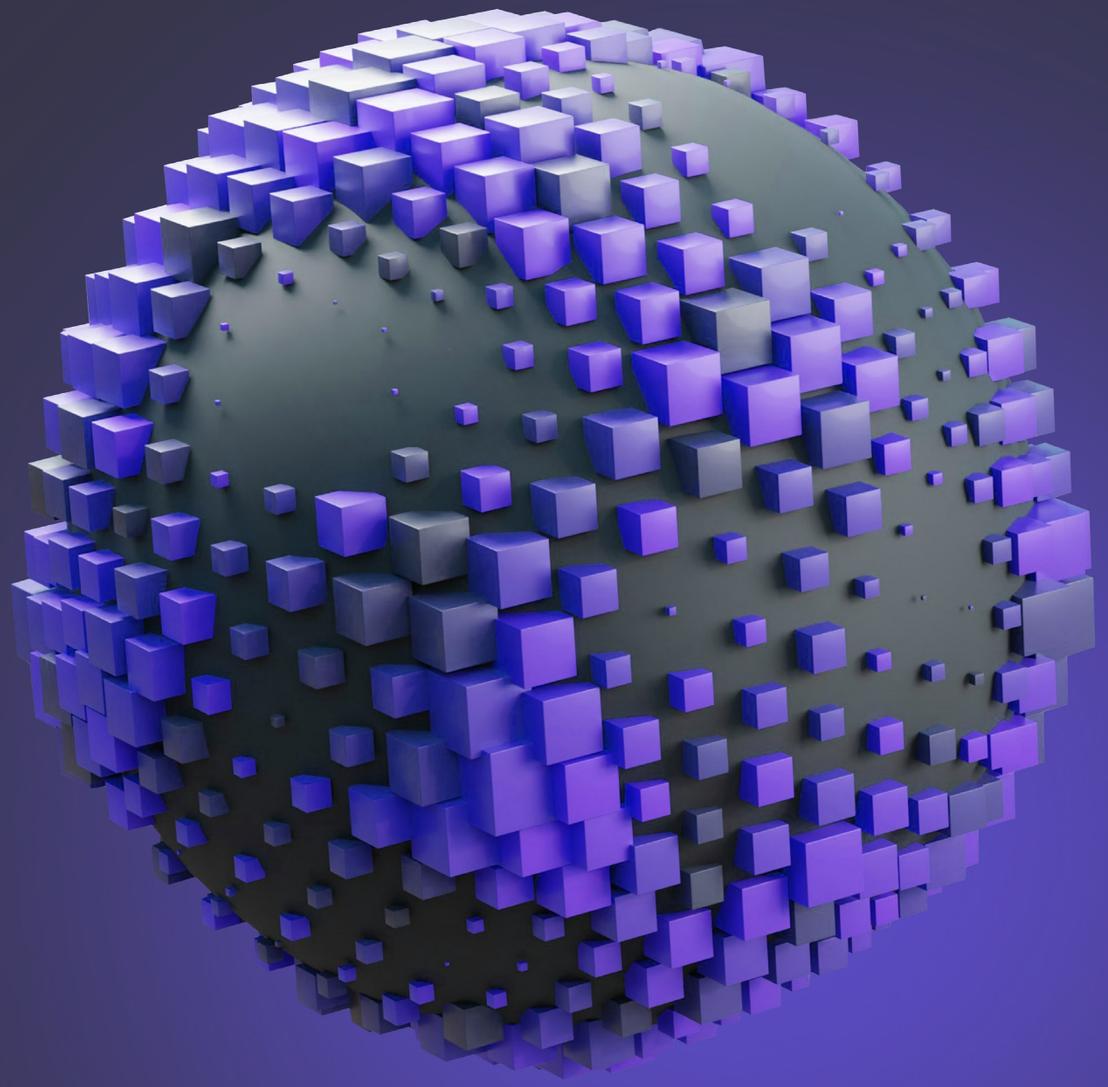
Softtek®

Data: what’s new?



Índice

04 - 09	**10 - 17**	**18 - 22**	**23 - 26**	**27 - 31**	**32 - 36**
Data democratization	Data Monetization & DaaS	Cloud Modernization/ New Cloud Data Ecosystems	Data Fabric	TinyML/ Small y Wide Data	AI Ubicua/ Edge AI
37 - 41	**42 - 46**	**47 - 51**	**52 - 56**	**57 - 58**	
Data Masking	DataOPs	La revolución Web3 y el almacenamiento descentralizado	Quantum Technologies	Conclusiones	

Prólogo

El futuro del panorama empresarial está marcado por la capacidad de las organizaciones para desbloquear el potencial de los datos. En un momento en que la recabación de datos ya es un problema superado, la transformación en información de estos parece ser un reto esquivo que solo las empresas más avanzadas en la aceleración de la transformación digital pueden alcanzar.

El escenario digital evoluciona para dar respuesta a este desafío, la innovación tecnológica está redefiniendo el rumbo de las nuevas estrategias que las organizaciones deben seguir para liderar este cambio de paradigma. **Los datos se han establecido como un aspecto clave en el ámbito empresarial debido a las posibilidades casi ilimitadas para generar valor de negocio, solo restringido por la privacidad de los datos.** En este punto, la protección cibernética se posiciona como un tema prioritario dentro del debate acerca de cómo aprovechar los avances tecnológicos y los datos en el desarrollo empresarial dentro del cumplimiento normativo y garantizando la seguridad.

Entonces, ¿cómo pueden las organizaciones transformar los datos en un activo estratégico que impulse la innovación? ¿Cómo pueden optimizar las operaciones y crear nuevas oportunidades a la vez que protegen la integridad y la confidencialidad de los datos? Para dar solución a este debate, los líderes y las organizaciones han de ir de la mano de la tecnología y tener una mentalidad flexible y resiliente dispuesta a abrazar el cambio.

Con esta investigación, buscamos ofrecer una **visión completa de las tendencias tecnológicas que están definiendo el curso del Data y que marcarán la evolución de los próximos años.** Adelantarse al futuro definirá el nuevo panorama empresarial.

01. Data democratization

En la era digital, los datos se tornan un elemento estratégico fundamental para las organizaciones permitiendo la creación y la evolución de la toma de decisiones, operaciones y acciones. Para garantizar el éxito del nuevo contexto transformador, las empresas no deben quedarse en lo puramente tecnológico, es esencial un cambio de mentalidad de los líderes para acelerar y garantizar la adopción del potencial del dato.

Para lograr una verdadera democratización de los datos, **los actores implicados deben obtener habilidades de alfabetización para poder verificar la exactitud de estos, garantizar su seguridad y su adecuado procesamiento y utilización.** Es imperativo que el formato seleccionado para interactuar con los datos permita una óptima usabilidad y proporcione una experiencia ágil al usuario. A menudo, la democratización de los datos se combina con la transparencia de los mismos, refiriéndose a procesos que garantizan la trazabilidad de dicha información proporcionando su ubicación o su origen.

La democratización de los datos está revolucionando la gestión y los procesos empresariales. Antiguamente, la gobernanza de los datos estaba reducida al equipo de IT y apenas se explotaba dicha información. El cambio que estamos viviendo no solo afecta al acceso, sino que se trata de **empoderar a todos los miembros de una organización y dotarlos de la capacidad y las herramientas para aprovechar este potencial fomentando una cultura ágil e informada.**

Evolución

Para convertirse en una organización Data Centric, **los datos deben trascender y penetrar en toda la estructura organizativa**; de esta forma, el dato adquiere la categoría de activo y puede ser explotado por todas las personas que componen la compañía mejorando los procesos y aumentando las interacciones. Para lograrlo, las empresas deben prever un modelo operativo basado en los datos y diseñar una base tecnológica que respalde toda la operación.

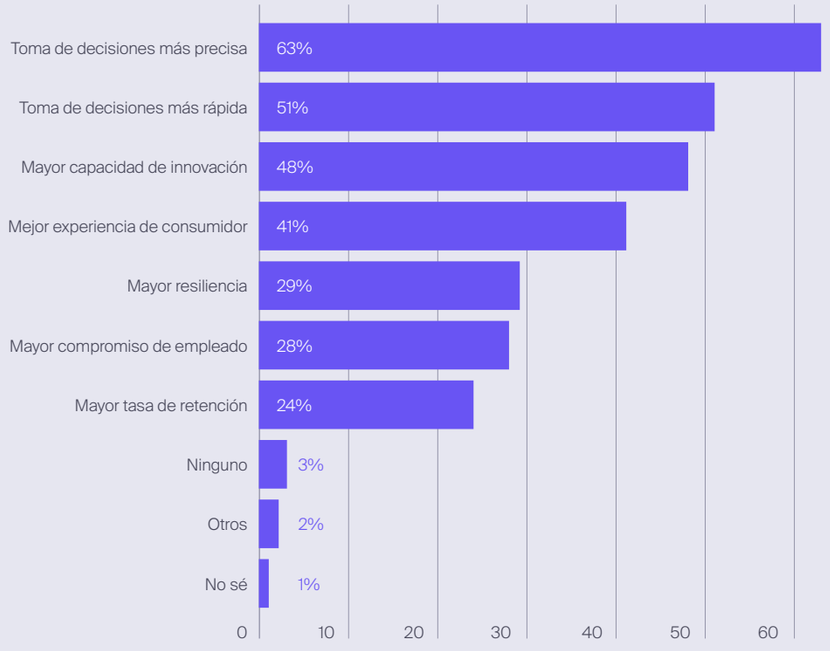
En este contexto en que **se evoluciona de datos aislados a ecosistemas de datos abiertos y accesibles**, las capacidades relacionadas con la gestión y el cambio cultural se vuelven críticas; por ello, las organizaciones están aumentando significativamente su inversión en competencias tecnológicas y en formación del talento para adaptarse al contexto digital.

La transformación trae consigo numerosos beneficios, el cambio progresivo hacia una nueva estrategia integral de datos supone la **reducción de sobrecostes** de infraestructura informática al adoptar plataformas modernas y desmantelar los antiguos sistemas. Asimismo, permite la aplicación de nuevas tecnologías inteligentes y procesos de automatización que maximizan el potencial de los resultados, obteniendo estrategias más punteras basadas en datos.

Investigaciones recientes arrojan información relevadora sobre el estado de la democratización de los datos: se estima que el 70% de los líderes empresariales que participan en iniciativas de mejora de sus habilidades en materia de datos experimentan una mejora de más del 60% en la rapidez y eficacia de su toma de decisiones, lo cual a su vez repercute en una mejora de la experiencia de cliente y las capacidades de innovación.

Valor de alfabetización de datos para la mejora de la fuerza laboral

Dimensiones en las que los empleados capacitados en datos superan a los que no están



*Según una encuesta a líderes de organizaciones a nivel global. Fuente: Softtek en base a datos de DataCamp

Background tecnológico

Automated Data Discovery and Categorization

- **Algoritmos de aprendizaje automático:** fragmentos de código que ayudan a los usuarios a explorar y analizar conjuntos de datos complejos detectando patrones. Ayudan a clasificar y etiquetar datos automáticamente facilitando su búsqueda, comprensión y procesamiento.

- **Capacidades de búsqueda semántica:** las mejoras en este campo, gracias a la rápida evolución de los modelos de procesamiento de lenguaje natural de aprendizaje automático (LLM), permiten al usuario realizar consultas en lenguaje natural, lo cual genera un proceso más intuitivo y ágil.

Robust Data Governance and Quality Frameworks

- **Catálogos de datos con metadatos mejorados:** proporcionan un contexto enriquecido, gracias al inventario de activos de datos, que permite a los usuarios encontrar, acceder e investigar la información de forma rápida y sencilla. Asimismo, incluye la trazabilidad de la información y anotaciones generadas por el usuario para ayudar a otros a comprender y confiar en los datos.

- **Comprobaciones automatizadas de calidad de los datos:** los sistemas automatizados garantizan que los usuarios trabajen siempre con datos de alta calidad.

Self-Service Data Preparation Tools

- **Interfaces intuitivas de gestión de datos:** permiten a los usuarios optimizar su interacción con los datos, desde buscar hasta combinar datos sin necesidad de codificación.

- **Transformación de datos mejorada por inteligencia artificial:** permite integrar una capa de inteligencia al tratamiento de datos basado en el comportamiento del usuario, lo que agiliza el proceso de preparación de datos.

Desafíos

Centralización de los datos

- **Infraestructura adecuada para la interacción de datos:** gracias a las nuevas tecnologías se están generando plataformas que permiten una interacción natural y ágil con los datos sin necesidad de formación técnica específica.

- **Calidad y coherencia de los datos**: el gobierno de datos es un enfoque riguroso, preciso y esencial para gestionar los datos durante su ciclo de vida, desde la adquisición hasta la eliminación, pasando por el uso para garantizar la exactitud, integridad, coherencia y relevancia de los datos.

Equilibrio entre accesibilidad y seguridad

- **Control de acceso basado en atributos (ABAC):** es un sistema de autorización que define el acceso en función de los atributos asociados a las entidades de seguridad, los recursos y el entorno de una solicitud de acceso, proporciona un control detallado en la administración de acceso en función de la asignación de roles.

- **Privacidad diferencial**: responde a la necesidad de equilibrio entre la explotación legítima de la información y el respeto a los derechos individuales. Esta estrategia está encaminada a preservar la utilidad de los datos al tiempo que se respeta la privacidad de las personas agregando "ruido" a los datos o consultas.

Garantizar una comprensión e interpretación coherentes

- **Programas de alfabetización de datos:** permiten dotar al talento de la capacidad integral de comprender, construir y comunicar los datos en su contexto, los métodos y técnicas analíticas aplicadas y la capacidad de producir una aplicación de valor.

- **Business Intelligence y analítica integrada:** es el conjunto de herramientas y software que permite el procesamiento y la transformación de datos en conocimiento a través de interfaces amigables para el usuario. Los usuarios, al analizar dicha información, obtienen insights que permiten la mejora del rendimiento de negocio.

Tendencias

Procesamiento y generación del lenguaje natural

- **Chatbots y asistentes de datos virtuales:** la explosión de la IA Generativa y los LLM ha revolucionado los modelos inteligentes gracias a su capacidad para mejorar la productividad y la eficiencia a través de consultas en lenguaje natural que permite a los usuarios obtener datos relevantes gracias a su capa de inteligencia.

- **Generación automatizada de información:** gracias a la implicación de la IA los sistemas ya no solo proporcionan datos, sino que detectan patrones, tendencias, anomalías e información clave de forma automatizada y ágil, permitiendo a los usuarios comprender rápidamente lo que están viendo sin necesidad de formación estadística profunda.

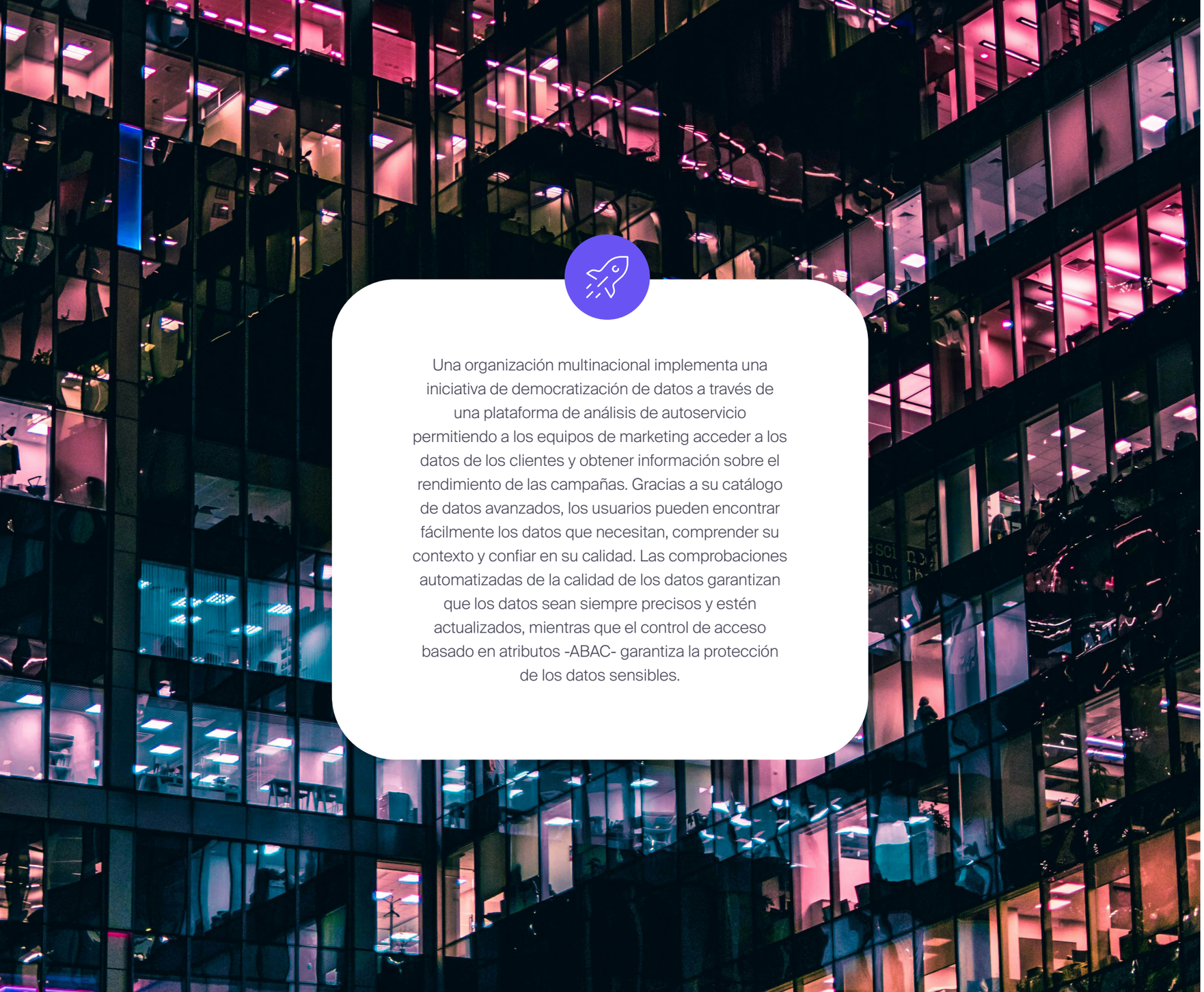
Augmented Data Management

- **Gestión de datos mejorada por el aprendizaje automático:** gracias a la IA y al ML, el software aprende de las interacciones con los usuarios, mejorando continuamente y automatizado el proceso de gestión de datos. De esta forma, el ecosistema de datos evoluciona de la mano de las necesidades de los usuarios.

Ecosistemas de datos descentralizados

- **Blockchain para compartir datos:** permite ampliar el potencial de la democratización de datos gracias al registro compartido que controlan los miembros de la red. Este mecanismo permite compartir información transparente y el almacenamiento de los datos en bloques que se vinculan entre sí en una cadena. De esta forma, se garantiza que los datos son consistentes debido a que no es posible eliminar ni modificar la cadena sin el consenso de la red.

Fuente: Softtek



Una organización multinacional implementa una iniciativa de democratización de datos a través de una plataforma de análisis de autoservicio permitiendo a los equipos de marketing acceder a los datos de los clientes y obtener información sobre el rendimiento de las campañas. Gracias a su catálogo de datos avanzados, los usuarios pueden encontrar fácilmente los datos que necesitan, comprender su contexto y confiar en su calidad. Las comprobaciones automatizadas de la calidad de los datos garantizan que los datos sean siempre precisos y estén actualizados, mientras que el control de acceso basado en atributos -ABAC- garantiza la protección de los datos sensibles.

Key insight

La democratización de los datos no se trata solo de hacer que los datos sean accesibles, sino de **crear una cultura en la que los datos sean una parte integral del papel de todos.** Al combinar tecnologías avanzadas con un enfoque en el empoderamiento y la educación de los usuarios, las organizaciones pueden liberar todo el potencial de sus datos, impulsando la innovación y el éxito en la era digital.

A medida que la tecnología siga avanzando, las posibilidades de la democratización de los datos no harán más que ampliarse, ofreciendo nuevas formas de empoderar a los usuarios y transformar las empresas.

