costes de financiación mediante el pago por uso (OpEX vs CapEX).**
- **Liberarse de las preocupaciones por actualizaciones de hardware, software y seguridad, ya que el proveedor del servicio se encarga de ello.**
- **Disponer de servicios con tolerancia a fallos y alta disponibilidad gracias a la arquitectura de almacenamiento y proceso distribuido.**
- **Mejora de tiempos de acceso, permitiendo el despliegue en distintas regiones.**
- **Acceder a la información desde cualquier dispositivo con acceso a internet, en cualquier momento y lugar.**

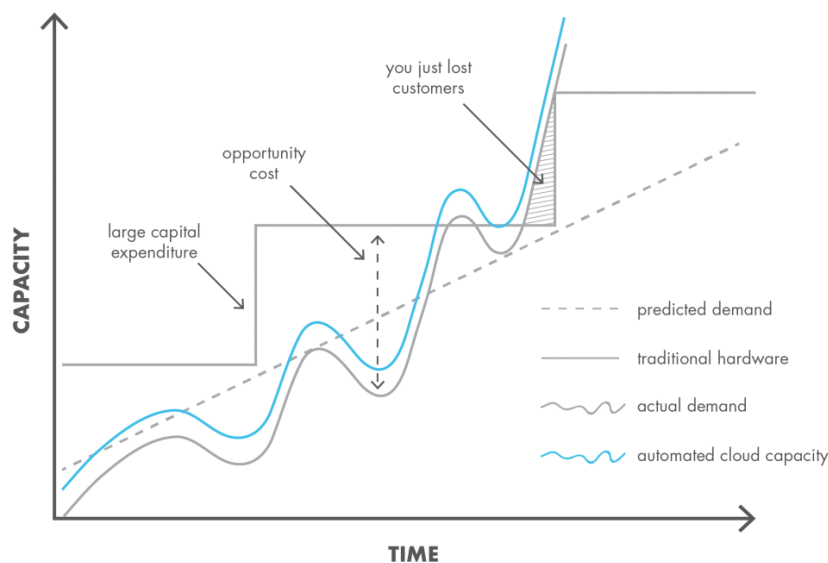
Permite analizar grandes volúmenes de datos para obtener información valiosa y tomar mejores decisiones, **con un coste optimizado y una gran flexibilidad.**

Así pues, la informática en la nube es un modelo basado en el consumo, lo que significa que los usuarios finales solo pagan por los recursos que usan, con las siguientes ventajas:

- Sin costes por adelantado.
- No es necesario comprar ni administrar infraestructuras costosas que es posible que los usuarios no aprovechen del todo, con lo cual el riesgo se reduce al mínimo.
- Se puede pagar para obtener recursos adicionales cuando se necesiten.
- Se puede dejar de pagar por los recursos que ya no se necesiten.

Con un planteamiento de infraestructura *in-house* / *on-premise* tenemos que estimar cual va a ser la necesidad de la empresa y adquirir la infraestructura por adelantado teniendo en cuenta que:

- hay que aprovisionar por encima de la demanda, lo que es un desperdicio económico.
- si la demanda crece por encima de la estimación, tendré un impacto negativo en la demanda con la consiguiente pérdida de clientes.



## Modelos de Servicio en la Nube: IaaS, PaaS y SaaS

Estos tres acrónimos representan diferentes modelos de servicio en Cloud Computing, cada uno con un nivel de abstracción distinto:

### 1. IaaS (Infraestructura como Servicio):

El proveedor ofrece acceso a recursos de infraestructura, como servidores, almacenamiento, redes y sistemas operativos.

El usuario tiene control total sobre el sistema operativo, las aplicaciones y la configuración del entorno.

Ejemplos:

- Máquinas Virtuales (VMs)
- Almacenamiento en bloque, en archivos

### 2. PaaS (Plataforma como Servicio):

El proveedor ofrece una plataforma completa para el desarrollo, ejecución y administración de aplicaciones. Incluye la infraestructura (IaaS) y servicios adicionales, como bases de datos, middleware y herramientas de desarrollo.

