Softtek®

Data: what's new?

Index

Prologue

The future of the business landscape is marked by the ability of organizations to unlock the potential of data. At a time when data collection has already been overcome as a problem, transforming this data into information seems to be an elusive challenge that only the most advanced companies in the acceleration of digital transformation can achieve.

The digital landscape evolves to respond to this challenge, where technological innovation is redefining the direction of the new strategies that organizations must follow to lead this paradigm shift. **Data has established itself as a key aspect in the business realm due to the almost limitless possibilities for generating business value, only restricted by data privacy.** At this point, cybersecurity positions itself as a priority issue within the debate on how to leverage technological advances and data in business development within regulatory compliance and ensuring security.

So, how can organizations transform data into a strategic asset that drives innovation? How can they optimize operations and create new opportunities while protecting the integrity and confidentiality of the data? To address this debate, leaders and organizations must go hand in hand with technology and have a flexible and resilient mindset ready to embrace change.

With this research, **we seek to offer a comprehensive view of the technological trends that are defining the course of Data and that will mark the evolution of the coming years.**

Anticipating the future will define the new business landscape.

01. Data democratization

In the digital age, data becomes a fundamental strategic element for organizations, enabling the creation and evolution of decision-making, operations, and actions. To ensure the success of the new transformative context, companies should not stay purely technological; a mindset change of the leaders is essential to accelerate and ensure the adoption of the data potential.

To achieve true data democratization, **involved parties must gain literacy skills to verify the accuracy of the data, ensure its security, and its proper processing** and use. It is imperative that the selected format for interacting with data allows for optimal usability and provides a seamless user experience. Often, data democratization is combined with transparency, referring to processes that ensure the traceability of such information by providing its location or origin.

Data democratization is revolutionizing management and business processes. Formerly, data governance was limited to the IT team and such information was barely exploited. The change we are experiencing not only affects access but **is about empowering all members of an organization and equipping them with the capability and tools to leverage this potential, promoting an agile and informed culture.**
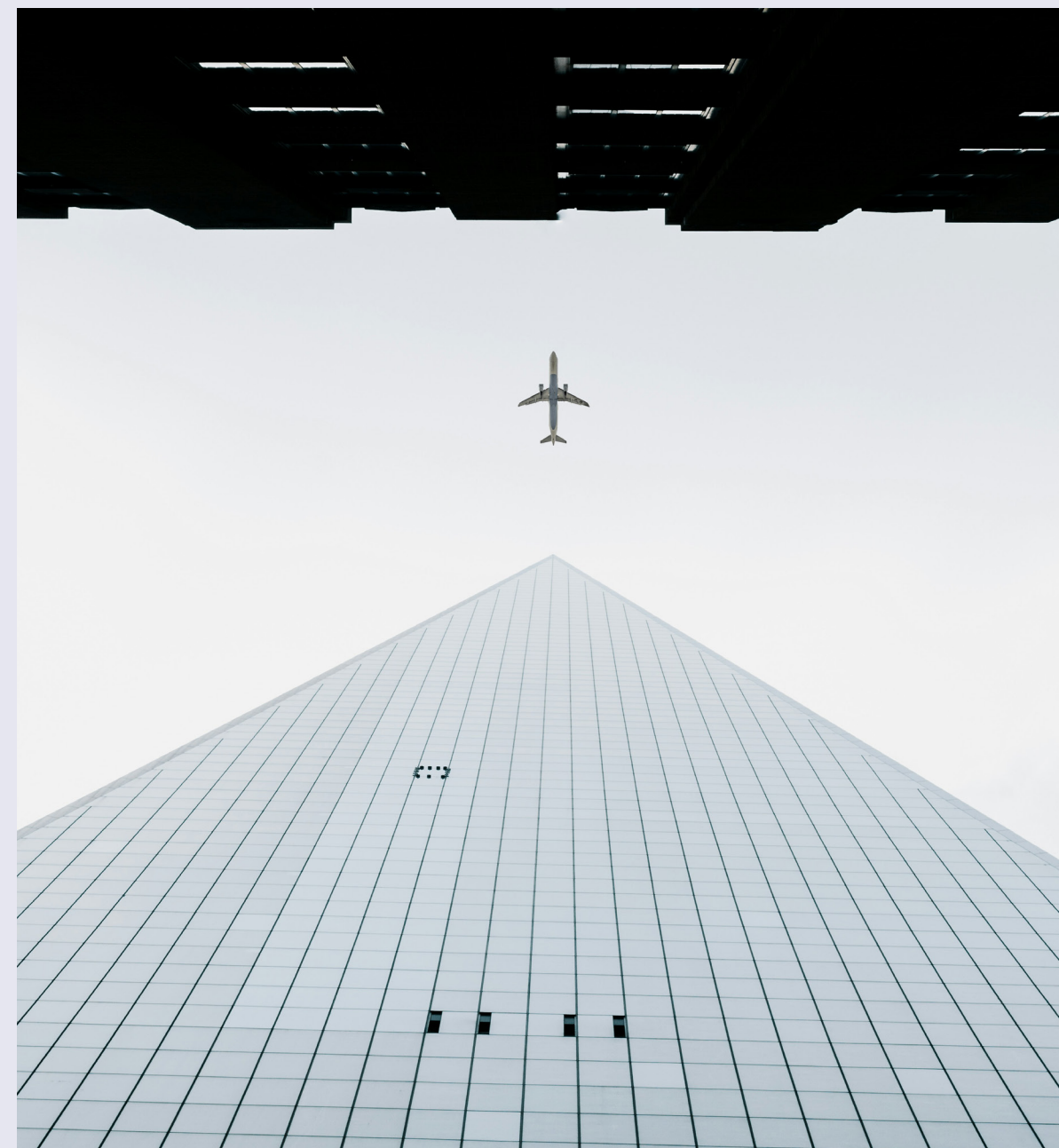
Evolution

To become a Data Centric organization, **data must transcend and penetrate the entire organizational structure**; in this way, data acquires the status of an asset and can be exploited by all the people who make up the company, improving processes and increasing interactions. To achieve this, companies must anticipate an operating model based on data and design a technological foundation that supports the entire operation.
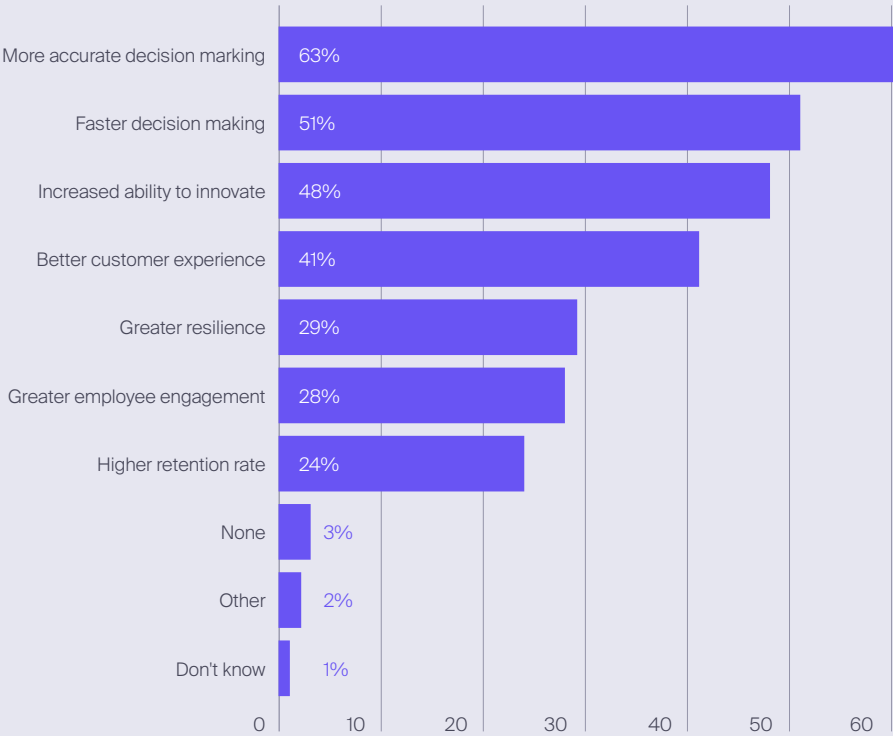
In this context, **where the shift is from isolated data to open and accessible data ecosystems**, capabilities related to management and cultural change become critical; therefore, organizations are significantly increasing their investment in technological competencies and talent training to adapt to the digital context.

The transformation brings numerous benefits, the progressive change towards a new comprehensive data strategy implies the **reduction of IT infrastructure overcosts** by adopting modern platforms and dismantling old systems. Likewise, it allows the application of new intelligent technologies and automation processes that maximize the potential of the outcomes, obtaining more cutting-edge strategies based on data.

Recent research provides revealing information about the state of data democratization: it is estimated that 70% of business leaders who participate in initiatives to improve their data skills experience an improvement of more than 60% in the speed and efficiency of their decision-making, which in turn has a positive impact on customer experience and innovation capabilities.

Value of data literacy for workforce improvement

Dimensions in which data-literate employees outperform non-data-literate employees.

Dimension	Percentage
More accurate decision marking	63%
Faster decision making	51%
Increased ability to innovate	48%
Better customer experience	41%
Greater resilience	29%
Greater employee engagement	28%
Higher retention rate	24%
None	3%
Other	2%
Don't know	1%

*According to a survey of global organization leaders. Source: Softtek based on data from Data Camp.

Technological Background

Automated Data Discovery and Categorization

- **Machine learning algorithms:** code snippets that assist users in exploring and analyzing complex data sets by detecting patterns. They help to classify and tag data automatically, facilitating its search, understanding, and processing.

- **Semantic search capabilities:** improvements in this field, thanks to the rapid evolution of machine learning natural language processing models (LLM), allow users to perform queries in natural language, making the process more intuitive and agile.

Robust Data Governance and Quality Frameworks

- **Data catalogs with enhanced metadata:** provide enriched context through the inventory of data assets that allows users to find, access, and investigate information quickly and easily. It also includes information traceability and user-generated annotations to help others understand and trust the data.

- **Automated data quality checks:** automated systems ensure that users always work with high-quality data.

Self-Service Data Preparation Tools

- **Intuitive data management interfaces:** allow users to optimize their interaction with data, from searching to combining data without the need for coding.

- **Data transformation enhanced by artificial intelligence:** integrates a layer of intelligence into data handling based on user behavior, streamlining the data preparation process.

Challenges

Data Centralization

- **Adequate infrastructure for data interaction:** Thanks to new technologies, platforms are being generated that allow a natural and agile interaction with data without the need for specific technical training.

- **Data quality and consistency**: Data governance is a rigorous, accurate, and essential approach to managing data throughout its lifecycle, from acquisition to disposal, including its use to ensure data accuracy, integrity, consistency, and relevance.

Balance between Accessibility and Security

- **Attribute-Based Access Control (ABAC):** This is an authorization system that defines access based on the attributes associated with security entities, resources, and the environment of an access request, providing detailed control in access management based on role assignment.

- **Differential Privacy**: Addresses the need for a balance between the legitimate exploitation of information and respect for individual rights. This strategy aims to preserve the utility of data while respecting individuals' privacy by adding "noise" to the data or queries.

Ensuring consistent Understanding and Interpretation

- **Data Literacy Programs:** These programs equip talent with the comprehensive ability to understand, construct, and communicate data in its context, the applied analytical methods and techniques, and the capability to produce a valuable application.

- **Business Intelligence and Integrated Analytics:** This is the collection of tools and software that enables the processing and transformation of data into knowledge through user-friendly interfaces. By analyzing such information, users gain insights that allow for business performance improvement.

Trends

**Natural Language Processing
and Generation**

- **Chatbots and Virtual Data Assistants:**
  The explosion of Generative AI and Large
  Language Models (LLM) has revolutionized
  intelligent models thanks to their ability
  to enhance productivity and efficiency
  through natural language queries, allowing
  users to obtain relevant data thanks to their
  intelligence layer.

- **Automated Information Generation:**
  Thanks to AI involvement, systems now not
  only provide data but also detect patterns,
  trends, anomalies, and key information in an
  automated and agile manner, enabling users
  to quickly understand what they are seeing
  without the need for deep statistical training.

**Augmented Data
Management**

- **Data Management Enhanced by
  Machine Learning:** Thanks to AI and
  ML, software learns from interactions
  with users, continuously improving and
  automating the data management process.
  In this way, the data ecosystem evolves
  alongside user needs.

Decentralized Data Ecosystems

- **Blockchain for Data Sharing:** Expands
  the potential for data democratization
  thanks to the shared ledger controlled by
  network members. This mechanism allows
  for the transparent sharing of information
  and the storage of data in blocks that are
  linked together in a chain. In this manner, it
  ensures that data is consistent because it
  is not possible to delete or modify the chain
  without the consensus of the network.
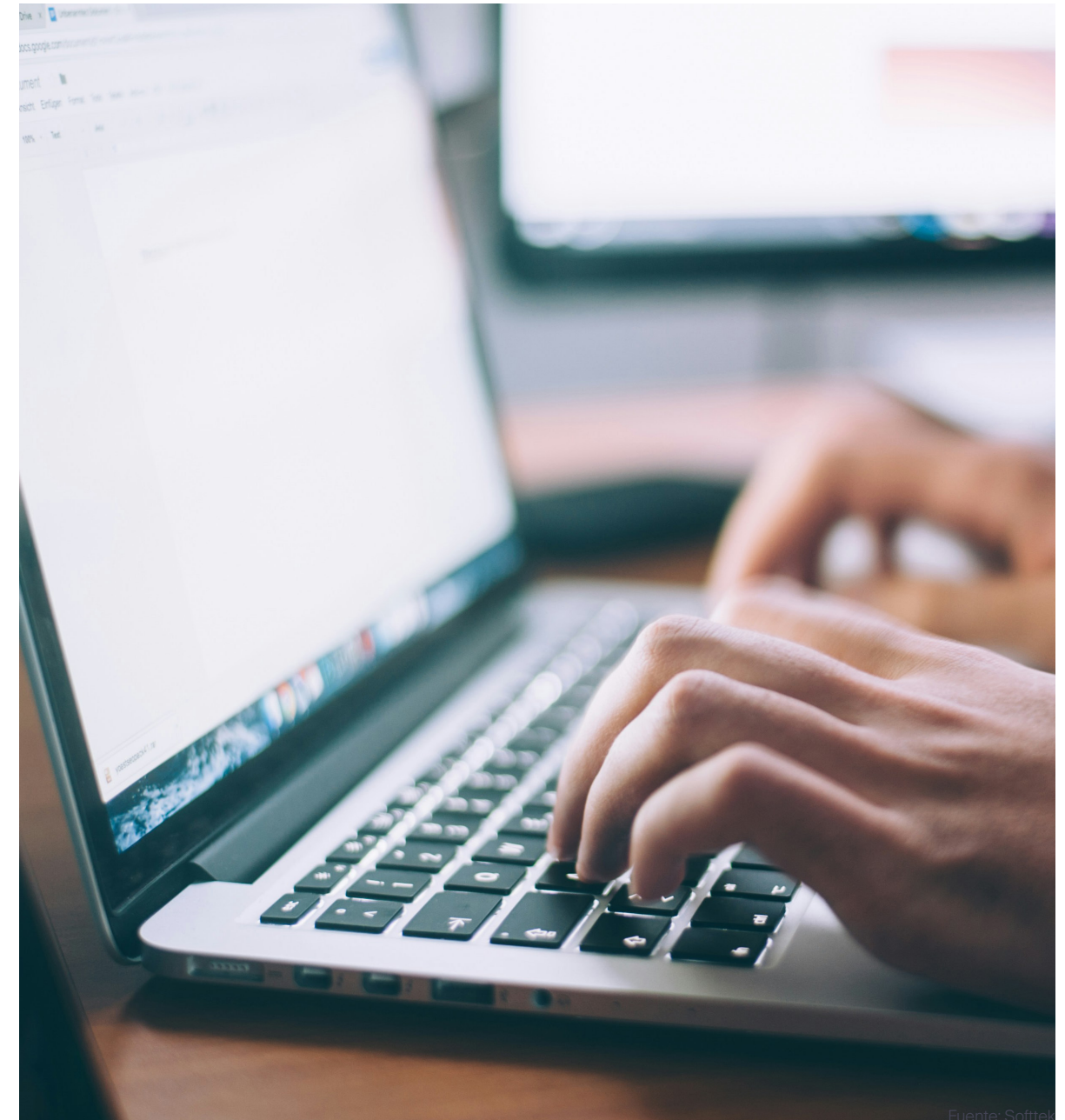
Fuente: Softtek

Key insight

A multinational organization implements a data democratization initiative through a self-service analytics platform, allowing marketing teams to access customer data and gain insights into campaign performance. Thanks to its advanced data catalog, users can easily find the data they need, understand its context, and trust its quality. Automated data quality checks ensure that the data is always accurate and up-to-date, while attribute-based access control -ABAC- ensures the protection of sensitive data.