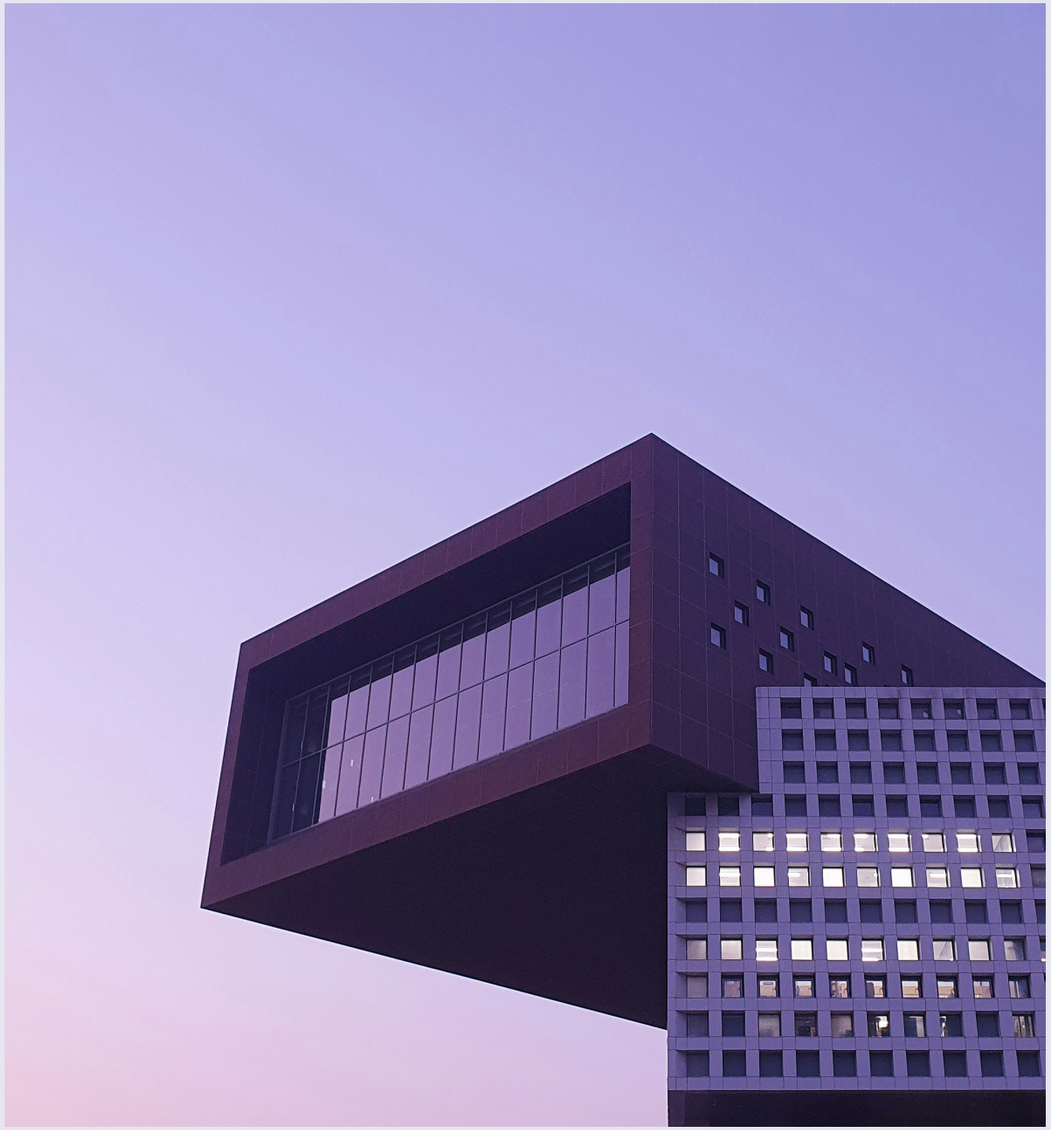
02. Data Monetization & DaaS

La rápida evolución de las tecnologías asociadas al Data ha ocasionado un nuevo paradigma en el que la diferenciación de las organizaciones proviene de su capacidad de adaptación y liderazgo en el nuevo ecosistema digital. La mayoría de las organizaciones ya están investigando e implementando estrategias relacionadas con el impacto que tienen los datos en el rendimiento del negocio para traducirlo en un valor monetario real.

La cuestión abordada por las organizaciones líderes acerca de cómo gestionar las funcionalidades que ofrece el dato ha proporcionado varias tendencias en el mercado. **Data Monetization engloba a su vez varias vertientes estratégicas para abordar este activo empresarial.**

Data Monetization hace referencia al proceso de **obtener un beneficio económico cuantificable de los datos.** Es, en definitiva, la capacidad de una organización para crear y obtener valor a partir de los datos y activos de inteligencia artificial. Esta idea enlaza a su vez con Data-as-a-Service, establece una diferenciación al contemplar otras formas de extraer valor de los datos y así aumentar el rendimiento empresarial.

De este cambio de mentalidad, surgen las **organizaciones Data-Driven, aquellas que basan sus decisiones estratégicas en las evidencias extraídas a través del análisis de datos.** Estas organizaciones integran en su cultura corporativa el valor estratégico de los datos transformando y optimizando sus estrategias en función del conocimiento generado.



Evolución

Si bien el concepto Big Data se remonta a las décadas de 1960 y 1970, cuando comienzan a funcionar los primeros centros de datos, no sería hasta 2005 cuando el mercado se percató de la cantidad de información que se producía a través de plataformas y servicios online. El apogeo se alcanzaría con la llegada de tecnologías como IoT, IA generativa, LLM y Machine Learning en esta década.

El potencial tecnológico y de negocio es evidente para todas las organizaciones que compiten por desarrollar modelos productivos. Gracias a dichas tecnologías las empresas se valen de **modelos predictivos altamente personalizados con las necesidades de los potenciales consumidores**; el mantenimiento predictivo ya es una realidad que ahorra costes en el sector industrial a la vez que reduce los accidentes; la experiencia del cliente o la experiencia del fan han evolucionado llegado a modelos futuristas conectados que únicamente eran posibles en la imaginación hace un pocos años, solo por citar algunas de las aplicaciones.

Background tecnológico

Robust DaaS Platforms

- **Arquitectura escalable:** las plataformas DaaS utilizan una infraestructura Cloud nativa que permite proporcionar soluciones escalables y flexibles de almacenamiento, procesamiento, integración y acceso a datos de fuentes internas y de terceros sin que se vea degradado el rendimiento al servir a numerosos clientes.

- **API-led Connectivity:** las API son la base de las plataformas DaaS ya que permiten un acceso seguro a los servicios de datos. Las dos API más utilizadas son API RESTful y GraphQL. La API RESTful es una interfaz de dos sistemas de computación para intercambiar información de manera segura a través de internet. GraphQL es un lenguaje de consulta que proporciona una sintaxis flexible e intuitiva para describir los requisitos e interacciones de los datos y que permite acceder a muchas fuentes en una sola solicitud. GraphQL y REST son dos enfoques distintos para diseñar una API para el intercambio de datos a través de Internet. REST permite a las aplicaciones cliente intercambiar datos con un servidor mediante verbos HTTP mientras que GraphQL define cómo la aplicación cliente debe solicitar los datos de un servidor remoto.

Background tecnológico

Medidas mejoradas de seguridad de datos y cumplimiento

- **Técnicas avanzadas de cifrado y anonimización:** el Reglamento General de Protección de Datos es el reglamento europeo relativo a la protección de las personas físicas en lo que respecta al tratamiento de sus datos personales y a la libre circulación de estos datos en la UE y el Espacio Económico Europeo, este reglamento y las diferentes leyes de los estados miembros hacen necesaria la implementación de técnicas, herramientas y metodologías que garanticen la confidencialidad. Destacan dos técnicas para abordar esta problemática: FPE y K-anonimidad para garantizar la privacidad de los datos sin perder su utilidad para el análisis y la generación de información. La técnica de cifrado de datos con preservación del formato (Format-Preserving Encryption, FPE) extiende los algoritmos de cifrados clásicos como AES, de manera que los textos conservan la longitud y formato originales. La K-anonimidad es una propiedad de los datos anonimizados que permite cuantificar hasta qué punto se preserva la anonimidad en un conjunto de datos en el que se han eliminado los identificadores, es una medida del riesgo de que agentes externos puedan obtener información de carácter personal a partir de datos anonimizados, las herramientas más utilizadas son ARX Data Anonymization Tool, UTD y Amnesia.

- **Blockchain para la integridad de los datos**: es una técnica de almacenamiento distribuido de información que hace uso de políticas de consenso para validar la información incluida y políticas de detección de integridad de los datos originalmente registrados. La tecnología Blockchain garantiza que todos los datos proporcionados a través de DaaS sean inmutables y rastreables. Estas cualidades la convierten en una defensa sólida contra las amenazas cibernéticas y la manipulación de información. La criptografía es esencial en el cifrado de información, un ejemplo de ello es la generación de hash, permite detectar cualquier alteración en los datos de un bloque. Asimismo, se utiliza para autentificar las transacciones y generar firmas digitales que verifican la identidad de las partes involucradas, asegurando así que las transacciones sean legítimas.

Herramientas de análisis de datos

- **AI y Machine Learning integrados:** las capacidades de la AI y el Machine Learning se integran cada vez más en DaaS para proporcionar información derivada de los datos. El aprendizaje automático usa algoritmos para identificar patrones en los datos y realizar modelos que ayudar a generar información.

- **Procesamiento de datos en tiempo real:** implica recopilar y procesar una secuencia de datos de varios orígenes en tiempo real para extraer un significado útil de los mismos. Algunos ejemplos son las aplicaciones móviles de comercio electrónico, información de las redes sociales o la telemetría de los dispositivos conectados. La transmisión de datos de múltiples dispositivos y las transformaciones ETL en grandes volúmenes de datos continuos y de alta velocidad en tiempo real permiten analizar los datos tan pronto como se producen.

Desafíos

Calidad y consistencia de los datos

- **Herramientas automatizadas de limpieza y enriquecimiento de datos:** son esenciales para la precisión del análisis de datos consiguiendo que estos sean precisos, consistentes y confiables gracias a la recopilación, limpieza, transformación y organización de datos sin procesar o incompletos. Las herramientas de preparación de datos ayudan a transformar y eficientar el proceso de tratamiento de datos. Las herramientas más conocidas del mercado son: Python, Tableau Desktop, OpenRefine, Apache Spark, Datameer, Alteryx, entre otros.

- **Gestión de datos maestros (MDM):** es el proceso de creación y mantenimiento de un único registro maestro para cada dato a través del cual las organizaciones obtienen una visión actualizada y confiable de los datos clave. La implementación de MDM garantiza que toda la organización haga referencia a un conjunto único y coherente de datos maestros, reduciendo los errores. En esta práctica varios proveedores ya están incluyendo los avances en IA y ML para automatizar el proceso de verificación de precisión de nuevos datos y compararlos con los registros actuales.

Estrategias de fijación de precios y valoración

- **Modelos de precios dinámicos:** aprovechan los algoritmos avanzados y tecnologías para analizar grandes volúmenes de datos como son el Big Data, la inteligencia artificial o el Machine Learning, para fijar dinámicamente el precio de los servicios en función de la demanda. Existen diferentes tipos de estrategias de Dynamic Pricing: Time-Based Pricing, Peak-pricing y de penetración de mercado orientadas a maximizar los ingresos sin dejar de ser competitivas.

- **Value-based Pricing:** esta estrategia implica establecer los precios en función del valor percibido del producto, el factor clave es determinar la percepción del valor por parte del cliente, pues permite a la organización la toma de decisiones inteligente y determinar, por ejemplo, si es adecuada una estrategia de precios premium.

Confianza y adopción por parte de los usuarios

- **Políticas de uso transparentes:** el Reglamento General de Protección de Datos señala un conjunto de principios que los responsables y encargados del tratamiento deben observar al tratar datos personales: principio de licitud, transparencia y lealtad; principio de finalidad, principio de minimización de datos, principio de exactitud, principio de limitación del plazo de conservación, principio de seguridad y principio de responsabilidad activa. Estos principios garantizan mantener la debida diligencia y el cumplimiento del RGPD y LOPDGD. Articular y comunicar adecuadamente el tratamiento de los datos genera la confianza a los usuarios de que sus derechos están siendo respetados.

- **Quality & Performance SLAs:** Service Level Agreement (SLA) es un contrato en el que se determina el servicio concreto y detallado y los compromisos de calidad para la prestación de este. Este documento es necesario para generar un marco de actuación clara por parte del proveedor acerca de plazos, calidad y cualquier concepto que permita entender con exactitud el servicio contratado. Entre otros aspectos, los acuerdos SLA deben incluir un apartado sobre gestión de datos, confidencialidad y calidad de los datos.

Desafíos

Estrategias innovadoras en la monetización de datos

- **Servicios de datos por niveles**

Niveles de acceso personalizados: con el fin de atender a una gama amplia de clientes, las organizaciones ofrecen un acceso escalonado a sus servicios de datos, los datos básicos están disponibles de forma gratuita o a bajo coste y los datos más detallados, estratégicos y en tiempo real requieren un nivel premium.

Servicios agrupados: otra forma de atender a los clientes es la agrupación de servicios adicionales como análisis avanzado, soporte o servicios de consultoría si se contratan suscripciones de nivel superior, lo que agrega más valor y justifica un aumento en los precios.

- **Servicios de enriquecimiento de datos**

Conocimientos de datos mejorados: los proveedores de DaaS van un paso más allá al ofrecer, además, datos enriquecidos gracias al procesamiento y análisis que los han convertido en información procesable. Un claro ejemplo de ello incluye el análisis de tendencias, el modelado predictivo o el análisis de engagement.

Integración con conjuntos de datos externos: genera datos enriquecidos que ofrecen un contexto más amplio y conocimientos más profundos que permiten satisfacer las demandas más especializadas de los clientes.

- **Modelos basados en API**

Modelos de pago por uso: los proveedores están ofreciendo modelos de acceso basado en API a sus servicios de datos con precios basados en el volumen de llamadas a la API o por la cantidad de datos consumidos. Este modelo permite un uso más flexible y escalable.

Mercado de API de datos: algunas organizaciones están creando o participando en mercado de API de datos, donde los desarrolladores y las empresas pueden suscribirse a una amplia gama de servicios de datos.

- **Monetización basada en el rendimiento**

Precios orientados a resultados: los proveedores más innovadores de DaaS están vinculado sus precios al rendimiento o a los resultados que sus datos ayudan a conseguir.

Ahorros compartidos y participación en los ingresos: en algunos casos, los proveedores celebran acuerdos de ahorro compartido o

Desafíos

reparto de ingresos, en los que reciben una parte de los ahorros en costes o de ingresos adiciones que sus datos ayudan a conseguir a sus clientes.

- **Sindicación de datos**

Productos de datos multicliente: suponen la agregación de datos de múltiples fuentes, incluidos los propios clientes, para proporcionar información valiosa para todos los participantes.

Anonimización y agregación: con el fin de garantizar la privacidad y el cumplimiento de la ley, los productos sindicados se basan en datos anónimos y agregados, lo que garantiza que los puntos de datos individuales no puedan rastrearse hasta su origen.

- **Productos de datos dinámicos y en tiempo real**

Flujos de datos en tiempo real: gracias al auge del IoT y la creciente demanda de datos, los proveedores ofrecen flujos de datos en tiempo real como un servicio premium para aquellos clientes que necesitan información actualizada constantemente para la realización de su función.

Servicios de datos basados en eventos: servicios de datos activados por eventos o condiciones específicas que proporcionan datos altamente relevantes a demanda.

Tendencias

Flujos de datos personalizados

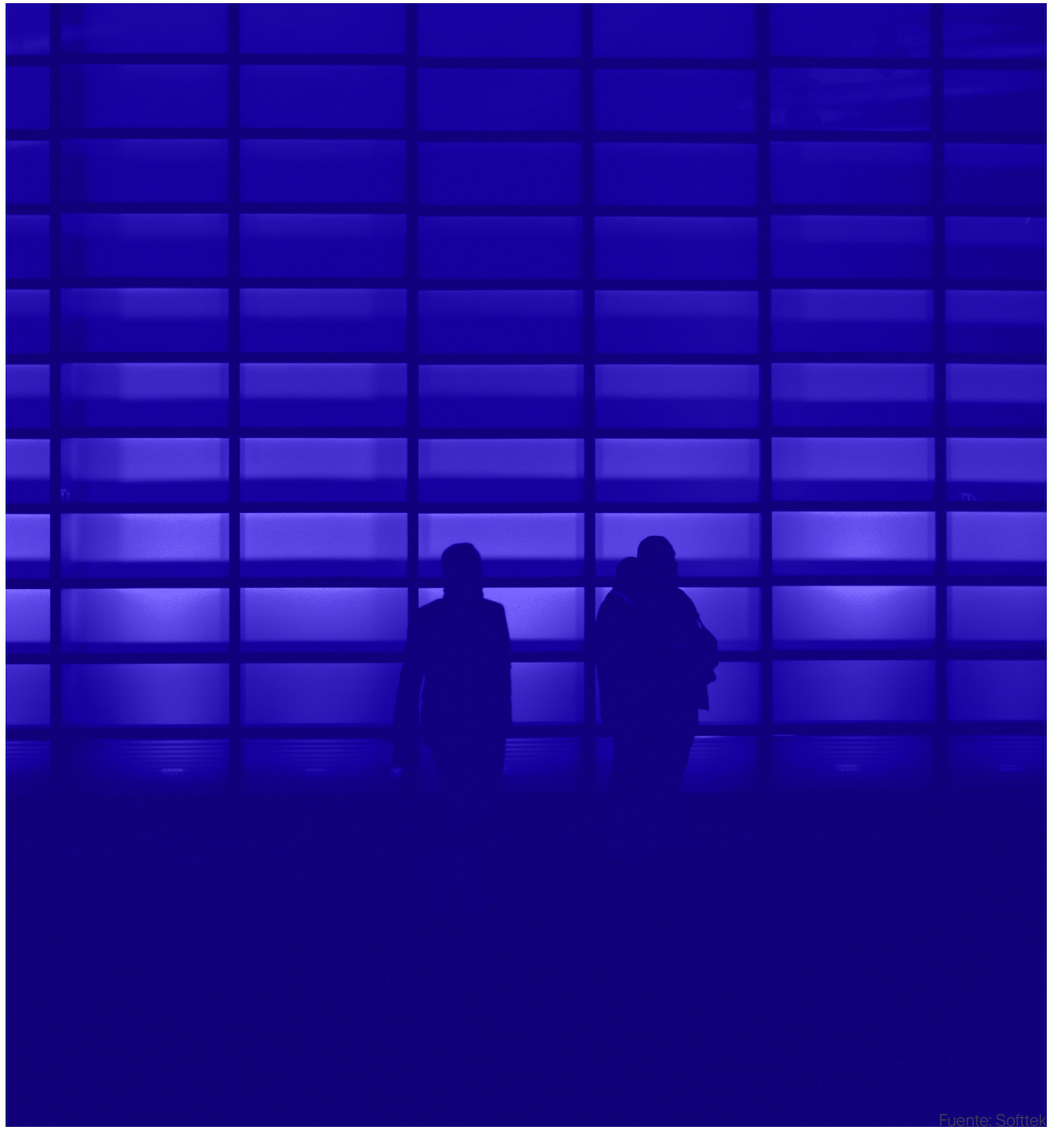
- **Fuentes de datos personalizables:** las futuras plataformas DaaS ofrecerán flujos de datos más personalizados, permitiendo al usuario seleccionar el tipo o la frecuencia de datos que recibe en base a sus necesidades y casos de usos.

Mercados de datos descentralizados

- **Plataformas habilitadas para Blockchain:** el auge de los mercados de datos descentralizados permitirá a las organizaciones comprar y vender datos de forma segura y transparente, garantizando la integridad de estas transacciones

Cumplimiento y gobernanza automatizados

- **Herramientas de cumplimiento impulsadas por IA:** las herramientas automatizadas basadas en la IA se volverán indispensables para garantizar el cumplimiento y la gobernanza continuos, ajustándose a las nuevas regulaciones en tiempo real.