El usuario se centra en el desarrollo de las aplicaciones, sin preocuparse por la gestión de la infraestructura subyacente.

Ejemplos:

- Bases de datos gestionadas
- Plataformas de streaming
- Kubernetes gestionado
- Entornos preconfigurados con lenguajes de programación

### 3. SaaS (Software como Servicio):

El proveedor ofrece aplicaciones de software completas, accesibles a través de Internet. Se encarga de todo el mantenimiento, las actualizaciones y la seguridad.

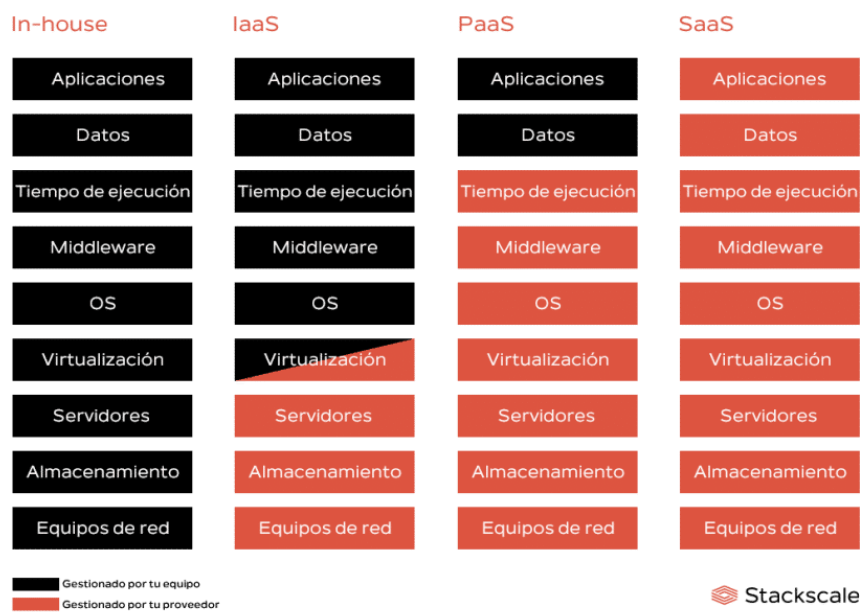
El usuario accede a la aplicación a través de un navegador web o una interfaz de cliente, sin necesidad de instalar software localmente.

Ejemplos:

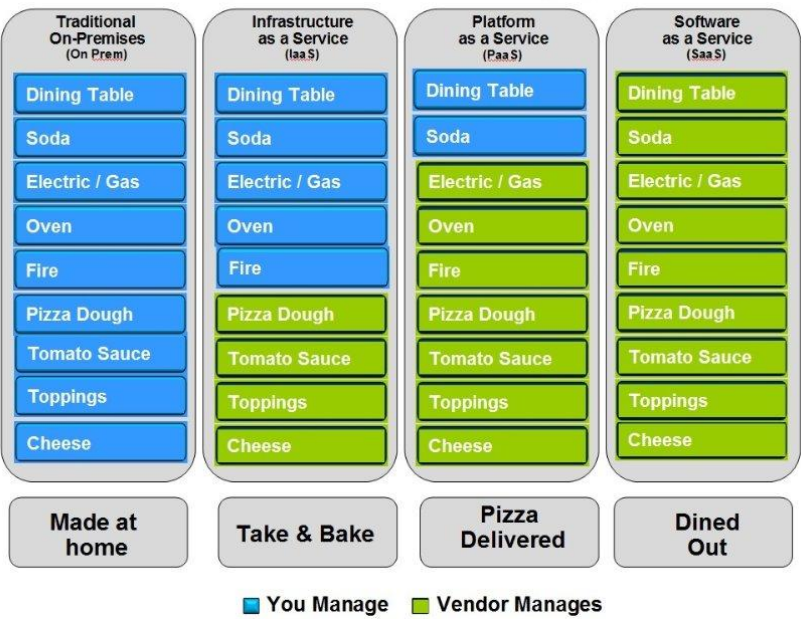
- Software de productividad: Google Workspace (Gmail, Docs, Drive), Microsoft Office 365.
- CRM: Salesforce, Dynamics 365
- Software de análisis de datos: entornos para trabajar directamente con Big Data, con herramientas para el procesamiento y análisis de datos.