Data democratization is not just about making data accessible, but about creating a culture in which data is an integral part of everyone's role. By combining advanced technologies with a focus on empowering and educating users, organizations can unlock the full potential of their data, driving innovation and success in the digital age.

As technology continues to advance, the possibilities for data democratization will only expand, offering new ways to empower users and transform businesses.

02. Data Monetization & DaaS

The rapid evolution of technologies related to Data has led to a new paradigm in which the differentiation of organizations stems from their ability to adapt and lead in the new digital ecosystem. Most organizations are already researching and implementing strategies related to the impact of data on business performance to translate it into real monetary value.

The issue addressed by leading organizations regarding how to manage the functionalities offered by data has provided several trends in the market. **Data Monetization encompasses several strategic aspects to address this business asset.**

Data Monetization refers to the **process of obtaining a quantifiable economic benefit from data.** It is, ultimately, an organization's ability to create and derive value from data and artificial intelligence assets. This idea is linked in turn to Data-as-a-Service, establishing a differentiation by considering other ways to extract value from data and thus increase business performance.

From this change in mindset, **Data-Driven organizations emerge, those that base their strategic decisions on the evidence extracted through data analysis.** These organizations integrate into their corporate culture the strategic value of data, transforming and optimizing their strategies based on the knowledge generated.

Evolution

Technological Background
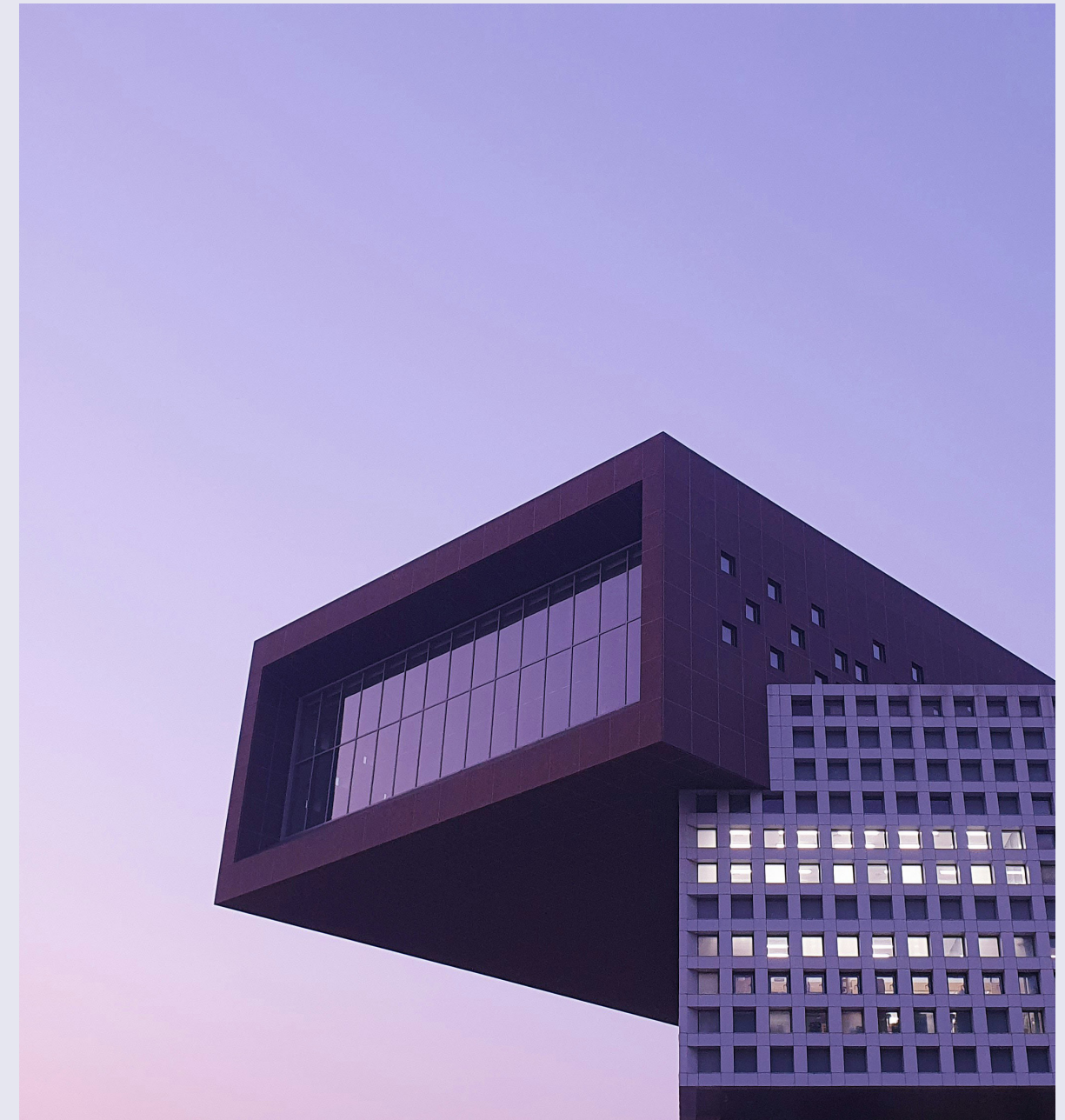
While the concept of Big Data dates back to the 1960s and 1970s, with the start of the first data centers, it was not until 2005 that the market became aware of the amount of information produced through online platforms and services. The peak would be reached with the arrival of technologies such as IoT, generative AI, LLM, and Machine Learning in this decade.

The technological and business potential is evident for all organizations competing to develop productive models. Thanks to these technologies, companies **utilize highly personalized predictive models tailored to the needs of potential consumers**; predictive maintenance is now a reality that saves costs in the industrial sector while reducing accidents; the customer experience or fan experience has evolved into futuristic connected models that were only possible in the imagination a few years ago, just to mention some of the applications.

Robust DaaS Platforms

- **Scalable architecture:** DaaS platforms utilize a Cloud-native infrastructure that allows for scalable and flexible solutions for storage, processing, integration, and access to data from both internal and external sources without degrading performance when serving numerous clients.

- **API-led Connectivity:** APIs are the foundation of DaaS platforms as they enable secure access to data services. The two most commonly used APIs are RESTful API and GraphQL. RESTful API is an interface between two computing systems to exchange information securely over the internet. GraphQL is a query language that provides a flexible and intuitive syntax for describing data requirements and interactions, allowing access to multiple sources in a single request.

GraphQL and REST are two different approaches to designing an API for data exchange over the Internet. REST allows client applications to exchange data with a server using HTTP verbs, while GraphQL defines how the client application should request data from a remote server.

Technological Background

**Enhanced Data Security
Measures and Compliance**

- **Advanced Encryption and Anonymization Techniques:** The General Data Protection Regulation is the European regulation concerning the protection of individuals with regard to the processing of their personal data and the free movement of such data within the EU and European Economic Area. This regulation and various member state laws necessitate the implementation of techniques, tools, and methodologies that ensure confidentiality.

Two techniques stand out to address this issue: Format-Preserving Encryption (FPE) and K-anonymity to guarantee data privacy without losing its utility for analysis and information generation. Format-Preserving Encryption (FPE) extends classic encryption algorithms like AES in a way that the texts retain their original length and format. K-anonymity is a property of anonymized data that allows quantifying the extent to which anonymity is preserved in a dataset from which identifiers have been removed. It is a measure of the risk that external agents can obtain personal information from anonymized data. The most used tools are the ARX Data Anonymization Tool, UTD, and Amnesia.

- **Blockchain for Data Integrity**: is a distributed information storage technique that uses consensus policies to validate included information and policies for detecting the integrity of originally recorded data. Blockchain technology ensures that all data provided through DaaS are immutable and traceable. These qualities make it a solid defense against cyber threats and information manipulation.

Cryptography is essential in encrypting information, an example of which is hash generation, which allows detecting any alteration in the data of a block. It is also used to authenticate transactions and generate digital signatures that verify the identity of the involved parties, thereby ensuring that the transactions are legitimate.

Data Analysis Tools

- **AI and Machine Learning Integration:** AI and Machine Learning capabilities are increasingly integrated into DaaS to provide insights derived from data. Machine learning uses algorithms to identify patterns in data and create models that help generate insights.

- **Real-time Data Processing:** Involves collecting and processing a stream of data from various sources in real time to extract meaningful information from it. Examples include e-commerce mobile applications, social media information, or telemetry from connected devices.

Streaming data from multiple devices and performing ETL transformations on large volumes of continuous, high-speed data in real time allows for the analysis of data as soon as it is generated.

Challenges

Data Quality and Consistency

- **Automated Data Cleaning and Enrichment Tools:** These are essential for the accuracy of data analysis, ensuring data is accurate, consistent, and reliable through the collection, cleaning, transformation, and organization of raw or incomplete data. Data preparation tools help transform and streamline the data processing procedure. The most well-known tools in the market include Python, Tableau Desktop, OpenRefine, Apache Spark, Datameer, Alteryx, among others.

- **Master Data Management (MDM):** This is the process of creating and maintaining a single master record for each piece of data through which organizations obtain an up-to-date and reliable view of key data. Implementing MDM ensures that the entire organization refers to a unique and consistent set of master data, reducing errors. In this practice, several providers are already

incorporating advancements in AI and ML to automate the process of verifying the accuracy of new data and comparing it with current records.

Pricing Strategies and Valuation

- **Dynamic Pricing Models:** Leverage advanced algorithms and technologies to analyze large volumes of data, such as Big Data, artificial intelligence, or Machine Learning, to dynamically set service prices based on demand. There are different types of Dynamic Pricing strategies: Time-Based Pricing, Peak-pricing, and market penetration strategies aimed at maximizing revenue while remaining competitive.

- **Value-based Pricing:** This strategy involves setting prices based on the perceived value of the product. The key factor is determining the customer's perception of value, as it allows the organization to make intelligent decisions and determine, for example, whether a premium pricing strategy is appropriate

Trust and User Adoption

- **Transparent Usage Policies:** The General Data Protection Regulation outlines a set of principles that data controllers and processors must observe when handling personal data: the principle of legality, transparency, and fairness; the purpose limitation principle; the data minimization principle; the accuracy principle; the storage limitation principle; the integrity and confidentiality principle; and the accountability principle. These principles ensure the due diligence and compliance with GDPR and LOPDGD. Articulating and properly

communicating data processing generates trust in users that their rights are being respected.

- **Quality & Performance SLAs:** A Service Level Agreement (SLA) is a contract that specifies the particular service and detailed quality commitments for its provision. This document is necessary to establish a clear framework of action by the provider regarding deadlines, quality, and any concept that allows for a precise understanding of the contracted service. Among other aspects, SLA agreements should include a section on data management, confidentiality, and data quality.

Challenges

Innovative Strategies in Data Monetization

- **Tiered Data Services**

Custom Access Levels: To cater to a broad range of customers, organizations offer tiered access to their data services, where basic data is available for free or at a low cost, and more detailed, strategic, and real-time data require a premium level.

Bundled Services: Another way to serve customers is by bundling additional services such as advanced analytics, support, or consulting services with higher-tier subscriptions, which adds more value and justifies a price increase.

- **Data Enrichment Services**

Enhanced Data Insights: DaaS providers go a step further by offering enriched data thanks to the processing and analysis that have turned it into actionable information. Clear examples include trend analysis, predictive modeling, or engagement analysis.

Integration with External Data Sets: Generates enriched data that offer a broader context and deeper insights, enabling the satisfaction of clients' more specialized demands.

- **API-Based Models**

Pay-Per-Use Models: Providers are offering API-based access models to their data services with pricing based on the volume of API calls or the amount of data consumed. This model allows for more flexible and scalable use.

Data API Marketplaces: Some organizations are creating or participating in data API marketplaces, where developers and businesses can subscribe to a wide range of data services.

Challenges

- **Performance-Based Monetization**

Outcome-Oriented Pricing: The most innovative DaaS providers are linking their pricing to the performance or outcomes that their data helps to achieve.

Shared Savings and Revenue Sharing: In some cases, providers enter into shared savings or revenue sharing agreements, where they receive a portion of the cost savings or additional revenues that their data helps their clients to achieve.

- **Data Syndication**

Multi-Client Data Products: Involve the aggregation of data from multiple sources, including the clients themselves, to provide valuable insights for all participants.

Anonymization and Aggregation: To ensure privacy and legal compliance, syndicated products are based on anonymized and aggregated data, ensuring that individual data points cannot be traced back to their origin.

- **Dynamic and Real- Time Data Products**

Real-Time Data Streams: Thanks to the rise of IoT and the growing demand for data, providers offer real-time data streams as a premium service for those customers who require constantly updated information for their operations.

Event-Based Data Services: Data services triggered by specific events or conditions that provide highly relevant data on demand.

Trends

Custom Data Flows

- **Customizable Data Sources:** Future DaaS platforms will offer more customized data flows, allowing users to select the type or frequency of data they receive based on their needs and use cases.

Decentralized Data Marketplaces

- **Blockchain-Enabled Platforms:** The rise of decentralized data marketplaces will allow organizations to buy and sell data securely and transparently, ensuring the integrity of these transactions.

Automated Compliance and Governance

- **AI-Driven Compliance Tools:** Automated tools based on AI will become indispensable for ensuring ongoing compliance and governance, adjusting to new regulations in real-time.
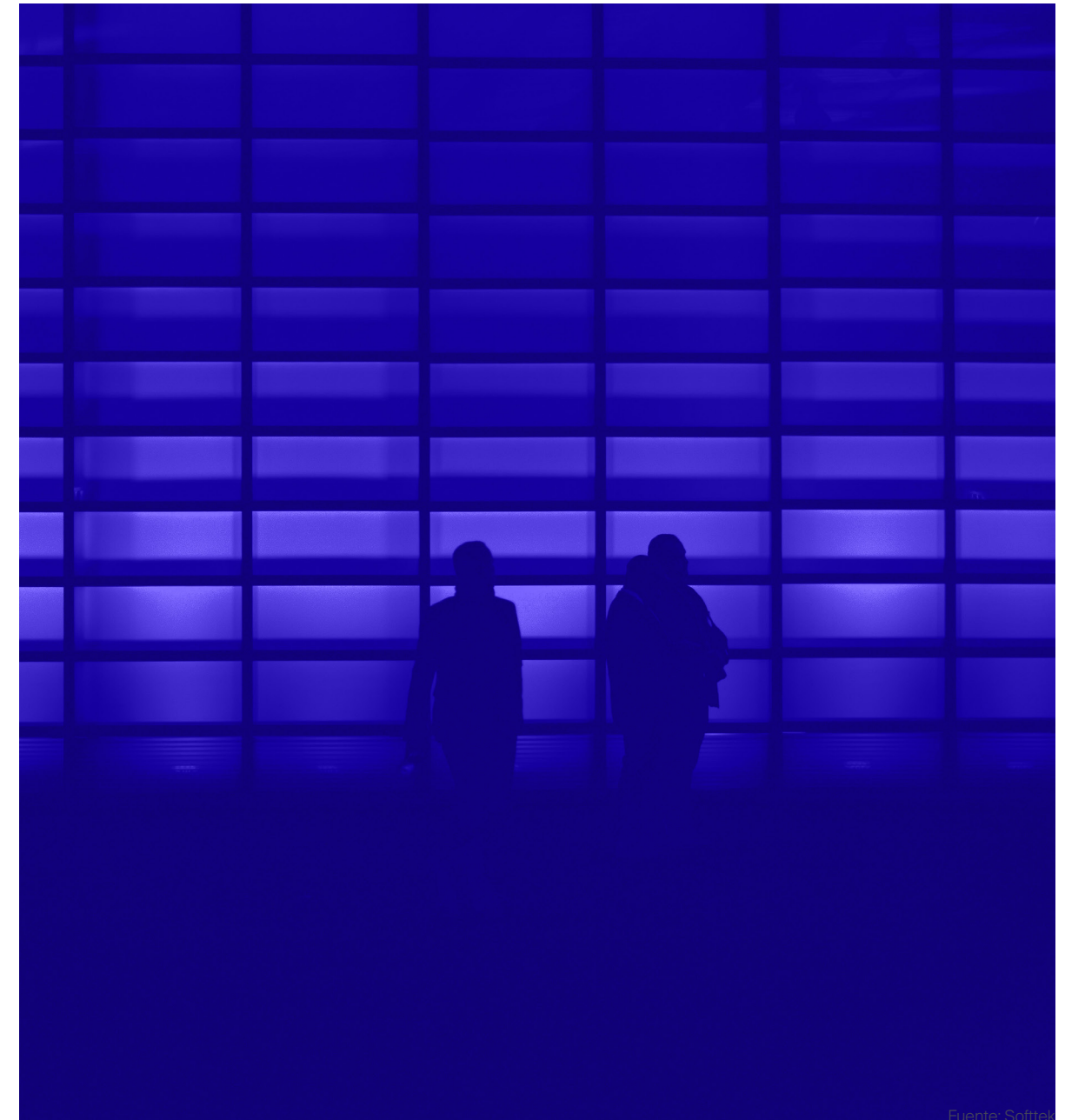
Fuente: Softtek

Key insight

A retail company adds value to its business through the implementation of a DaaS platform. Using advanced data collection methods, they aggregate and anonymize data from various sources, ensuring privacy. The platform also employs machine learning to provide predictive insights into consumer trends, helping retailers to optimize inventory and improve sales. Dynamic pricing models ensure that the service remains competitively priced, and solid SLAs guarantee data performance and quality.

Data monetization and DaaS are at the forefront of the data economy, transforming raw data into valuable assets. By adopting advanced technologies and strategies, organizations can unlock new revenue streams, provide enhanced services, and gain a competitive edge.

As the field continues to evolve, staying abreast of innovations and trends will be crucial for those looking to lead in the data-driven future.

03. Cloud Modernization/ New Cloud Data Ecosystems

The surge in the adoption of DaaS models is expected to simultaneously spark the creation of new data cloud ecosystems designed to leverage online-generated information and achieve higher efficiency through more agile, simple, and efficient storage.

This process has come to be known as Cloud Modernization and involves updating or restructuring all types of applications to fully leverage native cloud architectures and services.

For example, cloud modernization might involve taking a monolithic application and refactoring it to use a microservices architecture, thus consuming fewer hosting resources.

In this way, a modernized application takes better advantage of the cloud than one that merely migrates data without altering its own architecture.

At this point, we can differentiate between three types of cloud modernization:

- **Modernizing an on-premise application** by redesigning it to take full advantage before moving it to the cloud.

- **Modernizing an application that is already hosted in the cloud** but has not yet been optimized for truly efficient operation.

- **Modernizing a legacy on-premise application and keeping it on-premise**, but hosting it in a native cloud environment.
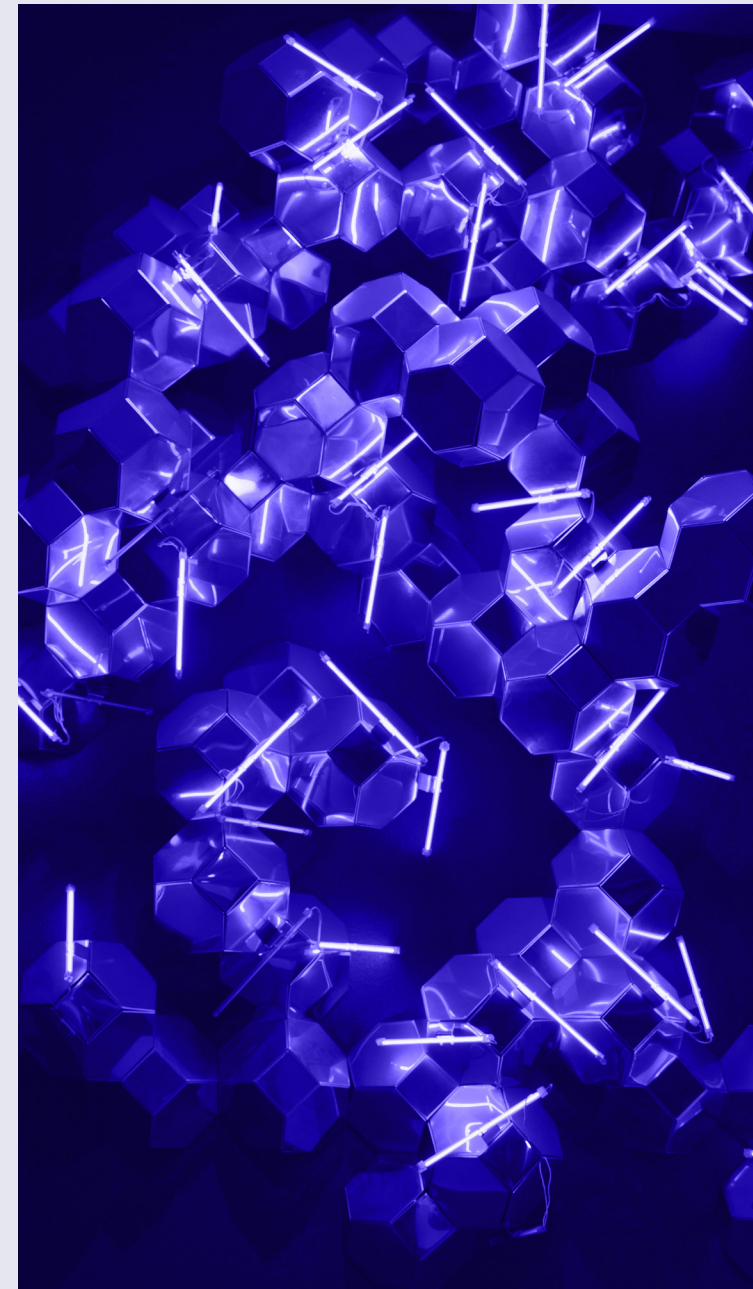
Evolution

The **evolution of cloud data ecosystems has been marked by two fundamental factors: data storage and processing.** The rapid advancement of Cloud technologies, artificial intelligence, Machine Learning, or Data presages a change as significant as the one Internet brought about in 1997.

The first Cloud services offered simple storage solutions, but as technology progresses, so does the ability to provide services. DaaS emerged as a critical component of this evolution, encapsulating the idea of delivering data to users on-demand, anywhere and anytime, without the burden of managing the underlying infrastructure.

Data ecosystems are moving from standalone software or combined implementations to solutions that are **100% Cloud-native**. It is expected that, throughout this year, new cloud system implementations will be based on a **cloud data ecosystem** rather than on ad hoc

integrated solutions. Organizations must evaluate data ecosystems based on their ability to integrate data from outside their environment and address challenges such as distributed data.

At this point, **the new capabilities of generative AI open the door to a new era of applications that can elevate the value of data to its highest level by creating, recommending, synthesizing, analyzing, or interacting with them naturally, ensuring transparency and accountability throughout the process.**

Technological Background

The latest generation DaaS ecosystems are based on several technical components::

- **API-First Architecture:** The DaaS approach focuses on provisioning data from a wide range of sources on-demand through an API. APIs are the backbone of this system, designed to simplify access to data and offer datasets or data streams ready to be consumed in various formats through the cloud efficiently.

- **Microservices & Containerization:** Microservices architecture refers to a style of architecture designed to develop applications. This methodology allows for dividing a large application into smaller, independent parts, each with its own scope of responsibility. Thus, an application based on microservices can call many internal microservices to form its response with the aim of delivering a single request to the user. Containers are a clear example of this architecture, as they allow

focusing on developing services without considering dependencies. Cloud-native applications are built this way, allowing, in this case, the development and deployment of DaaS applications. This way, each service can be scaled, updated, and managed independently, reducing downtime and accelerating innovation.

- **Data Virtualization & Federation:** Data federation is a process that allows multiple databases to function as one by providing a single data source for applications. Data federation is part of the data virtualization framework. It includes metadata repositories, data abstraction, read and write access to the source data systems, and advanced security. These technologies enable DaaS platforms to integrate data from multiple sources, providing users with a unified view of the data without the need to replicate or physically move them.

- **Integration of Artificial Intelligence and Machine Learning:** Advanced analytical capabilities are an integral part of modern DaaS platforms. The value of this service lies in delivering analysis and insights to offer information and predictive models, where the integration of intelligence and machine learning marks the step towards the future.
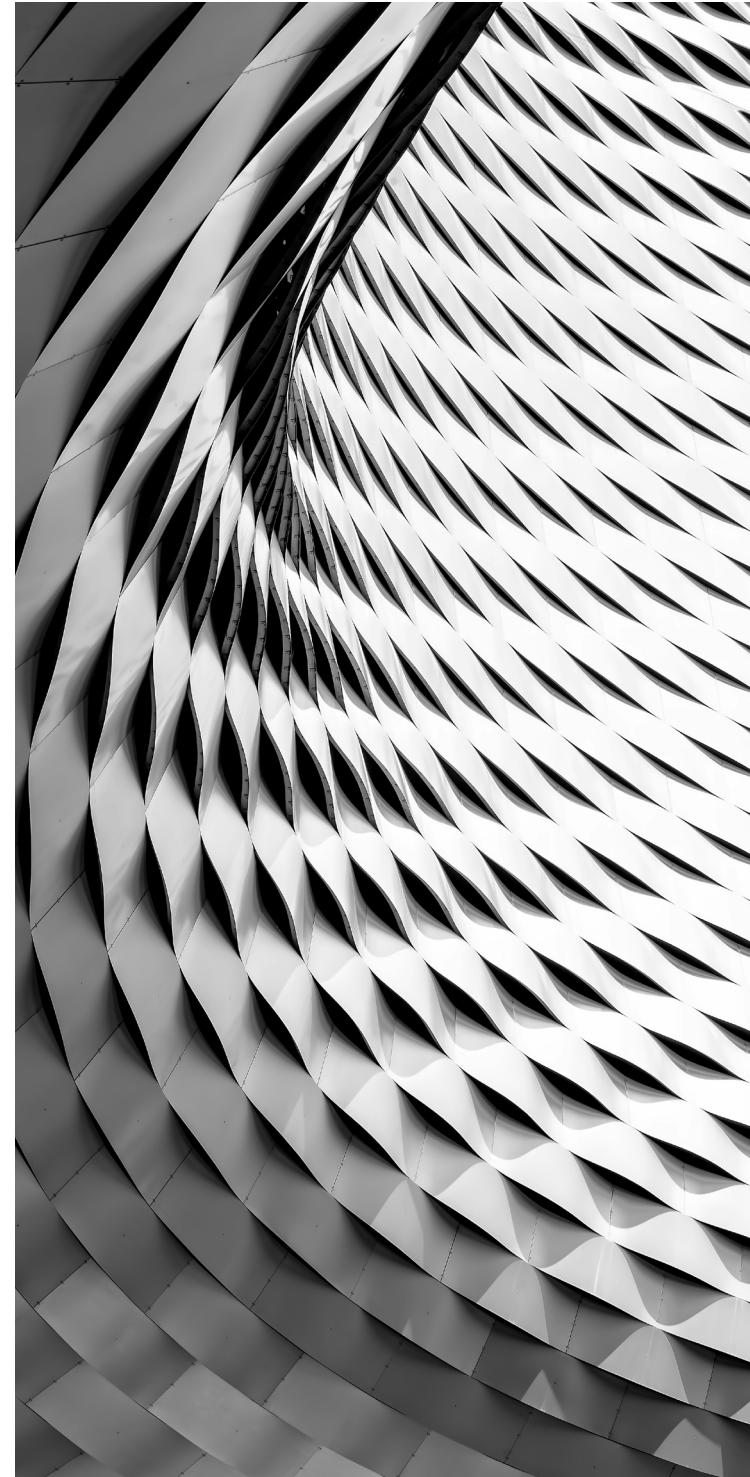
Challenges

- **Data Privacy and Security:** An AI and Cloud First focused data ecosystem allows an organization to generate greater business value from unique data assets, but it also creates the need to ensure their security. Advanced encryption solutions, such as those offered by Microsoft Azure among other providers, identify attacks faster, reduce risks, and design security based on customer needs.

- **Data Quality and Consistency:** As businesses integrate artificial intelligence and automation into their workflows, high-quality data becomes crucial for the effective adoption of these tools. In the last decade, advancements in Cloud, AI, IoT, and Edge Computing have exponentially evolved the possibilities of Big Data. As a result, data management has become a more crucial and complex practice. Organizations must put efforts into data quality management to ensure that Business Intelligence initiatives are accurate. In this regard, two determinants are found, governance and data quality. Therefore, organizations must establish data management

policies and programs and set mechanisms to solve problems that may arise throughout the data lifecycle. This involves sophisticated data validation, cleansing processes, and real-time monitoring of anomalies.

- **Integration Complexity:** Lies in the challenge of integrating with legacy systems. The goal of integration is to create a broader, more complete, and standardized system. Therefore, to address this process, the use of standardized APIs and the adoption of a solid integration strategy are key to overcoming these challenges.

DaaS platforms incorporate several advanced tools and techniques in their services and solutions:

- **Predictive Analytics:** Involves processing historical and current data through mathematical, statistical, artificial intelligence, and Machine Learning techniques to analyze the data and predict future trends or events, thus assisting organizations in strategic decision-making.

- **Sentiment Analysis:** Is the process of determining the emotional tone behind a series of words. Sentiment analysis tool is an automated technique that helps extract meaningful information from stakeholders about their attitudes, emotions, or opinions regarding the organization or the subject being evaluated. This tool uses advanced artificial intelligence technologies like natural language processing, text analysis, or data science to extract information. It is

particularly useful for conducting marketing strategies and monitoring social media, as it allows experts to understand customer sentiment and trends.

- **Real-Time Data Processing:** Refers to the ability to analyze and act upon data as it is generated, providing insights and responses in real time, reacting to specific events that have just occurred with appropriate logic. This is crucial for all applications requiring immediate information such as emergency response systems or financial trading.

Trends

Integration of Edge Computing: Allows for decentralized data processing close to the sources of data, reducing latency and bandwidth usage, especially critical for real-time data services.

Blockchain for Data Security: Is revolutionizing data security and transparency in DaaS, offering immutable and traceable data transactions.

Custom AI Services: As AI technology advances, DaaS platforms will offer more personalized services, providing businesses with information and models specific to their needs and challenges.

A healthcare company needs to provide real-time clinical data to hospitals and research institutions. To accomplish this project, a cloud-based DaaS platform was implemented, redefining the way users access and analyze information and improving patient care and advancing medical research.

Key insight

The **new cloud data ecosystems,** particularly focused on DaaS, **are redefining how data is accessed, analyzed, and monetized.** They represent not just a technological advance but a **strategic asset** that can drive innovation, efficiency, and new business models. As these ecosystems continue to evolve, they promise to unlock unprecedented opportunities for businesses ready to embrace the future of data.

04. Data Fabric

Achieving governance over all the data sources an organization has at its disposal, whether in local systems, Multi Cloud environments, or SaaS (Software-as-a-Service) applications, and doing so through a single virtualization layer, without the need to copy data from one place to another, is imperative **for any business that wishes to improve its data architecture approach in the coming years.**

Integration strategies based on the Data Fabric approach will play a key role in building more robust architectures, as they facilitate this end-to-end integration through the application of intelligent and automated systems.

The design of architectures based on the Data Fabric concept works by creating an integrated layer of data and connected processes. This layer performs continuous analysis on metadata to be able to reuse the information generated across all Cloud environments, including hybrid and Multi Cloud platforms, but also in new networks of sensors and connected IoT devices, or in applications of artificial intelligence, virtual reality, or augmented reality.

Through the Data Fabric approach, it is possible to intertwine data generated in all these different environments and treat them as a single reusable unit, providing a unified and coherent view.

Moreover, Data Fabric offers advanced data governance and security capabilities as it helps ensure that all data is being managed and protected under appropriate policies, and complying with regulations, regardless of its source of origin.
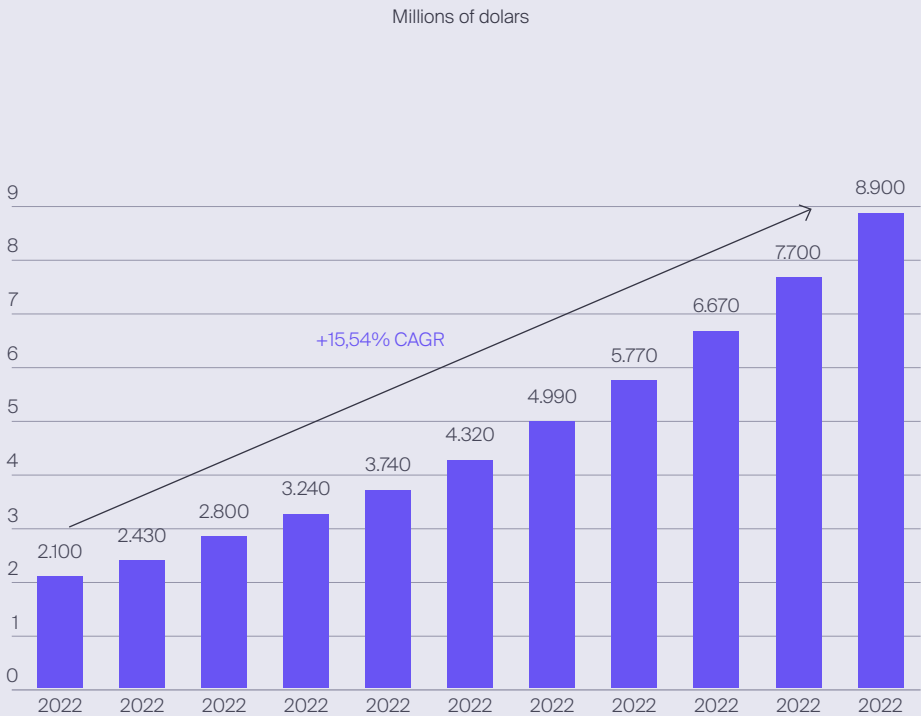
Evolution

Data Fabric emerges as an innovative concept from the need of companies to analyze and process data to leverage collective intelligence in their strategies. It represents the redesign of organizational architectures, with data as the main axis of operation supported by a technological approach.

From a semantic perspective, the term could be translated as a data fabric, referring to **the fibers that interrelate to form a structure that, with the right level of organization, provides a competitive advantage.** The ultimate goal is to maximize the value of data and accelerate digital transformation. Data Fabric simplifies and integrates data management in the cloud by connecting multiple locations, types, and sources of data through a single platform.

The concept of Data Fabric has evolved in response to the growing complexity of data environments. Traditionally, data was managed through centralized systems, but due to the explosion of data sources and cloud computing, the approach needed to be more flexible, scalable, and integrated to improve data quality, information availability, and automated insights, allowing for the anticipation of future challenges and opportunities in data management.

Throughout this decade, **the design of Data Fabric architectures will experience a market growth of more than 15.5% annually**, thanks to the opportunity it offers organizations to orchestrate data flows, processes, and operations more efficiently and consistently.

Evolution of Data Fabric market size at global level

Millions of dolars

+15,54% CAGR

Year	Value
2022	2.100
2022	2.430
2022	2.800
2022	3.240
2022	3.740
2022	4.320
2022	4.990
2022	5.770
2022	6.670
2022	7.700
2022	8.900

Source: Softtek based on Precedence Research data

Technological Background

Challenges

Data Fabric implements two high-level features:

- **Data Flows:** Provide a low-code interface that processes data from hundreds of sources and leverages hundreds of artificial intelligence-based transformations.

- **Data-Pipeline:** Maximizes the functionalities of enriched data orchestration to create flexible and continuous workflows, ensuring accessibility wherever and whenever needed.

Additionally, it consists of several layers that perform a specific function:

- **Data Source Layer:** Functions to dynamically catalog data, providing a unified view of all available data regardless of where it is hosted or its origin, facilitating its search and access.

- **Data Integration Layer:** Or semantic layer, performs all necessary tasks to connect and prepare data for use, from data cleaning, standardization of formats, or indexing to make it understandable and usable by the entire organization.

- **Data Access Layer:** Provides access to data through an API or easy-to-use user interface so that users can make queries or analysis without the interference of underlying complexity.

- **Automated Data Security and Governance Layer:** Responsible for maintaining the privacy and integrity of the data throughout its lifecycle and compliance with associated regulations and policies.

Integration Complexity: Integrating data across various systems and data sources can be complex. It's essential to audit current processes and systems to understand existing sources, systems, and data flows. Once the architecture is designed, integration through the use of middleware and standardized APIs can achieve the integration of systems into a unified view infrastructure.

Maintaining Data Quality: To ensure that data remains accurate and consistent across all systems, data management and control must be planned through defining policies, processes, standards, and tools for real-time monitoring and validation of data quality.

Scalability: As data volumes grow within organizations, the architecture must scale accordingly. Therefore, adopting a modular approach based on microservices ensures the capability to adapt fluidly to new requirements and helps unify processes to be quick and accurate with changes.

Data Fabric enhances data analysis by incorporating advanced tools and techniques:

Predictive Modeling: Thanks to access to a wide variety of data, predictive models can forecast trends and behaviors with greater accuracy.

Integration of Machine Learning: Data Fabric feeds machine learning algorithms, improving their accuracy.

Graph Analytics: This tool is especially useful in a Data Fabric context to understand relationships and patterns across diverse data sets.

Trends

Key insight

Greater Integration of AI: As artificial intelligence and machine learning technologies advance, they will be more deeply integrated into Data Fabric to provide more sophisticated, efficient, and strategic insights and automation.

Management of Edge Data Centers: With the rise of edge computing, Data Fabric will increasingly need to manage data at the network's edge, not just in centralized clouds or data centers.

Increased Focus on Real-Time Data: This will become a standard feature of Data Fabric, enabling more immediate insights and actions.

A global retail chain implements a Data Fabric architecture to integrate data from its physical and online stores. Thanks to this integration, the company gained a comprehensive view of customer behavior and inventory levels, leading to improved customer experiences and more efficient inventory management.

Data Fabric represents the future of data management, as it offers a flexible, integrated, and efficient approach to managing the constantly growing and evolving data needs of modern organizations. **By adopting Data Fabric, companies can ensure they are prepared to face the challenges of the digital age, making the most of their data to drive decision-making, innovation, and growth.**

26

05. TinyML/ Small y Wide Data

Technological advancements in TinyML, Small Data, and Wide Data are revolutionizing the field of data analytics and machine learning. As we generate more data, the ability to process and extract value in real time positions itself as a key competitive advantage for all organizations.

Tiny Machine Learning (TinyML) focuses on running machine learning models on low-power devices and machines, enabling them to handle complete machine learning systems thanks to the small data approach that provides strategies for extracting meaningful information from data sets.

Unlike traditional machine learning applications, TinyML does not rely on components like microprocessors and Graphics Processing Units (GPUs). Despite limited capacity, it uses the same tools to achieve deep learning through sensors, algorithms, and data analysis tools. These functions are scaled through a quantization process that reduces bit size. TinyML algorithms offer reduced storage functionalities that bring artificial intelligence and machine learning capabilities directly to Edge devices, processing data closer to the source where it originates and allowing much faster and more efficient decision-making.

They also offer advantages in terms of latency, privacy, and efficiency but pose challenges in effectively training models. A common issue is overfitting, or generalization, when the model adapts too well to the training data, including noise or random fluctuations in the data, and performs poorly when using new data not part of the initial training set. A clear example of TinyML application is the activation word detection model like "Ok, Google," which requires only 14 kilobytes to run.
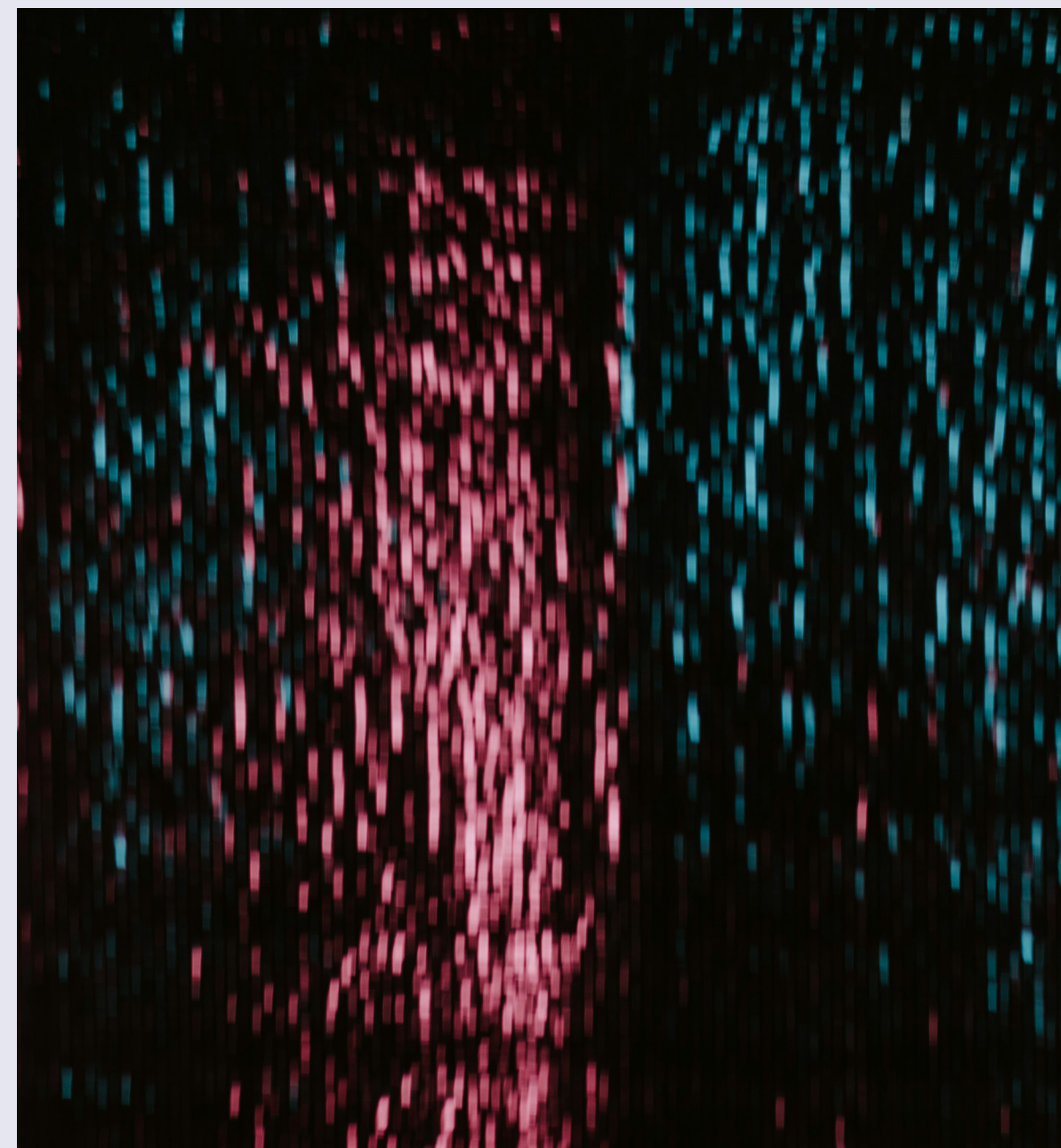
Small Data refers to the collection and analysis of data sets from individual organizations or based on individual problem-solving examples. It focuses on the importance of using limited and selected data sets to ensure their value relative to objectives for developing machine learning models. To achieve this, it employs time series analysis techniques, synthetic data, or self-supervised learning to find valuable information for the company in smaller data sets.

Wide Data involves relating different data sources to each other from a wide range to achieve meaningful analysis. The Wide Data approach handles a larger volume of data than Small Data but does not pursue massive volumes like Big Data approaches. It allows analyzing a wide variety of sources and both structured and unstructured data. This, in turn, enables finding links between various formats.

Thanks to Small Data and Wide Data, we find data more suitable for extracting more specific and comprehensive knowledge than with Big Data.

Both Small and Wide Data offer multiple opportunities for all those sectors where hyper-personalization is a fundamental variable of the business, such as in online retailing. But it also applies to any product or service that requires technologies facilitating quick decision-making. An example is autonomous feature cars, where fast and intelligent processing through Edge devices is essential to determine actions based on small (but very precise) volumes of data recorded in real time.

Evolution

Since the 1960s and 1970s, Big Data had not had a worthy competitor. Traditionally, machine learning has relied on large centralized data sets, Big Data. However, due to the proliferation of IoT devices and the need for real-time processing, there has been a shift towards incorporating intelligence. That is why **the importance of using smaller and more diverse sets of data alongside new strategic approaches such as Small Data and Wide Data is increasingly being recognized.**

In the coming years, we will see continued experimentation with these models to seek advanced techniques that can overcome these barriers and extract all possible value from the treatment of limited data sets. **By 2025, it is expected that 70% of organizations will have shifted their data analysis focus from Big Data to Small and Wide Data.**

Additionally, the installation of TinyML devices is expected to increase from nearly 2 billion globally in 2022 to over 11 billion in 2027. The main driver of this growth will be the trend to bring machine learning almost everywhere. For example, TinyML algorithms are especially useful in sound and environment sensors, and there are many possible use cases, which will drive the significant increase in installations of devices with TinyML capabilities.

Global adoption of TinyML devices

Millions of devices

2022	2027
2.000	11.000
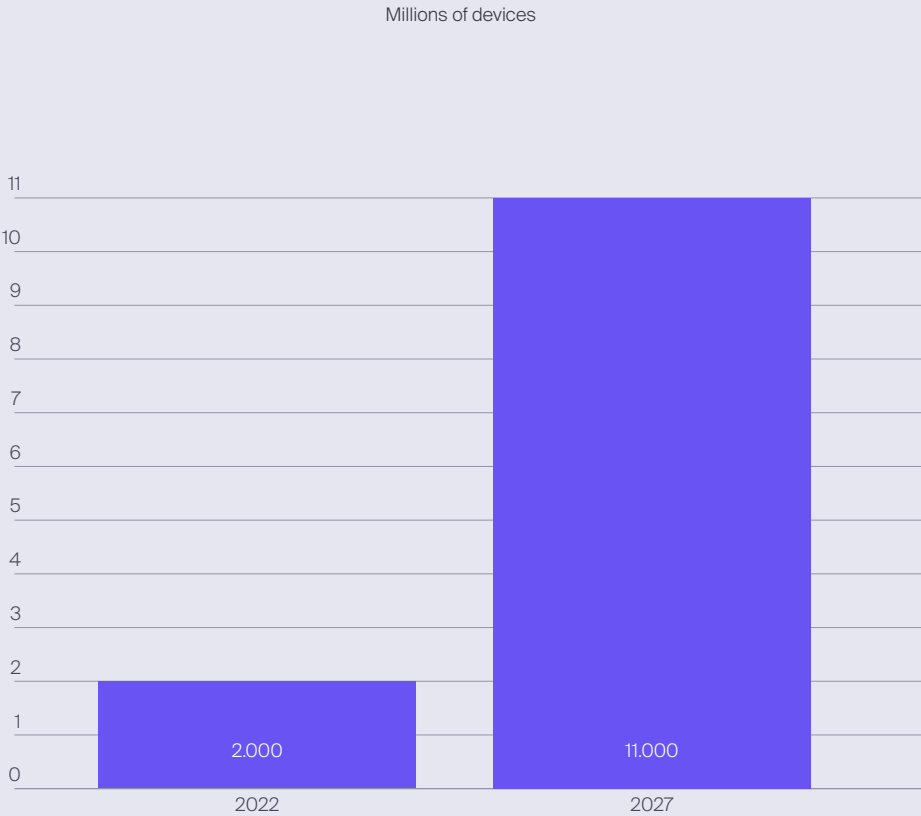
Source: Softtek based on ABI Research data

Technological Background

- **Advances in Hardware:** Progress in the design of ultra-low-power hardware allows for complex calculations on small devices such as microcontrollers.

- **Model Optimization:** Techniques such as model distillation, model pruning, model quantization, or dataset distillation enable the optimization of machine learning to create compact AI models without compromising performance.

- **Data Strategies:** As we generate a large amount of data, strategies to convert it into information proliferate. Depending on the capabilities and needs of organizations, models will be chosen. While Small Data focuses on the use of smaller, but well-selected datasets, Wide Data involves the integration of various types of data sources, requiring sophisticated data fusion techniques.

Challenges

Resource Constraints: TinyML devices often have limited memory and computational power, which is addressed through algorithm and model optimization solutions to reduce their footprint while maintaining accuracy.

Data Quality and Diversity: Ensuring the quality and representativeness of small and wide data is a challenge. Techniques that help overcome this challenge include robust data preprocessing, anomaly detection, and the incorporation of domain knowledge into the analysis.

Real-Time Processing: Both TinyML and Wide Data require real-time or near real-time processing. Efficient Edge Computing architectures and streaming data processing frameworks are key to addressing this need.

Advanced Analytical Tools:

Edge Analytics: Tools designed for Edge Computing can analyze data on the device, reducing latency and bandwidth requirements.

Federated Learning: Is a machine learning technique that trains an algorithm across a decentralized architecture of multiple devices, each holding their own local and private data, preserving privacy and reducing the need for data transfer.

Synthetic Data Generation: When real data are scarce or sensitive, synthetic data become a valuable tool, especially in the context of Small Data. The process of creating artificially generated synthetic data involves a generation or modeling process that must maintain analytical value for use cases.

Trends

Key insight

- **Integration with 5G:** The deployment of 5G networks will enhance the capabilities of TinyML and Wide Data applications, providing faster and more reliable data transfer.

- **Advances in Neuromorphic Computing:** This emerging technology mimics the structures and processes of the human brain, potentially revolutionizing the efficiency of TinyML.

- **More Sophisticated Multi-modal Data Integration:** As Wide Data strategies become more common, tools to integrate and make sense of various types of data will become increasingly sophisticated.

Imagine a smart farming scenario where TinyML is used to monitor and analyze soil and crop conditions in real time. Sensors equipped with TinyML models detect early signs of disease or dehydration, triggering timely interventions. This scenario leverages Small Data (localized and specific data sets) and Wide Data (integrating weather, soil, and crop data) to provide comprehensive insights.

The approaches of **TinyML, and Small and Wide Data are set to transform the way we think about and use data and machine learning.** By enabling intelligent processing on a smaller scale, in more diverse contexts, and closer to where data is generated, they offer **the promise of more personalized, timely, and efficient insights and actions.**

As these fields evolve, they will open up new possibilities for innovation and understanding across virtually every domain.

06. AI Ubicua/ Edge AI

The revolution brought about by generative AI is leading to a new perception of what the technological future of society will be. **The changes experienced at the level of operational work and human-machine interaction allow us to glimpse a new paradigm in which AI will extend to all daily activities of humans,** being present in all the applications and connected devices we use on a daily basis.

The tremendous impact of thinking about this new idea of ubiquitous artificial intelligence, where all our daily activity is driven by intelligent algorithms, makes it possible to appreciate in any environment the importance of structuring data sets, bringing about **the need for new approaches in data matters.**

In this aspect, the concept of **Edge AI is going to become very important, given its ability to allow data processing at the very point of its creation, helping organizations to obtain real-time information by analyzing data as close as possible to the place where it originates.**

The concept of Edge AI encompasses all those deployments of AI algorithms and models carried out directly on Edge devices. That is, developed on the same sensors, cameras, connected devices, and ultimately, any device connected to an IoT network.

EdgeAI global market value

Millions of $

Year	Value
2022	5.000
2032	70.000

24,8% CAGR
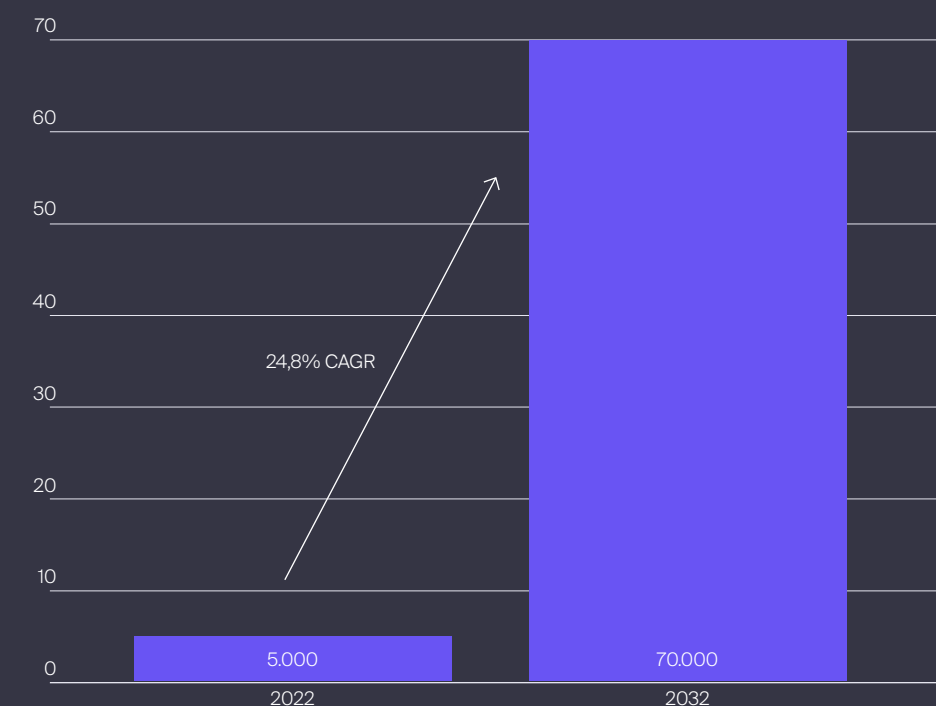
Source: Softtek based on data from Global Market Insights

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:

- **Security systems:** industrial surveillance applications equipped with computer vision or intelligent alarm and intrusion detection systems.

- **Industrial robots:** machinery capable of performing tasks autonomously and adapting to changes in the environment without human intervention.

- **Smart health devices:** connected medical devices that process data in real time allowing for automatic diagnoses and treatments with a high level of precision and personalization.

- **Autonomous vehicles:** features that allow vehicle systems to make decisions with greater precision and agility than human drivers, such as avoiding collisions and changing lanes, improving road safety.

- **Process control systems:** all types of analysis and control tools based on IoT device networks that can be applied for multiple purposes of safety, operational efficiency, sustainability, or productivity optimization.

The main variable that will direct the growth of the Edge AI approach will be the adoption of 5G networks on a massive scale. The latest generation wireless communication networks will provide much faster and more reliable connectivity, reducing latency in the transmission and processing of data generated on edge device networks. Additionally, 5G technology will offer new advanced security features that will facilitate the encrypted data transmission, another crucial aspect for Edge AI applications to be practically implemented in areas where very sensitive data is handled, such as in the health, financial, government, transportation, or defense sectors.

Evolution

Ubiquitous AI refers **to the deployment of data processing algorithms and AI directly on devices at the network edge, rather than in a centralized cloud or data center.** This approach enables real-time data processing, latency reduction, and enhanced privacy and security of sensitive data. It is particularly crucial in scenarios where rapid decision-making is essential, such as autonomous vehicles, healthcare monitoring, and industrial automation.

The journey toward Ubiquitous AI began with the centralization of data and processing power in large data centers and clouds. However, the explosion in the number and variety of smart devices generating data at the network's edge necessitated a shift. The need for real-time processing, bandwidth limitations, and privacy concerns have driven the evolution of edge AI. This evolution is marked by the miniaturization of hardware, advances in AI algorithms, and the development of frameworks and platforms specifically designed for edge computing.

Technological Background

Challenges
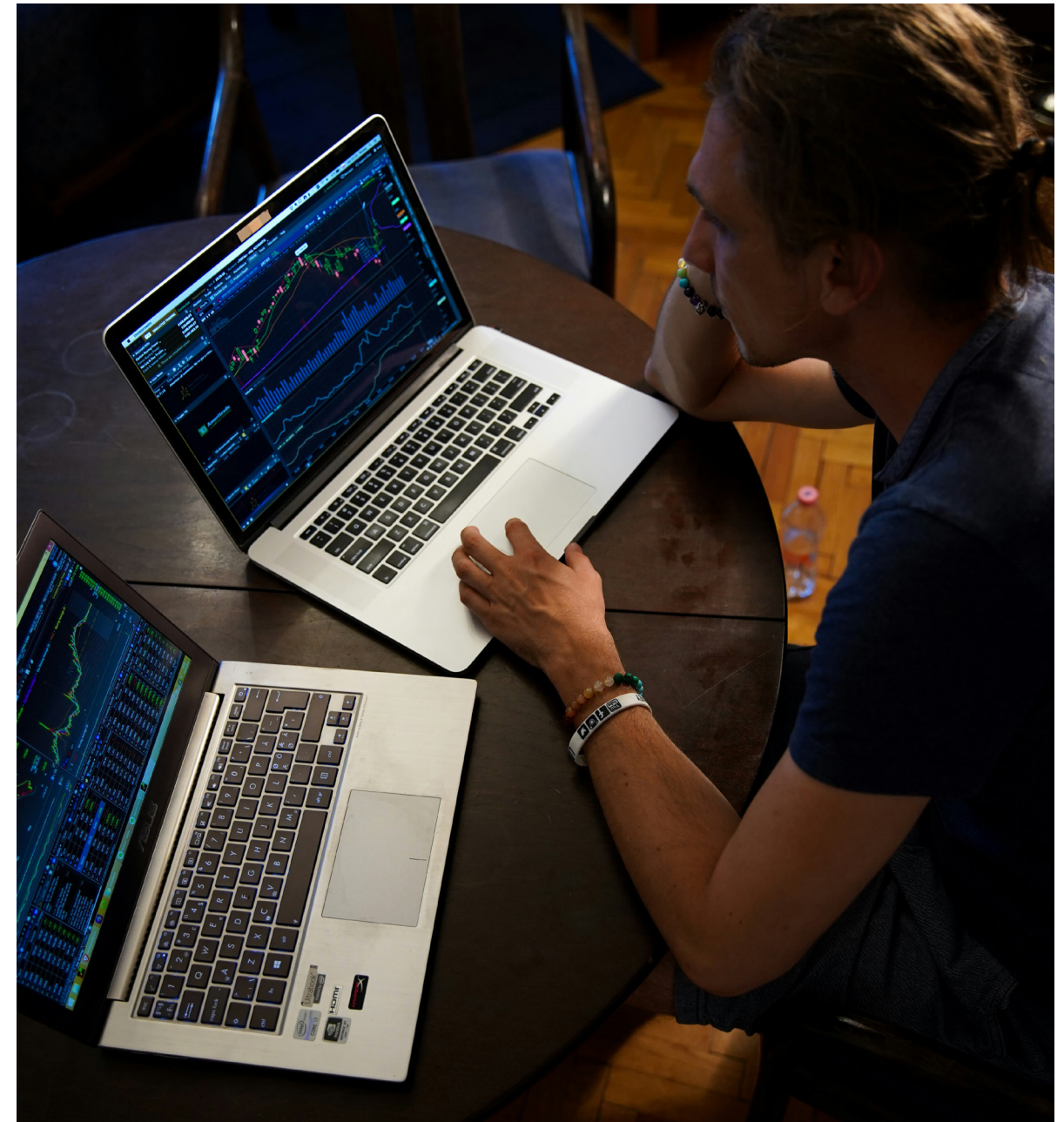
- **Edge Computing Infrastructure:** A processing model that allows data produced by systems to be processed where it is generated, to avoid sending this information to distant Cloud infrastructures, causing latency and security issues. It can be understood as the implementation of small distributed Data Centers. This requires the necessary hardware and network architecture to process data at the edge, with devices ranging from simple sensors to powerful edge servers.

- **Distributed Data Strategies:** This involves managing and processing data across multiple edge devices and possibly synchronizing with centralized systems, requiring solid data management and synchronization techniques.

- **Optimized AI Models:** AI models for the edge are designed or specifically adapted to require less computational power and memory, using techniques such as compression, quantization, and model pruning, as previously mentioned.

Resource Limitations: Edge devices often have limited processing power and memory. Therefore, the solution lies in the optimization of models and artificial intelligence algorithms to improve efficiency or in the consideration of new computing paradigms such as neuromorphic computing.

Data Security and Privacy: To ensure data security at the edge, techniques involving the use of secure enclaves for confidential computations are used, while end-to-end encryption is implemented for data in transit.

Intermittent Connectivity: Edge devices are not always connected to a network, hence, the design must include systems that can operate independently and synchronize with central systems when connectivity is available.

Advanced Analytical Tools in Data AI and Edge AI

- **Stream Processing Frameworks:** Essential for analyzing real-time data at the edge. They can handle high-throughput data streams and provide timely insights.

- **Federated Learning:** An approach to machine learning where the model is trained across many decentralized devices containing local data samples, improving model performance without compromising privacy.

- **Edge Analytics Databases:** These specialized databases are designed to run on edge devices, providing fast data storage and retrieval for real-time analysis.

Trends

- **5G Integration:** Will significantly enhance edge AI capabilities by providing faster and more reliable connectivity for edge devices, enabling more complex and collaborative AI applications.

- **Autonomous Edge Devices:** Will become more autonomous, capable of making decisions and acting on them independently, further reducing the need for central oversight.

- **AI-Powered Edge Security:** The future of security will focus on AI-driven security measures at the edge, providing dynamic and real-time defense mechanisms.

An industrial manufacturing company implemented Edge AI to monitor the condition of equipment in real-time. Sensors on the machinery collect data on performance metrics, which are processed on-site by edge devices using AI models to predict potential failures. This approach allows for immediate intervention, reducing downtime and maintenance costs.

Key insight

Data & Edge AI is transforming how and where we process and analyze data. By bringing AI capabilities closer to where data is generated, **it allows for faster, more efficient, and private data processing.**

As technology continues to advance, the potential applications of Data & Edge AI will expand, offering unprecedented opportunities for innovation across various sectors.

36

07. Data Masking

The need for organizations to protect the integrity, confidentiality, and privacy of data is going to become increasingly complex in the new paradigm of smart businesses, and techniques such as **Data Masking and Dynamic Data Masking will continue to emerge as a trend due to their ability to offer an effective solution for protecting sensitive data during development, testing, and information sharing.**

Furthermore, these techniques can be used to simplify the design and coding of security. By applying Data Masking techniques, organizations can mask or hide information in a database or test environment with the aim of limiting the exposure of sensitive data

to violations, security breaches, or regulatory non-compliance. This means that, although the data remains unchanged in storage, its access is limited and only the minimum amount necessary is revealed to the user.

Data masking can be structured into two main blocks: static and dynamic data masking. **Static Data Masking (SDM) permanently replaces sensitive data by altering data at rest, while Dynamic Data Masking (DDM) aims to replace sensitive data in transit, leaving the original data at rest intact and unchanged.**

Evolution

Data masking is the evolution of encryption techniques. The main difference between the two techniques is that encryption protects data against privacy breaches when it is at rest, but to process, analyze, or manipulate it, decryption is necessary, creating a security gap. **However, with data masking, operations can be performed on the masked data while maintaining its structure and behavior.**

In the coming years, Data Masking will become increasingly relevant, especially in development and testing environments, where production data is used for testing, but it is crucial to protect confidential information.

The most advanced trends in this technology include an increasing adoption of Data Masking solutions hosted in the cloud, with a special focus on applications for regulatory compliance and privacy policies.

Similarly, organizations are in search of real-time Data Masking solutions capable of protecting data during processes of integration, sharing, or migration.

Another trend we will see expanding in the next decade has to do with **the integration of artificial intelligence and Machine Learning algorithms** designed to increase the efficiency and effectiveness of Data Masking techniques, capable of scaling the volume of data handled and reducing the need for supervision.

Global market size of Data Masking tools and solutions

Millions of $

Year	Value
2023	0.820
2032	1.630

14,7% CAGR
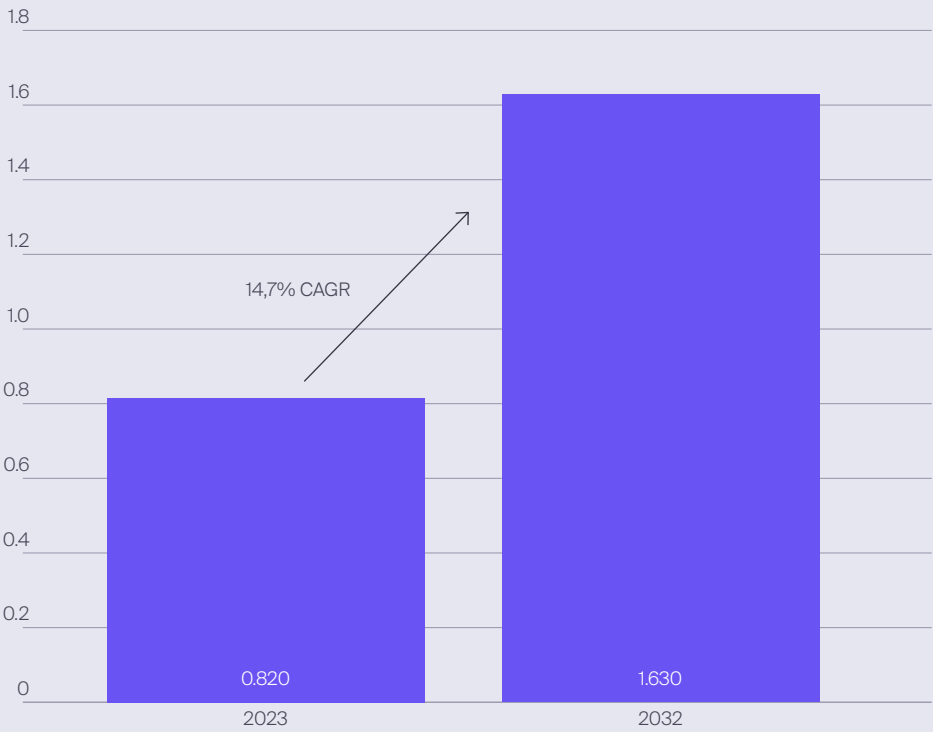
Source: Softtek based on data from Mordor Intelligence

Technological Background

The ultimate goal is to ensure that private data, such as financial credentials, account numbers, tax identifiers, or email addresses, cannot be exposed to unauthorized persons or illicit uses. To this end, it encompasses the use of different techniques:

- **Static Masking:** Involves replacing the original data with fictitious or altered data. For example, hiding the first 12 digits of a credit card number and showing only the last 4 digits.

- **Dynamic Masking:** Applied in real-time, where the original data is replaced with fictitious data each time it is accessed or displayed.

- **Subset Masking:** Involves hiding only certain parts of the data while keeping other parts visible. This can be useful for maintaining data integrity without revealing sensitive information.

- **Feature Masking:** Replaces the original data with similar, but not identical, values. For example, altering a person's age by adding or subtracting a small random number.

- **Format-Preserving Masking:** Maintains the format and appearance of the original data, but the actual data is replaced with fictitious or pseudo-anonymized values. This model is expected to experience the most growth in the near future.

Through these techniques, organizations can create fictitious but realistic representations of company data, developing a new version that is impossible to interpret or decipher through reverse engineering.

Challenges

Implementation Complexity

- **Automated Detection and Masking:** Through tools that automatically identify sensitive data and apply the most appropriate masking rules for each situation, helping to simplify the implementation process.

- **Customizable Masking Techniques:** Allow organizations to choose the method that best suits their specific type of data and use case.

Performance Overhead

- **Optimized Algorithms:** The use of efficient algorithms that minimize computational overhead is vital, especially for dynamic data masking.

- **Selective Masking:** Allows only the necessary data to be masked, significantly reducing the impact on performance.

Maintaining Data Utility

- **Format-Preserving Encryption (FPE):** This method encrypts data in a way that preserves the original data format, ensuring that the masked data can still be useful for certain types of analysis.

- **Synthetic Data Generation:** Involves generating entirely new synthetic data based on the characteristics of the original data. It can be used for testing and development without risking exposure of sensitive information.

Advanced Analytical Tools in Data Masking

- **Context-Aware Masking:** Advanced tools understand the context of the data and apply different masking rules based on the data's confidentiality, the user's role, and the specific scenario in which the data is used.

- **AI and Machine Learning:** These technologies can be used to improve the identification of sensitive data and to create more sophisticated masking techniques that better preserve the utility of the data.

- **Data Masking as a Service (DMaaS):** Some providers offer data masking as a cloud-based service, simplifying the implementation process and providing access to the latest technologies and techniques.

Trends

Key insight

- **Increased Regulation for Data Protection:** As privacy regulations become stricter, the techniques and tools that allow for data masking and protection will evolve to lead solutions that maximize the power of data while complying with current legislation.

- **Integration with Data Fabric:** Integrating data masking capabilities will be crucial to ensure that sensitive data is protected across the data landscape.

- **Advances in Real-Time Masking:** The need for real-time data access in analysis and decision-making will drive advances in dynamic data masking, making it more efficient and less intrusive.
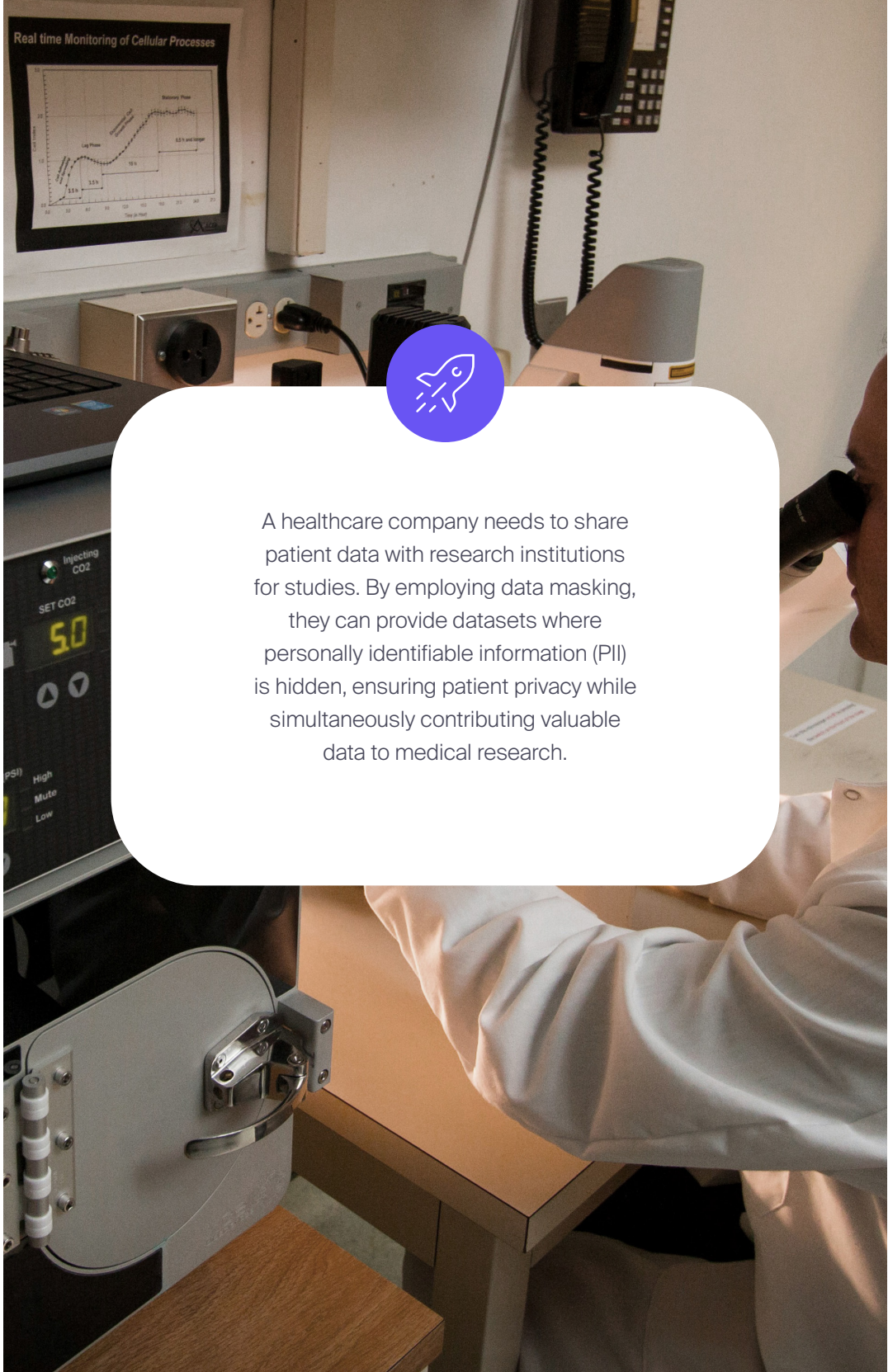
A healthcare company needs to share patient data with research institutions for studies. By employing data masking, they can provide datasets where personally identifiable information (PII) is hidden, ensuring patient privacy while simultaneously contributing valuable data to medical research.

Data masking is an essential strategy for protecting confidential information and maintaining privacy. By hiding specific data elements, organizations can securely share and use data, meeting both operational needs and compliance requirements.

As technology advances, so will techniques and tools for effective data masking, **ensuring that data privacy keeps pace with the increasing reliance on information and data-driven decision-making.**

08. DataOPs

Due to the growing need for agility in data handling, and the increased complexity of system integration for cross-functional collaboration, **DataOps is becoming an increasingly necessary technique for organizations.**

Combining data and operations is proving to be an effective practice for designing and implementing data pipelines in a more efficient and collaborative manner, and it is expected to be increasingly used to improve data delivery, quality, and availability, while accelerating the development lifecycle and operations in analytics environments, optimizing data collection and preparation.

On the other hand, **data observability provides the visibility and insights needed on the data flows generated within an organization**; and DataOps offers the framework and processes to operate, improve, and automate these flows.

In this way, DataOps **is proposed as a complementary methodological approach to DevOps**, aiming to optimize and accelerate the delivery and management of developments based on data systems or applications. Moreover, it facilitates data engineering teams in providing analysts and other stakeholders with reliable data to drive subsequent decision-making.

The great advantage of applying a DataOps approach to data **management is that it significantly promotes collaboration among different functions within the organization**, contributing to the elimination of traditional data silos, preventing development cycles from unnecessarily prolonging, and consequently, keeping projects agile.

Evolution

Originally derived from the principles of agile development, DevOps, and statistical process control, DataOps has become a distinct discipline. It is driven by the need for more dynamic and **real-time data processing and analysis** in an era where data volumes and complexity continue to grow exponentially.

The advances that DataOps techniques will experience in the coming years will go hand in hand with the expansion of artificial intelligence and machine learning algorithms, leading to **next-generation platforms that include predictive analytics features.**

On the other hand, we will see how these new solutions are mostly hosted in the cloud, with the purpose of allowing organizations greater scalability, flexibility, and cost reduction. These **Cloud DataOps platforms** also offer greater compatibility and integration capacity with different formats and data sources in distributed environments.

Interfunctional collaborative DataOps practices will continue to emerge, focused on promoting the existing **synergies between teams of engineers and data scientists and purely operational areas.**

Ultimately, an increasing awareness of the potential benefits of DataOps for optimizing data flows, and the need to reduce response time and improve data quality, will lead to a surge in demand for these advanced solutions.

Global market size of DataOps platforms

Millions of $

22% CAGR

	2022	2032
	3.400	25.000
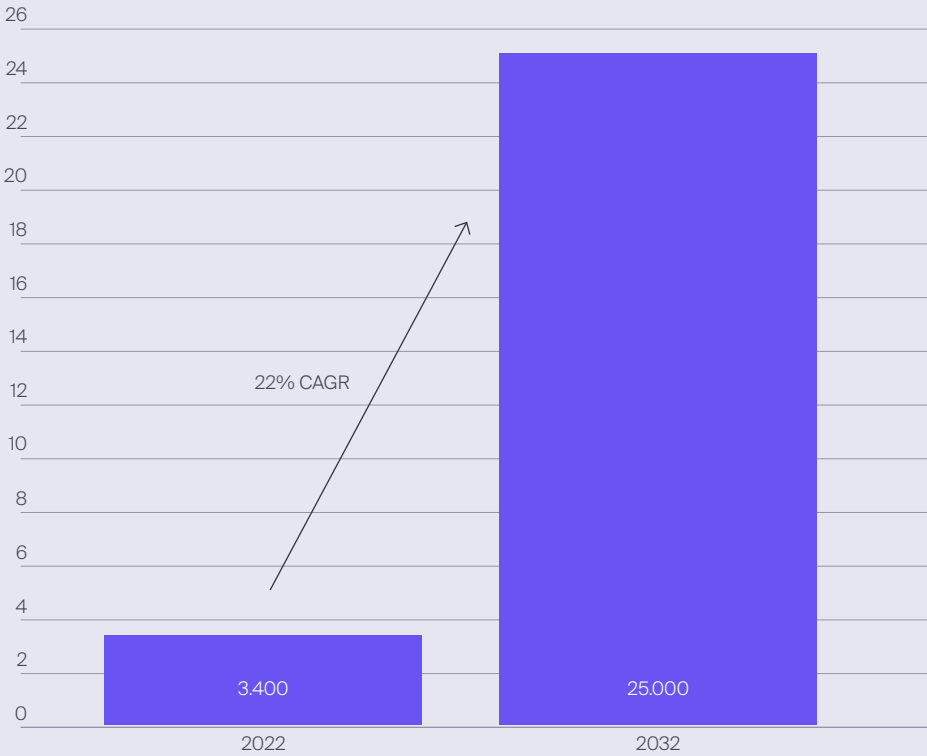
Source: Softtek based on data from Global Market Insights

Technological Background

Automated Data Quality Testing

- **Machine Learning for Anomaly Detection:** Advanced algorithms can automatically detect anomalies and inconsistencies in data, significantly reducing the time and effort required to ensure data quality.

- **Continuous Data Quality Metrics:** Provide ongoing monitoring and reporting of data quality, ensuring that issues are identified and addressed promptly.

Advanced Data Pipeline Orchestration

- **Dynamic Pipeline Adjustment:** Based on real-time data flows and user demand, ensures optimal performance and resource utilization.

- **Self-Healing Data Pipelines:** Innovations in AI allow data pipelines to detect issues and autonomously correct them, minimizing downtime and manual intervention.

Enhanced Collaboration Tools

- **Integrated Development Environments (IDE) for DataOps:** These specialized IDEs offer features tailored to DataOps teams, such as version control, test suites, and collaboration tools, all within a unified environment.

- **ChatOps for DataOps:** The integration of chat platforms with DataOps tools allows teams to collaborate more effectively, with bots and automated alerts keeping everyone informed and engaged.

Challenges

Integration Complexity

- **Standardization of Tools and Processes:** Adopting industry-standard tools and defining clear processes can help reduce the complexity and learning curve associated with integrating new DataOps innovations.

- **Modular Architecture:** Designing data systems and tools in a modular way allows for easier integration and replacement of components as needed.

Maintaining Data Security and Governance

- **Automated Compliance Monitoring:** Automatically monitoring and reporting on data governance policy compliance ensures that innovations do not lead to security or compliance failures.

- **Role-Based Access Controls in Tools:** Ensuring that DataOps tools implement strict access controls will help maintain data security as the number and complexity of data operations grow.

Leading Innovation

- **Continuous Learning and Development:** Fostering a culture of continuous learning within DataOps teams ensures they stay up-to-date with the latest innovations.

- **Partnerships with Technology Providers:** Collaborating closely with DataOps tool and platform providers can help organizations stay at the forefront and influence the development of new features that meet their needs.

Trends

Key insight

- **Integration of Artificial Intelligence and Machine Learning:** As they evolve, they will be more integrated into DataOps, enhancing data processing and management capabilities, leading to next-generation platforms in predictive analytics.

- **DataOps as a Service (DaaS):** Cloud-based DataOps platforms will offer organizations the ability to deploy sophisticated DataOps strategies without the need for extensive internal infrastructure and expertise.

- **Predictive Data Operations:** Leveraging predictive analytics, future DataOps tools will not only react to issues as they occur but will anticipate and prevent them before they happen, further improving the efficiency and reliability of data operations.

A financial services company implements an advanced DataOps strategy to handle real-time market data. By employing machine learning for anomaly detection, they significantly reduce the incidence of data errors that affected their trading algorithms. Automated data pipeline orchestration allows them to dynamically scale their data processing capabilities, ensuring they can always meet the demand for real-time analysis.

Recent innovations in DataOps are transforming the landscape of data management and analysis. By adopting these advances, **organizations can ensure their data operations are as efficient, reliable, and effective as possible**, allowing them to maximize the value of their data assets.

As DataOps continues to evolve, it will undoubtedly play an increasingly crucial role in data-driven decision-making that underpins modern business strategies.

09. The Web3 Revolution and Decentralized Storage

The expansion of Web3 is going to represent a comprehensive transformation of the Internet, moving away from a controlled and monolithic environment to become a secure, **decentralized ecosystem focused on the user.** This paradigm shift will occur as a result of the expansion of Blockchain, DApps (decentralized applications), and the popularization of cryptographic assets as a means of online exchange between individuals or organizations.

The arrival of this new phase in the evolution of the Internet will allow people to take greater control of their digital interactions, eliminating intermediaries and, therefore, preserving the ownership of the data generated.

Evolution

Web3 emerges as a response to these challenges, advocating for a **decentralized Internet in which users control their data.** In this paradigm, decentralized storage is not just an option but a fundamental component that ensures data remains secure and accessible in a distributed manner.

Over the next decade, **the market generated through Web3 is expected to experience exponential growth with a CAGR of 44.9% annually.**

In terms of storage, this trend will also entail a change of enormous dimensions **since individuals or groups can contribute their space to create new storage networks that escape the control of current centralized clouds** (such as AWS or Google Cloud).

This phenomenon will result **in a drastic reduction in data storage costs and the creation of new decentralized market economies,** based on the trade of data or crypto-assets in the form of tokens or NFTs.

Furthermore, the Blockchain technology underpinning Web3 offers the possibility to create immutable and transparent records of any operation or transaction. This not only increases security but is also a very effective method **to ensure traceability of all types of commercial and quality processes.**

For example, the possibility of conducting a complete and reliable tracking of all stages and processes within a product's supply chain, through unalterable smart contracts actionable only under the conformity of all parties involved, will cause a significant change in the way data related to quality certifications, sustainability, materials, etc., are recorded and managed.

On the other hand, Web3 **will promote greater interoperability between applications and platforms**, as it does not rely on proprietary formats. This will facilitate the exchange and integration of data across multiple systems, improving efficiency and collaboration.

At the monetization level, individual users will find more opportunities to directly monetize their data. This participation of people in decentralized economies (DeFi) will bring new competition for brands that currently have the monopoly on large volumes of customer data.

However, Web3 will still need to address significant challenges in terms of **governance and regulatory compliance**, and for many of these models to be practically implemented, it is necessary to close certain **legislative gaps**, such as those that exist, for example, in the financial domain.

Global market volume generated by Web 3.0

Millions of $

	2022	2032
Value	0.400	5.500

44,9% CAGR
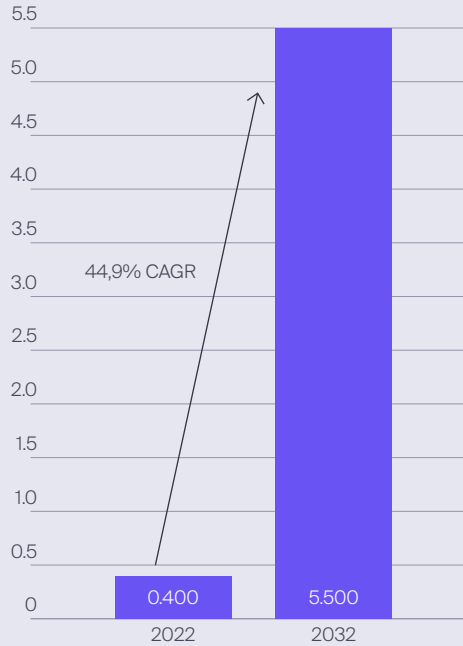
Source: Softtek based on data from Markets & Markets

Technological Background

Blockchain-Based Storage Solutions

- **Immutable Ledgers for Metadata:** Blockchain provides an immutable record of metadata, ensuring the integrity and traceability of data stored across the network.

- **Smart Contracts for Data Management:** These are self-executing contracts with the terms written directly into code, enabling automated, decentralized, and secure data transactions.

InterPlanetary File System (IPFS)

- **Content Addressing:** IPFS uses a block storage model with content addressing, where files are addressed by the hash of their content, ensuring data is uniquely identifiable and resistant to manipulation.

- **Distributed Hash Tables (DHT):** IPFS uses DHT for efficient routing and data location across a distributed network, improving data accessibility and resilience.

Decentralized Identifiers (DID)

- **Self-Sovereign Identity:** DIDs provide a new type of identifier that enables verifiable, self-sovereign digital identities, allowing users to control their data without relying on a central authority.

- **Interoperability:** DIDs are designed to be interoperable across different networks and systems, ensuring users can seamlessly access and manage their data across various Web3 services.
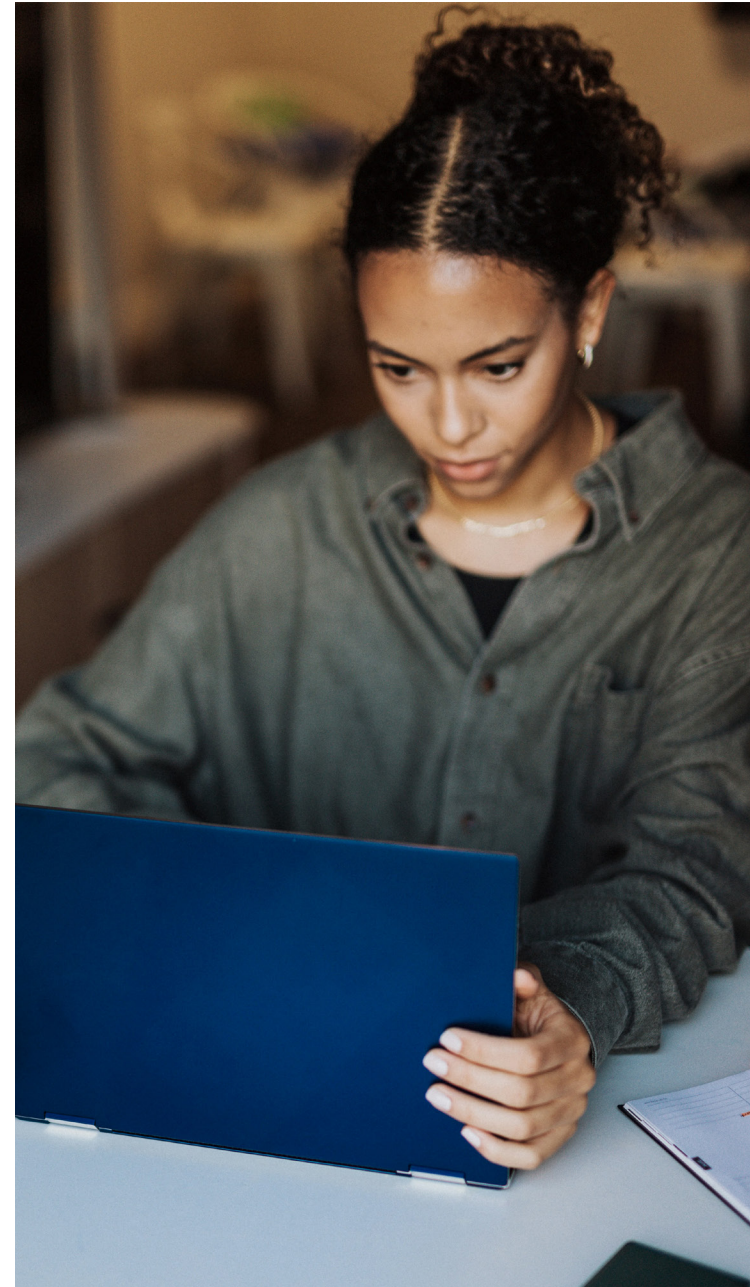
Challenges

Scalability

- **Sharding:** Involves dividing data into smaller parts and distributing them across the network, enabling parallel processing and improving scalability.

- **Layer 2 Solutions:** Technologies like state channels and sidechains can help scale decentralized storage solutions by handling transactions off the main blockchain.

Efficiency in Data Retrieval

- **Caching Mechanisms:** Implementing caching strategies at strategic network nodes ensures frequently accessed data can be quickly retrieved.

- **Incentive Models:** Encouraging nodes to store and serve data efficiently through economic incentives can significantly improve data retrieval times.

Privacy and Security

- **Encryption:** End-to-end encryption ensures data remains private and secure as it is stored and retrieved from the decentralized network.

- **Zero-Knowledge Proofs:** Allow one party to prove to another that a statement is true without revealing any information beyond the validity of the statement itself.

Trends

Key insight
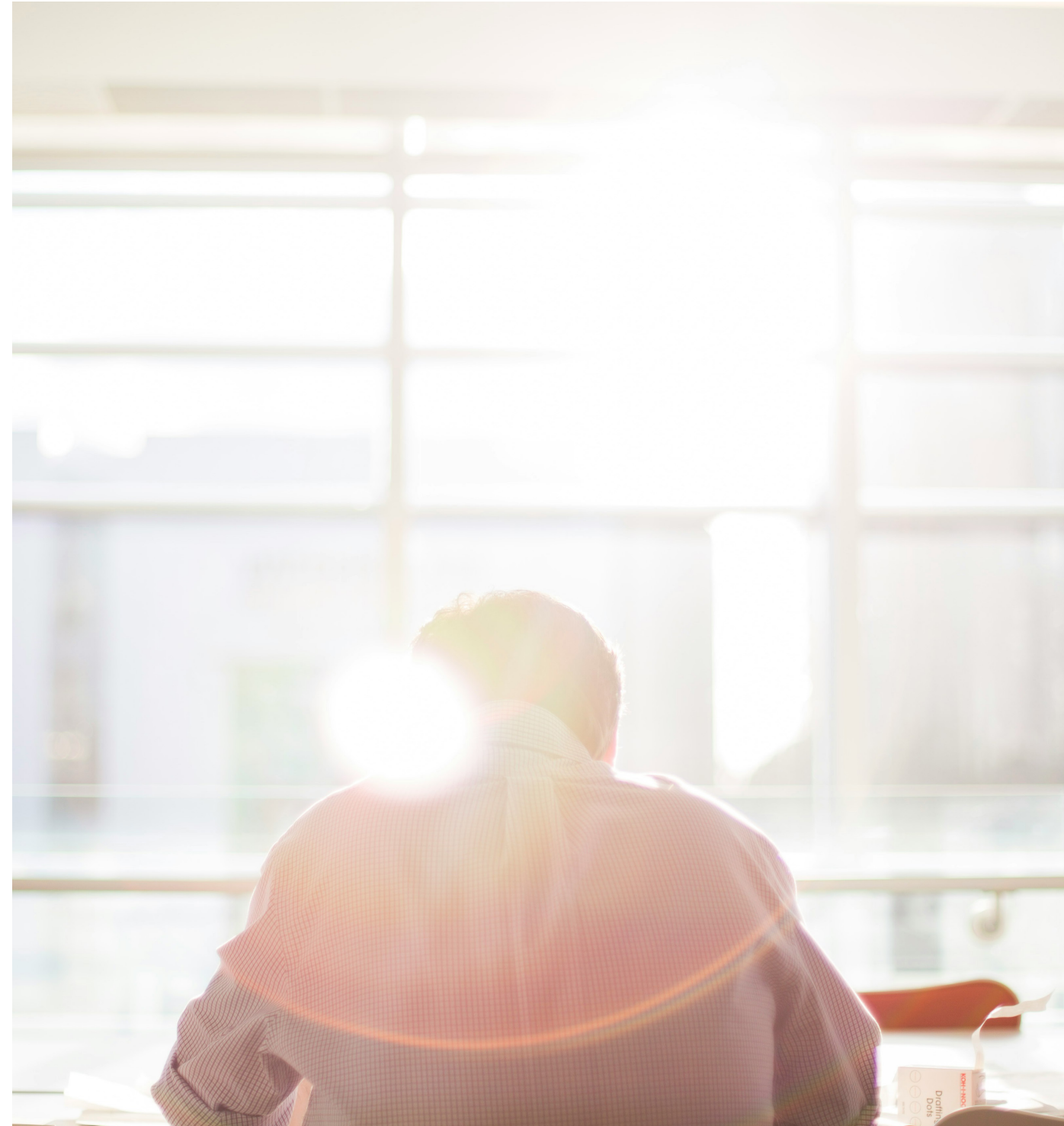
- **Mainstream Adoption:** As tools and platforms become easier to use and practical in addressing regulatory and technical challenges, wider adoption of Web3 and decentralized storage is expected.

- **Integration with AI and Machine Learning:** Decentralized storage will increasingly be used to securely store and share the large amounts of data needed to train AI and machine learning models.

- **Evolution of the Data Economy:** The combination of Web3 and decentralized storage will lead to new data markets and economies, where users have more control over their data and can monetize it securely and transparently.

A social media platform that utilizes decentralized storage would ensure that both posts and personal information are stored securely and that access is restricted to the users themselves or those they grant permission to. This platform would be further strengthened if it also implements Blockchain technology for transparent and tamper-proof governance processes.

Web3 and decentralized storage are laying the foundations for a new era in data management, where **decentralization, user control, and security are paramount.**

These technologies offer a promising solution to the limitations of traditional centralized storage, providing a **more resilient, private, and user-centric data ecosystem.** As the technology matures and adoption grows, we can expect a profound impact on the way data is stored, managed, and utilized on the Internet.

10. Quantum Technologies

Quantum technology is based on two main concepts: entanglement and superposition. **Superposition holds that matter can exist simultaneously in more than one quantum state, while the theory of entanglement posits that particles become dependent on the quantum states of others, behaving as individual entities.** This could translate, in the future, into significant advancements in computing and data management capabilities. **Unlike classical computing based on bits, quantum computing uses quantum bits or qubits, which can represent both 0 and 1 simultaneously.** This ability, along with phenomena such as entanglement and superposition, allows quantum computers to process large amounts of data at unprecedented speeds, offering revolutionary approaches to data management and security.

Quantum mechanics, the underlying principle, has been explored since the early 20th century, but only in recent decades have we begun to harness its principles for computing and data management. **Innovations in quantum algorithms, error correction, and hardware have gradually brought us closer to realizing the potential of quantum technologies in real-world applications.**

Evolution

Although quantum technology applications are still in a stage of development and experimentation, in the coming years we will begin to see the first steps revealing their potential to cause true disruption in data management.

Quantum computers **have the capability to perform much faster calculations**, which offers a window of opportunities to accelerate the processing and analysis of large data sets, **greatly boosting deep learning practices and AI-based predictive analytics.**

On the other hand, **it will also transform cybersecurity standards**, as they offer advanced encryption and security methods that are inherently secure against cyber attacks, which could significantly improve the protection of data stored across all types of systems and networks.

Proof of the promising potential of this new technology is the volume of startups that are already working on experimenting with business models applied to the areas of computing, sensors, and quantum communications:

Quantum technology offers multiple synergies with advancements in IoT network construction and ecosystems of connected sensors, **already allowing for the exponential acceleration of data transmission, reduced latency, and substantially improved quality with which data is processed in real-time.** This revolutionary field will open numerous doors for experimentation in the realm of industrial, environmental, scientific, and surveillance monitoring systems. Another area where quantum technology is applicable is in simulation, 3D modeling, and digital twins, as it will facilitate **the implementation of more**
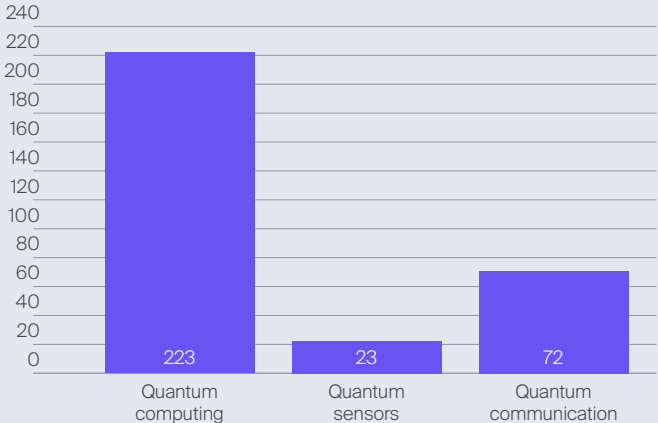
efficient and precise methods for optimizing manufacturing or maintenance processes, opening the door to much more complex and real-time data analysis and processing.

The next two decades will be crucial for the development of this emerging trend, and estimates suggest we will see exponential growth in the market size generated by solutions based on quantum technology.

Number of startups working in the field of quantum technology globally

Total number by segment, in 2022

Quantum computing	Quantum sensors	Quantum communication
223	23	72

Source: Softtek based on Statista data

Estimated global quantum technologies market size, by segment

Millions of $

● 2022
● 2040

Quantum computing	Quantum sensors	Quantum communication
9.000 / 93.000	1.000 / 7.000	1.000 / 6.000

Source: Softtek based on Statista data
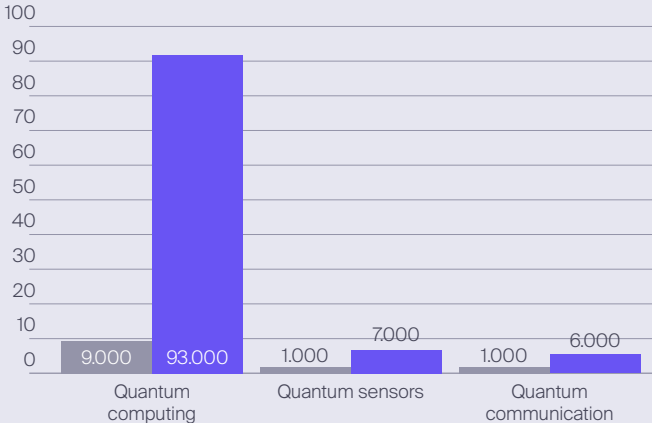
Technological Background

Quantum Key Distribution (QKD)

- **Unbreakable Encryption:** QKD uses the principles of quantum mechanics to create secure communication channels. In theory, it is impossible to intercept or decrypt data encrypted with QKD without alerting the sender and recipient.

- **Early Adoption and Networks:** Although still in its early stages, QKD has been successfully implemented in some specialized communication networks, heralding a new era of secure communication.

Quantum Algorithms for Data Processing

- **Acceleration of Complex Calculations:** Algorithms such as Shor's algorithm for factoring large numbers and Grover's algorithm for database searching promise to complete tasks in a fraction of the time it would take the most powerful classical computers.

- **Optimization of Large-Scale Problems:** Quantum algorithms are particularly suited for optimization problems in logistics, finance, and materials science, where they can analyze and optimize large and complex systems more efficiently than classical algorithms.

Quantum Machine Learning

- **Handling Complex Datasets:** Quantum machine learning algorithms can handle complex and high-dimensional datasets more efficiently than classical algorithms.

- **Enhanced Pattern Recognition:** They may also be capable of identifying patterns and correlations in data that are beyond the reach of classical computers.

Challenges

Technical Complexity

- **Developing Robust Qubits:** Creating stable qubits that can maintain their quantum state (coherence) long enough to perform meaningful calculations is a significant challenge. Ongoing research into various qubit technologies aims to overcome this.

- **Error Correction:** Quantum error correction techniques are being developed to address the errors that inherently arise in quantum computations due to decoherence and other quantum noise.

Integration with Classical Systems

- **Hybrid Systems:** In the short term, quantum computers are likely to be used alongside classical computers, creating hybrid systems that leverage the strengths of each for different tasks.

- **Standardization and Interoperability:** The development of standardized protocols and interfaces for quantum-classical integration is crucial for the practical implementation of quantum technologies.

Security Implications

- **Post-Quantum Cryptography:** The potential for quantum computers to break current encryption methods is driving research into post-quantum cryptography, which aims to develop encryption techniques that quantum computers cannot decrypt.

- **Quantum-Safe Security Practices:** Organizations are beginning to evaluate their cryptographic infrastructure and practices to prepare for the era of quantum computing.

Trends

Key insight

Generalized Quantum Computing: As the technology matures, quantum computers are expected to become more accessible, with cloud-based quantum computing services making them available to organizations that cannot afford their quantum hardware.

Quantum Networks and Quantum Internet: The development of quantum repeaters and satellites will enable the creation of long-range quantum networks, leading to a quantum internet that offers unprecedented levels of security and connectivity.

Revolution in Data Analysis: Quantum technologies are likely to revolutionize data analysis by more effectively handling complex and large-scale datasets, leading to new insights in fields ranging from climate science to healthcare.

A financial institution wants to explore quantum computing to optimize its investment portfolios. Using quantum algorithms, it can analyze a wide range of possible investment strategies and market scenarios in real-time, identifying optimal asset allocations that would be impractical to calculate with classical computers.

Quantum technologies are poised to dramatically transform the landscape of data management and security. With their unprecedented computational power and novel encryption approaches, they promise to address some of the most challenging problems in these fields.

While significant challenges remain, particularly in creating stable and scalable quantum systems, **the potential applications and benefits of quantum technologies make them one of the most exciting areas of development in data management and security.**

Conclusions

In today's fast-paced technological world, staying up-to-date and adapting to emerging data trends is crucial for any organization aspiring to be a leader in its sector.

Efficient data management and utilization have become not just a matter of technology but **a business strategy that can determine an organization's success or failure in the current market.**

Data and analytics leaders, CDOs, and CDAOs, need to lay the strategic foundations and operational models to create a roadmap that allows **the organization to build a data-driven culture, propelling the organization's goals.** Leaders' main concern is to support their organization's decision-making; therefore, they need reliable and high-quality data that, along with new technology, ensures the success of the process.

As organizations transform into data-driven cultures, technology leaders play a crucial role. They must promote **the implementation and use of platforms, tools, and processes that support the data strategy**, while fostering a data-driven culture at both vertical and horizontal levels within the organization.

Data directors and other leaders have a unique opportunity to make a significant impact on their organizations thanks to the rapid evolution of technologies associated with exploiting data value, such as artificial intelligence or machine learning. When applied to the Data techniques analyzed in this paper, they exponentially evolve their capabilities.

To achieve this, organizations must ensure they have solid foundations to support innovation. This includes a confluence between the overall data strategy and the capacity for experimentation and regulatory adaptation. Key approaches to efficient data management include:

1. **Data Democratization:**

    Democratization will allow a greater number of team members to access and understand data, promoting more informed and agile decision-making. For an organization, this means driving innovation from all levels and ensuring all teams are aligned with strategic objectives, fostering true interfunctional collaboration.

2. **Data Monetization (DaaS):**

    Transforming data into a marketable asset can open new revenue sources and business models. Companies adopting this approach will diversify their revenue streams and strengthen their market position.

3. **Cloud Modernization:**

    Migrating data ecosystems to modern cloud solutions allows unprecedented scalability, flexibility, and security. Organizations can leverage cloud capabilities to accelerate innovation and optimize operational costs.

4. **Data Fabric:**

    Building a cohesive and unified data infrastructure is essential to ensure data integrity, accessibility, and security. A well-designed data architecture will facilitate collaboration and improve operational efficiency.

5. TinyML:

Intelligent algorithms based on Small and Wide Data techniques will enable data analysis and processing on Edge devices with limited resources, enhancing IoT applications and high-efficiency embedded systems.

6. EdgeAI:

Bringing artificial intelligence capabilities to the network's edge will enable real-time analysis scenarios and autonomous decision-making systems, crucial for expanding IoT environments and monitoring critical applications.

7. Data Masking:

In a world where data privacy is becoming increasingly paramount, the ability to protect sensitive information without compromising its utility is fundamental. Implementing data masking techniques will ensure regulatory compliance while strengthening consumer trust.

8. DataOps:

This methodology emphasizes collaboration among development, operations, and data teams, promoting automation and continuous integration. Adopting DataOps practices will allow organizations to accelerate the data development lifecycle and improve the quality of the final product.

9. Web3:

With the evolution of the internet into a decentralized and Blockchain-based ecosystem, organizations must be prepared to innovate in areas like smart contracts, cryptocurrencies, or decentralized applications (DApps). Web3 will redefine online interactions and offer new business opportunities associated with data treatment.

10. Quantum Technologies:

Although still in development, quantum computing has the potential to revolutionize how we process and analyze data. Organizations investing in research and development in this field will be at the forefront of technological innovation, finding multiple business generation and performance optimization avenues.

For any organization looking to lead in data matters, the ability to anticipate and capitalize on these innovations will determine not just short-term success but also relevance and competitiveness in the digital future.

Over the next decade, marked by data positioning at the business core, **investing in emerging technologies and adopting best data management practices will be essential to ensure sustainable growth and a leadership position in the market.**