**¿Qué es Big Data?**

El Big Data engloba conjuntos extremadamente grandes y diversos de datos estructurados, sin estructurar y semiestructurados que siguen creciendo exponencialmente a lo largo del tiempo. Estos conjuntos de datos son tan enormes y complejos en volumen, velocidad y variedad que los sistemas de gestión de datos tradicionales no pueden almacenarlos, procesarlos ni analizarlos.

La cantidad y la disponibilidad de datos están creciendo rápidamente gracias a los avances en tecnología digital, como la conectividad, la movilidad, el Internet de las cosas (IoT) y la inteligencia artificial (IA). A medida que los datos siguen creciendo y proliferando, están surgiendo nuevas herramientas de Big Data que ayudan a las empresas a recoger, procesar y analizar datos a la velocidad necesaria para obtener el máximo valor de ellos.

El Big Data describe conjuntos de datos grandes y diversos que tienen un gran volumen y que también crecen rápidamente de tamaño con el tiempo. El Big Data se utiliza en el aprendizaje automático, la modelización predictiva y otras analíticas avanzadas para solucionar problemas empresariales y tomar decisiones fundamentadas.

Sigue leyendo para conocer la definición del Big Data, conocer algunas de las ventajas de las soluciones de Big Data, los desafíos más habituales del Big Data y descubrir cómo ayuda Google Cloud a las empresas a [crear sus nubes de datos](https://cloud.google.com/data-cloud) para que les saquen más partido.

Empezar gratis

Ejemplos de Big Data

Los datos pueden ser el recurso más valioso de las empresas. Usar Big Data para descubrir información valiosa te puede ayudar a comprender las áreas que afectan a tu empresa, como las condiciones del mercado, los comportamientos de compra de los clientes o tus procesos empresariales.

A continuación tienes algunos ejemplos de Big Data que están ayudando a transformar organizaciones de todos los sectores:

* Hacer un seguimiento del comportamiento y los hábitos de compra de los consumidores para ofrecer [recomendaciones de productos muy personalizadas](https://cloud.google.com/blog/products/ai-machine-learning/ikea-uses-google-cloud-recommendations-ai) para clientes particulares
* Monitorizar los patrones de pago y analizarlos comparándolos con el historial de actividad de los clientes para [detectar fraudes en tiempo real](https://cloud.google.com/blog/products/databases/how-ravelin-scales-fraud-detection-with-bigtable).
* Combinar datos e información de todas las fases del recorrido de envío de un pedido con estadísticas del tráfico hiperlocal para [ayudar a los operadores de flotas a optimizar la entrega en el último kilómetro](https://cloud.google.com/blog/products/maps-platform/introducing-last-mile-fleet-solution-maximize-what-your-fleet-can-do-start-finish).
* Usar tecnologías basadas en IA, como el [procesamiento del lenguaje natural para analizar datos médicos no estructurados](https://cloud.google.com/blog/topics/healthcare-life-sciences/natural-language-processing-nlp-healthcare-insights-clinical-research-data-cloud) (como informes de investigación, notas clínicas y resultados de laboratorios), para obtener nuevas estadísticas con las que mejorar el desarrollo de tratamientos y la atención al paciente.
* Usar datos de imágenes de cámaras y sensores, así como datos GPS, para [detectar baches y mejorar el mantenimiento de las carreteras en las ciudades](https://cloud.google.com/blog/products/ai-machine-learning/video-intelligence-machine-learning-improves-pothole-detection).
* Analiza conjuntos de datos públicos de imágenes por satélite y conjuntos de datos geoespaciales para visualizar, monitorizar, medir y predecir [el impacto social y medioambiental de las operaciones de la cadena de suministro](https://cloud.google.com/blog/topics/consumer-packaged-goods/sustainable-sourcing-for-consumer-brands-with-google-cloud)

Estas son solo algunas de las formas en que las empresas están usando Big Data para basarse más en los datos y adaptarse mejor a las necesidades y las expectativas de sus clientes y del mundo que les rodea.

Las Vs del Big Data

Las definiciones de Big Data pueden variar ligeramente, pero siempre se describirán en términos de volumen, velocidad y variedad. Estas características del Big Data suelen denominarse "las tres uves del Big Data" y Gartner las definió por primera vez en el 2001.

**Volumen**

Como su nombre indica, la característica más común asociada al Big Data es su gran volumen. Esto describe la enorme cantidad de datos que se pueden recopilar y que se generan a partir de diversas fuentes y dispositivos de forma continua.

**Velocidad**

La velocidad de Big Data hace referencia a la velocidad a la que se generan los datos. Hoy en día, los datos se suelen generar en tiempo real o casi en tiempo real y, por lo tanto, también deben procesarse, consultarse y analizarse al mismo ritmo para tener un impacto significativo.

**Variedad**

Los datos son heterogéneos, es decir, pueden proceder de muchas fuentes diferentes y pueden estar estructurados, no estructurados o semiestructurados. Los datos estructurados más tradicionales (como los datos de hojas de cálculo o bases de datos relacionales) se complementan ahora con texto, imágenes, audio y archivos de vídeo no estructurados, o con formatos semiestructurados como los datos de sensores que no se pueden organizar en un esquema de datos fijo.

Además de las tres Vs originales, otras tres que a menudo se mencionan en relación con el aprovechamiento del potencial del Big Data: la **veracidad**, la **variabilidad** y el **valor**.

* **Veracity** el Big Data puede resultar desordenado, con mucho ruido y es propenso a errores, lo que dificulta el control de la calidad y la precisión de los datos. Los conjuntos de datos grandes pueden resultar difíciles de manejar y confusos, mientras que los conjuntos más pequeños pueden presentar una imagen incompleta. Cuanto mayor sea la veracidad de los datos, más fiable será.
* **Variabilidad:** el significado de los datos recogidos cambia constantemente, lo que puede provocar incoherencias a lo largo del tiempo. Estos cambios incluyen no solo los cambios de contexto y de interpretación, sino también en los métodos de recogida de datos basados en la información que las empresas quieren recoger y analizar.
* **Valor:**es fundamental determinar el valor empresarial de los datos que recoges. Los Big Data deben contener los datos adecuados y luego analizarse de manera eficaz para extraer información valiosa que ayude a tomar decisiones fundamentadas.

¿Cómo funciona el Big Data?

El concepto principal del Big Data es que cuanta más visibilidad tengas sobre cualquier aspecto, más eficazmente podrás obtener información valiosa para tomar mejores decisiones, descubrir oportunidades de crecimiento y mejorar tu modelo de negocio.

Para que el Big Data funcione, hay que hacer tres cosas:

* **Integración:**el Big Data recoge terabytes (e incluso petabytes) de datos en bruto de muchas fuentes que deben recibirse, procesarse y transformarse al formato que necesitan los usuarios empresariales y analistas para empezar a analizarlos.
* **Gestión:**el Big Data necesita mucho almacenamiento, ya sea en la nube, on‐premise o de ambas formas. Los datos también deben almacenarse en el formato necesario. También se tiene que procesar y estar disponible en tiempo real. Cada vez más, las empresas recurren a soluciones en la nube para aprovechar la computación y la escalabilidad ilimitadas.
* **Análisis:**el último paso es analizar Big Data y actuar en consecuencia. De lo contrario, la inversión no merecerá la pena. Además de examinar los datos en sí, también es fundamental comunicar y compartir información valiosa con toda la empresa de un modo que todo el mundo la pueda entender. Esto incluye el uso de herramientas para crear visualizaciones de datos, como gráficos y paneles.

Ventajas del Big Data

Mejoras en la toma de decisiones

El Big Data es el elemento clave para que una empresa se base en los datos. Cuando puedes gestionar y analizar el Big Data, puedes descubrir patrones y obtener información valiosa que te permita mejorar y tomar mejores decisiones operativas y estratégicas.

Mayor agilidad e innovación

Big Data te permite recoger y procesar datos en tiempo real y analizarlos para adaptarte rápidamente y obtener una ventaja competitiva. Esta información valiosa puede guiar y acelerar la planificación, la producción y el lanzamiento de nuevos productos, funciones y actualizaciones.