Comparación:

	IaaS	PaaS	SaaS
Control	Alto	Medio	Bajo
Flexibilidad	Alta	Media	Baja
Costes	Variables	Suscripción mensual/anual	Suscripción mensual/anual
Mantenimiento	Usuario	Proveedor	Proveedor



Pizza as a Service



## Amazon Web Services

Amazon Web Services (AWS) es una plataforma de computación en la nube que ofrece una amplia gama de servicios, abarcando desde infraestructura básica hasta soluciones especializadas en Big Data, Machine Learning y otras. Estos servicios funcionan a modo de piezas de un puzzle, de manera que uniendo unos con otros podemos diseñar la arquitectura necesaria para nuestras aplicaciones.



## Servicios Fundamentales

- **Computación:**
  - **Amazon EC2 (Elastic Compute Cloud):** Permite alquilar máquinas virtuales (VMs) con diferentes capacidades de hardware y software para ejecutar aplicaciones y servicios.
  - **Contenedores:** AWS ofrece servicios como ECS (Elastic Container Service) y EKS (Elastic Kubernetes Service) para gestionar aplicaciones en contenedores, proporcionando escalabilidad y flexibilidad.
  - **Computación sin Servidor:** AWS Lambda permite ejecutar código sin necesidad de administrar servidores, lo que simplifica el desarrollo y reduce costes.
- **Almacenamiento:**
  - **Amazon EBS (Elastic Block Store):** Ofrece volúmenes de almacenamiento en bloque que se pueden conectar a instancias EC2, actuando como discos duros virtuales.



- **Amazon EFS (Elastic File System):** Proporciona un sistema de archivos de red escalable y gestionado que se puede compartir entre varias instancias EC2.
  - **Amazon S3 (Simple Storage Service):** Permite almacenar y recuperar cualquier cantidad de datos desde cualquier lugar de la web, ideal para almacenamiento de objetos y archivos.
  - **Servicios de Transferencia de Datos:** Snowball y Snowmobile facilitan la transferencia de grandes volúmenes de datos hacia o desde la nube de AWS.
- **Redes:**
    - **Amazon VPC (Virtual Private Cloud):** Permite crear redes virtuales privadas dentro de la nube de AWS, ofreciendo control total sobre el entorno de red.
    - **AWS Direct Connect:** Establece conexiones dedicadas entre la infraestructura local y la nube de AWS, mejorando el rendimiento y la seguridad.
    - **Amazon Route 53:** Un servicio de DNS (Domain Name System) escalable y altamente disponible que facilita la gestión de nombres de dominio y la asignación de direcciones IP.
    - **Amazon CloudFront:** Una red de entrega de contenido (CDN) global que distribuye contenido web, como páginas web, videos e imágenes, de forma rápida y eficiente.
- **Bases de Datos:**
    - **Amazon RDS (Relational Database Service):** Ofrece bases de datos relacionales gestionadas, incluyendo MySQL, PostgreSQL, Oracle, SQL Server y MariaDB, simplificando la administración y el escalado.
    - **Amazon Aurora:** Una base de datos relacional compatible con MySQL y PostgreSQL de alto rendimiento y disponibilidad.
    - **Bases de Datos NoSQL:** Amazon DynamoDB (base de datos clave-valor y de documentos) y Amazon DocumentDB (compatible con MongoDB) ofrecen soluciones NoSQL escalables y flexibles.
    - **Amazon ElastiCache:** Proporciona un servicio de almacenamiento en caché en memoria para mejorar el rendimiento de las aplicaciones al almacenar en caché los datos de acceso frecuente.