Fuente: Softtek

Una empresa de retail genera valor añadido a su negocio gracias a la implantación de una plataforma DaaS. Utilizando métodos avanzados de recopilación de datos, agregan y anonimizan datos de varias fuentes, lo que garantiza la privacidad. La plataforma, además, emplea el aprendizaje automático para proporcionar información predictiva sobre las tendencias de los consumidores, lo que ayuda a los minoristas a optimizar el inventario y mejorar las ventas. Los modelos de precios dinámicos garantizan que el servicio siga teniendo un precio competitivo, y los SLA sólidos garantizan el rendimiento y la calidad de los datos.

Key insight

La **monetización de datos y DaaS están a la vanguardia de la economía de los datos, transformando los datos sin procesar en activos valiosos.** Al adoptar tecnologías y estrategias avanzadas, las organizaciones pueden desbloquear nuevas fuentes de ingresos, proporcionar servicios mejorados y obtener una ventaja competitiva.

A medida que el campo continúa evolucionando, mantenerse al tanto de las innovaciones y tendencias será crucial para aquellos que buscan liderar en el futuro basado en datos.

03. Cloud Modernization/ New Cloud Data Ecosystems

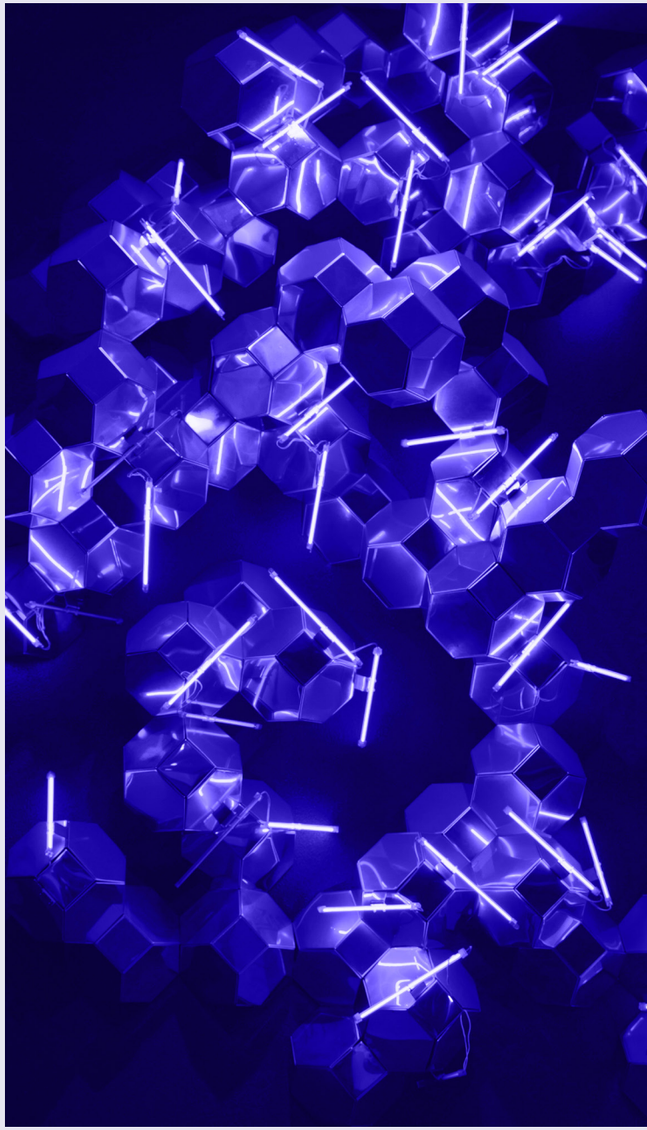
El auge en la adopción de modelos DaaS se espera que genere a su vez una eclosión de nuevos ecosistemas de nubes de datos creados con la finalidad de aprovechar información generada online y sacar un mayor rendimiento de ella gracias a un almacenamiento más ágil, sencillo y eficiente.

Este proceso se ha venido a conocer como **Cloud Modernization e implica la actualización o reestructuración de todo tipo de aplicaciones para aprovechar al máximo las arquitecturas y servicios nativos en la nube.**

Por ejemplo, una modernización en la nube puede consistir en tomar una aplicación monolítica y refactorizarla para utilizar una arquitectura de microservicios, haciendo que así consuma menos recursos de alojamiento. De esta forma, una aplicación modernizada aprovecha mejor la nube que otra que únicamente migra datos sin hacer cambios en su propia arquitectura.

En este punto, podemos diferenciar entre tres tipos de modernización en la nube:

- **Modernizar una aplicación en local (on-premise)** rediseñándola para aprovechar al máximo antes de trasladarla a la nube.
- **Modernizar una aplicación que ya está alojada en la nube** pero que aún no ha sido optimizada para un funcionamiento realmente eficiente.
- **Modernizar una aplicación heredada en local (on-premise) y mantenerla en local, pero alojarla en un entorno nativo de la nube.**



Evolución

La **evolución de los ecosistemas de datos** en la nube ha estado marcada por dos factores fundamentales, el **almacenamiento** y el **procesamiento de datos.** El avance acelerado de las tecnologías Cloud, inteligencia artificial, Machine Learning o Data presagian un cambio tan importante como el que provocó Internet en 1997. Los primeros servicios Cloud ofrecían soluciones de almacenamiento simple, pero **a medida que progresa la tecnología lo hace la capacidad de proporcionar servicios.** DaaS surgió como un componente crítico de esta evolución encapsulando la idea de entregar datos a los usuarios bajo demanda en cualquier lugar y momento, sin la carga de administrar la infraestructura subyacente.

Los ecosistemas de datos están pasando de software autónomo o implementaciones combinadas a **soluciones cien por cien nativas Cloud.** Se espera que, a lo largo de este año, las implementaciones de nuevos sistemas en la nube se basen en un **ecosistema de datos en la nube** en lugar de en soluciones puntualmente integradas. Las organizaciones deben evaluar los ecosistemas de datos en función de su capacidad para integrar datos de fuera de su entorno y resolver desafíos como los datos distribuidos.

En este punto **las nuevas capacidades de la IA generativa abren la puerta a una nueva era de aplicaciones que pueden elevar el valor de los datos a su máximo nivel al crear, recomendar, sintetizar, analizar o interactuar con ellos de manera natural, garantizando la transparencia y responsabilidad en todo el proceso.**

Background tecnológico

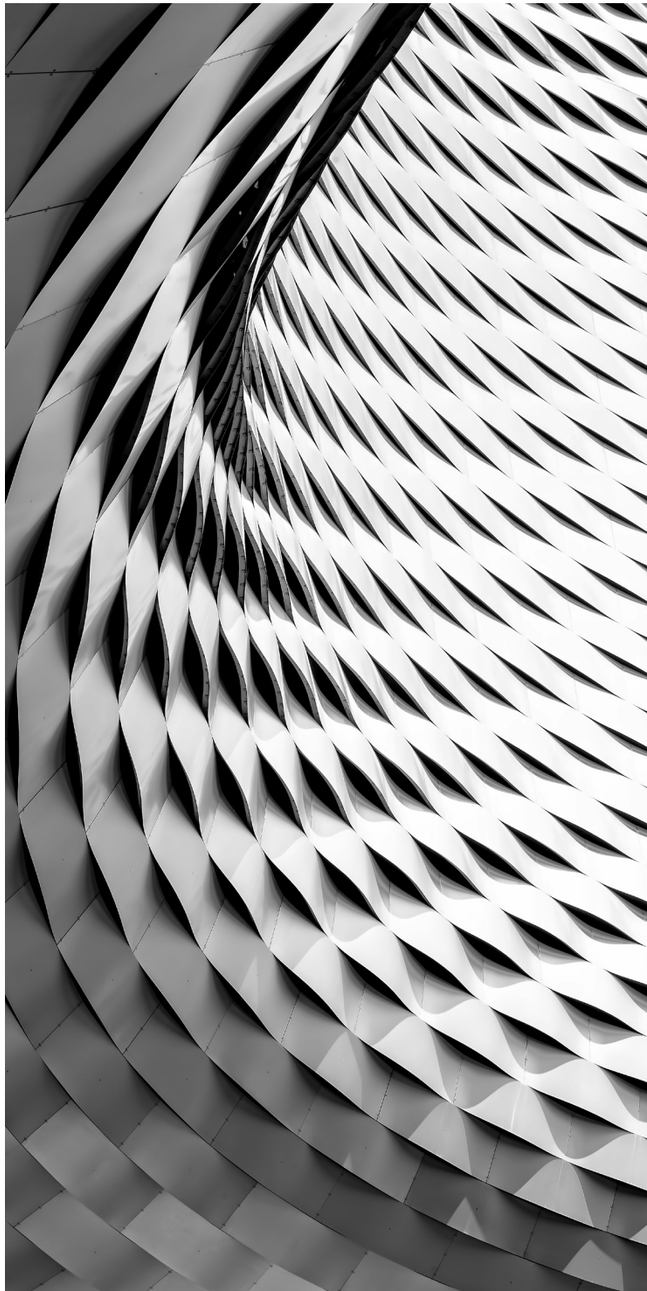
Los ecosistemas DaaS de última generación están basados en varios componentes técnicos:

- **Arquitectura API-First:** el enfoque de DaaS está centrado en el aprovisionamiento de datos de una amplia gama de fuentes bajo demanda a través de una API. Las API son la columna vertebral de este sistema, diseñado para simplificar el acceso a datos y ofrecer conjuntos de datos o flujos de datos preparados para ser consumidos en varios formatos a través de la nube de manera eficiente.

- **Microservices & Containerization:** la arquitectura de microservicios hace referencia a un estilo de arquitectura pensado para desarrollar aplicaciones. Esta metodología permite dividir una aplicación grande en partes más pequeñas e independientes con su propio ámbito de responsabilidad. De esta forma, una aplicación basada en microservicios puede llamar a muchos microservicios internos para formar su respuesta con la finalidad de entregar una única solicitud al usuario. Los contenedores son un claro ejemplo de esta arquitectura, ya que permiten enfocarse en desarrollar los servicios sin tener en cuenta las dependencias. Las aplicaciones nativas Cloud se compilan de esta forma permitiendo, en este caso, el desarrollo y la implementación de aplicaciones DaaS. De esta forma, cada servicio se puede escalar, actualizar y administrar de forma independiente, lo que reduce el tiempo de inactividad y acelera la innovación.

- **Data Virtualization & Federation:** la federación de datos es un proceso que permite que varias bases de datos funcionen como una sola proporcionando una única fuente de datos para las aplicaciones. La federación de datos es parte del marco de virtualización de datos. Incluye repositorios de metadatos, abstracción de datos, acceso de lectura y escritura a los sistemas de datos de origen y seguridad avanzada. Estas tecnologías permiten que las plataformas DaaS integren datos de múltiples fuentes, proporcionado a los usuarios una visión unificada de los datos sin necesidad de replicarlos o moverlos físicamente.

- **Integración de inteligencia artificial y aprendizaje automático:** las capacidades analíticas avanzadas son parte integral de las plataformas DaaS modernas. El valor de este servicio reside en la entrega de análisis y percepciones para ofrecer información y modelos predictivos, es aquí donde la integración de la inteligencia y el aprendizaje automático marca el paso hacia el futuro.



Desafíos

- **Privacidad y seguridad de los datos:** un ecosistema de datos basado en la IA y con enfoque Cloud First permite a la organización generar un mayor valor empresarial a partir de activos de datos únicos, pero también genera la necesidad de garantizar su seguridad. Las soluciones de cifrado avanzado, como las que ofrece Microsoft Azure, entre otros proveedores, identifican los ataques más rápido, reducen los riesgos y diseñan la seguridad en base a las necesidades del cliente.

- **Calidad y coherencia de los datos:** a medida que las empresas integran la inteligencia artificial y la automatización en sus flujos de trabajo, los datos de alta calidad se vuelven cruciales para la adopción efectiva de dichas herramientas. En la última década, los avances en Cloud, IA, IoT y Edge Computing han evolucionado de manera exponencial las posibilidades del Big Data. Como resultado, la gestión de datos se ha vuelto una práctica más crucial y compleja. Las organizaciones deben poner sus esfuerzos en la gestión de calidad de los datos para garantizar que las iniciativas de Business Intelligence son correctas. En este sentido, encontramos dos aspectos determinantes, la gobernanza y la calidad de los datos. Por ello, las organizaciones deben establecer políticas y programas de administración de datos y establecer mecanismos para resolver los problemas que puedan surgir a lo largo de todo el ciclo de vida de los datos. Esto implica una sofisticada validación de datos, procesos de limpieza y monitoreo en tiempo real de anomalías.

- **Complejidad de la integración:** reside en el reto de integrar con los sistemas heredados- El objetivo de la integración es generar un sistema más amplio, completo y estandarizado, por ello, para abordar este proceso el uso de API estandarizadas y la adopción de una estrategia de integración sólida son clave para superar estos desafíos.

Las plataformas DaaS incorporan varias herramientas y técnicas avanzadas en sus servicios y soluciones:

- **Análisis predictivo:** consiste en procesar datos históricos y actuales a través de técnicas matemáticas, estadísticas, inteligencia artificial y Machine Learning para analizar los datos y determinar las tendencias o eventos futuros, ayudando así a las organizaciones en la toma de decisiones estratégicas.

- **Sentiment Analysis:** es el proceso que ayuda de determinar el tono emocional de una serie de palabras. La herramienta de análisis de sentimiento es una técnica automatizada que ayuda a extraer la información significativa de los stakeholders acerca de sus actitudes, emociones u opiniones acerca de la organización o el tema que se esté evaluando. Esta herramienta utiliza tecnologías avanzadas de inteligencia artificial como el procesamiento del lenguaje natural, análisis de texto o ciencia de datos para extraer la información. Es particularmente útil para la realización de estrategias de marketing y el monitoreo de las redes sociales, al permitir a los expertos comprender el sentimiento y las tendencias de los clientes.

- **Procedimiento de datos en tiempo real:** se refiere a la capacidad de analizar y actuar sobre los datos a medida que se generan, proporcionando información y respuestas en tiempo real, reaccionando ante eventos específicos que acaben de ocurrir con la lógica adecuada. Esto es crucial para todas las aplicaciones que requieren de información inmediata como son los sistemas de respuestas a emergencias o comercio financiero.

Tendencias

Integración de Edge Computing: permite el procesamiento de datos descentralizado cerca de las fuentes de datos, reduciendo la latencia y el uso del ancho de banda, especialmente crítico para los servicios de datos en tiempo real.

Blockchain para la seguridad de los datos: está revolucionando la seguridad y la transparencia de los datos en DaaS, ofreciendo transacciones de datos inmutables y rastreables.

Servicios de IA personalizados: a medida que avanza la tecnología de la IA, las plataformas DaaS ofrecerán servicios más personalizados, proporcionando a las empresas información y modelos específicos para sus necesidades y desafíos.

Una empresa de atención médica necesita proporcionar datos clínicos en tiempo real a hospitales e instituciones de investigación. Para llevar a cabo este proyecto se implementó una plataforma DaaS basada en la nube, redefiniendo la forma en la que los usuarios acceden y analizan la información y mejorando la atención al paciente e impulsando la investigación médica.

Key insight

Los **nuevos ecosistemas de datos en la nube**, que se centran especialmente en DaaS, **están redefiniendo la forma en que se accede, analiza y monetiza a los datos.** Representan no solo un avance tecnológico, sino un **activo estratégico** que puede impulsar la innovación, la eficiencia y los nuevos modelos de negocio. A medida que estos ecosistemas continúan evolucionando, prometen desbloquear oportunidades sin precedentes para las empresas listas para adoptar el futuro de los datos.

04. Data Fabric

Conseguir gobernar todas las fuentes de datos que una organización tiene a su alcance, ya sea en sistemas locales, entornos Multi Cloud o aplicaciones SaaS (Software-as-a-Service), y hacerlo a través de una única capa de virtualización, sin necesidad de copiar los datos de un lugar a otro, es una máxima para todo negocio que desee mejorar su enfoque de arquitectura de datos en los próximos años.

Las estrategias de integración basadas en el enfoque Data Fabric van a desempeñar un papel fundamental a la hora de construir arquitecturas más sólidas, ya que facilitan esta integración end-to-end a través de la aplicación de sistemas inteligentes y automatizados.

El diseño de arquitecturas basadas en el concepto de Data Fabric funciona creando una capa integrada de datos y procesos conectados. Esta capa realiza análisis continuos sobre los metadatos para poder reutilizar la información generada en todos los entornos Cloud, incluyendo las plataformas híbridas y Multi Cloud, pero también en las nuevas redes de sensores y dispositivos conectados IoT o en las aplicaciones de inteligencia artificial, realidad virtual o realidad aumentada.

A través del enfoque Data Fabric se consigue entrelazar los datos generados en todos estos entornos diferentes y tratarlos mediante una única unidad reutilizable, proporcionando una vista unificada y coherente.

Data Fabric proporciona, además, capacidades avanzadas de gobernanza de datos y seguridad ya que ayuda a garantizar que todos los datos se están gestionando y protegiendo bajo unas adecuadas políticas, y cumpliendo con las regulaciones, independientemente de su fuente de origen.

Evolución

Data Fabric nace como un concepto innovador de la necesidad de las compañías de analizar y procesar datos para aprovechar la inteligencia colectiva en sus estrategias. Supone el rediseño de las arquitecturas de las organizaciones, teniendo como eje principal de la operación los datos apoyados por un enfoque tecnológico.

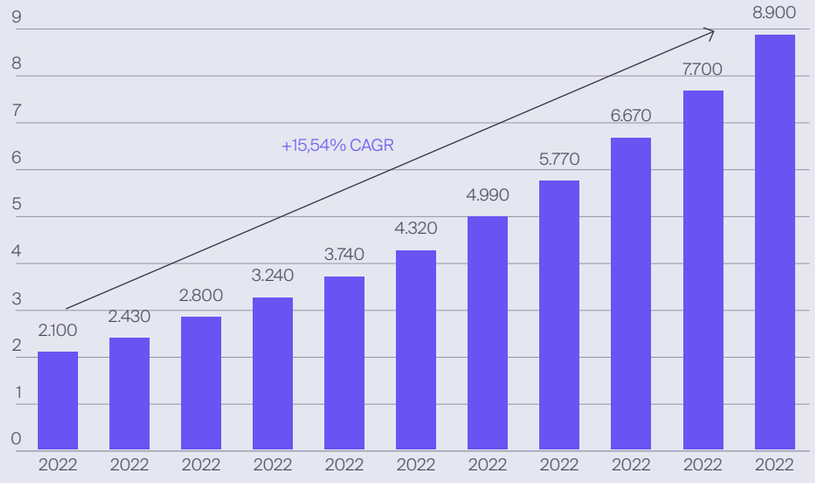
Desde una perspectiva semántica del término se podría traducir como tejido de datos, aludiendo a las fibras que de forma interrelacionada conforman una estructura que con el nivel adecuado de organización brindan una ventaja competitiva. El objetivo final es maximizar el valor de los datos y acelerar la transformación digital, Data Fabric simplifica e integra la gestión de datos en la nube conectando múltiples lugares, tipos y fuentes de datos a través de una única plataforma.

El concepto de Data Fabric ha evolucionado como respuesta a la creciente complejidad de los entornos de datos. Tradicionalmente, los datos se administraban a través de sistemas centrados, pero debido a la explosión de fuentes de datos y computación Cloud, el enfoque debía ser más flexible, escalable e integrado para mejorar la calidad de los datos, la disponibilidad de información y conocimientos automatizados, permitiendo anticipar los desafíos y oportunidades futuros en la gestión de datos.

Durante esta década, el diseño de arquitecturas Data Fabric experimentará un crecimiento de mercado de más del 15,5% interanual gracias a la oportunidad que ofrece a las organizaciones para orquestar los flujos de datos, procesos y operaciones de manera más eficiente y consistente.

Evolución del tamaño de mercado de Data Fabric a nivel global

millones de dólares



Fuente: Softtek en base a datos de Precedence Research

Background tecnológico

El Data Fabric implementa dos características de alto nivel:

- **Flujos de datos:** proporcionan una interfaz de bajo código que procesa datos de cientos de orígenes y permite aprovechar cientos de transformaciones basadas en inteligencia artificial.

- **Data-Pipeline:** elevan a su máximo exponente las funcionalidades de orquestación de datos enriquecidos para crear flujos de trabajos flexibles y continuos garantizando que sea accesible donde y cuando se necesiten.