Experiencia de cliente mejorada

Al combinar y analizar fuentes de datos estructurados con otras sin estructurar, obtendrás información más útil para entender y personalizar las fuentes de datos de los consumidores, así como para optimizar tu experiencia y así satisfacer mejor las necesidades y expectativas de los consumidores.

Inteligencia continua

El Big Data te permite integrar flujos de datos automatizados en tiempo real con analíticas de datos avanzadas para recoger datos de forma continua, conseguir nuevas estadísticas y descubrir nuevas oportunidades de crecimiento y valor.

Operaciones más eficientes

El uso de herramientas y funciones de análisis de Big Data te permite procesar los datos más rápido y generar información valiosa que puede ayudarte a determinar en qué áreas puedes reducir costes, ahorrar tiempo y aumentar la eficiencia general.

Mejor gestión de riesgos

El análisis de grandes cantidades de datos ayuda a las empresas a evaluar mejor los riesgos, lo que facilita la identificación y supervisión de todas las posibles amenazas, así como la generación de informes con información valiosa que les lleve a estrategias de control y mitigación más sólidas.

Retos de implementar analíticas de Big Data

Aunque el Big Data ofrece muchas ventajas, plantea algunos retos que las empresas deben estar preparadas para abordar a la hora de recoger y gestionar grandes cantidades de datos, así como actuar en consecuencia.

Estos son algunos de los retos más comunes relacionados con el Big Data:

* **Falta de talento y destrezas en materia de datos.**Los científicos, los analistas e ingenieros de datos tienen una escasez de personal y son algunos de los profesionales más codiciados y muy demandados del sector informático. La falta de conocimientos y experiencia en Big Data con herramientas avanzadas de datos es uno de los principales obstáculos para obtener valor de los entornos de Big Data.
* **Acelera el crecimiento de los datos.**Por naturaleza, el Big Data cambia y aumenta constantemente. Sin una infraestructura sólida que pueda gestionar tus necesidades de procesamiento, almacenamiento, red y seguridad, puede resultar extremadamente difícil de gestionar.
* **Problemas con la calidad de los datos.**La calidad de los datos influye directamente en la calidad de la toma de decisiones, las analíticas de datos y las estrategias de planificación. Los datos sin procesar son desordenados y pueden ser difíciles de seleccionar. El Big Data no garantiza obtener resultados a menos que los datos sean precisos, relevantes y estén organizados adecuadamente para analizarse. Esto puede ralentizar la generación de informes, pero si no se corrige, puedes acabar obteniendo resultados engañosos y estadísticas poco útiles.
* **Infracciones de cumplimiento.**El Big Data contiene una gran cantidad de datos sensibles e información sensible, por lo que es una tarea complicada para asegurar que el tratamiento y el almacenamiento de los datos de forma continua cumplan los requisitos normativos y de privacidad de los datos, como la localización y las leyes de residencia de datos.
* **Complejidad de la integración.**La mayoría de las empresas trabajan con datos aislados en varios sistemas y aplicaciones. Integrar fuentes de datos dispares y poner los datos a disposición de los usuarios empresariales es un proceso complejo, pero fundamental si esperas obtener valor a partir de tu Big Data.
* **Problemas de seguridad**El Big Data contiene información valiosa sobre las empresas y los clientes, por lo que el Big Data almacena objetivos de gran valor para los atacantes. Dado que estos conjuntos de datos son variados y complejos, puede resultar más difícil implementar estrategias y políticas integrales para protegerlos.

¿Cuál es el rendimiento de las empresas basadas en datos?

Algunas organizaciones se muestran recelosas de apostar de lleno en el Big Data debido al tiempo, el esfuerzo y el compromiso que requieren para aprovecharlo con éxito. En concreto, las empresas tienen dificultades para rediseñar los procesos establecidos y para facilitar el cambio cultural necesario para que los datos sean la piedra angular de cada decisión.

Pero vale la pena para convertirse en una empresa basada en datos. Estudios recientes demuestran que:

* El [58 % de las empresas](https://www.ciodive.com/news/data-driven-companies-revenue-coronavirus-covid19/578159/) que toman decisiones basadas en datos tienen más probabilidades de superar sus objetivos de ingresos que las que no lo hacen
* Las organizaciones que cuentan con funciones empresariales basadas en estadísticas avanzadas tienen [2,8 veces más probabilidades](https://www.forrester.com/blogs/move-from-insight-to-impact-data-strategy-insights-2020/) de registrar un crecimiento de dos dígitos interanuales.
* Las empresas basadas en datos generan, de media, [más del 30 %](https://www.accenture.com/nl-en/blogs/insights/data-driven-enterprise) de crecimiento al año.

Las empresas que tomen medidas ahora y progresen de forma significativa para implementar Big Data serán las ganadoras en el futuro.

Estrategias y soluciones de Big Data

Para desarrollar una estrategia de datos sólida, primero hay que entender lo que se quiere conseguir, identificar casos prácticos concretos y qué datos se pueden usar en la actualidad. También tendrás que evaluar qué datos adicionales podrían necesitar para cumplir tus objetivos de negocio y los nuevos sistemas o herramientas que necesitarás para alcanzarlos.

A diferencia de las soluciones tradicionales de gestión de datos, las tecnologías y herramientas de Big Data están diseñadas para ayudarte a lidiar con conjuntos de datos grandes y complejos para extraer valor de ellos. Las herramientas de Big Data pueden ayudarte con el volumen de datos que se recogen, la velocidad a la que están disponibles para analizarlos y con su complejidad o variedad.

Por ejemplo, los [data lakes](https://cloud.google.com/learn/what-is-a-data-lake) ingieren, procesan y almacenan datos estructurados, sin estructurar y semiestructurados a cualquier escala en su formato nativo. Los lagos de datos son la base para ejecutar diferentes tipos de analíticas inteligentes, como visualizaciones, analíticas en tiempo real y [aprendizaje automático](https://cloud.google.com/learn/what-is-machine-learning).

Es importante tener en cuenta que, en lo que respecta al Big Data, no hay una estrategia que se adapte a todos los casos. Puede que lo que funciona para una empresa no sea la estrategia adecuada para las necesidades específicas de tu organización.

A continuación te presentamos cuatro conceptos clave que han aprendido nuestros clientes de Google Cloud sobre cómo dar forma a una estrategia eficaz de Big Data:

**Mostrar**

Hoy en día, las organizaciones necesitan libertad para construir lo que quieren con las herramientas y las soluciones que desean. A medida que crecen las fuentes de datos y surgen nuevas innovaciones tecnológicas, la realidad del Big Data está constituida por varias interfaces, pilas de tecnología de código abierto y nubes. Los entornos de Big Data tendrán que diseñarse de manera que sean abiertos y adaptables para que las empresas puedan crear las soluciones y obtener los datos que necesitan para obtener buenos resultados.

**Inteligente**

Big Data requiere funciones de datos que les permitan aprovechar las analíticas inteligentes y las tecnologías de IA y aprendizaje automático para ahorrar tiempo y esfuerzo, a la hora de ofrecer información valiosa que mejore las decisiones empresariales y gestionar tu infraestructura general de Big Data. Por ejemplo, deberías plantearte automatizar procesos o habilitar las analíticas de autoservicio para que tus empleados puedan trabajar con datos por su cuenta sin la ayuda de otros equipos.

**Flexibilidad**

Las analíticas de Big Data deben fomentar la innovación, no dificultarla. Para ello, es necesario crear una base de datos que ofrezca acceso bajo demanda a los recursos de computación y almacenamiento, y que unifique los datos para que se pueda descubrir y acceder a ellos fácilmente. También es importante poder elegir tecnologías y soluciones que puedan combinarse y utilizarse fácilmente para crear los conjuntos de herramientas de datos perfectos que se adapten a la carga de trabajo y al caso práctico.

**De confianza**

Para que el Big Data sea útil, es necesario que sea fiable. Por tanto, es fundamental generar confianza en tus datos; es decir, que sean precisos, relevantes y estén protegidos. Independientemente de la procedencia de los datos, deben ser seguros de forma predeterminada, y tu estrategia también deberá tener en cuenta qué funciones de seguridad serán necesarias para asegurar el cumplimiento, la redundancia y la fiabilidad