## Servicios de Analítica de Datos

- **Amazon Redshift:** Un servicio de almacenamiento de datos a escala de petabytes que permite realizar consultas y análisis rápidos de grandes conjuntos de datos.
- **AWS Glue:** Simplifica la preparación de datos para análisis, permitiendo descubrir, limpiar, transformar y cargar datos en otros servicios de AWS.
- **Amazon Athena:** Permite realizar consultas interactivas en datos almacenados en S3 utilizando SQL estándar, sin necesidad de configurar servidores o clústeres.
- **Amazon Kinesis:** Una familia de servicios para recopilar, procesar y analizar datos en tiempo real, incluyendo Kinesis Data Streams, Kinesis Data Firehose y Kinesis Data Analytics.
- **Amazon QuickSight:** servicio de inteligencia empresarial que permite crear visualizaciones interactivas, paneles y análisis a partir de datos de AWS.
- **Amazon EMR (Elastic MapReduce):** Un servicio gestionado que permite procesar grandes conjuntos de datos utilizando marcos de Big Data como Apache Hadoop y Apache Spark.

## Servicios de Machine Learning

- **Amazon SageMaker:** Una plataforma de Machine Learning completa que permite a los desarrolladores y científicos de datos construir, entrenar e implementar modelos de aprendizaje automático.
- **Servicios de IA Preentrenados:** AWS ofrece APIs para tareas de IA comunes, como reconocimiento de imágenes (Amazon Rekognition), procesamiento de lenguaje natural (Amazon Comprehend), traducción de idiomas (Amazon Translate), transcripción de voz a texto (Amazon Transcribe) y generación de voz (Amazon Polly).

## Servicios de Seguridad

- **AWS IAM (Identity and Access Management):** Permite gestionar el acceso a los recursos de AWS, incluyendo la autenticación, autorización y auditoría.
- **AWS Security Hub:** Proporciona una vista centralizada del estado de seguridad de la cuenta de AWS, incluyendo alertas de seguridad y recomendaciones.

- **Amazon GuardDuty:** Un servicio de detección de amenazas que analiza de forma continua los registros de actividad de AWS para identificar actividad maliciosa.
- **AWS Inspector:** Un servicio de evaluación de vulnerabilidades que analiza las instancias EC2 en busca de vulnerabilidades de seguridad.
- **AWS Shield:** Un servicio de protección contra DDoS que protege las aplicaciones que se ejecutan en AWS contra ataques de denegación de servicio distribuidos.

## Azure

### Servicios Fundamentales

- **Computación:**
  - **Azure Virtual Machines:** Permite crear y gestionar máquinas virtuales (VMs) en la nube de Azure.
  - **Azure Container Instances (ACI):** Ofrece la forma más simple y rápida de ejecutar contenedores en la nube de Azure.
  - **Azure Kubernetes Service (AKS):** Un servicio gestionado para la orquestación de contenedores utilizando Kubernetes.
  - **Azure Functions:** servicio de computación sin servidor.
  - **Azure Stack:** Permite extender las capacidades de Azure a entornos locales.
- **Almacenamiento:** Azure ofrece una variedad de servicios de almacenamiento para satisfacer diferentes necesidades, incluyendo:
  - **Azure Blob Storage:** Un servicio de almacenamiento de objetos, similar a Amazon S3.
  - **Azure Data Lake Storage:** almacenamiento optimizado para grandes volúmenes de datos en su formato original.
  - **Azure File Storage:** Permite compartir archivos en la nube de Azure.
- **Redes:** Azure proporciona servicios de red para conectar recursos en la nube y con entornos locales. Algunos ejemplos son:
  - **Azure Virtual WAN (V-WAN):** Permite crear redes de área amplia virtuales para conectar diferentes ubicaciones y recursos de Azure.
  - **Azure ExpressRoute:** Ofrece conexiones privadas y dedicadas entre la infraestructura local y la nube de Azure.

- **Bases de Datos:** Azure, al igual que otras plataformas cloud, ofrece una gama de servicios de bases de datos, tanto relacionales como NoSQL. Las fuentes mencionan:
  - **Azure SQL Database:** base de datos relacional gestionada.
  - **Azure Cosmos DB:** base de datos NoSQL globalmente distribuida y multimodelo. Permite elegir entre diferentes modelos de datos NoSQL, como documentos, grafos y clave-valor, brindando flexibilidad para diferentes casos de uso.
  - **Azure Table Storage:** base de datos NoSQL para almacenar datos estructurados.