Asimismo, se compone de varias capas que desarrollan una función específica:

- **Capa de fuentes de datos:** tiene como función la catalogación dinámica de datos que proporciona una vista unificada de todos los datos disponibles con independencia de dónde estén alojados o su origen, facilitando su búsqueda y acceso.

- **Capa de integración de datos:** o capa semántica, realiza todas las tareas necesarias para conectar y preparar los datos para su utilización, desde la limpieza de datos, estandarización de formatos o la indexación para que sean comprensibles y utilizables por toda la organización.

- **Capa de acceso a datos:** proporciona el acceso a los datos a través de una API o interfaz de usuario de fácil manejo para que los usuarios puedan hacer consultas o análisis sin la interferencia de la complejidad subyacente.

- **Capa de seguridad y gobierno de datos automatizado:** se encarga de mantener la privacidad y la integridad de los datos a lo largo de su ciclo de vida y del cumplimiento de las regulaciones y políticas asociadas.

Desafíos

Complejidad de la integración: la integración de datos en diversos sistemas y fuentes de datos puede ser compleja. Es esencial auditar los procesos y sistemas actuales para comprender las fuentes, sistemas y flujos de datos existentes. Una vez diseñada la arquitectura, la integración gracias al uso del middleware y API estandarizadas, consigue integrar los sistemas en una infraestructura de vista unificada.

Mantener la calidad de los datos: para garantizar que los datos sigan siendo precisos y coherentes en todos los sistemas, se debe planificar la gestión y control de los datos a través de la definición de las políticas, procesos y estándares y herramientas de monitoreo y validación de la calidad de los datos en tiempo real.

Escalabilidad: a medida que crecen los volúmenes de datos en las organizaciones, la arquitectura debe escalar en consecuencia. Por ello, la adopción de un enfoque modular basado en microservicios garantiza la capacidad para adaptarse con fluidez a las nuevas necesidades y ayuda a unificar sus procesos para ser rápidos y precisos con los cambios.

El Data Fabric mejora el análisis de los datos mediante la incorporación de herramientas y técnicas avanzadas:

Modelo predictivo: gracias al acceso a una amplia variedad de datos, los modelos predictivos pueden pronosticar tendencias y comportamientos con mayor precisión.

Integración de aprendizaje automático: Data Fabric alimenta los algoritmos de aprendizaje automático mejorando su precisión.

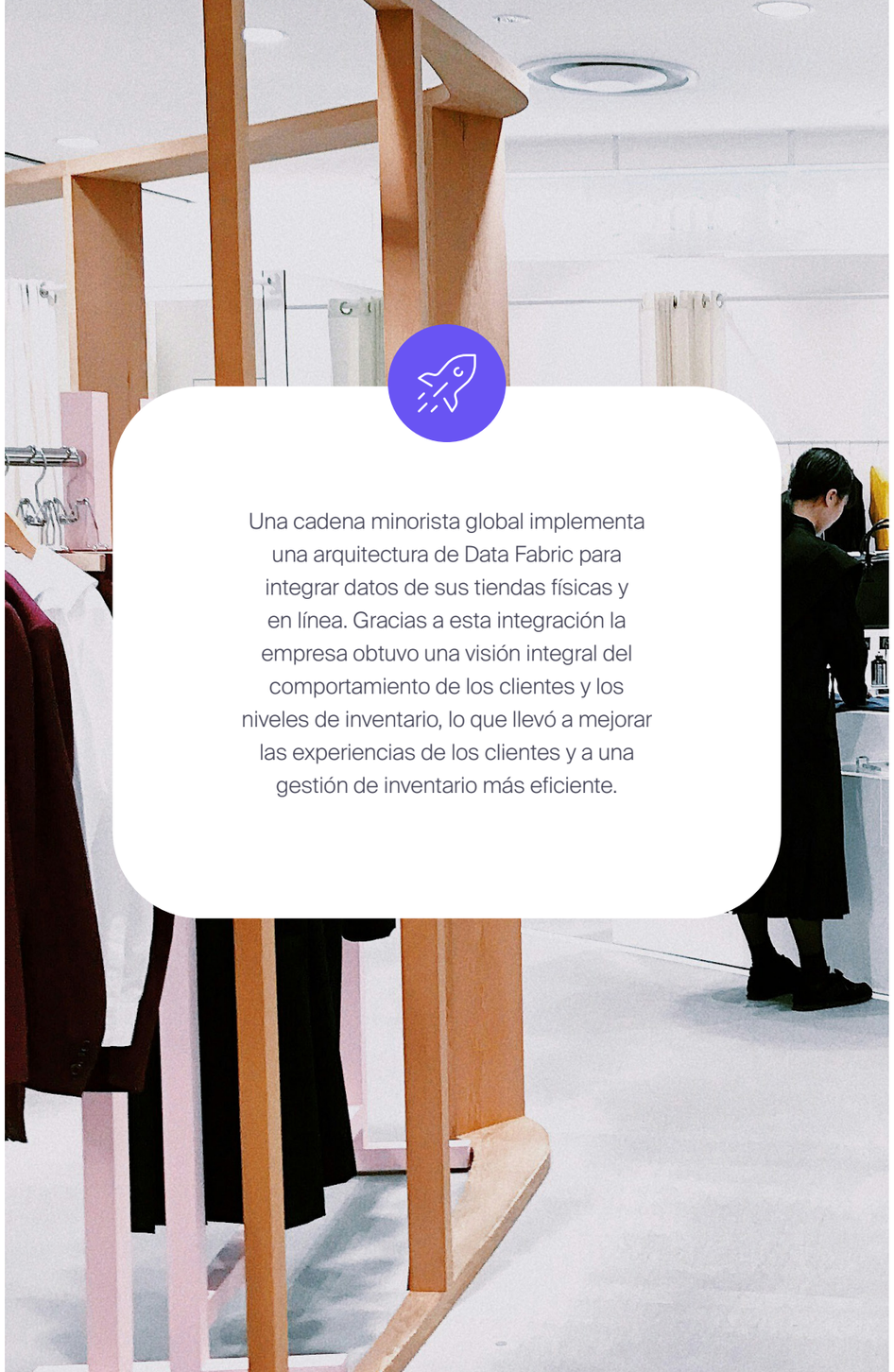
Graph analytics: esta herramienta es especialmente útil en un contexto de Data Fabric para comprender las relaciones y los patrones en diversos conjuntos de datos.

Tendencias

Mayor integración de la IA: a medida que avancen las tecnologías de inteligencia artificial y aprendizaje automático, se integrarán con más profundidad en Data Fabric con el fin de proporcionar información y automatización más sofisticadas, eficientes y estratégicas.

Gestión de Edge Data Centers: con el auge de la computación perimetral Data Fabric necesitará cada vez más gestionar los datos en el perímetro de las redes y no solo en nubes centralizadas o centros de datos.

Mayor enfoque en los datos en tiempo real: se convertirá en una característica estándar de Data Fabric, lo que permitirá obtener información y acciones más inmediatas.



Una cadena minorista global implementa una arquitectura de Data Fabric para integrar datos de sus tiendas físicas y en línea. Gracias a esta integración la empresa obtuvo una visión integral del comportamiento de los clientes y los niveles de inventario, lo que llevó a mejorar las experiencias de los clientes y a una gestión de inventario más eficiente.

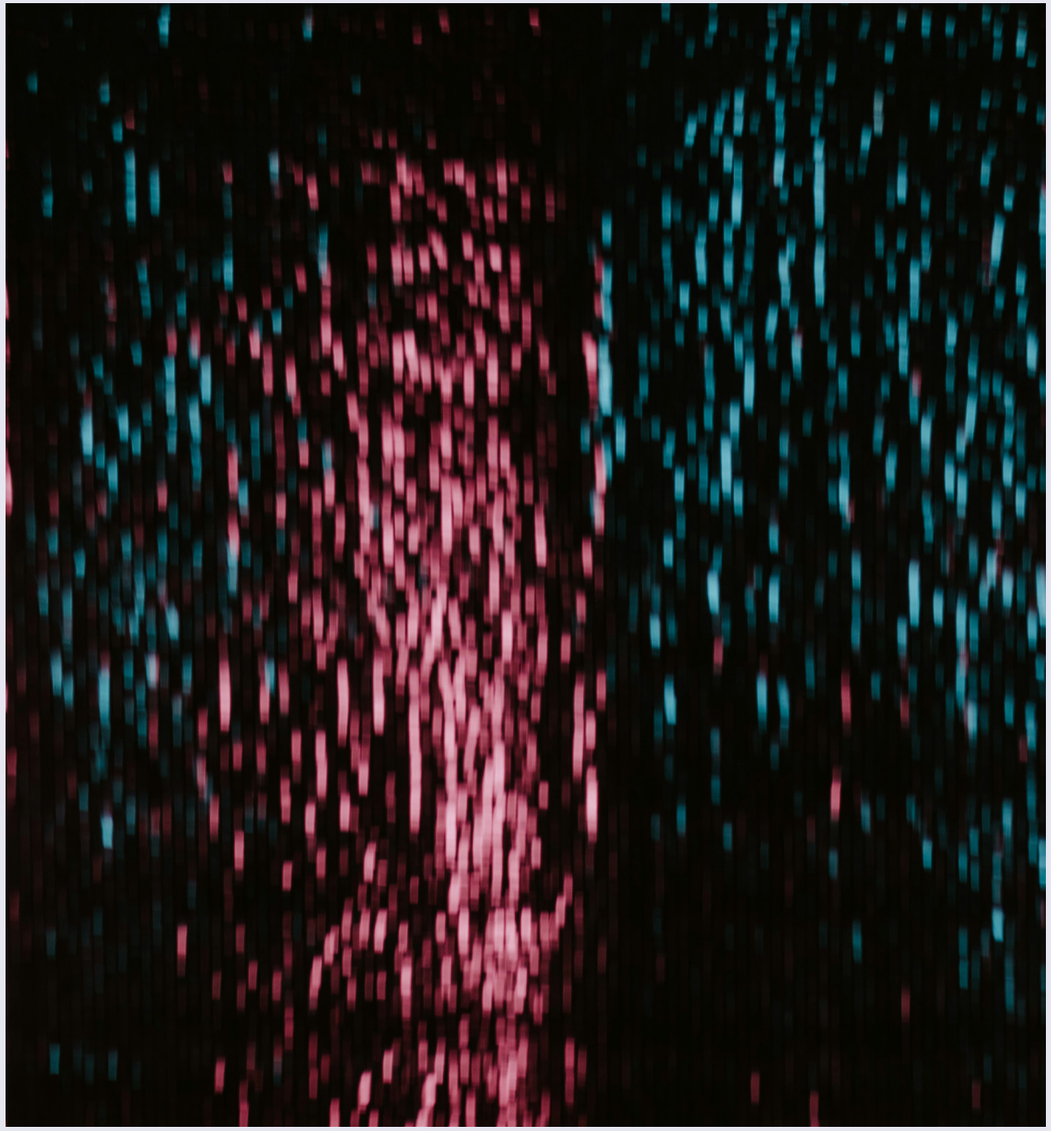
Key insight

Data Fabric representa el futuro de la gestión de datos, ya que ofrece un enfoque flexible, integrado y eficiente para gestionar las necesidades de datos en constante crecimiento y evolución de las organizaciones modernas. **Al adoptar Data Fabric, las empresas pueden asegurarse de que están preparadas para afrontar los retos de la era digital, aprovechando al máximo sus datos para impulsar la toma de decisiones, la innovación y el crecimiento.**

05. TinyML/ Small y Wide Data

Los avances tecnológicos TinyML, Small Data y Wide Data están revolucionando el campo de la analítica de datos y el aprendizaje automático. A medida que generamos más datos, la capacidad para procesarlos y extraer valor en tiempo real se posiciona como una ventaja competitiva clave para todas las organizaciones.

Tiny Machine Learning (TinyML) se centra en la ejecución de modelos de aprendizaje automático en dispositivos y máquinas de bajo consumo, para que adquieran la capacidad suficiente para manejar sistemas completos de aprendizaje automático gracias al enfoque de datos pequeños que proporcionan estrategias para extraer información significativa de conjuntos de datos. A diferencia de las aplicaciones tradicionales de aprendizaje automático, TinyML no depende de elementos como microprocesadores y unidades de procesamiento de gráficos (GPU). A pesar de la capacidad limitada, utiliza las mismas herramientas para conseguir el aprendizaje profundo como sensores, algoritmos y herramientas de análisis de datos. Estas funciones se escalan mediante un proceso de cuantificación que permite la reducción del tamaño de los bits. Los algoritmos TinyML, ofrecen funcionalidades de almacenamiento más reducido que permiten llevar las capacidades de la inteligencia artificial y el aprendizaje automático directamente a los dispositivos Edge, procesando datos más cerca de la fuente donde se originan y permitiendo una toma de decisiones mucho más rápida y eficiente. También ofrecen ventajas en términos de latencia, privacidad y eficiencia, pero también presentan desafíos a la hora de entrenar los modelos de forma efectiva. Una problemática habitual es el sobreajuste (overfitting), o la generalización, cuando el modelo se adapta demasiado bien a los datos de entrenamiento, incluyendo el ruido o las fluctuaciones aleatorias presentes en los datos y presentando un rendimiento deficiente cuando utiliza nuevos datos que no formaron



parte del conjunto de entrenamiento inicial. Un claro ejemplo de la aplicación de TinyML es el modelo de detección de palabras de activación como "Ok, Google" que apenas necesita 14 kilobytes para ejecutarse.

Small Data se refiere a la recopilación y el análisis de conjuntos de datos procedentes de organizaciones individuales o fundamentados en ejemplos individuales de resolución de problemas. Se centra en la importancia de utilizar conjuntos de datos limitados y seleccionados para garantizar su valor respecto a los objetivos para desarrollar modelos de aprendizaje automático. Para conseguir este propósito, emplea técnicas de análisis de series de tiempo, datos sintéticos o aprendizaje autosupervisado para encontrar información de valor para la empresa en conjuntos de datos más pequeños.

Wide Data consiste en relacionar fuentes de datos diferentes entre sí partiendo de una amplia gama para lograr un análisis significativo. En el enfoque Wide Data se maneja un volumen de datos mayor que en el Small Data, pero sin perseguir volúmenes masivos como ocurre en los enfoques Big Data. Permite analizar gran variedad de fuentes y datos tanto estructurados como no estructurado. Esto a su vez, permite encontrar enlaces entre diversos formatos.

Gracias a Small Data y Wide Data, encontramos datos más apropiados para extraer un conocimiento más específico y exhaustivo que con Big Data.

Tanto el Small como el Wide Data ofrecen múltiples oportunidades para todos aquellos sectores en los que la hiperpersonalización sea una variable fundamental del negocio, como es el caso del retailing online. Pero también aplica para cualquier producto o servicio que requiera de tecnologías que faciliten la toma de decisiones rápidas. Un ejemplo es el de los coches con prestaciones autónomas. donde es imprescindible un procesamiento rápido e inteligente a través de dispositivos Edge para determinar acciones en base a pequeños (pero muy precisos) volúmenes de datos registrados en tiempo real.

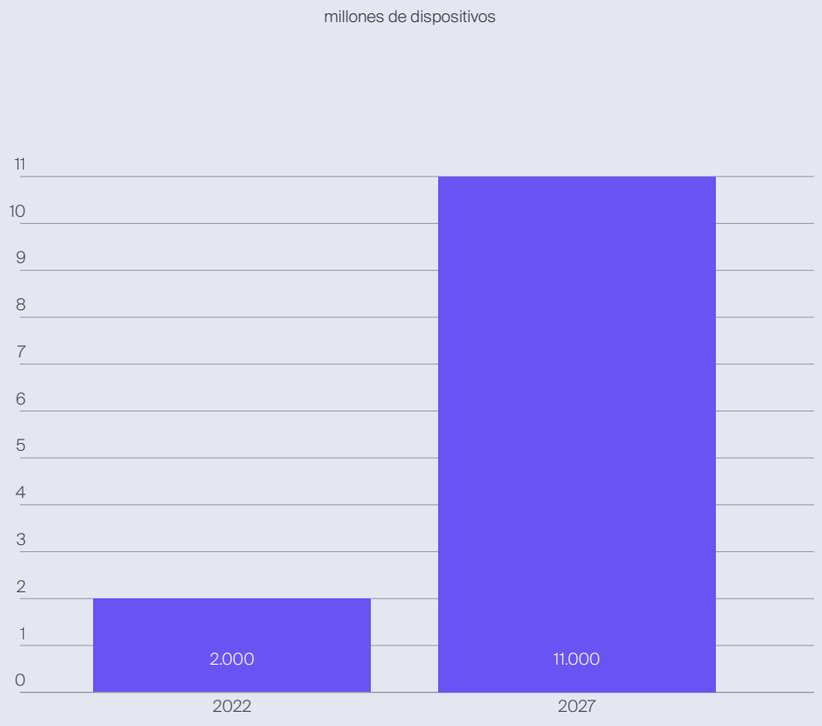
Evolución

Desde las décadas de 1960 y 1970 el Big Data no había tenido un competidor digno. Tradicionalmente, el aprendizaje automático se ha basado en grandes conjuntos de datos centralizados, Big Data. Sin embargo, debido a la proliferación de dispositivos IoT y la necesidad de procesamiento en tiempo real, se ha producido un cambio hacia la incorporación de la inteligencia. Es por ello que, cada vez más, se está se está reconociendo la **importancia de utilizar conjuntos más reducidos y diversos de la mano de nuevos enfoques estratégicos como el Small Data y Wide Data.**

En los próximos años, veremos cómo continua la experimentación con estos modelos para buscar técnicas avanzadas que permitan solventar dichas barreras y extraer todo el valor posible del tratamiento de conjuntos de datos limitados. **Para el año 2025, se espera que el 70% de las organizaciones hayan cambiado su enfoque de análisis de datos desde el Big Data al Small y Wide Data.**

Asimismo, se espera que la instalación de dispositivos TinyML aumente desde los casi 2.000 millones que existían a nivel global en 2022 hasta los más de 11.000 millones en 2027. La variable principal de este crecimiento va a ser la tendencia a llevar el aprendizaje automático en casi todas partes. Por ejemplo, los algoritmos TinyML son especialmente útiles en sensores de sonido y ambiente, y existen muchos casos de uso posibles, lo cual impulsará el gran aumento en las instalaciones de dispositivos con prestaciones TinyML.

Adopción de dispositivos TinyML a nivel global



Fuente: Softtek en base a datos de ABI Research

Background tecnológico

- **Avances en hardware:** el progreso en el diseño de hardware de ultra bajo consumo permite realizar cálculos complejos en dispositivos pequeños como microcontroladores.

- **Optimización de modelos:** técnicas como el modelo de destilación, modelo de poda, cuantificación de modelos o destilación de conjuntos de datos permiten la optimización del aprendizaje automático que permiten crear modelos de IA compactos sin comprometer el rendimiento.

- **Estrategias de datos:** a medida que generamos gran cantidad de datos, las estrategias para convertirlos en información proliferan. Dependiendo de las capacidades y necesidades de las organizaciones, se escogerán los modelos. Mientras que el Small Data se centra en el uso de conjuntos de datos más pequeños, pero bien seleccionados, Wide Data implica la integración de diversos tipos de fuentes de datos, lo que requiere técnicas sofisticadas de fusión de datos.

Desafíos

Restricciones de recursos: los dispositivos TinyML suelen tener una memoria y potencia computacional limitadas que se solventan gracias a las soluciones de optimización de algoritmos y modelos para reducir su huella mientras se mantiene la precisión.

Calidad y diversidad de datos: garantizar la calidad y representatividad de los datos pequeños y amplios es un desafío. Las técnicas que ayudan a superar el reto incluyen el preprocesamiento robusto de datos, detección de anomalías e incorporación de conocimientos de dominio en el análisis.

Procesamiento en tiempo real: tanto TinyML como Wide Data requieren de procesamiento en tiempo real o casi en tiempo real. Las arquitecturas eficientes de Edge Computing y los marcos de procesamiento de datos de transmisión son clave para abordar esta necesidad.

Herramientas analíticas avanzadas:

Edge Analytics: herramientas diseñadas para Edge Computing pueden analizar datos en el dispositivo, lo que reduce la latencia y los requisitos de ancho de banda.

Federated Learning: es una técnica de aprendizaje automático que entrena un algoritmo a través de una arquitectura descentralizada formada por múltiples dispositivos que contienen sus propios datos locales y privados, preservando la privacidad y reduciendo la necesidad de transferencia de datos.

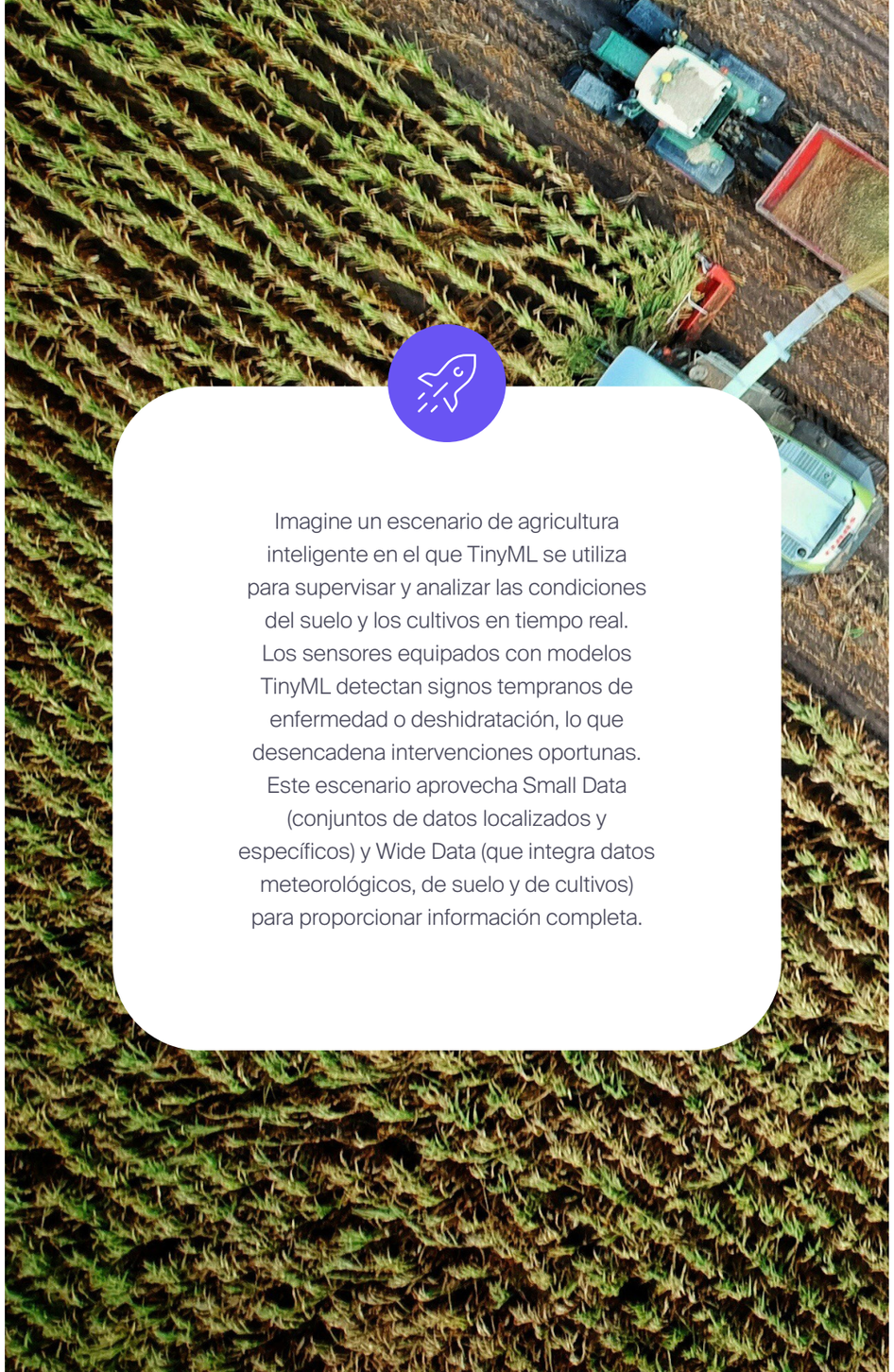
Generación de datos sintéticos: cuando los datos reales son escasos o sensibles, los datos sintéticos son una herramienta valiosa, especialmente en el contexto de Small Data. El proceso de creación de datos sintéticos generados artificialmente implica un proceso de generación o modelado que debe conseguir la preservación del valor analítico para los casos de uso.

Tendencias

- **Integración con 5G:** el despliegue de las redes 5G mejorará las capacidades de las aplicaciones TinyML y Wide Data, proporcionando una transferencia de datos más rápida y fiable.

- **Avances en Neuromorphic Computing:** esta tecnología emergente imita las estructuras y procesos del cerebro humano, lo que podría revolucionar la eficiencia de TinyML.

- **More Sophisticated Multi-modal Data Integration:** a medida que las estrategias de datos amplios se vuelven más comunes, las herramientas para integrar y dar sentido a diversos tipos de datos serán cada vez más sofisticadas.



Imagine un escenario de agricultura inteligente en el que TinyML se utiliza para supervisar y analizar las condiciones del suelo y los cultivos en tiempo real. Los sensores equipados con modelos TinyML detectan signos tempranos de enfermedad o deshidratación, lo que desencadena intervenciones oportunas. Este escenario aprovecha Small Data (conjuntos de datos localizados y específicos) y Wide Data (que integra datos meteorológicos, de suelo y de cultivos) para proporcionar información completa.

Key insight

Los enfoques de **TinyML y Small and Wide Data están destinados a transformar la forma en que pensamos y utilizamos los datos y el aprendizaje automático.** Al permitir el procesamiento inteligente a menor escala, en contextos más diversos y más cerca de donde se generan los datos, ofrecen la promesa de **información y acciones más personalizadas, oportunas y eficientes.**

A medida que estos campos evolucionen, abrirán nuevas posibilidades para la innovación y la comprensión en prácticamente todos los dominios.

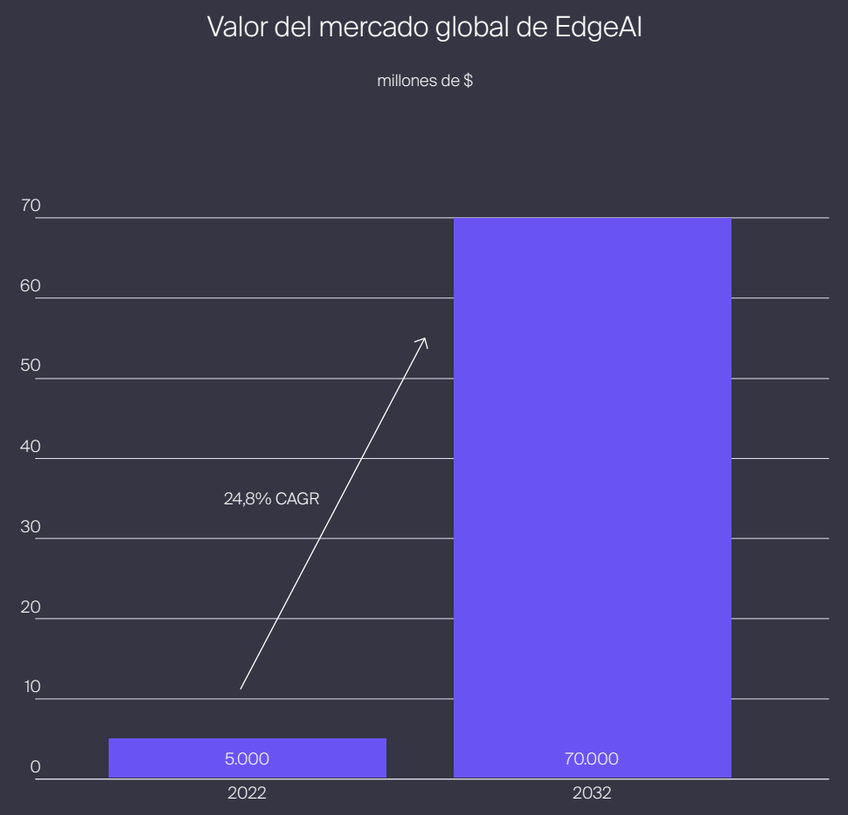
06. AI Ubicua/ Edge AI

La revolución provocada por la IA generativa está dando lugar a una nueva percepción de lo que va a ser el futuro tecnológico de la sociedad. **Los cambios experimentados a nivel de trabajo operativo e interacción humano-máquina nos permiten atisbar un nuevo paradigma en el que la IA se va a extender a todas las actividades cotidianas del ser humano**, estando presente en todas las aplicaciones y dispositivos conectados que utilizamos en el día a día.

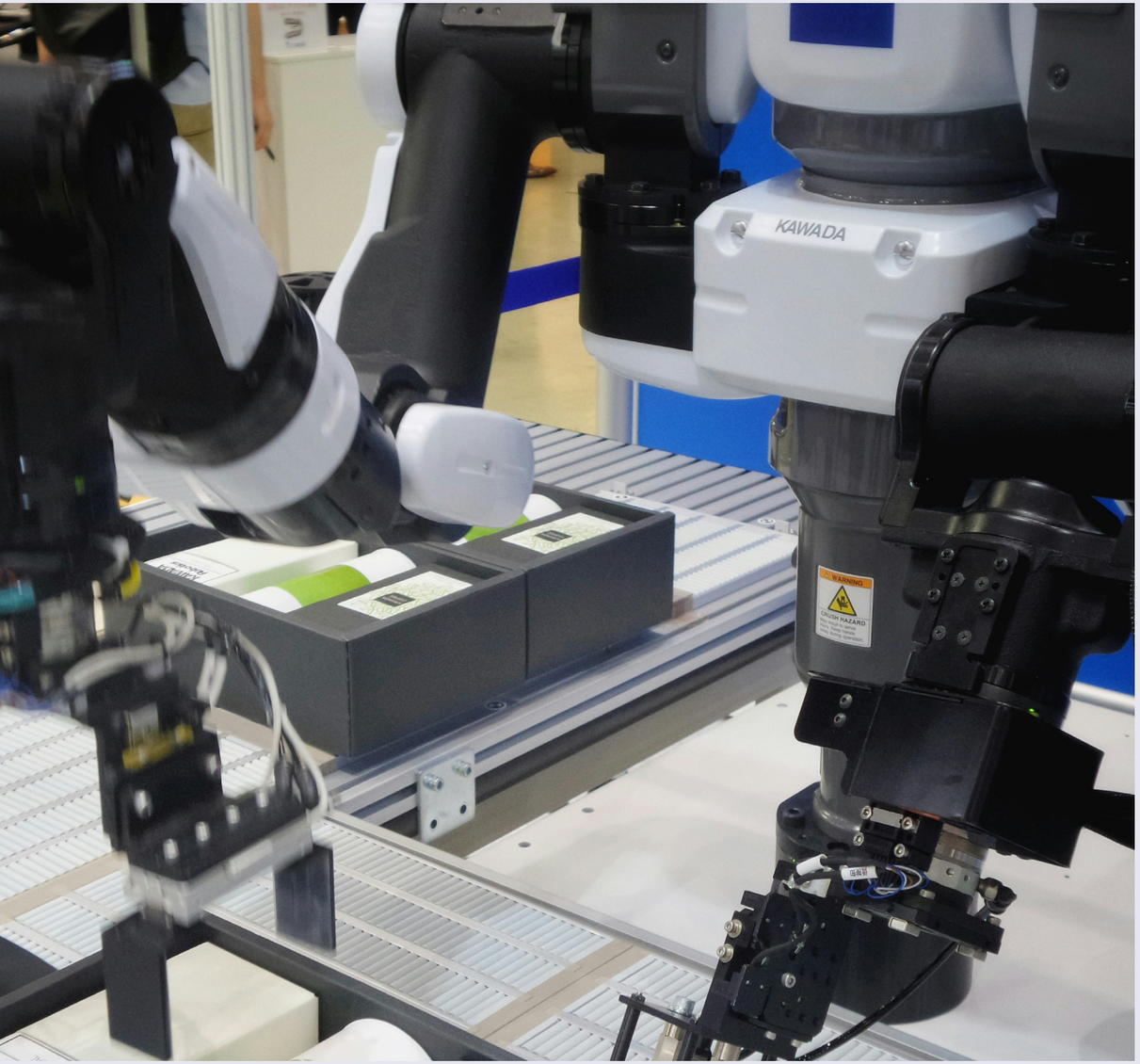
El tremendo impacto que supone pensar en esta nueva idea de inteligencia artificial ubicua, en la que toda nuestra actividad diaria está impulsada por algoritmos inteligentes que hace que se pueda apreciar en cualquier entorno la importancia de estructurar conjuntos de datos, acarrea la **necesidad de plantear nuevos enfoques en materia de datos.**

En este aspecto, va a cobrar mucha importancia el concepto de **Edge AI, dada su capacidad para permitir el procesamiento de datos en el punto mismo de su creación, ayudando a las organizaciones a obtener información en tiempo real, analizando datos lo más cerca posible al lugar en el que se originan.**

El concepto de Edge AI abarca todos aquellos despliegues de algoritmos y modelos de IA realizados directamente en dispositivos Edge. Es decir, desarrollados en los mismos sensores, cámaras, dispositivos conectados y, en definitiva, cualquier aparato conectado a una red IoT.



Fuente: Softtek en base a datos de Global Market Insights



Durante la próxima década veremos cómo el enfoque Edge AI va desplazando poco a poco a los tradicionales sistemas centralizados en la nube. Y especialmente se popularizará en casos de uso como:

- **Sistemas de seguridad:** a través de aplicaciones de vigilancia industrial equipadas con visión artificial o sistemas de alarma y detección inteligente de intrusiones.

- **Robots industriales:** maquinaria capaz de realizar tareas de manera autónoma y adaptarse a cambios en el entorno sin necesidad de intervención humana.

- **Dispositivos de salud inteligente:** aparatos médicos conectados que procesan datos en tiempo real permitiendo diagnósticos y tratamientos automáticos con gran nivel de precisión y personalización.

- **Vehículos autónomos:** prestaciones que permiten a los sistemas del vehículo tomar decisiones con mayor capacidad de precisión y agilidad que los conductores humanos, como evitar colisiones y cambiar de carril, mejorando la seguridad en la carretera.

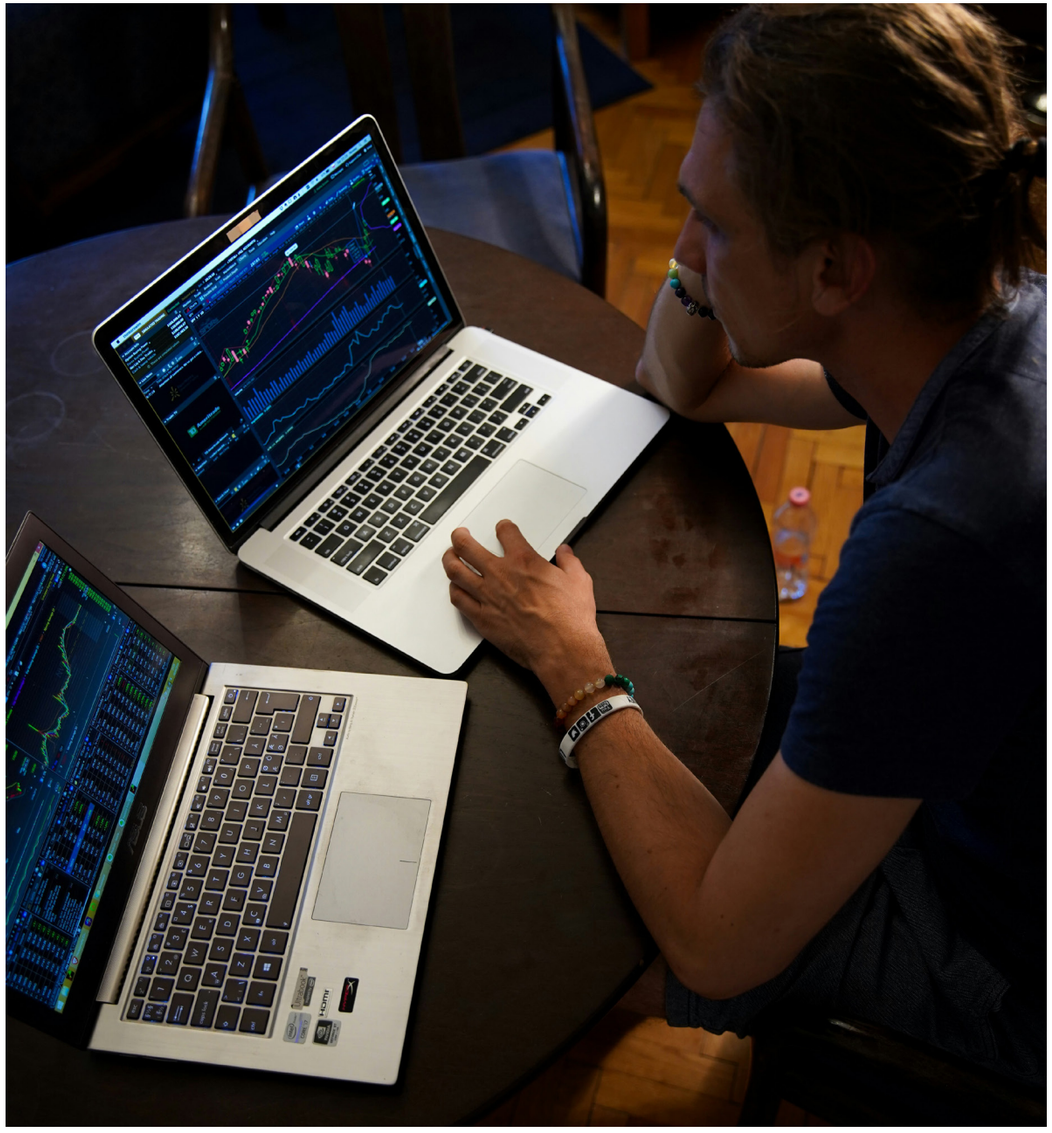
- **Sistemas de control de procesos:** todo tipo de herramientas de análisis y control basadas en redes de dispositivos IoT que pueden aplicarse con múltiples finalidades de seguridad, eficiencia operativa, sostenibilidad u optimización de la productividad.

La principal variable que dirigirá el crecimiento del enfoque EdgeAI va a ser la adopción de las redes 5G a un nivel masivo. Las redes de comunicación wireless de última generación van a proporcionar una conectividad mucho más rápida y fiable, reduciendo la latencia en la transmisión y el procesamiento de los datos generados en las redes de dispositivos edge. Adicionalmente, la tecnología 5G va a ofrecer nuevas prestaciones de seguridad avanzada que facilitarán la transmisión de datos encriptados, otro aspecto crucial para que las aplicaciones EdgeAI puedan llevarse de forma real a la práctica en áreas donde se manejan datos muy sensibles como pueden ser los sectores sanitarios, financiero, gubernamental, de transporte o de defensa.

Evolución

La IA Ubicua se refiere al **despliegue de algoritmos de procesamiento de datos e IA directamente en los dispositivos en el borde de la red, en lugar de en una nube centralizada o centro de datos.** Este enfoque permite el procesamiento de datos en tiempo real, la reducción de la latencia y la mejora de la privacidad y la seguridad de los datos confidenciales. Es particularmente crucial en escenarios donde la toma de decisiones rápida es esencial, como los vehículos autónomos, el monitoreo de la atención médica y la automatización industrial.

El viaje hacia la IA Ubicua comenzó con la centralización de los datos y la potencia de procesamiento en grandes centros de datos y nubes. Sin embargo, la explosión en el número y la variedad de dispositivos inteligentes que generan datos en el borde de las redes requirió un cambio. La necesidad de procesamiento en tiempo real, las limitaciones de ancho de banda y las preocupaciones sobre la privacidad han impulsado la evolución de la IA perimetral. Esta evolución está marcada por la miniaturización del hardware, los avances en los algoritmos de IA y el desarrollo de marcos y plataformas diseñados específicamente para la computación periférica.



Background tecnológico

- **Infraestructura de Edge Computing:** modelo de procesamiento que permite que los datos producidos por sistemas se procesen donde se generan para evitar enviar esta información a infraestructuras Cloud lejanas, provocando problemas de lactancia y seguridad. Puede entenderse como la implementación pequeños Data Centers distribuidos. Para ello, se necesita el hardware y la arquitectura de red necesarios para procesar los datos en el borde, los dispositivos pueden ir desde simples sensores hasta potentes servidores perimetrales.

- **Modelos de IA optimizados:** los modelos de IA para el perímetro se diseñan o adaptan específicamente para requerir menos potencia computacional y memoria, utilizando técnicas como la compresión, la cuantificación y la poda de modelos, anteriormente señalados.

- **Estrategias de datos distribuidos:** esto implica la gestión y el procesamiento de datos a través de múltiples dispositivos perimetrales y, posiblemente, la sincronización con sistemas centralizados, lo que requiere técnicas sólidas de gestión y sincronización de datos.

Desafíos

Limitaciones de recursos: los dispositivos perimetrales suelen tener una potencia de procesamiento y memoria limitadas. Por ello, la solución está en la optimización de modelos y algoritmos de inteligencia artificial para mejorar la eficiencia o en la consideración de nuevos paradigmas informáticos como la computación neuromórfica.

Seguridad y privacidad de los datos: para garantizar la seguridad de los datos en el perímetro se utilizan técnicas de uso de enclaves seguros para cálculos confidenciales, mientras que para los datos en tránsito se utiliza la implementación de cifrado de extremo a extremo.

Conectividad intermitente: los dispositivos perimetrales no siempre están conectados a una red, por ello, el diseño debe contemplar sistemas que puedan funcionar de forma independiente y sincronizarse con los sistemas centrales cuando la conectividad esté disponible.

Herramientas analíticas avanzadas en IA de datos y Edge AI

- **Marcos de procesamiento de flujos:** son esenciales para analizar datos en tiempo real en el perímetro. Pueden manejar flujos de datos de alto rendimiento y proporcionar información oportuna.

- **Federated Learning:** se trata de un enfoque de aprendizaje automático en el que el modelo se entrena en muchos dispositivos descentralizados que contienen muestras de datos locales, lo que mejora el rendimiento del modelo sin comprometer la privacidad.

- **Edge Analytics Databases:** estas bases de datos especializadas están diseñadas para ejecutarse en dispositivos perimetrales, lo que proporciona un almacenamiento y recuperación rápida de datos para el análisis en tiempo real.

Tendencias

- **Integración 5G:** mejorará significativamente las capacidades de IA perimetral, al proporcionar una conectividad más rápida y confiable para los dispositivos periféricos, lo que permitirá aplicaciones de IA más complejas y colaborativas.

- **Autonomous Edge Devices:** serán más autónomos, capaces de tomar decisiones y actuar sobre ellas de forma independiente, lo que reducirá aún más la necesidad de supervisión central.

- **AI-Powered Edge Security:** el futuro de la seguridad se centrará en medidas de seguridad impulsadas por la inteligencia artificial en el borde, proporcionando mecanismos de defensa dinámicos y en tiempo real.

Una empresa de fabricación industrial implementó Edge AI para supervisar el estado de los equipos en tiempo real. Los sensores de la maquinaria recopilan datos sobre las métricas de rendimiento, que son procesados in situ por dispositivos periféricos utilizando modelos de IA para predecir posibles fallos. Este enfoque permite una intervención inmediata, reduciendo el tiempo de inactividad y los costes de mantenimiento.

Key insight

Data & Edge AI está transformando cómo y dónde procesamos y analizamos los datos. AI acercar las capacidades de IA al lugar donde se generan los datos, **permiten un procesamiento de datos más rápido, eficiente y privado.**

A medida que la tecnología continúa avanzando, las aplicaciones potenciales de Data & Edge AI se expandirán, ofreciendo oportunidades sin precedentes para la innovación en varios sectores.

07. Data Masking

La necesidad de las organizaciones de proteger la integridad, la confidencialidad y la privacidad de los datos se va a volver cada vez más compleja en el nuevo paradigma de los negocios inteligentes, y técnicas como el **Data Masking y el Dynamic Data Masking van a continuar emergiendo como tendencia debido a su capacidad para ofrecer una solución efectiva de protección de datos sensibles** durante el desarrollo, las pruebas y el intercambio de información. Asimismo, estas técnicas se pueden utilizar para la simplificación del diseño y la codificación de la seguridad. Aplicando técnicas de Data Masking, las organizaciones pueden enmascarar u ocultar la información en una base de datos o entorno de prueba con la finalidad de limitar la exposición de los datos sensibles a violaciones, brechas de seguridad o incumplimientos normativos. Esto significa que, aunque los datos sigan almacenados sin cambios, su acceso se limita y se revela solo la cantidad mínima necesaria para el usuario.

El enmascaramiento de datos se puede estructurar en dos grandes bloques: enmascaramiento de datos estático y dinámico. **El enmascaramiento de datos estático (SDM) reemplaza permanentemente los datos sensibles al alterar los datos en reposo, mientras que el enmascaramiento dinámico de datos (DDM) tiene como objetivo reemplazar los datos confidenciales en tránsito dejando los datos originales en reposo intactos y sin cambios.**

Evolución

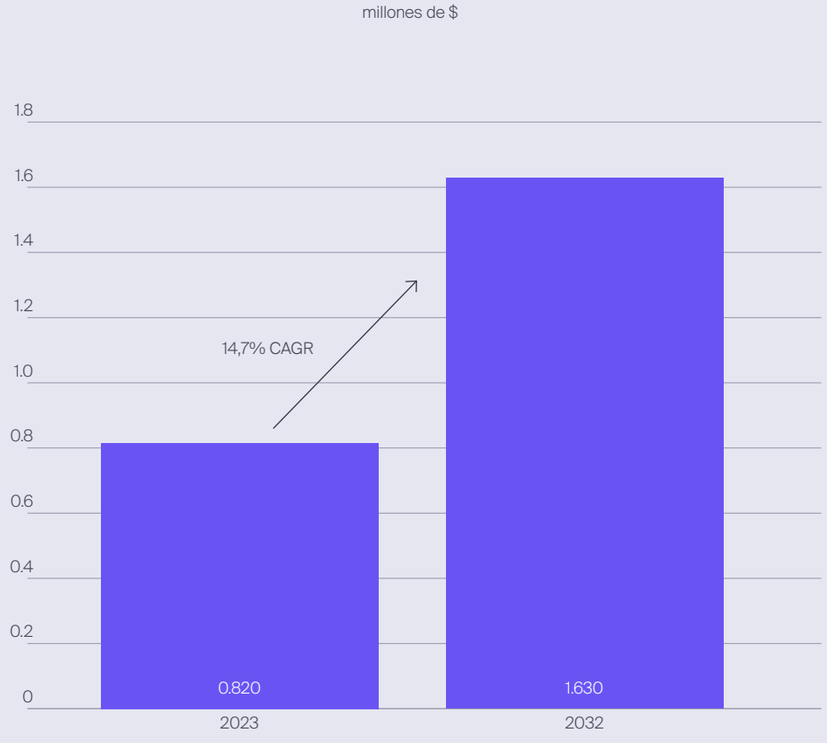
El enmascaramiento de datos es la evolución de las técnicas de encriptación. La principal diferencia entre ambas técnicas es que la encriptación protege los datos contra un ataque a la privacidad cuando está en reposo, pero para procesarlos, analizarlos o manipularlos es necesario proceder a su desencriptación, lo que genera una brecha en la seguridad. Sin embargo, **gracias al enmascaramiento de datos se pueden realizar operaciones con los datos enmascarados manteniendo su estructura y comportamiento.**

En los próximos años, el Data Masking se va a volver cada vez más relevante, en especial en los entornos de desarrollo y testing, donde los datos de producción se utilizan para pruebas, pero es crucial proteger la información confidencial.

Las tendencias más avanzadas en esta tecnología incluyen una adopción creciente de soluciones de Data Masking alojadas en la nube, con un especial foco en las aplicaciones para cumplimiento regulatorio y las políticas de privacidad. De igual forma, las organizaciones están en busca de soluciones de Data Masking en tiempo real capaces de proteger los datos durante procesos de integración, compartición o migración.

Otra tendencia que veremos expandirse en la próxima década tendrá que ver con la **integración de algoritmos de inteligencia artificial y Machine Learning** diseñados para aumentar la eficiencia y la eficacia de las técnicas de Data Masking, capaces de escalar el volumen de datos manejados y reducir la necesidad de supervisión.

Tamaño de mercado global de herramientas y soluciones de Data Masking



Fuente: Softtek en base a datos de Mordor Intelligence

Background tecnológico

El objetivo último es garantizar que datos privados, como son las credenciales financieras, números de cuenta, identificadores fiscales o direcciones de correo, no puedan quedar expuestas a personas no autorizadas o usos ilícitos. Para ello, abarca el empleo de diferentes técnicas:

- **Enmascaramiento estático:** consistente en reemplazar los datos originales con datos ficticios o modificados. Por ejemplo, ocultar los primeros 12 dígitos de un número de tarjeta de crédito y mostrar solo los últimos 4 dígitos.

- **Enmascaramiento dinámico:** de aplicación en tiempo real, donde los datos originales se reemplazan por datos ficticios cada vez que se acceden o se muestran.

- **Enmascaramiento de subconjuntos:** consistente en ocultar solo ciertas partes de los datos mientras se mantienen otras partes visibles. Esto puede ser útil para mantener la integridad de los datos sin revelar información sensible.

- **Enmascaramiento de características:** reemplaza los datos originales con valores similares, pero no idénticos. Por ejemplo, modificar la edad de una persona sumando o restando un pequeño número aleatorio.

- **Enmascaramiento con mantenimiento de formato:** conserva el formato y la apariencia de los datos originales, pero los datos reales se reemplazan por valores ficticios o pseudoanonimizados. Este modelo se espera que sea el que mayor auge experimente en el futuro próximo.

A través de estas técnicas, las organizaciones pueden crear representaciones ficticias pero realistas de los datos de la compañía, desarrollando una nueva versión, imposible de interpretar o descifrar a través de ingeniería inversa.

Desafíos

Complejidad en la implementación

- **Detección y enmascaramiento automatizados:** mediante herramientas que identifican de forma automática los datos confidenciales y aplican las reglas de enmascaramiento más adecuadas para cada situación, ayudando a simplificar el proceso de implementación.

- **Técnicas de enmascaramiento personalizables:** permiten a las organizaciones elegir el método que mejor se adapta a su tipo específico de dato y a su caso de uso.

Gastos generales de rendimiento

- **Algoritmos optimizados:** el empleo de algoritmos eficientes que minimicen la sobrecarga computacional es vital, especialmente para el enmascaramiento dinámico de datos.

- **Enmascaramiento selectivo:** permite enmascarar solo los datos necesarios, reduciendo significativamente el impacto en el rendimiento.

Mantenimiento de la utilidad de datos

- **Cifrado de preservación de formato (FPE):** este método cifra los datos de manera que conserva el formato de datos original, lo que garantiza que los datos enmascarados puedan seguir siendo útiles para ciertos tipos de análisis.

- **Generación de datos sintéticos:** consiste en generar datos sintéticos completamente nuevos basados en las características de los datos originales. Se puede utilizar para pruebas y desarrollo sin arriesgarse a exponer información confidencial.

Herramientas analíticas avanzadas en el enmascaramiento de datos

- **Context-Aware Masking:** las herramientas avanzadas comprenden el contexto de los datos y aplican diferentes reglas de enmascaramiento en función de la confidencialidad de los datos, el rol del usuario y el escenario específico en el que se utilizan los datos.

- **IA y aprendizaje automático:** estas tecnologías se pueden utilizar para mejorar la identificación de datos confidenciales y para crear técnicas de enmascaramiento más sofisticadas que preserven mejor la utilidad de los datos.

- **Enmascaramiento de datos como servicio (DMaaS):** algunos proveedores ofrecen el enmascaramiento de datos como un servicio basado en la nube, lo que simplifica el proceso de implementación y proporciona acceso a las últimas tecnologías y técnicas.

Tendencias

- **Aumento de la regulación para la protección de datos:** a medida que las regulaciones de privacidad se vuelven más estrictas, las técnicas y herramientas que permiten el enmascaramiento y la protección de datos evolucionarán para liderar soluciones que permitan aprovechar al máximo el poder del dato y cumplir con las legislaciones vigentes.

- **Integración con Data Fabric:** la integración de las capacidades de enmascaramiento de datos será crucial para garantizar que los datos confidenciales estén protegidos en todo el panorama de datos.

- **Avances en el enmascaramiento en tiempo real:** la necesidad de acceso a datos en tiempo real en el análisis y la toma de decisiones impulsará los avances en el enmascaramiento dinámico de datos, haciéndolo más eficiente y menos intrusivo.

Una empresa de atención médica necesita compartir datos de pacientes con instituciones de investigación para realizar estudios. Al emplear el enmascaramiento de datos, pueden proporcionar conjuntos de datos en los que se oculta la información de identificación personal (PII), lo que garantiza la privacidad del paciente y, al mismo tiempo, contribuye con datos valiosos a la investigación médica.

Key insight

El enmascaramiento de datos es una estrategia esencial para proteger la información confidencial y mantener la privacidad. Al ocultar elementos de datos específicos, las organizaciones pueden compartir y utilizar datos de forma segura, satisfaciendo tanto las necesidades operativas como los requisitos de cumplimiento.

A medida que avanza la tecnología, también lo harán las técnicas y herramientas para un enmascaramiento de datos eficaz, lo que **garantiza que la privacidad de los datos siga el ritmo de la dependencia cada vez mayor de la información y la toma de decisiones basadas en datos.**

08. DataOPs

Debido a la creciente necesidad de agilidad en el manejo de datos, y la mayor complejidad de integración de sistemas para la colaboración interfuncional, **DataOps se va a volver una técnica cada vez más necesaria para las organizaciones.**

Combinar datos y operaciones está demostrando ser una práctica eficaz para diseñar e implementar pipelines de datos de una forma más eficiente y colaborativa, y se espera que cada vez se utilice más para mejorar la entrega, calidad y disponibilidad de los datos, al tiempo que se acelera el ciclo de vida de desarrollo y las operaciones en entornos de análisis, optimizando las recopilación y preparación de los datos.

Por otro lado, **la observabilidad de datos proporciona la visibilidad y los insights necesarios sobre los flujos de datos que se generan en una organización**; y DataOps ofrece el marco y los procesos para operar, mejorar y automatizar dichos flujos.

De esta forma, DataOps se plantea como un **enfoque metodológico complementario a DevOps**, que busca optimizar y agilizar la entrega y gestión de desarrollos basados en sistemas o aplicaciones de datos. Además, facilita que los equipos de ingeniería de datos proporcionen a los analistas, y otros interesados, datos fiables para impulsar la posterior toma de decisiones.

La gran ventaja de aplicar un enfoque DataOps para la gestión de datos es **que impulsa enormemente la colaboración entre diferentes funciones dentro de la organización**, contribuyendo a eliminar los tradicionales silos de datos, evitando que los ciclos de desarrollo se prolonguen innecesariamente y, en consecuencia, que los proyectos pierdan agilidad.

Evolución

Originalmente derivado de los principios del desarrollo ágil, DevOps y el control estadístico de procesos, DataOps se ha convertido en una disciplina distinta. Está impulsado por la **necesidad de un procesamiento y análisis de datos más dinámicos y en tiempo real** en una era en la que los volúmenes y la complejidad de los datos siguen creciendo exponencialmente.

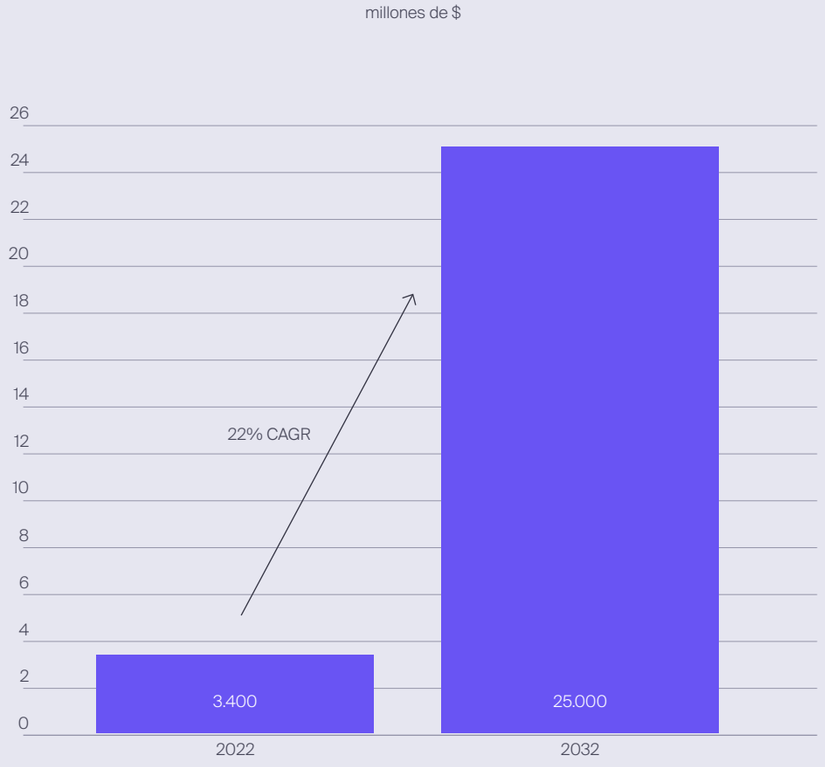
Los avances que experimentarán las técnicas de DataOps en los próximos años van a ir de la mano de la expansión de la inteligencia artificial y los algoritmos de aprendizaje automático, dando lugar a **plataformas de última generación que incluyen prestaciones de analítica predictiva.**

Por otro lado, veremos cómo estas nuevas soluciones se presentan mayoritariamente alojadas en la nube, con la finalidad de permitir a las organizaciones una mayor escalabilidad, flexibilidad y reducción de costes. Estas **plataformas Cloud de DataOps** ofrecen además una mayor compatibilidad y capacidad de integración con diferentes formatos y fuentes de datos en entornos distribuidos.

Las prácticas DataOps de colaboración interfuncional continuarán emergiendo enfocadas a promover las **sinergias existentes entre los equipos de ingenieros y científicos de datos y las áreas puramente operativas.**

En última instancia, un conocimiento cada vez mayor de los beneficios potenciales del DataOps para la optimización de los flujos de datos, y la necesidad de reducir el tiempo de respuesta y mejorar la calidad de los datos harán que se dispare la demanda de estas soluciones avanzadas.

Tamaño de mercado global de plataformas DataOps



Fuente: Softtek en base a datos de Global Market Insights

Background tecnológico

Pruebas automatizadas de calidad de datos

- **Aprendizaje automático para la detección de anomalías:** los algoritmos avanzados pueden detectar automáticamente anomalías e inconsistencias en los datos, lo que reduce significativamente el tiempo y el esfuerzo necesarios para garantizar la calidad de los datos.

- **Métricas continuas de calidad de datos:** proporcionan monitoreo e informes continuos de la calidad de los datos, lo que garantiza que los problemas se identifiquen y aborden con prontitud.

Advanced Data Pipeline Orchestration

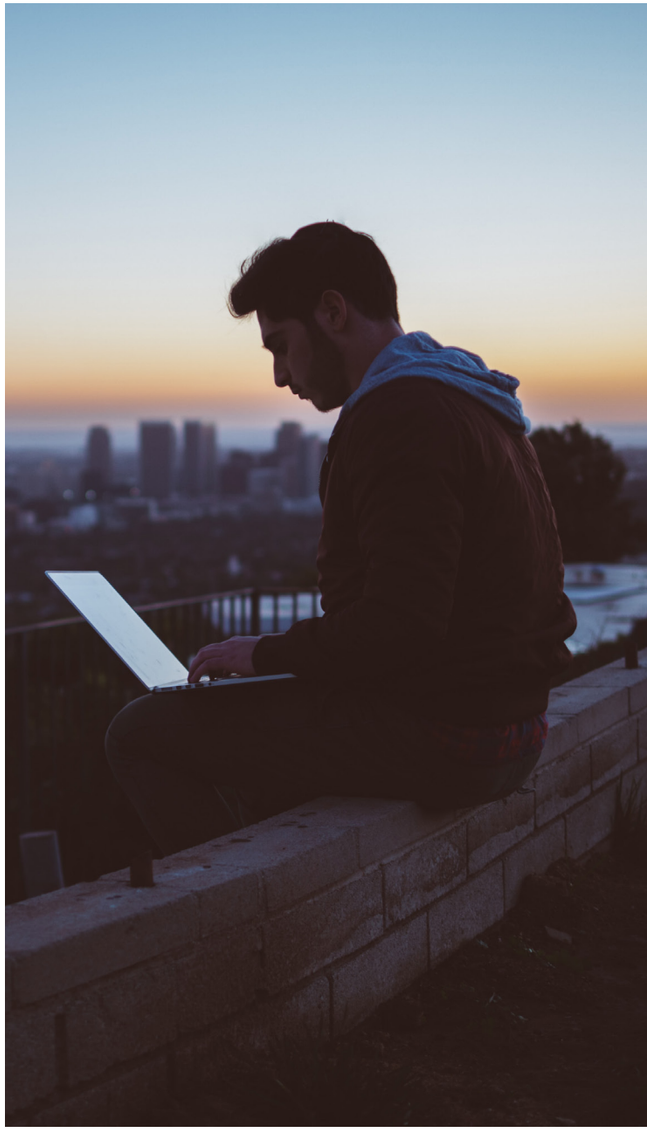
- **Dynamic Pipeline Adjustment:** en función de los flujos de datos en tiempo real y la demanda de los usuarios, garantiza un rendimiento y una utilización de recursos óptimos.

- **Self-Healing Data Pipelines:** las innovaciones en IA permiten que las canalizaciones de datos detecten problemas y los corrijan de forma autónoma, minimizando el tiempo de inactividad y la intervención manual.

Herramientas de colaboración mejoradas

- **Entornos de desarrollo integrados (IDE) para DataOps:** estos IDE especializados ofrecen funciones adaptadas a los equipos de DataOps, como control de versiones, conjuntos de pruebas y herramientas de colaboración, todo ello dentro de un entorno unificado.

- **ChatOps para DataOps:** la integración de las plataformas de chat con las herramientas de DataOps permite a los equipos colaborar de manera más efectiva, con bots y alertas automatizadas que mantienen a todos informados y comprometidos.



Desafíos

Complejidad en la integración

- **Estandarización de herramientas y procesos:** la adopción de herramientas estándar de la industria y la definición de procesos claros pueden ayudar a reducir la complejidad y la curva de aprendizaje asociadas con la integración de nuevas innovaciones de DataOps.

- **Arquitectura modular:** el diseño de sistemas y herramientas de datos de forma modular permite una integración y sustitución más sencillas de los componentes según sea necesario.

Mantenimiento de la seguridad y la gobernanza de los datos

- **Supervisión automatizada del cumplimiento:** supervisar e informar automáticamente sobre el cumplimiento de las políticas de gobernanza de datos, lo que garantiza que las innovaciones no provoquen fallos en la seguridad o el cumplimiento.

- **Controles de acceso basados en roles en las herramientas**: garantizar que las herramientas de DataOps apliquen controles de acceso estrictos ayudará a mantener la seguridad de los datos incluso a medida que crezca el número y la complejidad de las operaciones de datos.

Liderar la innovación

- **Aprendizaje y desarrollo continuos:** fomentar una cultura de aprendizaje continuo dentro de los equipos de DataOps garantiza que se mantengan actualizados con las últimas innovaciones.

- **Asociaciones con proveedores de tecnología:** colaborar estrechamente con los proveedores de herramientas y plataformas de DataOps puede ayudar a las organizaciones a mantenerse a la vanguardia e influir en el desarrollo de nuevas funciones que satisfagan sus necesidades.

Tendencias

- **Integración de inteligencia artificial y aprendizaje automático:** a medida que evolucionen se integrarán más en DataOps aumentando sus capacidades de procesamiento y gestión de datos, dando lugar a plataformas de última generación en analítica predictiva.

- **DataOps as a Service (DaaS):** las plataformas DataOps basadas en la nube ofrecerán a las organizaciones la capacidad de implementar estrategias sofisticadas de DataOps sin la necesidad de una amplia infraestructura y experiencia interna.

- **Operaciones predictivas de datos:** aprovechando el análisis predictivo, las futuras herramientas de DataOps no solo reaccionarán a los problemas a medida que ocurran, sino que los anticiparán y prevendrán antes de que sucedan, mejorando aún más la eficiencia y confiabilidad de las operaciones de datos.

Una empresa de servicios financieros implementa una estrategia avanzada de DataOps para operar datos de mercado en tiempo real, al emplear el aprendizaje automático para la detección de anomalías, reducen significativamente la incidencia de errores de datos que afectaban a sus algoritmos de trading. La orquestación automatizada de la canalización de datos les permite escalar sus capacidades de procesamiento de datos de forma dinámica, lo que garantiza que siempre puedan satisfacer la demanda de análisis en tiempo real.

Key insight

Las innovaciones recientes en DataOps están transformando el panorama de la gestión y el análisis de datos. Al adoptar estos avances, **las organizaciones pueden garantizar que sus operaciones de datos sean lo más eficientes, confiables y efectivas posible**, lo que les permite obtener el máximo valor de sus activos de datos.

A medida que DataOps siga evolucionando, sin duda desempeñará un papel cada vez más crucial en la toma de decisiones basada en datos que sustenta las estrategias empresariales modernas.

09. La revolución Web3 y el almacenamiento descentralizado

La expansión de la Web3 va a suponer una transformación integral de Internet, que dejará de funcionar como un entorno controlado y monolítico para pasar a ser un **ecosistema seguro, descentralizado y focalizado en el usuario.** Este cambio de paradigma se va a producir como consecuencia de la expansión de Blockchain, las DApps (aplicaciones descentralizadas) y la popularización de los activos criptográficos como medio de intercambio online entre particulares u organizaciones.

La llegada de esta nueva fase en la evolución de Internet va a permitir que las personas tomen un mayor control de sus interacciones digitales, eliminando intermediarios y, por tanto, preservando la propiedad de los datos que se generan.

Evolución

La Web3 surge como respuesta a estos desafíos, abogando por un **Internet descentralizado en la que los usuarios controlen sus datos**. En este paradigma, el almacenamiento descentralizado no es solo una opción, sino un componente fundamental que garantiza que los datos permanezcan seguros y accesibles de manera distribuida.

Durante la próxima década se espera que el **mercado generado a través de la Web3 experimente un crecimiento exponencial con una tasa CAGR del 44,9% interanual.**

En términos de almacenamiento, esta tendencia va a suponer también un cambio de enormes dimensiones ya que las **personas o grupos de individuos pueden contribuir con su espacio para crear nuevas redes de almacenamiento que escapen al control de las nubes centralizadas actuales** (como AWS o Google Cloud).

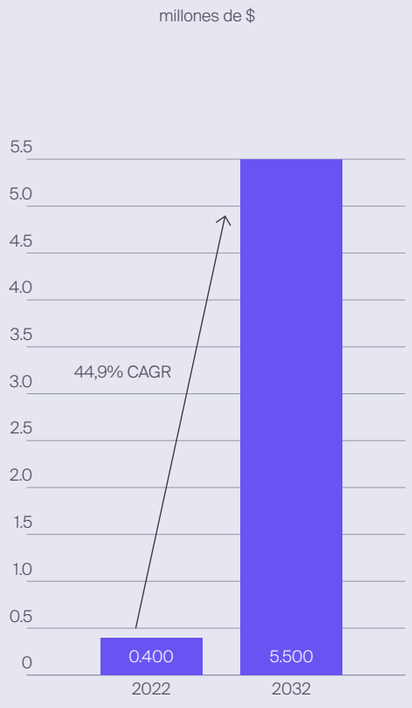
Este fenómeno derivará en una **reducción drástica de los costes de almacenamiento de datos** y en la creación de nuevas economías de mercado descentralizadas, basadas en el comercio de datos o criptoactivos en forma de tokens o NFTs. Por otro lado, la tecnología Blockchain que sustenta la Web3 ofrece la posibilidad de crear registros inmutables y transparentes de cualquier operación o transacción. Esto no solo aumenta la seguridad, sino que es un método muy eficaz para **garantizar la trazabilidad de todo tipo de procesos comerciales y de calidad.**

Por ejemplo, la posibilidad de realizar un seguimiento completo y fiable de todas las etapas y procesos dentro de la cadena de suministro de un producto, a través de contratos inteligentes inalterables y accionables únicamente bajo la conformidad de todas las partes implicadas, va a provocar un gran cambio en la forma en que se registran y gestionan los datos referentes a certificaciones de calidad, sostenibilidad, materiales, etc.

Por otro lado, la Web3 va a promover una **mayor interoperabilidad entre aplicaciones y plataformas**, al no depender de formatos propietarios. Esto va a facilitar el intercambio y la integración de datos entre múltiples sistemas, mejorando la eficiencia y la colaboración. A nivel de monetización, los usuarios individuales van a encontrar más oportunidades para monetizar directamente sus datos. Esta participación de las personas en las economías descentralizadas (DeFi) va a suponer la llegada de una nueva competencia para las marcas que actualmente cuentan con el monopolio de los grandes volúmenes de datos de clientes.

Sin embargo, la Web3 todavía tendrá que solventar importantes retos en términos de **gobernanza y cumplimiento regulatorio** y, para que muchos de estos modelos puedan implementarse de forma práctica, es necesario que se cierren ciertas **lagunas legislativas, como las que existen,** por ejemplo, en el ámbito financiero.

Volumen de mercado global generado a través de la web 3.0



Fuente: Softtek en base a datos de Markets & Markets

Background tecnológico

Soluciones de almacenamiento basadas en Blockchain

- **Immutable Ledgers for Metadata:** Blockchain proporciona un registro inmutable de metadatos, lo que garantiza la integridad y la trazabilidad de los datos almacenados en toda la red.

- **Smart Contracts for Data Management:** son contratos autoejecutables con los términos escritos directamente en el código, lo que permite transacciones de datos automatizadas, descentralizadas y seguras.

InterPlanetary File System (IPFS)

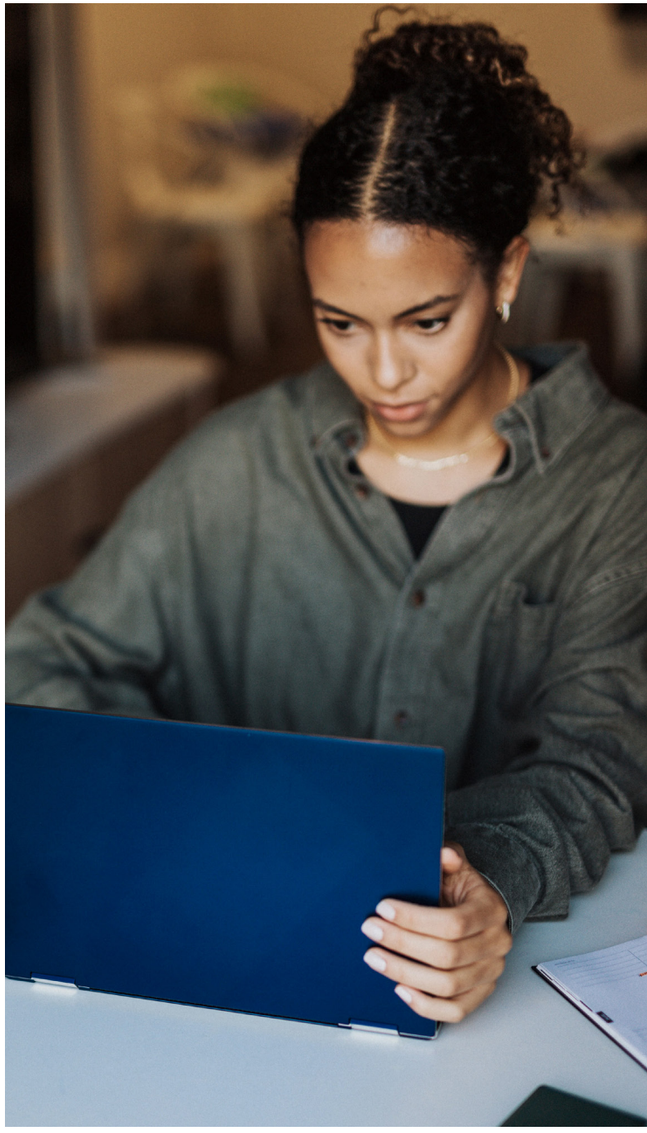
- **Content Addressing:** IPFS utiliza un modelo de almacenamiento en bloque con direccionamiento de contenido, en el que los archivos se direccionan mediante el hash de su contenido, lo que garantiza que los datos sean identificables de forma única y resistentes a la manipulación.

- **Distributed Hash Tables (DHT):** IPFS utiliza DHT para el enrutamiento eficiente y la localización de datos a través de una red distribuida, lo que mejora la accesibilidad y la resiliencia de los datos.

Identificadores descentralizados (DID)

- **Self-Sovereign Identity:** los DID proporcionan un nuevo tipo de identificador que permite identidades digitales verificables y autosoberanas, lo que permite a los usuarios controlar sus datos sin depender de una autoridad central.

- **Interoperabilidad:** los DID están diseñados para ser interoperables en diferentes redes y sistemas, lo que garantiza que los usuarios puedan acceder y gestionar sin problemas sus datos a través de varios servicios Web3.



Desafíos

Escalabilidad

- **Sharding:** consiste en dividir los datos en partes más pequeñas y distribuirlos por toda la red, lo que permite el procesamiento paralelo y mejora la escalabilidad.

- **Layer 2 Solutions:** tecnologías como los state channels y las sidechains pueden ayudar a escalar las soluciones de almacenamiento descentralizadas mediante el manejo de transacciones fuera de la cadena de bloques principal.

Eficiencia en la recuperación de datos

- **Mecanismos de almacenamiento en caché:** la implementación de estrategias de almacenamiento en caché en los nodos estratégicos de la red garantiza que los datos a los que se accede con frecuencia se puedan recuperar rápidamente.

- **Modelos de incentivos**: alentar a los nodos a almacenar y servir datos de manera eficiente a través de incentivos económicos puede mejorar significativamente los tiempos de recuperación de datos.

Privacidad y seguridad

- **Cifrado:** el cifrado de extremo a extremo garantiza que los datos permanezcan privados y seguros a medida que se almacenan y recuperan de la red descentralizada.

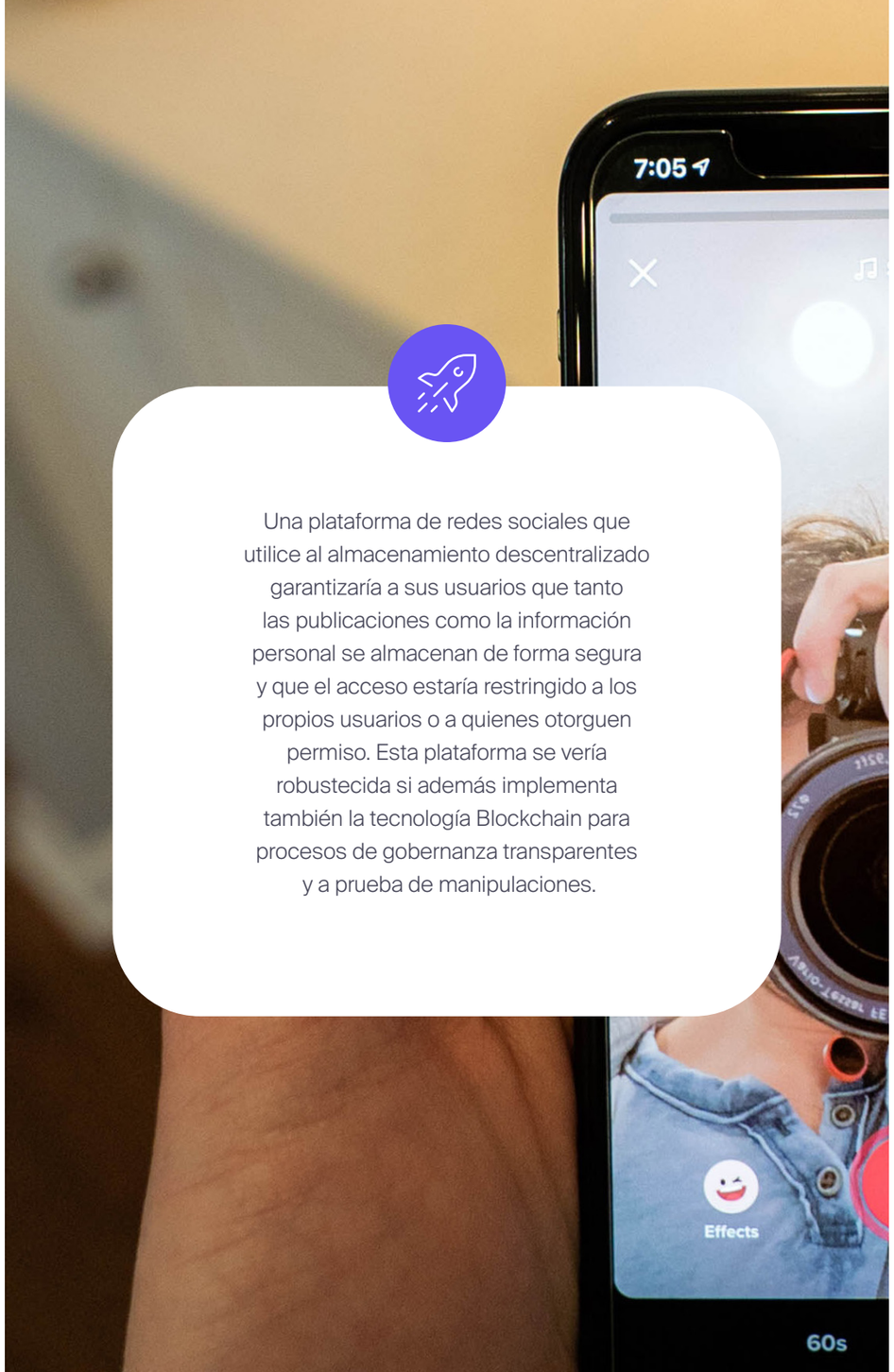
- **Zero-Knowledge Proofs:** permiten a una parte demostrar a otra que una declaración es verdadera sin revelar ninguna información más allá de la validez de la declaración en sí.

Tendencias

- **Mainstream Adoption:** a medida que las herramientas y plataformas se vuelven más fáciles de utilizar y prácticas a la hora de abordar los desafíos normativos y técnicos, se espera ver una adopción más generalizada de Web3 y almacenamiento descentralizado.

- **Integración con IA y aprendizaje automático:** el almacenamiento descentralizado se utilizará cada vez más para almacenar y compartir de forma segura las grandes cantidades de datos necesarios para entrenar modelos de IA y aprendizaje automático.

- **Evolución del Data Economy:** la combinación de Web3 y almacenamiento descentralizado dará lugar a nuevos mercados y economías de datos, en los que los usuarios tendrán más control sobre sus datos y podrán monetizarlos de forma segura y transparente.



Una plataforma de redes sociales que utilice al almacenamiento descentralizado garantizaría a sus usuarios que tanto las publicaciones como la información personal se almacenan de forma segura y que el acceso estaría restringido a los propios usuarios o a quienes otorguen permiso. Esta plataforma se vería robustecida si además implementa también la tecnología Blockchain para procesos de gobernanza transparentes y a prueba de manipulaciones.

Key insight

La Web3 y el almacenamiento descentralizado están sentando las bases para una nueva era en la gestión de datos, en la que **la descentralización, el control de los usuarios y la seguridad son primordiales.**

Estas tecnologías ofrecen una solución prometedora a las limitaciones del almacenamiento centralizado tradicional, proporcionando un **ecosistema de datos más resistente, privado y centrado en el usuario.** A medida que la tecnología madura y crece la adopción, podemos esperar un profundo impacto en la forma en que se almacenan, administran y utilizan los datos en Internet.

