

# Tema 1. Procesamiento de datos escalable

## Personalización y recomendaciones

### ► Descripción:

AWS EMR puede ayudar a construir sistemas de recomendaciones personalizadas basados en datos históricos y patrones de usuario.

### ► Ejemplo:

### ► Contexto: un servicio de *streaming* quiere ofrecer recomendaciones personalizadas sobre películas.

### ► Solución:

- Spark en EMR procesa datos de visualización de usuarios y utiliza algoritmos de filtrado colaborativo.
- Se generan listas de películas recomendadas para cada usuario.
- Los resultados se integran directamente en la aplicación del servicio.

# Tema 1. Procesamiento de datos escalable

## Indexación y búsqueda de datos

### ► Descripción:

Herramientas como Elasticsearch pueden integrarse con EMR para indexar grandes volúmenes de datos para búsquedas rápidas.

### ► Ejemplo:

- Contexto: un portal de empleo necesita indexar millones de currículums para búsquedas rápidas por parte de reclutadores.

### ► Solución:

- Hadoop en EMR se utiliza para procesar y transformar los datos.
- Los datos transformados se cargan en Elasticsearch.
- Los reclutadores obtienen resultados instantáneos al buscar candidatos por habilidades o experiencia.

# Tema 1. Procesamiento de datos escalable

## **Análisis de *logs***

### ▶ Descripción:

EMR es ampliamente utilizado para analizar *logs* de servidores, aplicaciones o redes, para identificar patrones o anomalías.

### ▶ Ejemplo:

### ▶ Contexto: una empresa quiere analizar *logs* de sus servidores para identificar problemas de rendimiento.

### ▶ Solución:

- Los *logs* se almacenan en S3 y se procesan en EMR con Apache Spark.
- Se generan informes que identifican picos de uso, errores frecuentes y tiempos de respuesta lentos.
- Los resultados ayudan al equipo de TI a optimizar la infraestructura.

# Tema 1. Procesamiento de datos escalable

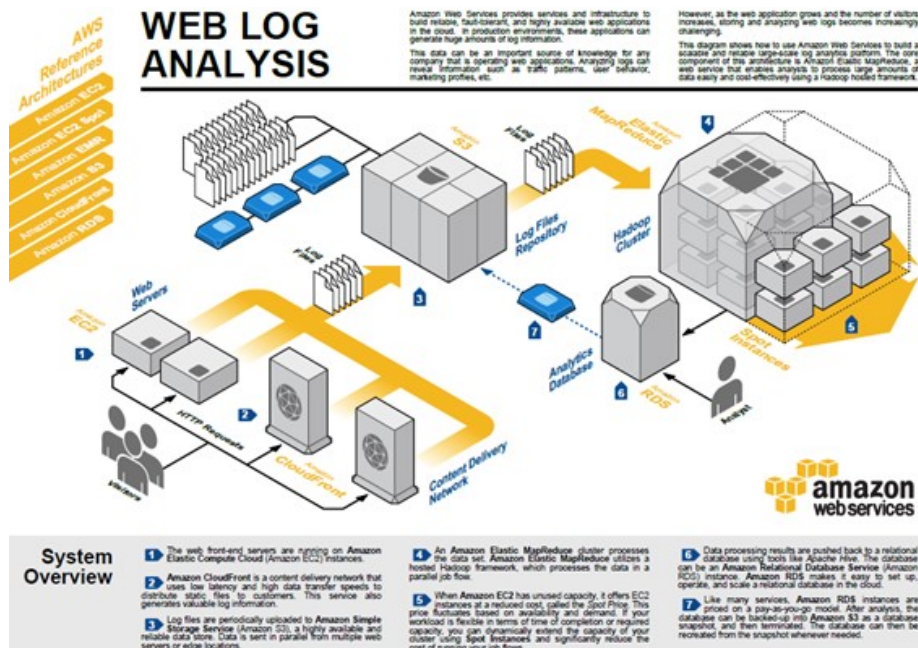


Figura 21. Infografía del procesamiento logs AWS services. Fuente: Kalakota, 2011.

# Tema 1. Procesamiento de datos escalable

## 1.4. Referencias bibliográficas

Amazon. (s. f.). *Tutorial: getting started with Amazon EMR.*  
<https://docs.aws.amazon.com/emr/latest/ManagementGuide/emr-gs.html>

Analytics. (s. f.). *All data international is confluent partner in Indonesia.*  
<https://www.adianalytics.com/confluent/>

Blancarte, Ó. (2017, 17 de marzo). Escalabilidad horizontal y vertical. Óscar Blancarte [blog]. <https://www.oscarblancarteblog.com/2017/03/07/escalabilidad-horizontal-y-vertical/>

Fortra. (s. f.). *What is batch processing?* Jams Scheduler.  
<https://www.jamsscheduler.com/solutions/batch-processing/>

Goller, A. (2019, 30 de julio). *Scaling up vs. scaling out your security segmentation.* LinkedIn. <https://www.linkedin.com/pulse/scaling-up-vs-out-your-security-segmentation-alexander-goller>

Grygoryan, K. y Peyer, S. (2021, 10 de agosto). *Power your Kafka streams application with Amazon MSK and AWS Fargate.* Amazon.  
<https://aws.amazon.com/es/blogs/big-data/power-your-kafka-streams-application-with-amazon-msk-and-aws-fargate/>

Hazel cast. (s. f.). *What is stream processing?*  
<https://hazelcast.com/foundations/event-driven-architecture/stream-processing/>

Kafka. (s. f.). *Architecture.*  
<https://kafka.apache.org/11/documentation/streams/architecture>

# Tema 1. Procesamiento de datos escalable

Kalakota, R. (2011, 13 de agosto). Analytics-as-a-Service: Understanding how Amazon.com is changing the rules. *Practical Analytics* [blog]. <https://practicalanalytics.wordpress.com/2011/08/13/analytics-as-a-service-understanding-how-amazon-com-is-changing-the-rules/>

Kumar, A. (2023, 26 de mayo). *Creación de clústeres de AWS EMR con Spark Big Data Processing. Python, PySpark, instrucciones paso a paso*. Medium. <https://awstip.com/aws-emr-cluster-creation-with-spark-big-data-processing-python-pyspark-step-by-step-instructions-c5eddc06dd6>

Maindola, A. y Yang, M. (2023, 18 de octubre). *Run Apache Hive workloads using Spark SQL with Amazon EMR on EKS*. Aws Amazon. <https://aws.amazon.com/es/blogs/big-data/run-apache-hive-workloads-using-spark-sql-with-amazon-emr-on-eks/>

nOps. (2024, 30 de mayo). *What is scalability in cloud computing? Types, benefits, and practical advice*. <https://www.nops.io/blog/cloud-scalability/>

Rao, Jun. (s. f.). *Overview of Kafka architecture*. <https://developer.confluent.io/courses/architecture/get-started/>

Rupert. (2024). *Data latency is an obvious concept to most data practitioners*. LinkedIn. [https://www.linkedin.com/posts/hirupert\\_data-latency-is-an-obvious-concept-to-most-activity-7130539982842146816-gXWM](https://www.linkedin.com/posts/hirupert_data-latency-is-an-obvious-concept-to-most-activity-7130539982842146816-gXWM)

Sakthivel, B. (2021, 15 de diciembre). *Cómo Goldman Sachs creó el etiquetado de personas con Apache Flink en Amazon EMR*. Noise. <https://noise.getoto.net/2021/12/15/how-goldman-sachs-built-persona-tagging-using-apache-flink-on-amazon-emr/>

# Tema 1. Procesamiento de datos escalable

Scylla DB. (s. f.). *Consistency models definition.*

<https://www.scylladb.com/glossary/consistency-models/>

Sodabatina, R., Aliabadi, A., Galleno, A. y Hom, J. (2021, 8 de octubre). How parametric built Audit surveillance using AWS data lake architecture.

<https://aws.amazon.com/es/blogs/architecture/how-parametric-built-audit-surveillance-using-aws-data-lake-architecture/>

Soft serve. (s. f.). Our aws MSK expertise. <https://www.softserveinc.com/en-us/amazon-msk-competency>

Why to easy learn. (s. f.). *Kafka features.* <https://waytoeasylearn.com/learn/kafka-features/>

# Tema 2. Almacenamiento escalable

## 2.1. Introducción y objetivos

Los objetivos generales de este tema son los siguientes:

- ▶ Comprender los fundamentos y características principales de las bases de datos NoSQL.
- ▶ Diferenciar entre los distintos tipos de bases de datos NoSQL: clave-valor, documentales, columnares y de grafos.
- ▶ Identificar casos de uso ideales para cada tipo de base de datos NoSQL, según sus ventajas y limitaciones.

Por su parte, los objetivos específicos son:

- ▶ Explicar las diferencias clave entre las bases de datos relacionales y NoSQL, incluyendo modelos de datos, escalabilidad y casos de uso.
- ▶ Analizar la arquitectura y el funcionamiento de las bases de datos clave-valor y su aplicación en sistemas de caché, sesiones y configuraciones.
- ▶ Describir cómo funcionan las bases de datos documentales, su flexibilidad estructural y su uso en aplicaciones web dinámicas y catálogos de productos.
- ▶ Comprender el modelo de almacenamiento columnar y su optimización para cargas de trabajo analíticas y *big data*.
- ▶ Explicar el modelo basado en grafos para almacenar relaciones complejas y sus aplicaciones en redes sociales, sistemas de recomendación y análisis de rutas.



## Tema 2. Almacenamiento escalable

- ▶ Implementar ejemplos prácticos básicos con herramientas populares como Redis, MongoDB, Apache Cassandra y Neo4j.
- ▶ Entender los conceptos fundamentales de Amazon DynamoDB y cómo usar esta base de datos.
- ▶ Entender los conceptos fundamentales de Amazon DocumentDB y cómo usar esta base de datos.

El almacenamiento escalable es un concepto fundamental en la gestión de datos modernos, que permite a las organizaciones **adaptarse** dinámicamente a las necesidades crecientes o cambiantes de almacenamiento **sin comprometer** el rendimiento, la seguridad o la accesibilidad de los datos.

Este enfoque es esencial en un mundo donde los **volúmenes de datos** crecen exponencialmente debido a la digitalización, el Internet de las Cosas (IoT), las redes sociales y otros avances tecnológicos.

### Definición básica

El almacenamiento escalable es la capacidad de un sistema de almacenamiento para aumentar o disminuir su capacidad y rendimiento de manera flexible, según las necesidades, sin interrumpir las operaciones existentes. En otras palabras, es la habilidad de **expandirse** (o contraerse) de forma eficiente, en respuesta a las **demandas** del negocio o de los usuarios.

# Tema 2. Almacenamiento escalable

## ¿Por qué es importante el almacenamiento escalable?

La importancia del almacenamiento escalable radica en su capacidad para ofrecer **soluciones adaptativas** y resilientes en un entorno tecnológico que evoluciona constantemente. Algunas de las razones clave incluyen:

- ▶ **Crecimiento exponencial de datos.** Se estima que la cantidad de datos generados en el mundo se duplica cada pocos años. Sin un sistema de almacenamiento que pueda crecer y adaptarse, sería imposible gestionar este volumen creciente.
- ▶ **Rendimiento y experiencia del usuario.** En sistemas modernos, como aplicaciones de *streaming* o videojuegos en línea, los usuarios esperan una experiencia fluida. Un almacenamiento escalable garantiza que los datos se entreguen rápidamente, incluso durante picos de tráfico.
- ▶ **Eficiencia operativa.** Permite a las empresas optimizar recursos, pagando solo por la capacidad que realmente necesitan en cada momento.
- ▶ **Soporte para nuevas tecnologías.** Tecnologías emergentes como la inteligencia artificial, el aprendizaje automático y la analítica avanzada requieren una infraestructura de almacenamiento que pueda manejar grandes volúmenes de datos de forma ágil y eficiente.
- ▶ **Continuidad del negocio.** Con almacenamiento escalable, las empresas pueden asegurar que sus sistemas se mantengan operativos durante situaciones imprevistas, como aumentos repentinos en la demanda o fallos en la infraestructura.

# Tema 2. Almacenamiento escalable

## Tipos de escalabilidad en el almacenamiento

Al igual que ocurre con el procesamiento, existen dos tipos de escalabilidad en el almacenamiento:

- ▶ **Escalabilidad horizontal** (*scale-out*). Consiste en añadir más nodos o unidades de almacenamiento al sistema para aumentar la capacidad total.
  - Ejemplo: en un sistema de almacenamiento distribuido, se pueden añadir nuevos servidores para gestionar más datos.
- ▶ **Escalabilidad vertical** (*scale-up*). Se logra aumentando la capacidad o el rendimiento de una unidad de almacenamiento existente, como añadir más discos duros o memoria a un servidor.
  - Ejemplo: ampliar la capacidad de un único servidor con discos SSD de mayor rendimiento.

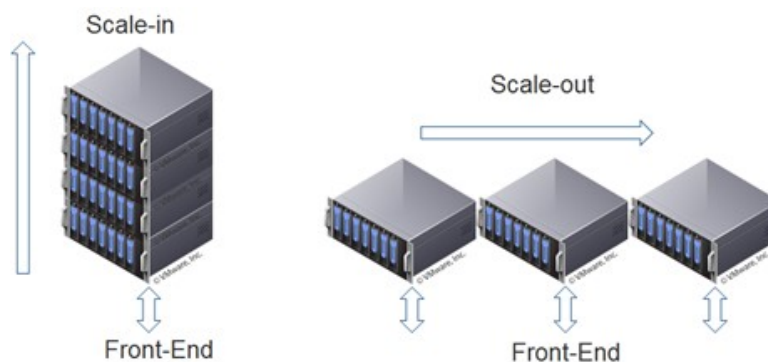


Figura 1. Imagen representativa de la escalabilidad horizontal-vertical de almacenamiento. Fuente: Mauro, 2014.

## Tema 2. Almacenamiento escalable

Veamos una tabla comparativa:

Escalabilidad	Ventajas	Desventajas
Horizontal	Alta tolerancia a fallos, expansible.	Puede ser compleja de gestionar.
Vertical	Más sencilla de implementar.	Limitada por las capacidades físicas.

Tabla 1. Comparativa entre el almacenamiento en la escalabilidad horizontal y en la vertical.

### Componentes clave del almacenamiento escalable

Para entender mejor el almacenamiento escalable, veamos sus componentes clave:

- 1.Capacidad dinámica.** Habilidad para ajustar la capacidad de almacenamiento sin interrupciones.
- 2.Alto rendimiento.** Mantener velocidades de lectura o escritura consistentes incluso al aumentar la capacidad.
- 3.Distribución inteligente.** Los datos se distribuyen de manera eficiente para evitar puntos de congestión.
- 4.Tolerancia a fallos.** Redundancia integrada para asegurar la disponibilidad de los datos, incluso si una parte del sistema falla.
- 5.Integración con la nube.** Muchas soluciones de almacenamiento escalable están diseñadas para trabajar de forma nativa con proveedores de servicios en la nube como AWS, Google Cloud o Azure.

# Tema 2. Almacenamiento escalable

## Almacenamiento escalable en la práctica

Ahora veremos algunos ejemplos prácticos de almacenamiento escalable que nos ayudarán a comprender mejor este concepto, para luego profundizar en ventajas, limitaciones, casos reales y el futuro del almacenamiento escalable.

### **Ejemplo 1. Empresas de *streaming***

Imagina una plataforma como Netflix o Spotify. Durante eventos especiales o estrenos populares, hay un aumento repentino en el tráfico. El almacenamiento escalable permite que estos datos se procesen y se entreguen a los usuarios sin interrupciones.

### **Ejemplo 2. Comercio electrónico**

Durante el Black Friday, una tienda en línea experimenta un incremento drástico en las ventas. Con almacenamiento escalable, los datos de transacciones, inventarios y usuarios se gestionan sin colapsar los sistemas.

Durante el Black Friday, una tienda en línea experimenta un incremento drástico en las ventas. Con almacenamiento escalable, los datos de transacciones, inventarios y usuarios se gestionan sin colapsar los sistemas.

### **Ejemplo 3. Internet de las cosas (IoT)**

En un sistema de ciudades inteligentes, los sensores recopilan datos constantemente. El almacenamiento escalable asegura que estos datos se almacenen y analicen eficientemente, a medida que se añaden más dispositivos.

# Tema 2. Almacenamiento escalable

## Ventajas del almacenamiento escalable

El almacenamiento escalable ofrece múltiples ventajas clave que impulsan la eficiencia y la flexibilidad de los sistemas modernos. Entre ellas destacan las siguientes:

- 1.Ahorro de costos.** Se paga solo por lo que se utiliza, evitando inversiones iniciales elevadas.
- 2.Rendimiento consistente.** Los usuarios finales no experimentan lentitud o interrupciones, incluso en momentos de alta demanda.
- 3.Adaptabilidad a nuevas tecnologías.** Facilita la integración de *big data*, inteligencia artificial y aprendizaje automático, que requieren grandes volúmenes de datos.
- 4.Redundancia y seguridad.** Mecanismos integrados para proteger datos críticos y garantizar su recuperación en caso de fallos.
- 5.Flexibilidad y agilidad.** Permite implementar nuevas aplicaciones y servicios sin preocuparse por las limitaciones de capacidad de almacenamiento.
- 6.Tolerancia a fallos.** Los sistemas escalables suelen incorporar redundancia, lo que mejora la resiliencia frente a fallos.

## Tema 2. Almacenamiento escalable

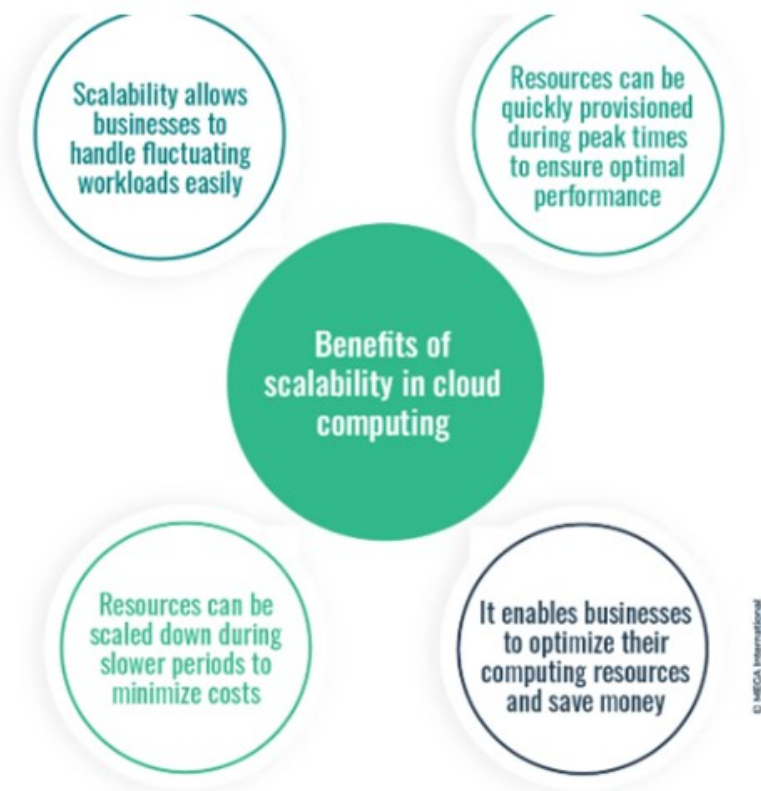


Figura 2. Beneficios de la escalabilidad en el almacenamiento en la nube. Fuente: Estrach, 2023.

### Limitaciones del almacenamiento escalable

Existe una serie de limitaciones que también es necesario tener en cuenta. Las vemos a continuación:

- 1.Complejidad de gestión.** En sistemas muy distribuidos, puede ser complicado garantizar que los datos estén sincronizados.
- 2.Costos operativos a largo plazo.** Aunque los costos iniciales son bajos, el uso continuo puede acumular gastos importantes.
- 3.Dependencia de proveedores.** Muchas soluciones de almacenamiento escalable dependen de servicios en la nube, lo que puede limitar la flexibilidad.

# Tema 2. Almacenamiento escalable

## Casos reales donde el almacenamiento escalable es crucial

En la era digital, donde los datos son el corazón de prácticamente todas las operaciones y decisiones, el almacenamiento escalable se ha convertido en un componente esencial para los sistemas modernos.

Desde aplicaciones empresariales hasta redes sociales y desde dispositivos del internet de las cosas (IoT) hasta plataformas de comercio electrónico, la capacidad de manejar grandes cantidades de datos de manera eficiente y flexible es un factor crítico para el éxito. Vamos a explicarlo con casos reales:

### Comercio electrónico y *retail*

- ▶ Durante eventos de alto tráfico, como el Black Friday o el Cyber Monday, las plataformas de comercio electrónico experimentan picos masivos en el volumen de transacciones.
- ▶ Un sistema de almacenamiento escalable asegura que estos datos se gestionen en tiempo real, evitando colapsos que puedan afectar a las ventas y a la confianza del cliente.

### Redes sociales

- ▶ Plataformas como Instagram, TikTok o Facebook procesan millones de fotos, vídeos y publicaciones diarias.
- ▶ Sin un sistema escalable, sería imposible manejar este flujo constante de datos de forma eficiente.



# Tema 2. Almacenamiento escalable

## Dispositivos IoT

1. Sensores en vehículos autónomos, ciudades inteligentes o dispositivos de monitoreo médico generan datos continuamente.
2. Un almacenamiento escalable permite que estos sistemas funcionen sin interrupciones, procesando datos en tiempo real.

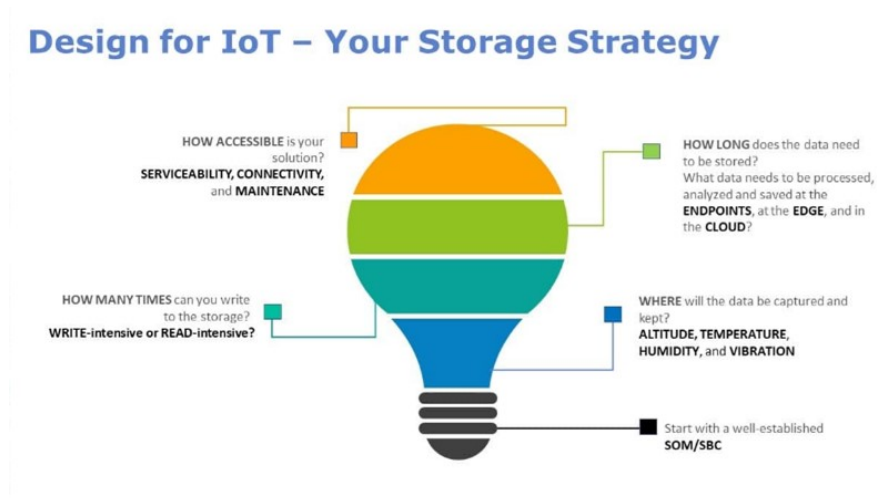


Figura 3. Imagen representativa del almacenamiento IoT. Fuente: Western Digital, 2020.

## Plataformas de *streaming*

Servicios como Netflix o Spotify deben gestionar enormes bibliotecas de contenido y adaptarse a millones de usuarios, reproduciendo vídeos o canciones simultáneamente.

# Tema 2. Almacenamiento escalable

## Impacto en la arquitectura de sistemas modernos

El almacenamiento escalable tiene un impacto significativo en los sistemas modernos, ya que permite que las organizaciones manejen de manera eficiente el rápido crecimiento de los datos, sin comprometer el rendimiento o la disponibilidad.

Este tipo de almacenamiento se adapta dinámicamente a las demandas cambiantes, lo que facilita el procesamiento de grandes volúmenes de datos en tiempo real, mejora la resiliencia ante fallos y optimiza los costes, al pagar solo por los recursos utilizados.

Además, habilita el uso de tecnologías avanzadas como *big data*, inteligencia artificial y análisis predictivo, proporcionando a las empresas la capacidad de innovar y ofrecer servicios más rápidos, inteligentes y personalizados, esenciales para competir en un entorno digital en constante evolución. Profundicemos sobre ello:

- ▶ Capacidad para procesar *big data*:
  - En sectores como la banca, el *marketing* digital o la salud, la capacidad de analizar grandes volúmenes de datos en tiempo real es vital.
  - El almacenamiento escalable permite gestionar datos distribuidos en múltiples ubicaciones, facilitando el análisis y la toma de decisiones.

## Tema 2. Almacenamiento escalable

- ▶ **Habilitación de sistemas en la nube:**
  - Soluciones como AWS, Google Cloud o Azure ofrecen almacenamiento escalable que se integra directamente con servicios en la nube, haciendo que las aplicaciones sean más accesibles y seguras.
- ▶ **Soporte para bases de Datos NoSQL:**
  - Muchas bases de datos modernas, como DynamoDB o Cassandra, están diseñadas para escalar horizontalmente, ofreciendo un rendimiento óptimo, incluso con grandes volúmenes de datos.

### Retos del almacenamiento escalable

A pesar de sus beneficios, implementar un sistema de almacenamiento escalable no está exento de desafíos. Algunos de los más comunes incluyen:

- ▶ **Complejidad de gestión.** En sistemas distribuidos, sincronizar y gestionar grandes cantidades de datos puede ser complicado.
- ▶ **Costos operativos.** Aunque reduce los costos iniciales, un uso prolongado de almacenamiento en la nube puede generar gastos elevados.
- ▶ **Latencia.** En sistemas altamente escalables, garantizar tiempos de respuesta rápidos en todas las operaciones puede ser un desafío técnico.
- ▶ **Seguridad.** La escalabilidad implica múltiples puntos de acceso, aumentando la superficie para posibles vulnerabilidades.

# Tema 2. Almacenamiento escalable

## Futuro del almacenamiento escalable

Con avances como la inteligencia artificial, el aprendizaje automático y el crecimiento continuo del IoT, la importancia del almacenamiento escalable no hará más que crecer. Las empresas que adopten esta tecnología estarán mejor preparadas para enfrentar los retos de la digitalización, ya que podrán ofrecer a sus usuarios servicios más rápidos, fiables y personalizados.

Algunos desarrollos clave en el horizonte incluyen:

- ▶ Mayor integración de inteligencia artificial para gestionar y optimizar recursos automáticamente.
- ▶ Tecnologías de almacenamiento basadas en *blockchain*, para garantizar mayor seguridad y transparencia.
- ▶ Innovaciones en *hardware*, como memorias cuánticas, que revolucionarán la forma en la que almacenamos y accedemos a los datos.

## Comparación entre almacenamiento tradicional y escalable

El almacenamiento de datos es una pieza fundamental en cualquier infraestructura tecnológica. A lo largo del tiempo, las necesidades de almacenamiento han evolucionado y, con ello, también las soluciones disponibles. Este apartado analiza las diferencias clave entre el almacenamiento tradicional y el almacenamiento escalable, destacando sus características, ventajas y limitaciones para comprender mejor cómo estas tecnologías se ajustan a los desafíos modernos.

# Tema 2. Almacenamiento escalable

## Definición básica de ambos

- ▶ Almacenamiento tradicional:
  - Modelo estático y limitado, donde los recursos de almacenamiento son predefinidos y no pueden ampliarse fácilmente sin grandes esfuerzos de reconfiguración.
  - Usualmente, se basa en servidores físicos o dispositivos de almacenamiento que requieren actualizaciones manuales para aumentar su capacidad.
- ▶ Almacenamiento escalable:
  - Modelo dinámico y flexible, diseñado para crecer o disminuir según la demanda.
  - Implementado generalmente en entornos distribuidos y a menudo habilitado por servicios en la nube.

## Comparativa: almacenamiento tradicional vs. escalable

Una tabla resumen de esta comparativa sería la siguiente:

## Tema 2. Almacenamiento escalable

Aspecto	Almacenamiento tradicional	Almacenamiento escalable
<b>Capacidad de crecimiento</b>	Limitada. Incrementar capacidad implica reemplazar <i>hardware</i> o añadir unidades manualmente.	Ilimitada. Escalabilidad vertical (añadiendo recursos al mismo servidor) u horizontal (añadiendo más servidores).
<b>Costo inicial</b>	Elevado. Requiere inversión en <i>hardware</i> y mantenimiento físico.	Bajo. Modelo de pago por uso en la nube o con capacidad ajustable, según necesidades.
<b>Flexibilidad</b>	Baja. Es difícil y lento adaptarse a cambios repentinos en la carga de trabajo.	Alta. Puede ajustarse automáticamente para manejar picos o reducciones en la demanda.
<b>Rendimiento</b>	Estático. El rendimiento depende de la configuración inicial y no escala bien con el crecimiento.	Dinámico. Optimizado para manejar cargas variables manteniendo la eficiencia.
<b>Mantenimiento</b>	Complejo. Requiere intervención manual y costosa, con actualizaciones frecuentes de <i>hardware</i> .	Sencillo. Servicios en la nube suelen incluir mantenimiento automatizado y actualizaciones transparentes.
<b>Ubicación</b>	Local. Los datos suelen estar en servidores físicos controlados por la empresa.	Distribuida. Los datos se alojan en múltiples ubicaciones (centros de datos en la nube).
<b>Tolerancia a fallos</b>	Limitada. Generalmente depende de sistemas de respaldo local.	Alta. Servicios escalables suelen incluir redundancia geográfica y recuperación ante desastres integrada.
<b>Tiempo de implementación</b>	Lento. La configuración inicial y las actualizaciones pueden tardar semanas o meses.	Rápido. La configuración puede completarse en minutos con servicios en la nube.
<b>Escenarios de uso</b>	Pequeñas empresas con necesidades de almacenamiento estables o aplicaciones <i>legacy</i> .	Empresas modernas, <i>big data</i> , analítica avanzada, IoT y cualquier aplicación con demanda de datos variable y creciente.

Tabla 2. Comparativa del almacenamiento tradicional vs. almacenamiento escalable. Fuente: elaboración propia.

# Tema 2. Almacenamiento escalable

## Ventajas del almacenamiento tradicional

Aunque el almacenamiento tradicional ha sido reemplazado en muchos contextos, sigue ofreciendo ventajas en escenarios específicos:

1. **Control completo:** ideal para organizaciones que requieren control total sobre sus datos, como aquellas con estrictas regulaciones de privacidad.
2. **Bajo costo operativo a largo plazo:** una vez implementado, no genera gastos recurrentes significativos, como sí lo hace el almacenamiento en la nube.
3. **Uso en ambientes *legacy*:** sigue siendo relevante para sistemas antiguos que no son compatibles con soluciones modernas.

## Ventajas del almacenamiento escalable

1. **Capacidad de adaptación:** perfecto para empresas que experimentan fluctuaciones en la demanda, como plataformas de *streaming* o comercio electrónico.
2. **Acceso global:** con el almacenamiento escalable en la nube, los datos están disponibles desde cualquier lugar, mejorando la colaboración y el acceso remoto.
3. **Reducción del riesgo:** la redundancia y la alta disponibilidad garantizan que los datos estén protegidos contra fallos y desastres.
4. **Integración con nuevas tecnologías:** es ideal para soportar aplicaciones modernas basadas en inteligencia artificial, análisis de *big data* y aprendizaje automático.

## Tema 2. Almacenamiento escalable

### Ejemplos comparativos

- ▶ **Caso de almacenamiento tradicional.** Una pequeña empresa utiliza un servidor local para almacenar los datos de sus clientes. Aunque este sistema funciona bien en sus primeros años, enfrenta problemas al intentar gestionar un aumento del 50 % en la cantidad de clientes. La empresa necesita adquirir y configurar nuevos discos duros, lo que implica tiempo, costo y complejidad adicional.
- ▶ **Caso de almacenamiento escalable.** Una empresa emergente de comercio electrónico utiliza Amazon S3 (almacenamiento escalable). A medida que aumentan las ventas durante una campaña de *marketing*, el sistema ajusta automáticamente la capacidad de almacenamiento para gestionar el incremento en los datos, asegurando una experiencia fluida para los usuarios. Además, al finalizar la campaña, la capacidad se reduce para minimizar costos.
- ▶ Limitaciones del almacenamiento tradicional:
  - Falta de flexibilidad: es incapaz de ajustarse dinámicamente a cambios en la carga de trabajo.
  - Altos costos de expansión: ampliar la capacidad requiere grandes inversiones en *hardware* y espacio físico.
  - Riesgo de fallos: en caso de desastres, la recuperación es lenta y depende de copias de seguridad físicas.



# Tema 2. Almacenamiento escalable

- ▶ Limitaciones del almacenamiento escalable:
  - Costo recurrente: aunque el costo inicial es bajo, el uso prolongado de almacenamiento en la nube puede volverse costoso.
  - Dependencia de proveedores: las empresas deben confiar en proveedores externos para la seguridad y disponibilidad de los datos.
  - Latencia en operaciones: en algunos casos, los sistemas distribuidos pueden experimentar retardos en el acceso a los datos.

## Ventajas del almacenamiento en la nube

El almacenamiento en la nube ha transformado la forma en la que las empresas y los usuarios gestionan sus datos. A diferencia de los métodos tradicionales, este modelo ofrece una flexibilidad y eficiencia incomparables al aprovechar la infraestructura de centros de datos distribuidos, accesibles a través de Internet.

En este apartado, exploraremos las principales ventajas del almacenamiento en la nube, enfocándonos en su impacto práctico y su valor en un mundo digital en constante cambio.

### 1. Escalabilidad dinámica

Una de las características más destacadas del almacenamiento en la nube es su capacidad para adaptarse a las **necesidades** de los usuarios en **tiempo real**.

- ▶ Crecimiento sin límites: los usuarios pueden aumentar o disminuir su capacidad de almacenamiento según la demanda, sin necesidad de comprar *hardware* adicional.
- ▶ Ejemplo práctico: durante eventos como campañas de *marketing* masivas o lanzamientos de productos, las empresas pueden incrementar temporalmente su capacidad de almacenamiento para gestionar picos de tráfico.

# Tema 2. Almacenamiento escalable

## 2. Acceso global y colaboración

El almacenamiento en la nube permite acceder a los datos desde cualquier lugar del mundo con conexión a Internet.

- ▶ Trabajo remoto optimizado: equipos distribuidos pueden colaborar en tiempo real sin barreras geográficas.
- ▶ Ejemplo práctico: un equipo internacional de diseño utiliza una herramienta como Google Drive para trabajar en un proyecto compartido, asegurando que todos los miembros tienen acceso a la última versión de los archivos.

## 3. Reducción de costos iniciales

El almacenamiento en la nube elimina la necesidad de invertir en costosos servidores físicos y su mantenimiento.

- ▶ Modelo de pago por uso: los usuarios solo pagan por la capacidad y los servicios que utilizan, reduciendo gastos innecesarios.
- ▶ Ejemplo práctico: una *startup* puede comenzar con una pequeña cantidad de almacenamiento en Amazon S3 y ampliarlo conforme crece su base de datos, evitando gastos iniciales elevados.

## Tema 2. Almacenamiento escalable

### 4. Seguridad avanzada

Los proveedores de almacenamiento en la nube implementan medidas de seguridad de nivel empresarial para proteger los datos.

- ▶ Protección robusta: incluye cifrado de datos, autenticación multifactorial y monitoreo continuo para prevenir brechas.
- ▶ Ejemplo práctico: una empresa de comercio electrónico asegura la información confidencial de sus clientes con cifrado de extremo a extremo, en un servicio como AWS o Azure.

### 5. Resiliencia y recuperación ante desastres

El almacenamiento en la nube está diseñado con redundancia y copias de seguridad distribuidas en múltiples ubicaciones.

- ▶ Tolerancia a fallos: incluso si un centro de datos experimenta problemas, los datos permanecen accesibles desde ubicaciones alternativas.
- ▶ Ejemplo práctico: una organización que utiliza Google Cloud recupera rápidamente sus datos tras un fallo en el servidor principal, asegurando continuidad en sus operaciones.

### 6. Integración con tecnologías modernas

La nube facilita el acceso a herramientas y servicios avanzados como inteligencia artificial, análisis de datos y *big data*.

- ▶ Entorno adaptable: los datos almacenados en la nube pueden ser utilizados directamente por aplicaciones modernas sin necesidad de transferencias adicionales.
- ▶ Ejemplo práctico: una empresa que utiliza almacenamiento en la nube puede integrar sus datos con servicios de análisis como Amazon Redshift, para generar informes en tiempo real.

## Tema 2. Almacenamiento escalable

### 7. Escalabilidad financiera y operativa

La nube permite a las empresas crecer sin las limitaciones operativas de los modelos tradicionales.

- ▶ Flexibilidad financiera: las empresas no necesitan comprometer grandes sumas en infraestructura física.
- ▶ Ejemplo práctico: una plataforma de *streaming* ajusta automáticamente la capacidad de su almacenamiento en función del número de usuarios activos.

### 8. Innovación constante

Los principales proveedores de almacenamiento en la nube actualizan constantemente sus plataformas con nuevas características y mejoras.

- ▶ Acceso a lo último: los usuarios pueden aprovechar tecnologías de última generación sin costos adicionales.
- ▶ Ejemplo práctico: una empresa accede a herramientas de *machine learning* en la nube, para optimizar sus procesos de negocio.

### Principales servicios de almacenamiento de AWS

Amazon Web Services (AWS) es una de las plataformas en la nube más utilizadas a nivel mundial, conocida por ofrecer un conjunto de servicios de almacenamiento diseñados para satisfacer las necesidades de diversos casos de uso, desde el almacenamiento de archivos básicos hasta soluciones avanzadas de *big data* y *machine learning*.

A continuación, se detallan los principales servicios de almacenamiento que AWS ofrece, explicando sus características, casos de uso y cómo se integran en un entorno escalable.

# Tema 2. Almacenamiento escalable

## 1. Amazon S3 (*simple storage service*)

- ▶ Descripción: Amazon S3 es un servicio de almacenamiento de objetos diseñado para almacenar y recuperar datos desde cualquier ubicación a través de Internet. Ofrece durabilidad y escalabilidad prácticamente ilimitadas.
- ▶ Características principales:
  - Durabilidad: 99,999999999 % (once nueves), asegurando la protección de los datos.
  - Escalabilidad automática: se adapta a la cantidad de datos almacenados sin intervención manual.
  - Clases de almacenamiento: ofrece opciones como S3 Standard, S3 Intelligent-Tiering, S3 Glacier (almacenamiento a largo plazo), entre otras.
  - Seguridad: cifrado de datos en reposo y en tránsito y controles de acceso basados en políticas.
- ▶ Casos de uso:
  - Almacenamiento de contenido estático para sitios web.
  - Copias de seguridad y recuperación de datos.
  - Almacenamiento de datos para análisis y *big data*.
- ▶ Ejemplo práctico: una empresa de comercio electrónico utiliza S3 para almacenar imágenes de productos, documentos PDF descargables y datos de usuarios de forma escalable y segura.

# Tema 2. Almacenamiento escalable

## 2. Amazon EBS (*elastic block store*)

- ▶ Descripción: Amazon EBS es un servicio de almacenamiento en bloque diseñado para ser usado con instancias EC2. Proporciona almacenamiento de baja latencia para aplicaciones que requieren acceso rápido y persistente.
- ▶ Características principales:
  - Alto rendimiento: diseñado para aplicaciones de base de datos, *big data* y sistemas transaccionales.
  - Persistencia de datos: los datos permanecen incluso si la instancia EC2 asociada se detiene o termina.
  - Opciones de volúmenes: general Purpose SSD, Provisioned IOPS SSD y HDD, para diferentes necesidades de rendimiento.
- ▶ Casos de uso:
  - Bases de datos relacionales (por ejemplo, MySQL o PostgreSQL).
  - Almacenamiento de datos críticos para aplicaciones empresariales.
  - Sistemas de archivos distribuidos.
- ▶ Ejemplo práctico: una empresa de desarrollo utiliza EBS para almacenar los datos transaccionales de su aplicación SaaS, garantizando velocidad y disponibilidad constante.