### Servicios de Analítica de Datos

- **Azure Synapse Analytics:** Un servicio unificado de análisis de datos para cargas de trabajo de bases de datos tradicionales y Big Data.
- **Azure Data Factory (ADF):** integración de datos.
- **Azure HDInsight:** análisis de Big Data.
- **Azure Databricks:** análisis de datos basado en Apache Spark.
- **Azure Event Hubs** y **Azure Stream Analytics** se pueden integrar con Kafka para la ingesta y el procesamiento de datos en tiempo real.

### Servicios de Inteligencia Artificial

- **Azure Machine Learning:** Un servicio de Machine Learning para crear, entrenar e implementar modelos de aprendizaje automático.
- **Azure Cognitive Services:** Una colección de servicios de IA preentrenados para tareas como el reconocimiento de imágenes, el procesamiento del lenguaje natural y la traducción de idiomas.
- **Azure OpenAI Service:** Proporciona acceso a modelos de lenguaje de gran tamaño como GPT-3 y Codex.

### Servicios de Seguridad

- **Azure Security Center:** seguridad centralizada para la nube de Azure.
- **Azure Active Directory (AAD):** gestión de identidades y accesos.
- **Azure Key Vault:** administración de claves y secretos de forma segura.

# Google Cloud Platform

## Servicios Fundamentales

- **Computación:** Google Cloud ofrece una variedad de servicios de computación para ejecutar aplicaciones y cargas de trabajo:
  - **Google Compute Engine (GCE):** Permite crear y administrar máquinas virtuales (VMs) en la nube de Google. GCE se asemeja a las instancias EC2 de AWS y a las Azure Virtual Machines. Se debe especificar el hardware (CPU, RAM, HD) y software (OS y apps) al aprovisionar una VM en GCE.
  - **Google Kubernetes Engine (GKE):** Un servicio completamente gestionado para implementar y administrar aplicaciones en contenedores utilizando Kubernetes. Kubernetes se desarrolló originalmente en Google y se lanzó como código abierto en 2014. GKE ofrece escalado automático, actualizaciones automáticas y seguridad integrada.
  - **Google Cloud Functions:** Permite ejecutar código sin necesidad de administrar servidores, lo que simplifica el desarrollo y reduce costos. Similar a AWS Lambda y Azure Functions.
  - **Google App Engine (GAE):** Una plataforma sin servidor completamente gestionada para desarrollar y alojar aplicaciones web a escala.
  - **Google Container Registry (GCR):** Almacena imágenes de contenedores Docker.
- **Almacenamiento:**
  - **Google Cloud Storage (GCS):** Un servicio de almacenamiento de objetos, similar a Amazon S3 y Azure Blob Storage.
  - **Cloud Memorystore:** Un servicio de almacenamiento en caché en memoria completamente administrado.
- **Redes:** servicios de red para conectar recursos en la nube y con entornos locales:
  - **Virtual Private Cloud (VPC):** Permite crear redes virtuales privadas dentro de la nube de Google, ofreciendo control total sobre el entorno de red.

- **Google Cloud Armor:** Un servicio de protección contra DDoS que protege las aplicaciones en Google Cloud contra ataques de denegación de servicio.
- **Google Cloud Transfer Service:** Facilita la transferencia de datos a GCS.
- **Bases de datos:**
  - **Cloud SQL:** Un servicio de base de datos relacional completamente administrado. Compatible con MySQL y PostgreSQL.
  - **Cloud Spanner:** Una base de datos relacional globalmente distribuida y de escala horizontal.
  - **Cloud Firestore:** Una base de datos de documentos NoSQL completamente administrada, diseñada para almacenar, sincronizar y consultar datos para aplicaciones móviles, web y de IoT. Almacena datos en documentos, utilizando estructuras similares a JSON, y proporciona actualizaciones en tiempo real.
  - **Cloud Bigtable:** Un servicio de base de datos NoSQL de alta disponibilidad y baja latencia, ideal para almacenar grandes conjuntos de datos.