10. Quantum Technologies

La tecnología cuántica está sustentada en dos principales conceptos: entrelazamiento y superposición. **La superposición sostiene que la materia puede existir simultáneamente en más de un estado cuántico mientras que la teoría del entrelazamiento postula que las partículas se vuelven dependientes de los estados cuánticos de otras**, comportándose como entidades individuales. Esto se traduciría, un futuro, un avance significativo en las capacidades de computación y gestión de datos. A diferencia de la computación clásica basada en bits, **la computación cuántica utiliza bits cuánticos o qubits, que pueden representar tanto 0 como 1 simultáneamente.** Esta capacidad, junto con fenómenos como el entrelazamiento y la superposición, permite a los ordenadores cuánticos procesar grandes cantidades de datos a velocidades sin precedentes, ofreciendo enfoques revolucionarios para la gestión y la seguridad de los datos.

La mecánica cuántica, el principio subyacente, se ha explorado desde principios del siglo XX, pero solo en las últimas décadas hemos comenzado a aprovechar sus principios para la computación y la gestión de datos. **Las innovaciones en algoritmos cuánticos, corrección de errores y hardware nos han acercado gradualmente a la realización del potencial de las tecnologías cuánticas en aplicaciones del mundo real.**

Evolución

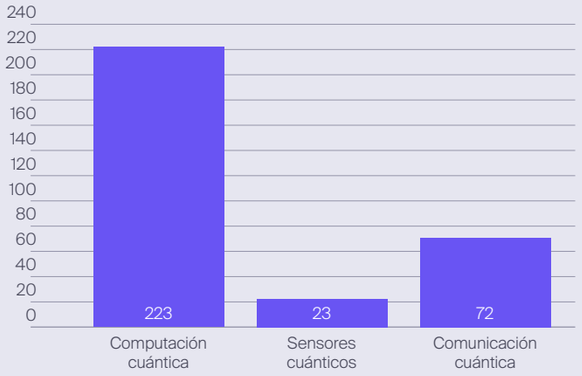
Aunque las aplicaciones de tecnología cuántica todavía se encuentran en una fase de desarrollo y experimentación, en los próximos años comenzaremos a ver los primeros pasos que revelan su potencial para provocar una verdadera disrupción en la gestión de datos. Los ordenadores cuánticos tienen **capacidad para realizar cálculos mucho más rápidos,** lo que ofrece una ventana de oportunidades para acelerar el procesamiento y análisis de grandes conjuntos de datos, impulsando enormemente las prácticas de **aprendizaje profundo y analítica predictiva basada en inteligencia artificial.** Por otro lado, también va a **transformar los estándares de ciberseguridad,** ya que ofrecen métodos de cifrado y seguridad avanzados que son inherentemente seguros contra ataques cibernéticos, lo que podría mejorar significativamente la protección de los datos almacenados en todo tipo de sistemas y redes. Prueba del prometedor potencial de esta nueva tecnología es el volumen de empresas emergentes que ya están trabajando en la experimentación con modelos de negocio aplicados a las áreas de computación, sensores y comunicaciones cuánticas:

La tecnología cuántica ofrece múltiples sinergias con los avances en construcción de redes IoT, y los ecosistemas de sensores conectados, ya **permiten acelerar exponencialmente la transmisión de datos, reducir la latencia y mejorar sustancialmente la calidad con la que procesan los datos en tiempo real.** Este terreno revolucionario va a abrir multitud de puertas para la experimentación en el ámbito de los sistemas de monitorización industrial, ambiental, científica y de vigilancia.

Otro espectro donde la tecnología cuántica tiene cabida es el de la simulación, el modelaje 3D y los gemelos digitales, ya que facilitará la **implementación de métodos más eficientes y precisos de optimización de procesos de fabricación o mantenimiento**, abriendo la puerta para un análisis y procesamiento de datos mucho más complejo y en tiempo real. Las próximas dos décadas serán cruciales para el desarrollo de esta tendencia emergente y las estimaciones apuntan a que veremos un crecimiento exponencial en el tamaño de mercado generado por soluciones basadas en tecnología cuántica:

Número de startups en el ámbito de la tecnología cuántica, a nivel global

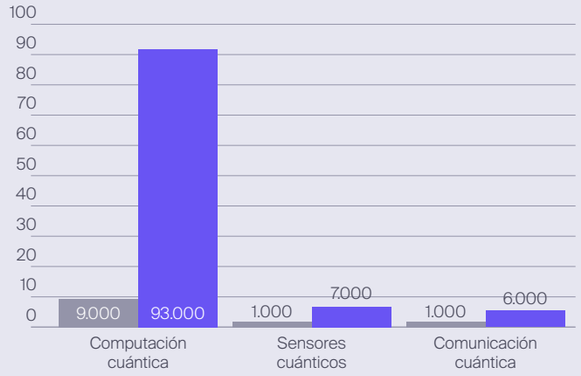
nº total por segmento, en 2022



Fuente: Softtek en base a datos de Statista

Estimación del tamaño de mercado global de tecnologías cuánticas, por segmento

millones de $



Fuente: Softtek en base a datos de Statista

Background tecnológico

Quantum Key Distribution (QKD)

- **Unbreakable Encryption:** QKD utiliza los principios de la mecánica cuántica para crear canales de comunicación seguros. En teoría, es imposible interceptar o descifrar datos cifrados con QKD sin alertar al remitente y al destinatario.

- **Early Adoption and Networks:** aunque todavía se encuentra en sus primeras etapas, QKD se ha implementado con éxito en algunas redes de comunicación especializadas, anunciando una nueva era de comunicación segura.

Algoritmos cuánticos para el procesamiento de datos

- **Aceleración de cálculos complejos:** algoritmos como el algoritmo de Shor para factorizar números grandes y el algoritmo de Grover para la búsqueda en bases de datos prometen completar tareas en una fracción del tiempo que tomarían las computadoras clásicas más poderosas.

- **Optimización de problemas a gran escala:** los algoritmos cuánticos son especialmente adecuados para problemas de optimización en logística, finanzas y ciencia de los materiales, donde pueden analizar y optimizar sistemas grandes y complejos de forma más eficiente que los algoritmos clásicos.

Aprendizaje automático cuántico

- **Handling Complex Datasets:** los algoritmos de aprendizaje automático cuántico pueden manejar conjuntos de datos complejos y de alta dimensión de manera más eficiente que los algoritmos clásicos.

- **Enhanced Pattern Recognition:** también pueden ser capaces de identificar patrones y correlaciones en los datos que están fuera del alcance de las computadoras clásicas.

Desafíos

Complejidad técnica:

- **Developing Robust Qubits:** la creación de Qubits estables que puedan mantener su estado cuántico (coherencia) el tiempo suficiente para realizar cálculos significativos es un desafío importante. La investigación en curso sobre varios tipos de tecnologías de Qubits tiene como objetivo superar esto.

- **Corrección de errores:** se están desarrollando técnicas de corrección de errores cuánticos para abordar los errores que surgen inherentemente en los cálculos cuánticos debido a la decoherencia y otros ruidos cuánticos.

Integración con sistemas clásicos:

- **Sistemas híbridos:** a corto plazo, es probable que las computadoras cuánticas se utilicen junto con las computadoras clásicas, creando sistemas híbridos que aprovechen las fortalezas de cada uno para diferentes tareas.

- **Estandarización e interoperabilidad:** el desarrollo de protocolos e interfaces estandarizados para la integración cuántica-clásica es crucial para la implementación práctica de las tecnologías cuánticas.

Implicaciones de seguridad:

- **Post-Quantum Cryptography:** el potencial de las computadoras cuánticas para romper los métodos de cifrado actuales está impulsando la investigación sobre criptografía poscuántica, que tiene como objetivo desarrollar técnicas de cifrado que las computadoras cuánticas no puedan descifrar.

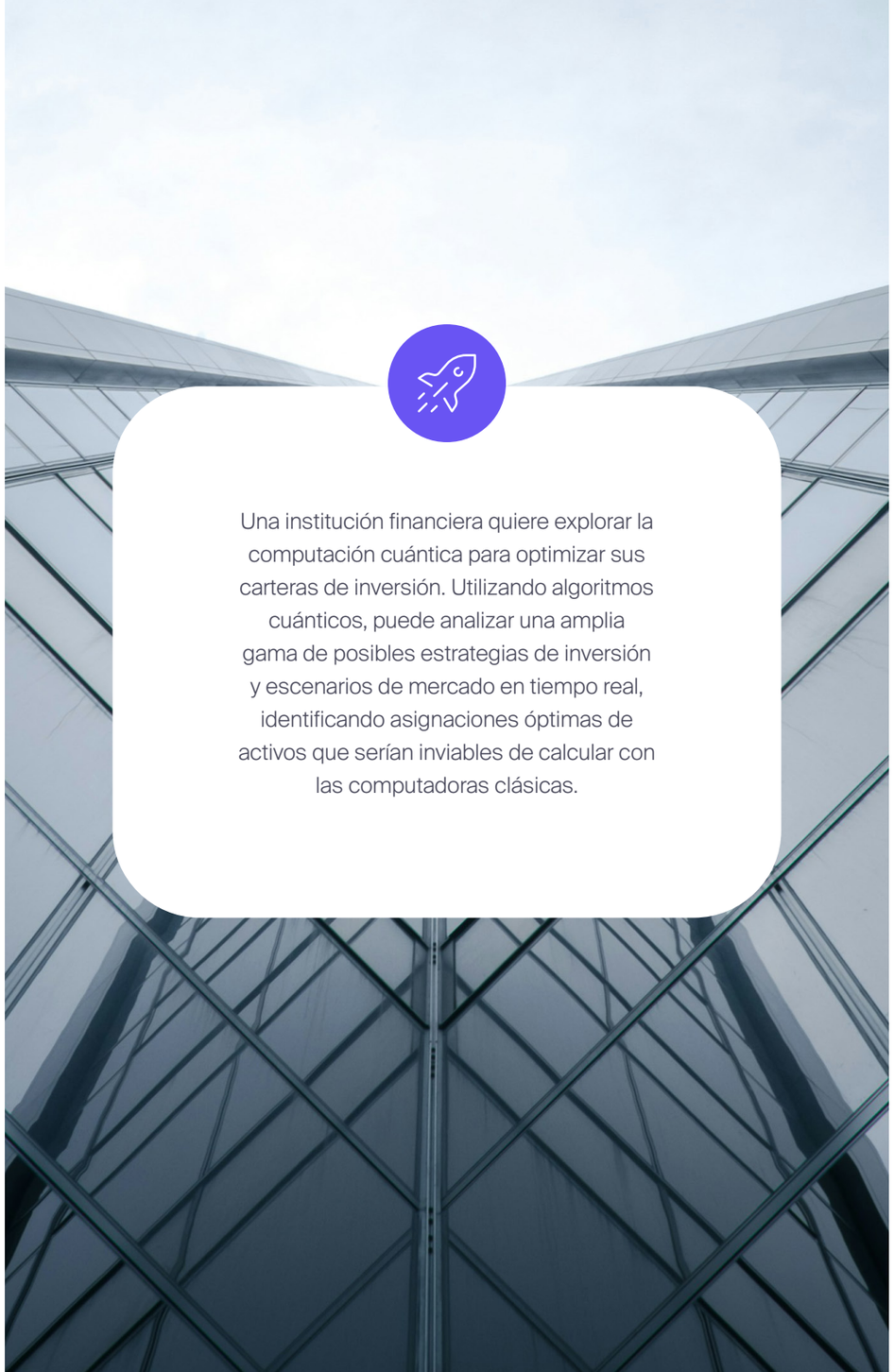
- **Quantum-Safe Security Practices:** las organizaciones están comenzando a evaluar su infraestructura y prácticas criptográficas para prepararse para la era de la computación cuántica.

Tendencias

- **Computación cuántica generalizada:** a medida que la tecnología madure, se espera que las computadoras cuánticas sean más accesibles, con servicios de computación cuántica basados en la nube que las pongan a disposición de las organizaciones que no pueden pagar su hardware cuántico.

- **Redes cuánticas e Internet cuántico:** el desarrollo de repetidores y satélites cuánticos permitirá la creación de redes cuánticas de largo alcance, lo que conducirá a un Internet cuántico que ofrezca niveles de seguridad y conectividad sin precedentes.

- **Revolución en el análisis de datos:** es probable que las tecnologías cuánticas revolucionen el análisis de datos al manejar conjuntos de datos complejos y a gran escala de manera más efectiva, lo que conducirá a nuevos conocimientos en campos que van desde la ciencia del clima hasta la atención médica.



Una institución financiera quiere explorar la computación cuántica para optimizar sus carteras de inversión. Utilizando algoritmos cuánticos, puede analizar una amplia gama de posibles estrategias de inversión y escenarios de mercado en tiempo real, identificando asignaciones óptimas de activos que serían inviables de calcular con las computadoras clásicas.

Key insight

Las tecnologías cuánticas están llamadas a transformar drásticamente el panorama de la gestión y la seguridad de los datos. Con su potencia computacional sin precedentes y sus novedosos enfoques de cifrado, prometen abordar algunos de los problemas más desafiantes en estos campos.

Si bien siguen existiendo desafíos significativos, particularmente en la creación de sistemas cuánticos estables y escalables, **las aplicaciones potenciales y los beneficios de las tecnologías cuánticas las convierten en una de las áreas de desarrollo más emocionantes en la gestión y seguridad de datos.**

Conclusiones

En el acelerado mundo tecnológico actual, permanecer actualizados y adaptarse a las tendencias emergentes en materia de datos es crucial para cualquier organización que aspire a ser líder en su sector.

La gestión y utilización eficiente de los datos ya no es solo una cuestión de tecnología, sino que ha pasado a convertirse en una **estrategia empresarial que puede determinar el éxito o el fracaso de la organización en el mercado actual.**

Los líderes de datos y análisis, CDO y CDAO, necesitan sentar las bases estratégicas y los modelos operativos para crear una hoja de ruta que permita a la organización **construir una cultura basada en datos impulsando los objetivos de la organización.** La mayor preocupación de los líderes es respaldar la toma de decisiones de su organización, es por ello, que necesitan datos confiables y de alta calidad que, de la mano de las nuevas tecnologías basadas permitan asegurar el éxito del proceso.

En el camino hacia la transformación de una organización hacia una cultura basada en datos los líderes tecnológicos desempeñan un papel fundamental. Es por ello por lo que deben promover **la implementación y uso de plataformas, herramientas y procesos que sustenten la estrategia de datos**, a la vez que fomentan dentro de la organización la cultura data driven a nivel vertical y horizontal.

Los directores de datos y otros líderes tienen una oportunidad única de generar un gran impacto en sus organizaciones gracias a la rápida evolución de las tecnologías asociadas a la explotación del valor de los datos como la inteligencia artificial o el aprendizaje automático que, aplicado a las técnicas de Data analizadas en este paper, evolucionan sus capacidades exponencialmente.

Para lograr esto, las organizaciones deben asegurarse de contar con fundamentos sólidos que sustenten la innovación. Esto incluye una confluencia entre la estrategia general de datos y la capacidad de experimentación y adaptación regulatoria.

Entre los enfoques clave para la gestión de datos eficiente se incluyen:

1. Democratización de Datos

La democratización permitirá que un mayor número de miembros del equipo acceda y comprenda los datos, promoviendo una toma de decisiones más informada y ágil. Para una organización, esto significa impulsar la innovación desde todos los niveles y asegurar que todos los equipos estén alineados con los objetivos estratégicos, fomentando una verdadera colaboración interfuncional.

2. Monetización de Datos (DaaS)

Transformar los datos en un activo comercializable puede abrir nuevas fuentes de ingresos y modelos de negocio. Las empresas que adopten este enfoque podrán diversificar sus flujos de ingresos y fortalecerán su posición en el mercado.

3. Cloud Modernization

Migrar los ecosistemas de datos a soluciones en la nube modernas permite una escalabilidad, flexibilidad y seguridad sin precedentes. Las organizaciones pueden aprovechar las capacidades de la nube para acelerar la innovación y optimizar costes operativos.

4. Data Fabric

La construcción de una infraestructura de datos cohesiva y unificada es esencial para garantizar la integridad, la accesibilidad y la seguridad de los datos. Una arquitectura de datos bien diseñada facilitará la colaboración y mejorará la eficiencia operativa.

5. TinyML

Los algoritmos inteligentes basados en técnicas Small y Wide Data permitirán realizar análisis y procesamiento de datos en dispositivos Edge con recursos limitados, potenciando aplicaciones IoT y sistemas embebidos de alta eficiencia.

6. EdgeAI

AI llevar las capacidades de inteligencia artificial al borde de la red, se habilitarán escenarios de análisis en tiempo real y sistewmas de toma de decisiones autónomas, cruciales para la expansión de los entornos IoT y la monitorización de aplicaciones críticas.

7. Data Masking

En un mundo donde la privacidad de los datos cada vez va a ser más primordial, la capacidad de proteger información sensible sin comprometer su utilidad es fundamental. Implementar técnicas de enmascaramiento de datos garantizará el cumplimiento normativo a la par que fortalecerá la confianza del consumidor.

8. DataOps

Esta metodología enfatiza la colaboración entre equipos de desarrollo, operaciones y datos, promoviendo la automatización y la integración continua. Adoptar prácticas DataOps va a permitir a las organizaciones acelerar el ciclo de vida del desarrollo de datos y mejora la calidad del producto final.

9. Web3

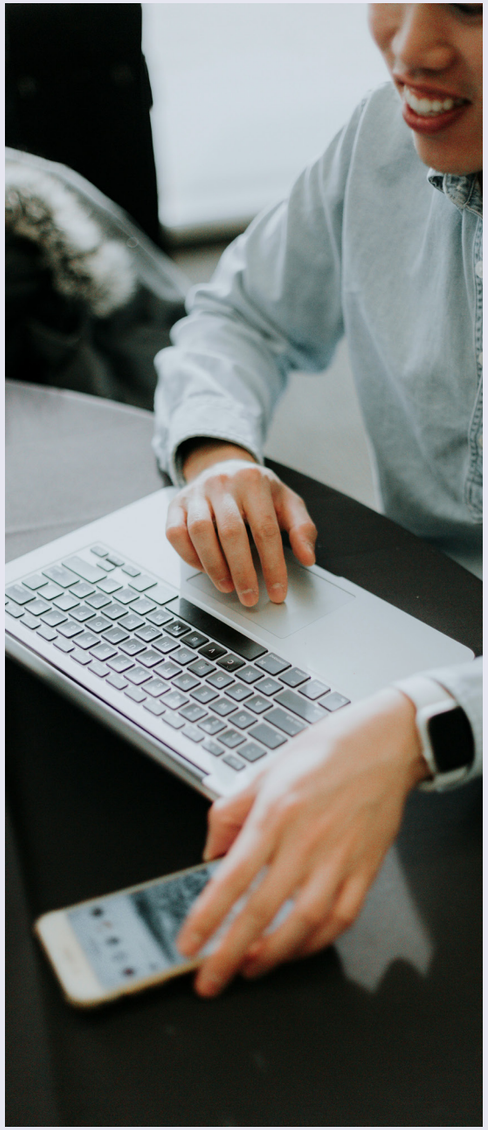
Con la evolución de internet hacia un ecosistema descentralizado y basado en Blockchain, las organizaciones deben estar preparadas para innovar en áreas como las de los contratos inteligentes, criptomonedas o las aplicaciones descentralizadas (DApps). La Web3 redefinirá la forma en que se realizan las interacciones en línea y ofrecerá nuevas oportunidades de negocio asociadas al tratamiento de datos.

10. Quantum Technologies

Aunque aún en desarrollo, la computación cuántica tiene el potencial de revolucionar la forma en que procesamos y analizamos datos. Las organizaciones que inviertan en investigación y desarrollo en este campo estarán a la vanguardia de la innovación tecnológica, encontrando múltiples vías de generación de negocio y optimización del rendimiento.

Para toda organización que busque ser puntera en materia de datos, la capacidad de anticiparse y capitalizar estas innovaciones determinará no solo el éxito a corto plazo, sino también la relevancia y competitividad en el futuro digital.

Durante la próxima década, marcada por un posicionamiento de los datos en el centro del negocio, la **inversión en tecnologías emergentes** y la adopción **de mejores prácticas** en gestión de datos va a ser fundamental para garantizar un crecimiento sostenible y una posición de liderazgo en el mercado.



softtek.com

Softtek®