## Servicios de Analítica de Datos

- **BigQuery:** Un almacén de datos a escala de petabytes completamente gestionado. Permite almacenar y analizar grandes cantidades de datos estructurados y no estructurados.
- **Cloud Dataflow:** Un servicio de procesamiento de datos completamente administrado, que permite la ejecución de canalizaciones de datos para el procesamiento por lotes y en tiempo real. Ofrece escalado automático, reequilibrio dinámico del trabajo y tolerancia a fallos.
- **Cloud Dataproc:** Un servicio gestionado de Hadoop y Spark. Permite procesar grandes conjuntos de datos de forma escalable.
- **Cloud Pub/Sub:** Un servicio de mensajería en la nube que permite a las aplicaciones intercambiar mensajes de forma fiable, rápida y asíncrona.
- **GCP Dataplex:** Una solución para la gestión de la calidad de los datos. Utiliza Auto Data Quality para proporcionar información sobre los datos, automatizar y simplificar la definición de la calidad con recomendaciones y flujos de trabajo.

## Servicios de Inteligencia Artificial

- **Vertex AI:** plataforma unificada de Machine Learning para construir, entrenar e implementar modelos de aprendizaje automático. Simplifica la gestión de los servicios de Google Cloud ML.
- **Google Cloud AutoML:** Permite crear modelos de aprendizaje automático personalizados sin necesidad de experiencia en codificación.
- **Google Cloud ML API:** Una suite de modelos de Machine Learning preentrenados que los desarrolladores pueden usar para agregar funciones avanzadas de IA a las aplicaciones sin entrenamiento de modelos de Machine Learning.

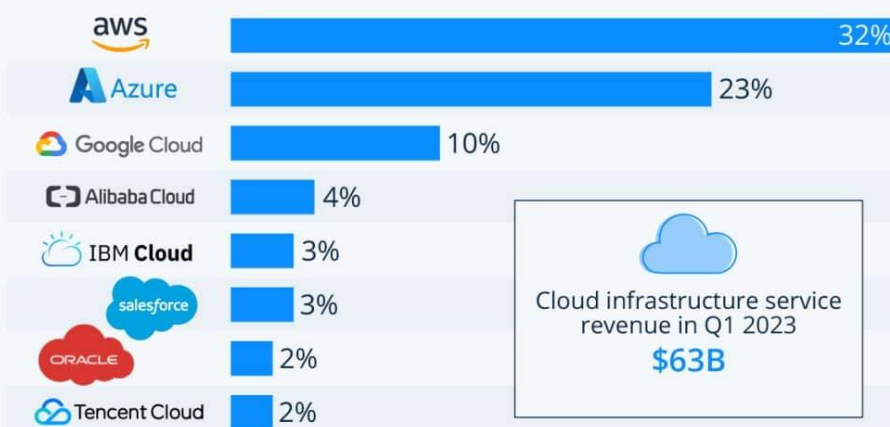
## Servicios de Seguridad

- **GCP IAM (Identity and Access Management):** Permite a las organizaciones administrar el acceso a los servicios y recursos de GCP. Proporciona control de acceso granular, administración centralizada de usuarios y acceso basado en políticas.
- **GCP IAP (Identity-Aware Proxy):** Proporciona acceso seguro a las aplicaciones que se ejecutan en GCP. Proporciona control de acceso contextual, autenticación multifactor y acceso remoto seguro.
- **GCP DLP (Data Loss Prevention):** Ayuda a las organizaciones a descubrir y proteger información sensible.
- **Google Cloud Security Command Center (SCC):** Una plataforma de gestión de seguridad integral que proporciona visibilidad, información y herramientas para ayudar a gestionar los riesgos de seguridad en los recursos y aplicaciones de GCP.

## Comparativa y evolución

### Big Three Dominate the Global Cloud Market

Worldwide market share of leading cloud infrastructure service providers in Q1 2023\*

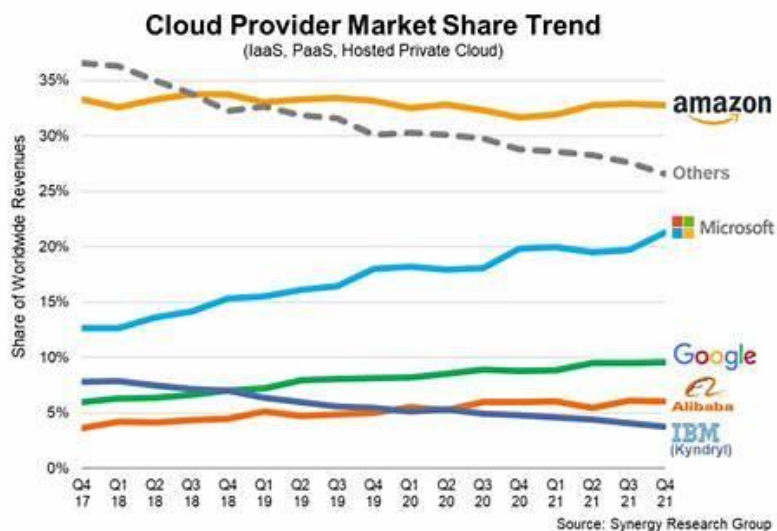


\* includes platform as a service (PaaS) and infrastructure as a service (IaaS) as well as hosted private cloud services

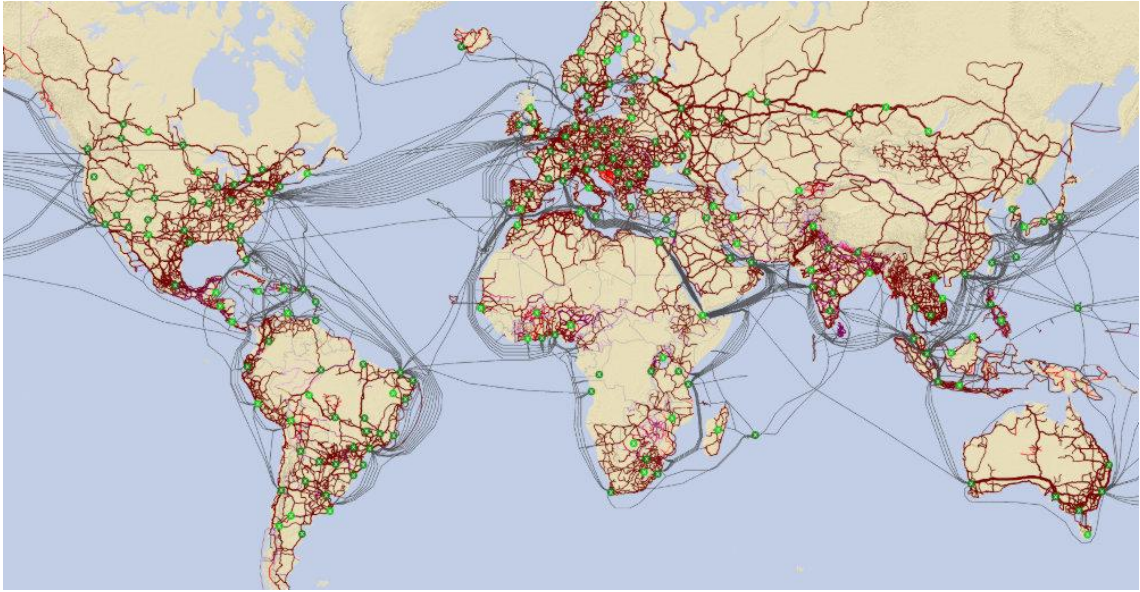
Source: Synergy Research Group



statista







[Submarine Cable Map](#)

[Data Centers - Database \(datacentermap.com\)](#)

## Ejercicios

1. Contesta a las siguientes preguntas justificando tus respuestas: